Interventions for promoting evidence-based guideline-consistent surgery in low back pain: a systematic review and meta-analysis of randomised controlled trials

Daniel L. Belavy1 · Scott D. Tagliaferri2 · Paul Buntine3,4 · Tobias Saueressig5 · Katja Ehrenbrusthoff2 · Xiaolong Chen6 · Ashish Diwan6 · Clint T. Miller2 · Patrick J. Owen2

Received: 15 May 2022 / Revised: 9 July 2022 / Accepted: 2 September 2022 / Published online: 17 September 2022
© The Author(s) 2022

Abstract

Purpose Examine the effectiveness of interventions to approach guideline-adherent surgical referrals for low back pain assessed via systematic review and meta-analysis.

Methods Five databases (10 September 2021), Google Scholar, reference lists of relevant systematic reviews were searched and forward and backward citation tracking of included studies were implemented. Randomised controlled/clinical trials in adults with low back pain of interventions to optimise surgery rates or referrals to surgery or secondary referral were included. Bias was assessed using the Cochrane ROB2 tool and evidence certainty via Grading of Recommendations Assessment, Development and Evaluation (GRADE). A random effects meta-analysis with a Paule Mandel estimator plus Hartung–Knapp–Sidik–Jonkman method was used to calculate the odds ratio and 95% confidence interval, respectively.

Results Of 886 records, 6 studies were included (N = 258,329) participants; cluster sizes ranged from 4 to 54. Five studies were rated as low risk of bias and one as having some concerns. Two studies reporting spine surgery referral or rates could only be pooled via combination of p values and gave evidence for a reduction (p = 0.021, Fisher’s method, risk of bias: low). This did not persist with sensitivity analysis (p = 0.053). For secondary referral, meta-analysis revealed a non-significant odds ratio of 1.07 (95% CI [0.55, 2.06], I² = 73.0%, n = 4 studies, Grading of Recommendations Assessment, Development and Evaluation [GRADE] evidence certainty: very low).

Conclusion Few RCTs exist for interventions to improve guideline-adherent spine surgery rates or referral. Clinician education in isolation may not be effective. Future RCTs should consider organisational and/or policy level interventions.

PROSPERO registration CRD42020215137.

Keywords Randomised controlled trial · Meta-analysis · Systematic review · Implementation science · Back pain · Low back pain · Sciatica · Radicular pain · Chronic pain · Controlled before-after studies · Interrupted time series analysis · Surgery · Operative procedures · Fusion · Discectomy · Decompression

Strengths and limitations of this study

- Strength: systematic review and meta-analytic approach.
- Strength: the analytical methods explore the heterogeneity in the main data estimates, by sensitivity analyses to examine outliers, outcome type, and assumptions on ICC values.
- Strength: implemented the most appropriate adjustment for the sample sizes of the included cluster RCTs for meta-analysis and used more efficient meta-analytic methods accounting for the smaller number of included studies.
- Limitation: limited evidence base and the differences of the interventions within included studies.
- Limitation: the outcomes examining secondary (specialist) referral are likely not solely surgical referral.

* Daniel L. Belavy
daniel.belavy@hs-gesundheit.de; belavy@gmail.com

Extended author information available on the last page of the article
Introduction

Low back pain (LBP) is the leading cause of disability worldwide [1]. It is estimated that up to 90% of people will suffer from LBP during their lifetime [2]. The societal cost of LBP is reported to be in excess of $USD100 billion per year in the USA [3], more than $AUD9 billion per year in Australia [4] and approximately €50 billion per year in Germany [5]. The development and resultant adoption of evidence-based clinical guidelines has been shown to reduce costs and may lead to improved patient outcomes: randomised controlled trials have reported either significant [6–8] or non-significant [9–11] reductions in costs favouring guideline adherent approaches. Patient outcomes have been seen to improve in some [10, 12, 13], but not all [14, 15] randomised controlled studies trialling guideline-consistent approaches.

International evidence-based clinical practice guidelines provide recommendations on specialist referral for back pain. The majority (69%) of international evidence-based clinical guidelines recommend specialist referral, including that of a neurosurgeon, in cases of serious radiculopathy, whereas 54% further recommend specialist referral if there is no improvement over time [16]. Conversely, neurosurgical referral is not recommended in patients with non-specific LBP [16]. Recommendations regarding the use of surgery for non-specific and radicular LBP are debated and subsequently 47% of guidelines do not provide any recommendation for or against. Of those that do provide recommendations, half are against the use of surgery for non-specific LBP [16].

Currently, clinical practice guidelines are not always followed. For example, increasing rates of surgical procedures for LBP have been observed internationally [17–19], despite no underlying change in the prevalence of the condition. Further, investigations of referral of patients to secondary specialist care show that these referrals can occur unnecessarily: three prior studies noted unnecessary or inappropriate referral in 10% (14 of 139) [20], 19% (10 of 54) [21], and 41% (12 of 29) [22] of the acute LBP patients included. Beyond the additional cost to healthcare systems of secondary care consultation, surgical management of LBP carries more potential harms than conservative management [18]. Overall, the data suggest that there is potential for cost savings and harm reduction by implementing changes in clinical practice to approach guideline-conform surgery and surgery referral in primary care. To date, no systematic review has assessed this topic.

Our aim was to systematically review and meta-analyse the randomised controlled or clinical trials (RCTs) that examined to the implementation of evidence-based guidelines on surgery rates or surgical/secondary referral for low back pain. In addition, to potentially inform future work, we also collated information on prospective non-randomised interventional studies relevant to the review.

Materials and methods

This review was completed in accordance with Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [23]. The review was registered prospectively with PROSPERO (CRD42020215137). The statistical code, data and extraction sheet are available at https://www.doi.org/10.17605/OSF.IO/APMH2.

Information sources and search strategy

Five online databases (MEDLINE [no limits], EMBASE [excluding MEDLINE], CINAHL [excluding MEDLINE], Web of Science Core Collection [excluding MEDLINE], CENTRAL [excluding MEDLINE, EMBASE and trial registrations]) were electronically searched for research published from database inception to 22nd of June 2020 and then revised on the 10th September 2021. The search terms and search strategy can be found in Supplementary Table 1. To locate additional relevant records, we also searched Google Scholar and included the reference lists of relevant systematic reviews (identified via the Cochrane Database of Systematic Reviews and Google Scholar) published in the last 10 years. The reference lists of included studies were checked for potentially relevant articles. In addition, forward citation tracking of included studies was performed by adding articles that cited the included studies in Web of Science to screening. Furthermore, reference lists (double screened) of studies excluded solely for not being an RCT (e.g. interrupted time series analyses, controlled before and after studies) were also screened for potentially relevant articles.

Study selection

All results of the search were screened to exclude duplicates. Two independent reviewers (KE, XC) who were blinded for each other’s assessment screened the titles and abstracts of the remaining studies against a predetermined eligibility criteria. The full-text reports of articles deemed potentially eligible after this first screening were screened once again by both independent reviewers using the previously mentioned inclusion and exclusion criteria. Any disagreements were adjudicated by SDT and discussed with the project team as necessary.
Eligibility criteria

Inclusion criteria followed the Participants, Interventions, Comparators, Outcomes and Study design (PICOS) framework [23].

Participants: adults (≥ 18 yrs of age) with LBP. See Supplemental Data 1 for further detail on definitions.

Interventions: No limitation was placed on interventions, and these were classified according to procedures [24] used by the Cochrane Effective Practice and Organisation of Care (EPOC) group into professional, financial, organisational, patient oriented, structural or regulatory interventions.

Comparators: were, per Cochrane EPOC procedures [24], no intervention control group, standard practice control group and/or untargeted activity.

Outcomes: Surgical referral (including referral to secondary care if deemed to have surgical component) or surgery rates were included outcomes. See Supplemental Data 1 for further detail on definitions.

Study design: only full-text peer-reviewed journal article (grey literature excluded) with a parallel arm (individual- or cluster-designed) randomised controlled or clinical trial design were eligible.

Data collection and data items

Data extraction was completed by two independent assessors (xxxx). Extracted information included: publication information (i.e. author, title, year, journal), study design, cluster details for cluster randomised trials (e.g. number of clusters), number of participants, participant characteristics (e.g. age and sex), intervention details (e.g. duration, type, frequency), outcome measures (e.g. surgical rates), study funding, and author conflict of interest. Where available, we extracted participants’ pain intensity, disability, and any adverse events from included trials. Extracted data were the number or percentage of surgery use or referrals and the total number of participants or appointment sessions. When data were presented in figures only, data were extracted by generating a screenshot and using ImageJ (version 1.48v https://imagej.nih.gov/ij/) to measure the length (in pixels) of the axes to calibrate, and then the length in pixels of the data points of interest [25]. In all instances where data required for meta-analysis were not available, authors were contacted a minimum of three times over a four-week period to request the information. Similarity between extracted data from the two independent assessors was evaluated through custom spreadsheets set-up in Google Sheets. Any discrepancies were discussed by the assessors with disagreements adjudicated by xxxx.

Risk of bias in individual studies

The Cochrane Collaboration Risk of Bias Tool version 2 [26] was for assessment of study quality. Further detail is provided in Supplemental Data 1. Two assessors (KE, XC) evaluated this independently and any disagreements were adjudicated by SDT.

Certainty of evidence

The assessment of certainty in the body of evidence was conducted in accordance with the Grading of Recommendations Assessment, Development and Evaluation (GRADE) guidelines [27] (see Supplementary Table 2) and performed by one person (SDT).

Statistical analysis

For statistical analysis, we created two categories of comparators: (1) multifaceted intervention and (2) no intervention control group, standard practice control group according to the EPOC guidelines [24]. The primary analysis approach was a random effects meta-analysis of odds ratios (ORs; as they are invariant to baseline risk [28]) with a Paule Mandel estimator for the between study variance $T^2$, an SSW estimator for the overall effect with weights that depended only on the studies’ effective sample sizes, and a 95% CI for the overall effect based on the Hartung–Knapp–Sidik–Jonkman method. We used this method as it outperforms the standard random effects method and other methods [29]. Cluster RCTs were handled by calculating a design effect to correct for clustering of the trials (see Supplemental Data 1 for further detail on this approach). Measures of heterogeneity used were Cochrane Q and the resulting chi-squared statistic and $I^2$. 95%-Prediction intervals were used to assess the amount of heterogeneity if there were at least 10 studies in the meta-analysis [30]. Publication bias (or small study bias) was assessed via funnel plots, Egger’s test, and trim and fill methods if at least 10 studies are included in the meta-analysis [31]. We performed sensitivity analysis via outlier identification and influence analysis [32].

If meta-analysis of effect size estimates was not possible due to missing data, following Cochrane Handbook recommendations [33], we performed a combination of P values. This can be considered where there is no or minimal information reported beyond P values and the direction of effect [33]. We used Fisher’s method to combine the P values and performed a sensitivity analysis with Stouffer’s method [33, 34]. All calculations and graphics were performed with the software R [35] and the extension packages Meta [36], dmetar [37] and metap [38].
Results

Study selection

A PRISMA flow chart of the systematic review process is shown in Fig. 1. There were 886 studies included in title and abstract screening, and 88 were identified as potentially meeting the eligibility criteria and included in the full-text screening. The examination of full texts resulted in 82 studies being excluded (Supplementary Table 3) and six [8, 11, 39–42] studies included (Tables 1 and 2). Of these, six [8, 11, 39–42] studies were eligible for quantitative analysis; two of which examined spine surgery referral rates and four which reported referral to secondary care. One publication required extraction from an image [41]. For one publication [41], the authors were contacted to request data, but the lead author had died and data were no longer available.

![Fig. 1 Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) diagram of the study screening process](image_url)
| Author       | Year Design | Study Design | N INT (baseline, study period) | N CON (baseline, study period) | Type LBP | N Clusters INT | N clusters CON | Cluster type | Int type | INT period | INT | CON | Funding | Study conclusions |
|--------------|-------------|--------------|--------------------------------|--------------------------------|----------|----------------|----------------|--------------|----------|------------|-----|-----|---------|-------------------|
| Cherkin      | 2018 Cluster RCT | Cluster RCT | 2106 (943 baseline, 1163 study period) | 2534 (1061 baseline, 1473 study period) | Mixed | 3 | 3 | Primary Care Clinics | Professional (ed) | 5 months | Six training sessions on the use of the STarT Back Tool and on diagnosis and management of pain | No intervention | None or public/non-profit | No significant difference between INT and CON for spine surgeon visits (neither within group over time nor between groups) |
| Goldberg     | 2001 Cluster RCT | Cluster RCT | 123,829 | 121,881 | Unclear | 5 | 5 | Communities | Professional (ed/pat.ed) | 1 year | Surgeon study groups, continuing medical education, conferences, academic endorsement, videodisc shared decision making, Working with hospital administrators and educating re: impact on profits | No intervention | None or public/non-profit | Significant difference in changes in surgery rates over time in INT vs CON during intervention period compared to baseline period (regression analysis performed) |
| Author    | Year | Study design | N INT | N CON | Type LBP | $N$ Clusters INT | $N$ clusters CON | Cluster type | Int type       | INT period | CON | Funding            | Study conclusions                                      |
|-----------|------|--------------|-------|-------|----------|------------------|-----------------|---------------|---------------|------------|-----|--------------------|-------------------------------------------------------|
| Coombs   | 2021 | Cluster RCT  | 1392  | 3233  | Mixed    | 4**              | 4**             |               |               | 20 months  |     | Multifaceted       | A multifaceted intervention for guideline implementa-    |
|          |      |              |       |       |          |                  |                 |               | Professional  |            |     |                    | tion did not improve specialist consultations in the  |
|          |      |              |       |       |          |                  |                 |               | (edlau.fe)    |            |     |                    | emergency department                                   |
|          |      |              |       |       |          |                  |                 |               | Hospital      |            |     |                    |                                                       |
|          |      |              |       |       |          |                  |                 |               | Emergency     |            |     |                    |                                                       |
|          |      |              |       |       |          |                  |                 |               | Departments   |            |     |                    |                                                       |
|          |      |              |       |       |          |                  |                 |               | Professional  |            |     |                    |                                                       |
|          |      |              |       |       |          |                  |                 |               | (edlau.fe)    |            |     |                    |                                                       |
| Dey      | 2004 | Cluster RCT  | 1049  | 1138  | Acute    | 12               | 11              |               |               | 8 months    |     | Outreach visits    | No significant differences existed between the         |
|          |      |              |       |       |          |                  |                 |               |               |            |     | with interactive  | intervention and control for changing referral to      |
|          |      |              |       |       |          |                  |                 |               |               |            |     | discussion between | secondary care                                        |
|          |      |              |       |       |          |                  |                 |               |               |            |     | the guideline     |                                                        |
|          |      |              |       |       |          |                  |                 |               |               |            |     | team and general   |                                                        |
|          |      |              |       |       |          |                  |                 |               |               |            |     | practitioners on   |                                                        |
|          |      |              |       |       |          |                  |                 |               |               |            |     | guidelines plus    |                                                        |
|          |      |              |       |       |          |                  |                 |               |               |            |     | a poster including  |                                                        |
|          |      |              |       |       |          |                  |                 |               |               |            |     | guideline recom-   |                                                        |
|          |      |              |       |       |          |                  |                 |               |               |            |     | menda-            |                                                        |
|          |      |              |       |       |          |                  |                 |               |               |            |     | tions             |                                                        |
|          |      |              |       |       |          |                  |                 |               |               |            |     |                    |                                                        |
| Author | Year | Study design | N INT | N CON | Type LBP | N Clusters INT | N clusters CON | Cluster type | Int type | INT period | INT | CON | Funding | Study conclusions |
|--------|------|--------------|-------|-------|----------|---------------|----------------|--------------|------------|-----------|-----|-----|---------|------------------|
| Riis [40] | 2016 | Cluster RCT | 539   | 562   | Mixed    | 28            | 26             | Practices    | Professional (edlau. felorg) | 1 year    | Visit from guideline facilitator, education of risk stratification tools and feedback on guideline adherence. Popups in electronic medical records system | Passive Guideline Dissemination | None or public/non-profit | There was a significant reduction in referrals to secondary healthcare in the intervention arm compared to the control |
| Schectman [8] | 2003 | Cluster RCT | 2232 (1014 baseline, 1218 study year) | 1834 (1006 baseline, 828 study year) | Acute Unclear | Unclear | Practices | Professional (edlissiu. fe) | 1 year | 90-min clinical guideline educational session, educational vignettes, audit report on care during baseline year and a copy of the guidelines | No intervention, Patient Education Materials (pamphlet and videotape) | None or public/non-profit | Change from 12 to 8.6% in the "Physician Education and Feedback" group, increase from 5.9 to 7.1% in "Physician Control Group". Unclear if difference was significant |

*Referral to secondary/specialist care was considered to include surgeons/orthopaedics

Specific classification of each intervention is described as 'ed': clinician education/workshops; 'diss': passive dissemination; 'au.fe': audit and feedback to clinicians; 're': reminders to clinicians; 'org': organisational change, administrative changes, electronic medical records system changes and/or government policy change; 'pat.ed': patient education
| Author       | Year | Outcome | Baseline INT | Follow-Up INT | n/N analysis INT | Baseline CON | Follow-Up CON | p value | n/N analysis CON | ICC  |
|-------------|------|---------|--------------|---------------|-----------------|--------------|---------------|---------|-----------------|------|
| Cherkin [39]| 2018 | Surgeon Visits: Number and Proportion | 943 (0.02) | 1163 (0.03) | 0.777 | 1061 (0.02) | 1473 (0.03) | – | – | – |
| Goldberg [41]| 2001 | Mean number of low back surgeries per 100,000 adults | Baseline Period: Year 89: 63.7/100,000 adults; Year 91: 59.0/100,000 adults | Intervention period: Year 92: 60.7/100,000 adults; Year 94: 54.5/100,000 adults | 0.01 | Baseline Period: Year 89: 52.1/100,000 adults; Year 91: 47.2/100,000 adults | Intervention period: Year 92: 44.6/100,000 adults; Year 94: 50.8/100,000 adults | – | – | – |
| Coombs [42]| 2021 | Consultation by a specialist in the emergency department | – | 52 (3.7) | 0.054 | – | 132 (4.1) | 132/3233 | 52/1392 | 0.037 |
| Dey [11]| 2004 | Referral to secondary care: number (%) | – | 36 (3.4%) | 0.12 | – | 26 (2.3%) | 26/1138 | 36/1049 | 0.001 |
| Riis [40]| 2016 | Number of patients referred to secondary care after 12 weeks; presentation of odds ratios | – | 27 (5.0%) | 0.020 | – | 59 (10.5%) | 59/562 | 27/539 | 0.015 |
| Schectman [8]| 2003 | Referral to specialty care: Total utilisation of clinical services Utilisation not consistent with guideline | Baseline Year: Total Utillisation: 12.0%; Utilisation not guideline consistent: 9.5% | Intervention Year: Total utilisation 8.6%, Utilisation not guideline consistent 7.1% | NR | Baseline Year: Total utilisation 5.9%; Utilisation not guideline consistent 4.0% | Intervention Year: Total utilisation 7.1%; Utilisation not guideline consistent 5.6% | 105/1218† | 59/828† | – |

*Referral to secondary/specialist care was considered to include surgeons/orthopaedics
†For Schectman 2003 the percentage of utilisation was multiplied by the total number of participants to get the raw number of secondary care referral. For intervention 0.086×1218 = 104.75; For control 0.071×828 = 58.79
**Study characteristics**

*Population* The sample sizes included in intervention phases of the studies ranged from 1101 to 245,710 and total number of patients included in the review was 258,329. Cluster sizes ranged from 4 through to 54, with the number of clusters in one study [8] being unclear. Attempts to contact authors for further information were not successful. Two studies examined patients with acute LBP, and [8, 11] two studies examined patients where the majority [39] or all [40] had less than three months of pain. In two studies [41, 42], the duration of pain was unclear.

*Interventions* All six studies [8, 11, 39–42] incorporated education and/or workshop component interventions for clinicians, with three studies [8, 40, 42] incorporating audit and/or feedback components, with three studies [11, 40, 42] incorporating some form of passive dissemination of materials to clinicians. One study [40] implemented changes in (electronic) medical records systems to remind clinicians of guideline-based approach. One US-based study [41] further worked with hospital administrators and educating them regarding negative impact on profits of non-adherent approaches and also implemented shared decision-making with patients.

*Comparator* Five studies [8, 11, 39, 41, 42] used a no-intervention control for clinicians and one study [40] used passive dissemination of guidelines. One study implemented patient education materials in a control group [8].

*Outcomes* Two studies [39, 41] reported data on spine surgery consultations [39] or spine surgery rates [41], and four studies reported referral more widely to specialist or secondary care but did not specifically report surgical consultation rates [8, 11, 40, 42]. Three studies reported a significant reduction in surgery rates [41] or secondary referral rates [11, 40], one study [8] reported a change in specialist referral rates (increase in control, decrease in intervention), but it was not significant (Fig. 2). Two studies reported no significant change in spine surgeon visits [39], referral rates [11] or specialist consultations [42].

*Study design* All included studies were cluster randomised controlled trials. Four studies had public or not-for-profit funding [8, 39, 41, 42], and one study [40] received funding from a medical insurer foundation and professional/community organisations and one study [11] did not report funding sources.

**Risk of bias and GRADE assessment**

Five studies [11, 39–42] were rated as low risk of bias and one [8] as having some concerns (Table 3). Considering individual domains, there was low risk of bias for the measurement of outcome, and some concerns in the remaining domains (Table 3). GRADE assessment was not

![Fig. 2](https://example.com/fig2.png)
implemented for the surgical referral or rate outcome as meta-analyses could not be conducted. For secondary referral, the certainty of the evidence was rated for meta-analytic outcomes as very low. Main reasons for downgrading the evidence were study quality, imprecision and inconsistency. Publication bias and/or small study bias could not be assessed because fewer than 10 studies were included [31], but the majority of studies were funded by not-for-profit entities leading us to expect that publication bias is unlikely to impact the evidence.

**Quantitative analysis**

Six studies [8, 11, 39–42] were included in quantitative analysis, two for surgical referral and four for secondary referral. Two studies [39, 41] reported surgery referral or surgery rates and the other studies [8, 11, 40, 42] reported on the outcome secondary referral (which may not exclusively include surgical referral).

For surgical referral or rates, we performed a combination of P values as only Cherkin et al. [39] reported an effect size (adjusted OR: 1.11; 95% CI [0.53, 2.32]) whereas the other study [41] did not. In this analysis, there was evidence for benefit of a reduction of surgery referral in at least one study (p = 0.021, Fisher’s method, n = 2 studies, risk of bias: low; GRADE assessment not possible as meta-analysis unable to be performed). However, sensitivity analysis with Stouffer’s method did not show a significant effect (p = 0.053) when applying an alpha level of 0.05.

For studies reporting the outcome of secondary referral, we performed a meta-analysis with four studies (Fig. 2) [8, 11, 40, 42]. A conservative ICC value of 0.015 was assumed for all studies that had no estimate and for the other studies the published estimate. Overall, meta-analysis revealed a non-significant OR 1.07; [95% CI (0.55, 2.06), I² = 73.0%, 95% CI (23.9%, 90.4%), n = 4 studies, GRADE: very low]. We performed sensitivity analyses on secondary referral (Table 4). First, we checked if there were potential outliers

### Table 3 Risk of bias for cluster randomised trials using the Cochrane ROB2 tool

| Author          | Year | Domain 1a randomisation | Domain 1b timing of randomisation | Domain 2 deviations from interventions | Domain 3 missing outcome data | Domain 4 measurement of outcome | Domain 5 selection of result | Overall |
|-----------------|------|------------------------|----------------------------------|----------------------------------------|-------------------------------|-------------------------------|-------------------------------|---------|
| Cherkin [39]    | 2018 | Low                    | Low                              | Low                                    | Low                           | Low                           | Low                           | Low     |
| Goldberg [41]   | 2004 | Low                    | Low                              | Low                                    | Low                           | Low                           | Low                           | Low     |
| Coombs [42]     | 2021 | Low                    | Low                              | Low                                    | Low                           | Low                           | Low                           | Low     |
| Dey [11]        | 2001 | Low                    | Low                              | Low                                    | Low                           | Low                           | Low                           | Low     |
| Riis [40]       | 2016 | Low                    | Low                              | Low                                    | Low                           | Low                           | Low                           | Low     |
| Schectman [8]   | 2003 | Some concerns          | Some concerns                    | Some concerns                          | Some concerns                | Low                           | Some concerns                | Some concerns |

### Table 4 Sensitivity analyses

| Outcome   | Type of sensitivity analysis | Excluded influential studies | Meta-analytic result of main analysis (OR, 95% CI, I², 95% CI, number of studies) | Result of sensitivity analysis (OR, 95% CI, I², 95% CI, number of studies) | Likely impact on meta-analytic result |
|-----------|------------------------------|-----------------------------|---------------------------------------------------------------------------------|---------------------------------------------------------------------------------|-------------------------------------|
| Secondary referral | Outlier and influential study analysis | Dey et al. [11] Riis et al. [40] | 1.07; [0.55; 2.06] I² = 73.0% [23.9%; 90.4%] N = 4 | 1.13; [0.63; 2.05] I² = 0.0% [NE] N = 2 | No relevant change |
| Secondary referral | Only studies with low ROB | Schectman et al | 1.07; [0.55; 2.06] I² = 34.0% [23.9%; 90.4%] N = 4 | 1.04; [0.43; 2.51] I² = 79.9% [36.4%; 93.7%] N = 3 | No relevant change |
| Secondary referral | ICC = 0.001 for Schectman et al. [8] | None | 1.07; [0.55; 2.06] I² = 73.0% [23.9%; 90.4%] N = 4 | 1.10; [0.60; 2.03] I² = 75.5% [32.3%; 91.1%] N = 4 | No relevant change |

NE not estimable

@Springer
or influential studies and what the impact of the removal of these studies would have on the overall summary effect size. We identified two influential studies [11, 40]. The removal of these studies had a slight effect in favour of the control group OR 1.13; [95% CI (0.63, 2.05), $I^2 = 0\%$, 95% CI (not estimable), $n = 2$ studies]. A sensitivity analysis was done to check if a value of the ICC would change the results which was not the case. One further sensitivity analysis was done by only including trials with a low risk of bias. This did not alter the results (Table 4). A sensitivity analysis for the study [41] where some data extraction occurred from an image is not applicable as the meta-analysis was performed using the $p$-value for this study (which was extracted from the text). We conclude that the results of the main analysis are robust regarding the performed sensitivity analyses.

**Discussion**

This review examined the effects of various interventions on optimising guideline-adherent surgery referral and rates for low back pain patients. To the best of our knowledge, our study is the first systematic review to consider interventions to improve guideline-adherent surgery and surgical referral in back pain patients. All six included studies were cluster RCTs examining clinician level interventions. There was some evidence for impact of the interventions on spine surgery rates or referral (two studies included), but this effect did not persist with sensitivity analysis. In considering referral to secondary care (four studies) we found very low-quality evidence that the interventions had no impact on the outcomes of interest. Individually, two studies [40, 41] showed a significant impact on the outcomes of interest whereas the remainder did not.

**Informing future efforts for guideline-adherent surgical referral for back pain**

Clinician education and/or workshops were a key component of the interventions in all studies [8, 11, 39–41] with two studies [8, 40] implementing a passive dissemination component (i.e. providing clinicians with information on guidelines), and the same two [8, 40] studies incorporated clinical audit and feedback components. Overall, on the basis of the data and results of quantitative analysis, we argue that clinician education is, in isolation, unlikely an effective approach. Notably, the two studies showing significant effects on the number of back surgeries [41] and referral to secondary care [40] also implemented organisational changes. Goldberg et al. [41], a US-based study, also implemented shared decision making with patients and further worked with hospital administrators to educate them regarding negative impact on profits of non-guideline-adherent approaches. The other study [40] implemented changes in (electronic) medical records systems to remind clinicians of guideline-based approach. This provides some suggestions that organisational change interventions may be more effective.

To inform discussion of potentially effective interventions, as part of this systematic review, we also collated and extracted studies that were prospective interventional designs but excluded solely for being non-randomised (Supplementary Table 4). However, this only identified one relevant study. Nonetheless, this controlled before and after study [43] conducted in a region of Denmark, implemented structural changes to patient flow with the introduction of non-surgical multidisciplinary spine clinics prior to patient referral to surgical care. This study showed a 52% reduction in the rates of lumbar disc surgeries per 100,000 inhabitants. This compared to a 12% increase in the remainder of the country. This further highlights the potential of structural and policy change as effective interventions.

**Implementation of behaviour change in clinical practice**

The implementation of such organisational and structural changes across an organisation are important considerations for effective transition to guideline-based behaviours. Behaviour change approaches for implementing evidence-based practice rely on the modification of the behaviour of all parties within an organisational structure. This includes clinicians, managers and administrators and requires an understanding of the determinants of current and desired behaviours [44]. Behaviour change is not a simple process and requires the application of evidence-based behaviour change principles to enhance efficacy [45]. It is rare for a single approach to be adequate; interventions typically require multiple elements that focus on different aspects of current behaviour. Future work should consider adopting a framework (such as the Theoretical Domains Framework [46]) to guide the design of interventions which consider the cognitive, affective, social, and environmental influences of behaviour within the context of the modified behaviours that are intended to be adopted [46, 47].

**Evidence gaps and recommendations for future primary research**

The findings of this systematic review highlight important considerations for future work. Pragmatically, randomised controlled cluster trials are likely the most feasible high-quality study design to implement in an organisational setting, as opposed to trials where patients are randomised to different interventions. Controlled before-and-after and interrupted time series designs are often easier to perform in
an organisational setting, but randomised designs are necessary to provide high-quality evidence for guiding practice in the future and informing guidelines. Importantly, both of these study designs allow opportunities for the thorough consideration of organisational wide behavioural change factors and transparent reporting of design principles. Based on our current findings, we recommend implementing future randomised controlled cluster trials of interventions that include key components of organisational (e.g. structured patient flow, changes to referral requirements and/or changes to electronic medical records systems) and ideally policy (e.g. funding model) change. Beyond this, studies evaluating the effect of organisational behavioural change on guideline adherence should report surgery rates and referral for surgical consultation. Of the six studies identified in this systematic review, two studies [39, 41] did not report these important outcomes. As recommended in Cochrane guidelines [48], future studies should report the total number of surgeries or referrals relative to the total number of appointments or patients seen to allow for appropriate meta-analysis. Furthermore, future cluster randomised controlled trials should always [48] report the intra-cluster correlation co-efficient (ICC) for their study. The ICC is a measure of how similar patients within the clusters are [49], and this information is necessary [48] for appropriate meta-analysis of cluster RCTs.

**Strengthens and limitations of the evidence included and of review processes**

Strengths of the present study include the systematic review and meta-analytic approach applied. Further, the analytical methods explore the heterogeneity in the main data estimates, by sensitivity analyses to examine outliers, outcome type, and assumptions on ICC values. We implemented the most appropriate adjustment for the sample sizes of the included cluster RCTs for meta-analysis [48]. Finally, we used more efficient [29] (i.e. better coverage probability and less bias estimating the between study variance Tau) meta-analytic methods. A strength of the included evidence is that they stem from randomised controlled trials, the highest level of primary evidence for the research question.

Limitations include the lack of eligible studies published to date for inclusion and the differences of the interventions within included studies. The limited pool of studies prevented further investigation of sub-groups and heterogeneity, for example via meta-regression. Of the included studies, some did not report the data in a way which enabled meta-analysis of odds ratios and only pooling of p values could be performed. This limits the quality and utility of the evidence we were able to glean from the literature. It was not possible to provide a GRADE assessment for when meta-analysis was based on combination of p values. The GRADE approach is based on estimated effect sizes [27] and given meta-analysis was not performed, it is impossible to rate p values with this framework. Due to the low number of studies, we could not provide a prediction interval for the meta-analysis as they do not provide a useful estimate of the amount of heterogeneity if the included number of studies is low (< 10) [50]. A subgroup analyses on the types of interventions were not possible and might explain why few showed beneficial impact of the intervention on the outcome of interest. Further, the outcomes examining secondary (specialist) referral are likely not solely surgical referral. Whilst we argue the majority will be referral to surgical care, this nonetheless limits our ability to comment on surgical referral rates. In terms of review processes, we did not include grey literature (e.g. conference abstracts) in our review; Cochrane guidelines [33] regard this as highly desirable but not mandatory.

**Conclusion**

There is a limited evidence base for interventions that improve guideline-adherent surgery rates or referral in back pain. Based on the current evidence, it is not possible to provide a clear recommendation for best-practice approach. Clinician education may, in isolation, be less effective and organisational and/or policy change interventions may be more effective. An evidence gap for future work is for RCTs to ideally assess these intervention components separately to better inform implementation efforts. We also provide recommendations for improving the reporting of cluster RCTs in the future.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s00586-022-07378-6.

**Acknowledgements** Nil.

**Author contributions** DLB and PJO contributed to conceptualization. SDT contributed to data curation. TS contributed to formal analysis. DLB contributed to funding acquisition. SDT, KE, XC contributed to methodology. SDT, PJO, DLB contributed to project administration. DLB, PJO, PB, AD contributed to resources. SDT contributed to software. DLB, PJO, CTM, PB, AD supervised the study. NA contributed to validation. TS contributed to visualization. DLB, SDT, TS contributed to writing–original draft. All contributed to writing–review and editing. All approved the final manuscript.

**Funding** Open Access funding enabled and organized by Projekt DEAL. This work was supported by internal institutional funding only. Scott Tagliaferri is supported by Australian Government Research Training Program (RTP) scholarship. Xiaolong Chen is supported by a RTP scholarship and a University Postgraduate Award from the Australian Government and UNSW.

**Availability of data and materials** Data underlying this study are available within the manuscript.
Declarations

Conflict of interest The authors have no related conflicts of interest to declare.

Ethical approval This study was a systematic review of publicly available data in prior publications and ethical approval is not required.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

References
1. Hartvigsen J, Hancock MJ, Kongsted A et al (2018) What low back pain is and why we need to pay attention. Lancet 391:2356–2367. https://doi.org/10.1016/S0140-6736(18)30480-X
2. Australian Institute of Health and Welfare (2020) Back problems. AIHW, Canberra
3. Martin BI, Deyo RA, Mirza SK et al (2008) Expenditures and health status among adults with back and neck problems. JAMA 299:656–664. https://doi.org/10.1001/jama.299.6.656
4. Walker BF, Muller R, Grant WD (2003) Low back pain in Australian adults: the economic burden. Asia-Pac J Public Health 15:79–87. https://doi.org/10.1177/101053950301500020
5. Wenig CM, Schmidt CO, Kohlmann T, Schweikert B (2009) Costs of back pain in Germany. Eur J Pain Lond Engl 13:280–286. https://doi.org/10.1016/j.ejpain.2008.04.005
6. Eccles M, Steen N, Grimshaw J et al (2001) Effect of audit and feedback, and reminder messages on primary-care radiology referrals: a randomised trial. Lancet 357:1406–1409. https://doi.org/10.1016/S0140-6736(00)04564-5
7. Kerry S, Oakeshott P, Dundas D, Williams J (2000) Influence of postal distribution of The Royal College of Radiologists’ guidelines, together with feedback on radiological referral rates, on X-ray referrals from general practice: a randomized controlled trial. Fam Pract 17:46–52. https://doi.org/10.1093/fampra/17.1.46
8. Schectman JM, Schroth WS, Verme D, Voss JD (2003) Randomized controlled trial of education and feedback for implementation of guidelines for acute low back pain. J Gen Intern Med 18:773–780. https://doi.org/10.1046/j.1525-1497.2003.10205.x
9. Mortimer D, French SD, McKenzie JE et al (2013) Economic evaluation of active implementation versus guideline dissemination for evidence-based care of acute low-back pain in a general practice setting. PLoS ONE 8:e75647. https://doi.org/10.1371/journal.pone.0075647
10. Becker A, Held H, Redaelli M et al (2012) Implementation of a guideline for low back pain management in primary care: a cost-effectiveness analysis. Spine 37:701–710. https://doi.org/10.1097/BRS.0b013e31822b01bd
11. Dey P, Simpson CWR, Collins SI et al (2004) Implementation of RCGP guidelines for acute low back pain: a cluster randomised controlled trial. Br J Gen Pract J R Coll Gen Pract 54:33–37
12. Becker A, Leonhardt C, Kochen MM et al (2008) Effects of two guideline implementation strategies on patient outcomes in primary care: a cluster randomized controlled trial. Spine 33:473–480. https://doi.org/10.1097/BRS.0b013e3181657ed
13. Bishop PB, Quon JA, Fisher CG, Dvorak MFS (2010) The Chiropractic Hospital-based Interventions Research Outcomes (CHIRO) study: a randomized controlled trial on the effectiveness of clinical practice guidelines in the medical and chiropractic management of patients with acute mechanical low back pain. Spine J 10:1055–1064. https://doi.org/10.1016/j.spine.2010.08.019
14. Scheel JB, Hagen KB, Herrin J et al (2002) Blind faith? The effects of promoting active sick leave for back pain patients: a cluster-randomized controlled trial. Spine 27:2734–2740. https://doi.org/10.1097/00007632-200221000-00014
15. Verbeek JH, van der Weide WE, van Dijk FJ (2002) Early occupational health management of patients with back pain: A randomised controlled trial. Spine 27:1844–1850. https://doi.org/10.1097/00007632-200209010-00006
16. Oliveira CB, Maher CG, Pinto RZ et al (2018) Clinical practice guidelines for the management of non-specific low back pain in primary care: an updated overview. Eur Spine J 27:2791–2803. https://doi.org/10.1007/s00586-018-6573-2
17. Xu Y, Yen D, Whitehead M et al (2019) Use of instrumented lumbar spinal surgery for degenerative conditions: trends and costs over time in Ontario, Canada. Can J Surg J Can Chir 62:393–401. https://doi.org/10.1503/cjs.017016
18. Atkinson L, Zacet J (2016) Surgical management of low back pain. Med J Aust 204:299–300. https://doi.org/10.5694/mja16.00038
19. Petzold T, Tesch F, Adler J-B et al (2018) 10-Jahres-Entwicklung operativer Eingriffe an der Wirbelsäule in Deutschland: Eine analyse von Krankenkassendaten im Zeitraum von 2005 bis 2014. Z Für Orthop Unfallchirurgie 156:298–305. https://doi.org/10.1055/s-0043-124768
20. Bishop P, Wing P (2003) Compliance with clinical practice guidelines in family physicians managing worker’s compensation board patients with acute lower back pain. Spine J 3:442–450. https://doi.org/10.1016/S1529-9430(03)00152-9
21. Fullen BM, Maher T, Bury G et al (2007) Adherence of Irish general practitioners to European guidelines for acute low back pain: a prospective pilot study. Eur J Pain 11:614–623. https://doi.org/10.1016/j.ejpain.2006.09.007
22. González-Urzelai V, Palacio-Elua L, López-de-Munain J (2003) Routine primary care management of acute low back pain: adherence to clinical guidelines. Eur Spine J 12:589–594. https://doi.org/10.1007/s00586-003-0567-2
23. Liberati A, Altman DG, Tetzlaff J et al (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol 62:e1-34. https://doi.org/10.1016/j.jclinepi.2009.06.006
24. McAuley L, Ramsey C (2002) Cochrane effective practice and organisation of care review group; data collection checklist. Institute of Population Health, University of Ottawa, Ottawa
25. Vucic K, Jelicic Kadic A, Puljak L (2015) Survey of Cochrane protocols found methods for data extraction from figures not mentioned or unclear. J Clin Epidemiol 68:1161–1164. https://doi.org/10.1016/j.jclinepi.2014.11.016
26. Sterne JAC, Savovic J, Page MJ et al (2019) RoB2: a revised tool for assessing risk of bias in randomised trials. BMJ 366:14898. https://doi.org/10.1136/bmj.j4898

European Spine Journal (2022) 31:2851–2865

2863
27. Guyatt G, Oxman AD, Akl EA et al (2011) GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 64:383–394. https://doi.org/10.1016/j.jclinepi.2010.04.026
28. Doi SA, Furuya-Kanamori L, Xu C et al (2020) Questionable utility of the relative risk in clinical research: a call for change to practice. J Clin Epidemiol 142:271–279. https://doi.org/10.1016/j.jclinepi.2020.08.019
29. Bakbergenuly I, Hoaglin DC, Kulinskaya E (2020) Methods for estimating between-study variance and overall effect in meta-analysis of odds ratios. Res Synth Methods 11:426–442
30. Borenstein M (2019) Common mistakes in meta-analysis: and how to avoid them. Biostat Inc, Englewood, NJ, USA
31. Mavridis D, Salanti G (2014) How to assess publication bias: funnel plot, trim-and-fill method and selection models. Evid Based Ment Health 17:30. https://doi.org/10.1136/eb-2013-101699
32. Viechtbauer W, Cheung MW-L (2010) Outlier and influence diagnostics for meta-analysis. Res Synth Methods 1:112–125. https://doi.org/10.1002/jrsm.11
33. Higgins JPT, Thomas J, Chandler J (2019) Cochrane handbook for systematic reviews of interventions. Wiley, Hoboken
34. Introduction to Meta-Analysis | Wiley Online Books. https://doi.org/10.1002/9780470743386. Accessed 27 Jan 2021
35. R Core Team (2017) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
36. Balduzzi S, Rücker G, Schwarzer G (2019) How to perform a meta-analysis with R: a practical tutorial. Evid Based Ment Health 22:153–160. https://doi.org/10.1136/ebmental-2019-300117
37. Harrer M, Cuijpers P, Furukawa T, Ebert DD (2019) Citation and authors. In: Metar companion R package guide doing meta-analysis. R R Package Version 009000 Httpdmetarprotectlaborg. https://dmetar.protectlab.org/authors.html. Accessed 11 Dec 2020
38. Dewey M (2020) metap: meta-analysis of significance values. R package version 1.3
39. Cherkin D, Balderson B, Wellman R et al (2018) Effect of low back pain risk-stratification strategy on patient outcomes and care processes: the MATCH randomized trial in primary care. J Gen Intern Med 33:1324–1336. https://doi.org/10.1007/s11606-018-4468-9
40. Riis A, Jensen CE, Bro F et al (2016) A multifaceted implementation strategy versus passive implementation of low back pain guidelines in general practice: a cluster randomised controlled trial. Implement Sci 11:143. https://doi.org/10.1186/s13012-016-0509-0
41. Goldberg HI, Deyo RA, Taylor VM et al (2001) Can evidence change the rate of back surgery? A randomized trial of community-based education. Eff Clin Pract ECP 4:95–104
42. Coombs DM, Machado GC, Richards B et al (2021) Effectiveness of a multifaceted intervention to improve emergency department care of low back pain: a stepped-wedge, cluster-randomised trial. BMJ Qual Saf 30:825–835. https://doi.org/10.1136/bmjqs-2020-012337
43. Rasmussen C, Nielsen GL, Hansen VK et al (2005) Rates of lumbar disc surgery before and after implementation of multidisciplinary nonsurgical spine clinics. Spine 30:2469–2473. https://doi.org/10.1097/01.brs.0000184686.70838.96
44. Grol R, Grimshaw J (2003) From best evidence to best practice: effective implementation of change in patients’ care. Lancet 362:1225–1230. https://doi.org/10.1016/S0140-6736(03)14546-1
45. Abraham C, Kelly MP, West R, Michie S (2009) The UK national institute for health and clinical excellence public health guidance on behaviour change: a brief introduction. Psychol Health Med 14:1–8. https://doi.org/10.1080/13548500802537903
46. Atkins L, Francis J, Islam R et al (2017) A guide to using the theoretical domains framework of behaviour change to investigate implementation problems. Implement Sci 12:77. https://doi.org/10.1186/s13012-017-0605-9
47. French SD, Green SE, O’Connor DA et al (2012) Developing theory-informed behaviour change interventions to implement evidence into practice: a systematic approach using the theoretical domains framework. Implement Sci 7:38. https://doi.org/10.1186/1748-5908-7-38
48. Higgins JPT, Thomas J, Chandler J et al (2020) 23.1.3 Methods of analysis for cluster-randomized trials. In: Cochrane handbook for systematic reviews of interventions. Wiley, Hoboken
49. Killip S, Mahlout Z, Pearce K (2004) What is an intracluster correlation coefficient? Crucial concepts for primary care researchers. Ann Fam Med 2:204–208. https://doi.org/10.1370/afm.141
50. Hedges LV, Vevea JL (1998) Fixed- and random-effects models in meta-analysis. Psychol Methods 3:486–504. https://doi.org/10.1037/1082-989X.3.4.486
51. Bardin LD, King P, Maher CG (2017) Diagnostic triage for low back pain: a practical approach for primary care. Med J Aust 206:268–273. https://doi.org/10.5694/mja16.00828
52. Clancy C, Quinn A, Wilson F (2017) The aetiologies of Failed Back Surgery Syndrome: A systematic review. J Back Musculoskelet Rehabil 30:395–402. https://doi.org/10.3233/BMR-150318
53. Nikolakopoulou A, Higgins JPT, Papakonstantinou T et al (2020) CINeMA: An approach for assessing confidence in the results of a network meta-analysis. PLOS Med 17:e1003082. https://doi.org/10.1371/journal.pmed.1003082
54. Guyatt GH, Oxman AD, Montori V et al (2011) GRADE guidelines: 5. Rating the quality of evidence—publication bias. J Clin Epidemiol 64:1277–1282

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Authors and Affiliations

Daniel L. Belavy¹ · Scott D. Tagliaferri² · Paul Buntine³,⁴ · Tobias Saueressig⁵ · Katja Ehrenbrusthoff² · Xiaolong Chen⁶ · Ashish Diwan⁶ · Clint T. Miller² · Patrick J. Owen²

Scott D. Tagliaferri
scott.tagliaferri@deakin.edu.au

Paul Buntine
paul.buntine@easternhealth.org.au

Tobias Saueressig
t.saueressig@physiomeetsscience.com

Katja Ehrenbrusthoff
katja.ehrenbrusthoff@hs-gesundheit.de

Xiaolong Chen
xiaolong.chen1@unsw.edu.au

Ashish Diwan
a.diwan@spine-service.org

Clint T. Miller
c.miller@deakin.edu.au

Patrick J. Owen
p.owen@deakin.edu.au

¹ Division of Physiotherapy, Department of Applied Health Sciences, Hochschule Für Gesundheit (University of Applied Sciences), Gesundheitscampus 6-8, 44801 Bochum, Germany

² School of Exercise and Nutrition Sciences, Institute for Physical Activity and Nutrition, Deakin University, Geelong, Australia

³ Emergency Department, Box Hill Hospital, Eastern Health, Melbourne, VIC, Australia

⁴ Eastern Health Clinical School, Monash University, Melbourne, VIC, Australia

⁵ Physio Meets Science GmbH, 69181 Leimen, Germany

⁶ Department of Orthopaedic Surgery, Spine Service, St. George Hospital, University of New South Wales, Sydney, Australia