Trial-Based Economic Evaluations in Occupational Health

Principles, Methods, and Recommendations

Johanna M. van Dongen, MSc, Marieke F. van Wier, PhD, Emile Tompa, PhD, Paulien M. Bongers, PhD, Allard J. van der Beek, PhD, Maurits W. van Tulder, PhD, and Judith E. Bosmans, PhD

To allocate available resources as efficiently as possible, decision makers need information on the relative economic merits of occupational health and safety (OHS) interventions. Economic evaluations can provide this information by comparing the costs and consequences of alternatives. Nevertheless, only a few of the studies that consider the effectiveness of OHS interventions take the extra step of considering their resource implications. Moreover, the methodological quality of those that do is generally poor. Therefore, this study aims to help occupational health researchers conduct high-quality trial-based economic evaluations by discussing the theory and methodology that underlie them, and by providing recommendations for good practice regarding their design, analysis, and reporting. This study also helps consumers of this literature with understanding and critically appraising trial-based economic evaluations of OHS interventions.

R esources for occupational health are scarce.[1,2] Therefore, decision makers in this field increasingly call upon advisors and researchers to demonstrate that occupational health and safety (OHS) interventions are not only effective but also efficient in terms of their resource implications. Economic evaluations provide information on the relative efficiency of two or more alternative interventions and are defined as "the comparative analysis of alternative courses of action in terms of both their costs and consequences."[1(p9)] The main aspects of any economic evaluation are to identify, measure, value, and compare the costs and consequences of alternatives.[1]

In the health care sector, economic evaluations are increasingly being conducted and play an important role in many countries when deciding whether (new) treatments should be covered by public funding.[1] Nevertheless, only a few of the studies that consider the effectiveness of OHS interventions take the extra step of considering

whether they are efficient in terms of their resource implications.[3] Moreover, the methodological quality of those that do is generally poor.[4–7] Reasons for this may be the distinct challenges that confront researchers when trying to identify the resource implications of OHS interventions, and a lack of recommendations on how to deal with these issues.[3] Many economic evaluation text books and articles are designed for use in health care settings and may be difficult to adapt to the occupational health context.[4]

Effectiveness trials are a commonly used vehicle for economic evaluations, as they provide a unique opportunity to reliably estimate the resource implications of a new intervention without substantially higher research expenses. Although some efforts have been undertaken to improve the quality of (trial-based) economic evaluations in occupational health,[3,8,9] more needs to be done to accomplish this. Therefore, this study aims to help occupational health researchers conduct high-quality trial-based economic evaluations by discussing the theory and methodology that underlie them, and by providing recommendations for good practice regarding their design, analysis, and reporting.

DESIGN OF AN ECONOMIC EVALUATION

Kind of Economic Evaluations

Choosing the appropriate kind of economic evaluation for a particular occupational health decision context can be a challenge as a result of the relative complexity of the decision-making context that generally includes multiple stakeholders (eg, workers, employers, insurance companies, public policymakers). Four kinds of economic evaluations are distinguished. There are similarities across the 4 kinds. The main difference is the metric used to measure the key outcome (health and/or safety, in the case of OHS interventions).[10]

1. *Cost-effectiveness analysis (CEA).* Costs and some consequences (eg, productivity, health care utilization implications) are measured in monetary units, whereas the key outcome is measured in natural units.[1]
2. *Cost–benefit analysis (CBA).* Both costs and consequences are measured in monetary units. In business administration, CBAs are sometimes described as return-on-investment (ROI) analyses.
3. *Cost-utility analysis (CUA).* Costs and some consequences are measured in monetary terms, whereas the key outcome is measured in utility units. Utilities are often expressed in terms of quality adjusted life years (QALYs).[1]
4. *Cost-minimization analysis.* Only costs are considered across alternatives, as it is assumed that the consequences are similar. Cost-minimization analyses are considered inappropriate if there is uncertainty regarding a possible difference in the magnitude of consequences.[1]

Which kind of economic evaluation is most appropriate depends on the stakeholders involved and the question being asked. Generally, employers are most interested in CBAs that can provide insight into the impact of an intervention on a company's bottom line, whereas public policymakers may be more interested in CEAs and CUAs, particularly if monetary measures do not adequately capture important health outcomes.[1,8,11] Therefore, it is recommended that researchers

conduct various kinds of economic evaluations within the same study to inform all relevant stakeholders.[3]

When to Undertake an Economic Evaluation?

Economic evaluations are often conducted alongside ("piggybacked" onto) trials evaluating the effectiveness of OHS interventions. Various design aspects are, therefore, typically determined by the requirements of the effectiveness trial (eg, alternatives, outcome measures). Nevertheless, to ensure that all relevant economic data are collected in a valid, reliable, and efficient way, it is important to consider the requirements for the economic evaluation at the earliest possible stage.[12–14]

Debate exists as to whether an economic evaluation should be included in a trial before the effectiveness of a new intervention is established. Nevertheless, not including an economic evaluation would risk losing the opportunity to simultaneously collect cost and effect data.[14] Also, the absence of statistically significant consequence/effect differences between the alternatives being compared does not necessarily imply that the new alternative is not cost-effective and/or cost-beneficial. Economic evaluations are about the joint distribution of costs and consequences and could demonstrate clear cost-effectiveness/cost–benefit when neither cost nor consequence differences are individually significant.[14] Also, cost savings might occur in the absence of health improvements and could thus be missed if an economic evaluation is not performed.

Trial Design

Pragmatic randomized controlled trials (RCTs) are generally acknowledged as the best vehicle for economic evaluations, because they enable the evaluation of the resource implications of OHS interventions under "real life" conditions. This setup increases the external validity of results, while the internal validity is guaranteed by the randomization of participants.[4,14] Within the occupational health setting, however, participant-level randomization may not always be feasible (eg, when interventions include organizational components). In such cases, randomization at the level of departments or locations might provide a more feasible approach (ie, cluster-RCTs).[3] To ensure that the results of an economic evaluation are generalizable to occupational health practice, trial conditions should resemble daily practice as much as possible. For example, participants should be similar to those who will experience the intervention if it is implemented broadly, monitoring should be done under routine circumstances, and interventions should be compared with usual practice.

Perspective

An essential aspect of an economic evaluation is its perspective. Perspective refers to the "point of view" taken to identify relevant costs and consequences for inclusion in the evaluation. The chosen perspective may be that of any relevant stakeholder or an aggregate of stakeholders such as a societal perspective. The perspective determines which costs and consequences are included. In the societal perspective, for example, all costs and consequences are considered irrespective of who pays or benefits, whereas only those borne by employers are included when the employer's perspective is applied. Given this fact, the perspective is a critical element in an analysis and should therefore be stated explicitly.[1]

The OHS interventions are typically initiated by company management, either to comply with the law, in an effort to save money (ie, reduced sickness absence costs), or for moral reasons.[11] Consequently, most economic evaluations of such interventions are performed from the employer's perspective,[4–7,15] but other perspectives may also be relevant, for example, worker's, insurer's, and societal perspective. When the employer's perspective is applied, key worker outcomes, such as the value of worker health, are often not included in the analysis, but simply the health-related expenses incurred by an employer (eg, productivity implications). This is a critical oversight,

as occupational health is essentially about worker health. A societal perspective is particularly useful to consider as the perspective in a study, as it provides insight into the net effect across all stakeholders. Hereby, it better ensures that the societal costs of an intervention are less than the benefits experienced by all stakeholders, rather than simply the company's costs being less than its benefits.[3] This information will ensure that there is a net societal benefit, rather than simply cost shifting from one stakeholder to another. In addition, the disaggregated information on costs and consequences from a societal perspective provides a good sense of their distribution across stakeholders. Such information can be the launch pad for bargaining between them.[1] This may be of particular importance in countries with dual-payer (eg, The Netherlands) and universal health care systems (eg, The United Kingdom), because employers generally bear most of the costs of OHS interventions, whereas in such jurisdictions the health care system and/or government reaps a large part of their benefits (ie, reduced medical spending).[16] Therefore, it is recommendable to supplement findings from the employer's perspective with those from other relevant perspectives, particularly the societal one.

Analytic Time Frame

Researchers also need to decide about the time frame over which costs and consequences are analyzed. The analytic time frame ought to cover the entire period over which costs and consequences flow from the alternatives under consideration.[12] This time frame generally extends beyond the follow-up needed to establish the effectiveness of a new intervention. To illustrate, the follow-up of an effectiveness trial may be terminated after the occurrence of the clinical event of interest (eg, incidence of repetitive strain injury). If this follow-up was used for the economic evaluation, all costs and consequences incurred during the course of the disorder or its recurrences would not be taken into account (eg, repetitive strain injury–related medication and/or operation costs), leading to an underestimation of the total costs and consequences. Although the optimal follow-up period is generally unknown, researchers and readers should at least feel confident that the most important costs and consequences are covered by the chosen analytic time frame. In addition, future costs and consequences that occur after the measurement period can be estimated using information and data from various sources. This is particularly important to do if future costs and consequences are expected to be substantial (eg, many of the [health] benefits of preventive interventions are thought to occur in the future).

Identification, Measurement, and Valuation of Resource Use

In economic evaluations, costs and some consequences are expressed in monetary units. For this purpose, relevant resource use categories should be identified, measured, and valued. As discussed earlier, relevant resource use categories for inclusion in an economic evaluation depend on its perspective. Other factors that might determine the relevance of a resource use category are, among others, the country or jurisdiction in which the study is undertaken and the nature of the alternatives being compared.

After relevant resource use categories are identified, researchers should determine how to cost them. Costing generally involves three steps: (1) the measurement of quantities of resources consumed (Q), (2) the assignment of unit prices (P), and (3) the valuation of resources consumed by multiplying their quantities by their respective unit prices ($Q*P$).[1] These estimates should be reported separately so that the reader can judge the relevance of these measures to his or her setting.[17]

Measurement of Quantities of Resources Consumed

Resource use data are ideally collected prospectively through a data collection process that is fully integrated into the effectiveness trial.[1,13] Also, when collecting self-reported resource use

data, researchers have to balance recall bias against completeness of information. Shorter recall periods reduce the risk of participants forgetting important information. Nevertheless, collecting data with relatively short recall periods (eg, a couple of weeks) over a longer period of time may be overly burdensome to participants and may thus increase the risk of missing data and dropouts. Therefore, it may be better to maximize completeness at the cost of some recall bias,[14] for example, by using 2- to 3-month recall periods in a trial with a long-term follow-up (≥12 months).[18] Also, care should be taken to collect resource use data continuously during follow-up and to avoid the need for extrapolation of resource use estimates between measurement periods.

Assignment of Unit Prices

Unit prices used for valuing resource use ought to reflect opportunity costs, that is, "the value of a resource in its most highly valued alternative use."[8(p56)] In a world of perfect markets, such costs are revealed by the market price of a good or service. Nevertheless, if a competitive market does not exist for a good or service, market prices often are an inaccurate measure of its value. For example, if a premium is paid for a good or service due to restricted market entry, market prices may overestimate the opportunity costs at the societal level. When the societal perspective is applied, an adjustment should, therefore, be made to the market price, for example, by using the price of a comparable good or service.[8] For the employer's perspective, the actual purchase costs incurred by the employer may be more appropriate, as they better represent the sum of money that is not available to the employer for its best alternative use.[12,19] Thus, appropriate unit prices may vary between perspectives, and researchers should ensure that they reflect the true resource implications to the decision maker at hand.[8]

A brief description of the methods used for measuring and valuing the most frequently used resource use categories in economic evaluations of OHS interventions is provided later. The most frequently used resource use categories are intervention, productivity, health care, and workers' compensation costs.[4–7,15]

Intervention Costs

Information on the market price of an intervention may be derived from vendors or company and/or research project records. Many trials, however, assess novel interventions that either have no predefined price weights associated with them or for which the use of market prices is inappropriate (eg, when the societal perspective is applied).[12] In such cases, the actual intervention costs can be assessed using a bottom-up micro-costing approach, in which detailed data regarding the quantities of resources consumed as well as their unit prices are collected per intervention component separately. Such resources may include intervention staff hours, materials used, depreciation, overhead activities, square feet of office space, and traveling.[1,3,12] Also, workers may be taken away from their regular production activities to participate in the intervention and this should be accounted for as well. Costs associated with the intervention's evaluation should not be included unless it is a condition of implementation.[8] Quantities of resources consumed can be measured using administrative databases, expert panels, surveys or interviews with intervention participants and/or providers, intervention operation logs, or observations.[20] Unit prices may be collected from administrative databases, scientific literature, vendors, and/or costing manuals (eg,[21]).

Health Care Costs

Ideally, all health care service use is measured to reduce the likelihood that (unexpected) shifts in health care utilization rates are missed. Although this approach will increase the validity of the results, it may not always be feasible. An alternative strategy is to limit data collection to those health care services that are related

to the alternatives and/or condition under study.[12] A description of the care path for the condition under study might provide researchers with a clear picture of what those health care services are. In all cases, care should be taken to include the most important cost drivers.

Health care utilization can be measured through various means, including retrospective questionnaires, prospective resource use diaries (ie, cost diaries), and insurance or hospital databases. Databases, however, may not always contain all required data, and their validity and reliability may not be very high.[10] Moreover, health care costs borne by participants (eg, copayments, over-the-counter medication) are typically not included in these databases. Therefore, researchers are often dependent on self-report data to measure these health care utilization items. To value health care utilization, unit prices may be either estimated using a micro-costing approach or based on predefined price weights, prices according to professional organizations, or tariffs. Typically, several methods are used simultaneously.[10,19]

Productivity Costs

For employers, an important benefit of OHS interventions is the resulting changes in productivity loss. Productivity loss can be defined as the company's output loss corresponding to reduced labor input (ie, time and efforts/skills of the workforce). According to this definition, to value productivity loss is to value the output loss.[22] Unfortunately, however, objective measurement of the true impact of reduced labor input on a company's output is often impossible to estimate. Therefore, researchers typically use proxies of productivity loss, which are often estimated using (self-reported) data on the participants' level of absenteeism (ie, sickness absence) and/or presenteeism (ie, reduced performance while at work). The methodologies used for measuring and valuing absenteeism and presenteeism are a fiercely debated topic in the field of economic evaluations. Later, a brief description of the most frequently used methods is provided. For more information about the main debates and developments regarding the identification, measurement, and valuation of productivity, we refer to other publications.[22,23]

The two main methods for estimating absenteeism costs are the Human Capital Approach (HCA) and the Friction Cost Approach (FCA). For both methods, the number of sickness absence days has to be collected, for which administrative databases, self-report (questionnaires), or reports by others can be used.[9] For the FCA, it is also important to identify the number and duration of different absence periods. According to the HCA, absenteeism costs are equal to the amount of money participants would have earned had they not been injured or ill.[4,21] Therefore, in the HCA, sickness absence days are typically valued using actual wage rates of participants (including employment overheads and benefits) and represent losses for the entire duration of absence.[1,19,24] It is argued that the HCA overestimates the true societal cost of sickness absence, as the possible replacement of workers with long-term sickness absence is not taken into account.[1,4] Therefore, the FCA was developed, in which production losses are assumed to be 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 (ie, friction period).[23] In the FCA, absenteeism is typically valued using age-, gender- and/or education-specific price weights.[25] The length of the friction period depends on the state (ie, the unemployment rate) and efficiency of the labor market. As such, friction periods typically differ between countries and should be estimated per country separately.[1] If there are important changes in the economic climate, it may be necessary to estimate the friction period anew. In the Netherlands, a friction period of 23 weeks is currently assumed.[21] Thus, if a sickness absence period exceeds 23 weeks, absenteeism costs are truncated at the costs of 23 weeks. Furthermore, as a reduction of labor input is often assumed to cause a less than proportional reduction in productivity, Koopmanschap et al[25] also proposed the application of an

elasticity factor of 0.8, which is often used in economic evaluations that apply the FCA. This elasticity factor implies that a 100% loss of labor input corresponds with an 80% reduction in productivity.[25]

In the economic evaluation literature, the need to consider presenteeism as a component of the costs incurred from productivity loss is increasingly being recognized.[9] Presenteeism is typically estimated using participant self-report or report by others. For this purpose, various instruments have been developed, including both generic[26–29] and disease-specific questionnaires.[30,31] Most of these questionnaires measure work performance in terms of points, percentages, or proportions.[32] These responses can then be used to estimate the total number of working days lost due to presenteeism by using the following equation:

$$P = (E - A)^* p$$

where P is full working days lost because of presenteeism, E is total working days, A is sickness absence days, and p is the proportion of lost work performance estimated by the instrument used in the study.[22] To value the number of lost working days due to presenteeism, actual wage rates of participants, or age-, gender-, and/or job-specific price weights can be used. Researchers should be aware, however, that the estimated number of work days lost because of presenteeism may vary widely between instruments. This suggests a lack of comparability among instruments, but it is still unclear which instrument provides the best presenteeism estimate.[22] Given its significance, however, ignoring presenteeism may lead to severe underestimations.[22] Therefore, researchers are recommended to include this resource use category whenever possible. To assess the possible influence of the choice of instrument, sensitivity analyses can be performed (see later).

Workers' Compensation Costs

Workers' compensation is an insurance program, offered in some countries (eg, Canada, the United States), through which workers may receive wage replacement and/or medical benefits in the event of an occupational injury or disease. Funding usually comes from premiums paid by employers.[8] To estimate workers' compensation costs, total claim costs per participant can be obtained from company and/or workplace insurance records. It is generally inadequate, however, to use workers' compensation costs as the sole cost category, as they do not reflect the full extent of work-related injuries and illnesses.[4] Many compensable injuries and illnesses go unreported and others are not compensable.[4] When supplementing health care and/or productivity costs with workers' compensation costs, double counting should be avoided. Also, insurance premium-related wage replacement benefits should be excluded for the societal perspective, as they constitute "transfer payments" from the employer via the insurer to the worker rather than depleted sources.[1,4]

Identification, Measurement, and Valuation of Outcomes

As noted previously, CEAs have the key outcome measured in natural units. The most appropriate outcome used for this purpose depends on the nature of the alternatives being compared, the condition under study, and/or the applied perspective. Sometimes, there may be some concern about whether the chosen outcome captures all relevant consequences. If this is a concern, it is advisable to conduct multiple CEAs using different outcomes.[8] In CUAs, the key outcome is measured in utility units, generally known as QALYs. They capture both the duration of survival and health-related quality of life in a single measure.[1,12,14] An advantage of QALYs is that they provide a general index score that allows decision makers to compare the consequences of a range of interventions for different health issues.[1,10] Nevertheless, even though QALYs are the preferred outcome measure when health care interventions for patients are evaluated from

the societal perspective,[13,21,33] they have not yet been frequently used in economic evaluations of OHS interventions.[4,6,7,34] This may be due to the fact that QALYs may not reflect what occupational health decision makers feel is most important in terms of outcomes. In the case of a workplace safety programs, for example, outcomes such as worker safety may be more meaningful to decision makers than a utility-weighted health measure.[11] Moreover, occupational health decision makers are generally unfamiliar with QALYs, and QALYs seem to lack sensitivity to mild conditions that are often the focus of OHS interventions (eg, of worksite health promotion programs).[35] Therefore, more sensitive utility measures are warranted for economic evaluations of OHS interventions and/or utility measures that are more applicable to the occupational health setting, for example, the recently conceptualized "Disease-Adjusted Working Years," which aims to express the amount of working years lost because of poor working conditions and associated illness.[36,37]

ANALYSIS OF AN ECONOMIC EVALUATION

Later, we discuss some important issues in the analysis of trial-based economic evaluations. To illustrate some of them, data are used from an economic evaluation that was previously performed alongside a 12-month pragmatic RCT, in which construction workers at risk for cardiovascular disease either received a lifestyle intervention or usual practice. A CEA in terms of kilogram body weight loss was performed from the societal perspective and a CBA from that of the employer. Resource use categories included intervention, health care, absenteeism, and sports costs and were expressed in 2008 Euros. More detailed information about this trial-based economic evaluation can be found elsewhere.[38]

Sample Size

Ideally, economic outcomes are used in the sample size calculation of a trial.[13] Nevertheless, although various techniques have been proposed to estimate the appropriate sample size for economic endpoints,[39–42] sample size calculations are typically performed on the basis of primary outcomes.[10,13,14] This is due to the fact that cost data are right skewed and therefore require larger sample sizes to detect relevant differences than (health) outcome data. A large sample size may be neither feasible nor ethically acceptable.[14,43] Also, a large number of parameters have to be specified to perform sample size calculations for economic endpoints (eg, variance parameters of effectiveness measures, cost measures, incremental cost-effectiveness ratios [ICER]), many of which are hard to predict a priori.[39,41,42] Consequently, trial-based economic evaluations are typically underpowered for economic outcomes.[10] Low-powered studies have imprecise and uncertain cost estimates and should be interpreted with caution.[43] Moreover, if studies are likely to be underpowered, researchers are recommended to use estimation rather than hypothesis testing (ie, by using confidence intervals rather than *P* values).[1]

Adjusting for Differential Timing

Interventions may have different time profiles of costs and consequences. Within occupational health, intervention costs are generally incurred immediately, while consequences such as productivity costs might extend into the future.[44] Two types of adjustments should be made to account for these differences in timing. The first concerns the adjustment of cost data for inflation, that is, "the general upward price movement of goods and services."[12] Because of inflation, prices drawn from different years are generally not comparable.[8] All prices should, therefore, be adjusted to the same reference year using consumer price indices and the applied reference year should be stated explicitly.[17] The second adjustment concerns the adjustment of cost and outcome data for time preferences of individuals when they are collected over a period of more than 1 year.[12] Even within a world with zero inflation, individuals have a preference for

receiving benefits today rather than in the future.[1] Therefore, costs and consequences incurred in different years have to be discounted at some rate to estimate their present value.[44] The appropriate discount rate depends on the borrowing cost of money and other contextual factors. Guidelines for discount rates used in public sector projects are provided by some jurisdictions. For example, in the Netherlands, cost data should be discounted at 4% and health outcomes at 1.5%, while both should be discounted at 3.5% in the United Kingdom.[21,33]

Intention-to-Treat and Missing Data

Guidelines for conducting trials prescribe that all participants should be included in the analyses, all retained in the group to which they were allocated (ie, intention-to-treat analysis).[45] Nevertheless, true intention-to-treat analyses are often hampered by missing data, which are generally inevitable in trials. For economic evaluations, this problem is even more pronounced, because total costs are typically the sum of numerous cost components. As such, cost data will already be incomplete if one component is missing.[13] Missing data itself may have no relation to observed and unobserved factors among participants (MCAR: missing completely at random), may only have a relationship to observed factors (MAR: missing at random), or may also have a relationship to unobserved factors (MNAR: missing not at random) (see Box 1 for a more detailed description).[46] Historically, complete-case analyses (ie, eliminating cases with missing data) were used to deal with missing data and this is still an often-used approach in trial-based economic evaluations.[47] Nevertheless, complete-case analyses reduce the power of a study and lead to biased estimates if missing data are not MCAR.[12,13] If the rate of missing data is smaller than 5%, complete-case analyses may be considered. If more than 5% of data are missing, researchers should use imputation techniques to fill in missing values. Nowadays, multiple imputation is generally recommended to impute missing data.[13,14] When using multiple imputation, multivariate regression techniques are used to predict missing values on the basis of observed factors.[12,14] To account for the uncertainty about the missing data, several different imputed data sets are created.[46] As a rule of thumb, White et al[48] suggested that the number of data sets should at least be equal to the percentage of incomplete cases. The imputed data sets are subsequently analyzed separately to obtain a set of parameter estimates, which can then be pooled using Rubin's rules to obtain overall estimates, variances, and 95% confidence intervals (95% CIs).[46,48,49] Multiple imputation leads to unbiased estimates if missing data are MAR.[12] Researchers should bear in mind, however, that cost and consequence estimates derived using multiple imputation are less reliable and precise than those based on a 100% complete data set.[14] Every endeavor should, therefore, be made to minimize the amount of missing data.

Incremental Analysis of Costs and Consequences

After costs and consequences have been quantified, their mean differences between the intervention and control group(s) as well as the statistical significance of these differences need to be assessed.[12]

As mentioned previously, cost data are typically right skewed. This is caused by the fact that only a small proportion of participants incur high costs and costs are naturally bound by zero (see Fig. 1).[1]

The skewed cost distribution complicates the analysis of cost data, as it violates the assumptions of standard statistical tests, such as independent *t* tests and linear regression analyses. A standard approach to describe skewed data is to provide a summary measure of the distribution in the form of a median. Nevertheless, this is inappropriate for cost data as decision makers need to be able to estimate the total cost of implementing a new intervention (total implementation costs = mean costs per participant × the number of participants). As such, the arithmetic mean is generally viewed as the most informative measure to describe cost data.[1,14,50] Various methods are currently used to compare cost data between study arms,

BOX 1. Types of Missing Data[46]

1. *MCAR.* The "missingness" of data has no relationship to observed and unobserved factors among participants. For example, sickness absence data may be missing because of problems with the registration of this data due to a temporary computer problem.
2. *MAR.* The "missingness" of data has a relationship to observed factors among participants, but not to unobserved factors. For example, missing sickness absence durations may be longer than available sickness absence durations but only because older employees may be more likely to have missing sickness absence data.
3. *MNAR.* Even after the observed data are taken into account, systematic differences remain between the missing values and the observed values. This means that the "missingness" of data also has a relationship to unobserved factors. For example, in trials relying on self-reported sickness absence, participants with longer sickness absence durations may be more likely to forget to return their cost diaries because they are not feeling well.
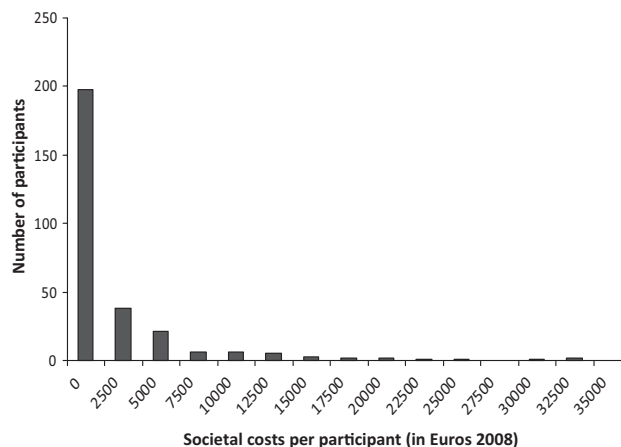
FIGURE 1. Distribution of the societal costs per participant to a trial-based economic evaluation of a lifestyle intervention for construction workers at risk for cardiovascular disease compared to usual practice.[38]

including standard nonparametric tests (eg, Mann–Whitney *U* test), *t* tests on log-transformed data, and nonparametric bootstrapping. Standard nonparametric tests compare the distribution of the data instead of means and are therefore inappropriate. Transformations to normalize the distribution are not straightforward and are often sensitive to departures from distributional assumptions.[13] Moreover, back-transformations are often complicated. Therefore, researchers increasingly favor the nonparametric bootstrap,[13,50] which can be used to estimate 95% CIs around mean cost differences while avoiding distributional assumptions (Box 2).[51]

Comparing Incremental Costs and Consequences

The core of any economic evaluation is the analysis of the relation between the costs and consequences of alternatives. The preferred methods for conducting such analyses differ between the types of economic evaluations and are discussed later.

CEA and CUA

In CEAs and CUAs, an ICER is calculated by dividing the mean difference in cost (Δ Cost) between study arms by that in

BOX 2. Nonparametric Bootstrapping

With nonparametric bootstrapping, statistical analyses are based on repeatedly sampling with replacement from the observed data. In short, a sample of N participants is repeatedly drawn with replacement from both the intervention and control groups separately, where N equals the number of participants per study arm. Every resample (ie, bootstrap sample) is the equivalent of a repetition of the trial. Because resamples have been drawn with replacement (ie, per sample, participants can be drawn more than once), these bootstrap samples differ from one another. Per bootstrap sample, the statistics of interest are calculated (eg, the difference in arithmetic mean costs and effects, incremental cost-effectiveness ratios, cost–benefit estimates). By doing so multiple times, a distribution for the statistics of interest is generated that provides an approximation of its population sampling distribution, which can then be used to estimate confidence intervals.[12] At least 2000 bootstrap samples are recommended and preferably more.[52] Various methods have been proposed to estimate bootstrapped 95% CIs, of which the bias corrected and accelerated method is currently the preferred one.[51,53] Nonparametric bootstrapping is available in many software packages, including SPSS, SAS, STATA, and R.

BOX 3. Calculation and Interpretation of the Incremental Cost-Effectiveness Ratio (ICER) of a Lifestyle Intervention for Construction Workers at Risk for Cardiovascular Disease Compared to Usual Practice[38]

During follow-up, intervention group participants significantly decreased their body weight by 2.02 kilogram compared to the control group (Δ Effect). Mean societal costs per participant were nonsignificantly higher in the intervention group than in the control group by €293 (Δ Cost). Using this information, the incremental cost-effectiveness ratio (ICER) can be calculated as follows:

ICER : €293/2.02 = €145

This ICER indicates that the society has to pay €145 per participant in the intervention group for each additional kilogram body weight loss compared with usual practice.

effect (Δ Effect). The ICER indicates the additional costs of a new intervention in comparison with a control condition per unit of effect gained.[1,12]

$$\frac{\text{Cost}_{\text{intervention}} - \text{Cost}_{\text{control}}}{\text{Effect}_{\text{intervention}} - \text{Effect}_{\text{control}}} = \frac{\Delta\text{Cost}}{\Delta\text{Effect}} = ICER$$

To illustrate, a description of the calculation and interpretation of the example trial's ICER is provided in Box 3.

Incremental cost-effectiveness ratios are generally hard to interpret. For example, negative ICERs might represent reduced costs and positive effects indicating a win–win situation or increased costs and negative effects indicating a lose–lose situation.[14] Therefore, ICERs are often graphically illustrated on cost-effectiveness planes (CE-planes), in which incremental effects are plotted on the x axis and incremental costs on the y axis (Fig. 2).[54,55]
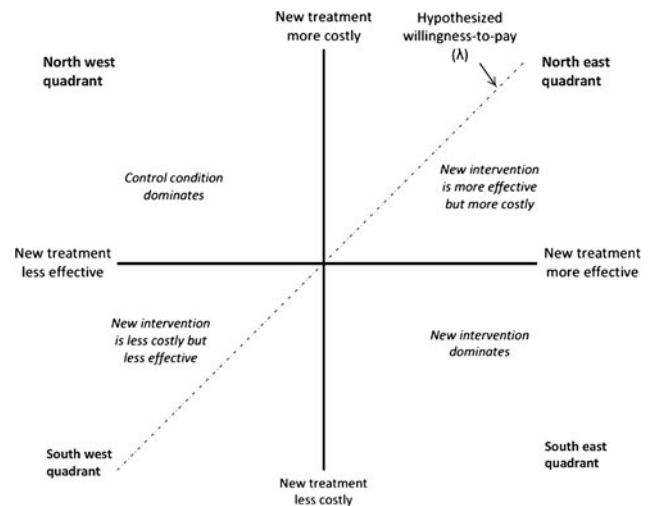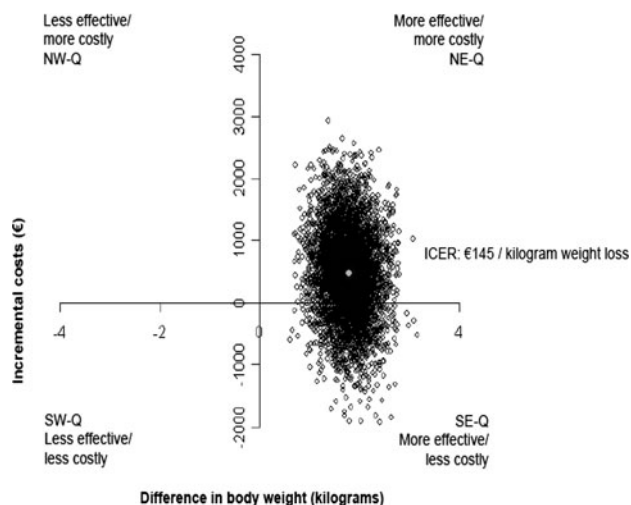

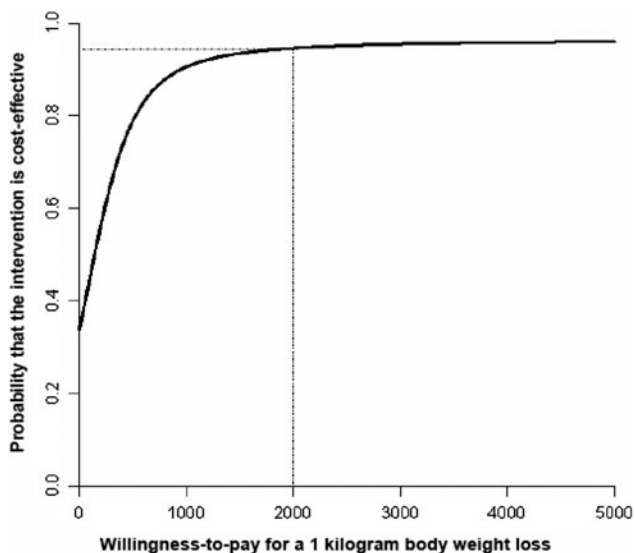**FIGURE 2.** Cost-effectiveness plane.

If an ICER is located either in the South East Quadrant (SE-Q) or in the North West Quadrant (NW-Q), the choice between alternatives is clear (assuming that there is no uncertainty surrounding the ICER). In the SE-Q, the new intervention is more effective and less costly than the control condition and is therefore said to dominate the control condition. In the NW-Q, the opposite is true and the new intervention is dominated by the control condition. If a new intervention is more effective and more costly (NE-Q: North East Quadrant) or less effective and less costly (SW-Q: South West Quadrant), the decision whether or not to adopt it depends on the so-called "willingness-to-pay" (λ). That is, the maximum amount of money decision makers are willing to pay for an additional unit of effect.[1] To illustrate, a hypothesized λ is depicted as the diagonal line in Figure 2 and divides the CE-plane into a cost-effective and a non–cost-effective halve. Incremental cost-effectiveness ratios located to the right of this line are considered acceptable, whereas ICERs located to the left are considered inacceptable.[14,54,55] The more decision makers are willing to pay for an additional unit of effect, the steeper the slope of this line.[14]

With participant-level data, it is natural to consider representing the uncertainty surrounding ICERs using 95% CIs. Nevertheless, as a ratio measure, estimating 95% CIs around ICERs is not straightforward and, more importantly, 95% CIs around ICERs suffer from the same interpretation problem as ICERs.[55] Therefore, alternative methods have been proposed to estimate the uncertainty surrounding ICERs. Current guidelines recommend using the bootstrap method described in Box 2. In this case, both incremental costs and effects are calculated per bootstrap sample. The uncertainty surrounding an ICER can then be graphically illustrated by plotting these bootstrapped incremental cost-effect pairs (CE-pairs) on a CE-plane. As indicated by the example trial's CE-plane provided in Figure 3, CE pairs commonly cover more than one quadrant.

Although CE planes give a good impression of the uncertainty surrounding the ICER, they do not provide a summary measure of the joint uncertainty of costs and effects.[56] Therefore, cost-effectiveness acceptability curves (CEACs) were introduced that provide insight into the probability that a new intervention is cost-effective compared to the control condition. This probability can be estimated by determining what proportion of CE pairs is located in the cost-effective half of the CE plane (ie, to the right of the previously mentioned line with the slope equal to λ) (Fig. 2). Because it is generally unknown what decision makers are willing to pay for an additional unit of effect, λ is varied between its natural bounds (range: 0 to ∞) and the

FIGURE 3. Cost-effectiveness plane for a lifestyle intervention for construction workers at risk for cardiovascular disease compared to usual practice.[38] Abbreviations: ICER, incremental cost-effectiveness ratio; NE-Q, North East Quadrant; NW-Q, North West Quadrant; SE-Q, South East Quadrant; SW-Q, South West Quadrant.

FIGURE 4. Cost-effectiveness acceptability curve for a lifestyle intervention for construction workers at risk for cardiovascular disease compared to usual practice.[38] This cost-effectiveness acceptability curve corresponds with the cost-effectiveness plane in Figure 3 and indicates the probability of cost-effectiveness for different values of willingness-to-pay per kilogram body weight loss.

probability that the new intervention is cost-effective compared with the control condition is estimated for a range of λs. These values can then be plotted on CEACs that show the probability of cost-effectiveness (*y* axis) for various λs (*x* axis).[55–57] To illustrate, the CEAC of the example trial is provided in Figure 4.

This CEAC indicates that if decision makers are not willing to pay anything to obtain an additional kilogram body weight loss (ie, λ = 0), there is a 0.33 probability that the new intervention is cost-effective compared to the control condition. If decision makers are

---

BOX 4. Calculation and Interpretation of the Cost–Benefit Estimates of a Lifestyle Intervention for Construction Workers at Risk for Cardiovascular Disease in Comparison to Usual Practice[38]

Mean intervention costs per participant were €605. During follow-up, average absenteeism costs per participant were €3302 in the intervention group and €3604 in the control group. Thus, the absenteeism benefits per participant were €302 (€3604 − €3302). Using this information, cost–benefit estimates can be calculated as follows:

NB : €302 − €605 = € − 303
BCR : €302/€605 = 0.50
ROI : ((€302/€605)/€302) ∗ 100 = −50%

These cost–benefit estimates indicate that the intervention resulted in a net loss to the employer of €303. Also, per Euro invested, the employer gained €0.50 and suffered a loss of 50%. Thus, the intervention cannot be regarded as cost-beneficial in terms of absenteeism costs.

---

willing to pay €2000 (ie, λ = 2000), this probability is 0.95. When interpreting CEACs, two approaches can be used by decision makers. If their willingness to pay is known, they have to judge whether the probability of cost-effectiveness at this ceiling ratio is acceptable. If their willingness to pay is unknown, they should consider whether the ceiling ratio at an acceptable probability of cost-effectiveness is acceptable to them. The latter might depend on the scale of the outcome measure and the prevalence of the condition under study.

CBA

In health economics and business administration, various measures exist for comparing costs and benefits. Of them, the net benefits (NBs), benefit cost ratio (BCR), and ROI are the most frequently used measures in occupational health research and can be estimated using the following equations[6]:

$$NB = Benefits − Costs$$

$$BCR = Benefits/Costs$$

$$ROI = (Benefits − Costs)/Costs[∗100]$$

where Costs are defined as intervention costs and Benefits as the difference in monetized outcomes between the intervention group and the control group (eg, difference in productivity costs). Benefits are estimated by subtracting the mean expenses incurred by the intervention group participants from those of the control group. Hereby, positive benefits indicate reduced spending. The NB indicates the amount of money gained after costs are recovered (ie, net loss or net savings). The BCR indicates the amount of money returned per monetary unit invested. The ROI indicates the percentage of profit per monetary unit invested.[58,59] Interventions can be regarded as cost saving if the following criteria are met: NB > 0, BCR > 1, and ROI > 0%. To illustrate, a description of the calculation and interpretation of the example trial's cost–benefit estimates are provided in Box 4.

Cost–benefit estimates, and BCRs and ROIs in particular, are typically presented without an indication of their uncertainty. If uncertainty is substantial and this is not taken into account, wrong conclusions could be drawn. Therefore, we recommend the use of the previously described bootstrap method (Box 2) to estimate the uncertainty surrounding cost–benefit estimates. In this case, the NB, BCR, and/or ROI are calculated per bootstrap sample. Subsequently, 95% CIs can be estimated using the bias corrected and accelerated

method.[51,53] Although BCRs and ROIs are ratio measures, estimating their 95% CIs is straightforward as the denominator (ie, intervention costs) is typically positive. Many occupational health decision makers, however, may lack the necessary statistical background to interpret 95% CIs.[11] A possible way to deal with this issue is to estimate the proportion of NBs, BCRs, and/or ROIs that indicate cost savings (ie, "the probability of financial return"). Occupational health decision makers can subsequently use this information to consider whether the established probability of financial return is acceptable to them.

When reporting CBA results, economists and policymakers prefer the NB, whereas the BCR and ROI are more familiar to business managers. As such, it is recommendable to report at least two of them (ie, NB and BCR/ROI), so that the results can be easily interpreted by all stakeholders. Another advantage of this approach is that it makes the results easily comparable with those of other studies, because different metrics are used in the literature to estimate whether OHS interventions generate cost savings.[6]

Sensitivity Analysis

Economic evaluations are often conducted in the context of incomplete information and uncertainty, which necessitates the use of proxy measures, and invariably, the need to make assumption about the methods and unit prices used for valuing resource use, the methods used for dealing with incomplete data, and the way in which adjustments are made for differential timing.[4,8] Therefore, sensitivity analyses should be undertaken to assess how study results would change for different key assumptions and parameter values (ie, the robustness of study results).[17,60] The ranges of values tested, and arguments for selecting these ranges, must be clearly described.[10,17] Various approaches to sensitivity analyses exist, including one-way, multiway, and probabilistic sensitivity analysis. One-way sensitivity analyses assess the impact of changes to a single parameter at a time, while multiple parameters are varied simultaneously in multiway sensitivity analyses.[61] These methods may indicate parameter values for which results could change, but do not provide an indication of the combined impact of the uncertainty surrounding these parameters.[60] The latter could be modeled using probabilistic sensitivity analyses.[62]

CONCLUDING REMARKS

Resources for occupational health are scarce. This makes it necessary for decision makers to have information on the relative efficiency of OHS interventions to allocate available resources to their best use. As such, economic evaluations of OHS interventions are becoming increasingly important, many of which are conducted alongside effectiveness trials. Trial-based economic evaluations provide a unique opportunity to reliably estimate the resource implications of OHS interventions at low incremental cost.[10,14] Nevertheless, it is critical that high-quality trial-based economic evaluations are performed when this information is used to inform allocation decisions.

Designing a high-quality trial-based economic evaluation requires close collaboration between occupational health specialists, individuals executing the trial, and health economists.[14] Careful considerations must be made regarding the perspective, the analytic time frame, the identification, measurement, and valuation of resource use and outcomes, as well as the methods used for calculating sample sizes, comparing costs and consequences, and handling missing data and uncertainty. The latter is of particular importance, as few economic evaluations in occupational health report on the uncertainty surrounding their incremental cost-consequence estimates.[4–7,15] Failing to estimate values under uncertainty makes it impossible to determine the certainty of results and could thus lead to inappropriate decision making. To quantify precision, nonparametric bootstrapping can be used as a statistical technique

for dealing with the right skewed nature of cost data.[1,7] An overview of our core recommendations for trial-based economic evaluations in occupational health can be found in the Appendix.

Trial-based economic evaluations may also have shortcomings, including limited sample sizes, limited comparators, and truncated time horizons.[14] To deal with the latter, researchers might consider extrapolating economic evaluation results beyond the follow-up of a trial by using decision analytic modeling, in which expected costs and consequences between alternatives are compared by synthesizing information from multiple sources (eg, scientific literature, study results).[1,13,14] For more detailed information about decision analytic modeling, we refer to other publications.[14,63] Also, even though we recommend a pragmatic (cluster-)RCT design for economic evaluations, we are aware that randomization itself may not always be feasible and/or desired in the occupational health setting. In those cases, well-executed nonrandomized studies may provide valuable information, but it is critical that efforts be made to control for selection bias (eg, by using propensity score matching).[64,65]

When interpreting economic evaluations of OHS interventions, it is important to bear in mind that their results may not be directly applicable to other countries and jurisdictions due to differences in health care, social security systems, and other factors. Verbeek et al[66] demonstrated that economic evaluation results can be generalized from one country to another. Nevertheless, to enable the necessary calculations, researchers need to provide an extensive description of the intervention, a detailed list of resource use as well as information of the health care system in the original study and the allocation of costs to various stakeholders.[66]

By simultaneously providing recommendations for good practice in the economic evaluation of OHS interventions and discussing the methods and principles that underlie them, this study aimed to help researchers in conducting and reporting high-quality trial-based economic evaluations. Such studies are expected to contribute to the development of a sound evidence base on the resource implications of OHS interventions,[3,4] which is a necessary prerequisite for evidence-based practices occurring in occupational health.[11] The present article may also be helpful to consumers of this literature with understanding and critically appraising trial-based economic evaluations of OHS interventions, which might help improve the uptake of their results.

REFERENCES

1. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the Economic Evaluation of Health Care Programmes*. New York: Oxford University Press; 2005.
2. Burdorf A. Economic evaluation in occupational health—its goals, challenges, and opportunities. *Scand J Work Environ Health*. 2007;33:161–164.
3. 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;36:313–318.
4. Tompa E, Dolinschi R, de Oliveira C. Practice and potential of economic evaluation of workplace-based interventions for occupational health and safety. *J Occup Rehabil*. 2006;16:367–392.
5. Uegaki K, de Bruijne MC, Lambeek L, et al. Economic evaluations of occupational health interventions from a corporate perspective—a systematic review of methodological quality. *Scand J Work Environ Health*. 2010;36:273–288.
6. van Dongen JM, Proper KI, van Wier MF, et al. Systematic review on the financial return of worksite health promotion programmes aimed at improving nutrition and/or increasing physical activity. *Obes Rev*. 2011;12:1031–1049.
7. van Dongen JM, Proper KI, van Wier MF, et al. A systematic review of the cost-effectiveness of worksite physical activity and/or nutrition programs. *Scand J Work Environ Health*. 2012;38:393-408.

         © 2014 American College of Occupational and Environmental Medicine

8. Tompa E, Culyer AJ, Dolinschi J. *Economic Evaluation of Interventions for Occupational Health and Safety: Developing Good Practice*. New York: Oxford University Press; 2008.

9. Uegaki K, de Bruijne MC, Anema JR, et al. Consensus-based findings and recommendations for estimating the costs of health-related productivity loss from a company's perspective. *Scand J Work Environ Health.* 2007;33:122–130.

10. Korthals-de Bos I, van Tulder M, van Dieten H, Bouter L. Economic evaluations and randomized trials in spinal disorders: principles and methods. *Spine.* 2004;29:442–448.

11. van Dongen JM, Tompa E, Clune LA, et al. Bridging the gap between the economic evaluation literature and daily practice in occupation health: a qualitative study among decision-makers in the healthcare sector. *Implement Sci.* 2013;8:57.

12. Glick HA, Doshi JA, Sonnad SS, Polsky D. *Economic Evaluations in Clinical Trials*. New York: Oxford University Press; 2007.

13. Ramsey S, Willke R, Briggs A, et al. Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA Task Force Report. *Value Health.* 2005;8:521–533.

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

15. Verbeek J, Pulliainen M, Kankaanpaa E. A systematic review of occupational safety and health business cases. *Scand J Work Environ Health.* 2009;35:403–412.

16. Downey AM, Sharp DJ. Why do managers allocate resources to workplace health promotion programmes in countries with national health coverage? *Health Promot Int.* 2007;22:102–111.

17. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the *BMJ. BMJ.* 1996;313:275–283.

18. 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;53:688–695.

19. Drummond M, Sculpher M. Common methodological flaws in economic evaluations. *Med Care.* 2005;43(suppl):5–14.

20. Frick FD. Microcosting quantity data collection methods. *Med Care.* 2009;47:S76–S81.

21. Hakkaart-van Roijen L, Tan SS, Bouwmans CAM. *Handleiding voor Kostenonderzoek. Methoden en Standaardkostprijzen voor Economische Evaluaties in de Gezondheidszorg*. Diemen: College voor zorgverzekeringen; 2010.

22. Zhang W, Bansback N, Anis AH. Measuring and valuing productivity loss due to poor health: a critical review. *Soc Sci Med.* 2011;72:185–192.

23. Krol M, Brouwer WBF, Rutten FFH. Productivity costs in economic evaluations: past, present, future. *Pharmacoeconomics.* 2013;31:537–549.

24. Koopmanschap MA, Rutten FFH. Indirect costs in economic studies: confronting the confusion. *PharmacoEconomics.* 1993;4:446–454.

25. 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.

26. Kessler RC, Barber C, Beck A, Berglund P, Cleary PD, McKenas D. The World Health Organization health and work performance questionnaire (HPQ). *J Occup Environ Med.* 2003;45:156–174.

27. 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.

28. Koopmanschap MA. PRODISQ: a modular questionnaire on productivity and disease for economic evaluation studies. *Expert Rev Pharmacoecon Outcomes Res.* 2005;5:23–28.

29. Van Roijen L, Essink-bot ML, Koopmanschap MA, Bonsel G, Rutten FFH. Labor and health status in economic evaluation of health care: The Health and Labor Questionnaire. *Int J Technol Assess Health Care.* 1996;12:405–415.

30. Wahlqvist P, Carlsson J, Stalhammar NO, Wiklund I. Validity of a Work Productivity and Activity Impairment Questionnaire for Patients with Symptoms of Gastro-Esophageal Reflux Disease (WPAI-GERD)—results from a cross-sectional study. *Value Health.* 2002;5:106–113.

31. Reilly MC, Bracco A, Ricci JF, Santoro J, Stevens T. The validity and accuracy of the Work Productivity and Activity Impairment Questionnaire—Irritable Bowel Syndrome Version (WPAI:IBS). *Aliment Pharmacol Ther.* 2004;20:459–467.

32. Prasad M, Wahlqvist P, Shikiar R, Shih YT. A review of self-report instruments measuring health-related work productivity. *Pharmacoeconomics.* 2004;22:225–244.

33. NICE. Guide to the methods of technology appraisal. Available at: http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf. Accessed January 11, 2013.

34. Hamberg-van Reenen HH, Proper KI, van den Berg M. Worksite mental health interventions: a systematic review of economic evaluations. *Occup Environ Med.* 2012;69:837–845.

35. Brazier J, Deverill M, Green C, Harper R, Booth A. A review of the use of health status measures in economic evaluation. *J Health Serv Res Policy.* 1999;4:174–184.

36. Uegaki K. *Economic Evaluation of Interventions for Occupational Health*. Amsterdam: Vrije Universiteit Amsterdam; 2010.

37. Eysink PED, Hamberg-van Reenen HH, van Gool CH, Hoeymans N, Burdorf A. *Meten van verloren arbeidsjaren door ziekte: Disease-Adjusted Working Years (DAWYs): Verkenning van een Nieuwe Maat*. Bilthoven: RIVM; 2010.

38. 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;53:610–617.

39. Briggs AH, Gray AM. Power and sample size calculations for stochastic cost-effectiveness analysis. *Med Decis Making.* 1998;18:S81-S92.

40. Gafni A, Walter SD, Birch S, Sendi P. An opportunity cost approach to sample size calculation in cost-effectiveness analysis. *Health Econ.* 2008;17:99–107.

41. Gardiner JC, Sirbu CM, Rahbar MH. Update on statistical power and sample size assessments for cost-effectiveness studies. *Expert Rev Pharmacoeconomics Outcomes Res.* 2004;4:89–98.

42. Al MJ, Van Hout BA, Michel BC, Rutten FFH. Sample size calculation in economic evaluations. *Health Econ.* 1998;7:327–335.

43. Andrew B. Economic evaluation and clinical trials: size matters. *BMJ.* 2000;321:1362–1363.

44. Goossens M, Evers S, Vlaeyen J, Rutten-van Mölken M, van der Linden S. Principles of economic evaluation for interventions of chronic musculoskeletal pain. *Eur J Pain.* 1999;3:343–353.

45. Moher D, Hopewell S, Schultz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ.* 2010;340:c869.

46. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ.* 2009;338:b2393.

47. Noble SM, Hollingworth W, Tilling K. Missing data in trial-based cost-effectiveness analysis: the current state of play. *Health Econ.* 2012;21:187–200.

48. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Statist Med.* 2011;30:377–399.

49. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons; 1987.

50. Simon GT, Julie AB. How should cost data in pragmatic randomised trials be analysed? *BMJ.* 2000;320:1197–1200.

51. Barber JA, Thompson SG. Analysis of cost data in randomized trials: an application of the non-parametric bootstrap. *Statist Med.* 2000;19:3219–3236.

52. Chaudhary MA, Stearns SC. Estimating confidence intervals for cost-effectiveness ratios: an example from a randomized trial. *Statist Med.* 1996;15:1447–1458.

53. Kelley K. The effects of nonnormal distributions on confidence intervals around the standardized mean difference: bootstrap and parametric confidence intervals. *Educ Psychol Meas.* 2005;65:51–69.

54. Black WC. The CE plane: a graphic representation of cost-effectiveness. *Med Decis Making.* 1990;10:212–214.

55. Briggs AH, O'Brien BJ, Blackhouse G. Thinking outside the box: recent advances in the analysis and presentation of uncertainty in cost-effectiveness studies. *Ann Rev Public Health.* 2002;23:377–401.

56. Fenwick E, Marshall D, Levy A, Nichol G. Using and interpreting cost-effectiveness acceptability curves: an example using data from a trial of management strategies for atrial fibrillation. *BMC Health Serv Res.* 2006;6:52.

57. Fenwick E, O'Brien BJ, Briggs A. Cost-effectiveness acceptability curves—facts, fallacies and frequently asked questions. *Health Econ.* 2004;13:405–415.

58. Phillips JJ. *Return on Investment in Training and Performance Improvement Programs*. Burlington: Elsevier; 2003.

59. Stone PW. Return-on-investment models. *Appl Nurs Res.* 2005;18:186–189.

60. Giffin SC. *Dealing With Uncertainty in the Economic Evaluation of Health care Technologies*. PhD Dissertation; 2010.

61. Briggs A, Sculpher M, Buxton M. Uncertainty in the economic evaluation of healthcare technologies: the role of sensitivity analysis. *Health Econ.* 1994;3:95–104.

62. Claxton K, Sculpher M, McCabe C, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. *Health Econ.* 2005;14:339–347.

63. Briggs A, Claxton K, Sculpher M. *Decision Modelling for Health Economic Evaluation.* New York: Oxford University Press; 2006.

64. Dehejia RH, Wahba S. Propensity score-matching methods for nonexperimental causal studies. *Rev Econ Stat.* 2002;84:151–161.

65. Caliendo M, Kopeinig S. Some practical guidance for the implementation of propensity score matching. *J Econ Surv.* 2008;22:31–72.

66. 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;36:305–312.

Appendix
Core Recommendations for Trial-Based Economic Evaluation in Occupational Health

DESIGN OF AN ECONOMIC EVALUATION

Kinds of Economic Evaluations

Perform various kinds of economic evaluations to inform all relevant stakeholders, for example, cost-effectiveness analysis, cost-benefit analysis, cost-utility analysis.

Timing

Consider economic evaluation requirements during an early phase of the design of a trial.

Trial Design

If possible, use randomization to allocate participants to study arms (ie, (cluster-)RCTs).

Trial conditions should resemble daily practice as much as possible.

Perspective

Apply various perspectives to inform all relevant stakeholders. The applied perspective(s) should be explicitly stated.

Analytic Time Frame

Ideally, the analytic time frame covers the entire period over which costs and consequences flow from the alternatives under study.

Identification, Measurement, and Valuation of Costs

Collect all resources that may influence the overall costs related to the applied perspective(s).

Appropriate unit prices may vary between perspectives. Researchers should, therefore, ensure that unit prices reflect the true resource implications to the decision maker(s) at hand.

Report aggregate costs, disaggregate resource use, and applied unit prices separately.

ANALYSIS OF AN ECONOMIC EVALUATION

Sample Size

Ideally, economic outcomes are used in the sample-size calculation of a trial. If this is not possible, use estimation rather than hypothesis testing.

Adjusting for Differential Timing

Prices drawn from different years should be adjusted for inflation using consumer prices indices and the applied reference year should be explicitly stated.

Costs and consequences collected over a period of more than one year should be discounted using discount rates pertaining to the jurisdiction in which the economic evaluation is performed to adjust for time preferences of individuals.

Missing Data

Use multiple imputation to impute missing values, particularly if 5% of data or more are missing.

Incremental Analysis of Costs and Consequences

Incremental costs and consequences should be reported as differences in arithmetic means.

Use nonparametric bootstrapping to quantify precision of cost data.

Comparing Incremental Costs and Consequences

The preferred method for comparing incremental costs and consequences depends on the kind of economic evaluation, that is, ICERs for cost-effectiveness analyses/cost-utility analyses, and NBs, BCRs, and/or Return on Investments for cost-benefit analyses.

To quantify the uncertainty surrounding incremental cost-consequence estimates, use nonparametric bootstrapping techniques.

Use cost-effectiveness planes to graphically illustrate the uncertainty surrounding ICERs and cost-effectiveness acceptability curves to provide a summary measure of the joint uncertainty of costs and effects/utilities. For cost–benefit estimates, use 95% confidence intervals and/or the probability of financial return.

Sensitivity Analysis

Perform sensitivity analyses to test the robustness of results.

The ranges of values tested, and arguments for selecting these ranges, should be described.