International Journal of Technology Assessment in Health Care, 34:5 (2018), 481–497.

doi:10.1017/S0266462318000533

COST-EFFECTIVENESS OF INTERVENTIONS BASED ON PHYSICAL ACTIVITY IN THE TREATMENT OF CHRONIC CONDITIONS: A SYSTEMATIC LITERATURE REVIEW

Marlène Guillon
Université Clermont Auvergne, CNRS, CERDI and Hospinnomics – PSE – École d'économie de Paris, Assistance publique Hôpitaux de Paris
marlene.guillon@uca.fr

Lise Rochaix
Université Paris 1 and Hospinnomics - PSE - École d'économie de Paris, Assistance publique Hôpitaux de Paris

Jean-Claude K. Dupont
Hospinnomics - PSE - École d'économie de Paris, Assistance publique Hôpitaux de Paris

Objectives: The aim of this study is to review evidence on the cost-effectiveness of exercise-based interventions in the treatment of chronic conditions a decade after the publication of Roine et al. in 2009 (Roine E, Roine RP, Räsänen P, et al. *Int J Technol Assess Health Care.* 2009;25:427–454).
Methods: We carried out a review of published articles in PUBMED and JSTOR between January 1, 2008, and December 31, 2016. Full economic evaluations of exercise programs targeting patients with a chronic condition were eligible for inclusion. Data on program, design, and economic characteristics were extracted using a predefined extraction form. The quality of the economic evaluations was appraised using the adjusted Consensus Health Economic Criteria List.
Results: A total of 426 articles were identified and thirty-seven studies were selected. Eleven studies dealt with musculoskeletal and rheumatologic disorders, ten with cardiovascular diseases, six with neurological disorders, three with mental illnesses, three with cancers, and four with diabetes, respiratory diseases, or pelvic organ prolapse. In total, 60 percent of exercise programs were dominant or cost-effective. For musculoskeletal and rheumatologic disorders, 72 percent of programs were dominant or cost-effective while this was the case for 57 percent of programs for cardiovascular diseases using a nonsurgical comparator.
Conclusions: There is clear evidence in favor of exercise-based programs for the treatment of musculoskeletal and rheumatologic disorders and, to a lesser extent, for the treatment of cardiovascular diseases. More research is needed to evaluate the cost-effectiveness of physical activity in the treatment of neurological disorders, mental illnesses, cancers, respiratory diseases, and diabetes/obesity.

Keywords: Exercise, Cost-benefit analysis, Chronic disease, Review

Effectiveness of physical activity in the treatment of chronic conditions such as psychiatric diseases, metabolic diseases, cardiovascular diseases, pulmonary diseases, musculoskeletal disorders, or cancers is now well established (1). Based on clinical evidence, in several countries such as Sweden or the United Kingdom, general practitioners can prescribe physical activity for at risk and chronically ill patients. Where there is a recognized clinical benefit to adding physical activity interventions to usual care in the treatment of a chronic condition, the cost-effectiveness of these interventions needs to be assessed for optimizing allocation of health care resources.

In a previous systematic literature review, Roine et al. (2) identified sixty-five studies focusing on the cost-effectiveness of exercise programs in the treatment of various diseases. Most of these studies focused on chronic conditions such as musculoskeletal and rheumatic disorders or cardiovascular diseases. The authors found large variations in the cost-effectiveness of interventions based on physical exercise. They nevertheless concluded that some kinds of exercise interventions can be cost-effective, especially for the treatment of cardiovascular diseases and low back pain, despite partly contradictory findings.

Since the work of Roine et al. in 2009 (2), no recent review gathered available evidence on the cost-effectiveness of exercise-based programs in the treatment of chronic conditions. Our objective is to provide an up-to-date literature review. We systematically review and evaluate the methodological quality of the economic evaluations of physical activity programs among chronically ill patients published since 2008.

METHODS

We used a predefined research protocol for inclusion criteria and methods of analysis. This protocol was not registered in the international prospective register of systematic reviews (PROSPERO) before the start of the study. The review was conducted according to PRISMA guidelines.

Information Sources and Search Strategy

To identify relevant articles published since 2008, two databases were searched using keywords for the period between January 1, 2008 and December 31, 2016 (last search November 23, 2017): PUBMED and JSTOR. JSTOR is a multidisciplinary database of academic content which was

searched to identify health economics articles from journals not indexed in PubMed. Search terms were classified in two categories relating to physical activity or economic analysis. In the two databases the following keywords were used as general search terms in titles: "cost-effectiveness", "cost-benefit", "cost-utility", "economic evaluation", "economic analysis" or "economic impact" for the economic analysis category and "physical activity", "sport", "exercise", "training", "strength", "fitness", "running", "walking", "swimming", or "gymnastics" for the physical activity category. Articles including at least one search term of each category in their title were selected for inclusion. In addition, we used the following major Medical Subject Headings (MeSH) terms to search for articles in PUBMED: "costs and cost analysis" and "exercise therapy", "sports", "exercise", "physical fitness". Supplementary File 1 provides the full search strategy for each database. References in literature reviews identified through the keywords search were screened and were included if they met all inclusion criteria. The result of the systematic search was recorded in Zotero®, in particular to remove duplicates.

Selection Criteria

Full economic evaluations of exercise programs, either based on the results of a randomized controlled trial (RCT) or using a model based on one or several RCTs, comparing costs and health benefits of two or more interventions targeting patients with chronic conditions were included. Studies were excluded if not in English or French or if published before 2008 or after 2016. Studies that did not report on original data (i.e., commentaries, editorials, case studies and study protocols) or studies already included in the previous literature review of Roine et al. (2) were excluded. Furthermore, studies were excluded if the exercise program targeted nonchronically ill individuals, if they compared programs including different types of physical activity or if they provided only indirect evaluations of physical activity programs (e.g., media campaigns promoting physical activity or counselling without participation in an exercise program). We chose to exclude multicomponent programs for which at least one component was not physical activity, unless strictly related to enhancing participation in the physical activity component. For instance, studies mixing physical activity with weight management or a psychological intervention were excluded. On the other hand, studies including techniques to enhance participation along with the exercise program under evaluation were included. We included studies using either general or disease-specific health benefit measures. Disease-specific measures can provide valuable information to compare health benefits within the same disease category and their use might be needed when generic benefit measures such as quality-adjusted life-year (QALY) lack sensitivity to capture the effects of the exercise program on health.

Study Selection

In a first step, M.G. undertook the systematic keywords search in the two databases and performed the first eligibility assessment based on titles and abstracts following the predefined inclusion and exclusion criteria determined by the three authors (M.G., L.R., J.C.K.D.). Articles were classified in two categories: "no" if the article clearly violated one of the inclusion criteria or met one of the exclusion criteria and "maybe" when there was uncertainty. Titles and abstracts of articles in the "maybe" category were read by the two other researchers (L.R. and J.C.K.D.) and the decision about their inclusion in the next step was agreed upon consensually. In the second step, full-text reading of potentially eligible articles was performed by M.G. to determine final eligibility. Articles excluded through this second step were read by the two other researchers (L.R. and J.C.K.D.) and in case of doubt the decision was made through consensus. Data were then extracted for all articles meeting inclusion criteria. Data extraction was performed independently by the three researchers on a sample of ten articles to identify lack of consistency in data extraction. Data extraction on the remaining articles was performed by M.G.

Data Collection and Quality Evaluation

Using a predefined form, we extracted data on the following main categories: pathology, characteristics of the study population, exercise program and comparator(s), type and measurement of costs, type and measurement of health outcomes, design, results and uncertainty analyses performed. We adopted three classification categories regarding the cost measurement perspective. We considered that a study used a health care perspective if resource use from the health system for the program and health care consumptions of patients over the study period were taken into account. We classified the perspective of the study as "health and social care" if resource use from social services was also included. Finally, the perspective was held as societal if the study took into account at least one of the following costs: productivity losses, opportunity cost of time spent exercising or cost of informal care. We classified the results of the economic evaluations based on the broader cost perspective in each study and according to the Incremental cost-utility ratio (ICUR) rather than Incremental cost-effectiveness ratios (ICERs) if both types of ratios were calculated. The physical activity program was considered cost-effective if its cost per QALY was below the lower bound £20,000 per QALY threshold referred to by the National Institute for Health and Care Excellence (NICE). We also mention when an exercise program was cost-effective only at the upper bound NICE threshold of £30,000 per QALY. To determine the cost-effectiveness of physical activity programs, all ICURs were converted in United States dollars (US$) using the Purchasing Power Parity (PPP) exchange rate of the price year used in the study and compared with the NICE thresholds converted in US$

https://doi.org/10.1017/S0266462318000533 Published online by Cambridge University Press

using the PPP exchange rate of the same year. Exchange rates were drawn from the Organisation for Economic Co-operation and Development (OECD) (3).

Using the predefined extraction form we also collected data on the characteristics of the RCTs, on the structural assumptions and validity checks of model-based studies and on the economic characteristics needed to assess the methodological quality of the study. Data that could not be retrieved from the economic evaluation study were gathered from the study protocol or the clinical evaluation study.

Based on the extracted data, we assessed the risk of bias of the RCTs (for RCT-based studies and for modelling studies based on the results of a single RCT) using the Cochrane Risk of Bias Tool for Randomized Controlled Trials (4). We assessed three criteria of the tool: random sequence generation, allocation concealment, and incomplete outcome data. The other criteria were not investigated given the nature of this review. Specifically, selective outcome reporting was not investigated as it can be legitimate to focus on outcomes such as QALYs in the cost-effectiveness analysis. Blinding of participants and personnel was not possible for physical activity programs and the blinding of outcome assessment criterion was not evaluated as all studies used self-reported health benefit measures. For the three criteria assessed, the RCT was classified as "low risk," "high risk," or "unclear risk." We scored each criterion as one for "low risk" and zero otherwise. The scores of the three criteria were summed to create an overall score of RCT quality ranging from 0 to 3.

We used the adjusted Consensus Health Economic Criteria (CHEC) list (5) to assess the methodological quality of the economic evaluations. The CHEC list was specifically designed for conducting systematic reviews based on economic evaluation studies and its use is recommended by the Cochrane Collaboration (6). The list was recently adapted to fit both model and trial-based economic evaluations (7;8). The adjusted CHEC list contains twenty yes/no questions on the methodology of the economic evaluations. To obtain an index of methodological quality, we scored each item as 1 if the adjusted CHEC list criterion was satisfactorily fulfilled (yes) and 0 (no) otherwise. One question (question 5) is specific to model-based studies while another question on discounting (question 15) only applies to studies with time horizons longer than 1 year. Thus, the maximum achievable score ranges between 18 and 20. We classified the methodological quality of the economic evaluations based on the percentage of the maximum achievable score they obtained: low (<50 percent), moderate (50–70 percent), and high (>70 percent).

RESULTS

Study Selection

Keywords search gave 431 hits. After removing duplicates, 426 studies were screened for inclusion. Exclusion was carried out in two steps. First, 333 studies were excluded after title and abstract reading. Second, full-text reading led to exclude fifty-five of the ninety-three remaining articles based on the criteria described in Figure 1. Among the thirty-eight articles retained for analysis, two studies reported on the same results from a unique physical activity program, leaving a total of thirty-seven different economic evaluations (9–45).

Overview of Included Studies

Table 1 presents the major program and economic characteristics of the thirty-seven studies. Further details can be found in Supplementary Files 2 and 3. The main disease categories were musculoskeletal and rheumatologic disorders (eleven studies, 29.7 percent), cardiovascular diseases (ten studies, 27 percent), neurological disorders (six studies, 18.2 percent), mental illnesses (three studies, 8.1 percent) and cancers (three studies, 8.1 percent). A majority of articles came from two countries: the United Kingdom with twelve articles (32.4 percent of total) and the Netherlands with eleven articles (29.7 percent of total). Among the thirty-seven included studies, thirty were RCT-based (81.1 percent), while seven (18.9 percent) were modeling studies based on one or several RCTs. Only two studies exclusively used disease-specific measures of health benefits (13;16). The time horizon of RCT-based studies ranged from 12 weeks (36) to 2.5 years (23). Model-based studies had longer time horizons as they most often simulated the long-term effects of exercise programs. A total of seventeen studies (45.9 percent) were conducted from a societal cost perspective, fifteen studies (40.6 percent) from a health care perspective and five studies (13.5 percent) from a health and social care perspective. The physical activity interventions differed in terms of type, volume, and duration of exercise performed, even within disease categories.

RESULTS BY DISEASE CATEGORY

Musculoskeletal and Rheumatologic Disorders

Among the eleven studies on musculoskeletal and rheumatologic disorders, eight programs included strength exercises, four included stretching exercises, two included aerobics, two included balance exercises, one included water-based exercises and one included yoga. Some exercise programs were very long and intensive with 104 hours of exercise over 8 months (11) while others were much shorter and less intensive with 4 hours of exercise over 2 weeks (14). Only three of eleven studies did not take usual care as the comparator. In these studies, the physical activity program was compared with self-care advice (9), leaflet provision (10), or therapeutic education on weight and exercise (16). All studies were based on RCTs while only two studies, both focusing on knee pain, did not include QALYs (13;16). Seven of eleven studies used a societal cost perspective. The time horizon for the evaluation ranged from 8 months (11) to 30 months (13).

https://doi.org/10.1017/S0266462318000533 Published online by Cambridge University Press

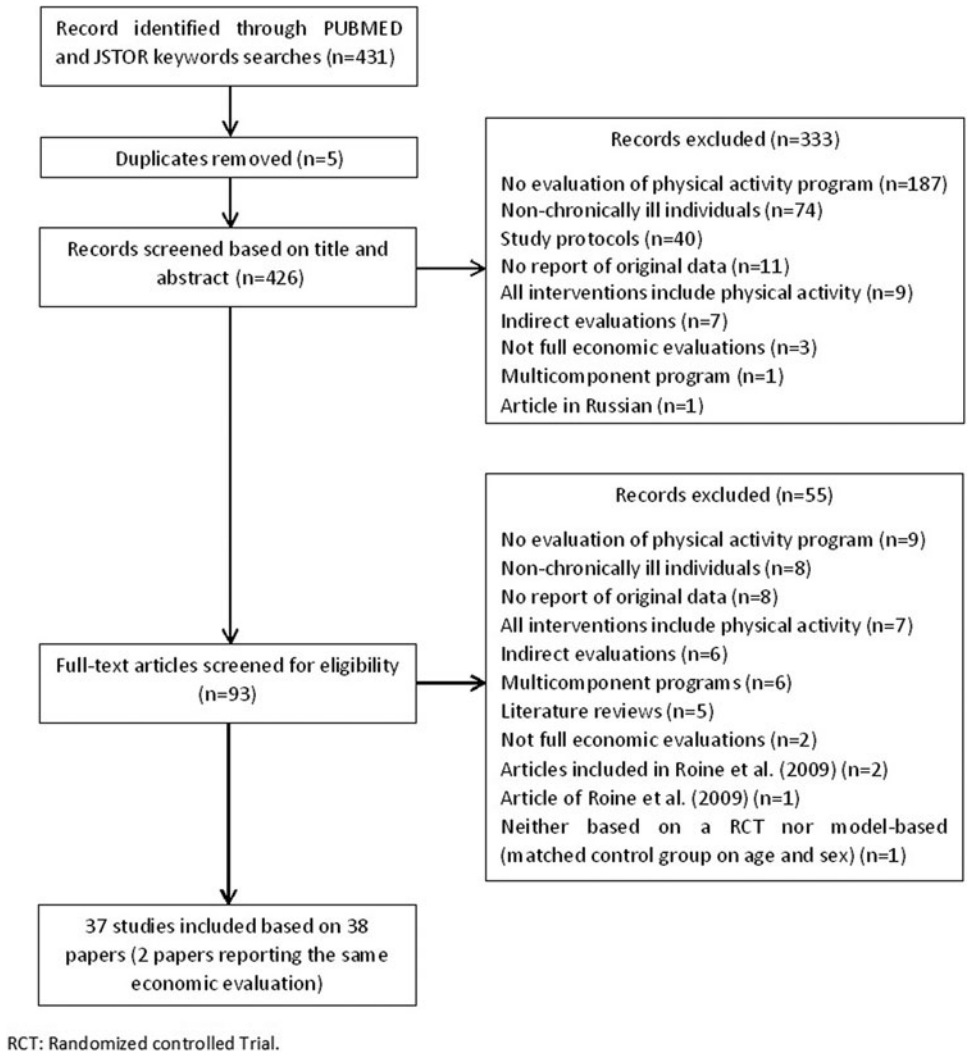

Fig. 1. Selection of included articles.

Results of the quality assessment of the RCTs are available in Supplementary File 4. Among the eleven studies on musculoskeletal and rheumatologic disorders, three (10;14;18) obtained the maximum RCT quality score of 3, six obtained a score of 2, and two obtained lower scores of 1 (9) or 0 (12). Table 2 gives the results of the methodological quality assessment of the economic evaluations. Among the eleven studies on musculoskeletal and rheumatologic disorders, five studies had a high score, five had a moderate score, and one had a low score (16).

Table 3 gives the results of the cost-effectiveness analyses. Results are reported for the broadest cost perspective adopted by the authors. Full results for all cost perspectives and health benefit measures are detailed in Supplementary File 5. For the eleven exercise programs focusing on musculoskeletal and rheumatologic disorders, five were dominant (cheaper and more effective) (9;13;14;17;18) and five represented intermediate cases as they were both more expensive and more effective than their comparators (10;11;12;15;19). Among the five intermediate cases, the exercise program was cost-effective at the £20,000 per QALY threshold in three studies (11;15;19) and not cost-effective in two studies (10;12). The cost-effectiveness of the exercise program was uncertain in one study that only used disease-specific measures of health benefits. This study did not report cost-effectiveness acceptability curves nor discuss the threshold values that should be used to judge the cost-effectiveness based on the disease-specific health benefit measures used (16).

Cardiovascular Diseases

Among the ten studies on cardiovascular diseases, four programs included walking, four included strength or stretch exercises, four included aerobics while one study reported no information on the exercise program (25). The duration and volume of the exercise program was highly variable. For example, in the case of intermittent claudication, the exercise program lasted from 12 weeks (22) to 12 months (27) or included 24 (26;28) to 78 hours (24;27) of supervised exercise. Usual care or optimal medical care was the comparator in six studies, while one study compared physical

https://doi.org/10.1017/S0266462318000533 Published online by Cambridge University Press

Table 1. Program Description and Economic Characteristics

Author	Country[a]	Disease	Design	Exercise type	Exercise volume and duration	Comparator	Benefit measure	Time horizon	Perspective[b]
–	–	–	RCT: 81.1%	–	–	Usual care: 64.9%	QALY: 94.6%	≤ 12 months: 64.9%	Societal: 45.9%
				Musculoskeletal and rheumatologic disorders (11 studies, 29.7%)					
Aboagye et al. (9)	SE	Low back pain	RCT	Yoga	Supervised: 6 weeks, 12 sessions	Self-care advice	QALY	12 months	Societal
Barton et al. (10)	GB	Knee pain	RCT	Strength and resistance quadriceps exercises	Unsupervised: 24 months, daily basis	Leaflet	QALY	24 months	Health care
Gusi et al. (11)	ES	Fibromyalgia	RCT	Water-based exercises	Supervised: 8 months, 104 sessions, 104 h	Usual care	QALY	8 months	Societal
Henchoz et al. (12)	CH	Low back pain	RCT	Strength, endurance and stretching exercises	Supervised: 12 weeks, 24 sessions, 36 h	Usual care + exercise advice	QALY	12 months	Societal
Hurley et al. (13)	GB	Chronic knee pain	RCT	Strength, endurance, and balance exercises	Supervised: 6 weeks, 12 sessions, 8 h	Usual care	WOMAC function score	30 months	Health and social care
Manning et al. (14)	GB	Rheumatoid arthritis	RCT	Functional and strength exercises	Supervised : 2 weeks, 6 sessions, 4 h Unsupervised : 12 weeks, daily basis	Usual care	QALY	36 weeks	Societal
Pinto et al. (15)	NZ	Hip or knee osteoarthritis	RCT	Aerobic, strength and stretching exercises	Supervised: 16 weeks, 9 sessions, 7.5 h Unsupervised: 16 weeks, 3 times per week	Usual care	QALY; WOMAC function score	12 months	Societal
Sevick et al. (16)	US	Knee osteoarthritis	RCT	Aerobic, walking, resistance training	Supervised: 4 months, 36 sessions, 36 h Unsupervised or supervised: Subsequent 14 months, 3 times per week	Education (on weight loss and exercise)	6MWD; WOMAC function score	18 months	Health care
Tan et al. (17)	NL	Patellofemoral pain syndrome	RCT	Quadriceps, balance and flexibility exercises	Supervised: 6 weeks, 9 sessions Unsupervised: subsequent 6 weeks	Usual care	QALY	12 months	Societal
Tan et al. (18)	NL	Hip arthritis	RCT	Strengthening and stretching exercises	Supervised: 9 months, 15 sessions, 7.5 h	Usual care	QALY	12 months	Societal
Williams et al. (19)	GB	Hand rheumatoid arthritis	RCT	Strengthening and stretching exercises	Supervised: 12 weeks, 5 sessions, 2.5–3.25 h	Usual care	QALY	12 months	Health and social care
				Cardiovascular diseases (10 studies, 27%)					
Hautala et al. (20)	FI	Acute coronary syndrome	RCT	Aerobic and strength exercises	Supervised: 6 months, 26 sessions, 13–19.5 h Unsupervised: 12 months	Usual care	QALY	12 months	Health and social care

https://doi.org/10.1017/S0266462318000533 Published online by Cambridge University Press

Table 1. Continued

Author	Country[a]	Disease	Design	Exercise type	Exercise volume and duration	Comparator	Benefit measure	Time horizon	Perspective[b]
–	–	–	RCT: 81.1%	–	–	Usual care: 64.9%	QALY: 94.6%	≤ 12 months: 64.9%	Societal: 45.9%
Kühr et al. (21)	BR	Heart failure	Model (13 RCTs)	Aerobic, cycling	Supervised: First year: 72 sessions, 72 h Subsequent years: 50 sessions, 50 h	Usual care	QALY; Life year saved	10 years	Health care
Mazari et al. (22)	GB	Intermittent claudication	RCT	Aerobic and stretching exercises	Supervised: 12 weeks, 36 sessions, 36 h	Angioplasty	QALY	12 months	Societal
Reed et al. (23)	US	Heart failure	RCT	Walking, treadmill and cycling	Supervised: 12 weeks, 36 sessions, 60–108 h	Usual care	QALY	2.5 years	Societal
Reynolds et al. (24)	US-CA	Intermittent claudication	Model (1 RCT)	Walking	Supervised: 6 months, 78 sessions, 78 h Unsupervised: subsequent 12 months	Optimal medical care	QALY	5 years	Societal
Rincón et al. (25)	CO	Heart failure	Model	Not specified	Supervised: 12 weeks, 36 sessions	Usual care	QALY; Life year gained	5 years	Health care
Spronk et al. (26)	NL	Intermittent claudication	RCT	Walking	Supervised: 24 weeks, 48 sessions, 24 h Unsupervised: 24 weeks, 72 sessions, 36 h	Endovascular revascularization	QALY	12 months	Societal
van Asselt et al. (27)	NL	Intermittent claudication	RCT	Walking	Supervised: 12 months, 156 sessions, 78 h	Walking advice + leaflet	QALY; Walking distance	12 months	Societal
van den Houten et al. (28)	NL	Intermittent claudication	Model (2 RCTs)	Strength and endurance exercises	Supervised: 12 months, 48 sessions, 24 h	Endovascular revascularization	QALY	5 years	Health care
Witham et al. (29)	GB	Heart failure	RCT	Aerobic, strength and resistance exercises	Supervised: 8 weeks, 16 sessions Unsupervised: subsequent 16 weeks	Usual care	QALY; 6MWD	24 weeks	Health care
Neurological disorders (6 studies, 18.2%)									
Farag et al. (30)	AU	Parkinson disease	RCT	Strengthening and balance exercises	Supervised: 6 months, 6 sessions, 4–6 h Unsupervised: 6 months, 78 sessions, 52–78 h	Usual care	QALY; Falls prevented	26 weeks	Health care
Fletcher et al. (31)	GB	Parkinson disease	RCT	Balance and strengthening exercises	Supervised: 10 weeks, 10 sessions, 10 h Unsupervised: subsequent 10 weeks, 20 sessions, 20 h	Usual care	QALY	20 weeks	Health and social care

Table 1. Continued

Author	Country[a]	Disease	Design	Exercise type	Exercise volume and duration	Comparator	Benefit measure	Time horizon	Perspective[b]
–	–	–	RCT: 81.1%	–	–	Usual care: 64.9%	QALY: 94.6%	≤ 12 months: 64.9%	Societal: 45.9%
McCrone et al. (32)	GB	Chronic fatigue	RCT	Aerobic, walking	Supervised: 12 months, up to 15 sessions	Specialist medical care	QALY; Chalder fatigue score	12 months	Societal
Sabes-Figuera et al. (33)	GB	Chronic fatigue	RCT	Walking	Supervised: 16 weeks, 8 sessions, 4.25 h	Usual care + leaflet	QALY; Chalder fatigue score	6 months	Health and social care
Slaman et al. (34)	NL	Cerebral palsy	RCT	Aerobic endurance, aerobic interval and strength training	Supervised: 12 weeks, 12 sessions Unsupervised: subsequent 12 weeks, 12 sessions	Usual care	QALY	12 months	Societal
Tosh et al. (35)	GB	Multiple sclerosis	RCT	Aerobic	Supervised: 12 weeks, 18 sessions, 18 h Unsupervised: 12 weeks, 18 sessions, 18 h	Usual care	QALY	9 months	Societal
Mental illnesses (3 studies, 8.1%)									
d'Amico et al. (36)	GB	Dementia	RCT	Walking	Supervised: 6 weeks, 42 sessions, 14–21 h Unsupervised: subsequent 6 weeks, 42 sessions, 14–21 h	Usual care	QALY; NPI score	12 weeks	Societal
Edwards et al. (37)	GB	Anxiety; depression	RCT	Not specified	Supervised: 16 weeks	Usual care + leaflet	QALY	12 months	Health care
Gusi et al. (38)	ES	Obesity; Depression	RCT	Walking, strengthening and stretching exercises	Supervised: 6 months, 72 sessions, 60 h	Usual care + exercise advice	QALY	6 months	Health care
Cancers (3 studies, 8.1%)									
Gordon et al. (39)	AU	Breast cancer	RCT	Aerobic and strength-based exercises	Supervised: 8 months, 16 sessions	Usual care	QALY; Quality of life	12 months	Societal
Mewes et al. (40)	NL	Breast cancer	Model (1 RCT)	Swimming, running and cycling	Supervised: one intake session, 1 h Unsupervised: 12 weeks, 30–36 h	Usual care	QALY	5 years	Health care
Retèl et al. (41)	NL	Head and neck cancers	Model (2 RCTs)	Swallowing stretch and strength exercises	Supervised: one intake session Unsupervised: 10 weeks, 3 times per day	Usual care	QALY	12 months	Health care
Other diseases (4 studies, 10.8%)									
Panman et al. (42)	NL	Pelvic organ prolapse	RCT	Pelvic floor muscle training	Supervised: 16 weeks, 7 sessions (median) Unsupervised: Up to 2 to 3 times each day	Watchful waiting	QALY; PFDI-20	2 years	Health care

Table 1. Continued

Author	Country[a]	Disease	Design	Exercise type	Exercise volume and duration	Comparator	Benefit measure	Time horizon	Perspective[b]
–	–	–	RCT: 81.1%	–	–	Usual care: 64.9%	QALY: 94.6%	≤ 12 months: 64.9%	Societal: 45.9%
Panman et al. (43)	NL	Pelvic organ prolapse	RCT	Pelvic floor muscle training	Supervised: 16 weeks, 7 sessions (median) Unsupervised: Up to 2 to 3 times each day	Pessary treatment	QALY; PFDI-20	2 years	Health care
Zwerink et al. (44)	NL	COPD	RCT	Cycling, walking, stairs climbing and weights lifting	Supervised: 11 months, 78 sessions Unsupervised: 11 months, 54 sessions	Self-management	QALY; Shuttle walk test; Daily step	2 years	Health care
Coyle et al. (45)	CA	Type 2 diabetes	Model (1 RCT)	Aerobic and resistance exercises	Supervised: 6 months, 11 sessions	Waiting list	QALY; Life Year Saved	40 years	Health care

[a]3-letter International Organization for Standardization country codes. AU: Australia; BR: Brazil; CA: Canada; CH: Switzerland; CO: Colombia; ES: Spain; FI: Finland; GB: The United Kingdom; NL: The Netherlands; NZ: New Zealand; SE: Sweden; US: The United States.
[b]Broader cost perspective reported by the authors.
6MWD, 6-Minute Walking Distance; COPD, chronic obstructive pulmonary disease; h, hours; NPI, Neuropsychiatric Inventory; PFDI-20, Pelvic-Floor-Distress-Inventory-20; QALY, quality-adjusted life-year; RCT, randomized controlled trial; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index.

activity with walking advice and leaflet provision (27). The comparator was surgery in three studies on intermittent claudication (22;26;28). Grouping the findings from studies using surgical and nonsurgical comparators might not be appropriate and the results are thereafter differentiated for these two types of studies.

Among the seven studies using nonsurgical comparators, four were based on a RCT while three were model-based. All seven studies included QALYs while only three (23;24;29) used a societal cost perspective. The time horizon for the evaluation ranged from 24 weeks (29) to 2.5 years (23). Among the three studies using a surgical comparator, one study was model-based (28), all studies used QALYs and two studies adopted a societal cost perspective (22;26). One study evaluated the exercise program over 10 years (21), while the time horizon was 12 months in the other two studies (22;26).

The RCT-quality could be rated for seven studies, including five studies using nonsurgical comparators. Among these five studies, two obtained a score of 2 (23;29) and three a score of 1 (20;24;27). Among studies using surgical comparators, one had a RCT quality score of 0 (22) and the other a score of 2 (26). For the seven studies using nonsurgical comparators, one (24) had a high score for the methodological quality of the economic evaluation, while five had a moderate score, and one had a low score (21). The three studies using surgical comparators obtained a moderate score for the quality of the economic evaluation.

Among the seven studies using nonsurgical comparators, the exercise program was dominant in two studies (22;29), cost-effective in two studies (24;25) and not cost-effective in three studies (21;23;27). If an alternative cost-effectiveness threshold at £30,000 per QALY were to be adopted, exercise programs would then be cost-effective in two studies (21;27). Among the three studies using surgical comparators, the exercise program was dominant in one study (20) and cost-effective in the other two studies (26;28).

Neurological Disorders

Among the six studies on neurological disorders, two studies (30;31) focused on Parkinson disease and evaluated programs of strengthening and balance exercises over the same time period (5 to 6 months). However, the first study (30) included 84 hours of exercise versus only 30 hours in the second study (31). Two studies focused on walking and aerobics or walking programs for patients with chronic fatigue. The first program (32) lasted longer (12 months) and included twice as many sessions compared with the second program which lasted 16 weeks (33). The last two studies evaluated aerobic programs among patients with cerebral palsy and multiple sclerosis with either twelve (35) or eighteen (35) supervised sessions over 12 weeks. Usual or specialist medical care, sometimes in combination with leaflet provision, was the comparator in all studies on neurological disorders. All studies in this category were based on RCTs and included QALYs. Only half of studies used a

https://doi.org/10.1017/S0266462318000533 Published online by Cambridge University Press

Table 2. Assessment of the Methodological Quality of the Economic Evaluations

	Adjusted Consensus Health Economic Criteria list items																				
	Study population described	Alternatives described	Well-defined research question	Appropriate design	Structural assumptions and validation methods described	Appropriate time horizon	Appropriate perspective	Important costs included	Appropriate measurement of costs	Appropriate valuation of costs	Important outcomes included	Appropriate measurement of outcomes	Appropriate valuation of outcomes	Appropriate incremental analysis	Appropriate discounting	Complete sensitivity analyses	Conclusions follow data reported	Discussion on the generalizability of the results	Declaration of interests	Discussion on ethical and distributional issues	% of maximum achievable score
Musculoskeletal and rheumatologic disorders																					
Aboagye et al. (9)	+	−	+	+	†	−	+	+	+	+	+	+	+	+	*	−	+	+	+	−	78
Barton et al. (10)	+	+	−	+	†	+	−	−	−	+	+	+	+	+	+	−	+	−	+	+	68
Gusi et al. (11)	+	−	+	−	†	−	+	+	−	+	+	+	+	+	*	−	+	+	−	−	61
Henchoz et al. (12)	+	+	+	+	†	−	+	+	+	+	+	+	−	+	*	−	+	−	−	+	72
Hurley et al. (13)	+	+	−	+	†	+	−	+	+	+	−	−	+	+	+	−	+	+	−	+	68
Manning et al.(14)	+	+	+	+	†	+	+	+	+	+	+	+	+	+	*	−	+	−	−	−	78
Pinto et al. (15)	+	+	−	+	†	−	+	+	+	−	+	−	−	+	*	−	+	+	+	−	61
Sevick et al. (16)	+	+	−	−	†	+	+	−	−	−	−	−	−	+	+	−	−	+	−	−	37
Tan et al. (17)	−	−	+	+	†	+	+	+	+	−	+	+	−	+	*	−	+	+	−	−	61
Tan et al. (18)	+	+	+	+	†	−	+	+	+	−	+	+	+	+	*	+	−	+	+	−	78
Williams et al. (19)	+	+	+	+	†	+	−	+	+	−	+	+	+	+	*	−	+	+	+	+	83
Cardiovascular diseases																					
Hautala et al. (20)	+	−	−	−	†	+	+	+	−	−	+	+	−	+	*	−	−	+	−	+	50
Kühr et al. (21)	−	−	−	+	+	−	−	+	+	+	+	−	−	+	−	−	+	+	−	−	45
Mazari et al. (22)	+	−	−	−	†	+	+	−	−	−	+	+	−	+	*	−	+	+	+	−	50
Reed et al. (23)	+	+	−	+	†	−	+	+	−	−	+	+	−	−	+	−	+	+	−	+	58
Reynolds et al. (24)	+	+	−	+	+	+	+	+	+	−	+	+	+	+	+	−	+	+	−	−	75
Rincón et al. (25)	−	−	−	+	+	−	−	+	+	+	+	+	−	+	+	−	−	+	+	+	60
Spronk et al. (26)	+	+	+	−	†	−	+	+	−	−	+	+	+	+	*	−	+	+	+	−	67
van Asselt et al. (27)	+	+	−	−	†	−	+	+	+	+	+	−	−	+	*	−	+	−	+	−	56
van den Houten et al. (28)	−	−	+	+	+	+	−	−	+	+	+	+	+	+	+	−	+	+	−	+	70
Witham et al. (29)	+	−	−	−	†	−	−	−	+	+	+	+	−	+	*	−	+	+	+	−	50
Neurological disorders																					
Farag et al. (30)	+	+	−	−	†	+	+	+	+	−	+	+	−	+	*	−	+	+	−	−	61
Fletcher et al. (31)	+	+	+	−	†	+	+	+	+	+	+	+	+	+	*	−	+	+	+	+	89
McCrone et al. (32)	+	−	+	−	†	+	+	+	+	+	+	−	+	+	*	−	+	+	+	−	72
Sabes-Figuera et al. (33)	+	+	+	−	†	+	−	−	+	−	−	−	+	+	*	−	+	+	−	−	50
Slaman et al. (34)	+	−	+	+	†	−	+	+	+	+	+	+	−	+	*	+	+	−	−	−	67
Tosh et al. (35)	+	−	+	+	†	+	+	+	−	+	+	+	−	+	*	+	+	+	−	−	72
Mental illnesses																					
d'Amico et al. (36)	+	+	−	−	†	−	+	−	−	+	+	−	+	+	+	−	+	−	+	−	53
Edwards et al. (37)	+	−	−	+	†	+	−	+	+	+	+	+	−	+	*	+	+	+	+	−	72
Gusi et al. (38)	+	+	+	−	†	−	+	−	−	+	+	+	+	+	*	−	+	+	−	−	61
Cancers																					
Gordon et al. (39)	+	−	+	−	†	−	+	+	−	−	+	+	+	+	*	−	−	+	−	−	50
Mewes et al. (40)	+	−	+	+	+	+	−	+	+	−	+	+	−	+	+	−	+	+	−	−	65
Retèl et al. (41)	−	+	+	+	+	−	−	+	+	−	+	−	−	+	*	−	−	+	+	−	53
Other diseases																					
Panman et al. (42)	+	+	−	+	†	−	−	+	−	+	+	+	+	+	−	−	+	+	+	−	63

Table 2. Continued

	Adjusted Consensus Health Economic Criteria list items																					
	Study population described	Alternatives described	Well-defined research question	Appropriate design	Structural assumptions and validation methods described	Appropriate time horizon	Appropriate perspective	Important costs included	Appropriate measurement of costs	Appropriate valuation of costs	Important outcomes included	Appropriate measurement of outcomes	Appropriate valuation of outcomes	Appropriate incremental analysis	Appropriate discounting	Complete sensitivity analyses	Conclusions follow data reported	Discussion on the generalizability of the results	Declaration of interests	Discussion on ethical and distributional issues	% of maximum achievable score
Panman et al. (43)	+	+	+	+	†	–	–	+	–	+	+	+	+	+	–	–	+	+	+	–	68
Zwerink et al. (44)	+	–	–	+	†	+	+	+	–	+	+	–	–	+	–	–	+	+	+	–	58
Coyle et al. (45)	+	–	–	+	+	+	–	–	+	–	+	–	+	+	+	–	–	+	–	+	55

Note. + The adjusted Consensus Health Economic Criteria list criterion is satisfactorily fulfilled. - The adjusted Consensus Health Economic Criteria list criterion is not satisfactorily fulfilled. † Not applicable: study based on a randomised controlled trial. * Not applicable: time horizon shorter or equal to 12 months.

societal cost perspective (32;34;35) while the time horizon for the evaluation ranged from 20 weeks (31) to 12 months (32;34).

On the six studies focusing on neurological disorders, two obtained a RCT quality score of 3 (30;35), two had a score of 2 (31;32), and two had low scores of 1 (33) or 0 (34). Regarding the quality of the economic evaluation, half of studies reached a high score (31;32;35), and the other half obtained a moderate score (30;33;34).

The exercise program was dominant in three studies on neurological disorders (31;32;34). The program was not cost-effective at the £20,000 per QALY threshold in two of the remaining studies (30;35) and the cost-effectiveness of the exercise program was uncertain in the third one (33). For one study, the exercise program was cost-effective only at the £30,000 per QALY threshold (35). In one study (33), only a disease-specific measure of health benefit, the Chalder fatigue score, was used and thus the classification of the cost-effectiveness result was difficult. However, the authors reported that the probability of cost-effectiveness was equal to 55 percent to 63 percent if the decision maker was willing to pay £1,000 to £2,500 per clinically significant improvement in fatigue (four-point variation in the Chalder fatigue scale).

Mental Illnesses

Among the three studies on mental illnesses, two exercise programs focused on walking (36;38) while the type of exercise performed was not mentioned in one study (37). The length of the supervised exercise program ranged from 6 weeks (36) to 6 months (38), while the supervised exercise volume was comprised between 14 (36) and 60 hours (38) over the whole program. Usual care was the comparator in all studies, in combination with leaflet provision or exercise advice. All studies were based on RCTs and used QALYs, but only one study adopted a societal cost perspective (36). The time horizon for the evaluation ranged from 12 weeks (36) to 12 months (37).

On the three studies dealing with mental illnesses, one obtained the minimal RCT quality score of 0 (36), and two had a score of 1 (37;38). For the methodological quality of the economic evaluation, one study (37) obtained a good score, while two studies had a moderate score (36;38).

The exercise program was found cost-effective in two studies among depressed patients (37;38) but not cost-effective in one study on dementia (36).

Cancers

Two studies focused on breast cancer patients (39;40). The first study (39) included sixteen sessions of aerobic and strength exercises spread over 8 months while women undertook mostly unsupervised running, swimming, or cycling in the second study (40). The third study examined a swallowing exercise program for head and neck cancer patients (41). All exercise programs were compared with usual care. Two

https://doi.org/10.1017/S0266462318000533 Published online by Cambridge University Press

Table 3. Results of Cost-Effectiveness Analyses

Author	Disease	ICER/ICUR (95% confidence interval)	Probability of cost-effectiveness of the exercise program	Result category[a]
		Musculoskeletal and rheumatologic disorders		
Aboagye et al. (9)	Low back pain	Dominant: ICUR not reported	Not calculated	Dominant
Barton et al. (10)	Knee pain	£49,146/QALY	Not calculated	Intermediate case (not cost-effective)
Gusi et al. (11)	Fibromyalgia	€7878/QALY (3559 ; 93,818)	p = 95% if decision maker willing to pay €28,300 to gain a QALY	Intermediate case (cost-effective)
Henchoz et al. (12)	Low back pain	€79,270/QALY	Not calculated	Intermediate case (not cost-effective)
Hurley et al. (13)	Chronic knee pain	Dominant: ICER not reported	p = 81% to 100% if decision maker willing to pay £0 to £9750 for a 1% increase in the proportion of patients improving on the WOMAC function score	Dominant
Manning et al. (14)	Rheumatoid arthritis	Dominant: -£185,068/QALY	Not reported for societal perspective	Dominant
Pinto et al. (15)	Hip or knee osteoarthritis	NZ$23,365/QALY (−102,356 ; 163,958) NZ$87 per improvement in WOMAC function score (−233 ; 6037)	Not reported for societal perspective	Intermediate case (cost-effective)
Sevick et al. (16)	Knee osteoarthritis	US$200 per percentage point of improvement in WOMAC function score US$10 per percentage point of improvement in 6MWD	Not calculated	Uncertain
Tan et al. (17)	Patellofemoral pain syndrome	Dominant: -€14,738 per QALY (−210,206 ; 178,822)	p = 73% if decision maker willing to pay €20,000 to gain a QALY	Dominant
Tan et al. (18)	Hip arthritis	Dominant: -€97,195/QALY	p = 68% if decision maker willing to pay €20,000 to gain a QALY	Dominant
Williams et al. (19)	Hand rheumatoid arthritis	£17,941/QALY	p = 52% to 59% if decision maker willing to pay £20,000 to £30,000 to gain a QALY	Intermediate case (cost-effective)
		Cardiovascular diseases		
Hautala et al. (20)	Acute coronary syndrome	Dominant: -€24,511/QALY	p = 100% with any value of willingness to pay	Dominant
Kühr et al. (21)	Heart failure	Int$26,461/QALY Int$21,169/Life Year Saved	p = 55% if decision maker willing to pay Int$27,495 to gain a QALY	Intermediate case (not cost-effective[b])
Mazari et al. (22)	Intermittent claudication	Dominant: -€13,450.35/QALY	Not calculated	Dominant
Reed et al. (23)	Heart failure	Not reported	p = 47.9% to 59.2% if decision maker willing to pay US $50,000 to US$100,000 to gain a QALY	Intermediate case (not cost-effective)
Reynolds et al. (24)	Intermittent claudication	U$24,070/QALY	p > 60% if decision maker willing to pay US$30,000 to US$80,000 to gain a QALY	Intermediate case (cost-effective)

https://doi.org/10.1017/S0266462318000533 Published online by Cambridge University Press

Table 3. Continued

Author	Disease	ICER/ICUR (95% confidence interval)	Probability of cost-effectiveness of the exercise program	Result category[a]
Rincón et al. (25)	Heart failure	US$998/QALY US$156/Life Year Gained	$p = 76\%$ if decision maker willing to pay US$21,000 to gain a QALY	Intermediate case (cost-effective)
Spronk et al. (26)	Intermittent claudication	Endovascular revascularization versus hospital-based exercise: €231,800/QALY	$p = 95\%$ if decision maker willing to pay €50,000 for a QALY gained	Intermediate case (cost-effective)
van Asselt et al. (27)	Intermittent claudication	€28,693/QALY €4.08 per extra metre on the treadmill test	$p = 64\%$ if decision maker willing to pay €40,000 to gain a QALY $p = 85\%$ if decision maker willing to pay €6 to gain a meter	Intermediate case (not cost-effective[b])
van den Houten et al. (28)	Intermittent claudication	Endovascular revascularization versus supervised exercise therapy: €91,600/QALY	$p = 73\%$ if decision maker willing to pay €40,000 to gain a QALY	Intermediate case (cost-effective)
Witham et al. (29)	Heart failure	Dominant: ICUR not reported	Not calculated	Dominant
		Neurological disorders		
Farag et al. (30)	Parkinson disease	AU$338,800/QALY gained AU$574 per fall prevented	$p < 20\%$ if decision maker willing to pay AU$100,000 to gain a QALY $p = 80\%$ if decision maker willing to pay AU$2000 to prevent a fall	Intermediate case (not cost-effective)
Fletcher et al. (31)	Parkinson disease	Dominant: -£1167/QALY	$p = 78\%$ if decision maker willing to pay £20,000 to gain a QALY	Dominant
McCrone et al. (32)	Chronic fatigue	Dominant: -£13,761/QALY ICERs not reported but negative for fatigue	$p = 34.8\%$ that graded exercise therapy is the second most cost-effective option after cognitive behaviour therapy if decision maker is willing to pay £30,000 to gain a QALY	Dominant
Sabes-Figuera et al. (33)	Chronic fatigue	£987 per clinical significant improvement (4 points) in Chalder fatigue scale	$p = 55\%$ to 63% if decision maker willing to pay £1000 to £2500 to have a clinical significant improvement in Chalder fatigue scale	Uncertain
Slaman et al. (34)	Cerebral palsy	Dominant: -€23,664/QALY (−167,992 ; 129,007)	$p = 86\%$ if decision maker willing to pay €20,000 to gain a QALY	Dominant
Tosh et al. (35)	Multiple sclerosis	£24,897/QALY	Not reported for societal perspective	Intermediate case (not cost-effective[b])
		Mental illnesses		
d'Amico et al. (36)	Dementia	£286,440/QALY £1263 for a meaningful improvement (3 points) in NPI	$p = 68\%$ if decision maker willing to pay £3000 to gain three-point in the NPI score	Intermediate case (not cost-effective)
Edwards et al. (37)	Anxiety; depression	£10,276/QALY (−40,659 ; 61,228)	$p = 80\%$ to 89% if decision maker willing to pay £20,000 to £30,000 to gain a QALY	Intermediate case (cost-effective)

Table 3. Continued

Author	Disease	ICER/ICUR (95% confidence interval)	Probability of cost-effectiveness of the exercise program	Result category[a]
Gusi et al. (38)	Obesity; Depression	€311/QALY (143 ; 394)	p = 99.9% if decision maker willing to pay €600 to gain a QALY	Intermediate case (cost-effective)
		Cancers		
Gordon et al. (39)	Breast cancer	Service provider[c]: AU$105,231/QALY AU$2644 per patient with meaningful improvement (8 points or more) in the quality of life scale	p = 44.4% if decision maker willing to pay AU$50,000 to gain a QALY	Intermediate case (not cost-effective)
Mewes et al. (40)	Breast cancer	€28,078/QALY	Physical exercise has the highest probability of being cost-effective if decision maker is willing to pay €26,000 to gain a QALY	Intermediate case (not cost-effective[b])
Retèl et al. (41)	Head and neck cancers	€3197/QALY	p = 83% if decision maker willing to pay €20,000 to gain a QALY	Intermediate case (cost-effective)
		Other diseases		
Panman et al. (42)	Pelvic organ prolapse	€31,983/QALY (−76,652 ; 88,078) €43 per additional point on the PFDI-20 (18 ; 237)	Not calculated	Intermediate case (not cost-effective[b])
Panman et al. (43)	Pelvic organ prolapse	Pessary treatment versus pelvic floor muscle training: -US$27,439/QALY (−91,974 ; 74,695) -US$77 per additional point on the PFDI-20 (−373 ; 351)	Not calculated	Dominated
Zwerink et al. (44)	COPD	€10,950/QALY €6257 per additional patient prevented deteriorating at least 47.5 meters on the walk test €1564 per additional patient improving the mean number of steps with at least 500 steps/day	Not calculated	Intermediate case (cost-effective)
Coyle et al. (45)	Type 2 diabetes	CA$37,782/QALY CA$28,494/Life-year saved	p = 55.5% that combined program is the most cost-effective if decision maker is willing to pay CA $50,000 to gain a QALY	Intermediate case (not cost-effective[b])

[a]Based on the £20,000 per QALY threshold used by the National Institute for Health and Care Excellence. Results are based on the ICUR rather than ICERs if both types of ratios were calculated and are reported for the broadest cost perspective adopted by the authors.
[b]Not cost-effective at the £20,000 per QALY threshold but cost-effective at the £30,000 per QALY threshold.
[c]The intervention is implemented by a community organisation and the physiologist is an employee of the organisation. The authors also report the cost-effectiveness results for a private model where exercise physiologists working privately integrate the intervention into their routine practice. In both cases, the exercise program is not cost-effective.
6MWD, 6-minute walking distance; AU, Australia; CA, Canada; COPD, chronic obstructive pulmonary disease; ICER, incremental cost-effectiveness ratio; ICUR, incremental cost-utility ratio; Int, international; NPI, Neuropsychiatric Inventory; NZ, New Zealand; p, probability; PFDI-20, Pelvic-Floor-Distress-Inventory-20; QALY, quality-adjusted life-year; US, United States of America; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index.

studies were model-based (40;41) and one was based on an RCT (39). QALYs were measured in all studies, while only one study (39) adopted a societal cost perspective. The time horizon for the evaluation ranged from 12 months (39;41) to 5 years (40).

The RCT quality was assessable for two studies that obtained scores of 2 (39) and 1 (40), respectively. The three studies obtained a moderate score for the methodological quality of the economic evaluation.

Among the three studies in the field of oncology, one swallowing exercise program for head and neck cancer patients was found to be cost-effective (41), while two exercise programs among women with breast cancer were not cost-effective at the £20,000 per QALY threshold (39;40). In one study on breast cancer (40), the exercise program was cost-effective when using the £30,000 per QALY threshold.

Other Diseases

Among the four remaining studies, two focused on pelvic floor muscle training among women with pelvic prolapse (42;43), one focused on a 2-year walking and cycling program among chronic obstructive pulmonary disease (COPD) patients (44) and one evaluated a 6-month aerobic and resistance exercise program for diabetic patients (45). The two articles on pelvic organ prolapse evaluated the same program with respect to different comparators, watchful waiting (42), or pessary treatment (43). Self-management was used as comparator in the study on COPD (44) while the comparator group was a waiting list control in the study with diabetic patients (45). One study was model-based (45), while the others were based on RCTs (42;43;44). All studies included QALYs and used a health care cost perspective. The time horizon of the evaluation was 40 years in the model-based study (45), while it was 2 years in the three RCT-based studies (42;43;44).

The COPD study (44) and the two studies on pelvic organ prolapse (42;43) achieved a RCT quality score of 1, while the study on type 2 diabetes (45) obtained a score of 2. All four studies attained a moderate score for the methodological quality of the economic evaluation.

The pelvic floor muscle training for women with pelvic organ prolapse was found to be dominated by pessary (43) treatment or not cost-effective at the £20,000 per QALY threshold against watchful waiting (42) (but cost-effective at the £30,000 per QALY threshold). The physical activity program was found cost-effective in the study on COPD (44), while a combined aerobic and resistance training among diabetic patients was found not cost-effective at the £20,000 per QALY level but cost-effective at the £30,000 per QALY level (45).

DISCUSSION

We identified thirty-seven studies evaluating the cost-effectiveness of exercise programs among chronically ill patients, mainly with musculoskeletal and rheumatologic disorders or cardiovascular diseases, published after 2008. Exercise programs were dominant or cost-effective in twenty-two studies (59.5 percent) when using a £20,000 per QALY threshold or in twenty-eight studies (75.7 percent) when using a £30,000 per QALY threshold. Exercise programs were not cost-effective in seven studies (18.9 percent) for either threshold, while cost-effectiveness of physical activity remained unclear in two studies (5.4 percent) given the use of disease-specific health benefit measures only.

Exercise programs were dominant or cost-effective at the £20,000 per QALY threshold in eight of eleven studies on musculoskeletal and rheumatologic disorders, in four of seven studies on cardiovascular diseases using a nonsurgical comparator, in three of six studies on neurological disorders, in one of three studies on cancers and in two of three studies on mental illnesses. If an alternative cost-effectiveness threshold at £30,000 per QALY were to be adopted, the exercise program would be cost-effective in two additional studies on cardiovascular diseases using a nonsurgical comparator, in one additional study on neurological disorders and in one additional study on cancers. Thus, available evidence shows that exercise programs are cost-effective for the most part for the treatment of musculoskeletal and rheumatologic disorders and for the treatment of cardiovascular diseases when considering the upper NICE cost-effectiveness threshold of £30,000 per QALY. This result is in line with the previous findings of Roine et al. (2) who found stronger evidence of cost-effectiveness for exercise programs in cardiac rehabilitation and in back pain patients.

Since the last review of Roine et al. (2) more studies were found in the fields of neurological disorders, mental illnesses and oncology. However, for these disease groups, economic evaluations of physical activity programs remain scarce and the few studies available show contradictory results. For other conditions, such as diabetes, obesity, and respiratory diseases, the exercise programs under evaluation are usually multicomponent and study designs most often do not allow isolation of the specific impact of exercise on costs and health outcomes.

The existing literature suffers from several limitations. First, included studies were of varying levels of methodological quality. For instance, five of eleven studies on musculoskeletal and rheumatologic disorders had a good score for the methodological quality of the economic evaluation, while this was the case for only one of seven studies on cardiovascular diseases using a nonsurgical comparator.

Progress has been made in the comparability of results between and within disease categories, because QALYs were used as a measure of health benefits in thirty-five of thirty-seven studies. Nevertheless, comparability of cost-effectiveness results is still limited. This arises mainly from the differences in methodologies across studies. The cost perspective adopted, the type of design used (RCT-based or model-based studies) or the

https://doi.org/10.1017/S0266462318000533 Published online by Cambridge University Press

time horizon chosen to evaluate the exercise program may all impact the cost-effectiveness results. The use of a larger cost perspective, that includes productivity losses or informal care, tends to increase the cost-effectiveness of exercise programs compared with a strict health care perspective. Conversely, including the opportunity cost of exercise or patients' out-of-pocket costs tends to reduce the cost-effectiveness of the exercise program compared with using a health care system perspective.

The choice of model-based analyses, which typically simulates the impact of the exercise program over a longer period of time, might also affect the results of the cost-effectiveness analysis. For example, among the seven studies focusing on cardiovascular diseases and using nonsurgical comparators, three studies (21;24;25) used modeling to assess the impact of exercise programs over several years and the exercise program was cost-effective in all three studies. In this review, cost-effectiveness results tend to be more often positive in modeling studies. Indeed, among the seven modeling studies, the exercise program was dominant or cost-effective in 71.4 percent of cases when using the £20,000 QALY thresholds, while among the thirty RCT-based studies, the exercise program was dominant or cost-effective in only 60 percent of cases.

The second factor impeding the comparability of cost-effectiveness results is the heterogeneity among interventions. Exercise programs under evaluation differed in terms of the type of exercise performed. For example, in the case of low back pain, the physical activity program included strength, endurance, and stretching exercises in one study (12), while it included yoga in another study (9).

The exercise programs also differed in terms of duration and volume of exercise performed. For instance, for intermittent claudication, the exercise programs lasted from 12 weeks (22) to 12 months (27) or included 24 (26;28) to 78 hours (24;27) of supervised exercise. No clear pattern of association between the volume and duration of the exercise program and its cost-effectiveness emerged from our analysis. Indeed, for low back pain, a yoga program of twelve sessions over 6 weeks (9) was found dominant, while a longer exercise program of twenty-four sessions spread over 12 weeks (12) was not cost-effective.

On the other hand, in the case of heart failure, two exercise programs both including thirty-six sessions over 12 weeks were found either cost-effective (25) or not cost-effective (23). The two studies focusing on intermittent claudication and using nonsurgical comparators included two walking programs of 78 hours. In the first study, the program was spread over 6 months and 78 sessions (24), while in the second study, the program was run in 156 sessions over 12 months (27). In this specific case, both programs were cost-effective at the £30,000 per QALY threshold even if the cost-effectiveness of the more intensive program was slightly superior.

In addition to the interventions' characteristics, patients' adherence to the program may also affect its health and economic impacts. We found that only seventeen studies (45.9 percent) reported patients' adherence to the exercise programs, with highly varying levels (Supplementary File 2). Finally, the differences in the clinical and demographic characteristics of patients may also impact the cost-effectiveness results. For example, among the four studies focusing on heart failure, two studies included patients with New York Heart Association class II or III heart failure (21;29), while one study also included class IV patients (23) and one gave no details on the severity of the diseases among included patients (24) (Supplementary File 2).

In conclusion, we identified thirty-seven studies evaluating the cost-effectiveness of exercise programs among chronically ill patients. Exercise programs for the treatment of musculoskeletal and rheumatologic disorders, and to a lesser extent for the treatment of cardiovascular diseases, appear cost-effective. More research is needed to investigate the cost-effectiveness of exercise programs in the treatment of cancers, mental disorders, diabetes, obesity, and respiratory diseases.

ACKNOWLEDGMENTS

This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found at https://doi.org/10.1017/S0266462318000533
Supplementary File 1:
https://doi.org/10.1017/S0266462318000533
Supplementary File 2:
https://doi.org/10.1017/S0266462318000533
Supplementary File 3:
https://doi.org/10.1017/S0266462318000533
Supplementary File 4:
https://doi.org/10.1017/S0266462318000533
Supplementary File 5:
https://doi.org/10.1017/S0266462318000533

CONFLICTS OF INTEREST

The authors have nothing to disclose.

REFERENCES

1. Pedersen BK, Saltin B. Exercise as medicine–evidence for prescribing exercise as therapy in 26 different chronic diseases. *Scand J Med Sci Sports*. 2015;25:1-72.
2. Roine E, Roine RP, Räsänen P, et al. Cost-effectiveness of interventions based on physical exercise in the treatment of various diseases: a systematic literature review. *Int J Technol Assess Health Care*. 2009;25:427-454.

3. OECD [Intenet]. Purchasing power parities (PPP) (indicator). [cited May 19, 2018]. https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm (accessed August 14, 2018).
4. Higgins JP, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343:d5928.
5. Evers S, Goossens M, De Vet H, Van Tulder M, Ament A. Criteria list for assessment of methodological quality of economic evaluations: Consensus on Health Economic Criteria. *Int J Technol Assess Health Care*. 2005;21:240-245.
6. Higgins JP, Green S, editors. *Cochrane handbook for systematic reviews of interventions*. Vol. 4. New York: John Wiley & Sons; 2011.
7. Odnoletkova I, Goderis G, Pil L, et al. Cost-effectiveness of therapeutic education to prevent the development and progression of type 2 diabetes. Systematic Review. *J Diabetes Metab*. 2014;5:438.
8. van Eeden M, van Heugten CM, van Mastrigt GA, Evers SM. Economic evaluation studies of self-management interventions in chronic diseases: A systematic review. *Int J Technol Assess Health Care*. 2016;32:16-28.
9. Aboagye E, Karlsson ML, Hagberg J, Jensen I. Cost-effectiveness of early interventions for non-specific low back pain: A randomized controlled study investigating medical yoga, exercise therapy and self-care advice. *J Rehabil Med*. 2015;47:167-173.
10. Barton GR, Sach TH, Jenkinson C, et al. Lifestyle interventions for knee pain in overweight and obese adults aged ≥ 45: Economic evaluation of randomised controlled trial. *BMJ*. 2009;339:b2273.
11. Gusi N, Tomas-Carus P. Cost-utility of an 8-month aquatic training for women with fibromyalgia: A randomized controlled trial. *Arthritis Res Ther*. 2008;10:R24.
12. Henchoz Y, Pinget C, Wasserfallen JB, et al. Cost-utility analysis of a three-month exercise programme vs usual care following multidisciplinary rehabilitation for chronic low back pain. *J Rehabil Med*. 2010;42:846-852.
13. Hurley MV, Walsh NE, Mitchell HL, et al. Economic evaluation of a rehabilitation program integrating exercise, self-management, and active coping strategies for chronic knee pain. *Arthritis Rheum*. 2007;57:1220-1229.
14. Manning VL, Kaambwa B, Ratcliffe J, et al. Economic evaluation of a brief education, self-management and upper limb exercise training in people with rheumatoid arthritis (EXTRA) programme: A trial-based analysis. *Rheumatology*. 2014;54:302-309.
15. Pinto D, Robertson MC, Abbott JH, et al. Manual therapy, exercise therapy, or both, in addition to usual care, for osteoarthritis of the hip or knee. *2: Economic evaluation alongside a randomized controlled trial*. Osteoarthritis Cartilage. 2013;21:1504-1513.
16. Sevick MA, Miller GD, Loeser RF, Williamson JD, Messier SP. Cost-effectiveness of exercise and diet in overweight and obese adults with knee osteoarthritis. *Med Sci Sports Exerc*. 2009;41:1167-1174.
17. Tan SS, van Linschoten RL, van Middelkoop M, et al. Cost-utility of exercise therapy in adolescents and young adults suffering from the patellofemoral pain syndrome. *Scand J Med Sci Sports*. 2010;20:568-579.
18. Tan SS, Teirlinck CH, Dekker J, et al. Cost-utility of exercise therapy in patients with hip osteoarthritis in primary care. *Osteoarthritis Cartilage*. 2016;24:581-588.
19. Williams MA, Williamson EM, Heine PJ, et al. Strengthening And stretching for Rheumatoid Arthritis of the Hand (SARAH). A randomised controlled trial and economic evaluation. *Health Technol Assess*. 2015;19:1-222.
20. Hautala AJ, Kiviniemi AM, Mäkikallio T, et al. Economic evaluation of exercise-based cardiac rehabilitation in patients with a recent acute coronary syndrome. *Scand J Med Sci Sports*. 2017;27:1395-1403.
21. Kühr EM, Ribeiro RA, Rohde LEP, Polanczyk CA. Cost-effectiveness of supervised exercise therapy in heart failure patients. *Value Health*. 2011;14:S100-S107.
22. Mazari FAK, Khan JA, Carradice D, et al. Economic analysis of a randomized trial of percutaneous angioplasty, supervised exercise or combined treatment for intermittent claudication due to femoropopliteal arterial disease. *Br J Surg*. 2013;100:1172-1179.
23. Reed SD, Whellan DJ, Li Y, et al. Economic evaluation of the HF-ACTION randomized controlled trial: An exercise training study of patients with chronic heart failure. *Circ Cardiovasc Qual Outcomes*. 2010;3:374.
24. Reynolds MR, Apruzzese P, Galper BZ, et al. Cost-effectiveness of supervised exercise, stenting, and optimal medical care for claudication: Results from the Claudication: Exercise Versus Endoluminal Revascularization (CLEVER) trial. *J Am Heart Assoc*. 2014;3:e001233.
25. Rincón M, Rojas MX, Romero VAR, et al. Economic evaluation of exercise-based cardiac rehabilitation programs for chronic heart failure patients in Colombia. *J Cardiopulm Rehabil Prev*. 2016;36:12-19.
26. Spronk S, Bosch JL, den Hoed PT, et al. Cost-effectiveness of endovascular revascularization compared to supervised hospital-based exercise training in patients with intermittent claudication: A randomized controlled trial. *J Vasc Surg*. 2008;48:1472-1480.
27. van Asselt ADI, Nicolaï SPA, Joore MA, Prins MH, Teijink JAW. Exercise Therapy in Peripheral Arterial Disease Study Group. Cost-effectiveness of exercise therapy in patients with intermittent claudication: Supervised exercise therapy versus a "go home and walk" advice. *Eur J Vasc Endovasc Surg*. 2011;41:97-103.
28. Van Den Houten MML, Lauret GJ, Fakhry F, et al. Cost-effectiveness of supervised exercise therapy compared with endovascular revascularization for intermittent claudication. *Br J Surg*. 2016;103:1616-1625.
29. Witham MD, Fulton RL, Greig CA, et al. Efficacy and cost of an exercise program for functionally impaired older patients with heart failure: A randomized controlled trial. *Circ Heart Fail*. 2012;5:209-216.
30. Farag I, Sherrington C, Hayes A, et al. Economic evaluation of a falls prevention exercise program among people With Parkinson's disease. *Mov Disord*. 2016;31:53-61.
31. Fletcher E, Goodwin VA, Richards SH, Campbell JL, Taylor RS. An exercise intervention to prevent falls in Parkinson's: An economic evaluation. *BMC Health Serv Res*. 2012;12:426.
32. McCrone P, Sharpe M, Chalder T, et al. Adaptive pacing, cognitive behaviour therapy, graded exercise, and specialist medical care for chronic fatigue syndrome: A cost-effectiveness analysis. *PLoS One*. 2012;7:e40808.
33. Sabes-Figuera R, McCrone P, Hurley M, et al. Cost-effectiveness of counselling, graded-exercise and usual care for chronic fatigue: Evidence from a randomised trial in primary care. *BMC Health Serv Res*. 2012;12:264.
34. Slaman J, van den Berg-Emons R, Tan SS, et al. Cost-utility of a lifestyle intervention in adolescents and young adults with spastic cerebral palsy. *J Rehabil Med*. 2015;47:338-345.
35. Tosh J, Dixon S, Carter A, et al. Cost effectiveness of a pragmatic exercise intervention (EXIMS) for people with multiple sclerosis: Economic evaluation of a randomised controlled trial. *Mult Scler J*. 2014;20:1123-1130.
36. d'Amico F, Rehill A, Knapp M, et al. Cost-effectiveness of exercise as a therapy for behavioural and psychological symptoms of dementia within the EVIDEM-E randomised controlled trial. *Int J Geriatr Psychiatry*. 2016;31:656-665.
37. Edwards RT, Linck P, Hounsome N, et al. Cost-effectiveness of a national exercise referral programme for primary care patients in Wales: Results of a randomised controlled trial. *BMC Public Health*. 2013;13:1021.
38. Gusi N, Reyes MC, Gonzalez-Guerrero JL, Herrera E, Garcia JM. Cost-utility of a walking programme for moderately depressed, obese, or overweight elderly women in primary care: A randomised controlled trial. *BMC Public Health*. 2008;8:231.

https://doi.org/10.1017/S0266462318000533 Published online by Cambridge University Press

39. Gordon LG, DiSipio T, Battistutta D, et al. Cost-effectiveness of a pragmatic exercise intervention for women with breast cancer: Results from a randomized controlled trial. *Psychooncology*. 2017;26:649-655.
40. Mewes JC, Steuten LMG, Duijts SFA, et al. Cost-effectiveness of cognitive behavioral therapy and physical exercise for alleviating treatment-induced menopausal symptoms in breast cancer patients. *J Cancer Surviv*. 2015;9:126-135.
41. Retèl VP, van der Molen L, Hilgers FJ, et al. A cost-effectiveness analysis of a preventive exercise program for patients with advanced head and neck cancer treated with concomitant chemo-radiotherapy. *BMC Cancer*. 2011;11:475.
42. Panman CM, Wiegersma M, Kollen BJ, et al. Two-year effects and cost-effectiveness of pelvic floor muscle training in mild pelvic organ prolapse: A randomised controlled trial in primary care. *BJOG*. 2017;124:511-520.
43. Panman CM, Wiegersma M, Kollen BJ, et al. Effectiveness and cost-effectiveness of pessary treatment compared with pelvic floor muscle training in older women with pelvic organ prolapse: 2-year follow-up of a randomized controlled trial in primary care. *Menopause*. 2016;23:1307-1318.
44. Zwerink M, Effing T, Kerstjens HAM, et al. Cost-effectiveness of a community-based exercise programme in COPD self-management. *COPD*. 2016;13:214-223.
45. Coyle D, Coyle K, Kenny GP, et al. Cost-effectiveness of exercise programs in type 2 diabetes. *Int J Technol Assess Health Care*. 2012;28:228-234.

https://doi.org/10.1017/S0266462318000533 Published online by Cambridge University Press