Economic evaluations of worksite health promotion programs

van Dongen, J.M.

2014

document version
Publisher's PDF, also known as Version of record

10

General discussion

GENERAL DISCUSSION

The prevalence of modifiable health risks among the population is high, which imposes a large economic burden on society as a whole and on employers in particular. The workplace presents a useful setting to offer behavior change interventions that aim to prevent and/or reduce such risk factors. Amongst others, because a large number of people can be reached, including many who would otherwise be unlikely to engage in preventive health behaviors. Furthermore, Dutch employers themselves may financially benefit from implementing such interventions through reductions in productivity-related costs (1-5).

In practice, numerous occupational health interventions exist, of which only a limited number can be provided with the resources available (6). Therefore, high quality evidence in the form of methodologically sound economic evaluations is needed to demonstrate their value. Nonetheless, this evidence is scarce, which is partly due to the fact that only a few of the studies that consider the effectiveness of worksite health promotion programs take the extra step of considering their resource implications, and the methodological quality of those that do is generally poor. Moreover, the uptake of those that have been performed in daily practice is likely to be limited. Therefore, the aim of this thesis was to contribute to the development of a sound evidence base on the resource implications of worksite health promotion programs as well as to improve the uptake of the results of such studies in daily practice. This was done by summarizing the current literature on the cost-effectiveness and financial return of worksite physical activity and/or nutrition programs (*Chapter 2* and *3*), generating new evidence by performing economic evaluations of various newly developed worksite health promotion programs (*Chapter 4, 5, 6,* and *7*), and developing and providing recommendations for good practice when conducting and disseminating economic evaluations in occupational health (*Chapter 8* and *9*).

This general discussion is divided into five parts. First, the main findings of the systematic reviews, the applied studies, as well as a qualitative study into the information needs of occupational health decision-makers will be summarized and discussed. Second, various considerations will be discussed that warrant further exploration in relation to the methodology of economic evaluations in occupational

health. Third, the present findings will be compared to the literature. Fourth, recommendations for practice and research will be presented. The discussion will end with concluding remarks.

Main findings

What is known about the cost-effectiveness and financial return of worksite physical activity and/or nutrition programs?

Chapter 2 and *chapter 3* describe two systematic reviews that summarize and critically appraise the current evidence of the cost-effectiveness and financial return of worksite physical activity and/or nutrition programs, respectively. From the review results described in *chapter 2*, no firm conclusions can be drawn about the cost-effectiveness of such interventions. This was due to the fact that the included studies used a broad range of outcome measures and analytic perspectives, which hampered pooling of their results. Also, most interventions were more costly and more effective in improving various health outcomes (e.g. body weight and cholesterol level reduction), whereas set levels as to how much decision-makers are willing to pay for these improvements are currently lacking. The review in *chapter 3* found that average financial return estimates of worksite physical activity and/or nutrition programs in terms of absenteeism benefits, medical benefits, or both, were positive in non-randomized studies, but negative in randomized controlled trials (RCTs). These results indicate that financial return estimates derived from non-randomized studies should be interpreted with great caution. Economic evaluations alongside RCTs with a low risk of bias, on the other hand, indicate that worksite physical activity and/or nutrition programs may not pay for themselves in terms of absenteeism and/or medical benefits during the first years after implementation. However, as such programs are thought to be associated with additional types of benefits (e.g. presenteeism benefits), which have not been measured in most of the studies included in the review, conclusions about their overall profitability cannot be made.

Methodological quality of economic evaluations of worksite physical activity and/or nutrition programs

In both of the aforementioned systematic reviews (C*hapter 2* and *3*), the methodological quality of the included studies was assessed using a consensus-based methodological quality checklist. Both assessments indicated that the methodological quality of the included studies was generally poor. Examples of quality criteria that were least fulfilled included those related to the description of the study population and alternatives under study, the identification, measurement, and valuation of resource use, as well as the performance of sensitivity analyses and discounting. Also, few studies reported on the uncertainty of their cost-effectiveness and/or financial return estimates. The latter is a critical oversight as failing to estimate values under uncertainty may lead to biased conclusions and could thus result in inappropriate decision-making.

Do the evaluated worksite health promotion programs provide good value?

In *chapter 4* through *chapter 7*, four economic evaluations of various newly developed worksite health promotion programs were presented. Three economic evaluations were conducted alongside RCTs (*Chapter 4, 5,* and *6*), whereas the fourth used a 2X2 factorial design (*Chapter 7*). All interventions were compared to usual practice, both their cost-effectiveness and financial return were evaluated, analyses were performed from both the societal and employer's perspective, and the follow-up duration of all studies was 12 months. The main findings of the studies were:

- Vital@Work study: The worksite vitality intervention for older hospital workers evaluated in *chapter 4* was neither cost-effective from the societal perspective in improving general vitality, work-related vitality, and need for recovery, nor did it result in financial savings for the employer.
- The Mindful VIP study: The mindfulness-based worksite intervention for knowledge workers evaluated in *chapter 5* was neither cost-effective from the societal perspective in improving work engagement and general vitality, nor from that of the employer in improving work engagement, job satisfaction, and work ability. Also, the intervention was not saving costs to the employer.

- The VIP in Construction study: The worksite physical activity and nutrition program for construction workers evaluated in *chapter 6* was not cost-effective from the employer's perspective in improving work-related vitality and job satisfaction. The intervention's cost-effectiveness in improving weight-related outcomes (societal perspective) and musculoskeletal disorders (employer's perspective) depends on the respective decision-makers' willingness-to-pay for these effects. Also, even though financial return estimates were positive, the intervention was not considered cost saving to the employer due to a high level of uncertainty.
- The Be Active & Relax VIP study: Whether the combined social and physical environmental intervention evaluated in *chapter 7* can be regarded as cost-effective in improving need for recovery from both the societal and employer's perspective depends on the respective decision-makers' willingness-to-pay for these effects. The separate interventions were not cost-effective in improving this outcome. Moreover, none of the interventions was cost-effective in improving general vitality (societal perspective) and job satisfaction (employer's perspective), nor did they result in financial savings for the employer.

Information needs of occupational health decision-makers

Chapter 9 presents the results of a qualitative study into the occupational health decision-making process and information needs of occupational health decision-makers in the Ontario healthcare sector. The study indicated that the decision-making process can be generally subdivided into three stages: 1) initiation stage, during which the need for an intervention is established; 2) pre-implementation stage, during which an intervention and its business case are developed in order to receive senior management approval; and 3) implementation and evaluation stage, during which an intervention is implemented and evaluated. Organizations were found to invest in occupational health and safety interventions for legal, financial, and moral reasons. Financial information, especially the employer's costs and benefits, was found to be particularly important at the front end of implementation decisions, and can be a key deciding factor of whether to go forward with a new intervention. Results also

indicated that occupational health decisions are currently not being made in an evidence-based manner and that there is a need to advance the decision-makers' economic evaluation skill set. Further research is needed to explore whether these results are generalizable to the Dutch occupational health context. Nonetheless, it seems reasonable to assume that Dutch occupational health decision-makers are also particularly interested in the interventions' costs and benefits to the employer and that implementation decisions are not being made in an evidence-based manner as well.

Methodological considerations

Many of the methodological strengths and limitations of the applied studies have been discussed in *chapter 4* through *chapter 7*. In addition, recommendations for good practice when conducting economic evaluations in the field of occupational health research have been described in *chapter 8*. However, a selection of methodological considerations in relation to the study population, analytic perspective, study power, missing data, the identification, measurement, and valuation of resource use, time horizon, as well as the generalizability of our results warrant further exploration.

Study design

Three studies were conducted alongside an RCT (*Chapter 4, 5,* and *6*), while the Be Active & Relax VIP study used a 2X2 factorial design (*Chapter 7*). All studies used a pragmatic design, meaning that the interventions were evaluated under circumstances that resembled routine practice conditions as much as possible (7).

The pragmatic design of the studies made it possible to evaluate the interventions' (resource) implications under "real world" circumstances. This facilitates the generalizability of our results (i.e. external validity), whereas the randomization of participants improved the studies' internal validity (i.e. the ability to draw true conclusions about causes and effects) (7). The importance of randomization, on the other hand, was underscored by the review presented in *chapter 3*, in which average financial return estimates were found to differ between studies with and without randomization. Nonetheless, many economic evaluations of worksite health promotion programs are currently performed alongside non-randomized studies

(See *chapter 2* and *3*), even though these are particularly prone to selection bias. Selection bias arises when allocation methods other than randomization are used, meaning that the intervention and control group are unlikely to be comparable (8). For example, due to the lack of randomization it is unclear whether program participants were healthier and/or more motivated to change their health behavior(s) to begin with than non-participants. The possible existence of such a priori differences makes it hard to attribute study results to the intervention and to rule out the possibility that they were caused by (baseline) differences between study arms (i.e. confounding caused by selection bias) (3;8). Some people question the applicability of RCT results to daily practice, because the same design aspects that contribute to their high internal validity (e.g. well-defined inclusion and exclusion criteria) may simultaneously hamper the generalizability of their results in an extended population and/or setting (7). However, although other research designs may add to the existing knowledge on worksite health promotion programs, RCTs should be viewed as the "gold standard" for evaluating their (resource) implications untainted by bias (8).

In three studies (*Chapter 4, 5,* and *6*), randomization was performed at the participant-level, whereas group allocation was performed at the department-level in the Be Active & Relax VIP study (*Chapter 7*). The latter was done because the intervention under study operated on the group-level rather than on the individual-level as well as to avoid contamination between study groups (9). Methods for economic evaluations alongside RCTs are relatively well established (10;11), and these methods were used to evaluate the data of such studies. A fundamental issue in clustered studies, however, is that costs and consequences within a cluster may be more similar to each other than costs and consequences from a different cluster. As a consequence, methods that ignore clustering in economic evaluations generally underestimate the statistical uncertainty and are likely to have inaccurate point estimates (9;12). Based on recent research findings on the optimal strategy to account for clustering in economic evaluations (12;13) we therefore used multilevel analyses to assess the cost-effectiveness and financial return of the Be Active & Relax VIP interventions.

Study population

All interventions were aimed at primary prevention. That is, they were directed at all employees of the participating companies, who on average were generally healthy, instead of high-risk individuals (3). This approach attempts to shift the whole distribution of exposure in a particular population in a favorable direction by controlling the determinants of a disease and by lowering the mean level of risks. A drawback of primary prevention is that it offers only small benefits to individuals at the short-term, because their absolute risk for a disease is generally low (14). Consequently, it is relatively hard to motivate them to change their unhealthy behavior(s), and thus to achieve sustained health improvements (14;15). This may partially explain the lack of, or relatively small, effects of the interventions. To produce better effects, a high-risk strategy may be needed, in which prevention efforts are solely aimed at high-risk individuals (e.g. overweight and/or obese construction workers in the case of the VIP in Construction intervention (*Chapter 6)).* Such an approach likely offers a more cost-effective use of limited resources, because it is generally more efficient to concentrate limited time and money where the need, and therefore also the benefits, are likely to be greatest (14). High-risk strategies, however, do not deal with the root of a problem and it is questionable whether employers are willing to discriminate between their employees by providing different worksite health promotion programs to different groups of high-risk individuals. Therefore, a combination of various prevention strategies may ultimately be needed to achieve a meaningful degree of prevention in the workplace (16).

In three of the applied studies (*Chapter 4, 5,* and *7*), participants had relatively good baseline values of the primary outcomes, which further reduced the interventions' ability to accomplish sustained health improvements (i.e. ceiling effect) (Table 1). Selective enrolment of healthy individuals is not uncommon in health promotion programs/studies and is explained by the fact that people with healthy lifestyle behaviors are also the ones who are generally most motivated to pursue and maintain health (17;18). In the VIP in Construction study (*Chapter 5*), on the other hand, a relatively large number of obese construction workers was included (Table 1). This might have resulted from the fact that occupational physicians, who played an import role in the recruitment process, may have been particularly inclined to

motivate obese construction workers for study participation. The selective inclusion of either healthy or unhealthy employees likely bears on the generalizability of our findings, and should thus be taken into account when making inferences about the interventions' resource implications in a broader working population/setting.

Table 1: Participants' baseline values of the primary outcomes in relation to their respective norm scores

Study	Primary outcome	Baseline scores participants	Norm scores
VIP in Construction Study	Body weight		
	Normal Weight (BMI ≤ 25 kg m^{-2})	30.1%	34.3%[1]
	Overweight (BMI ≥ 25 kg m^{-2} and	47.4%	48.8%
	BMI < 30 kg m^{-2})	22.4%	16.9%
	Obesity (BMI ≥ 30 kg m^{-2})		
Vital@Work Study	Work-related vitality (Range: 0-6) [Mean (SD)]	4.88 (0.85)	4.01[2]
Mindful VIP study	Work engagement (Range: 0-6) [Mean (SD)]	4.10 (0.89)	3.82[3]
Be Active & Relax VIP study	Need for recovery (Range: 0-100, with lower scores indicating a lower need for recovery) [Mean (SD)]	33.2 (29.3)	38.1[4]

Abbreviations: BMI: Body Mass Index, SD: Standard Deviation, n: number
[1]Percentage of normal weight, overweight, and obese Dutch construction workers (100)
[2] Average work-related vitality score among the general Dutch working population (101)
[3] Average work engagement score among the general Dutch working population (101)
[4] Average need for recovery score among the general Dutch working population (102)

Analytic perspective

All studies applied a so-called two-perspective approach, in which analyses were performed from both the societal and employer's perspective (19;20). In the societal perspective, all costs and consequences are taken into account irrespective of who pays or benefits, whereas only those borne by, or accruing to, employers are included when the employer's perspective is applied. The main advantage of the employer's perspective is that its results are directly interpretable for those who we are trying to aid with our economic evaluations, namely occupational health decision-makers. A disadvantage of this perspective is that it does not provide an indication of whether the "local rationality" of the company is in line with societal optimality (i.e. maximizing

the welfare of society as a whole with the resources available) (19). To deal with this issue, analyses were also performed from the societal perspective, which provides insight into the interventions' net societal effects. Even though occupational health decision-makers themselves may view societal perspective results as externalities, having them ignorant of these results may lead to non-optimal resource allocation decisions at the aggregate level (19;21). Another advantage of the societal perspective is that its disaggregate information on costs and consequences gives a good sense of their distribution across stakeholders, which could provide a starting point for bargaining between them (11). Moreover, the application of the societal perspective improves the transferability of our results to countries with different (occupational) health and welfare systems. For example, U.S. employers who typically bear most of the healthcare costs of their employees, can extract this information from the disaggregate information on costs and consequences from the societal perspective.
Even though it was not the case in the applied studies, it is important to mention that economic evaluations from the societal and employer's perspective may provide conflicting results. For example, worksite health promotion programs whose benefits fall entirely on employees in the form of improved health, but do not have a positive impact on productivity and/or occupational health costs, may be justified in social terms, but may not be in any company's financial interest to implement (22). In case of such a scenario, other stakeholders (e.g. "*the Dutch Ministry of Health, Welfare, and Sports*") may wish to consider giving incentives to companies to ensure that a socially preferred program goes ahead (11;22). If the opposite is true (i.e. a new intervention is cost-effective from the employer's perspective, but not from the societal one), it is of utmost importance that occupational health decision-makers are made aware of the fact that an intervention which benefits their goals is unattractive to other stakeholders and society as a whole in order to discourage them from implementing such an intervention (19).

Study power

All sample sizes were based on detecting relevant differences in health and/or work-related outcomes, and not to detect relevant cost differences. However, as only a small proportion of participants incur high costs and costs are naturally bound

by zero, cost data have the tendency to follow a rightly skewed distribution. As a consequence, economic evaluations generally require much larger sample sizes than their corresponding effect analyses in order to achieve sufficient power to detect relevant cost differences (23;24). Thus, all of the applied studies are likely to be underpowered. This is a common problem in trial-based economic evaluations and is often due to various factors. First, many economic evaluations are "piggybacked" onto effectiveness trials, and power calculations are therefore typically performed before the economic evaluation requirements are considered (23). Second, a large number of parameters has to be specified in order to perform sample size calculations for economic endpoints, many of which are hard to forecast a priori (25). Third, and most importantly, if studies would be sufficiently powered to detect relevant cost differences, they typically become infeasible with extremely large sample sizes and very high research expenses (20;26).

If studies are likely to be underpowered, it is recommended to use estimation and/or decision uncertainty rather than hypothesis testing (11;23). Therefore, economists typically focus on estimating cost and effect differences and assessing the probability of an intervention being cost-effective (i.e. "*How confident are we that an intervention is cost-effective?*"), rather than testing a particular hypothesis (e.g. *"Are the cost-effectiveness outcomes statistically significant?"*) (26;27). In line with this recommendation, confidence intervals around cost and effect differences as well as financial return estimates were presented, and the interventions' probabilities of cost-effectiveness were explored at different ceiling ratios (i.e. the maximum amount of money decision-makers are willing to pay for an additional unit of effect) (11). Although confidence intervals around financial return estimates are relatively straightforward to interpret for researchers, many occupational health decision-makers lack the required economic and/or statistical background (See *chapter 9*). Therefore, the concept of the "probability of financial return" was introduced in *chapter 8*. This probability provides an indication of the likelihood that, given the data, a new intervention is cost saving. Occupation health decision-makers can subsequently use this information to consider whether the established probability of financial return is acceptable to them.

Missing data

All studies had some missing data, ranging from 12% to 41% on the effect measures and from 29% to 62% on the cost measures. Missing data are often inevitable in trials due to participant drop-out and/or non-response (11;28). In economic evaluations, the problem of missing data is even more pronounced, because cost data are generally the sum of numerous components and relatively short recall periods (and thus more measurement points) are needed to reliably estimate them (26;29). When data are missing, the key challenge is to maximize usage of available data while minimizing the bias introduced by the elements that are missing (28). Simply eliminating participants who have missing data (i.e. complete-case analysis) is inefficient, as it ignores available data of incomplete cases and produces a reduced-sized dataset of complete-cases, and thus a loss of power (11;28). On top of that, complete-case analyses may be biased when systematic differences exist between the missing and observed values (28;30). In all studies, multiple imputation was therefore used to fill in missing values. Multiple imputation is currently preferred over so-called naive methods (e.g. last-observation carried forward), because it accounts for the uncertainty associated with filling in the missing values (30;31).

Within a study, results derived using multiple imputation may differ from those of a complete-case analysis. To a greater or lesser extent, this was also the case in the applied studies. For example, excluding participants with incomplete data in the Vital@Work study (*Chapter 4*) resulted in positive financial return estimates, whereas the reverse was the case when multiple imputation was applied. On the basis of the aforementioned reasons, we always considered the results derived from the multiple imputed datasets to be more reliable than those of the complete-case analysis. Nonetheless, it is important to bear in mind that multiple imputation is based on the assumption that data are missing at random (MAR; missing data has a relation to observed factors and not to unobserved factors), an assumption that may not necessarily hold true but cannot be tested. Therefore, having a complete dataset is always preferred and every endeavor should be made in future studies to reduce the amount of missing data. Amongst others, this may be accomplished by minimizing the length of the questionnaires, using incentives, systematically contacting participants when their responses are missing, unclear, and/or incorrect,

and using modern data collection technologies, such as online questionnaires and mobile apps, to reduce the burden of the data collection process. When doing so, it is advisable to use a so-called mixed approach, since a strategy that may limit non-response among one type of participant, may not be effective for another (32).

Identification of resource use

As has been explained earlier, relevant resource use categories for inclusion in an economic evaluation depend on its analytic perspective. Other factors that might determine their relevance are, amongst others, the country or jurisdiction in which the study is undertaken, the nature of the alternatives being compared, and the relative order of magnitude of the resource use categories (11). From the societal perspective, resource use from the healthcare, alternative care, and occupational health sector, as well as that of employees, and changes in paid productivity were included. The latter were expressed in terms of changes in lost production due to sickness absence (i.e. absenteeism) as well as reduced performance while at work (i.e. presenteeism). The inclusion of presenteeism costs in economic evaluations is a much debated topic, particularly because a sound methodological framework for their assessment is currently lacking (21;29;33). After some consideration, we decided to include this resource use category in all studies, because presenteeism seems to account for the largest component of paid productivity changes and efforts to improve health were found to have a more immediate effect on presenteeism than on absenteeism (29;33-35). Resource use of family members and changes in unpaid productivity, on the other hand, were not included, as our economic evaluation results were expected to be unaffected by them (11). When the employer's perspective was applied, analyses were restricted to resource use from the occupational health sector and changes in paid productivity.

Measurement of resource use

Resource use data can be collected through a variety of means, including the use of insurance records, company databases, questionnaires, and prospective cost diaries. Of them, more objective measurement strategies are favoured over those that rely on participant self-report, because they minimize the possible influence of

recall bias (i.e. bias due to inaccurate and/or incomplete recollections of events) (37). Unfortunately, however, objective measurements were not always feasible and/or preferred in the applied studies.

Questionnaires were used in all studies to assess healthcare utilization, because collecting health insurance claim data of participants was practically infeasible and would not have provided all required information. To illustrate, Dutch employees can buy insurance packages from over 30 different insurance companies, most insurance companies offer various levels of supplementary insurance packages, and people can buy basic and supplementary insurance packages from different insurance companies (36). Even if all insurance companies would have been willing to provide data, which is highly unlikely, healthcare claim data would not have been comparable between employees, because the treatments covered (and claimed) differ between them. Furthermore, health insurance records often lack detailed resource use information and information on the healthcare services borne by employees themselves (e.g. co-payments, over-the-counter medication) are typically not included (37).

As it was not feasible to objectively measure on-the-job productivity, presenteeism data were collected using questionnaires as well. For this purpose, the *"World Health Organization – Health and Work Performance Questionnaire"* (WHP-HPQ) was used, which has shown good concordance with archival performance data (38;39). It should be noted, however, that numerous instruments exist for assessing presenteeism and that their estimates may vary widely. This suggests a lack of comparability among instruments, but it is still unclear which instrument provides the best estimates (29). We opted for the WHO-HPQ, because it is the most frequently used instrument in economic evaluations of similar interventions, and thus increases the comparability of our results (See *chapter 3*).

Questionnaires were also used for assessing sickness absence in the Vital@Work study (*Chapter 4*), whereas sickness absence data were extracted from company records in all other studies (*Chapter 5, 6,* and *7*). Research indicates that absenteeism estimates may differ extensively between both methods, and that the accuracy of self-reported sickness absence estimates strongly decreases with an increasing recall period (40;41). Given the available evidence on the optimal recall period for absenteeism, Zang et al. (2011) recommended the application of a 3-month recall

period in order to balance loss in precision and the increase in research costs and participant burden (29). As this recall period was used in the Vital@Work study as well, we do not expect that its results are severely distorted by recall bias. Nonetheless, as most employers systematically track employee sickness absence and sickness absence data are relatively easy to collect when conducting studies at the workplace, future economic evaluations of worksite health promotion programs are recommended to use company records whenever possible.

As indicated above, questionnaires may be prone to recall bias. However, as it seems highly unlikely that the extent of impairment in recall systematically differed between study groups, we do not expect that our reliance on them severely biased our results (42). When having to rely on participant self-report, the possible influence of recall bias may be reduced by reducing a questionnaire's recall period (e.g. 3 months for absenteeism and healthcare utilization data (29;43) and 2 weeks for presenteeism data (29)) or by using a more accurate data collection method, such as a prospective cost diary. Provided that participants truly complete such diaries in a prospective way, they are thought to result in a minimum recall error and therefore in a better and more complete reporting of resource use (37).

Valuation of resource use

One of the most important challenges when valuing resource use is the identification of the "best" price weight for translating units of resource use into monetary values. Such price weights should be based on the true opportunity cost of a good or service (i.e. the amount of money that is not available for its best alternative use), and should be reflective of the analytic perspective (11;22). Our ideas about the "best" price weights, as well as the most appropriate methods for valuing resource use, have slightly evolved over the course of this thesis and will be discussed below.

In the Vital@Work study (*Chapter 4*), intervention costs were estimated using a so-called bottom-up micro-costing approach for both the societal and employer's perspective. This means that we estimated the cost of the Vital@Work intervention by collecting detailed data regarding the resources consumed as well as their unit prices (11). In doing so, we aimed to best reflect the true cost of the intervention, meaning that profit margins and transfer payments were excluded as much as possible. In

the subsequent studies, however, we deviated from this approach in that bottom-up micro-costing was solely used for the societal perspective, whereas market prices were used when the employer's perspective was applied. This was done because we are of the opinion that market prices better reflect an intervention's true value at the company level (i.e. the amount of money that is not available to the company for its best alternative use).

Healthcare utilization, which was only included when the societal perspective was applied, was valued using standard price weights whenever possible. Such standard price weights are preferred over market prices, because market prices are an inaccurate reflection of its societal opportunity cost if a perfect market does not exist for a healthcare service. For example, if a healthcare provider has a local monopoly, its charges are often an overestimation of their true (societal) value because monopolists have the power to set their own price (11;36). Healthcare provider fees may not be an accurate reflection of the time and relative skill level that is needed for different procedures. Moreover, drug prices are often set in negotiations between the government and pharmaceutical companies, where the pharmaceutical company's commitment to research and the provision of employment might be taken into account, as well as the costs of discovery, production, and distribution of the drug in question (11).

In all studies, sports costs were based on the participants' self-reported expenses on sports membership fees and sports equipment. We considered this gross-costing approach to be appropriate, because the impact of changes in sports costs on the resulting cost-effectiveness and/or financial return estimates was expected to be low (11).

Occupational health costs were only considered in the VIP in Construction (*Chapter 6*) and Mindful VIP study (*Chapter 5*). In the VIP in Construction study, they solely included employer-provided gym membership subsidies, and were valued using data derived from financial department staff. In the Mindful VIP study, on the other hand, occupation health costs consisted of a broad range of occupational health services and in-company health promotion activities of the participating companies. In line with our methods for estimating intervention costs, micro-costed price weights were used for the societal perspective, whereas marked prices were used when the employer's perspective was applied.

In three studies, changes in productivity were valued using gross salaries of participants (*Chapter 4, 5,* and *7*). In the VIP in Construction study (*Chapter 6*), on the other hand, we had to use the average salary of construction workers, because the participating construction company did not provide permission to collect participant salary data. Even though the use of age- and gender-specific price weights may have improved the generalizability of our results (44), we decided to rely on participant salary data instead in order to account for the fact that the magnitude of production losses is likely to be greater among employees with higher incomes.

Another important issue when valuing changes in productivity is the method used for estimating absenteeism costs. In the first two economic evaluations (*Chapter 4* and *5*), the "Friction Cost Approach" (FCA) was used for both the societal and employer's perspective. The FCA is recommended by the "*Dutch Manual of Costing*" and assumes that production losses are confined to the time-span companies need to replace a sick worker by a formerly unemployed person to restore the company's initial production level (i.e. friction period, which is estimated to be 23 weeks in the Netherlands) (21;44;45). In the subsequent studies (*Chapter 6* and *7*), we deviated from this approach in that the FCA was only used for the societal perspective, whereas the "Human Capital Cost approach" (HCA) was used when the employer's perspective was applied. This was done because Dutch employers are obliged to pay at least 70% of the salary of sick employees for a period of two years, and most of them top up the wage payments from 70% to 100% during the first year of sickness absence (46). Thus, although the initial production level of a Dutch company may be restored after the friction period, employers still bear the additional cost of having to pay the salary of the sick worker.

It should be noted that it is unclear how accurate our productivity-related cost estimates are. First, we may have underestimated the actual productivity-related costs, because the applied methods do not account for the negative effect of absenteeism and presenteeism on co-workers in team-dependent production. The productive output of a full team may namely be jeopardized by one member's reduced labour input, and this may be especially relevant when substitutes are less productive and/or unavailable (i.e. "The multiplier effect"). Until now, some attempts have been made in the U.S. to construct "job-dependent multipliers" that account

for the (average) effect on co-worker absenteeism and presenteeism in specific job types (21;47). However, future studies are needed to establish the validity of these multipliers and to investigate their transferability across countries and/or jurisdictions (21). Conversely, we may have overestimated the actual productivity-related costs, because productivity losses may partly be compensated during normal working hours (41;48). For example, work that is normally performed by the sick employee in question may be completed by colleagues or made up by the sick employee itself after return to work (20). Currently, it is unknown what the best method is for correcting for such possible compensations of productivity losses. As such, compensation adjustments are typically uncommon in economic evaluations. Even though we may have already included some form of correction for such compensations by factoring in the 0.8 elasticity factor when using the FCA, whether this elasticity factor indeed represents compensation during normal working hours is currently unknown. Therefore, as various studies indicate that over half of the lost work is compensated during normal working hours, further research in this area is warranted (41;48;49).

Time horizon

All studies applied a follow-up of one year. As many of the (health) benefits of preventive interventions, such as ours, are thought to occur in the future, this follow-up is probably insufficient to capture all costs and consequences flowing from the interventions under study (11). Decision analytic modeling may be used to bridge the gap between what has been observed in the applied studies and what would be expected to happen over a longer time horizon (11). The validity of such modeling studies, however, strongly relies on the quality of the information used for constructing the model (11;22). Amongst others, there is a risk of overstating the benefits, especially if there is the possibility of decreased intervention effectiveness over time (22). Evidence indicates that the latter is often the case in health promotion studies (50), and this phenomenon was also observed in some of the applied studies. Therefore, when trying to extrapolate the present findings, various scenarios for the sustainability of the effects should be used. One should bear in mind, however, that it is highly unlikely that the longer-term cost-effectiveness and/or financial return of

the evaluated interventions would be much more favorable than those observed in the applied studies, because most of them did not result in statistically significant (health) improvements at one-year follow-up. Furthermore, it is questionable whether employers would wish to implement interventions that only generate financial savings after an extensive number of years. Employees typically switch employers a couple of times during their working life, and many of the benefits are therefore likely to accrue to future employers and/or the public (i.e. "The free rider problem") (51;52).

Generalizability of results

Some factors influencing the generalizability of our findings have been mentioned earlier, including the pragmatic design of the applied studies as well as the selective enrollment of healthy and unhealthy individuals. Furthermore, most studies were performed within a single company and the worksite health promotion programs themselves were specifically tailored to the needs of stakeholders involved. As a consequence, it is unknown to what extent the results may be generalized to other companies, work settings or the general working population. Nonetheless, we at least assume that they are generalizable to other companies with similar employee populations, with similar health issues. Also, the companies' participation in the current health promotion trials may be reflective of their degree of problem recognition, and thus their current workplace culture, available policies for improving employee health, the health status, sickness absence, and work performance of their employees, as well as their motivation to improve the current situation. As such, the participating companies may represent an optimal setting and any of our effect, cost-effectiveness, and/or financial return estimates could thus be an overestimation (20). The generalizability of our findings to other countries may be limited by differences in (occupational) healthcare and social security systems (53). In the Netherlands, for example, most healthcare costs are borne by health insurance companies and the government, whereas in countries with employer-provided health insurance (e.g. the United States (U.S.)) they typically accrue to the employer. As such, our employer's perspective findings are mainly of interest to countries with similar healthcare systems. Another factor that should be noted is that healthcare expenditure levels

may differ extensively between countries. For example, per capita spending on healthcare in the U.S. is double that of most European countries, leaving more room for improvements in healthcare costs (54). The generalizability of our productivity-related cost estimates, on the other hand, may be hampered by the fact that income rates, friction periods, and sickness absence behaviors may differ between countries as well. For example, it is reasonable to expect that Dutch employees are more inclined to report sick than, for example, U.S. employees, because Dutch employees generally get paid during sickness absence, while many U.S. employees are not (22;46). Other factors that may contribute to different resource use patterns include differences in the organization of (occupational) healthcare as well as the incidence of the health risk factors in question (20;55).

The easiest way to transfer economic evaluation results from one country to another would be to recalculate the monetary value of resource use for the target country where the results are to be applied and then recalculate the cost-effectiveness and/or financial return estimates of interest (53). This approach, however, is probably too simple, as additional adjustments are likely to be needed to account for differences in healthcare utilization patterns and sickness absence behaviors (56). Therefore, Manca and Willan (2006) proposed an algorithm based on the availability of data for choosing the appropriate analytic strategy for adapting economic evaluation results from one country to another. If the country of interest has participated in a multinational clinical trial in which data on costs and consequences have been collected, the preferred strategy would be to analyze the individual patient data of the trial. Such studies, however, have neither been performed in the present thesis nor in the field of worksite health promotion research. Therefore, a modeling approach would be required according to Manca and Willan (2006), in which as much cost and consequence data as possible are used from the jurisdiction of interest (56;57).

Comparison with the literature

During the last 15 years, Pelletier published a series of reviews of the clinical and cost-effectiveness of comprehensive worksite health promotion and disease management programs (58-64). The most recent review in this series reported fairly consistent positive effects on employee health and costs, as well as improvements

in the number and quality of studies (59). However, effect sizes were generally small and only seven of the 27 most recently performed studies used an RCT design. Based on these results, he concluded that there was "*guarded cautious optimism*" about the clinical and/or cost-effectiveness of comprehensive health promotion and disease management programs, a conclusion that is not necessarily supported by the findings of the review presented in *chapter 2*.

Up until now, various reviews have been conducted on the financial return of worksite health promotion programs in general. For example, a 1999 review of early worksite health promotion studies, mostly conducted in the 1980s and early 1990s, found financial return estimates in terms of healthcare benefits, productivity-related benefits, or both, to range from $1.4 to $3.1 per Dollar invested in the program (65). In 2001, Aldana performed a comprehensive review of the financial return of worksite health promotion programs. Seven of the included studies reported both costs and healthcare benefits, with an average financial return of $3.5 per Dollar spent. Only three of the included studies reported financial return estimates in terms of absenteeism benefits, which ranged from $2.5 to $10.1 (66). In a more recent review, Baicker et al. (2010) found that healthcare and absenteeism costs fell by $3.3 and $2.7 per Dollar invested in the program, respectively (67). Moreover, based on a review of 62 worksite health promotion studies conducted during the last three decades, Chapman (2012) reported that participants to worksite health promotion programs had 25.1% lower absenteeism costs and 24.5% lower healthcare costs than non-participants. Twenty-five of the included studies reported financial return estimates in terms of various types of benefits, with an average of $5.6 per Dollar spent (68). All of these reviews included a broad range of worksite health promotion programs (e.g. smoking cessation, stress reduction, physical activity, and/or nutrition programs) and most of them were evaluated using non-randomized studies, of which many even lacked a comparison group. Moreover, even though all review authors reported that the quality of the included studies was less than optimal, none of them assessed their methodological quality using a consensus-based checklist, nor did they explore the possible difference in results between non-randomized studies and RCTs. The latter, however, was explored in a recent review of U.S. worksite health promotion studies published after 2000 (69). The authors found that only one of the

seven studies showing cost savings utilized an RCT. In line with the review presented in *chapter 3*, they concluded that strong evidence of cost savings is currently lacking due to the general use of weak evaluation designs, and thus the possible distortion of results by selection bias.

Implications for practice

In order to prevent spending already scarce resources on ineffective and/or inefficient strategies, worksite health promotion program implementation and continuation decisions should be made in an evidence-based manner. That is, methodologically sound scientific evidence on their (financial) implications should be consulted before program implementation and sound ex-post program evaluations ought to be performed to inform continuation decisions.

Even though multiple reviews showed favorable, albeit small, effects of worksite health promotion programs on various health-related outcomes (70-75), the present thesis indicated that (strong) evidence of their cost-effectiveness and financial return is currently lacking. The latter is in contrast to the findings of most of the aforementioned reviews (65-68), which generally concluded that wider adoption of worksite health promotion programs could prove beneficial for company budgets. These reviews, however, mainly included non-randomized studies with a high risk of bias, while the review presented in *chapter 3* found financial return estimates to systematically differ between studies with and without randomization (i.e. positive in non-randomized studies and negative in RCTs). Moreover, none of the interventions evaluated in the present thesis were found to generate cost savings to the employer. Therefore, widespread implementation of existing worksite health promotion programs in an effort to generate cost savings is not recommended. It should be noted, however, that some of the evaluated interventions (*Chapter 6* and *7*) may be considered cost-effective if decision-makers are willing to invest a certain amount of money to improve employee health (52). Whether the latter is the case, however, is currently unknown.

Our recommendation is in contrast to the current widespread advertisement and implementation of worksite health promotion programs. Many program vendors advertise them by implying that they are an evidence-based strategy for reducing

healthcare and/or productivity-related costs. Advertisement statements such as *"With careful planning, efficient and effective wellness programs offered to employees and their families can shrink both the waistline and the bottom line"* are common (76), but not supported by methodologically sound evidence (i.e. high quality (cluster-)RCT-based economic evaluations). Nonetheless, a recent industry survey indicated that about 50 percent of Dutch employers invest in preventive strategies at the workplace, including stress management and lifestyle interventions (77). Although some of these employers may implement such programs purely to improve employee health, controlling costs seems to be their most important motivation (See also *Chapter 9*) (52;78). Moreover, more than half (52%) of the U.S. employers that offered worksite health promotion programs in 2012 were found to believe that they were effective in reducing the company's health care costs (79). As such, the present findings indicate that an innovative and dynamic industry appears to have outpaced the underlying evidence (69;80).

Next to the fact that worksite health promotion programs are generally thought to result in financial savings, they are also expected to result in various intangible corporate benefits that cannot be considered in a return on investment analysis. Examples of such intangible corporate benefits are improved job satisfaction, employee morale, and in-role performance (i.e. behavior required by formal job descriptions) (81;82). Moreover, worksite health promotion programs are hypothesized to strengthen a company's ability to attract new talent in a competitive market place, because healthy lifestyle benefits may entice younger employees. Among existing employees, on the other hand, worksite health promotion programs are thought to improve overall perceptions of the company, engender a greater sense of commitment and trust, and thus improve employee retention (83). However, the hypothesis regarding the positive effect of worksite health promotion programs on job satisfaction is not supported by the findings presented in *chapter 5* through *chapter 7*, and strong evidence of their favorable impact on other types of intangible corporate benefits is currently lacking.

Implications for research

Future research efforts in the field of worksite health promotion should be directed towards two important gaps in knowledge. First, the relatively small effects and lack of evidence of cost savings associated with existing worksite health promotion programs, does not negate the value of improving employee health. Therefore, more research is needed to explore what attributes of worksite health promotion programs are most important and how such interventions should be optimally designed (67). Second, researchers should help ensure that worksite health promotion program implementation and/or continuation decisions are made in an evidence-based manner, because a lack thereof may result in inappropriate decision-making and thus a waste of scarce resources.

Future directions of worksite health promotion programs

The absence of, or relatively small, effects of the evaluated interventions as well as their lack of cost savings is in line with the findings of other high-quality studies on primary prevention strategies in the workplace (84-88). This raises the question of whether primary prevention programs are indeed the "optimal" strategy for improving employee health and costs. The adoption of a high-risk approach may be more likely to be cost-effective and/or cost saving, as it is generally more efficient to concentrate limited resources where the need, and therefore also the benefits, are likely to be greatest (14). As such, future worksite health promotion programs are recommended to shift their focus from primary prevention for all employees towards prevention programs that are aimed at high-risk individuals (89). A possible way to do this is by offering more comprehensive worksite health promotion programs, in which all employees are screened for various health risks, after which only those with high-risks are referred to the necessary prevention and/or treatment programs. Amongst others, such comprehensive worksite health promotion programs may be aimed at tobacco cessation, physical activity promotion, stress management, weight management, and nutritional guidance (83). In addition, most of the evaluated interventions were mainly targeted at individual determinants of behavior (e.g. through health education and communication) (see *chapter 4, 5*, and *6*), whereas interventions targeted at both individual and environmental determinants are

expected to be more effective in achieving (health) behavior change (90;91). Therefore, future worksite health promotion programs are recommended to include both individual and environmental modifications. Examples of environmental modifications are healthy canteen food and physical activity promoting adaptations to the workplace, such as standing conference tables and the introduction of exercise balls. Moreover, a necessary prerequisite for any successful worksite health promotion program is a high level of participation, because "*nothing happens until [people] participate*" (3). Research indicates that participation levels are often far from optimal in worksite health promotion programs (92), and this was also the case in the applied studies. Possible means to improve program participation include the use of incentives, the provision of a variety of program modalities (e.g. coaching, health information), the use of multi-component programs, as well the integration of health promotion into the company's culture (3;70;92). Furthermore, as many worksite health promotion programs are associated with decreased effectiveness over time, future interventions are recommended to include follow-up contacts and/or booster sessions after their completion in order to better maintain their initial results. The cost-effectiveness and/or financial return of such "optimally" designed interventions should subsequently be established by performing (cluster-)RCT-based economic evaluations.

Improving evidence-based practice in the worksite health promotion field

Two important factors currently hinder worksite health promotion program implementation and/or continuation decisions from being made in an evidence-based manner, namely the poor methodological quality of most economic evaluations of worksite health promotion programs (*Chapter 2* and *3*) and the lack of uptake of their results (*Chapter 9*). To prevent inappropriate decision-making, researchers should ensure that both issues are addressed.

Improving the methodological quality of economic evaluations of worksite health promotion programs

Recommendations for improving the quality of economic evaluations of worksite health promotion programs have been extensively provided and discussed in *chapter*

8 as well as in the methodological considerations section of this chapter. Our most important recommendations include:

- Future economic evaluations should be conducted alongside (cluster-)RCTs to minimize the possible influence of selection bias.
- Future economic evaluations should be performed from both the employer's and societal perspective. This approach ensures that the results are directly interpretable for occupational health decision-makers and provides an indication of whether the "local rationality" of the company is in line with societal optimality.
- Future economic evaluations should assess the uncertainty surrounding their cost-consequence estimates, as failing to evaluate values under uncertainty may lead to biased conclusions and may thus result in inappropriate decision-making.
- Ideally, future economic evaluations base their sample sizes on economic endpoints. If this is not possible, researchers should use estimation and/or decision uncertainty rather than hypothesis testing (i.e. providing confidence intervals and assessing the probability of cost-effectiveness and/or financial return).
- Future economic evaluations should use multiple imputation for handling missing data, as study results may be biased when systematic differences exist between missing and observed values.
- Future economic evaluations should use price weights for valuing resource use that represent their true opportunity cost to the decision-maker at hand.

Moreover, methodological issues that warrant further inquiry include the methods for economic evaluations of clustered data, the measurement and valuation of changes in on-the-job productivity, the conceptualization of multipliers and compensation mechanisms in the valuation of changes in paid productivity, as well as the transferability of economic evaluation results across countries and jurisdictions.

Improving the uptake of economic evaluation results

In order to improve the uptake of economic evaluation results, researchers should ensure that their products are in line with the information needs of occupational

health decision-makers (93;94). The qualitative study included in *chapter 9* provided some initial clues as to what these information needs are at the company level. Namely, return-on-investment analyses performed from the employer's perspective were found to form the basis of business cases for worksite health promotion programs. Within these analyses, hard cost items (e.g. equipment costs, employee training costs) were of particular importance and reduced sickness absence-related costs were viewed as one of the most important benefits. Furthermore, decisions typically have to be made within a limited time frame and many decision-makers lack the skill set required to determine what economic evaluation results are most reliable, and what information should be considered, under which circumstances (95). Therefore, it is advisable to provide them with easy-to-use critical summaries of published studies (96). In the Netherlands, such critical summaries may be distributed through (applied) research institutes and/or employers' associations, or published in easily accessible journals, newsletters, or websites. Improving the economic evaluation skill set of occupational health decision-makers may be accomplished by educating them through a variety of means, including the development of handbooks and workshops on economic evaluation methods, integrating these topics into management, occupational health, and/or worksite health promotion training programs, and involving occupational health decision-makers in the process of commissioning studies (95;97;98). Participation in scientific studies is namely closely linked to the uptake of their results and may simultaneously lead to an improved economic evaluation skill set (96).

To further advance the development of a solid evidence base on the resource implications of worksite health promotion programs and to facilitate the uptake of their results, it is recommendable to develop a set of consensus-based guidelines for good practice when conducting and reporting economic evaluations of interventions in the workplace. In order to be successful, such guidelines must be based on sound economic principles and meet the needs of all stakeholders (99). As such, they are ideally developed through a close cooperation between economists, occupational health researchers, workplace parties, policy-makers, and all other possibly relevant stakeholders.

Concluding remarks

The present thesis indicated that (strong) evidence of the cost-effectiveness and/or financial return of worksite health promotion programs is currently lacking. Therefore, widespread implementation of such interventions in an effort to generate cost savings is not recommended, while some of them may be considered cost-effective if decision-makers are willing to invest a certain amount of money to improve employee health. Whether the latter is the case, however, is currently unknown.

The lack of evidence of cost savings associated with existing worksite health promotion programs, does not negate the value of improving employee health. Therefore, more research is needed to explore what attributes of worksite health promotion programs are most important and how such interventions should be optimally designed. Amongst others, existing worksite health promotion programs may be improved by using a so-called high-risk approach, including environmental modifications, incorporating strategies to improve program participation, and including follow-up contacts and booster sessions after their completion in order to better maintain their initial effects. The cost-effectiveness and/or financial return of such "optimally" designed interventions should subsequently be established by performing (cluster-) RCT-based economic evaluations. Furthermore, the methodological quality of economic evaluations of worksite health promotion programs is generally poor, as is the uptake of their results in daily practice. To prevent inappropriate decision-making, researchers should ensure that both issues are addressed and recommendations have been provided in this thesis as to how this may be established.

REFERENCES

1. Glasgow RE, McCaul KD, Fisher KJ. Participation in worksite health promotion: A critique of the literature and recommendations for future practice. *Health Educ Q* 1993;20(3):391-408.
2. McDaid D. The economics of mental health in the workplace: what dowe know and where do we go? *Epidemiol Psichiatr Soc* 2007;16:294-298.
3. Goetzel RZ, Ozminkowski RJ. The health and cost benefits of work site health-promotion programs. *Annu Rev Public Health* 2008;29(1):303-323.
4. Goetzel RZ, Juday TR. What's the ROI? A systematic review on return of investment (ROI) studies of corporate health and productivity management initiatives. *AWPH's Worksite Health* 1999;(6):12-21.
5. Bull SS, Gillette C, Glasgow RE, Estabrooks P. Work site health promotion research: to what extent can we generalize the results and what is needed to translate research to practice? *Health Educ Behav* 2003;30(5):537-549.
6. Burdorf A. Economic evaluation in occupational health-its goals, challenges, and opportunities. *Scan J Work Environ Health* 2007;33(3):161-164.
7. Patsopoulos NA. A pragmatic view on pragmatic trials. *Dialogues Clin Neurosci* 2011;13(2):217-224.
8. Higgins JPT, Altman DG. Chapter 8: Assessing risk of bias in included studies. In: Higgins J.P.T., Green S., editors. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0. (updated March 2011) ed. The Cochrane Collaboration; 2011.
9. Gomes M, Grieve R, Nixon R, Edmunds WJ. Statistical methods for cost-effectiveness analyses that use data from cluster randomized trials: A systematic review and checklist for critical appraisal. *Med Decis Making* 2012;32(1):209-220.
10. Willan AR, Briggs AH, Hoch JS. Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectiveness data. *Health Econ* 2004;13(5):461-475.
11. Drummond MF, Sculpher M.J., Torrance G.W., O'Brien B.J., Stoddart G.L. *Methods for the Economic Evaluation of Health Care Programmes. 3rd ed*. Oxford University Press: New York, 2005.
12. Grieve R, Nixon R, Thompson SG. Bayesian hierachical models for cost-effectiveness analyses that use data from cluster randomized trials. *Med Decis Making* 2010;30(2):163-175.
13. Gomes M, Grieve R, Nixon R, Ng ES, Carpenter J, Thompson SG. Methods for covariate adjustments in cost-effectiveness analysis that use cluster randomised trials. *Health Econ* 2012;21(9):1101-1118.
14. Rose G. Sick individuals and sick populations. *Int J Epidemiol* 2001;30(3):427-432.
15. Gatchel R, Kishino N. *Conceptual Approaches to Occupational Health and Wellness*: An Overview. In: Gatchel RJ, Schultz IZ, editors. Handbook of Occupational Health and Wellness. Springer US; 2012. p. 3-21.
16. Verweij LM. Occupational Health Guideline for Preventing Weight gain among Employees. Thesis (PhD), VU University Amsterdam, 2012.
17. Neve MJ, Collins CE, Morgan PJ. Dropout, nonusage attrition, and pretreatment predictors of nonusage attrition in a commercial Web-based weight loss program. *J Med Internet Res* 2010;12:e69.

18. van Dongen J, van Poppel M, Milder I, van Oers H, Brug J. Exploring the reach and program use of hello world, an email-based health promotion program for pregnant women in the Netherlands. *BMC Research Notes* 2012;5(1):514.

19. Brouwer WB, van Exel NJ, Baltussen RM, Rutten FF. A dollar is a dollar is a dollar--or is it? *Value Health* 2006;9(5):341-347.

20. Uegaki K. Economic evaluation of interventions for occupational health. Thesis (PhD), Vrije Universiteit Amsterdam, 2010.

21. Krol M, Brouwer W, Rutten F. Productivity costs in economic evaluations: past, present, future. *Pharmacoeconomics* 2013;31(7):537-549.

22. Tompa E, Culyer AJ, Dolinschi R. *Economic evaluations of interventions for occupational health and safety. Developing good practice.* Oxford University Press: New York, 2008.

23. Briggs A. Economic evaluation and clinical trials: size matters. The need for greater power in cost analyses poses an ethical dilemma. *BMJ* 2000;321(7273):1362-1363.

24. Bosmans JE. Cost-effectiveness of treatment of depression in primary care. Thesis (PhD). Vrije Universiteit Amsterdam, 2007.

25. Briggs AH, Gray AM. Power and Sample Size Calculations for Stochastic Cost-Effectiveness Analysis. *Med Decis Making* 1998;18(2):S81-S92.

26. Petrou S, Gray A. Economic evaluation alongside randomised controlled trials: design, conduct, analysis, and reporting. BMJ 2011;342:d1548.

27. Glick HA, Doshi JA, Sonnad SS, Polsky D. Economic evaluations in clinical trials. Oxford University Press: New York, United States, 2007.

28. Annemans L, Ollendorf LD. The Missing Link: Managing Missing Data in Economic Evaluations*. The Official News & Technical Journal Of The International Society For Pharmacoeconomics And Outcomes Research* 2007.

29. Zhang W, Bansback N, Anis AH. Measuring and valuing productivity loss due to poor health: A critical review*. Soc Sci Med* 2011;72(2):185-192.

30. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice*. Statist Med* 2011;30(4):377-399.

31. Sterne JAC., White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009;338:b2393.

32. van de Kerckhove W, Krenze T, Mohadjer L. Approaches to a Nonresponse Bias Analysis in an Adult Literacy Survey*. Proceedings of the Survey Research Methods Section, American Statistical Association* 2006;3790-3795.

33. Søgaard R, Sørensen J, Linde L, Hetland ML. The significance of presenteeism for the value of lost production: the case of rheumatoid arthritis. *Clinicoecon Outcomes Res* 2010;2:105-112.

34. Caverley N, Cunningham JB, MacGregor JN. Sickness Presenteeism, Sickness Absenteeism, and Health Following Restructuring in a Public Service Organization. *Journal of Management Studies* 2007;44(2):304-319.

35. Burton WN, Conti DJ, Chen CY, Schultz AB, Edington DW. The Economic Burden of Lost Productivity Due to Migraine Headache: A Specific Worksite Analysis. J Occup Environ Med 2002;44(6):523-529.

36. *Economie van de Gezondheidszorg*. Reed Business education: Amsterdam, 2012.

37. Goossens M, Rutten-van Mölken M, Vlaeyen J, van der Linden S. The cost diary: a method to measure direct and indirect costs in cost-effectiveness research. *J Clin Epidemiol* 2000;53(7):688-695.

38. Kessler R, Barber C, Beck A, Berglund P, Cleary P, McKenas D. The World Health Organization Health and Work Performance Questionnaire (HPQ). *J Occup Environ Med* 2003;45:156-174.

39. Kessler RC, Ames M, Hymel PA, Loeppke R, McKenas DK, Richling DE. Using the world health organization health and work performance questionnaire (HPQ) to evaluate the indirect workplace costs of illness. *J Occup Environ Med* 2004;46:S23-S37.

40. van Poppel MNM, de Vet HCW, Koes BW, Smid T, Bouter LM. Measuring sick leave: a comparison of self-reported data on sick leave and data from company records. *Occupl Med (Lond)* 2002;52(8):485-490.

41. Severens JL, Mulder J, Laheij RJ, Verbeek AL. Precision and accuracy in measuring absence from work as a basis for calculating productivity costs in The Netherlands. *Soc Sci Med* 2000;51(2):243-9.

42. Coughlin SS. Recall bias in epidemiologic studies. *J Clin Epidemiol* 1990;43(1):87-91.

43. Bhandari A, Wagner T. Self-reported utilization of health care services: improving measurement and accuracy. *Med Care Res Rev* 2006;63(2):217-235.

44. Hakkaart-van Roijen L, Tan SS, Bouwmans CAM. *Handleiding Voor Kostenonderzoek*. Methoden en Standaardkostprijzen Voor Economische Evaluaties in de Gezondheidszorg. Diemen, the Netherlands: College Voor Zorgverzekeringen; 2010.

45. Koopmanschap MA, Rutten FFH, van Ineveld BM, van Roijen L. The friction cost method for measuring indirect costs of disease. *J Health Econ.* 1995;14:171–189.

46. OECD. Sickness and Disability Schemes in the Netherlands: Country memo as a background paper for the OECD Disability Review. http://www.oecd.org/social/soc/41429917.pdf

47. Nicholson S, Pauly MV, Polsky D, Sharda C, Szrek H, Berger ML. Measuring the effects of work loss on productivity with team production. *Health Econ* 2006;15(2):111-123.

48. Krol M, Brouwer WBF, Severens JL, Kaper J, Evers SMAA. Productivity cost calculations in health economic evaluations: Correcting for compensation mechanisms and multiplier effects. *Soc Sci Med* 2012;75(11):1981-1988.

49. Jacob-Tacken KHM, Koopmanschap MA, Meerding WJ, Severens JL. Correcting for compensating mechanisms related to productivity costs in economic evaluations of health care programmes. *Health Econ* 2005;14(5):435-443.

50. Barte JCM, Ter Bogt NCW, Bogers RP, Teixeira PJ, Blissmer B, Mori TA, et al. Maintenance of weight loss after lifestyle interventions for overweight and obesity, a systematic review. *Obes Rev* 2010;11(12):899-906.

51. Herring B. Suboptimal provision of preventive healthcare due to expected enrollee turnover among private insurers. *Health Econ* 2010;19(4):438-448.

52. Horwitz JR, Kelly BD, DiNardo JE. Wellness Incentives In The Workplace: Cost Savings Through Cost Shifting To Unhealthy Workers. *Health Aff (Millwood)* 2013;32(3):468-476.

53. Verbeek J, Pulliainen M, Kankaanpaa E, Taimela S. Transferring results of occupational safety and health cost-effectiveness studies from one country to another - a case study. *Scand J Work Environ Health* 2010;36(4):305-312.

54. Reinhardt UE, Hussey PS, Anderson GF. U.S. Health Care Spending In An International Context. *Health Aff (Millwood)* 2004;23(3):10-25.

55. Sculpher M, Pang F, Manca A, Drummond M, Golder S, Urdahl H. Generalisability in economic evaluation studies in healthcare: a review and case studies. *Health Technol Assess* 2004;8(49):206.

56. Drummond M, Barbieri M, Cook J, Glick HA, Lis J, Malik F, et al. Transferability of economic evaluations across jurisdictions: ISPOR good research practices task force report. *Value Health* 2009;12(4):409-418.

57. Manca A, Willan AR. Lost in translation: accounting for between-country differences in the analysis of multinational cost-effectiveness data. *Pharmacoeconomics* 2006;24:1101-1119.

58. Pelletier KR. A review and analysis of the clinical and cost-effectiveness studies of comprehensive health promotion and disease management programs at the worksite: update VI 2000-2004. *J Occup Environ Med* 2005;47(10):1051-1058.

59. Pelletier KR. A Review and Analysis of the Clinical and Cost-effectiveness Studies of Comprehensive Health Promotion and Disease Management Programs at the Worksite: Update VIII 2008 to 2010. *J Occup Environ Med* 2011;53(11):1310-1331.

60. Pelletier KR. A review and analysis of the clinical and cost-effectiveness studies of comprehensive health promotion and disease management programs at the worksite: update VII 2004-2008. *J Occup Environ Med* 2009;51(7):822-837.

61. Pelletier KR. A review and analysis of the health and cost-effective outcome studies of comprehensive health promotion and disease prevention programs at the worksite: 1991-1993 update. *Am J Health Promot* 1993;8(1):50-62.

62. Pelletier KR. A review and analysis of the health and cost-effective outcome studies of comprehensive health promotion and disease prevention programs at the worksite: 1993-1995 update. *Am J Health Promot* 1996;10(5):380-388.

63. Pelletier KR. A review and analysis of the clinical and cost-effectiveness studies of comprehensive health promotion and disease management programs at the worksite: 1995-1998 update (IV). *Am J Health Promot* 1999;13(6):333-345.

64. Pelletier KR. A review and analysis of the clinical- and cost-effectiveness studies of comprehensive health promotion and disease management programs at the worksite: 1998-2000 update. *Am J Health Promot* 2001;16(2):107-116.

65. Goetzel RZ, Juday TR, Ozminkowski RJ. What's the ROI? A systematic review on return of investment (ROI) studies of corporate health and productivity management initiatives. *AWPH's Worksite Health* 1999;6:12-21.

66. Aldana SG. Financial impact of health promotion programs: a comprehensive review of the literature. *Am J Health Promot* 2001;15(5):296-320.

67. Baicker K, Cutler D, Song Z. Workplace wellness programs can generate savings. *Health Aff (Millwood)* 2010;29(2):304-11.

68. Chapman LS. Meta-Evaluation of Worksite Health Promotion Economic Return Studies: 2012 Update. *Am J Health Promot* 2012;26(4):TAHP-1-TAHP-12.

69. Osilla KC, van Busum K, Schnyer C, Larkin JW, Eibner C, Mattke C. Systematic review of the impact of worksite wellness programs. *Am J Manag Care* 2012;18(2):e68-e81.

70. Soler RE, Leeks KD, Razi S, Hopkins DP, Griffith M, Aten A, et al. A systematic review of selected interventions for worksite health promotion: The assessment of health risks with feedback. *Am J Prev Med* 2010;38(2, Supplement 1):S237-S262.

71. Anderson LM, Quinn TA, Glanz K, Ramirez G, Kahwati LC, Johnson DB, et al. The effectiveness of worksite nutrition and physical activity interventions for controlling employee overweight and obesity: a systematic review. *Am J Prev Med* 2009;37(4):340-357.

72. Proper KI, Koning M, van der Beek AJ, Hildebrandt VH, Bosscher RJ, van Mechelen W. The effectiveness of worksite physical activity programs on physical activity, physical fitness, and health. *Clin J Sport Med* 2003;13(2):106-117.

73. Engbers LH, van Poppel MNM, Chin AP, van Mechelen W. Worksite Health Promotion Programs with Environmental Changes: A Systematic Review. *Am J Prev Med* 2005 Jul;29(1):61-70.

74. Groeneveld IF, Proper KI, van der Beek AJ, Hildebrandt VH, van Mechelen W. Lifestyle-focused interventions at the workplace to reduce the risk of cardiovascular disease--a systematic review. *Scand J Work Environ Health* 2010;36(3):202-215.

75. Verweij LM, Coffeng J, Van Mechelen W, Proper KI. Meta-analyses of workplace physical activity and dietary behaviour interventions on weight outcomes. *Obes Rev* 2011;12(6):406-429.

76. Nordin M. Corporate Wellness Programs Are a Secure Investment (Part 4). 2013. http://www.huffingtonpost.com/melissa-nordin/corporate-wellness-progra_b_2862830.html

77. Hooftman W, van der Klauw M, Klein Hesselink J, Jongen M, Kraan K, Wevers C, et al. *Arbobalans 2011.* Hoofddorp: TNO; 2011.

78. Downey AM, Sharp DJ. Why do managers allocate resources to workplace health promotion programmes in countries with national health coverage? *Health Prom Int* 2007;22(2):102-111.

79. KFF/HRET. Employer health benefits: 2012 Annual Survey. 2012. Menlo Park, Calif: Chicago, Ill, Kaiser Family Foundation/Health Research and Education Trust.

80. Mattke S, Seid M, Ma S. Evidence for the effect of disease management: is $1 billion a year a good investment? *Am J Manag Care* 2007;13(12):670-676.

81. Riketta M. Attitudinal organizational commitment and job performance: a meta-analysis. *J Organiz Behav* 2002 1;23(3):257-266.

82. Anderson DR, Sexner SA, Gold DB. Conceptual Framework, Critical Questions, and Practical Challenges in Conducting Research on the Financial Impact of Worksite Health Promotion. *Am J Health Promot* 2001;15(5):281-288.

83. Shaw W, Reme S, Boot CRL. Health and Wellness Promotion in the Workplace. In: Gatchel RJ, Schultz IZ, editors. Handbook of Occupational Health and Wellness. Springer US; 2012. p. 365-332.

84. Robroek SJW, Polinder S, Bredt FJ, Burdorf A. Cost-effectiveness of a long-term Internet-delivered worksite health promotion programme on physical activity and nutrition: a cluster randomized controlled trial. *Health Educ Res* 2012;27(3):399-410.

85. Driessen M, Bosmans J, Proper K, Anema J, Bongers P, Van der Beek A. The economic evaluation of a Participatory Ergonomics programme to prevent low back and neck pain. *Work: A Journal of Prevention, Assessment and Rehabilitation* 2012;41(0):2315-2320.

86. Meenan RT, Vogt TM, Williams AE, Stevens VJ, Albright CL, Nigg C. Economic evaluation of a worksite obesity prevention and intervention trial among hotel workers in Hawaii. *J Occup Environ Med* 2010;52(Suppl 1):S8-S13.

87. Groeneveld IF, van Wier MF, Proper K, Bosmans JE, Van Mechelen W, van der Beek A. Cost-effectiveness and cost-benefit of a lifestyle intervention for workers in the construction industry at risk for cardiovascular disease. *J Occup Environ Med* 2011;56:610–617.

88. van Wier MF, Verweij LM, Proper KI, Hulshof CTJ, van Tulder MW, van Mechelen W. Economic evaluation of an occupational health guideline for preventing overweight among employees. *J Occup Environ Med* 2013;55(9):1000-1109.

89. Oude Hengel OH. Sustainable employability of Construction Workers. Thesis (PhD). VU University Amsterdam, 2013.

90. McLaren L, Hawe P. Ecological perspectives in health research. *J Epidemiol Community Health* 2005;59:6-14.

91. Richard L, Gauvin L, Raine K. Ecological Models Revisited: Their Uses and Evolution in Health Promotion Over Two Decades. *Annu Rev Public Health* 2011;32(1):307-326.

92. Robroek S, van Lenthe F, van Empelen P, Burdorf A. Determinants of participation in worksite health promotion programmes: a systematic review. *Int J Behav Nutr Phys Act* 2009;6:26.

93. Nutbeam D. Achieving 'best practice' in health promotion: improving the fit between research and practice. *Health Educ Res* 1996;11(3):317-326.

94. Walshe K, Rundall TG. Evidence-based Management: From Theory to Practice in Health Care. *Milbank Q* 2001;79(3):429-457.

95. Clancy CM, Cronin K. Evidence-Based Decision Making: Global Evidence, Local Decisions. *Health Aff (Millwood)* 2005;24(1):151-162.

96. Hoffmann C, Stoykova BA, Nixon J, Glanville JM, Misso K, Drummond MF. Do health-care decision makers find economic evaluations useful? The findings of focus group research in UK health authorities. *Value Health* 2002 Mar;5(2):71-78.

97. Drummond M, Brown R, Fendrick AM, Fullerton P, Neumann P, Taylor R, et al. Use of Pharmacoeconomics Information: Report of the ISPOR Task Force on Use of Pharmacoeconomic/Health Economic Information in Health-Care Decision Making. *Value Health* 2003;6(4):407-416.

98. Ross J. The use of economic evaluation in health care: Australian decision makers' perceptions. *Health Policy* 1995;31(2):103-110.

99. Tompa E, Verbeek J, van Tulder MW, de Boer A. Developing guidelines for good practice in economic evaluation of occupational health and safety intervention. *Scand J Work Environ Health* 2010;36(4):313-318.

100. Arbouw. Bedrijfstaksatlas 2012. http://www.arbouw.nl/pdf/tools/bedrijfstak-atlas-2012.pdf

101. Schaufeli WB, Bakker AB. *Utrecht Work Engagement Scale*. Utrecht, the Netherlands: Occupational Health Psychology, Unit Utrecht University; 2003.

102. Jansen NWH, Kant IJ, van den Brandt PA. Need for Recovery in the Working Population: Description and Associations With Fatigue and Psychological distress. Int J Behavl Med 2001;9(4):322-340.