

Running head: Systematic review of health-economic evaluation models of Osteoarthritis

A Systematic Review of the Evolution of Health-Economic Evaluation Models of Osteoarthritis

Ting Zhao¹, PhD candidate; Hasnat Ahmad¹, PhD candidate; Barbara de Graaff¹, PhD, Research fellow; Qing Xia¹, PhD candidate; Tania Winzenberg¹, PhD, Professor; Dawn Aitken¹, PhD, Research fellow; Andrew J. Palmer^{1,2}, MBBS, Professor

1. Menzies Research Institute Tasmania, University of Tasmania

2. School of Population and Global Health, The University of Melbourne

Correspondence to: *Andrew J. Palmer*, Menzies Research Institute Tasmania, University of Tasmania, Medical Science 1 Building, 17 Liverpool St, Hobart, Tasmania 7000, Australia

E-mail: Andrew.palmer@utas.edu.au

Tel: +61 3 6226 7729

No financial support for this study was received, all authors declare that they have no competing interests.

Word count for the manuscript: 3,996

Abstract

Objective

To comprehensively synthesise the evolution of health-economic evaluation models (HEEMs) of all OA interventions including preventions, core treatments, adjunct non-pharmacological interventions, pharmacological and surgical treatments.

Methods

The literature was searched within health-economic/biomedical databases. Data extracted included: OA type, population characteristics, model setting/type/events, study perspective, comparators; and the reporting quality of the studies was assessed. The review protocol was registered at the International Prospective Register of Systematic Reviews (registration: CRD42018092937).

Results

Eighty-eight studies were included. Pharmacological and surgical interventions were the focus in 51% and 44% studies, respectively. Twenty-four studies adopted a societal perspective (with increasing popularity after 2013), however most (63%) did not include indirect costs. Quality-adjusted life years (QALYs) was the most popular outcome measure since 2008. Markov models were used by 62% of studies, with increasing popularity since 2008. Until 2010, most studies used short-to-medium time horizons; subsequently a lifetime horizon became popular. Eighty-six percent of studies reported discount rate(s) (predominantly between 3% and 5%). Studies published after 2002 had a better coverage of OA-related adverse events (AEs). Reporting quality significantly improved after 2001.

Conclusions

OA HEEMs have evolved and improved substantially over time, with focus shifting from short-to-medium-term pharmacological decision-tree models to surgical-focused lifetime Markov models. Indirect costs of OA are frequently not considered, despite using a societal perspective. There was lack of reporting sensitivity of model outcome to input parameters including discount rate, OA definition, and population parameters. Whilst the coverage of OA-related AEs has improved over time, it is still not comprehensive.

Key Words: Osteoarthritis; Health-economic evaluation; Markov Models, Decision-tree Models

Significance and Innovations

- This is the first study comprehensively reviewing the evolution of health-economic evaluation models of all OA interventions including preventions, core treatments, adjunct non-pharmacological interventions, pharmacological and surgical treatments.
- It identifies the key strengths and limitations facing existing OA health-economic evaluation models and provides directions for improvement in the current modelling practice.
- Modelled OA health-economic evaluations have evolved substantially over time, with focus shifting from short-to-medium-term pharmacological decision-tree models to surgical-focused lifetime Markov models.
- Indirect costs are mostly not considered in the OA health-economic evaluation models, even when the study is conducted from a societal perspective. The coverage of OA-related adverse events is not comprehensive.

Introduction

Approximately 240 million people globally were affected by osteoarthritis (OA) in 2016 (1) and its prevalence is projected to rise steadily (1-3). OA is characterised by joint pain, stiffness, swelling, loss of function and disability; which in turn, negatively impacts individuals' health-related quality of life (HRQoL) (4) and poses a significant economic burden to patients and society in terms of both direct (healthcare) and indirect costs (from lost productivity [early retirement/absenteeism/presenteeism]) (5-10).

Whilst there is no cure for OA, there are treatments available to ease OA symptoms and postpone disability progression. According to recent OA management guidelines (11), treatments include lifestyle (e.g.: exercise, weight management), non-medical (e.g.: heat packs, manual therapy), medical, and surgical interventions. In chronic diseases like OA, healthcare policy decisions should be informed by evidence of the long-term health and economic impacts. Given the scarcity of healthcare resources, it is critically important that the most cost-effective interventions are chosen.

Health-economic evaluations (HEEs) compare alternative treatment options in terms of both economic costs and clinical effectiveness to identify the interventions that are best value for money. The models used for these incorporate clinical, health-economic and epidemiological data. HEEs include full (e.g.: cost-benefit, cost-effectiveness and cost-utility analysis) and partial (e.g.: cost of illness analysis) evaluations (Appendix 1) (12). Since the first model-based HEE of OA treatments was performed in 1994, numerous modelling studies have been conducted, particularly in Western developed nations (13-15). OA models vary in their methodological framework, model structures, division of model events, and input data sources (16-19). The availability of better quality clinical and epidemiological data, and

methodological advances have contributed to the evolution of modelled HEEs over time (16, 17, 20).

The scope of previous systematic reviews of OA health-economic evidence was often limited to a specific treatment type (e.g.: surgical, pharmacological, physical) (21-23). None have synthesised the evolution of health-economic evaluation models (HEEMs) of all OA interventions. As the comprehensive investigation of the diverse HEEMs could lead to significant improvements in modelling practice (24, 25), we aimed to explore the evolution of HEEMs used for all forms of OA interventions, with an emphasis on their strengths and weaknesses and study gaps to inform the future development of an improved and overarching HEEM of OA.

Methods

This study is reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines (26). The protocol was registered at the International Prospective Register of Systematic Reviews (PROSPERO) (registration number: CRD42018092937).

Literature search

Three biomedical (Medline via OvidSP, Embase via Ovid, and China National Knowledge Infrastructure) and three health-economic/economic databases (American Economic Association, the Centre for Reviews and Dissemination and the Cost-effectiveness Analysis Registry) were searched according to a search strategy that was defined in consultation with co-authors and a research librarian (Appendix 2). The literature was searched from each database's inception to July 2018. Reference lists of included studies and relevant reviews were hand searched.

Screening criteria

Title/abstract screening and full-text screening were performed in Covidence (27) by two reviewers (TZ and QX) independently based on predefined inclusion and exclusion criteria. Screening conflicts between the two reviewers were resolved through discussion with senior researchers (AP, HA and BdeG).

The inclusion criteria were studies: 1) in humans; 2) that reported the construction/application/validation of partial or full OA HEEM; 3) that were available as full-text; and 4) that were published in English/Chinese/German. Studies were included if they focused on arthritis populations including OA if they did not report the proportion of OA participants or if they reported on a cohort with $\geq 90\%$ OA. Review articles, conference abstracts, comments and books were excluded. (Appendix 3)

Data extraction

A Microsoft Excel spreadsheet designed by co-authors was used to extract data by two reviewers (TZ and HA) independently. Discrepancies were resolved by consensus, and an additional reviewer (AP) was consulted in cases of no consensus. Data extracted included authors, publication year, study setting, OA type, targeted interventions and comparators, and information related to HEEMs (simulated population characteristics, study perspective, time horizon, discount rate, clinical effectiveness measures, cost inputs, model type, modelling software, health states, health events, and sensitivity analysis type). (Appendix 4)

Assessment of reporting quality

We assessed each study's reporting quality using the 24-item Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist (28). We adapted the evaluation methods published by de Graaff et al. (29). Twenty-four items were equally weighted with '1' referring to the item being well performed, and '0' otherwise. As not all items were applicable to all studies (e.g.: for a cost of illness analysis, "effectiveness" was not applicable), the quality

scores were converted to percentages, adjusting the denominator to reflect the different number of applicable items. Studies were categorized into low ($\leq 50\%$), moderate (50%-75%) and high reporting quality ($>75\%$) groups (29).

Strategy for data synthesis

We adopted a narrative, descriptive synthesis approach (30) to assess and outline the evolution of HEEMs of OA interventions over time.

Results

Screening results

As shown in Figure 1, our search identified 1,683 potential references (1,498 from biomedical and 185 from economic databases). After removal of duplicates ($n=424$), 1,259 were left for title and abstract screening, which excluded 1,080. Of the 179 left, 96 were excluded during full-text screening (Figure 1). We identified an additional 5 studies through hand-searching, resulting in a total of 88 included studies (Appendix 5).

Year of publication

The first model-based HEE of OA treatment was published in 1994 (31) and more than half (51%, $n=45$) were published between 2013 and 2018 (Figure 2a).

Study settings

Almost half of the studies (49%, $n=43$) were conducted in the Americas, followed by Europe (36%, $n=32$) (Figure 2b). Most studies were performed in the United States (US) (40%, $n=35$), followed by the United Kingdom (UK) (18%, $n=16$) and Canada (7%, $n=6$). Only two studies were conducted in Australia and/or New Zealand. Seven did not report the study location. Studies from Australia, New Zealand, Saudi Arabia and China were published only after 2011.

OA types and intervention options

Thirty-seven studies (42%) focused on knee OA and 13 (15%) on hip OA. One study each focused on ankle arthritis and glenohumeral OA. Thirty-six (41%) studies did not specify OA type, most of which targeted non-surgical treatments. Most studies focusing on a specific type of OA evaluated surgical treatments. Seventy-five percent (48/64) of post-2005 publications focused on a specific type of OA, compared to only 17% (4/24) of such studies pre-2005. (Figure 3a)

Of the 88 studies, 45 (51%) focused on pharmacological interventions and 39 (44%) focused on surgical interventions (Figure 3a, Appendix 6 [for the full list of interventions by OA type, intervention categorization, frequency, and references]). Four of 88 included studies assessed OA preventions (32-35), all of which were performed after 2014. Before 2005, all but two (16, 36) compared alternative pharmacological treatments, particularly cyclo-oxygenase 2 (COX-2) inhibitors selective nonsteroidal anti-inflammatory drugs (NSAIDs) with non-selective NSAIDs. Studies focusing on OA surgical treatments became popular thereafter.

Targeted populations

The target population was generally patients aged ≥ 40 years. Three studies specifically focused on older patients with mean ages of 74, 78 and ≥ 80 years (37-39). Of 45 pharmacological-focused studies, 17 analysed two patient subgroups separately: 1) low risk patients (younger and without prior history of upper gastrointestinal [GI] events); and 2) high risk patients (older and with upper GI history).

Study perspectives and reporting of costs

Study perspectives were reported in 79 (90%) studies. Of these, 38 (48%) adopted a national healthcare perspective and 24 (30%) a societal perspective, however, 15 (63%) of these 24 studies did not consider indirect costs. Eleven studies adopted a third-party payer perspective,

and single studies used patient and veteran health administration perspectives. In addition, four studies considered more than one perspective, all of which were published after 2013 (Figure 3b).

For pharmacological-focused studies, direct medical costs generally comprised of drug acquisition and adverse events (AEs) treatment costs, while for surgical-focused studies, direct costs were from primary/revision surgery. Indirect costs included productivity losses (n=6), time lost from work (n=2), lost wages (n=3), and caregivers' expenses (n=1).

Effectiveness measurements

Eighty-nine percent (78/88) of studies reported the measurement of effectiveness, most commonly quality-adjusted life years (QALYs) (n=62), followed by multiple measures (n=7), and disease-specific effectiveness measures (n=6) (e.g.: AEs/complications averted, revision-free life years) (Figure 4a). The use of QALYs almost doubled, from 48% of studies published pre-2008 to 93% for those published after 2008 (Figure 4b).

Model types and computational software

Model type was reported in 87 (99%) studies. There were three key types: 1) Markov models (n=54, 62%), 2) decision-tree models (n=30, 34%), and 3) discrete-event simulation models (n=3, 3%) (Figure 5a). Markov models predominantly focused on surgical treatments, while pharmacological treatments were evaluated in 87% of decision-tree models (Figure 5a). The popularity of Markov models increased over time from 7% (pre-2008 period) to 88% (after 2008) (Figure 5b). There were four commonly used OA model structures : 1) OA policy (OAPoL) model (a Markov model to simulate the natural history of knee OA and predominantly used in the US) (19, 35, 38, 40-42); 2) the National Institute for Health and Care Excellence (NICE) model (a Markov model originally developed to compare NSAID/COX-2 inhibitor oral analgesics and subsequently extended to incorporate dose titration,

discontinuation and AEs in addition to GI and cardiovascular (CV) AEs) (14, 18, 20, 43-46); 3) a model developed by Fitzpatrick (a Markov model aimed at evaluating hip OA surgical treatments) (47-51); and 4) a decision-tree model developed by Burke to compare NSAID/COX-2 inhibitor oral analgesics (52-55) (Appendix 7).

The adopted modelling software was mentioned in 57% (n=49) studies, with TreeAge being the most common (n=38), followed by Microsoft Excel (n=8), and Arena (n=1). Two studies used more than one software, and no clear time trend was observed in the choice of software (Appendix 4).

Time horizon

Forty-eight (55%) studies ran the model over a lifetime horizon while 39 (45%) used short-to-medium-term horizons (from 2 weeks to 30 years). Markov model-based evaluations mostly adopted lifetime horizons (78%) while decision-tree evaluations largely used a pre-defined short-to-medium-term horizon (83%) (Figure 5c). The lifetime modelling horizon became more prevalent after 2010 (Figure 6a).

Discount rates for costs and outcomes

Seventy-six (86%) studies reported the discount rate, which was most commonly 3% (n=39, 51%) for costs and outcomes (Figure 6b). Ten (13%) and six (8%) studies utilised 3.5% and 5% discount rates, respectively. Thirty (of 39) studies using a 3% discount rate were based in the US, and 9 (of 10) studies using a 3.5% discount rate were based in the UK. Seven used alternative discount rates for costs and outcomes. Twelve studies did not report use of a discount rate, with 11 of these published before 2010 (Figure 6c). The impact of varying discount rates on model outcomes was assessed by 15 of 88 (17%) included studies.

Model health states

Pharmacological-focused studies defined health states based on: 1) the occurrence of AEs/complications (43); 2) OA severity, the presence of joint pain, obesity and comorbidities (41); and 3) the Kellgren-Lawrence radiographic scale of OA (56). For surgical-focused studies, health states were defined based on: 1) the American College of Rheumatologists (ACR) functional status classification (47); 2) the event pathway following surgical treatment (57); 3) pain severity, postoperative complications, and subsequent surgical procedures (58); and 4) the Western Ontario and McMaster Osteoarthritis Index (WOMAC) scores (13).

All surgical-focused studies considered revision surgery, and 23 considered two or more surgical revisions. Nine studies also considered surgery-related complications (e.g.: infection, bleeding, dislocation).

For pharmacological-focused studies, the most important modelling event was medical AEs/regimen toxicity. Studies conducted before 2002 only included GI AEs, while post-2002 studies also considered CV AEs and renal toxicity. Twenty studies included CV AEs, and eleven considered discontinuation (mostly published after 2013). Only three studies considered treatment adherence (55, 59, 60).

Uncertainty analysis

Uncertainty analysis was reported in 86 (98%) studies, with most (n=47) conducting more than one type of sensitivity analysis. The adopted methods included: (a) deterministic (univariate and/or multivariate) sensitivity analysis (DSA) (n=68), (b) probability sensitivity analysis (PSA) (n=43), (c) threshold analysis (n=4), and (d) scenario analysis (n=2). Five studies did not report the type of sensitivity analysis. Whilst DSA was popular across the reporting periods, PSA was first adopted in 2001 and became increasingly common thereafter. The most commonly evaluated parameters included costs, health state utilities (HSUs), probabilities of AEs, and treatment efficacy.

Reporting quality assessment

The reporting quality of all studies was assessed, with the exception of one study for which this was not applicable (61). The mean (standard deviation) score of all studies was 81% (9%). The reporting quality improved after 2001 and remained relatively high thereafter (Appendix 8a). Overall, 64 out of 87 studies exhibited a high reporting quality (>75%) (Appendix 8b). Whilst no study fully met CHEERS criteria, three achieved the highest reporting quality score of 96%. The CHEERS criteria were mostly met for the items of introduction, comparators, time horizon, resources and cost estimation, and uncertainty characterisation. However, the title (in terms of describing the interventions compared), abstract (in terms of the study perspective, setting, study inputs, and uncertainty analyses), and effectiveness measurement were commonly under-reported.

Discussion

This is the first study comprehensively reviewing the evolution of HEEMs for all forms of OA. Our review found that OA modelled evaluations are of a wide variety and have evolved substantially over time, with their emphasis and complexity shifting from pharmacological-focused short-to-medium term decision-tree models to surgical-focused lifetime Markov models. Existing HEEMs have limitations related to the choice of model input parameters, discount rates, and model health states/events. For instance, indirect costs related to OA were mostly not considered. Discount rates were mostly consistent with local guidelines, however, most studies failed to gauge the sensitivity of the model outcomes to discount rate changes. Most studies failed to consider important model events (e.g.: CV AEs), therapeutic adherence and treatment discontinuation. Despite clear guidelines (11), studies failed to pay adequate attention to lifestyle management, non-drug treatments and preventions. The reporting quality

of included studies was reasonably satisfactory, however, the title, abstract, and effectiveness measures were mostly reported inadequately.

Cost categories considered in OA models should be consistent with the perspective (62). In this review, however, more than half (15 of 24) of the studies using a societal perspective failed to incorporate indirect costs, and of the 14 focused on patients of working age (< 65 years), only six considered productivity losses. Similarly, when modelling is for older populations (37-39), special consideration should be given to include the potentially large contribution of informal care costs (63), but only one study in this review included this. Given the increasing popularity of the societal perspective in recent years, we recommend future studies include all relevant costs, particularly the indirect costs from lost wages/productivity (64).

The focus of modelled evaluations of OA interventions changed over time. Prior to 2005, studies focused on pharmacological treatments; subsequently the focus shifted to surgical treatments. Our finding is consistent with the increasing popularity of surgical treatment for OA since 2000 (65) but the availability of relevant longitudinal data from national joint replacement registries may also have contributed to this rise (66). Most pharmacological studies focused on all types of OA combined and/or general arthritis patients, as there is no evidence of different treatment effects of drugs between OA types (43). In contrast, most studies of surgery focused on a specific type of OA (e.g.: knee and/or hip OA). Other (e.g.: ankle and glenohumeral) joints attracted limited attention, as might be expected due to the relatively low prevalence of OA at those joint sites (67, 68). As the studies focusing on ankle and glenohumeral joints were published before 2010, updated studies of these joint sites should be on the agenda for future research.

A limited number of HEEMs focused on lifestyle, non-drug treatments and preventions. This is despite guidelines (11, 69) recommending such interventions before pharmacological and

surgical treatments. Because OA is a preventable, non-curable and progressive condition, it is critical that future studies investigate the impact of prevention and non-drug and non-surgical treatments on clinical and health-economic outcomes.

The most commonly used effectiveness measure was QALYs, with the proportion of studies reporting this increasing from 48% before 2008 to 93% thereafter. This is in accordance with national and international guidelines that recommend the use of QALYs (70). Importantly, the calculation of QALYs relies on HSUs that can be obtained from a variety of sources such as primary studies (37), systematic reviews (47) or randomized controlled trials (71). As different populations may value health states differently, caution is required when using non-locally derived HSUs (37, 72). Whilst disease-specific effectiveness measures (e.g.: revision free life years) provide information in more clinically relevant terms, these measures are not preference-based and also suffer from other limitations including the lack of comparability, and difficulty in trading-off across different diseases (73). Thus, the transition to use of QALYs reflects best practice, and use of QALYs based on local HSUs is recommended in future studies.

More than half studies applied lifetime horizon; however, this only became popular after 2010. This change could be related to two factors: first, Markov models became more popular in recent years which tend to adopt lifetime horizons due to their ability to take into account the re-occurrence of model events (74); and second, CV AEs of pharmacological treatments were taken into account in recent years which, comparing with GI AEs, have a larger impact on mortality (60). As OA is a chronic condition with on-going medical management, we recommend use of lifetime horizons in future studies, especially when using a Markov model structure (75) and in situations where an intervention is expected to influence mortality rates (76).

Discount rates varied between study settings. The use of 3% and 3.5% discount rates in most US and UK based studies aligned well with the local guidelines (77, 78). However, 64% (9 out of 14) of studies conducted in other nations did not use discount rates from national guidelines (15, 47, 60, 79). Reporting of the discount rate has improved over time, which is as expected from the recent introduction of CHEERS statement for the development and appraisal of health economic evaluations (28). A significant majority (83%) of included studies did not conduct sensitivity analyses for discount rates, which contrasts recommendations (80). Future studies should choose a discount rate that is in-line with local guidelines and assess the sensitivity of the model outcomes to discount rate changes.

The choice of model events varied between studies and over time. Pharmacological-focused models evolved in their complexity by incorporating more AEs (such as CV AEs) and regimen discontinuation. However, the models of surgical treatments (with revision surgery being the most important model event) did not evolve to the same extent. Differences existed between studies in terms of the number of considered revisions. Considering the ten-yearly cumulative re-revision rates of primary total knee (22.8%) and hip replacement (21.5%) and the factors influencing these rates (81), the times of revision surgery should be decided based on the age of the target population and the surgery techniques of interest. The high rates of OA medical AEs and possibility of revision surgery impose significant additional costs (40), likely impacting the outcomes of cost-effectiveness analyses (22). Future HEEMs should therefore incorporate all relevant OA and treatment related events and complications/revisions.

The choice of modelling method varied between studies and over time, with Markov models becoming more popular after 2008. This is consistent with numerous decision analytic modelling guidelines (74, 76, 82), and suits the chronic nature of OA with the possibility of recurrent health events (75). Consistent with previous findings (83, 84), TreeAge was the single

most popular modelling software followed by Microsoft Excel at all periods. Both software have been found equally reliable in conducting health-economic evaluations, and choice of software can depend on software availability, implementation skills, time constraints and end-user requirements (83, 84). We identified four popular OA HEEM structures, each with their own strengths and limitations. The original form of the NICE model included GI and CV AEs; however, treatment discontinuation and adherence were not included. The NICE model was extended in two studies by incorporating dose titration, discontinuation and additional AEs (44, 46). However, data on discontinuation is not easily available. The OAPoL model accounts for the inter-relationships among key variables such as the function of pain, obesity, and comorbidities. However, the inherent dependence of this model on scarce data for key variables limits its widespread use. The model developed by Fitzpatrick can be easily adapted to suit alternative settings and study objectives. However, it is considered too simple to fully evaluate the various outcomes between alternative surgeries (37, 85). Lastly, the model developed by Burke considered the severity of GI AEs and can easily be adapted to different settings (53, 54). However, it fails to consider other events such as CV AEs, discontinuation, and its time horizon of <1-year is not well suited to the chronic nature of OA.

PSA showed increasing popularity after 2001 which could be the result of an increasing awareness of PSA's importance in health economics over time (86) and advances in computational technologies. Limited studies evaluated the sensitivity of model outcomes to important input parameters including OA definitions and population parameters. We recommend considering these relatively neglected aspects in future modelling studies.

We recommend the use of Markov models due to their ability to incorporate repetitive (short and long-term) health events, including important medical AEs, therapeutic adherence, and discontinuation. Importantly, the probabilities of medical AEs, and time to events (e.g.:

decision for joint replacement, revision surgery) depend on the history of previous states; however, no study in our review considered this important dimension. We therefore recommend future studies to integrate memory into their models to avoid problems associated with the Markov assumption of memorylessness. Future HEEMs should also benefit from the recent availability of MRI-based data on OA definitions, progression and MRI-based markers (19), and advances in new data science (that have enabled the use of machine learning-based patient-specific prediction models) (87).

This review found the reporting quality of studies has improved and has been reasonably satisfactory over time. Nonetheless, further improvements could be made, particularly, in relation to the quality of title, abstract, and effectiveness measures. As poor reporting may lead to costly decisions, future studies should ensure high transparency and reporting quality in all areas.

The strengths of our review include its comprehensive nature and the incorporation of assessment of reporting quality. The inclusion of all OA therapies builds on two existing systematic reviews of OA oral therapies and surgical interventions (21, 22). Furthermore, our review summarised the development of OA models in terms of various model characteristics, which will help to evaluate the existing OA HEEMs and guide the development of a comprehensive gold standard HEEM of OA. Lastly, our assessment of reporting quality should be of interest to future researchers in improving the reporting quality of their health-economic modelling studies. A limitation is that our study did not cover studies published in languages other than English, Chinese and German. However, this may have a minimal impact on our key conclusions as only a small number of studies (n=20) were subject to language exclusions. A further limitation is that as we only had a small number of studies comparing OA preventative interventions, the review is predominantly focused on OA treatments. Future

studies will need to consider the synthesis of OA prevention models as more evidence becomes available.

Conclusion

OA HEEMs are of a wide variety and evolved substantially over time. Furthermore, the number of modelled OA evaluations have rapidly increased in recent years. The focus of OA HEEs has shifted from short-to-medium-term pharmacological-focused decision-tree models to surgical-focused lifetime Markov models. We recommend future HEEMs use life-time Markovian model structures with memory integration and should also incorporate all relevant costs, model events, therapeutic adherence, discontinuation, appropriate discount rate and time horizon, and conduct sensitivity analyses for input parameters. Finally, we recommend improvements in relation to the reporting quality of (1) title, (2) abstract, and (3) effectiveness measurement.

Authors' contributions

TZ, BdeG, DA, TW and AP conceived and designed the methods. TZ and QX conducted literature search and screening, TZ and HA conducted the data extraction and reporting quality assessment. All authors were involved in the manuscript preparation. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to thank the librarian researcher Michaela for her help on building the literature search strategy and Covidence Help Centre for facilitating the progression of literature screening.

References

1. Osteoarthritis Research Society International. Osteoarthritis: a serious disease. https://www.oarsi.org/sites/default/files/docs/2016/oarsi_white_paper_oa_serious_disease_121416_1.pdf.
2. Hales CM, Carroll MD, Fryar CD, Ogden CL. Prevalence of obesity among adults and youth: United States, 2015-2016. NCHS Data Brief. 2017(288):1-8.
3. Sisson SB, Katzmarzyk PT. International prevalence of physical activity in youth and adults. *Obes Rev*. 2008;9(6):606-14.
4. Wilson R, Blakely T, Abbott JH. Radiographic knee osteoarthritis impacts multiple dimensions of health-related quality of life: data from the Osteoarthritis Initiative. *Rheumatology (Oxford)*. 2018;57(5):891-9.
5. Huang Z, Ding C, Li T, Yu SP. Current status and future prospects for disease modification in osteoarthritis. *Rheumatology (Oxford)*. 2018;57(suppl_4):iv108-iv23.
6. Abbott JH, Usiskin IM, Wilson R, Hansen P, Losina E. The quality-of-life burden of knee osteoarthritis in New Zealand adults: A model-based evaluation. *PLoS One*. 2017;12(10):e0185676.
7. Henderson JV, Harrison CM, Britt HC, Bayram CF, Miller GC. Prevalence, causes, severity, impact, and management of chronic pain in Australian general practice patients. *Pain Med*. 2013;14(9):1346-61.
8. March LM, Bachmeier CJ. Economics of osteoarthritis: a global perspective. *Baillieres Clin Rheumatol*. 1997;11(4):817-34.
9. Puig-Junoy J, Ruiz Zamora A. Socio-economic costs of osteoarthritis: A systematic review of cost-of-illness studies. *Semin Arthritis Rheum*. 2015;44(5):531-41.

10. Sharif B, Garner R, Hennessy D, Sanmartin C, Flanagan WM, Marshall DA. Productivity costs of work loss associated with osteoarthritis in Canada from 2010 to 2031. *Osteoarthritis Cartilage*. 2017;25(2):249-58.
11. The Royal Australian College of General Practitioners. Guideline for the management of knee and hip osteoarthritis. 2nd edn. East Melbourne, Vic: RACGP, 2018.
12. Rabarison KM, Bish CL, Massoudi MS, Giles WH. Economic evaluation enhances public health decision making. *Frontiers in public health*. 2015;3:164.
13. Losina E, Walensky RP, Kessler CL, Emrani PS, Reichmann WM, Wright EA, et al. Cost-effectiveness of total knee arthroplasty in the United States: patient risk and hospital volume. *Arch Intern Med*. 2009;169(12):1113-21; discussion 21-2.
14. Brereton N, Winn B, Akehurst R. The cost-effectiveness of celecoxib vs diclofenac in the treatment of osteoarthritis in the UK; an update to the NICE model using data from the CONDOR trial. *J Med Econ*. 2012;15(3):465-72.
15. Higashi H, Barendregt JJ. Cost-effectiveness of total hip and knee replacements for the Australian population with osteoarthritis: discrete-event simulation model. *PLoS One*. 2011;6(9):e25403.
16. Chang RW, Pellisier JM, Hazen GB. A cost-effectiveness analysis of total hip arthroplasty for osteoarthritis of the hip. *Jama*. 1996;275(11):858-65.
17. Haglund U, Svarvar P. The Swedish ACCES model: predicting the health economic impact of celecoxib in patients with osteoarthritis or rheumatoid arthritis. *Rheumatology (Oxford)*. 2000;39 Suppl 2:51-6.
18. Latimer N, Lord J, Grant RL, O'Mahony R, Dickson J, Conaghan PG, et al. Cost effectiveness of COX 2 selective inhibitors and traditional NSAIDs alone or in combination with a proton pump inhibitor for people with osteoarthritis. *BMJ*. 2009;339:b2538.

19. Losina E, Daigle ME, Suter LG, Hunter DJ, Solomon DH, Walensky RP, et al. Disease-modifying drugs for knee osteoarthritis: can they be cost-effective? *Osteoarthritis Cartilage*. 2013;21(5):655-67.
20. Latimer N, Lord J, Grant RL, O'Mahony R, Dickson J, Conaghan PG. Value of information in the osteoarthritis setting: cost effectiveness of COX-2 selective inhibitors, traditional NSAIDs and proton pump inhibitors. *Pharmacoeconomics*. 2011;29(3):225-37.
21. Kamaruzaman H, Kinghorn P, Oppong R. Cost-effectiveness of surgical interventions for the management of osteoarthritis: a systematic review of the literature. *BMC Musculoskelet Disord*. 2017;18(1):183.
22. Wielage RC, Myers JA, Klein RW, Happich M. Cost-effectiveness analyses of osteoarthritis oral therapies: a systematic review. *Appl Health Econ Health Policy*. 2013;11(6):593-618.
23. Pinto D, Robertson MC, Hansen P, Abbott JH. Cost-effectiveness of nonpharmacologic, nonsurgical interventions for hip and/or knee osteoarthritis: systematic review. *Value Health*. 2012;15(1):1-12.
24. Si L, Winzenberg TM, Palmer AJ. A systematic review of models used in cost-effectiveness analyses of preventing osteoporotic fractures. *Osteoporos Int*. 2014;25(1):51-60.
25. Si L, Winzenberg TM, Jiang Q, Palmer AJ. Screening for and treatment of osteoporosis: construction and validation of a state-transition microsimulation cost-effectiveness model. *Osteoporos Int*. 2015;26(5):1477-89.
26. Moher D, Liberati A, Tetzlaff J, Altman DG, Group atP. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*. 2009;151(4):264-9.

27. Better systematic review management C, accessed: <https://www.covidence.org/home>, dated 29 January 2019.
28. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated health economic evaluation reporting standards (CHEERS) statement. *Bmj*. 2013;346:f1049.
29. de Graaff B, Neil A, Sanderson K, Si L, Yee KC, Palmer AJ. A systematic review and narrative synthesis of health economic studies conducted for hereditary haemochromatosis. *Appl Health Econ Health Policy*. 2015;13(5):469-83.
30. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the conduct of narrative synthesis in systematic reviews. A product from the ESRC methods programme Version. 2006;1:b92.
31. McKell C, Stewart A. Cost-minimisation analysis comparing topical versus systemic NSAIDs in the treatment of mild osteoarthritis of the superficial joints. *British Journal of Medical Economics*. 1994;7:137-46.
32. Faucett SC, Geisler BP, Chahla J, Krych AJ, Kurzweil PR, Garner AM, et al. Meniscus root repair vs meniscectomy or nonoperative management to prevent knee osteoarthritis after medial meniscus root tears: clinical and economic effectiveness. *Am J Sports Med*. 2018:363546518755754.
33. Rongen JJ, Govers TM, Buma P, Grutters JP, Hannink G. Societal and economic effect of meniscus scaffold procedures for irreparable meniscus injuries. *Am J Sports Med*. 2016;44(7):1724-34.
34. Feeley BT, Liu S, Garner AM, Zhang AL, Pietzsch JB. The cost-effectiveness of meniscal repair versus partial meniscectomy: A model-based projection for the United States. *The Knee*. 2016;23(4):674-80.

35. Losina E, Burbine SA, Suter LG, Hunter DJ, Solomon DH, Daigle ME, et al. Pharmacologic regimens for knee osteoarthritis prevention: can they be cost-effective? *Osteoarthritis Cartilage*. 2014;22(3):415-30.
36. SooHoo NF, Kominski G. Cost-effectiveness analysis of total ankle arthroplasty. *J Bone Joint Surg Am*. 2004;86-a(11):2446-55.
37. Kunkel ST, Sabatino MJ, Kang R, Jevsevar DS, Moschetti WE. The cost-effectiveness of total hip arthroplasty in patients 80 years of age and older. *J Arthroplasty*. 2018;33(5):1359-67.
38. Katz JN, Smith SR, Collins JE, Solomon DH, Jordan JM, Hunter DJ, et al. Cost-effectiveness of nonsteroidal anti-inflammatory drugs and opioids in the treatment of knee osteoarthritis in older patients with multiple comorbidities. *Osteoarthritis Cartilage*. 2016;24(3):409-18.
39. Slover J, Espehaug B, Havelin LI, Engesaeter LB, Furnes O, Tomek I, et al. Cost-effectiveness of unicompartmental and total knee arthroplasty in elderly low-demand patients. A Markov decision analysis. *J Bone Joint Surg Am*. 2006;88(11):2348-55.
40. Losina E, Usiskin IM, Smith SR, Sullivan JK, Smith KC, Hunter DJ, et al. Cost-effectiveness of generic celecoxib in knee osteoarthritis for average-risk patients: a model-based evaluation. *Osteoarthritis Cartilage*. 2018;26(5):641-50.
41. Smith SR, Katz JN, Collins JE, Solomon DH, Jordan JM, Suter LG, et al. Cost-effectiveness of tramadol and oxycodone in the treatment of knee osteoarthritis. *Arthritis Care Res (Hoboken)*. 2017;69(2):234-42.
42. Losina E, Michl G, Collins JE, Hunter DJ, Jordan JM, Yelin E, et al. Model-based evaluation of cost-effectiveness of nerve growth factor inhibitors in knee osteoarthritis: impact of drug cost, toxicity, and means of administration. *Osteoarthritis Cartilage*. 2016;24(5):776-85.

43. Nasef SA, Shaaban AA, Mould-Quevedo J, Ismail TA. The cost-effectiveness of celecoxib versus non-steroidal anti-inflammatory drugs plus proton-pump inhibitors in the treatment of osteoarthritis in Saudi Arabia. *Health Econ Rev.* 2015;5(1):53.
44. Wielage RC, Patel AJ, Bansal M, Lee S, Klein RW, Happich M. Cost effectiveness of duloxetine for osteoarthritis: a Quebec societal perspective. *Arthritis care & research.* 2014;66(5):702-8.
45. Brereton N, Pennington B, Ekelund M, Akehurst R. A cost-effectiveness analysis of celecoxib compared with diclofenac in the treatment of pain in osteoarthritis (OA) within the Swedish health system using an adaptation of the NICE OA model. *J Med Econ.* 2014;17(9):677-84.
46. Wielage RC, Bansal M, Andrews JS, Klein RW, Happich M. Cost-utility analysis of duloxetine in osteoarthritis: a US private payer perspective. *Appl Health Econ Health Policy.* 2013;11(3):219-36.
47. Mujica-Mota RE, Watson LK, Tarricone R, Jager M. Cost-effectiveness of timely versus delayed primary total hip replacement in Germany: A social health insurance perspective. *Orthop Rev (Pavia).* 2017;9(3):7161.
48. Pulikottil-Jacob R, Connock M, Kandala NB, Mistry H, Grove A, Freeman K, et al. Has metal-on-metal resurfacing been a cost-effective intervention for health care providers?—A registry based study. *PLoS One.* 2016;11(11):e0165021.
49. Pulikottil-Jacob R, Connock M, Kandala NB, Mistry H, Grove A, Freeman K, et al. Cost effectiveness of total hip arthroplasty in osteoarthritis: comparison of devices with differing bearing surfaces and modes of fixation. *Bone Joint J.* 2015;97-b(4):449-57.
50. Mota RE. Cost-effectiveness analysis of early versus late total hip replacement in Italy. *Value Health.* 2013;16(2):267-79.

51. Fitzpatrick R, Shortall E, Sculpher M, Murray D, Morris R, Lodge M, et al. Primary total hip replacement surgery: a systematic review of outcomes and modelling of cost-effectiveness associated with different prostheses. *Health Technol Assess.* 1998;2(20):1-64.
52. Chancellor JV, Hunsche E, de Cruz E, Sarasin FP. Economic evaluation of celecoxib, a new cyclo-oxygenase 2 specific inhibitor, in Switzerland. *Pharmacoeconomics.* 2001;19(1):59-75.
53. Zabinski RA, Burke TA, Johnson J, Lavoie F, Fitzsimon C, Tretiak R, et al. An economic model for determining the costs and consequences of using various treatment alternatives for the management of arthritis in Canada. *Pharmacoeconomics.* 2001;19 Suppl 1:49-58.
54. You JH, Lee KK, Chan TY, Lau WH, Chan FK. Arthritis treatment in Hong Kong--cost analysis of celecoxib versus conventional NSAIDS, with or without gastroprotective agents. *Aliment Pharmacol Ther.* 2002;16(12):2089-96.
55. Al MJ, Maniadakis N, Grijseels EW, Janssen M. Costs and effects of various analgesic treatments for patients with rheumatoid arthritis and osteoarthritis in the Netherlands. *Value Health.* 2008;11(4):589-99.
56. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. *Ann Rheum Dis.* 1957;16(4):494-502.
57. Chawla H, Nwachukwu B, Van Der List J, Eggman A, Pearle A, Ghomrawi H. Cost effectiveness of patellofemoral versus total knee arthroplasty in younger patients. *The bone & joint journal.* 2017;99(8):1028-36.
58. Konopka JF, Gomoll AH, Thornhill TS, Katz JN, Losina E. The cost-effectiveness of surgical treatment of medial unicompartmental knee osteoarthritis in younger patients: a computer model-based evaluation. *The Journal of bone and joint surgery American volume.* 2015;97(10):807.

59. Capel M, Tornero J, Zamorano JL, Oyaguez I, Casado MA, Sanchez-Covisa J, et al. Efficiency of naproxen/esomeprazole in association for osteoarthritis treatment in Spain. *Reumatol Clin*. 2014;10(4):210-7.
60. Bessette L, Risebrough N, Mittmann N, Roussy J-P, Ho J, Zlateva G. Cost-utility of celecoxib use in different treatment strategies for osteoarthritis and rheumatoid arthritis from the Quebec healthcare system perspective. *Journal of medical economics*. 2009;12(3):246-58.
61. Wilson R, Abbott JH. Development and validation of a new population-based simulation model of osteoarthritis in New Zealand. *Osteoarthritis Cartilage*. 2018;26(4):531-9.
62. Garrison LP, Pauly MV, Willke RJ, Neumann PJ. An overview of value, perspective, and decision context—a health economics approach: an ISPOR Special Task Force report [2]. *Value in Health*. 2018;21(2):124-30.
63. Gupta S, Hawker GA, Laporte A, Croxford R, Coyte PC. The economic burden of disabling hip and knee osteoarthritis (OA) from the perspective of individuals living with this condition. *Rheumatology (Oxford)*. 2005;44(12):1531-7.
64. Koenig L, Zhang Q, Austin MS, Demiralp B, Fehring TK, Feng C, et al. Estimating the societal benefits of THA after accounting for work status and productivity: a Markov model approach. *Clinical Orthopaedics and Related Research®*. 2016;474(12):2645-54.
65. Maradit Kremers H, Larson DR, Crowson CS, Kremers WK, Washington RE, Steiner CA, et al. Prevalence of total hip and knee replacement in the United States. *J Bone Joint Surg Am*. 2015;97(17):1386-97.
66. Ayers DC, Franklin PD. Joint replacement registries in the United States: a new paradigm. *The Journal of bone and joint surgery American volume*. 2014;96(18):1567.
67. Murray C, Marshall M, Rathod T, Bowen CJ, Menz HB, Roddy E. Population prevalence and distribution of ankle pain and symptomatic radiographic ankle osteoarthritis

in community dwelling older adults: A systematic review and cross-sectional study. PLoS One. 2018;13(4):e0193662.

68. Cameron ML, Kocher MS, Briggs KK, Horan MP, Hawkins RJ. The prevalence of glenohumeral osteoarthritis in unstable shoulders. The American journal of sports medicine. 2003;31(1):53-5.

69. Conditions NCCfC, Excellence NifC, editors. Osteoarthritis: national clinical guidelines for care and management in adults 2008: Royal College of Physicians.

70. Hiligsmann M, Cooper C, Arden N, Boers M, Branco JC, Brandi ML, et al., editors. Health economics in the field of osteoarthritis: an expert's consensus paper from the European Society for Clinical and Economic Aspects of Osteoporosis and Osteoarthritis (ESCEO). Seminars in arthritis and rheumatism; 2013: Elsevier.

71. Bove AM, Smith KJ, Bise CG, Fritz JM, Childs J, Brennan GP, et al. Exercise, manual therapy, and booster sessions in knee osteoarthritis: cost-effectiveness analysis from a multicenter randomized controlled trial. Phys Ther. 2018;98(1):16-27.

72. Peersman G, Jak W, Vandenlangenbergh T, Jans C, Cartier P, Fennema P. Cost-effectiveness of unicondylar versus total knee arthroplasty: a Markov model analysis. Knee. 2014;21 Suppl 1:S37-42.

73. Gray AM, Clarke PM, Wolstenholme JL, Wordsworth S. Applied methods of cost-effectiveness analysis in healthcare: Oxford University Press; 2011.

74. Barton P, Bryan S, Robinson S. Modelling in the economic evaluation of health care: selecting the appropriate approach. Journal of health services research & policy. 2004;9(2):110-8.

75. Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment.

NIHR Health Technology Assessment programme: Executive Summaries: NIHR Journals Library; 2004.

76. Drugs CAf, Health Ti. Guidelines for the economic evaluation of health technologies: Canada. Guidelines for the economic evaluation of health technologies: Canada: CADTH; 2006.

77. Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the Panel on Cost-effectiveness in Health and Medicine. *Jama*. 1996;276(15):1253-8.

78. Smith WB, Steinberg J, Scholtes S, Mcnamara IR. Medial compartment knee osteoarthritis: age-stratified cost-effectiveness of total knee arthroplasty, unicompartmental knee arthroplasty, and high tibial osteotomy. *Knee Surgery, Sports Traumatology, Arthroscopy*. 2017;25(3):924-33.

79. Attema AE, Brouwer WBF, Claxton K. Discounting in economic evaluations. *Pharmacoeconomics*. 2018;36(7):745-58.

80. Adalsteinsson E, Toumi M. Benefits of probabilistic sensitivity analysis—a review of NICE decisions. *Journal of market access & health policy*. 2013;1(1):21240.

81. Australian Orthopaedic Association, National Joint Replacement Registry, Hip and Knee Arthroplasty, <https://aoanjrr.sahmri.com/documents/10180/172286/Annual+Report+2014>, Accessed: March 15 2019.

82. Soto J. Health economic evaluations using decision analytic modeling: principles and practices—utilization of a checklist to their development and appraisal. *International journal of technology assessment in health care*. 2002;18(1):94-111.

83. Tsoi B, O'Reilly D, Jegathisawaran J, Tarride J, Blackhouse G, Goeree R. Systematic narrative review of decision frameworks to select the appropriate modelling approaches for health economic evaluations. *BMC research notes*. 2015;8(1):244.

84. Menn P, Holle R. Comparing three software tools for implementing markov models for health economic evaluations. *Pharmacoeconomics*. 2009;27(9):745-53.
85. Ghomrawi HM, Eggman AA, Pearle AD. Effect of age on cost-effectiveness of unicompartmental knee arthroplasty compared with total knee arthroplasty in the US. *The Journal of bone and joint surgery American volume*. 2015;97(5):396.
86. Baio G, Dawid AP. Probabilistic sensitivity analysis in health economics. *Statistical Methods in Medical Research*. 2015;24(6):615-34.
87. Jamshidi A, Pelletier J-P, Martel-Pelletier J. Machine-learning-based patient-specific prediction models for knee osteoarthritis. *Nature Reviews Rheumatology*. 2019;15(1):49-60.

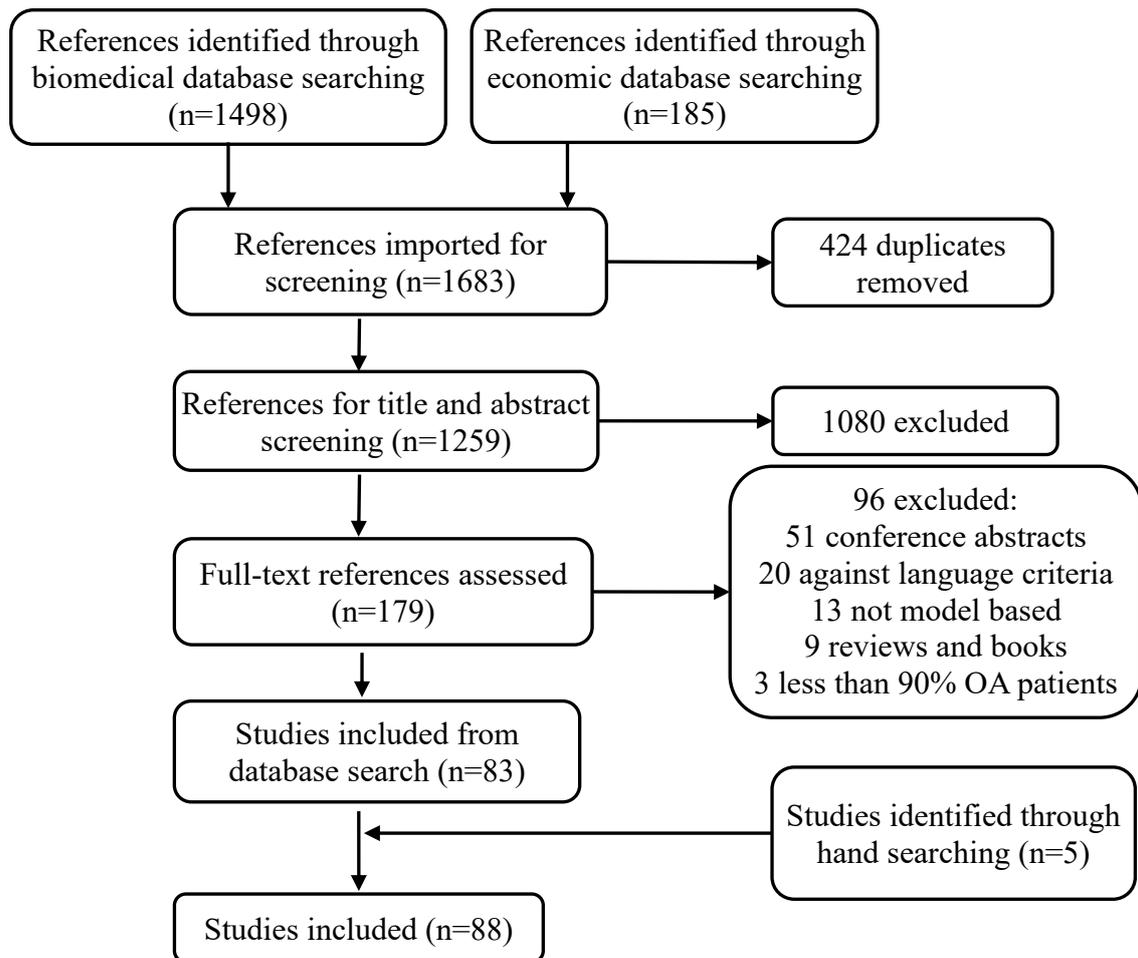


Figure 1. Flow chart results of study search based on Preferred Reporting Items for Systematic Reviews and Meta-Analyses methodology.

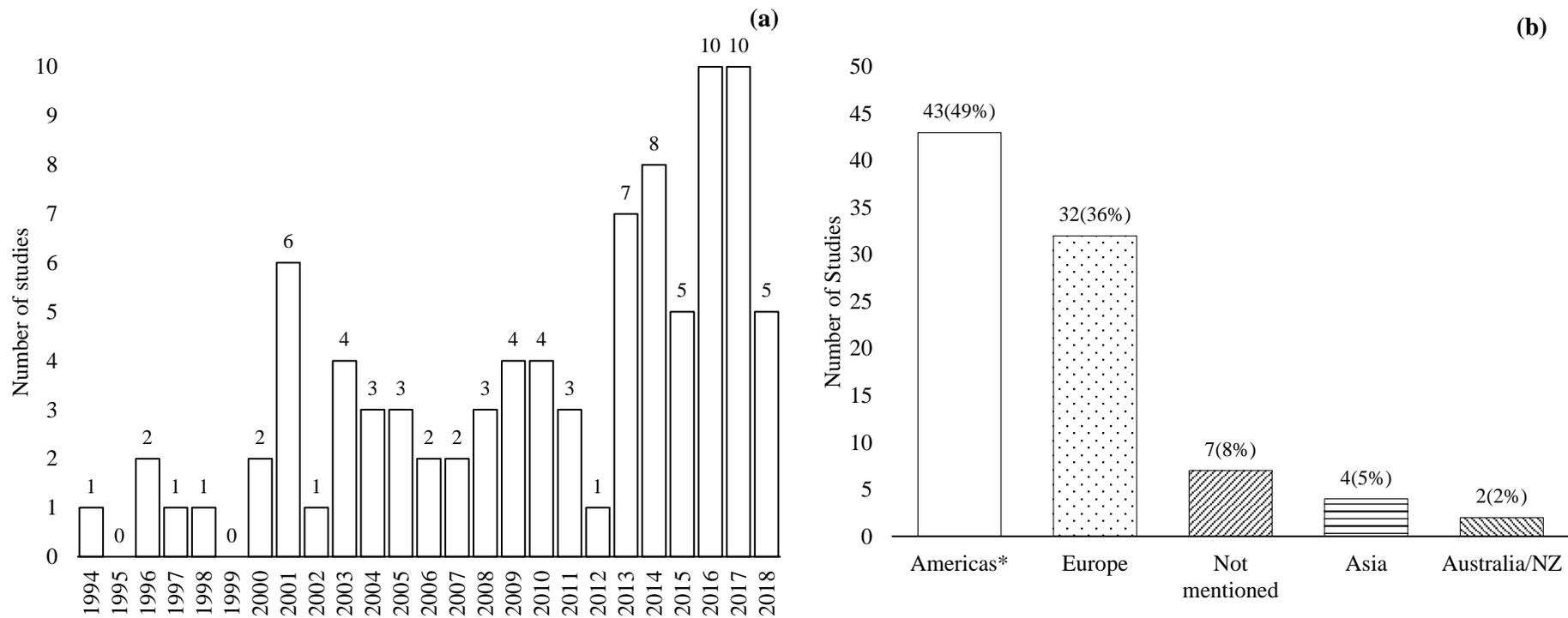


Figure 2. The distribution of published model-based osteoarthritis (OA) health economic evaluations by years of publication (a) and study settings (b). *Americas included both North and South America. NZ=New Zealand.

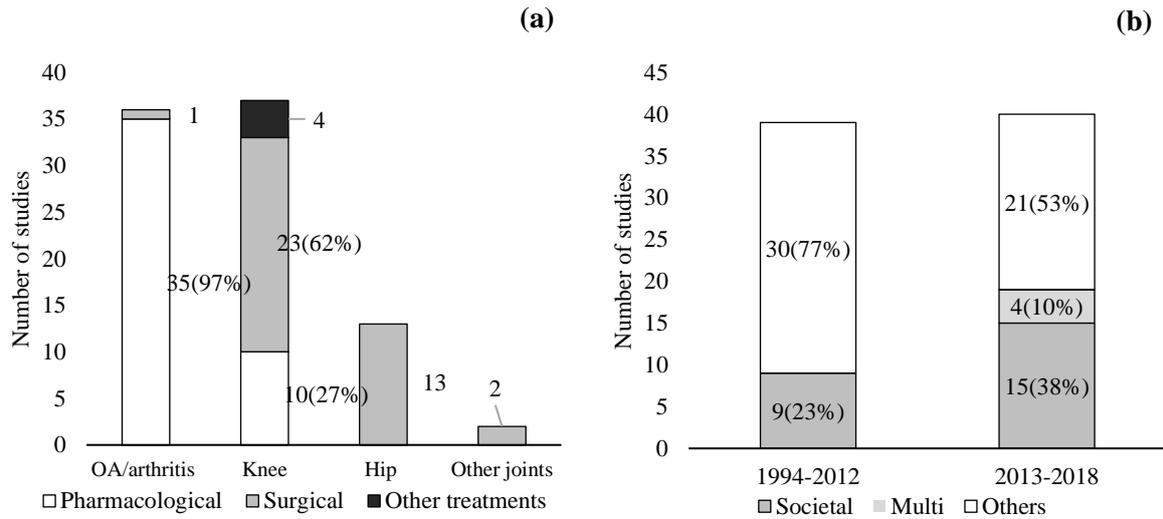


Figure 3. The distribution of included studies by focused OA types and intervention options (a) and the distribution of adopted perspectives by published time (n=79) (b). OA/arthritis includes OA without a specified site and arthritis comprising OA and rheumatoid arthritis. OA=osteoarthritis.

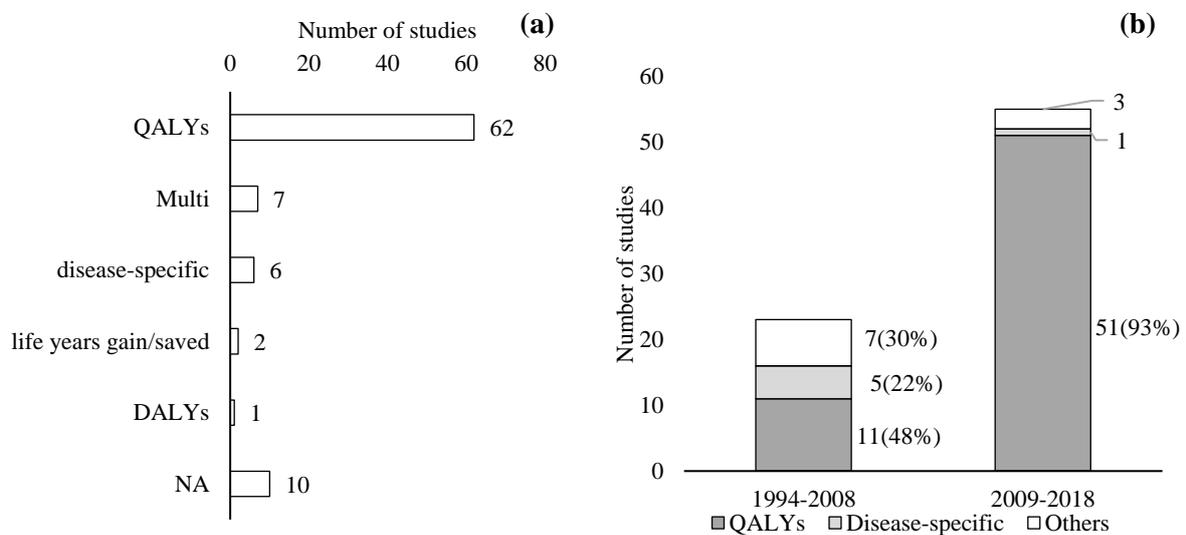


Figure 4. The effectiveness measures adopted in the included studies (a) and distribution by published time (n=78) (b). QALYs=quality-adjusted life years, DALYs=disability-adjusted life years, NA=not applicable.

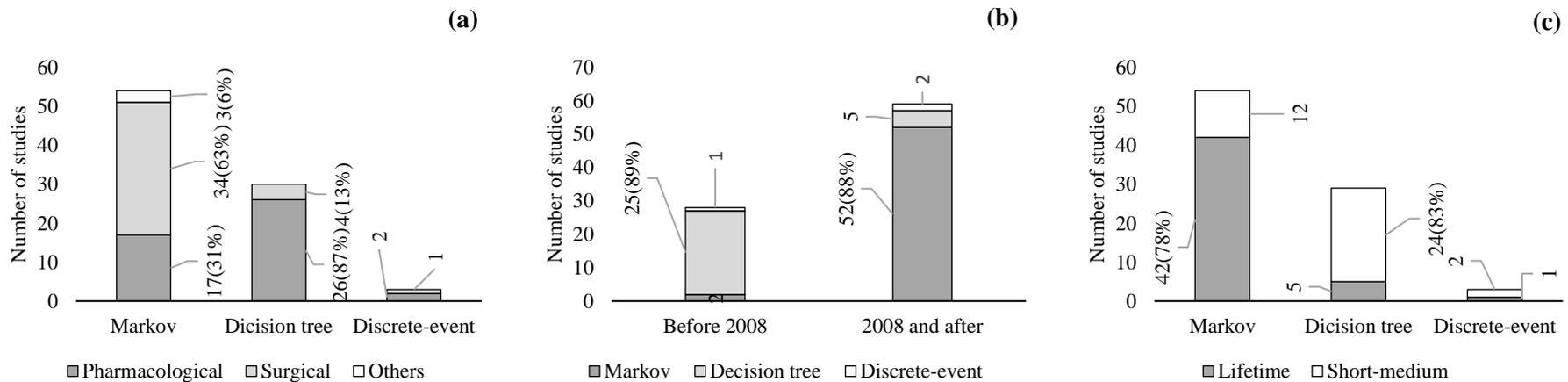


Figure 5. The number distribution of model types and treatments (a), by published time (b), and time horizon (c).

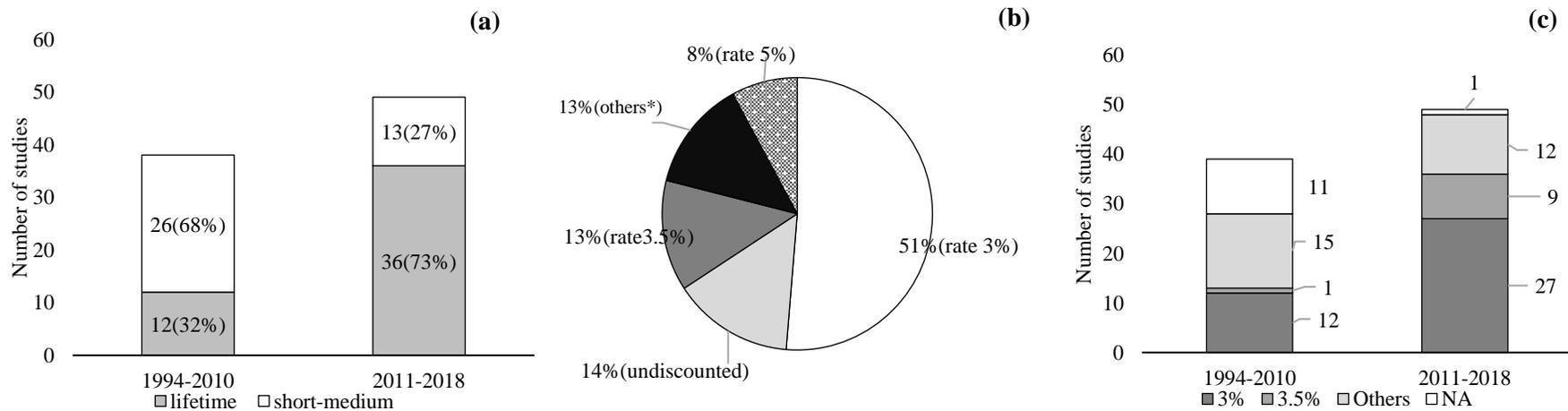


Figure 6. The distribution of adopted time horizon in included studies by published time (a), the percentage distribution of the adopted discount rates in the included modelling studies (n=76) (b), and the distribution of the adopted discount rates by published time (c). *Others included alternative discount rate for cost and outcomes, discount rate of 4%, 4.76%, and 6%.