Exploring Local Literature Bias: A Critical Evaluation of Iranian Trials on the Efficacy of Psychotherapy for Chronic Pain

Ali Zia-Tohidi 1,2*, Zahra Shamshiri 1, Fatemeh Askari 1,2, Manijeh Firoozi 1,2

1 Faculty of Management and Accounting, Farabi Campus, University of Tehran, Qom, Iran
2 Department of Psychology, Faculty of Psychology and Education, University of Tehran, Tehran, Iran

*Corresponding Author: Ali Zia-Tohidi, Ph.D., Faculty of Management and Accounting, Farabi Campus, University of Tehran, Qom, Iran. Tel: + 98-9109615720, Email: towhidi@ut.ac.ir

Received January 16, 2023; Accepted February 5, 2023; Online Published March 15, 2023

1. Background
Before the 1960s, Chronic pain (CP) was mostly regarded as purely physiological. A large body of evidence has since emerged leading to a consensus of experts on a biopsychosocial perspective. This shift has put psychological interventions at the core of CP management, with hundreds of Randomized Clinical Trials (RCTs) supporting these interventions.

However, current empirical evidence on psychotherapy for CP is mostly from American and European studies, which may not be generalizable to culturally distinct populations (e.g. Middle Eastern societies such as Iran), especially considering the psychosocial aspects of CP and the culturally sensitive nature of psychotherapy. This highlights the need for local evidence on efficacy.

In Iran, dozens of RCTs have evaluated the efficacy of psychological interventions in CP. Within the last five years, several meta-analyses had summarized this evidence, reporting effect sizes that were far larger than non-Iranian estimates. On the surface, this may seem promising, but we doubt whether such large differences are attributable to cultural factors.

1.1. Is the Difference Genuine or a Function of Bias?
The difference between the estimates from Iranian and non-Iranian studies can be genuine. Iranian CP patients may benefit more from psychological interventions, due to some characteristics in patients, therapists, and/or environment. However, there is an alternative explanation: the difference may be a function of bias in the estimates of the efficacy the treatment.

While it is tempting to believe that psychotherapy is as effective as our local estimates indicate, we have enough reason to be skeptical. The pattern of local evidence producing more favorable results is not limited to this topic and not to Iranian studies. Several investigations on different health-related topics suggest that compared to...
international evidence, studies from some countries tend to produce larger effects and higher rates of significant results.\textsuperscript{13–16} Generally, trials conducted in less-developed countries seem to produce more favorable results.\textsuperscript{17}

It can be mentioned that, it is not easy to determine whether the difference between local and international estimates is real or induced by biased estimates on one side. Indeed, unless we have multinational RCTs, we may not be able to directly address such a question, as controlling for confounding factors is very difficult, if not impossible. However, we can explore different sources of bias that can lead to such differences.

1.2. Methods to Examine the Presence of Bias in Local Evidence

To explore bias, we can use internal and external information. External information comes from the outside of the studies under investigation. For instance, if we have evidence that the acceptance of a manuscript in Iranian journals is more affected by the reported effect size or \(p\)-value, we may suspect that publication bias has inflated the effect size from Iranian trials, leading to a difference between their effect size and the international estimates. Internal information, on the other hand, comes from the collective patterns among the studies under investigation, such as using funnel plot to explore publication bias.

Collecting external information on local evidence bias in a given topic requires familiarity with such potential sources of bias. Current literature suggests several sources. First, publication bias may be greater in some countries—i.e., publishing a non-significant result may be more difficult in some countries.\textsuperscript{13} Second, research quality is not similar across countries,\textsuperscript{19} and lower quality is shown to be associated with inflated effect size.\textsuperscript{18} Finally, while questionable research practice (e.g., \(p\)-hacking) and misconduct (i.e., data falsification and data fabrication) are worldwide issues, they may not be equally common across countries. There is reliable evidence that the rate for such behaviors is higher in low- and middle-income countries.\textsuperscript{19}

Regarding the internal information, it is well known that publication bias can be evaluated in a meta-analysis. A common tool for assessing and visualizing publication bias is the funnel plot. The funnel plot is a simple scatter plot, in which the effect size is plotted against a measure of precision (such as sample size or other related measures).\textsuperscript{20} If publication bias exists, we expect an asymmetry in the plot, because small trials tend to get published only if they report significant results while large trials are likely to get published regardless of the statistical significance of their results (for an example, see Figure 2 in the Results section).

What is less known is that misconduct and questionable research practice can also cause asymmetry in the funnel plot.\textsuperscript{21} Assume that for a given treatment, the true effect size is \(SMD = 0.5\). In some small RCTs with an \(SMD\) estimate of around 0.5, the statistical tests may not reach significance. A researcher may feel tempted to fraudulently change the data to elevate the effect size and get a significant result. Another researcher may exclude some valid cases to reach statistical significance. These behaviors lead to a systematic change in the observed effect sizes from small trials and produce an asymmetry in the funnel plot.

Furthermore, study quality can also cause asymmetry in the funnel plot. Larger studies usually have higher quality, and high-quality trials tend to report lower effect sizes; putting differently, low-quality studies tend to have higher effect sizes due to biased estimation.\textsuperscript{21} This leads to an asymmetry in the plot.

Besides using the funnel plot (or other similar methods), there is a second source of internal evidence: When we carefully review the studies under investigation, we may find some evidence of problematic quality or misconduct/questionable research practice. Also, some moderation analyses can provide further evidence. For instance, if dissertation projects, which are published in some databases regardless of the significance of their results, show smaller effect sizes than other studies, it may indicate the presence of publication bias.

1.3. Obtaining Sensible Estimates from Biased Evidence

As bias is a systematic deviation in the estimate, it can be corrected, if we know its direction and magnitude. The methods that have been devised to evaluate publication bias (e.g., the funnel plot or Egger’s regression) can provide an estimate of the direction and magnitude of bias. Although they are commonly assumed to assess merely publication bias, they indeed assess funnel plot asymmetry, whether it is caused by publication bias, questionable research practice, misconduct, or lower quality of small studies. The corresponding corrections also provide less biased estimates, regardless of the source of bias.

While useful, these methods have a requirement. They are based on the relation between sample size (or other related concepts) and effect size. To capture such a relationship, we need to have sufficient variability in the study size, and we need some trials with fairly large samples; otherwise, the logic behind these methods does not stand. In such a situation, a meta-analyist may use such methods and falsely assume that the bias is corrected. Therefore, in our previous meta-analyses on similar topics, we did not perform such analyses.\textsuperscript{10–12,22} However, if we find a context where we can apply such methods, we may be able to provide a reliable answer to whether psychotherapy for Iranian CP patients is as effective as the previous meta-analytic estimates suggest,\textsuperscript{10–12,22} or that those estimates were partly functions of local evidence bias.
1.4. The Current Study
The aim of the present study was to critically evaluate the evidence on the efficacy of psychological interventions for improving quality of life and reducing disability in Iranian adults with CP. The choice of these outcomes is based on several points: (a) we have large trials that have evaluated these outcomes, allowing proper evaluation and correction for bias, (b) there are a large number of trials that have reported outcomes on these variables, and (c) these outcomes have not yet been examined systematically. We will provide a corrected estimate for the treatment efficacy. Applying the new advanced meta-analytic methods, we will decompose the within- and between-study variability to evaluate how different sources of variance have contributed to the total variance. Finally, we will provide some estimates on pre-posttest improvement in experimental and control conditions, which is not meant to provide evidence of the efficacy but to provide some information for practitioners regarding the expected improvement.

2. Methods
This study was a systematic review and a meta-analysis reporting the following PRISMA guidelines.

2.1. Protocol and Registration
The protocol for this review was registered on PROSPERO, Record [deleted for blind review]. However, two changes were made. First, we decided not to perform a risk of bias assessment. We came to this decision after further evaluation of our previous assessments for similar topics. On one hand, previous assessments suggested minimal variability across trials; almost all trials had an overall low or unclear risk of bias. Due to the lack of variability, reliably exploring the relation between study quality and effect size was not possible. On the other hand, the descriptive purpose of such an assessment was already achieved by previous assessments. Considering it all together, it seemed that such an assessment is of little use and is likely a waste of resources. Second, we added three moderators: scale type (QoL vs. disability), scale domain (e.g., physical, psychological), and follow-up duration. To note, these additional variables were specified before data analysis—i.e., they were not post hoc analyses.

2.2. Eligibility Criteria
2.2.1. Participants and Settings
Iranian adults (age >16) with a diagnosis of any CP condition were eligible, except for pain related to cancer or Multiple Sclerosis (MS). No setting restriction was considered.

2.2.2. Interventions
Psychological interventions were eligible, from any type and in any delivery format (e.g., individual or group therapy). We excluded interventions with merely medical education (e.g., postural training for back pain) or substantial physical elements (e.g., yoga), as there may be some doubt about whether these should be classified as psychological interventions. To reduce heterogeneity, we also excluded brief interventions that were delivered in less than four sessions.

2.2.3. Type of Control
Active or nonactive control groups were eligible.

2.2.4. Outcome Measures
All measures of quality of life or disability were eligible.

2.2.5. Study Design
Randomized trials published in peer-reviewed journals, theses/dissertations, and full-text conference articles were included.

2.3. Information Sources, Search Strategy, and Study Selection
On May 2021, we searched Elm-Net, PubMed, and ProQuest for eligible records. Different combinations of the following words (or their Persian synonyms) were used: pain, fibromyalgia, irritable bowel syndrome, disability, and quality of life. For Pubmed and ProQuest, some terms were added to restrict the search to Iranian studies: Iran, Iranian, Persian, and Farsi. We also searched the citations from previous meta-analyses with related topics. The titles and abstracts were screened by one reviewer. The full-text assessments were done by two independent reviewers. Among the 67 reports, their initial decisions were consistent in 58 records (86%; Cohen’s kappa = .49). The disagreements were resolved through discussion.

2.4. Data Preparation and Synthesis Methods
2.4.1. Calculating Within-study Effect Sizes
Three types of Standardized Mean Difference (SMD) were calculated for each outcome: the between-group post-test difference and the within-group pre-posttest differences in experimental and control groups. Whenever possible, the effect size was calculated from the results of the analysis of covariance, which provides a more precise estimate of the intervention effect. If data on subscales were available, they were preferred over the overall scale. All the effect sizes were corrected for small sample bias.

2.4.2. Data Screening
The previous investigations on related topics in our literature suggest the necessity of careful examination of included RCTs for inconsistency and practically impossible results. In this study, we first screened the
Figure 1. PRISMA Diagram for Search and Screening Processes
<table>
<thead>
<tr>
<th>Study</th>
<th>Pain type</th>
<th>Intervention (format, sessions)</th>
<th>Measure</th>
<th>N</th>
<th>Num. of ESs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tabatabaei 2014<strong>44</strong></td>
<td>IBS</td>
<td>metacognitive therapy (gp, 8)</td>
<td>SF–36</td>
<td>21</td>
<td>8</td>
</tr>
<tr>
<td>Zomorodi 2013<strong>55</strong></td>
<td>IBS</td>
<td>CBT (gp, 8), MBSR (gp, 8)</td>
<td>IBS-Qol–34</td>
<td>36</td>
<td>4</td>
</tr>
<tr>
<td>Sadeghi 2015<strong>56</strong></td>
<td>LBP</td>
<td>Emotion regulation training (gp, 8), Coping skills training (gp, 8)</td>
<td>SF–36</td>
<td>60</td>
<td>16</td>
</tr>
<tr>
<td>Tavakoli 2019<strong>57</strong></td>
<td>NCCP</td>
<td>ACT (gp, 8)</td>
<td>SF–36</td>
<td>40</td>
<td>4</td>
</tr>
<tr>
<td>Firooz 2020<strong>58</strong></td>
<td>LBP</td>
<td>positive thinking-based CBT (cellphone, 30 days)</td>
<td>WHOQL–26</td>
<td>62</td>
<td>4</td>
</tr>
<tr>
<td>Rezaie 2014<strong>59</strong></td>
<td>RA</td>
<td>emotion-focused CBT (gp, 8), schema therapy (gp, 8)</td>
<td>Qol-RA–8</td>
<td>30</td>
<td>4</td>
</tr>
<tr>
<td>Masoumian 2013<strong>60</strong></td>
<td>LBP</td>
<td>MBSR (gp, 8)</td>
<td>SF–36</td>
<td>18</td>
<td>8</td>
</tr>
<tr>
<td>Ghotbinejad 2019<strong>61</strong></td>
<td>CP</td>
<td>positive psychotherapy (gp, 15), psychodrama (gp, 10)</td>
<td>SF–36</td>
<td>45</td>
<td>4</td>
</tr>
<tr>
<td>Haghayegh 2010<strong>62</strong></td>
<td>IBS</td>
<td>CBT (gp, 8)</td>
<td>IBS-Qol–34</td>
<td>24</td>
<td>2</td>
</tr>
<tr>
<td>Salarian 2020<strong>63</strong></td>
<td>CP</td>
<td>mindfulness (gp, 8)</td>
<td>WHOQL–26</td>
<td>26</td>
<td>1</td>
</tr>
<tr>
<td>Abazari 2016<strong>64</strong></td>
<td>LBP</td>
<td>hypnotherapy (ind, 6)</td>
<td>RMQ–24</td>
<td>28</td>
<td>1</td>
</tr>
<tr>
<td>Sadeghi 2020<strong>65</strong></td>
<td>RA</td>
<td>ACT (gp, 8)</td>
<td>ODI–10(v2.0)</td>
<td>40</td>
<td>1</td>
</tr>
<tr>
<td>Parhizgar 2020<strong>66</strong></td>
<td>IBS</td>
<td>mindfulness integrated CBT (gp, 8)</td>
<td>IBS-Qol–34</td>
<td>24</td>
<td>1</td>
</tr>
<tr>
<td>Shafiee Fard 2015<strong>67</strong></td>
<td>CP</td>
<td>CBT (gp, 12)</td>
<td>SF–36</td>
<td>30</td>
<td>8</td>
</tr>
<tr>
<td>Naddafnia 2019<strong>68</strong></td>
<td>IBS</td>
<td>CBT (gp, 12)</td>
<td>IBS-Qol–34</td>
<td>23</td>
<td>16</td>
</tr>
<tr>
<td>Farokh 2019<strong>69</strong></td>
<td>IBS</td>
<td>CBT (gp, 8)</td>
<td>IBS-Qol–34</td>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td>Kamkar 2011<strong>70</strong></td>
<td>IBS</td>
<td>stress management CBT (gp, 8)</td>
<td>IBS-Qol–34</td>
<td>42</td>
<td>16</td>
</tr>
<tr>
<td>Haghreagh 2012<strong>71</strong></td>
<td>IBS</td>
<td>emotion focused therapy (gp, 8), DBT (gp, 8)</td>
<td>IBS-Qol–34</td>
<td>60</td>
<td>4</td>
</tr>
<tr>
<td>Mohamadi 2017<strong>72</strong></td>
<td>IBS</td>
<td>DBT (gp, 8)</td>
<td>IBS-Qol–34</td>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td>Azizi 2018<strong>73</strong></td>
<td>IBS</td>
<td>mindfulness (gp, 8)</td>
<td>IBS-Qol–34</td>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td>Dabbaghi 2015<strong>74</strong></td>
<td>IBS</td>
<td>MBSR (gp, 8)</td>
<td>IBS-Qol–34</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>Solati 2009<strong>75</strong></td>
<td>IBS</td>
<td>CBT (gp, 8), life style education (gp, 8), relaxation (gp, 8), metacognitive therapy (gp, 8),</td>
<td>IBS-Qol–34</td>
<td>64</td>
<td>6</td>
</tr>
<tr>
<td>Pashang 2019<strong>76</strong></td>
<td>IBS</td>
<td>metacognitive therapy (gp, 8), ACT (gp, 8)</td>
<td>SF–36</td>
<td>45</td>
<td>4</td>
</tr>
<tr>
<td>Kheirabadi 2010<strong>77</strong></td>
<td>IBS</td>
<td>Coping skills training (gp, 8)</td>
<td>IBS-Qol–34</td>
<td>46</td>
<td>1</td>
</tr>
<tr>
<td>Shojaie 2017<strong>78</strong></td>
<td>LBP</td>
<td>Multidisciplinary (gp, 4)</td>
<td>QDS–20</td>
<td>125</td>
<td>1</td>
</tr>
<tr>
<td>Tavafian 2011<strong>79,52,53</strong></td>
<td>LBP</td>
<td>Multidisciplinary (gp, 5)</td>
<td>SF–36, RMQ–24, QDS–20</td>
<td>197</td>
<td>60</td>
</tr>
<tr>
<td>Chamani 2019<strong>80</strong></td>
<td>IBS</td>
<td>CBT (gp, 10)</td>
<td>WHOQL–26</td>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td>Anvari 2012<strong>81</strong></td>
<td>CP</td>
<td>ACT (gp, 8)</td>
<td>PDI–7</td>
<td>17</td>
<td>2</td>
</tr>
<tr>
<td>N. Vakili 2009<strong>82</strong></td>
<td>LBP</td>
<td>CBT (gp, 8)</td>
<td>SF–36</td>
<td>24</td>
<td>2</td>
</tr>
<tr>
<td>Irandoost 2014<strong>83</strong></td>
<td>LBP</td>
<td>ACT (unclear, 8)</td>
<td>SF–36</td>
<td>40</td>
<td>2</td>
</tr>
<tr>
<td>Rezaei 2014<strong>84</strong></td>
<td>PP</td>
<td>ACT (gp, 8)</td>
<td>PDI–7</td>
<td>22</td>
<td>2</td>
</tr>
<tr>
<td>Dadollahi 2016<strong>85</strong></td>
<td>IBS</td>
<td>MBSR (gp, 9)</td>
<td>IBS-Qol–34</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>Fouladi 2018<strong>86</strong></td>
<td>LBP</td>
<td>CBT (ind, 8)</td>
<td>IBS-Qol–34</td>
<td>32</td>
<td>2</td>
</tr>
<tr>
<td>Jafari 2020<strong>87</strong></td>
<td>CP</td>
<td>MBCT (gp, 8)</td>
<td>WHOQL–26</td>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td>Mansoobi 2020<strong>88</strong></td>
<td>LBP</td>
<td>ACT (gp, 8)</td>
<td>WHOQL–26</td>
<td>20</td>
<td>3</td>
</tr>
<tr>
<td>T. Vakili 2020<strong>89</strong></td>
<td>IBS</td>
<td>Reality therapy (gp, 8)</td>
<td>WHOQL–26</td>
<td>30</td>
<td>4</td>
</tr>
<tr>
<td>Soltanian 2016<strong>90</strong></td>
<td>RA</td>
<td>CBT (gp, 8)</td>
<td>WHOQL–26</td>
<td>20</td>
<td>4</td>
</tr>
<tr>
<td>Ayoughi 2019<strong>91</strong></td>
<td>RA</td>
<td>stress inoculation training (gp, 10)</td>
<td>SF–36</td>
<td>30</td>
<td>8</td>
</tr>
<tr>
<td>Boloorsaz 2017<strong>92</strong></td>
<td>FS</td>
<td>Islamic-based coping strategy training (gp, 10)</td>
<td>WHOQL–26</td>
<td>30</td>
<td>4</td>
</tr>
<tr>
<td>Pourmahmohseni 2017<strong>93</strong></td>
<td>IBS</td>
<td>MBCT (gp, 8)</td>
<td>SF–36</td>
<td>40</td>
<td>7</td>
</tr>
<tr>
<td>Mahdipoor 2012<strong>94</strong></td>
<td>IBS</td>
<td>mixed (gp, 12)</td>
<td>WHOQL–26</td>
<td>30</td>
<td>3</td>
</tr>
<tr>
<td>Esmaeilpoor 2017<strong>95</strong></td>
<td>IBS</td>
<td>Emotional Intelligence Training (gp, 12)</td>
<td>WHOQL–26</td>
<td>36</td>
<td>1</td>
</tr>
</tbody>
</table>

ES = effect size; IBS = irritable bowel syndrome; LBP = low back pain; NCCP = noncardiac chest pain; RA = rheumatoid arthritis; CP = chronic pain (general); PP = pelvic pain; FS = fibromyalgia syndrome; CBT = cognitive-behavioral therapy; MBSR = mindfulness-based stress reduction; ACT = acceptance and commitment therapy; DBT = dialectical behavioral therapy; MBCT = mindfulness-based cognitive therapy; SF–36 = Short Form Health Survey; IBS-Qol–34 = Irritable Bowel Syndrome–Quality of Life; WHOQL–26 = World Health Organization Quality of Life; Qol-RA = Quality of Life–Rheumatoid Arthritis; RMQ–24 = Roland–Morisk Disability Questionnaire; QDS–20 = Quebec Disability Scale; PDI–7 = Pain Disability Inventory.

Data for such cases. Note that this procedure was not an exploration of outliers or influential cases, which is usually performed after initial analysis. We in fact excluded these studies as such results are likely due to some errors in their reports. For each study that was excluded, we explained the reason.

2.4.3. Synthesis Method
For data synthesis, a five-level meta-analytic model was implemented. Effect sizes for different measurement occasions (i.e., the posttests and follow-ups; Level 1) were nested within outcome measures (e.g., subscales of SF–36; Level 2), nested within treatment comparison (for
2.4.4. Additional Analyses

Seven potential moderators were investigated in the present study: pain type (musculoskeletal vs. IBS), intervention type, researcher’s education, publication type (dissertation or not), measurement type (i.e., QoL vs. disability measure), measurement domain (i.e., psychological, physical), and length of follow-up period. As the intervention types are numerous, they were classified into five main categories: cognitive behavioral, mindfulness-based, coping skills training, multidisciplinary interventions, and other approaches. If needed, we checked the intervention protocol for accurate classification.

2.4.5. Exploring Publication and Other Related Sources Bias

To investigate publication and other related sources of bias and to provide corrected estimates on the treatment efficacy, we used two methods. First, we used a recently-developed extension of the regression test, which includes...
the inverse of effective sample size as a moderator in a multilevel meta-regression. The test for the slope of the moderator examines the presence of bias, and the intercept provides a corrected estimate of treatment efficacy. It is based on the premise that when the lack of precision is associated with higher effect sizes, the predicted value at perfect precision (i.e., when the inverse of the effective sample size is zero—the intercept) is a corrected estimate. Second, we aggregated the effect sizes within each study, and then generated a standard funnel plot, along with a trim-and-fill method to correct bias.

2.4.6. Statistical Software
All analyses were conducted in R using the package "Metafor," except for calculating SMD from analysis of covariance, which was made using package "compute.es".

3. Results
The initial search yielded 731 records. Twelve further records were found from other sources. Of these records, 42 RCTs met our criteria, comparing 50 experimental groups to a control condition (Figure I). Except for two, the control condition was Treatment As Usual (TAU). Overall, 225 between-group effect sizes were extracted. The most common pain type was IBS (n = 21, 50%), followed by LBP (n = 9, 21%). The most common intervention was CBT (n = 9, 18%), followed by ACT (n = 8, 16%). The total sample was 1639. A summary of the characteristics of included studies is provided in Table 1.

Table 2. The Moderating Effect of Pain-, Intervention-, and Measurement-Related Factors and Follow-Up Duration

<table>
<thead>
<tr>
<th>Moderator</th>
<th>n. ESs</th>
<th>SMD</th>
<th>95% CI</th>
<th>Test for moderator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Musculoskeletal</td>
<td>130</td>
<td>1.08</td>
<td>0.77, 1.39</td>
<td>F(1, 32) = 0.00, P = 0.962, R² = 0</td>
</tr>
<tr>
<td>IBS</td>
<td>55</td>
<td>1.09</td>
<td>0.77, 1.41</td>
<td></td>
</tr>
<tr>
<td>Intervention type</td>
<td></td>
<td></td>
<td></td>
<td>F(4, 29) = 20.9, P &lt; 0.001, R² = .01</td>
</tr>
<tr>
<td>CBT</td>
<td>40</td>
<td>0.87</td>
<td>0.51, 1.24</td>
<td></td>
</tr>
<tr>
<td>Coping strategy training</td>
<td>29</td>
<td>1.01</td>
<td>0.38, 1.65</td>
<td></td>
</tr>
<tr>
<td>Mindfulness-based</td>
<td>41</td>
<td>1.36</td>
<td>0.93, 1.78</td>
<td></td>
</tr>
<tr>
<td>Multidisciplinary</td>
<td>61</td>
<td>0.46</td>
<td>0.44, 0.49</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>14</td>
<td>1.13</td>
<td>0.93, 1.32</td>
<td></td>
</tr>
<tr>
<td>Researcher’s education</td>
<td></td>
<td></td>
<td></td>
<td>F(2, 31) = 0.2, P = 0.788, R² = 0</td>
</tr>
<tr>
<td>PhD</td>
<td>111</td>
<td>0.99</td>
<td>0.62, 1.37</td>
<td></td>
</tr>
<tr>
<td>MA</td>
<td>66</td>
<td>1.14</td>
<td>0.82, 1.45</td>
<td></td>
</tr>
<tr>
<td>Unclear</td>
<td>8</td>
<td>1.19</td>
<td>0.60, 1.78</td>
<td></td>
</tr>
<tr>
<td>Publication Type</td>
<td></td>
<td></td>
<td></td>
<td>F(1, 30) = 0.02, P = 0.896, R² = 0</td>
</tr>
<tr>
<td>Thesis (or extracted from thesis)</td>
<td>74</td>
<td>1.12</td>
<td>0.79, 1.45</td>
<td></td>
</tr>
<tr>
<td>Journal article</td>
<td>104</td>
<td>1.09</td>
<td>0.77, 1.42</td>
<td></td>
</tr>
<tr>
<td>Type of measurement</td>
<td></td>
<td></td>
<td></td>
<td>F(1, 32) = 26.6, P &lt; 0.001, R² = 0</td>
</tr>
<tr>
<td>Quality of Life</td>
<td>167</td>
<td>1.09</td>
<td>0.87, 1.32</td>
<td></td>
</tr>
<tr>
<td>Disability</td>
<td>18</td>
<td>1.02</td>
<td>0.80, 1.24</td>
<td></td>
</tr>
<tr>
<td>Measurement domain</td>
<td></td>
<td></td>
<td></td>
<td>F(4, 29) = 2.4, P = 0.074, R² = 0</td>
</tr>
<tr>
<td>General</td>
<td>38</td>
<td>1.17</td>
<td>0.86, 1.48</td>
<td></td>
</tr>
<tr>
<td>Physical</td>
<td>65</td>
<td>0.96</td>
<td>0.62, 1.29</td>
<td></td>
</tr>
<tr>
<td>Psychological</td>
<td>69</td>
<td>1.04</td>
<td>0.65, 1.44</td>
<td></td>
</tr>
<tr>
<td>Social</td>
<td>9</td>
<td>1.23</td>
<td>0.87, 1.59</td>
<td></td>
</tr>
<tr>
<td>Environmental</td>
<td>4</td>
<td>1.03</td>
<td>0.53, 1.53</td>
<td></td>
</tr>
<tr>
<td>Follow-up duration (month)</td>
<td>185</td>
<td></td>
<td></td>
<td>F(1, 32) = 351.4, P &lt; 0.001, R² = 0, B = 0.003 [0.0028, 0.0034]</td>
</tr>
</tbody>
</table>

The tests for moderation are based on robust variance estimation. As the models are multilevel models, the R² is in fact pseudo-R², which extends the logic to the multilevel model framework.

3.1. Initial Data Screening
The initial screening revealed three studies with SMDs higher than 5, which is probably due to some errors in their reports, because an average improvement of more than 5 SDs is not practically possible. Another study reported unusually different effect sizes across SF-36 subscales, with the SMDs ranging from -1.34 to 5.92, indicating some errors in the report, as it is not practically possible to have an SMD of 5.92 in one outcome and an SMD of -1.34 in another highly related outcome. Another study had seven effect sizes, consistently very large, which is probably due to some errors in their reports, because an average improvement of more than 5, 24. These six studies were excluded from further analysis. Note that excluding these studies was not because of them being outliers, but rather because of strong evidence of some errors in the reports that make the extracted effect sizes very likely invalid.

3.2. Psychological Interventions vs. Treatment As Usual
The overall SMD across 185 effect sizes was 1.08 [87, 1.30]. The total true variance across all levels was σ²(total) = 0.34, F(total) = 87%. To provide an easily readable forest plot, we calculated one aggregated effect size per study and used these to create the plot (Figure 2).
Decomposing the true variance showed that the majority of the variance is at the study level, $\sigma^2_{\text{between-study}} = 0.29 \; [0.0, 0.58]$, $I^2_{\text{between-study}} = 76\%$, and the variance due to within-study levels was fairly small, $\sigma^2_{\text{between-group}} = 0.04 \; [0.0, 0.38]$, $I^2_{\text{between-group}} = 10\%$; $\sigma^2_{\text{between-outcome}} = 0.01 \; [0.0, 0.04]$, $I^2_{\text{between-outcome}} = 1\%$; $\sigma^2_{\text{between-timepoint}} = 0.0 \; [0.0, 0.01]$, $I^2_{\text{between-timepoint}} = 0\%$. The results of seven moderation analyses have been presented in Table 2. The influence of the moderators were nonsignificant (publication type, measurement domain, and researcher’s education), negligible (follow-up duration), or difficult to interpret (intervention type; see Discussion).

### 3.3. Publication and other Sources of Bias

To examine publication and other related sources of bias, the inverse of the effective sample size was included as a moderator in a multilevel meta-regression (Figure 3, Panel a). The test for the slope was significant, $F(1, 32) = 8.8$, $p = .006$, indicating the presence of bias. The intercept, a corrected estimate of treatment efficacy, was 0.45 $[0.04, 0.87]$, which is much smaller than our original estimate. We also generated a funnel plot, based on the aggregated SMDs for each study (Figure 3, Panel b). The asymmetry of the observed SMDs (i.e., the black dots) was clear. Using the trim-and-fill method, 11 studies were imputed at the left side (the white dots in Figure 3, Panel b), reducing the SMD to 0.75 $[0.48, 1.0]$. While this is noticeably lower than our original estimate, it is higher than the corrected estimate from the meta-regression model.

### 3.4. Pre-posttest Improvement in Psychological Interventions

The pooled pre-posttest Standardized Mean Change (SMC) from 214 effect sizes only including experimental groups was 1.00 $[0.79, 1.22]$. The total true variance was $\sigma^2_{\text{total}} = 0.43$, $I^2 = 95\%$.

### Table 3. Direct Contrasts Among Different Psychotherapeutic Approaches

<table>
<thead>
<tr>
<th>Study</th>
<th>Contrast</th>
<th>SMD</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zomorodi 2013</td>
<td>CBT – MBSR</td>
<td>-0.78</td>
<td>-1.06, -0.5</td>
</tr>
<tr>
<td>Naddafnia 2019</td>
<td>CBT – spiritual therapy</td>
<td>0.54</td>
<td>0.31, 0.76</td>
</tr>
<tr>
<td>Solati Dehkordi 2009</td>
<td>CBT – lifestyle education</td>
<td>-0.03</td>
<td>-0.23, 0.17</td>
</tr>
<tr>
<td></td>
<td>CBT – relaxation training</td>
<td>-0.13</td>
<td>-0.33, 0.07</td>
</tr>
<tr>
<td></td>
<td>Relaxation – lifestyle education</td>
<td>0.1</td>
<td>-0.1, 0.29</td>
</tr>
<tr>
<td>Rezaee 2014</td>
<td>Emotion-focused CBT – schema therapy</td>
<td>0.00</td>
<td>-0.31, 0.31</td>
</tr>
<tr>
<td>Pashang 2019</td>
<td>Metacognitive therapy – ACT</td>
<td>0.45</td>
<td>0.23, 0.66$^{(a)}$</td>
</tr>
<tr>
<td>Ghobinejad 2019</td>
<td>Positive psychotherapy – psychodrama</td>
<td>-0.03</td>
<td>-0.24, 0.18</td>
</tr>
<tr>
<td>Sadeghi Mazidi 2015</td>
<td>Emotion regulation – coping skills training</td>
<td>0.42</td>
<td>0.28, 0.55</td>
</tr>
</tbody>
</table>

The SMDs are unweighted aggregate estimates from each study, and the confidence intervals are calculated based on an intercorrelation of 0.6 among the effect sizes. However, repeating the analysis based on intercorrelation values of 0.3 and 0.9 did not change the significance levels. (a) This study reported no significant difference based on a mixed analysis of variance, but based on the SMDs from the posttest and follow-up, the difference was significant.

### 3.5. Pre-posttest Improvement in Usual Care

Two studies did not have a TAU control, and three studies did not report sufficient information for calculating pre-post change scores. The pooled SMC from 160 effect sizes was 0.06 $[-0.09, 0.21]$. The total true variance was $\sigma^2_{\text{total}} = 0.18$, $I^2 = 95\%$, indicating
Interestingly high variability in the control condition.

3.6. Direct Comparison of Psychotherapeutic Approaches

Seven studies reported data on nine direct contrasts between different therapeutic approaches. The results of these direct comparisons have been presented in Table 3. However, we could not find any sensible pattern among these contrasts to suggest a reliable conclusion.

4. Discussion

This study aimed to synthesize the evidence on the efficacy of psychological interventions for improving the quality of life in Iranian adults with CP. Our original estimate indicated a large effect for these interventions; this is fairly similar to previous meta-analytic estimates from Iranian trials on other CP outcomes. However, we had evidence on the presence of bias; when we corrected for the bias, the estimated showed a medium effect size. No reliable evidence was found favoring one approach over another, and the efficacy was evident for patients with musculoskeletal pain and IBS. Our data suggested a slight enhancement in the treatment effect over time, but the evidence was mainly influenced by a single RCT with a large sample size and long follow-up periods.

The existing meta-analytic evidence from international literature suggests a small to medium effect of psychological intervention on the QoL in chronic pain patients (the SMDs are usually around 0.5). While our original estimate substantially departs from these estimates, our corrected estimate was relatively comparable.

Noticeably, our corrected estimate was close to the effect size from one large Iranian RCT. Altogether, the efficacy of psychological interventions in Iranian CP patients is likely fairly comparable to its efficacy in America and Europe, and if there is some cultural difference, it is less than what is suggested by the previous meta-analytic estimate from our local RCTs.

4.1. Current Evidence on the Sources of Bias in Iranian CP Trials

The substantial gap between the original and corrected estimates of treatment efficacy can indicate a strong bias. While we cannot directly determine the source of bias, we can search for supporting evidence.

Regarding publication bias, we hypothesized that if such publication bias is present, we expect the effect size from theses/dissertations (or articles that are extracted from them) to be lower than journal articles. However, our data showed very similar effect sizes from both types of RCTs. Also, exploring direct evidence on publication bias suggests that compared to international literature, such bias is not higher in Iranian literature. Therefore, it cannot explain why Iranian trials have higher effect sizes.

Regarding the bias due to the quality of trials, in two systematic reviews by two independent groups of researchers, the majority of Iranian CP trials were rated as having low quality. Their quality is substantially lower than international literature, especially in some aspects such as randomization bias. Therefore, it seems the evidence supports the possibility that the higher effect size from Iranian RCTs may be caused by the lower quality of trials.

With regard to questionable research practice and misconduct, one study suggested that misconduct (including data falsification, data fabrication, and plagiarism) is not rare in dissertations; and in the current study, at least 41% were extracted from dissertations. A second piece of evidence comes from a previous meta-analysis, in which the average SMD from investigations conducted by master’s degree researchers was more than twice the effect size from studies conducted by PhD researchers. This large difference is beyond what is expected from publication bias or study quality. Finally, some features in our research community can elevate the frequency of such behaviors. A notable example is the low rate of article retraction. Our search for Persian retracted papers in Google Scholar retrieved only nine papers in all disciplines. This can indicate the lack of scrutiny for detecting such malpractice, which in turn can elevate such behaviors, as the risk of exposure is minimal. Also, our personal experience from students is that they sense an unwritten obligation to report significant results; this is usually observable when we are asked for statistical consult.

Finally, the difference between average Iranian and non-Iranian effect size is not limited to the topic of psychotherapy for CP. A similar pattern is observable in other topics, such as the efficacy of exercise on quality of life in multiple sclerosis. This further supports the notion that the difference between the estimates is not all genuine; local evidence bias plays a role.

4.2. The Heterogeneity among Trials

Besides moderation analysis, which assesses potential sources of heterogeneity, the multilevel model allowed us to decompose the true heterogeneity across within- and between-study levels. It suggested that the variability is mainly related to the between-study variance, and the within-study variance was minimal, regardless of its source (i.e., intervention type, measurement tool, and follow-up period). Therefore, exploring other sources of variability, such as therapists’ expertise and treatment fidelity, may be more fruitful.

Our results on the effect of the intervention type showed substantially lower SMD for multidisciplinary programs, which is surprising. However, only two RCTs used such interventions, and they were the largest RCTs we have. Therefore, we highly suspect that the difference between these intervention types is similar to the
difference between the local and international evidence, especially as the corrected estimate was close to the SMD from these two trials. These results, along with the results from the direct comparisons, suggest no reliable preference among different therapeutic approaches.

5. Conclusion
Our results support the use of psychological interventions for improving quality of life and reducing disability in Iranian adults with CP. We observed strong bias in the observed effect sizes, with current evidence holding the low quality of trials and questionable research practice as the main source of bias. While publication bias is another possible source of bias, the supporting evidence is weak. The bias-corrected estimate suggested a moderate effect for these interventions, fairly close to the estimates from international evidence. In our situation, this corrected estimate seems to be our best guess, especially as it is supported by the estimate from a large RCT. 45,52,53

5.1. Limitations and Future Directions
Due to the lack of information, we did not include treatment fidelity in our analyses. This is an important limitation, especially as fidelity is likely an important aspect of heterogeneity. Another limitation is the lack of assessment for risk of bias. In the Methods section, we have explained that in such situation, an assessment is likely of little use. This study has several implications for future RCTs on this topic. First, for power calculation, it is probably better to assume an SMD of around 0.4–0.45 as the predicted effect size. Second, enhancing the quality of trials is vital for the further development of reliable local evidence. Third, treatment fidelity and therapist expertise are two important yet neglected aspects of our literature. Finally, we need more rigorous examinations of RCTs in our local journals; we suggest that each submitted study be scrutinized for evidence on research malpractice.

Author Contributions
ZA designed the study and wrote the protocol, collaborated with the data collection, conducted the statistical analysis, interpreted the results, and wrote the manuscript. ZSH collaborated with the data collection. FA collaborated with data collection. MF collaborated with the design and provided consult. All the authors approved the final version of the manuscript for submission.

Conflict of Interest Disclosures
All authors declared that they have no conflict of interest.

Ethical Approval
Not applicable.

Funding/Support
None.

Research Highlights

What Is Already Known?
• International literature suggests a small to medium effect size for psychotherapy for CP.
• Iranian literature suggests a large effect size for psychotherapy for CP.
• The difference between the Iranian and international estimates can be genuine, or it can be caused by local evidence bias.

What Does This Study Add?
• The difference does not seem genuine, as our bias-corrected estimate is fairly comparable to the international estimates.
• Sufficient evidence indicates notable bias in our randomized trials on the efficacy of psychotherapy for CP.
• Current evidence holds low quality of trials and/or questionable research practice as the main sources of bias in our local literature.

References

Zia-Tohidi et al

Hospital Practices and Research 2023;8(1):177-188
6. Panagiotou OA, Contopoulos-Ioannidis DG, Ioannidis JPA. Comparative effect sizes in randomised trials from less developed and more developed countries: meta-epidemiological assessment. BMJ. 2013;346:f707. doi:10.1136/bmj.f707
12. [Reference deleted for blind review]

Local Evidence Bias in Pain Psychotherapy Trials


54. Tabatabae Shahrabad A. The Effectiveness of Metacognitive Therapy on Anxiety, Depression and Quality of life of Patients with Irritable Bowel Syndrome (Thesis) [Internet]. Ferdowsi University of Mashhad; 2014. https://ganj.irandoc.ac.ir/#/articles/5eaba7c222a8b3e725e8f58e4ec4e485bf [Accessed January 5, 2022]


59. Rezaie F. A comparison of the effectiveness of emotion focused cognitive behavioral therapy, and schema therapy on pain, pain catastrophizing, emotion regulation and quality of life in rheumatoid arthritis patients (PHD thesis) [Persian] [Internet]. 2014. https://ganj.irandoc.ac.ir/#/articles/9d43ddbe2e0b0a3e13a3e037154648db [Accessed January 5, 2022]

60. Masumian S. The Effect of Mindfulness-Based Stress Reduction (MBSR) on Decreasing Depression, increasing Quality of life and self-efficacy of the Chronic Low Back Pain Patients(CLBP) [Internet]. Allameh Tabataba'i university; 2013. https://ganj.irandoc.ac.ir/#/articles/a36b0b3e27a4e64725cd814e9261aba [Accessed January 5, 2022]


65. Kamkar A. The Effectiveness of group Cognitive – Behavioral Stress Management on physical symptoms of Patients with Irritable Bowel Syndrome with and without comorbidity of anxiety and depression and on increased of their quality of life (Thesis) [Internet]. Allameh Tabataba'i University; 2011. https://ganj.irandoc.ac.ir/#/articles/955e6f87b8a646c279de0e838d54e3 [Accessed January 5, 2022]


68. Dabbagh Zafir S. Effectiveness of mindfulness based stress reduction (MBSR) on mental health and quality of life in Irritable Bowel Syndrome (IBS) cases. Payam Noor University; 2015.

69. Solati-Dehkordi K. A comprehensive study of cognitive-behavioral therapy, life style training, relaxation and medication therapies on quality of life, psychological symptoms, severity and frequency of symptoms IBS-d patients (Thesis) [Internet]. University of Isfahan; 2009. https://ganj.irandoc.ac.ir/#/articles/e3e319fcd99ff7be8095245bab35258 [Accessed January 5, 2022]


75. Vakili N. effectiveness of group cognitive behavior therapy on depression, quality of life of patients with low back pain (thesis) [Internet]. university of Isfahan; 2009. https://ganj.irandoc.ac.ir/#/articles/b2b709c6f660dcd929e74de3784ab8/ [Accessed January 5, 2022]


