Big Data and Pain
Review

OPEN

PAIN REPORTS

Real-world data and evidence in pain research: a qualitative systematic review of methods in current practice

Jan Vollert[a,b,c,d,*], Bethea A. Kleykamp[e], John T. Farrar[f], Ian Gilron[g], David Hohenschurz-Schmidt[a], Robert D. Kerns[h], Sean Mackey[i], John D. Markman[j], Michael P. McDermott[k], Andrew S.C. Rice[a], Dennis C. Turk[l], Ajay D. Wasan[m], Robert H. Dworkin[n]

Abstract

The use of routinely collected health data (real-world data, RWD) to generate real-world evidence (RWE) for research purposes is a growing field. Computerized search methods, large electronic databases, and the development of novel statistical methods allow for valid analysis of data outside its primary clinical purpose. Here, we systematically reviewed the methodology used for RWE studies in pain research. We searched 3 databases (PubMed, EMBASE, and Web of Science) for studies using retrospective data sources comparing multiple groups or treatments. The protocol was registered under the DOI:10.17605/OSF.IO/KGVRM. A total of 65 studies were included. Of those, only 4 compared pharmacological interventions, whereas 49 investigated differences in surgical procedures, with the remaining studying alternative or psychological interventions or epidemiological factors. Most 39 studies reported significant results in their primary comparison, and an additional 12 reported comparable effectiveness. Fifty-eight studies used propensity scores to account for group differences, 38 of them using 1:1 case:control matching. Only 17 of 65 studies provided sensitivity analyses to show robustness of their findings, and only 4 studies provided links to publicly accessible protocols. RWE is a relevant construct that can provide evidence complementary to randomized controlled trials (RCTs), especially in scenarios where RCTs are difficult to conduct. The high proportion of studies reporting significant differences between groups or comparable effectiveness could imply a relevant degree of publication bias. RWD provides a potentially important resource to expand high-quality evidence beyond clinical trials, but rigorous quality standards need to be set to maximize the validity of RWE studies.

1. Introduction

Real-world evidence (RWE) is defined as "information on health care that is derived from multiple sources outside typical clinical research settings".[77] It is a rapidly expanding field of interest: technological advances of the past decades, especially the wide availability of large databases and computational methods to search them, have enabled secondary research use of data not initially collected for this purpose. It has even been suggested that RWE studies can—in limited settings—serve as a complement to randomized controlled trials (RCTs).[17] Although increased value

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

[a] *Pain Research, Department of Surgery and Cancer, Imperial College London, London, United Kingdom,* [b] *Division of Neurological Pain Research and Therapy, Department of Neurology, University Hospital of Schleswig-Holstein, Campus Kiel, Germany,* [c] *Department of Anaesthesiology, Intensive Care and Pain Medicine, University Hospital Muenster, Muenster, Germany,* [d] *Neurophysiology, Mannheim Center of Translational Neuroscience (MCTN), Medical Faculty Mannheim, Heidelberg University, Heidelberg, Germany,* [e] *BAK and Associates, LLC, Baltimore, MD, USA,* [f] *Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, Philadelphia, PA, USA,* [g] *Departments of Anesthesiology & Perioperative Medicine, Biomedical & Molecular, Sciences, Centre for Neuroscience Studies, and School of Policy Studies, Queen's University and Kingston Health Sciences Centre, Kingston, ON, Canada,* [h] *Departments of Psychiatry, Neurology and Psychology, Yale University, New Haven, CT, USA,* [i] *Department of Anesthesiology, Perioperative and Pain Medicine, Stanford University School of Medicine, Stanford, CA, USA,* [j] *Department of Neurosurgery, University of Rochester, Rochester, NY, USA,* [k] *Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY, USA,* [l] *Chair of Anesthesiology & Pain Research, UW Medicine, Department of Anesthesiology & Pain Medicine, University of Washington, WA, USA,* [m] *Department of Anesthesiology and Perioperative Medicine, University of Pittsburgh, Pittsburgh, PA, USA,* [n] *Departments of Anesthesiology and Perioperative Medicine, Neurology, and Psychiatry, University of Rochester, Rochester, NY, USA*

**Corresponding author. Address: Pain Research, Department of Surgery and Cancer, Imperial College London, Chelsea and Westminster Campus, 369 Fulham Rd, London SW10 9NH, United Kingdom. Tel.: +44 (0)20 3315 8816; fax: +44 (0)20 3315 51525. E-mail address: j.vollert@imperial.ac.uk (J. Vollert).*

PR9 8 (2023) e1057

http://dx.doi.org/10.1097/PR9.0000000000001057

of routine data would be generally welcomed, the lack of randomization along with often limited data quality and quality control (eg, incomplete data, incorrect data), and potential confounding that can have large effects, emphasize that valid RWE can only be drawn from well-designed, carefully conducted studies using well-curated data and accounting for data quality issues.[86]

As part of an ACTTION (Analgesic, Anesthetic, and Addiction Clinical Trial Translations, Innovations, Opportunities, and Networks) IMMPACT (Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials) effort, this qualitative systematic review aims to identify approaches used to assess effectiveness of pain treatments in RWE studies and to provide an overview of methods used to date to design, conduct, and analyze RWE studies in pain research. Because this is a review of methods, we explicitly do *not* aim to assess results of these studies or perform a meta-analysis thereof. We focus on design of studies using retrospective data comparing 2 or more groups, to focus on the challenging aspects of retrospective design and how providing valid causal inferences about the interventions in the setting of such noncomparability can be made despite no randomization. Our discussion excludes 2 large fields of potential RWE studies: prospective trials, which was covered previously in our work on "pragmatic trials",[33] and single-arm cohort studies because they contribute to a separate trail of evidence.

2. Methods

2.1. Protocol and deviations

We registered a protocol for this review under the DOI 10.17605/OSF.IO/KGVRM on the Open Science Framework. There were no major deviations from the protocol.

2.2. Search strategy

For our systematic search, PubMed, EMBASE, and Web of Science were queried combining various terms from 3 domains: data sources, analytic methods, and pain research. At least one term of each domain had to be included. Additional studies were included as solicited from the author group.

Thus, the general search string was as follows:

("Real-world data" OR "Claims data" OR "Billing data" OR "clinical data" OR "pharmacy data" OR "Administrative data" OR "Electronic medical records" OR "Electronic health records" OR "Health system" OR "Registry" OR "Insurance" OR "Third-party payer" OR "retrospective cohort") AND

("Real-world evidence" OR "Causal inference" OR "Propensity score" OR "Predictive model" OR "Confounding factors" OR "Time-varying confounding" OR "Risk set matching" OR "Path analysis") AND "Pain." The search string was optimized for each of the 3 databases.

Both title and abstract and full-text screening were performed in duplicate (by JV and BK). Disagreements between reviewers were mediated by a consensus discussion. Data extraction was done in singular (by JV).

Duplicates were identified before screening, based on PubMed ID, DOI, and title, journal, and author list, using automated methods. Screening was based on abstracts only and aimed for sensitivity over specificity (ie, excluding only articles that are clearly out of scope) at this stage. During full-text screening and annotation, secondary exclusion was conducted for articles included at abstract screening while not fitting the inclusion criteria on full-text stage.

Additional studies identified by search of the reference list of included studies or solicited by the author group were included if not found in the systematic search.

2.3. Inclusion and exclusion criteria

All full-text original research on real-world data and evidence on effectiveness or comparative or comparable effectiveness of treatments where pain was the primary outcome criterion were included. Studies with pain as a secondary outcome were included if pain was central to the aim of the study, ie, if (1) the primary outcome was a composite outcome including pain or (2) pain was a necessary inclusion criterion. Only studies comparing 2 or more groups were included. Reviews, conference proceedings, book chapters, and abstracts were excluded. Studies focussing on other health aspects and only peripherally reporting pain were excluded. Articles for which no full texts could be retrieved through online access, interlibrary loan, or by contacting authors directly were excluded. Furthermore, articles written in languages with which the authors were not fluent and for which no native or fluent speaker could be recruited through the wider network of the authors were excluded.

2.4. Extraction items

Extraction was focussed on methodological items and general study characteristics, such as condition studied, pain type (nociceptive/nociplastic/neuropathic/postsurgery and acute/chronic), use of hospital records vs registry data, single or multicentre data, number of patients screened and included, and equal or unequal group size. We extracted statistical design aspects specifically focussing on mention of propensity scores, use of multiple regression (outside of propensity scores), instrumental variables, sensitivity analysis, and mention of any other inference methods.

3. Results

A total of 536 studies were screened (**Fig. 1** for inclusion flowchart). Based on our inclusion/exclusion criteria, through full-text screening, we identified 61 studies for inclusion.[1,3,6–8,11,13–15,21,22,24–31,34–41,44–46,49–61,63–67,69–73,75,78–80,82–84,87] We included no additional studies through reference search and 4 additional studies solicited from coauthors that were not otherwise included,[19,23,43,81] resulting in a total sample of 65 studies, all of which were published in the English language; hence, no studies were excluded based on language.

A list of all extraction items with summary statistics can be found in **Table 1**. The studies identified were remarkably similar: 49 of 65 reported on surgical interventions; of the remaining16; 3 studied alternative treatments[8,25,29]; 4 studied epidemiological risk factors like obesity,[22] old age,[21] opioid abuse,[23] and smoking[43,59]; only 4 studies investigated pharmacological interventions[28,67,71,81]; the remaining studied patient–professional–interaction or behavioural medicine,[19,66] radiotherapy,[49] implants,[69] or compared 2 disease progressions under routine care.[26] Of the 49 studies on surgical interventions, 23 focussed on postsurgical pain, with 3 investigating chronic postsurgical pain.[11, 65, 80]

Very few studies mentioned a study registration[13] or provided a link or identifier of a publicly accessible protocol in a central register.[46,51,53,63] The majority (39 of 65) reported significant differences in their primary group comparison; of the remaining

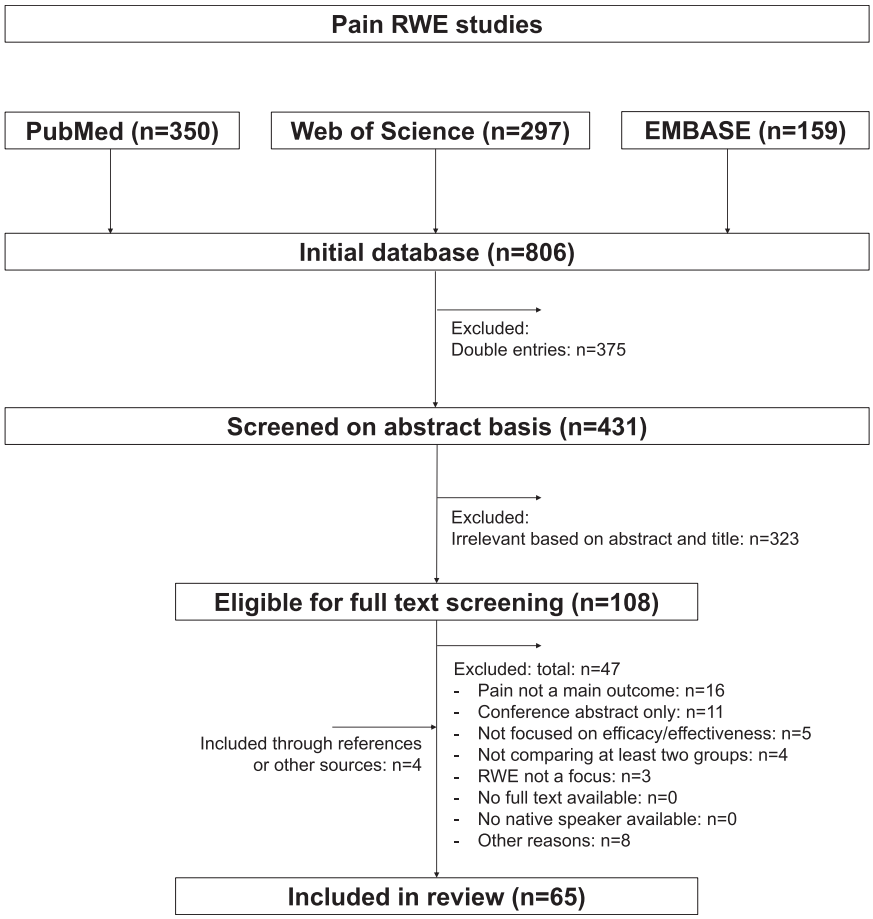

Figure 1. PRISMA flowchart.

26, 12 reported comparable effectiveness based on nonsignificant *P* values.[3,6,28,30,31,34,39,49,75,78–80]

A majority of 58 studies used propensity scores as a means of adjusting for potential confounders, with only 7 studies not reporting use of propensity scores.[8,21,23,28,58,59,81] A propensity score is the probability of being assigned to a particular intervention group given a set of potentially confounding baseline variables. It reduces the possibly large set of patient or clinical characteristics (some or all of which could confound the relationship under study, eg, age, social status, ethnicity, sex) to a single variable (or $k - 1$ variables if there are k intervention groups). The propensity score(s) can be used in various ways to adjust for potential confounding without having to explicitly include all of these confounders in the statistical model,[12] including propensity score matching, stratification, and inverse probability weighting.[9] Of the studies using propensity scores, most used propensity score matching, in which for each case, exactly one control (with a similar propensity score to that of the case) is drawn from a usually larger pool of potential controls. The minority of studies (20 of 58) used propensity scores in regression analysis with unequal group sizes.[1,7,19,25,29,30,34,36,37,43,46,61,65,69,71,80,82–84,87] Multiple regression techniques were used, instead or in addition to propensity scores, by 13 studies.[8,19,23,29,43,53,58,59,61,71,81,83,84] We could find no mention of instrumental variables in the studies included. Only 17 of 65 studies provided sensitivity analyses to demonstrate robustness of their findings to violation of assumptions in the primary model.[1,11,14,23–25,27,29,30,52,56,66,67,71,79–81]

Roughly half of the studies (28/65) included hospital or outpatient records instead of publicly accessible repositories. Of these, 6[15,28,36,54,69,82] were multicentre studies, the remaining 22 single-centre studies, whereas all 33 studies using publicly accessible repositories were using multicentre data. For multicentre studies, if the number of participating centres was given, it ranged between 2[15] and 524.[38–41] Study sample size differed significantly, from 56[13] to moer than 300,000[25] (median: 560) patients included. Screening of record numbers ranged between 56[13] and more than 5 million records[27] (median: 1,828). Included patient number was higher for registry vs hospital record-based studies (median: n = 1,741 vs n = 170), multicentre (median: n = 1,397) vs single-centre studies (median: n = 172), and studies without 1:1 case matching (unequal group sizes: median n = 1,452, equal group sizes: median n = 375). This picture was similar for screened records.

Surprisingly, in this systematic search trying to identify real-word data and evidence studies, only 6 reports used the term "real-world data" in the full text,[26,28,29,71,79,87] only one used the term "real-world evidence",[28] and only 2 studies used the term "real-world" in the title.[28,71] Despite no time filter in our search, most studies were conducted recently, with only one published in 2009,[66] 7 in 2014 to 16, 31 in 2017 to 2019, and 26 since 2020.

4. Discussion

In this review, we summarize the current practice of studies using real-world data in pain research, focusing on studies comparing at least 2 groups.

Table 1

Extraction items and summary.

Year published	2009–2021
Condition studied (free text)	
Pain type	11 neuropathic, 28 nociceptive, 22 post-surgical
Chronicity	35 chronic, 26 acute
Pain description (free text)	
Data source	33 registry, 28 hospital records
Data source description (free text)	
Single center?	22 yes
n Participating centers (can be left blank if single center = yes)	Median: 36
n Patients (total screened)	Median: 1828
n Patients (total included)	Median: 560
Groups of equal size?	37 yes
n group 1	Median: 195
n group 2 (if sizes are equal, leave blank)	Median: 196
n group 3 (leave blank if only 2 groups)	Median: 462
n group 4 (leave blank if 3 or 2 groups)	Median: 108
Use of propensity scoring	58 yes
Use of multiple regression models	13 yes
Use of instrumental variable models	0 yes
Use of mediation analysis	0 yes
Use of other inference or correction models	1 yes
Use of sensitivity analyses	17 yes
Use of term "real-world data"	8 yes
Use of term "real-world evidence"	1 yes
Registration mentioned	5 yes
Protocol available	4 yes
Primary hypothesis confirmed	39 yes
If no, noninferior?	12 yes

4.1. An evolving field

Although we did not use a time filter for our search, it is apparent that the field is moving fast: of all studies included, only one predated 2010, and more than one-third were published since 2020, with more studies on the way, as evidenced by published protocols.[10,47,62,68] This reflects large databases becoming publicly available recently, improved search methods and data base indexing, and growing awareness of using routinely collected health data for research purposes. Although some sources have been created specifically for future research, like the Spine Tango Registry[6,61,79,87] and the Collaborative Health Outcomes Information Registry (CHOIR),[5,23,43,74] in other cases, large data sets of national health bodies were made accessible to enable research. These often include naturally large unified systems, like the US Department of Veterans Affairs[85] or the United Kingdom's National Health Service, a single-payer system, under which UK residents have single identification numbers under which multiple records across multiple health services are identifiable.[16] Such secondary use of data should certainly be welcomed because it can increase research value without additional burden on patients.

4.2. Terminology is ill defined

One of the surprising results of this review was how rarely the terms "real-world data" and "real-world evidence" were used in reports, with fewer than 6 articles mentioning "real-world data" as a phrase, just one article using the term "real-world evidence",[28] and only 2 studies using "real-world" as a phrase in the title. At the same time, the term "real-world" is frequently used to describe investigations like pragmatic trials that we would not necessarily classify as "real-world evidence" studies, as a means of distancing them from laboratory settings. There is little consensus regarding what the term RWE should be used for, and if it is the best term to use after all.[77] This aspect is nontrivial because if RWE studies are to be used for evidence appraisal as an alternative to conventional RCT data for treatment decisions, studies need to be accessible and robust for evidence synthesis by means of systematic reviews and meta-analyses. However, clinical trials use clearly defined terms that are distinct and clear. Terms like "randomized, placebo-controlled, and double-blind" will be in the title of all randomized, placebo-controlled, double-blind clinical trials, thanks to initiatives like CONSORT.[16] This is not the case for RWE studies. Searching for sensitivity (eg, searching PubMed for "pain AND (real-world OR real world)") will lead to thousands of findings, increasing the workload of systematic reviews of the field, with a very low specificity. "In addition, the term "real world" will not be used by many relevant studies, suggesting that even such a search string would not be entirely sensitive. This also indicates that the search conducted here was likely not exhaustive but can only provide a partial picture of real-world studies."

Therefore, it will be critical for the field to agree on standard terminology and quality of methods and assessment, to lead to a body of work that can contribute to evidence synthesis in medicine.

4.3. A monoculture of statistics

More than 90% of the studies included in our analysis used propensity scores to account for potential confounding, making it by far the dominant method. This may partly be the result of our search string (which included "propensity score" as a term). However, other statistical methods in our string were not picked up at all. Propensity score matching or adjustment is an appropriate method to reducing confounding effects. However, these methods depend on the measurement and inclusion of all important confounders. More surprising to us was the wide use of

propensity matching over other uses of propensity scores, such as stratification or inverse probability weighting, especially in large databases. The absence of more modern methods, such as marginal structural models, was conspicuous, possibly because of the relative simplicity of implementing propensity score-based methods. The use of appropriate statistical methods for drawing valid causal inferences is a crucial element for the success of RWE studies, and the potential of emerging methods has been shown.[2,4,18,76] The fast-emerging field of causal inference develops methods designated to drawing high degrees of evidence from nonexperimental data[32] and can be especially used in RWE studies.[48]

4.4. Registration and group differences

Most studies that met the eligibility criteria for this review reported a statistically significant difference in the primary comparison. We would argue that this is than what should be expected. The 2 principal sources of unexpectedly high rates of significant findings are (1) reporting of false-positives, due to a failure to properly account for potential confounding, and (2) the pressure of publishing "positive" findings, ie, publication bias.

The risk of making multiple comparisons and selectively publishing significant results or HARKing (hypothesizing after the results are known[42]) is increased in retrospective data studies, where the research question can be more easily changed, or a secondary question elevated to the primary one. Currently, it is not expected nor standard practice to register and publish protocols, as shown by the low number of protocols available for the studies in this analysis. However, the necessary tools are widely available, and we encourage authors to make use of them voluntarily. Going forward, a rigorous mandate for registration of studies, including protocols and hypotheses being publicly available, could improve this situation. However, this will be partly dependent on time-stamped access to data repositories and proof that registration took place before data access. Currently, we believe that the risk of publication bias or HARKing is too high to allow for RWE studies to be treated alongside RCTs in evidence synthesis. Sensitivity analyses are also not part of standard recommendations or practice. These will be of a similar importance moving forward because they show the robustness of results to violations of the assumption that are critical for the validity of causal inference methods.

4.5. Limitations

Although we aimed for a systematic, comprehensive approach to capturing methods used in the field, we cannot assume that this review was in all ways comprehensive. As explained above, constructing a search string for RWE studies can be challenging, making it difficult to find all potentially relevant studies. It is likely that there are studies meeting our inclusion criteria that we did not discover using our search strategy. The fact that we included additional studies suggested by authors implies that there will likely be additional studies missing from the search. In addition, we focused on studies comparing at least 2 groups, which excludes a large proportion of RWE studies. We did so because we were specifically interested in methodological approaches to comparing groups. Our eligibility criteria were partly based on subjective judgment (eg, excluding studies only "peripherally reporting pain"). We acknowledge that this may introduce a bias and decrease generalizability. Moreover, we found that most included studies were published relatively recently, but this could be partly because of the search strategy, which used terms that have become popular only in the past decade. Although we found a high proportion of studies reporting significant results, we cannot exclude the possibility of this finding being linked to an imperfect search strategy as well.

5. Way forward

Despite current limitations, RWE studies in pain research already contribute information to the evidence base. This is often the case in questions where RCTs are less frequently conducted, especially in surgery and alternative medicine. In these fields, RWE studies are not considered in competition with RCTs but are rather seen as a complementary source of evidence.[20] The use of rigorous causal inference methods will allow for high-level evidence to be drawn from RWD studies.[32,48] As is to be expected with emergent technologies, methodological improvements and increased rigor are needed. Statisticians and epidemiologists should be included from the early planning stage, preregistration as well as transparent project timelines should be mandatory, and a widely accepted standard terminology will be needed to make these works accessible and the evidence generated of high quality.

Disclosures

The authors have no conflicts of interest to declare.

Acknowledgement

Financial support was provided by the Analgesic, Anesthetic, and Addiction Clinical Trial Translations, Innovations, Opportunities, and Networks (ACTTION) public–private partnership with the US Food and Drug Administration (FDA), which has received research contracts, grants, or other revenue from the FDA, multiple pharmaceutical and device companies, philanthropy, royalties, and other sources. The views expressed in this article are those of the authors and no official endorsement by the FDA or the pharmaceutical and device companies that provided unrestricted grants to support ACTTION should be inferred.

Article history:
Received 28 May 2022
Received in revised form 21 October 2022
Accepted 12 November 2022

References

[1] Adogwa O, Huang MI, Thompson PM, Darlington T, Cheng JS, Gokaslan ZL, Gottfried ON, Bagley CA, Anderson GD, Isaacs RE. No difference in postoperative complications, pain, and functional outcomes up to 2 years after incidental durotomy in lumbar spinal fusion: a prospective, multi-institutional, propensity-matched analysis of 1,741 patients. Spine J 2014;14:1828–34.

[2] Alter BJ, Anderson NP, Gillman AG, Yin Q, Jeong J-H, Wasan AD. Hierarchical clustering by patient-reported pain distribution alone identifies distinct chronic pain subgroups differing by pain intensity, quality, and clinical outcomes. PLoS One 2021;16:e0254862.

[3] Austevoll IM, Gjestad R, Brox JI, Solberg TK, Storheim K, Rekeland F, Hermansen E, Indrekvam K, Hellum C. The effectiveness of decompression alone compared with additional fusion for lumbar spinal stenosis with degenerative spondylolisthesis: a pragmatic comparative non-inferiority observational study from the Norwegian Registry for Spine Surgery. Eur Spine J 2017;26:404–13.

[4] Azizoddin DR, Schreiber K, Beck MR, Enzinger AC, Hruschak V, Darnall BD, Edwards RR, Allsop MJ, Tulsky JA, Boyer E, Mackey S. Chronic pain severity, impact, and opioid use among patients with cancer: an analysis of biopsychosocial factors using the CHOIR learning health care system. Cancer 2021;127:3254–63.

[5] Bhandari RP, Feinstein AB, Huestis SE, Krane EJ, Dunn AL, Cohen LL, Kao MC, Darnall BD, Mackey SC. Pediatric-Collaborative Health Outcomes Information Registry (Peds-CHOIR): a learning health system to guide pediatric pain research and treatment. Pain 2016;157:2033–44.
[6] Bieri KS, Goodwin K, Aghayev E, Riesner HJ, Greiner-Perth R. Dynamic posterior stabilization versus posterior lumbar intervertebral fusion: a matched cohort study based on the spine Tango registry. J Neurol Surg A Cent Eur Neurosurg 2018;79:224–30.
[7] Brouwer ME, Reininga IHF, El Moumni M, Wendt KW. Outcomes of operative and nonoperative treatment of 3- and 4-part proximal humeral fractures in elderly: a 10-year retrospective cohort study. Eur J Trauma Emerg Surg 2019;45:131–8.
[8] Centeno C, Pitts J, Al-Sayegh H, Freeman M. Efficacy of autologous bone marrow concentrate for knee osteoarthritis with and without adipose graft. Biomed Res Int 2014;2014:370621.
[9] D'Agostino RB. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Statist Med 1998;17:2265–81.
[10] Darnall BD, Mackey SC, Lorig K, Kao M-C, Mardian A, Stieg R, Porter J, DeBruyne K, Murphy J, Perez L, Okvat H, Tian L, Flood P, McGovern M, Colloca L, King H, van Dorsten B, Pun T, Cheung M. Comparative effectiveness of cognitive behavioral therapy for chronic pain and chronic pain self-management within the context of voluntary patient-centered prescription opioid tapering: the EMPOWER study protocol. Pain medicine (malden, mass.) 2020;21:1523–31.
[11] De Oliveira GS Jr, Bialek JM, Nicosia L, McCarthy RJ, Chang R, Fitzgerald P, Kim JY. Lack of association between breast reconstructive surgery and the development of chronic pain after mastectomy: a propensity matched retrospective cohort analysis. Breast 2014;23:329–33.
[12] Desai RJ, Franklin JM. Alternative approaches for confounding adjustment in observational studies using weighting based on the propensity score: a primer for practitioners. BMJ (Clinical research ed.) 2019;367:l5657.
[13] El Desoky M, El Nakeeb A, El Sorogy M, Hamed H, Attia M, Ezzat H, El Hemly M, El-Geidi A, Moneer A. Comparative study between open and laparoscopic total proctocolectomy with ileal pouch-anal anastomosis for ulcerative colitis: a propensity score-matched study. Egypt J Surg 2020; 39:985–91.
[14] Faber S, Angele P, Zellner J, Bode G, Hochrein A, Niemeyer P. Comparison of clinical outcome following cartilage repair for patients with underlying varus deformity with or without additional high tibial osteotomy: A propensity score-matched study based on the German cartilage registry (KnorpelRegister DGOU). Cartilage 2021;13: 1206S-1216S.
[15] Fang CX, Liu R, Yee DKH, Chau J, Lau TW, Chan R, Woo SB, Wong TM, Fang E, Leung F. Comparison of radiological and clinical outcomes, complications, and implant removals in anatomically pre-contoured clavicle plates versus reconstruction plates - a propensity score matched retrospective cohort study of 106 patients. BMC Musculoskelet Disord 2020;21:413.
[16] Foley T, Lie M. Clinical review: The impact of data released through the data access request service. United Kingdom: National Health Service, 2019.
[17] Franklin JM, Schneeweiss S. When and how can real world data analyses substitute for randomized controlled trials?. Clin Pharmacol Ther 2017; 102:924–33.
[18] Gilam G, Cramer EM, Webber KA, Ziadni MS, Kao M-C, Mackey SC. Classifying chronic pain using multidimensional pain-agnostic symptom assessments and clustering analysis. Sci Adv 2021;7:eabj0320.
[19] Gillman A, Zhang Di, Jarquin S, Karp JF, Jeong J-H, Wasan AD. Comparative effectiveness of embedded mental health services in pain management clinics vs standard care. Pain Med 2020;21:978–91.
[20] Gilron I, Blyth F, Smith BH. Translating clinical trials into improved real-world management of pain: convergence of translational, population-based, and primary care research. PAIN 2020;161:36–42.
[21] Goh GS, Tay YWA, Liow MHL, Gatot C, Ling ZM, Fong PL, Soh RCC, Guo CM, Yue WM, Tan SB, Chen JL. Elderly patients undergoing minimally invasive transforaminal lumbar interbody fusion may have similar clinical outcomes, perioperative complications, and fusion rates as their younger counterparts. Clin Orthop Relat Res 2020;478:822–32.
[22] Goh GS, Zeng GJ, Tay DK, Lo NN, Yeo SJ, Liow MHL. Does obesity lead to lower rates of clinically meaningful improvement or satisfaction after total hip arthroplasty? A propensity score-matched study. Hip Int 2020: 1120700020974656.
[23] Hah JM, Sturgeon JA, Zocca J, Sharifzadeh Y, Mackey SC. Factors associated with prescription opioid misuse in a cross-sectional cohort of patients with chronic non-cancer pain. J Pain Res 2017;10:979–87.
[24] Han DG, Koh W, Shin JS, Lee J, Lee YJ, Kim MR, Kang K, Shin BC, Cho JH, Kim NK, Ha IH. Cervical surgery rate in neck pain patients with and without acupuncture treatment: a retrospective cohort study. Acupunct Med 2019;37:268–76.
[25] Han L, Goulet JL, Skanderson M, Bathulapalli H, Luther SL, Kerns RD, Brandt CA. Evaluation of complementary and integrative health approaches among US Veterans with musculoskeletal pain using propensity score methods. Pain Med 2019;20:90–102.
[26] Harrold LR, Shan Y, Rebello S, Kramer N, Connolly SE, Alemao E, Kelly S, Kremer JM, Rosenstein ED. Disease activity and patient-reported outcomes in patients with rheumatoid arthritis and Sj├Âgren's syndrome enrolled in a large observational US registry. Rheumatol Int 2020;40:1239–48.
[27] Hayes CJ, Krebs EE, Hudson T, Brown J, Li C, Martin BC. Impact of opioid dose escalation on pain intensity: a retrospective cohort study. Pain 2020;161:979–88.
[28] Helwig U, Mross M, Schubert S, Hartmann H, Brandes A, Stein D, Kempf C, Knop J, Campbell-Hill S, Ehehalt R. Real-world clinical effectiveness and safety of vedolizumab and anti-tumor necrosis factor alpha treatment in ulcerative colitis and Crohn's disease patients: a German retrospective chart review. BMC Gastroenterol 2020;20:211.
[29] Herman PM, Yuan AH, Cefalu MS, Chu K, Zeng Q, Marshall N, Lorenz KA, Taylor SL. The use of complementary and integrative health approaches for chronic musculoskeletal pain in younger US Veterans: an economic evaluation. PLoS One 2019;14:e0217831.
[30] Hermansen E, Romild UK, Austevoll IM, Solberg T, Storheim K, Brox JI, Hellum C, Indrekvam K. Does surgical technique influence clinical outcome after lumbar spinal stenosis decompression? A comparative effectiveness study from the Norwegian Registry for Spine Surgery. Eur Spine J 2017;26:420–7.
[31] Hirsch BP, Khechen B, Patel DV, Cardinal KL, Guntin JA, Singh K. Safety and efficacy of revision minimally invasive lumbar decompression in the ambulatory setting. Spine 2019;44:E494–e499.
[32] Ho M, van der Laan M, Lee H, Chen J, Lee K, Fang Y, He W, Irony T, Jiang Q, Lin X, Meng Z, Mishra-Kalyani P, Rockhold F, Song Y, Wang H, White R. The current landscape in biostatistics of real-world data and evidence: causal inference frameworks for study design and analysis. Stat Biopharm Res 2021:1–14.
[33] Hohenschurz-Schmidt D, Kleykamp BA, Draper-Rodi J, Vollert J, Chan J, Ferguson M, McNicol E, Phalip J, Evans SR, Turk DC, Dworkin RH, Rice ASC. Pragmatic trials of pain therapies: a systematic review of methods. PAIN 2022;163:21–46.
[34] Hu CG, Zheng K, Liu GH, Li ZL, Zhao YL, Lian JH, Guo SP. Effectiveness and postoperative pain level of single-port versus two-port thoracoscopic lobectomy for lung cancer: a retrospective cohort study. Gen Thorac Cardiovasc Surg 2021;69:318–25.
[35] Il Kim J, Kim YT, Jung HJ, Lee JK. Does adding corticosteroids to periarticular injection affect the postoperative acute phase response after total knee arthroplasty? Knee 2020;27:493–9.
[36] Jeong HJ, Kim HS, Rhee SM, Oh JH. Risk factors for and prognosis of folded rotator cuff tears: a comparative study using propensity score matching. J Shoulder Elbow Surg 2021;30:826-35.
[37] Jung H, Lee KH, Jeong Y, Yoon S, Kim WH, Lee HJ. Effect of fentanyl-based intravenous patient-controlled analgesia with and without basal infusion on postoperative opioid consumption and opioid-related side effects: a retrospective cohort study. J Pain Res 2020;13:3095–106.
[38] K├Âckerling F, Bittner R, Kofler M, Mayer F, Adolf D, Kuthe A, Weyhe D. Lichtenstein versus total extraperitoneal patch plasty versus transabdominal patch plasty technique for primary unilateral inguinal hernia repair: a registry-based, propensity score-matched comparison of 57, 906 patients. Ann Surg 2019;269:351–7.
[39] K├Âckerling F, Koch A, Adolf D, Keller T, Lorenz R, Fortelny RH, Schug-Pass C. Has shouldice repair in a selected group of patients with inguinal hernia comparable results to lichtenstein, TEP and TAPP techniques?. World J Surg 2018;42:2001–10.
[40] K├Âckerling F, Lammers B, Weyhe D, Reinpold W, Zarras K, Adolf D, Riediger H, Kr├╝ger CM. What is the outcome of the open IPOM versus sublay technique in the treatment of larger incisional hernias?: a propensity score-matched comparison of 9091 patients from the Herniamed Registry. Hernia 2021;25:23–31.
[41] K├Âckerling F, Simon T, Adolf D, K├Âckerling D, Mayer F, Reinpold W, Weyhe D, Bittner R. Laparoscopic IPOM versus open sublay technique for elective incisional hernia repair: a registry-based, propensity score-matched comparison of 9907 patients. Surg Endosc 2019;33:3361–9.
[42] Kerr NL. HARKing: hypothesizing after the results are known. Personal Soc Psychol Rev official J Soc Personal Soc Psychol Inc 1998;2: 196–217.

[43] Khan JS, Hah JM, Mackey SC. Effects of smoking on patients with chronic pain: a propensity-weighted analysis on the Collaborative Health Outcomes Information Registry. PAIN 2019;160:2374–9.
[44] Kim MK, Kang H, Choi GJ, Kang KH. Robotic thyroidectomy decreases postoperative pain compared with conventional thyroidectomy. Surg Laparosc Endosc Percutan Tech 2019;29:255–60.
[45] Kim MK, Yi MS, Kang H, Choi GJ. Effects of remifentanil versus nitrous oxide on postoperative nausea, vomiting, and pain in patients receiving thyroidectomy: propensity score matching analysis. Medicine (Baltimore) 2016;95:e5135.
[46] Krau├ƒ M, Heinzel-Gutenbrunner M, Kr├Ânung L, Hanisch E, Buia A. Comparing large pore lightweight mesh versus small pore heavyweight mesh in open mesh plug repair of primary and recurrent unilateral inguinal hernia - a questionnaire study for a retrospective analysis of a cohort of elective groin hernia patients using propensity score matching. Int J Surg 2020;75:93–8.
[47] Krauss I, Mueller G, Haupt G, Steinhilber B, Janssen P, Jentner N, Martus P. Effectiveness and efficiency of an 11-week exercise intervention for patients with hip or knee osteoarthritis: a protocol for a controlled study in the context of health services research. BMC Public Health 2016;16:367.
[48] Lee H-J, Wong JB, Jia B, Qi X, DeLong ER. Empirical use of causal inference methods to evaluate survival differences in a real-world registry vs those found in randomized clinical trials. Stat Med 2020;39:3003–21.
[49] Leoni MLG, Schatman M, Demartini L, Lo Bianco G, Terranova G. Genicular nerve pulsed dose radiofrequency (PDRF) compared to intra-articular and genicular nerve PDRF in knee osteoarthritis pain: a propensity score-matched analysis. J Pain Res 2020;13:1315–21.
[50] Li L, Zhang J. Application value of ERAS in perioperative period of precise hepatectomy for hepatocellular carcinoma patients. J Buon 2020;25: 965–71.
[51] Li XW, Wang CY, Zhang JJ, Ge Z, Lin XH, Hu JH. Short-term efficacy of transvaginal specimen extraction for right colon cancer based on propensity score matching: a retrospective cohort study. Int J Surg 2019;72:102–8.
[52] Liu CW, Bhatia A, Buzon-Tan A, Walker S, Ilangomaran D, Kara J, Venkatraghavan L, Prabhu AJ. Weeding out the problem: the impact of preoperative cannabinoid use on pain in the perioperative period. Anesth Analg 2019;129:874–81.
[53] Liu H, Tang X, Chang Y, Li A, Li Z, Xiao Y, Zhang Y, Pan Z, Lv L, Lin M, Yin L, Jiang H. Comparison of surgical outcomes between video-assisted anal fistula treatment and fistulotomy plus seton for complex anal fistula: a propensity score matching analysis - retrospective cohort study. Int J Surg 2020;75:99–104.
[54] Liu Y, Cai C, Aquino A, Al-Mousawi S, Zhang X, Choong SKS, He X, Fan X, Chen B, Feng J, Zhu X, Al-Naimi A, Mao H, Tang H, Jin D, Li X, Cao F, Jiang H, Long Y, Zhang W, Wang G, Xu Z, Yin S, Zeng G. Management of large renal stones with super-mini percutaneous nephrolithotomy: an international multicentre comparative study. BJU Int 2020;126:168–76.
[55] Luzzi L, Corzani R, Ghisalberti M, Meniconi F, Leonibus Lde, Molinaro F, Paladini P. Robotic surgery vs. open surgery for thymectomy, a retrospective case-match study. J Robotic Surg.
[56] MacDowall A, Skeppholm M, Lindhagen L, Robinson Y, Lofgren H, Michaelsson K, Olerud C. Artificial disc replacement versus fusion in patients with cervical degenerative disc disease with radiculopathy: 5-year outcomes from the National Swedish Spine Register. J Neurosurgery-Spine 2019;30:159–67.
[57] Mao Y, Lan Y, Cui F, Deng H, Zhang Y, Wu X, Liang W, Liu J, Liang H, He J. Comparison of different surgical approaches for anterior mediastinal tumor. J Thorac Dis 2020;12:5430–9.
[58] Mekhail N, Costandi S, Mehanny DS, Armanyous S, Saied O, Taco-Vasquez E, Saweris Y. The impact of tobacco smoking on spinal cord stimulation effectiveness in complex regional pain syndrome patients. Neuromodulation 2020;23:133–9.
[59] Mekhail N, Mehanny D, Armanyous S, Saweris Y, Costandi S. The impact of obesity on the effectiveness of spinal cord stimulation in chronic spine-related pain patients. Spine J 2019;19:476–86.
[60] Mohammad HR, Matharu GS, Judge A, Murray DW. A matched comparison of revision rates of cemented oxford unicompartmental knee replacements with single and twin peg femoral components, based on data from the national joint registry for england, wales, northern Ireland and the Isle of Man. Acta Orthop 2020;91:420–5.
[61] Munting E, R├Âder C, Sobottke R, Dietrich D, Aghayev E. Patient outcomes after laminotomy, hemilaminectomy, laminectomy and laminectomy with instrumented fusion for spinal canal stenosis: a propensity score-based study from the Spine Tango registry. Eur Spine J 2015;24:358–68.
[62] Neuderth S, Schwarz B, Gerlich C, Schuler M, Markus M, Bethge M. Work-related medical rehabilitation in patients with musculoskeletal disorders: the protocol of a propensity score matched effectiveness study (EVA-WMR, DRKS00009780). BMC Public Health 2016;16:804.
[63] Niedermayer S, Heyn J, Guenther F, K├╝chenhoff H, Luchting B. Remifentanil for abdominal surgery is associated with unexpectedly unfavorable outcomes. Pain 2020;161:266–73.
[64] Oh TK, Ji E, Na HS. The effect of neuromuscular reversal agent on postoperative pain after laparoscopic gastric cancer surgery: comparison between the neostigmine and sugammadex. Medicine (Baltimore) 2019; 98:e16142.
[65] Oma E, Bisgaard T, Jorgensen LN, Jensen KK. Nationwide propensity-score matched study of mesh versus suture repair of primary ventral hernias in women with a subsequent pregnancy. World J Surg 2019;43: 1497–504.
[66] Pape E, Hagen KB, Brox JI, Natvig B, Schirmer H. Early multidisciplinary evaluation and advice was ineffective for whiplash-associated disorders. Eur J Pain 2009;13:1068–75.
[67] Pearlman M, Covin Y, Schmidt R, Mortensen EM, Mansi IA. Statins and lower gastrointestinal conditions: a retrospective cohort study. J Clin Pharmacol 2017;57:1053–63.
[68] Petrou PA, Leong MS, Mackey SC, Salmasi V. Stanford Pragmatiec Effectiveness Comparison (SPEC) protocol: comparing long-term effectiveness of high-frequency and burst spinal cord stimulation in real-world application. Contemp Clin trials 2021;103:106324.
[69] Qin J, Zhu HD, Guo JH, Pan T, Lu J, Ni CF, Wu P, Xu H, Mao AW, Teng GJ. Comparison of 125 iodine seed-loaded stents with different diameters in esophageal cancer: a multicenter retrospective cohort study. Dysphagia 2020;35:725–32.
[70] R├Âder C, Baumg├ñrtner B, Berlemann U, Aghayev E. Superior outcomes of decompression with an interlaminar dynamic device versus decompression alone in patients with lumbar spinal stenosis and back pain: a cross registry study. Eur Spine J 2015;24:2228–35.
[71] Reed GW, Gerber RA, Shan Y, Takiya L, Dandreo KJ, Gruben D, Kremer J, Wallenstein G. Real-world comparative effectiveness of tofacitinib and tumor necrosis factor inhibitors as monotherapy and combination therapy for treatment of rheumatoid arthritis. Rheumatol Ther 2019;6:573–86.
[72] Reinpold W, Schr├Âder M, Berger C, Nehls J, Schr├Âder A, Hukauf M, K├Âckerling F, Bittner R. Mini- or less-open sublay operation (milos): a new minimally invasive technique for the extraperitoneal mesh repair of incisional hernias. Ann Surg 2019;269:748–55.
[73] Roeb MM, Wolf A, Gr├╝ber SS, Mei├ƒner W, Volk T. Epidural against systemic analgesia: an international registry analysis on postoperative pain and related perceptions after abdominal surgery. Clin J Pain 2017;33:189–97.
[74] Scherrer KH, Ziadni MS, Kong J-T, Sturgeon JA, Salmasi V, Hong J, Cramer E, Chen AL, Pacht T, Olson G, Darnall BD, Kao M-C, Mackey S. Development and validation of the collaborative health outcomes information registry body map. Pain Rep 2021;6:e880.
[75] Schrittwieser R, K├Âckerling F, Adolf D, Hukauf M, Gruber-Blum S, Fortelny RH, Petter-Puchner AH. Small and laterally placed incisional hernias can be safely managed with an onlay repair. World J Surg 2019; 43:1921–7.
[76] Sharifzadeh Y, Kao M-C, Sturgeon JA, Rico TJ, Mackey S, Darnall BD. Pain catastrophizing moderates relationships between pain intensity and opioid prescription: nonlinear sex differences revealed using a learning health system. Anesthesiology 2017;127:136–46.
[77] Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, LaVange L, Marinac-Dabic D, Marks PW, Robb MA, Shuren J, Temple R, Woodcock J, Yue LQ, Califf RM. Real-world evidence - what is it and what can it tell us?. New Engl J Med 2016;375:2293–7.
[78] Staartjes VE, Battilana B, Schr├Âder ML. Robot-guided transforaminal versus robot-guided posterior lumbar interbody fusion for lumbar degenerative disease. Neurospine 2020;18:98-105.
[79] Staub LP, Ryser C, R├Âder C, Mannion AF, Jarvik JG, Aebi M, Aghayev E. Total disc arthroplasty versus anterior cervical interbody fusion: use of the Spine Tango registry to supplement the evidence from randomized control trials. Spine J 2016;16:136–45.
[80] Steward R, Carney P, Law A, Xie L, Wang Y, Yuce H. Long-term outcomes after elective sterilization procedures - a comparative retrospective cohort study of Medicaid patients. Contraception 2018;97:428–33.
[81] Stundner O, Poeran J, Ladenhauf HN, Berger MM, Levy SB, Zubizarreta N, Mazumdar M, Bekeris J, Liu J, Galatz LM, Moucha CS, Memtsoudis S. Effectiveness of intravenous acetaminophen for postoperative pain management in hip and knee arthroplasties: a population-based study. Reg Anesth pain Med 2019;44:565–72.
[82] Takenaka S, Mukai Y, Tateishi K, Hosono N, Fuji T, Kaito T. Clinical outcomes after posterior lumbar interbody fusion. Clin Spine Surg 2017; 30:E1411–E1418.
[83] Tosi D, Bonitta G, Mazzucco A, Righi I, Mendogni P, Palleschi A, Rocco G, Mancuso M, Pernazza F, Refai M, Bortolotti L, Rizzardi G, Gargiulo G,

Dolci GP, Perkmann R, Zaraca F, Benvenuti M, Gavezzoli D, Cherchi R, Ferrari P, Mucilli F, Camplese P, Melloni G, Mazza F, Cavallesco G, Maniscalco P, Voltolini L, Gonfiotti A, Stella F, Argnani D, Pariscenti GL, Surrente C, Lopez C, Droghetti A, Giovanardi M, Breda C, Lo Giudice F, Alloisio M, Bottoni E, Spaggiari L, Gasparri R, Torre M, Rinaldo A, Nosotti M, Rosso L, Negri GP, Bandiera A, Stefani A, Natali P, Scarci M, Pirondini E, Curcio C, Amore D, Baietto G, Casadio C, Nicotra S, Dell'amore A, Bertani A, Russo E, Ampollini L, Carbognani P, Puma F, Vinci D, Andreetti C, Poggi C, Cardillo G, Margaritora S, Meacci E, Luzzi L, Ghisalberti M, Crisci R, Zaccagna G, Lausi P, Guerrera F, Fontana D, Della Beffa V, Morelli A, Londero F, Imperatori A, Rotolo N, Terzi A, Viti A, Infante M, Benato C. Uniportal and three-portal video-assisted thoracic surgery lobectomy: analysis of the Italian video-assisted thoracic surgery group database. Interactive Cardiovasc Thorac Surg 2019;29:714–21.

[84] Turi S, Gemma M, Braga M, Monzani R, Radrizzani D, Beretta L. Epidural analgesia vs systemic opioids in patients undergoing laparoscopic colorectal surgery. Int J Colorectal Dis 2019;34:915–21.

[85] Wandner LD, Fenton BT, Goulet JL, Carroll CM, Heapy A, Higgins DM, Bair MJ, Sandbrink F, Kerns RD. Treatment of a large cohort of Veterans experiencing musculoskeletal disorders with spinal cord stimulation in the Veterans health administration: veteran characteristics and outcomes. J Pain Res 2020;13:1687–97.

[86] Yuan H, Ali MS, Brouwer ES, Girman CJ, Guo JJ, Lund JL, Patorno E, Slaughter JL, Wen X, Bennett D. Real-world evidence: what it is and what it can tell us according to the international society for pharmacoepidemiology (ISPE) comparative effectiveness research (CER) special interest group (SIG). Clin Pharmacol Ther 2018;104: 239–41.

[87] Zweig T, Enke J, Mannion AF, Sobottke R, Melloh M, Freeman BJ, Aghayev E. Is the duration of pre-operative conservative treatment associated with the clinical outcome following surgical decompression for lumbar spinal stenosis? A study based on the Spine Tango Registry. Eur Spine J 2017;26:488–500.