Review Article

Long-Term Opioid Therapy for Chronic Noncancer Pain: A Systematic Review and Meta-Analysis of Efficacy and Safety

Meredith Noble, MS, Stephen J. Tregear, DPhil, Jonathan R. Treadwell, PhD, and Karen Schoelles, MD, SM
ECRI Institute, Evidence-Based Practice Center and Health Technology Assessment Group, Plymouth Meeting, Pennsylvania, USA

Abstract

Opioid therapy for chronic noncancer pain (CNCP) is controversial due to concerns regarding long-term efficacy and adverse events (including addiction). We systematically reviewed the clinical evidence on patients treated with opioids for CNCP for at least six months. Of 115 studies identified by our search of eleven databases (through April 7, 2007), 17 studies (patients \( n = 3,079 \)) met inclusion criteria. Studies evaluated oral (studies \( k = 7; n = 1,504 \)), transdermal (\( k = 3; n = 1,993 \)), and/or intrathecal (\( k = 8; n = 177 \)) opioids. Many patients withdrew from the clinical trials due to adverse effects (oral: 32.5% [95% confidence interval (CI), 26.1%–39.6%]; intrathecal: 6.3% [95% CI, 2.9%–13.1%]; transdermal: 17.5% [95% CI, 6.5%–39.0%]), or due to insufficient pain relief (oral: 11.9% [95% CI, 7.8%–17.7%]; intrathecal: 10.5% [95% CI, 3.5%–27.4%]; transdermal: 5.8% [95% CI, 4.2%–7.3%]). Signs of opioid addiction were reported in only 0.05% (1/2,042) of patients and abuse in only 0.43% (3/685). There was an insufficient amount of data on transdermal opioids to quantify pain relief. For patients able to remain on oral or intrathecal opioids for at least six months, pain scores were reduced long-term (oral: standardized mean difference [SMD] 1.99, 95% CI, 1.17–2.80; intrathecal: SMD 1.33, 95% CI, 0.97–1.69). We conclude that many patients discontinue long-term opioid therapy due to adverse events or insufficient pain relief; however, weak evidence suggests that oral and intrathecal opioids reduce pain long-term in the relatively small proportion of individuals with CNCP who continue treatment. J Pain Symptom Manage 2008;35:214–228. © 2008 U.S. Cancer Pain Relief Committee. Published by Elsevier Inc. All rights reserved.

Key Words

Pain, intractable, opioids, narcotics, chronic pain, noncancer pain, nonmalignant pain
**Background**

Chronic pain is an important and common medical concern worldwide. In the United States, pain is the most common complaint that leads patients to seek medical care. A systematic review of four international studies conducted in developed countries found prevalence rates of any type and severity level of chronic pain ranging from 10.5% to 55.2% of the population.\(^1\) The Pain in Europe survey of 46,000 individuals showed that one in five people reported suffering from chronic pain. In this survey, chronic pain sufferers reported seven years of chronic pain on average, with some reporting pain lasting more than 20 years.\(^2\) A World Health Organization survey of primary care patients seeking care at 15 centers in 14 countries across Asia, Africa, Europe, South America, and North America found that 22% of primary care patients reported pain lasting more than six months.\(^3\) An estimated 9% of Americans\(^4\) and 19% of Europeans\(^2\) have moderate-to-severe chronic noncancer pain (CNCP). Older people and women are more likely to experience chronic pain.\(^2,5\)

Although opioid use for acute/postsurgical pain and for palliative care is accepted in the United States, there is debate about whether opioids are appropriate for the treatment of CNCP,\(^8,9\) although acceptance may be growing.\(^10\) The most commonly prescribed opioids in the United States are codeine, fentanyl, hydrocodone, hydromorphone, levorphanol, meperidine, methadone, morphine, and oxycodone.\(^11\) The efficacy of opioids for CNCP has been demonstrated in short-term trials,\(^12\) including for neuropathic pain,\(^13\) but little is known about whether these agents continue to be effective over the many months or years typical of CNCP. Concerns have also been raised about adverse effects that may arise with long-term use, including the development of psychological addiction and/or abuse. The purpose of this systematic review is to summarize the evidence pertaining to the efficacy and safety of long-term opioid therapy for CNCP.

**Methods**

**Search Strategy**

Eleven databases were searched, including EMBASE (1980 through April 7, 2007) and PubMed (1966 through April 7, 2007) and all Cochrane databases and registries (inception through Issue 2, 2007). We used search terms including the following: chronic pain (pain, intractable [major heading(mh)]) OR (pain AND (chronic OR intractable OR refractory OR persistent OR chronic disease[mh])), a list of painful conditions, and all opioids in current use by generic and proprietary names. Full search strategies are available upon request from the contact author. We also examined reference lists from identified studies and reviewed gray literature for additional studies not identified by other means.

**Selection of Studies**

Systematic reviews aim to identify the best available evidence, which is often in the form of randomized controlled trials (RCTs). However, no placebo-controlled, long-term RCTs on the efficacy and safety of opioids for CNCP exist, and only one long-term controlled trial (which compares two opioids)\(^14\) exists. Currently, the best available long-term evidence is from open-label, uncontrolled, time-series studies.

Open-label, uncontrolled, time-series studies are susceptible to bias. Nonetheless, we chose to analyze data from these studies because (1) the patients in these studies had pain refractory to treatment for at least three months before enrollment (i.e., International Association for the Study of Pain [IASP] definition); and (2) all patients had failed previous pharmacotherapy before beginning opioids (in accordance with the World Health Organization [WHO] analgesic ladder). These factors suggest that patients’ pain levels probably would not have lessened without active treatment.

We searched for studies that (1) collected data on patients after at least six months of opioid therapy; (2) were published in English; (3) were reported as full-text articles; (4) did not include patients also reported on in other included studies; (5) were prospective; (6) enrolled at least 10 patients; (7) enrolled only patients who had CNCP, defined as pain lasting at least three months as defined by IASP. We used two additional criteria for pain outcomes: (1) pain outcomes must have been patient-reported; and (2) outcome data must not have been collected
retrospectively (e.g., post-treatment surveys/questionnaires), because reports based on memory of pain may differ from reports given at the time that pain is experienced.\textsuperscript{15, 16}

**Data Collection**

Data on patient withdrawal from the study due to adverse events/effects, the incidence and type of adverse drug effects (including addiction and abuse rates), withdrawal due to insufficient pain relief and pain relief were independently extracted by two researchers. Discrepancies were settled by consensus.

**Analysis**

Two methodologists independently assessed the quality of all identified studies using a 14-item quality assessment instrument (Table 1). This quality assessment scale was developed by ECRI Institute to assess the internal validity of studies using domains identified as important factors by experts in the field.\textsuperscript{17–19} We assessed the quality of each included study for each outcome of interest and used the median quality of the studies for each outcome to describe the overall evidence-base quality (refer to Table 1).

We analyzed the available evidence using systematic a priori protocols.\textsuperscript{20} We pooled data using meta-analysis only when at least three studies per mode of opioid administration addressed a particular outcome of interest. These protocols were used to formulate both qualitative (i.e., the direction of a treatment effect) and quantitative (i.e., the size of a treatment effect) conclusions.

Follow-up times in included studies ranged from 6 to 48 months. We did not pool data for specific time points, but chose instead to use the last reported time point of each study and then performed meta-regression to assess whether the treatment duration affected the outcomes. We took this approach because too few studies reported data at any given time point to enable multiple time point analyses. Reporting of data in intrathecal studies differed from oral and transdermal opioid studies in that means of results from patients’ last reported follow-up visits were presented as results for the mean length of follow-up.

Fixed-effects and random-effects meta-analyses were used to pool data from different studies.\textsuperscript{21–25} We used the $I^2$-statistic to quantify the magnitude of important differences in the findings of different studies (i.e., heterogeneity).\textsuperscript{21, 26, 27} When no substantial heterogeneity ($I^2 \leq 50\%$) was detected, we used the fixed-effects model. Whenever appropriate, heterogeneity was explored with meta-regression techniques using a permutation test.\textsuperscript{28, 29} We decided a priori to investigate duration of treatment as a covariate whenever we detected substantial heterogeneity, given our assumption that length of treatment would be of particular importance. We acknowledge, however, that the small number of studies limits the power to detect an effect. For a meta-regression using patient characteristics such

<table>
<thead>
<tr>
<th>Quality Assessment Item</th>
<th>ECRI Institute Before—After Study Quality Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Does the study specifically state that previous treatments for this disease, disorder, or condition had been attempted and were unsuccessful?</td>
<td></td>
</tr>
<tr>
<td>2. Was the study prospective?</td>
<td></td>
</tr>
<tr>
<td>3. Did the study enroll ALL patients or CONSECUTIVE patients?</td>
<td></td>
</tr>
<tr>
<td>4. Were the criteria for including and excluding patients based on objective laboratory and/or clinical findings?</td>
<td></td>
</tr>
<tr>
<td>5. Were the patient inclusion/exclusion criteria established a priori?</td>
<td></td>
</tr>
<tr>
<td>6. Was the same treatment given to all patients enrolled? (If treatment was specifically tailored to individual patients, and was reported by authors to be in the same category of treatment, answer “Yes.”)</td>
<td></td>
</tr>
<tr>
<td>7. Was the study free of confounding interventions?</td>
<td></td>
</tr>
<tr>
<td>8. Were the outcome measures objective?</td>
<td></td>
</tr>
<tr>
<td>9. Did $\geq 85%$ of patients complete the study?</td>
<td></td>
</tr>
<tr>
<td>10. Were the characteristics of those who did and did not complete the study compared, and were these characteristics similar?</td>
<td></td>
</tr>
<tr>
<td>11. Was the analysis conducted on an intent-to-treat basis, or was data presented in such a way that one could be performed?</td>
<td></td>
</tr>
<tr>
<td>12. Did the study have at least 80% statistical power to detect a clinically important change in the outcome variable after the intervention?</td>
<td></td>
</tr>
<tr>
<td>13. Was the funding for this study derived from a source that does not have a financial interest in its results?</td>
<td></td>
</tr>
<tr>
<td>14. Were the author’s conclusions, as stated in the abstract or the article’s discussion section, supported by the data presented in the article’s results section?</td>
<td></td>
</tr>
</tbody>
</table>
as age, diagnosis, or duration of pain, patient-level data would be preferable, because it would increase statistical power and would be less susceptible to the ecological fallacy. However, such data were not available to this project. Therefore, we decided against exploring additional possible sources of heterogeneity (e.g., type of pain, patient characteristics, type and dosing of opioid, degree of flexibility in dosing, concomitant therapies, and use of preventive treatment for common adverse events). When heterogeneity was present that could not be explained by duration of follow-up, we used the random effects model for meta-analysis.

We calculated the standardized mean difference (SMD) for continuous and categorical data (e.g., visual analog scale [VAS] pain scores and Likert scales). Because time series data are not independent, SMDs were adjusted by assuming a correlation of 0.50. If means and standard deviations were not available for continuous data, every effort was made to determine an estimate of treatment effect from reported statistics (e.g., $t$-values, $f$-values) or from $F$-values using methods described in detail elsewhere. Event rates (odds) were calculated for dichotomous data. Results are presented as percentages or numeric pain scores and 95% confidence intervals (CIs).

Cumulative meta-analysis and impact meta-analysis (in which we removed and replaced one study at a time) were performed for all outcomes. For continuous outcomes, assumptions about correlation coefficients were tested by sensitivity analyses in which we varied the correlation coefficient from 0.00 to 1.00 to determine whether the summary effect size was stable, in the method of Thiessen Philbrook and colleagues. Because patients who withdraw may differ in important ways from those who remain in a study, we also performed sensitivity analyses in which we inflated baseline standard deviations for pain scores by 100% and 200%. We considered more than a 25% change in summary effect size upon manipulation of the correlation coefficient and/or the baseline standard deviations to indicate that the quantitative estimate of the meta-analysis was not stable. For all fixed-effects meta-analyses, we also computed a random effect meta-analysis in addition to a fixed-effect meta-analysis, to determine whether the type of meta-analysis model used substantially affected the summary effect size and/or CIs. In evidence bases without substantial heterogeneity ($I^2 \leq 50\%$) we used the “trim-and-fill” method to test for funnel plot asymmetry, which suggests missing studies, possibly due to publication bias.

Comprehensive Meta-Analysis (Biostat, Englewood, NJ) software was used for most statistical analyses. Meta-regression with permutation tests was performed using STATA (StataCorp LP, Bryan/College Station, TX) software.

**Findings**

**Study Selection**

Our searches identified 3,672 abstracts from which we retrieved 115 studies; 17 of those studies met our inclusion criteria (Fig. 1). All but one of the 17 studies were open-label, single-arm time series. The other study included one group in which patients were treated with transdermal fentanyl and another in which patients were treated with oral extended-release morphine.

**Study Description**

All of the evidence bases considered in this systematic review were of low quality. Their low quality was primarily due to failure to describe patient recruitment methods, failure to compare the characteristics of patients who completed the study to patients who did not, high patient withdrawal rates, and failure to conduct intent-to-treat (ITT) analyses in addition to completer analyses. For pain outcomes, studies met an average of 8.2 quality criteria on average (range 7–10). For withdrawal outcomes, studies met an average of 9.5 quality criteria (range 9–11) (see Table 1).

Time-series data for 18 separate treatment groups from 17 separate studies that enrolled a total of 3,808 patients with CNCP formed the complete evidence base for this review (Table 2). This evidence base was stratified by mode of opioid administration: oral (7 treatment groups, $n = 1,504$), transdermal (3 treatment groups, $n = 1,399$), and intrathecal pump (8 treatment groups, $n = 177$). Patients in all treatment groups reported moderate-to-severe pain at baseline due to nociceptive or neuropathic pain, or both. Types of pain included general or mixed CNCP (studies $[k] = 2$).
neuropathic pain \( (k = 2),^{40,41} \) osteoarthritis \( (k = 3),^{42-44} \) back pain \( (k = 8),^{14,45-51} \) low back and leg pain \( (k = 1),^{52} \) and neuropathic and back pain \( (k = 1).^{53} \) Across all included studies, the mean ages of these patients ranged from 44 to 60 years. The percentage of females enrolled ranged from 35% to 62.5%.\(^50\) The duration of pain prior to study enrollment was reported by only five studies, but where reported, ranged from 19 months\(^51\) to 10.3 years.\(^14\)

Patient withdrawal rates were generally high in the included studies. For example, although the RCTs of oral opioids were short-term (six weeks or less),\(^41,43,54\) drop-out rates were high and reduced the number of patients eligible to continue in the open-label phase.\(^42\) In the four studies of this type, 918 patients enrolled in the RCT portions of these studies, 450 (49.0%) participated in open-label extension studies, and 162 (17.6%) reported pain scores usable in the analysis (at follow-up times ranging from 6 to 18 months). At six months, withdrawal rates were lowest in intrathecal studies (range 0% to 17%).\(^45\) Reported six-month withdrawal rates were higher in oral opioid studies (range 35% to 86%) and transdermal opioid studies (33% and 44%). At one year, a similar trend was observed (oral: 52% to 88%; transdermal: 43% to 67%; intrathecal: 0% to 33%).\(^39,38\) In intrathecal studies, withdrawal rates ranged from 0% to 33% at months 18 and 24, and from 61% to 73% at 36 months. The longest reported follow-up time of any of the intrathecal studies was 48 months, at which time 95% withdrawal was observed.\(^40\) Only one study each of oral and transdermal administration presented data beyond one year, with both reporting 88% withdrawal at 18 months.\(^39,34\) In the

---

**Fig. 1. Flow diagram of study selection.**
<table>
<thead>
<tr>
<th>Mode of Administration</th>
<th>Reference</th>
<th>Opioid</th>
<th>Predominant CNCP Type(s)</th>
<th>n Enrolled</th>
<th>Outcomes Used in Evidence Synthesis</th>
<th>Continuous/Categorical Pain</th>
<th>≥30% Relief</th>
<th>≥50% Relief</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral</td>
<td>Allan et al. 14</td>
<td>Morphine&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Low back pain</td>
<td>342</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td></td>
<td>Caldwell et al. 12</td>
<td>Morphine&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Osteoarthritis</td>
<td>(295)</td>
<td>181</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td></td>
<td>Harati et al. 41</td>
<td>Tramadol</td>
<td>Diabetic neuropathy</td>
<td>(131)</td>
<td>117</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td></td>
<td>Fredheim et al. 48</td>
<td>Methadone</td>
<td>Low back pain</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>McIlwain and Ahdieh 43</td>
<td>Extended-release oxymorphone</td>
<td>Osteoarthritis</td>
<td>(491)</td>
<td>133</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Roth et al. 54</td>
<td>Controlled-release oxycodone</td>
<td>Osteoarthritis</td>
<td>(133)</td>
<td>106</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zenz et al. 55</td>
<td>Dihydrocodeine&lt;sup&gt;a&lt;/sup&gt;, buprenorphine, or morphine&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Neuropathic or back pain</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transdermal</td>
<td>Allan et al. 14</td>
<td>Fentanyl</td>
<td>Low back pain</td>
<td>338</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td></td>
<td>Milligan et al. 38</td>
<td>Fentanyl</td>
<td>Unspecified</td>
<td>532</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td></td>
<td>Mystakidou et al. 39</td>
<td>Fentanyl</td>
<td>Unspecified</td>
<td>329</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>Intrathecal</td>
<td>Anderson and Burchiel 45</td>
<td>Morphine</td>
<td>Nociceptive and neuropathic</td>
<td>30</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td></td>
<td>Anderson et al. 46</td>
<td>Morphine</td>
<td>Failed back surgery syndrome (FBSS)</td>
<td>27</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td></td>
<td>Angel et al. 47</td>
<td>Morphine</td>
<td>Failed back syndrome</td>
<td>11</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td></td>
<td>Hassenbusch et al. 40</td>
<td>Morphine or sufentanil citrate</td>
<td>Neuropathic</td>
<td>18</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td></td>
<td>Kanoff 49</td>
<td>Morphine</td>
<td>Mixed</td>
<td>15</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td></td>
<td>Mironer and Tollison 50</td>
<td>Methadone</td>
<td>FBSS</td>
<td>24</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td></td>
<td>Rainov et al. 51</td>
<td>Morphine (with bupivacaine, clonidine, or midazolam)</td>
<td>FBSS</td>
<td>26</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td></td>
<td>Tutak and Doleys 52</td>
<td>Morphine or fentanyl</td>
<td>Unspecified</td>
<td>26</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
</tbody>
</table>

*Not sustained release.
<sup>a</sup>Sustained release.
<sup>b</sup>Not analyzed because number of patients at follow-up times not reported.
<sup>c</sup>N in parentheses denotes number of patients randomized in original RCT; second number is that enrolled in open-label extension.
<sup>d</sup>Not meta-analyzed because reported units are statistically incompatible with the three other studies meeting inclusion criteria.
<sup>e</sup>Not analyzed because data were reported for fewer than ten patients at follow-up times.
<sup>f</sup>Not analyzed because instrument used not validated.
transdermal study, the withdrawal rate was 95% by 48 months (the longest reported follow-up time). Insufficient data were provided to statistically analyze prognostic factors for withdrawal for any reason.

Evidence Synthesis

The number of studies and enrolled patients in all analyses in this systematic review are presented in Table 2. Because patient withdrawal rates were so high, we first analyzed withdrawal rates for the two most common reasons cited for leaving a study: adverse events and insufficient pain relief. Therefore, the patients who reported long-term pain outcomes represent only a subset of the patients initially enrolled in the studies. For the four studies that were open-label extensions of RCTs, patients who withdrew from the active treatment group during the RCT portion were included in our analyses.

Withdrawal from Clinical Study

Due to Adverse Events

Among individuals with CNCP taking oral opioids, approximately one-third did not continue long-term treatment (follow-up time range 6–18 months) because of intolerable adverse effects (32.5% [95% CI, 26.1%–39.6%]). Intrathecal studies reported the lowest drop out rates due to adverse events at longest follow up (6–mean of 29 months): 6.3% (95% CI, 2.9%–13.1%). A small amount of funnel plot asymmetry attributed to two missing studies was detected among the intrathecal studies. Adjusted by the trim-and-fill method, the event rate is slightly higher, 8.0% (95% CI, 4.1–15.2%). Although we did not consider this to be a substantial difference, we acknowledge the small number of studies analyzed (k = 6) limits the power of the statistic to detect an effect. Transdermal trials reported intermediate adverse event-related drop-out rates at longest follow-up (range 12–48 months), at 17.5% (95% CI, 6.5%–39.0%). Forest plots of these findings are shown in Fig. 2. No substantial heterogeneity was detected among intrathecal studies ($\hat{I}^2 < 0.001\%$); however, substantial heterogeneity was detected for both oral ($\hat{I}^2 = 82.9\%$) and transdermal trials ($\hat{I}^2 = 98.2\%$). In neither case could this heterogeneity be explained by differences across studies in the longest follow-up time, which ranged widely.

Adverse Events

Given the high rate of withdrawal from the clinical trials due to adverse events, we examined adverse events reported by all included studies. The variability in thresholds for reporting adverse events, failure to report the absence of unobserved but potential adverse
events in some studies, inconsistent reporting or use of definitions of events/effects, and absence of control groups precluded our ability to pool data for meta-analysis.

The most commonly reported adverse events included gastrointestinal effects (i.e., constipation, nausea, and dyspepsia), headache, fatigue/lethargy/somnolence, and urinary complications (i.e., retention, hesitancy, and “disturbance”). Adverse events unique to the intrathecal mode of administration included pump and catheter malfunctions and malpositioning, surgical complications, and postsurgical complications. The percentage of patients whose device complications required reoperation was as high as 35%. These studies included a total of 2,042 patients (not including an additional 259 patients in the original six-week RCT preceding the extension reported by McIlwain and Ahdieh, for whom addiction rates were not reported). Reporting of addiction and abuse was inconsistent because not all studies reported this outcome. Among studies that did report addiction and abuse, which diagnostic criteria were used, and whether addiction/abuse monitoring was prospectively planned were generally unclear. Therefore, we included any data potentially suggesting addiction or abuse, although we recognize that this could lead to overestimation of actual addiction and/or abuse rates. Of the 2,042 patients in the 17 studies that reported signs of addiction and/or addictive behaviors, only one patient was reported as having possibly developed opioid addiction (in Anderson 1999, in the form of “drug-seeking behavior”). Therefore, in these seven studies the overall addiction rate was 0.042%. For the other 10 studies (n = 680) that met general inclusion criteria, it is unclear whether addiction was not mentioned because it was not an outcome of interest, or because it was not observed. Milligan and colleagues reported three cases of opioid abuse (defined as taking opioids in inappropriate doses, or inappropriately with other drugs) among the 532 patients who enrolled in this 13-month study. Although drug abuse is not mentioned in the article describing their original six-week RCT (an additional 259 treated with opioids), McIlwain and Ahdieh reported that no opioid abuse was observed in any of the patients (n = 153) enrolled in their 13-month, open-label extension study. If the original RCT population from McIlwain and Ahdieh is not included, the total abuse rate from these two studies is 3/685 (0.43%). Given the widespread concern about abuse, it is unclear why it was not mentioned in the other 15 studies. If none occurred, the rate would be lower.

Withdrawal from Clinical Study Due to Insufficient Pain Relief

As with withdrawal from clinical studies due to adverse events, the highest summary event rate of discontinuation due to inadequate pain relief was observed in long-term (range 6–18 months) studies of oral opioids (11.9% [95% CI, 7.8%–17.7%]). The next highest summary discontinuation rate occurred among individuals enrolled in long-term (range 6 months to a mean of 29 months) intrathecal opioid studies (10.5% [95% CI, 3.5%–27.4%]). The lowest summary discontinuation rate due to insufficient pain control was observed in long-term (range 12–48 months) transdermal opioid studies; however, CIs overlap with those of oral opioids (5.8% [95% CI, 4.2%–7.3%]). Forest plots showing these effect sizes are shown in Fig. 3. As with withdrawal from clinical studies due to adverse events, substantial heterogeneity was not found for withdrawal due to insufficient pain relief among the studies of transdermal opioid use (I² = 48.9%), but was detected in oral opioid (I² = 85.5%) and intrathecal (I² = 63.3%) studies. This heterogeneity could not be explained by differences across studies in the widely ranging longest follow-up times.

Pain

We could not formulate evidence-based conclusions on long-term efficacy of transdermal opioids for CNCP because only two studies contributed data to this analysis. For oral and intrathecal administration of opioids, data were sufficient for analyzing long-term pain score data. The included studies assessed pain severity with a visual analog scale (VAS) (0–10 or 0–100) or Likert scales (0–3 or 0–4). We combined the studies in meta-analysis and found
that patients treated with long-term (range 6–mean of 29 months) intrathecal opioids had an approximately 38.0% mean reduction in pain scores (SMD = 1.33 [95% CI, 0.97–1.69]), and patients treated with long-term (range 6–18 months) oral opioids had approximately a 63.4% mean reduction in pain scores (SMD = 1.99 [95% CI, 1.17–2.80]). Substantial heterogeneity was observed in both oral studies ($I^2 = 86.6\%$) and intrathecal studies ($I^2 = 51.3\%$). The heterogeneity of oral and intrathecal opioid studies could not be resolved using meta-regression by follow-up time.

The summary effect estimate of pain relief from oral and intrathecal opioids was not robust to sensitivity analyses manipulating baseline standard deviations, and the summary effect size of intrathecal opioids was also not robust to changes in the correlation coefficient. These findings suggest that the variability in pain scores of patients at baseline may affect the amount of pain relief attained. Due to this lack of robustness upon sensitivity analysis and unexplained heterogeneity, the quantitative estimates of the amount of pain relief associated with opioid therapy may be unstable. However, it was still evident that long-term opioids are associated with some degree of pain relief because even the smallest summary effect sizes yielded by the sensitivity analyses remained large ($\geq 0.853$ SMD for oral opioids, and $\geq 0.917$ for intrathecal opioids) and statistically significant ($P < 0.001$).

Among patients treated with intrathecal opioids, an estimated 48.9% (95% CI, 35.7%–62.3%) of patients had at least a 30% reduction in pain and an estimated 40.0% (95% CI, 23.0%–59.8%) had at least a 50% reduction in pain by the longest follow-up time. In both bodies of evidence, longest follow-up ranged from six months to a mean of 29 months. Findings from studies reporting at least a 30% reduction in pain were not substantially heterogenous ($I^2 < 0.001\%$) and no funnel plot asymmetry was detected using the trim-and-fill method. Heterogeneity ($I^2 = 69.8\%$) was detected among studies reporting at least a 50% reduction in pain, but could not be explained by duration of treatment time.

**Conclusions and Discussion**

Many patients in the included studies were so dissatisfied with adverse events or insufficient pain relief from opioids that they withdrew from the studies. For patients able to continue on opioids, evidence (albeit weak) suggests that their pain scores were lower than before therapy began and that this relief could be maintained long-term (≥6 months). However, data describing long-term safety and efficacy of opioids for CNCP are limited in terms of quantity and quality, precluding the formation of evidence-based conclusions.
supported by strong qualitative or stable quantitative evidence. An evidence base of low quality provides only weak evidence from which to draw qualitative conclusions and only low-stability evidence from which to draw quantitative conclusions. Some of the quantitative estimates were not robust upon sensitivity analyses, which means that an estimate of the size of a treatment effect cannot be accurately estimated with the currently available evidence. These quality ratings indicate that the evidence supporting our conclusions is highly subject to change, and that the likelihood is high that findings of future studies may overturn these conclusions. There is insufficient data to determine whether transdermal opioids maintain long-term pain relief. These findings are summarized in Table 3.

The generalizability of findings of these studies to “real-world” patients with chronic noncancer pain in general is unclear. Prescreening of patients in intrathecal studies for opioid responsiveness prior to commencement of treatment may limit the generalizability of the findings of these studies to patients who are not prescreened (neither the oral nor the transdermal studies had comparable prescreening requirements). In addition, the follow-up data on pain scores reported reflect outcomes for only a portion of the patients initially enrolled: many patients withdrew from studies that administered oral or transdermal opioids before they reported efficacy outcomes, and the majority of those studies did not report intent-to-treat outcome data. To investigate the importance of the possibility that the baseline characteristics of completers may differ in important ways from the baseline characteristics of the intent-to-treat population, we explored the effect of varying some input in sensitivity analyses. When baseline pain level standard deviations were changed, the summary effect size for oral opioids and intrathecal studies was not robust. When the assumed correlation coefficient was varied, the summary effect size of pain relief from intrathecal opioids was not robust either.

Whether publication bias or other possible causes of missing studies (and, therefore, missing patients and outcomes data) influenced our results was not demonstrated by a funnel plot or trim-and-fill analysis. However, there may be a particular risk of publication bias in uncontrolled time-series studies, as journals may be less likely to accept studies of this design, investigators may be less likely to submit them for publication (given the lesser financial investment in conducting them), and no clinical trial registry for uncontrolled studies is currently in widespread use.

Among outcomes for which qualitative conclusions were possible, substantial unexplained heterogeneity limited the number of quantitative conclusions we could make. Heterogeneity was encountered most frequently in the evidence bases for oral and transdermal opioid studies. Intrathecal evidence bases were generally not substantially heterogenous, perhaps due to the more rigorous prescreening for study enrollment and automation of dose administration. We postulated that at least some of the heterogeneity might be explained by differences in duration of treatment; however, meta-regression did not support this possibility for any outcome. Additional explorations of heterogeneity were not undertaken, given the limited number of available studies, and poor reporting. Whether the study was sponsored by drug or device manufacturers with a financial interest in the outcome would also be an interesting covariate; however, in many of these studies, financial sponsorship (or lack thereof) was not reported. Researchers whose studies are funded by parties without a conflict of interest in the findings should clearly state so in their articles. When more data become available, meaningful covariates for investigation of heterogeneity would include specific opioid administered, mode of administration, flexibility in dosing, use of adjuvant therapies, and etiology and baseline severity of pain. For the investigation of patient characteristics, publication of patient-level data would be even more desirable, because it would increase statistical power for meta-analysis and may be less influenced by study characteristics. In addition, the high attrition rates call into question whether the patients reporting pain relief outcomes are representative of the patient population originally enrolled in the study. We found no studies that attempted to identify prognostic factors for drop out. Studies designed to examine patient and treatment-related factors predicting long-term success with opioid therapy would be extremely useful.
Table 3
Summary of Findings

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Mode of Administration</th>
<th>K</th>
<th>N</th>
<th>$I^2$</th>
<th>Meta-analysis Unit</th>
<th>Fixed Effects Meta-analyses (95% CI)</th>
<th>Funnel Plot Asymmetry?</th>
<th>Random Effects Meta-analyses (95% CI)</th>
<th>Quantitatively Robust to All Sensitivity Analyses?</th>
<th>P-Value?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Withdrawal due to adverse events</td>
<td>Oral</td>
<td>6</td>
<td>1,404</td>
<td>82.9%</td>
<td>Event rate</td>
<td>32.5% (26.1%–39.6%)</td>
<td>Yes</td>
<td>No</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Transdermal</td>
<td>3</td>
<td>1,399</td>
<td>98.2%</td>
<td>Event rate</td>
<td>17.5% (6.5%–39.0%)</td>
<td>No</td>
<td>No</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intrathecal</td>
<td>6</td>
<td>154</td>
<td>0.00%</td>
<td>Event rate</td>
<td>6.3% (2.9%–13.1%)</td>
<td>Yes</td>
<td>Yes</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Withdrawal due to insufficient pain relief</td>
<td>Oral</td>
<td>7</td>
<td>1,504</td>
<td>85.5%</td>
<td>Event rate</td>
<td>11.9% (7.8%–17.7%)</td>
<td>Yes</td>
<td>Yes</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Transdermal</td>
<td>3</td>
<td>1,399</td>
<td>48.9%</td>
<td>Event rate</td>
<td>5.8% (4.2%–7.3%)</td>
<td>Yes</td>
<td>Yes</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intrathecal</td>
<td>5</td>
<td>110</td>
<td>63.3%</td>
<td>Event rate</td>
<td>10.5% (3.5%–27.4%)</td>
<td>Yes</td>
<td>Yes</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Pain</td>
<td>Oral</td>
<td>3</td>
<td>162</td>
<td>86.6%</td>
<td>SMD on VAS (0–100), Likert (0–3; 0–4)</td>
<td>1.99 (1.17–2.80)</td>
<td>No</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Transdermal</td>
<td>1</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intrathecal</td>
<td>6</td>
<td>125</td>
<td>51.3%</td>
<td>SMD on VAS (0–10; 0–100)</td>
<td>1.33 (0.97–1.69)</td>
<td>No</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥30% Pain relief</td>
<td>Oral</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Transdermal</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intrathecal</td>
<td>3</td>
<td>55</td>
<td>0.00%</td>
<td>Event rate</td>
<td>48.9% (35.7%–62.3%)</td>
<td>No</td>
<td>48.9% (35.7%–62.3%)</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>≥50% Pain relief</td>
<td>Oral</td>
<td>2</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Transdermal</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intrathecal</td>
<td>5</td>
<td>104</td>
<td>69.8%</td>
<td>Event rate</td>
<td>40.0% (23.0%–59.8%)</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviations:** $K = \text{studies}$, $N = \text{patients}$, CI = Confidence Interval, SMD = Standardized Mean Difference, VAS = Visual Analog Scale.

*Denotes that this statistic could not be computed for that outcome because of the presence of heterogeneity.

$^b$P-values not calculated for dichotomous outcomes, given absence of control groups.

$^c$In accordance with a priori protocols, evidence was not pooled when the evidence base was composed of fewer than three studies.
Opioid addiction is of great concern medically and societally. Fear that an individual with CNCP may develop dependence on the drug during long-term administration represents a potential barrier to treatment. However, the rate of observed signs of opioid addiction was extremely low (one case from a total of 2,042 enrollees) in the body of evidence considered in this review. This rate may be even lower if no addictive behaviors occurred among the 683 patients enrolled in the seven studies that did not mention addiction rates at all. At less than 1%, the addiction rate is comparable to the findings of Porter and Jick's oft-cited 1980 retrospective chart review of a general population of 11,882 hospitalized patients who were treated with narcotics for any duration.51 Opioid abuse was also rare (conservatively, 3/685). These rates do not support the contention that potential iatrogenic opioid addiction and abuse should limit therapy for well-selected patients.

The low rates of addiction and abuse reported in these studies may only be generalizable to patients without a history of addictive/abusive behaviors. Twelve of the studies in this review screened out patients with a history of opioid or substance addiction or abuse, and two more screened for unspecified psychological contraindications (which may have included addictive/abusive history). Additional prospective studies are needed to address iatrogenic addiction/abuse in other patient populations. Given the importance of abuse and addiction issues, all studies on the efficacy and safety of opioids should prospectively collect data on abuse and addiction using validated diagnostic criteria.

Given the complexity of definitively diagnosing opioid addiction (see Ballentyne56) and in the interest of capturing the overall effect of opioid therapy on quality of life, we sought to analyze health-related quality-of-life outcomes in this review. However, insufficient data were available to enable quantitative analysis for any mode of administration, according to our protocols.

Additional Thoughts and Discussion from the Pain Research Community

The Mayday Fund (web site: http://www.painandhealth.org/mayday/mayday-home.html), a foundation sponsoring research on the treatment of pain, asked ECRI Institute to perform a review of efficacy and safety of opioids in addition to attitudinal and legal barriers to treatment with opioids. The Milbank Memorial Fund collaborated in that effort by inviting an international group of pain researchers from clinical, social, and legal disciplines to help formulate key questions for the systematic review. When the original work was completed, representatives from these organizations and additional pain researchers were asked to critique the review and to make recommendations for future research.

Many ethical and practical difficulties are associated with long-term RCTs for CNCP. Allowing rescue opioids (with the amount and frequency of dosage as an outcome) in placebo or nonopioid groups may circumvent ethical problems of withholding analgesics from individuals with CNCP. While RCTs would be ideal, adequately powered and well-designed prospective cohort studies can provide very useful information.57–59 Ideally, a study protocol would specify uniform diagnostic and data collection criteria (e.g., pain etiology, drugs prescribed, dosing regimens), and would mimic clinical practice (e.g., drug combinations could be used, drug changes could be made, drug dosage could be titrated slowly and adjusted, and adverse effects could be aggressively managed). Reasons for discontinuing opioid therapy must be carefully documented. Data must also be collected on additional long-term effects of opioids that are currently not well understood, including androgen deficiency and changes in immune function. Practical/pragmatic RCTs that collect this information would be even more useful, but costs might be prohibitive.60

Existing regulatory and clinical databases could be used to look for long-term dose-dependent efficacy and adverse effects. This could include data from the European Medicines Evaluation Agency (EMEA), Food and Drug Administration (FDA), Australia’s Therapeutic Goods Administration, and other regulatory agencies. Studies on prescribing of opioids for noncancer pain performed using pharmacy and clinical data from the U.S. Veterans Health-care Administration are examples of retrospective analyses.61–64 Prospective data collection in such integrated systems could provide more
reliable estimates of long-term efficacy and safety. Data from more broadly representative populations, however, would be important.

There is agreement that the “best possible evidence” for therapeutic efficacy is provided by systematic reviews of high-quality RCTs. This systematic review has shown that such evidence is currently lacking for the efficacy and safety of long-term opioid therapy in chronic noncancer pain. As noted above, there are many practical difficulties in undertaking long-term RCTs in this patient population and there is increasing recognition in the evidence-based medicine community that other study designs (e.g., practical clinical trials) may be necessary for answering certain clinical questions. Although additional prospective controlled trials of long-term opioid treatment would provide valuable information, a prospective cohort study of suitable size has the potential to provide practical answers to key efficacy and safety questions.

Acknowledgments

The authors thank Vivian H. Coates, MBA, and Jeffrey Lerner, PhD, of ECRI Institute for overseeing this project from inception to completion, Eileen Erinoff of ECRI Institute for designing and conducting the systematic literature searches, and Clarisse Akafomo, MD, MPH, of ECRI for assistance in data acquisition. They also thank Philip J. Wiffen, MSc, MRPharmS, of the Cochrane Pain, Palliative and Supportive Care Review Group, Pain Research Unit, for contributions to the conception and design of the review and for reviewing earlier drafts of it, and Charles M. Turkelson, PhD for his contributions to the original project. They appreciate the members of the pain research, legal, social science, and regulatory community for their valuable insights during the earlier systematic review, and thank Dr. James Campbell, Dr. Howard Field, and Dr. Rollin Gallagher for a number of helpful comments on the scientific and clinical context for the review. The authors also wish to thank the following people who reviewed and provided comments on the original draft report: Steven I. Altchuler, PhD, MD, Daniel B. Carr, MD, Roger Chou, MD, MPH, Michael Clarke, DPhil, Wilson M. Compton, MD, MPE, Richard A. Denisso, MD, MPH, Kathleen M. Foley, MD, James Giglio, Geoff Gourlay, PhD, Marcia Meldrum, PhD, Robert Meyer, MD, Charles D. Ponte, PharmD, CDE, BGPS, Russell Portenoy, MD, Andrew S.C. Rice, MD, Linda Simoni-Wastila, PhD, Philip J. Wiffen, MSc, MRPharmS, and James Zaczyn, PhD, and those who participated in teleconferences on setting the research agenda: Philip J. Wiffen, MSc, MRPharmS, Chair, Mitchell Max, MD, Elan Eisenberg, MD, Ian Gilron, MD, MSc, FRCP (C), Geoff Gourlay, PhD, Art Lipman, PhD, Marcia Meldrum, PhD, Kathleen Foley, MD, Roger Goucke, FANZCA., FFFPMANZCA, FACP, and Henry McQuay, MD. The full report of this review will be published in the Cochrane Library.

References


