Available online at www.sciencedirect.com

ScienceDirect

International Journal of Project Management 37 (2019) 549–564

International Journal of Project Management

www.elsevier.com/locate/ijproman

Assessing public projects' value for money: An empirical study of the usefulness of cost–benefit analyses in decision-making

Gro Holst Volden

Concept Research Program, Norwegian University of Science and Technology, 7491 Trondheim, Norway

Received 3 June 2018; received in revised form 3 February 2019; accepted 4 February 2019
Available online xxxx

Abstract

Value for money, as measured by cost–benefit analyses (CBAs), is a crucial part of the business case for major public investment projects. However, the literature points to a range of challenges and weaknesses in CBAs that may cause their degree of usefulness in decision-making to be limited. The paper presents an empirical study of CBA practice in Norway, a country that has made considerable efforts to promote quality and accountability in CBAs of public projects. The research method is qualitative, based on a case study of 58 projects. The results indicate that the studied CBAs are largely of acceptable quality and heeded by decision-makers. Appraisal optimism seems to have been reduced by the introduction of external quality assurance of CBAs. However, there is need for a more consistent assessment of the non-monetized benefits, and distinguishing them from other decision perspectives such as the achievement of political goals. The paper offers a set of practical recommendations to increase CBA usefulness further.

Keywords: Project value; Project appraisal and evaluation; Cost–benefit analysis; Business case

1. Introduction

1.1. Projects ought to be good value for money

The project management community has increasingly shifted its attention beyond the 'iron triangle' of cost, time, and quality, to take a wider, strategic view of projects. Projects are implemented to deliver benefits and create value for users, the parent organization, and/or society at large (Morris, 2013; Samset and Volden, 2012). Accordingly, project governance has become an important issue in project research and practice. It refers to the processes, systems, and regulations that the financing party must have in place to ensure that relevant and viable projects are chosen and delivered efficiently (Müller, 2009; Volden and Samset, 2017b).

Williams and Samset (2010) refer to the choice of project concept as the most important decision that project owners make. The choice of concept ought to be approved on the basis of a business case, in which the expected benefits and strategic outcomes are described (Jenner, 2015). The business case provides a rationale for the preferred solution, and is therefore crucial for future benefits and cost management (Musawir et al., 2017; Serra and Kunc, 2015).

This paper focuses on the cost–benefit analysis (CBA) which is often a crucial part of the business case. The CBA concerns the relationship between resources invested and the benefits that can be achieved and is a tool to determine the project's value for money (i.e. whether it is profitable for society). Specifically, the aim of a CBA is to compute the net present value (NPV) of a project or various project alternatives, as defined by Eq. (1):

$$NPV = \sum_{t=0}^{N} \frac{B_t - C_t}{(1+i)^t} \quad (1)$$

where B is social benefit, C is social cost, i is social discount rate, t is time, and N is the period of analysis. It can be used to rank

E-mail address: gro.holst.volden@ntnu.no.

https://doi.org/10.1016/j.ijproman.2019.02.007
0263-7863/00

projects unambiguously (Boardman et al., 2011). The decision rule is to adopt a project if the NPV is positive, or in the case of several alternatives, to select the project with the highest NPV. Alternative criteria such as the benefit–cost ratio or internal rate of return can be applied too, but the NPV is normally recommended as a metric.

The CBA is particularly relevant for state-funded projects, as they are regarded in an overall national perspective, rather than the perspective of particular agencies, regions, or stakeholder groups. The benefits are interpreted in terms of the affected people's willingness to pay for them, and the costs are defined by the value of the alternative uses of the resources (Boardman et al., 2011).

The aim of the CBA is to be comprehensive in terms of the coverage of a project's impacts (Sager, 2013), and to monetize them as far as possible. Various techniques have been developed to elicit the willingness to pay (WTP) for non-market goods. However, remaining impacts that cannot be monetized must be described and presented in other ways, to enable decisions to be made as to whether they will be likely to improve or depreciate the NPV. In some cases, if analysts are unable or unwilling to attribute a monetary value to key benefits, they may be forced to apply cost-effectiveness analyses. In such cases, the intention is to minimize a ratio involving the benefit in physical units and monetary costs (e.g. cost per life saved). Unlike the CBA, the cost-effectiveness analysis does not make it possible for the analyst to conclude that the given project will contribute to social welfare (Boardman et al., 2011). It is thus a subordinate or second-best measure of value for money. Additionally, various multicriteria analyses are sometimes used, but they are not measures of value for money. In this paper we focus on value for money as measured by the CBA and not on project analysis in general.

A number of authors have highlighted the value for money perspective and the CBA (e.g. Jenner, 2015; Laursen and Svejvig, 2016; Terlizzi et al., 2017). Governments and professional project management bodies all require assurance of value for money, such as the Association for Project Management (2018), the (former) Office of Government Commerce (2009), and the Project Management Institute (2017). Volden and Samset (2017a) studied project governance frameworks in six OECD countries, and found that all of the frameworks highlighted the CBA in the front-end of projects. This is a dominant method of appraisal in the transport sector, for which many countries have developed guidelines (HEATCO, 2006; Mackie et al., 2014). Similarly, highlighting the CBA in the front-end has been used to assess development aid projects for decades, and is referred to as one of the World Bank's signature issues (World Bank, 2010). The appraisal method is also increasingly used in other sectors.

1.2. The research gap

However, the attention paid to the quality and utility of CBAs is limited in project research. The broad but fragmented literature on CBAs, which discusses a number of challenges and weaknesses, is rarely cited in project management and project governance literature. This is surprising, as we would normally expect that the quality of an analysis affects the extent to which CBAs are used, their recommendations followed and social benefits realized. We claim that it is not sufficient to require a CBA to be performed, but that also its usefulness must be ensured as part of project governance frameworks. A number of studies have documented the limited impact of CBAs on political decisions (e.g. Annema, 2013; Eliasson et al., 2015; Nyborg, 1998). For example, a review of World Bank projects shows that CBAs are rarely mentioned in policy documents, and that the percentage of projects justified following CBAs is declining (World Bank, 2010).

The explanations given in the literature are multifaceted and involve both analytical and political issues. For example, the World Bank report notes that only 54% of CBAs were of acceptable quality, but also that high-quality CBAs were often disregarded by decision-makers (World Bank, 2010).

In this paper we focus on the analytical issues in terms of the weaknesses that materialize in CBA reports. Other authors have focused on issues such as adverse incentives at the decision-making level that may result in the value for money aspect being played down when decisions are made (e.g. Sager, 2016). Decision-making in a democratic setting is inherently complex, frequently unpredictable, and influenced by other decision logics than just the rational economic ones. Therefore, as noted by Samset and Volden (2015), the greatest potential for improvement might be to strengthen the analytical processes.

1.3. This study

The aim of this study is to increase knowledge about the quality and usefulness of CBAs as basis for project selection. We take the perspective of the financing party (the true owner) who, in the case of public projects, is the entire society and its taxpayers, as represented by the Cabinet.

We define seven research questions (RQs) that together cover the main weaknesses in CBAs that have been discussed in academic literature (cf. Section 2). We want to learn about the relative prevalence of these weaknesses and to consider to what extent they reduce the quality and usefulness of analyses. The seventh and last research question, about whether CBA recommendations are actually followed (RQ7), is therefore of particular interest, and we consider it in relation to the other six questions. The seven questions are as follows:

RQ1: Are the CBAs consistent across projects with respect to which impacts are included, whether a valuation has been performed, and parameters and assumptions applied?
RQ2: Are non-monetized impacts assessed and presented consistently?
RQ3: Are associated uncertainties identified and presented?
RQ4: Are distributional impacts presented as supplementary information?
RQ5: Are CBAs unbiased? Specifically, is there a difference between CBAs done by project promoters and CBAs done by an independent party?
RQ6: Is transparency and clarity acceptable in the reports?
RQ7: Do decision-makers follow the advice presented in the CBAs?

To answer these research questions, we apply high-quality empirical data from Norway. Since 2005, CBAs have been compulsory in appraisals of the country's largest public investment projects under the Ministry of Finance's Quality Assurance (QA) scheme. The scheme is presented and discussed by Volden and Samset (2017b).

The QA scheme applies to public infrastructure projects that exceed an estimated threshold cost of NOK 750 million (USD 100 million). In those projects, external quality assurance (QA) of decision documents is required before the Cabinet makes its choice of project concept. As a basis for the external QA, the sectoral ministry or agency prepares a *conceptual appraisal* (CA) document. The CA is the business case and must include an assessment of needs and overall requirements, a possibility study that results in at least three alternative project concepts, including the zero-investment alternative, and a CBA of these concepts. The QAs are performed by private consultants contracted by the Ministry of Finance. The QA team should review the CA and thereafter present its own independent CBA, with alternatives ranked on the basis of their estimated value for money. This implies that for each project there will be two value for money assessments, one produced by the initiating ministry or agency and the other by the external quality assurer.

The QA team includes economists who are experts on CBA. Additionally, the ministries and agencies use highly qualified people to prepare the CBAs. The CA-QA process takes place at the same stage in all projects' life cycle, namely the end of the pre-study phase. The Norwegian Ministry of Finance has issued guidelines with a set of overall requirements for CBAs that we consider to be in line with best practice internationally (Finansdepartementet, 2005, 2014).

We considered Norway an interesting research case because of the efforts made to ensure that CBAs are of high quality. According to Flyvbjerg's (2006) categorization of case study research, Norway is a 'critical case' (here understood as an assumed best case). Our findings should be relevant beyond the Norwegian context, our thinking being that CBA weaknesses observed in this country, with a project governance scheme that requires high-quality and quality assured CBAs, will most likely also be a problem in countries without such a scheme. That said, there may be cultural and other differences between countries that influence project practices. In a case study, we must always present reservations concerning transferability of results across countries.

In Section 2 we present a review of the literature on weaknesses in CBAs. The review forms the basis for the framework of analysis applied to study the case CBAs. The framework is presented in Section 3, and a description of the study data and methodology is provided in Section 4. In Section 5, we present and discuss the findings with respect to each research question. Lastly, in Sections 6 and 7 we present our conclusions and recommendations, and discuss possibilities for further work.

2. Literature review

Today it is widely recognized that not only programs and portfolios, but also individual projects, should be linked to higher-order goals and strategies. The project management community has been increasingly concerned with how projects create value and reap benefits (Shenhar et al., 2001; Zwikael and Smyrk, 2012; Morris, 2013; Breese et al., 2015; Hjelmbrekke et al., 2017). Whereas some authors focus on the front-end, others discuss benefits management throughout the project life-cycle (e.g., Serra and Kunc, 2015; Musawir et al., 2017).

However, this part of the project management literature is still young. As noted by Laursen and Svejvig (2016) the definitions of project benefits and value are sometimes vague and depend on the perspective chosen. Baccarini (1999) suggested a distinction between two levels of project success, i.e. project management success, which concerns delivery, and product success, which concerns the outcome. Samset (2003) suggested a triple-level performance test concerning project outputs, first-order effects for users, and long-term effects for society. A similar chain of benefits has been suggested by Zwikael and Smyrk (2012) and Serra and Kunc (2015) and is also largely in line with PRINCE2®.[1] In the framework suggested by Zwikael and Smyrk (2012) it is also specified who should be responsible for project success on each level. The project manager is responsible for success at the operational level (project management success), the project owner is responsible for success at the tactical level (project ownership success) and the funder is responsible for success at the strategic level (project investment success).

In this paper we focus on the highest level of project success (i.e., project investment success, in Zwikael and Smyrk's terminology) where benefits and costs are compared to determine the effective 'return' on the investment. The CBA takes an overall societal perspective where all benefits and costs to affected parties nation-wide ought to be included, and (to the extent possible) translated into the monetary amount that people are willing to exchange. This is not the only possible definition of project investment success (as discussed further in Section 2.1) but at least it provides a very clear definition.

The project management community has not devoted much attention to the specificities of the CBA thus far, and we therefore had to search for other types of literature. The 'CBA literature' is large, with publications in transport sector journals as well as journals in economics, public policy and other social sciences.

Many weaknesses and challenges have originated in both theory and practice regarding the use of CBAs, to the extent that decision-makers do not find them useful or trustworthy. Such weaknesses may remain undisclosed due to the complexity and often low transparency of the methodology. In the following subsections we synthesize the literature on the various weaknesses in CBAs, which may explain decision-makers' lack of confidence in this metric. The literature is fragmented in the sense that different authors focus on entirely different issues. However, we suggest the following categorization of the weaknesses in CBAs: (1) criticism of the CBAs' normative fundament, (2) discussion of various measurement problems, and (3) challenges relating to appraisal optimism.

[1] Projects IN Controlled Environments, see www.axelos.com.

2.1. The CBA – Its normative fundament

The CBA is a powerful project evaluation tool, primarily because it is not based on political preferences, and therefore it can be characterized as a 'neutral tool' (van Wee and Rietvold, 2013). However, this strength is also a weakness because the CBA only recognizes people's preferences in their role as consumers. By contrast, analysis of people's preferences in their role as citizens may give a different result (Mouter and Chorus, 2016), as may the use of either planners' preferences or decision-makers' preferences (Mackie et al., 2014). Thus, the CBA is a framework for measuring efficiency, not equity, alignment with political goals, or any other definition of social desirability. Inevitably, the use of WTP implies that more weight is attached to high-income groups than to low-income groups (Nyborg, 2014). Furthermore, by focusing on the *aggregate* WTP, the CBA disregards the fact that some groups may be *worse off* after project completion than they were previously. The use of aggregate WTP is justified by the Kaldor-Hicks efficiency criterion, according to which a new resource allocation would be an improvement for society if the winners could *hypothetically* compensate the losers and still be better off. However, there is no requirement for such compensation to be given (Nyborg, 2014).

Thus, the CBA is of little help in cases in which the public sector has clear policy objectives that differ from consumers' preferences. Nyborg (1998) found this an important reason why some Norwegian politicians did not trust the CBA, with politicians on the left of the political axis being most sceptical. Mouter (2017) has reported similar responses from Dutch politicians.

A related critique is that the CBA systematically downplays the welfare of future generations. Decision-makers are increasingly concerned with investments' sustainability (Eskerod and Huemann, 2013; Haavaldsen et al., 2014), which requires a more holistic and long-term perspective than taken in CBAs. In particular, the use of a discount rate in CBAs implies that impacts on future generations have low worth today, and this weakness has been criticized by a number of authors (e.g. Ackerman, 2004; Næss, 2006; Pearce et al., 2006).

Some researchers have suggested that the CBA should be replaced by some form of multicriteria analysis that is based on the preferences of planners or decision-makers, at least in cases with moral dimensions (Browne and Ryan, 2011, van Wee, 2013). Others have noted that a multicriteria analysis has weaknesses too, which makes it more subjective and manipulable (Dobes and Bennett, 2009). In our view, both types of analysis can supplement each other, as they measure different things. For all projects that either directly or indirectly aim to contribute to economic growth, the CBA should at least be partly relevant.

The solution to this weakness most often recommended by authors is that all the costs and benefits should be presented in a disaggregated and transparent form that shows how they are distributed, not just their aggregated effect. When relevant, a separate overview and discussion of significant distributional impacts, both within and between generations, should be provided in the report. In that way, decision-makers would be able to decide for themselves whether the distributional impacts are acceptable. The CBA could also be included more systematically in a broader project evaluation framework that includes other perspectives than efficiency, such as the Five Case Model in the UK, in which the economic case is one of the five cases (HM Treasury, 2013). Another framework, one that has been very influential in evaluations of development assistance projects, comprises the five OECD-DAC criteria of efficiency, effectiveness, impact, relevance, and sustainability (Samset, 2003). A variant of the latter framework has been applied in ex post evaluations of Norwegian projects (Volden, 2018).

2.2. Measurement problems

Even if the ethical and normative premises on which the CBA rests were accepted, the credibility and usefulness of the results might be low due to various measurement problems (Atkins et al., 2017). At an early stage, information about the effects of a project is sparse and depends on many assumptions (Samset and Volden, 2015). Thus, an early CBA will have many sources of error, such as omitted impacts, forecasting errors, and valuation errors. Several studies have indicated that cost estimates and demand forecasts are highly inaccurate (i.e. Flyvbjerg et al., 2003; Kelly et al., 2015; Nicolaisen and Driscoll, 2014; van Wee, 2007). For example, Nicolaisen and Driscoll (2014) reviewed 12 studies conducted within the transport sector in various countries and concluded that traffic forecasts were unreliable, largely due to weaknesses in the model specifications, combined with low transparency, which made it difficult for others to observe what had been done.

Prediction and valuation of non-market goods such as health, safety, and the environment are a particular challenge. Different studies have revealed very different estimates of people's willingness to pay for such goods: for example, research conducted for a recently published doctoral thesis revealed huge variation in the estimates of the value of a statistical life (Elvik, 2017). It should also be noted that valuation methods differ in what they measure. For example, while stated preference (SP) methods are designed to capture the *total value*, revealed preference (RP) methods estimate only *use values* (Boardman et al., 2011). In many cases, inferior approaches that violate the principle of consumer sovereignty are used, such as implicit valuation, whereby analysts use the government's WTP as a proxy for the population's WTP. As discussed by Sager (2013) and by Mouter and Chorus (2016), a related challenge is that the population's preferences may be unstable, and the difference between consumer values and political opinions may be blurred.

Thus, it is crucial that the uncertainty involved in estimation is not downplayed (Flyvbjerg et al., 2003). Additionally, transparency is crucial: Wachs (1989) recommends that all details of the models and parameters should be available to anyone who might wish to replicate, verify, or merely critique the uses of the technical procedures. This implies that the

findings must be presented in a disaggregated form and not only as a summary indicator (Nyborg, 1998; Næss, 2006).

A further challenge is that the CBA is normally based on a partial equilibrium model and only measures direct effects. This is acceptable as long as other markets are competitive, but following the publication of the SACTRA report in the UK (Standing Advisory Committee on Trunk Road Assessment, 1999), attention has been paid to market imperfections that may mean that the full benefits of a transport investment fail to be included in the CBA. Some authors have indicated that such wider economic benefits may be considerable (Venables, 2007; Vickerman, 2008), while others have noted that they may also be negative (Næss et al., 2017; Small, 1999). Given that these impacts are not included in the NPV, they must be identified, discussed, and potentially quantified separately.

More generally, some impacts are inherently difficult to quantify and monetize. In particular, environmental effects are often substantially underestimated or ignored in practice, despite being possible to measure in principle (Ackerman, 2004; Browne and Ryan, 2011; Kelly et al., 2015; Næss et al., 2017). CBA textbooks and guides make it clear that non-monetized impacts must be identified, described, and balanced against the NPV, yet few textbooks give specific guidance on how this should be done. In practice, the treatment of non-monetized impacts tends to be random or politically driven as noted by some authors (e.g. Ackerman, 2004; Mackie and Preston, 1998).

2.3. Appraisal optimism

The third and last weakness of CBAs is that they are inherently at risk of bias and manipulation. For example, Mouter (2017) interviewed decision-makers who said that they knew how easy it was to affect results by 'shifting the buttons in the model' (Mouter, 2017, p. 1134). As noted by Wachs (1989), planning is not just analytical, and 'the most effective planner is sometimes the one who can cloak advocacy in the guise of scientific or technical rationality' (Wachs, 1989, p. 477).

Mackie and Preston (1998) list 21 sources of error and bias in transport project appraisals and conclude that appraisal optimism is one of the most important sources. Empirically, it has been shown that not only are CBAs inaccurate, but also they are often biased on the optimistic side (Flyvbjerg et al., 2003; Kelly et al., 2015; Nicolaisen and Driscoll, 2014; van Wee, 2007; World Bank, 2010).

Significant research has focused on explaining leaders' and entrepreneurs' optimism bias as a feature inherent in human behaviour. Such people are self-confident and tend to exaggerate their own abilities and control over a situation. While some authors describe this behaviour as unconscious (e.g. Lovallo and Kahneman, 2003), others argue that the persistence of bias is intentional and driven by a persistent excess demand for project finance (e.g. Bertisen and Davis, 2008). The persistence of bias can also be explained in terms of a principal–agent problem (Eisenhardt, 1989), such as when project promoters, who themselves are not responsible for funding, compete for discretionary grants from a limited budget (Samset and Volden, 2015). However, it is difficult to find conclusive empirical evidence of manipulation, as noted by Andersen et al. (2016).

A common recommendation to avoid appraisal optimism, whether or not it is intentional, is to ensure an outside view (Flyvbjerg, 2009; Lovallo and Kahneman, 2003; Mackie and Preston, 1998). This can be done by, for example, applying historical data (e.g. through reference class forecasting) and/or by having an independent third party perform or review the CBA. Additionally, systematic ex post evaluations should be performed to learn about the costs and benefits that can be expected (Flyvbjerg et al., 2003; Mackie and Preston, 1998; Volden, 2018).

Additionally, incentives for true speech must be in place. In this respect, Flyvbjerg et al. (2003) and Samset and Volden (2015) all recommend that project promoters are made accountable for financing, risk, and benefits realization, and that the appraisals are transparent and open to scrutiny. Mouter (2017) points out that the CBA is often complex and lacks transparency, which makes it particularly difficult to discover manipulation. More generally, an overall project governance framework that takes the risk of front-end agency problems into account should be in place.

3. Conceptual framework

We argue that the three strands of literature discussed in the previous section give rise to three broad explanations for why CBAs may not be considered useful by decision-makers. A simple conceptual framework is presented in Fig. 1.We have chosen 'CBA usefulness' as the main outcome variable. It is a multifaceted term that, in meaning, partly overlaps other terms such as trustworthiness, validity, and credibility (see Patton, 1999, and Scriven, 2015, for a discussion of criteria of merit by which analyses and evaluations ought to be evaluated). Since the CBA is specifically intended for decision support, CBA usefulness is considered from decision-makers' perspective. To some extent, the assessment of CBA usefulness will be subjective and depend on each decision-maker's preferences, competencies, and other abilities, but our focus is on assessments with which most decision-makers are likely to agree.

In line with the three categories of weaknesses of CBAs presented in Sections 2.1–2.3, we argue that CBA usefulness is threatened when (1) the analysis is too narrow in terms of relevant aspects being included in the business case (only the CBA alone), (2) the analysis is inconsistent, incomplete, and uncertainties are underestimated, and (3) the analysis is biased on the part of the analyst. By contrast, CBA usefulness is high when these weaknesses are not present.

The next step is to develop a framework for the empirical analysis, based on the conceptual framework in Fig. 1. In practice, the relative significance of the weaknesses in CBAs is largely unknown, as is the extent to which CBAs adhere to the recommendations provided in the literature to avoid or mitigate the weaknesses. To date, few empirical studies have

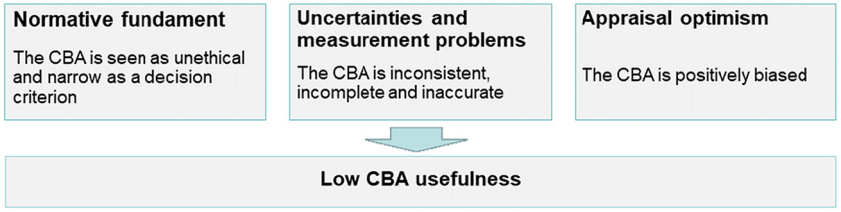

Fig. 1. Three types of weaknesses that lead to low CBA usefulness – a simple conceptual framework.

systematically reviewed CBA reports with respect to their overall relevance, quality, and credibility. This raises the question of whether it is possible for governments, through guidelines, quality standards, and other governance mechanisms, to ensure that CBAs are of high quality and useful to decision-makers. An interesting case is a recent study of the quality of CBAs of public projects in the UK (Atkins et al., 2017), in which the authors mainly focus on the second and third categories of weaknesses discussed above. The UK has taken steps to improve project competencies in central government and has introduced various governance arrangements to improve project performance (Volden and Samset, 2017a). Atkins et al. (2017) find that the CBAs are largely of acceptable quality, but that some challenges remain, the most important of which concern the lack of consistency across projects, and poor transparency and communication. They are also concerned about possible bias in the cost estimates, especially in cases in which decisions have been based on early estimates.

We draw on the most essential recommendations provided in literature, which, if adhered to, could increase CBA usefulness. Authors who criticize the normative foundations of the CBA (cf. Section 2.1) typically recommend that value for money assessments are supplemented by analyses of the project's impact on, for example, equity and sustainability. Those who discuss measurement problems (cf. Section 2.2) recommend a certain level of standardization, proper treatment of non-monerized impacts and uncertainty analyses. Lastly, those who are worried about appraisal optimism (cf. Section 2.3) recommend an outside view, and measures to ensure accountability. Common to all of the aforementioned three groups of authors is that they recommend transparent CBAs. Fig. 2 shows our framework for the empirical analysis, including the seven research questions presented in Section 1.

The use of the CBA in practice, understood as adherence to its recommendations, is a relevant indicator of CBA usefulness and is applied in this study (RQ7). We expect, ceteris paribus, that a CBA is more often adhered to when it is of high quality. However, it should be noted that adherence is not a perfect indicator of usefulness. As noted by Scriven (2015), there may be a number of reasons for lack of adherence to the results of a high-quality analysis. A thorough treatment of these issues would lead us beyond the analytical process and into political decision-making. Hence, for the purpose of this study, we merely assume that an instrumental decision logic or the 'rational ideal' is applied on the part of decision-makers (Samset and Christensen, 2017) and therefore disregard problems on the decision-making level, such as self-interest, the practice of 'horse trading', positioning, and power.

The final step in the outcome chain would be 'realized value for money'. This, too, would be an interesting indicator (although similar caution is required). Unfortunately, we do not have access to ex post data, and therefore this is not a topic of the empirical study.

4. Methodology

The empirical part of this study is largely qualitative, with the purpose of exploring, describing, and evaluating CBA practice within the Norwegian QA scheme. It is a multiple-case study of 58 Norwegian projects, based on a document review, interviews, and a review of the decisions made by the Cabinet. Although we refer to the cases as 'projects', all of the investments are studied in their early phases, in which they

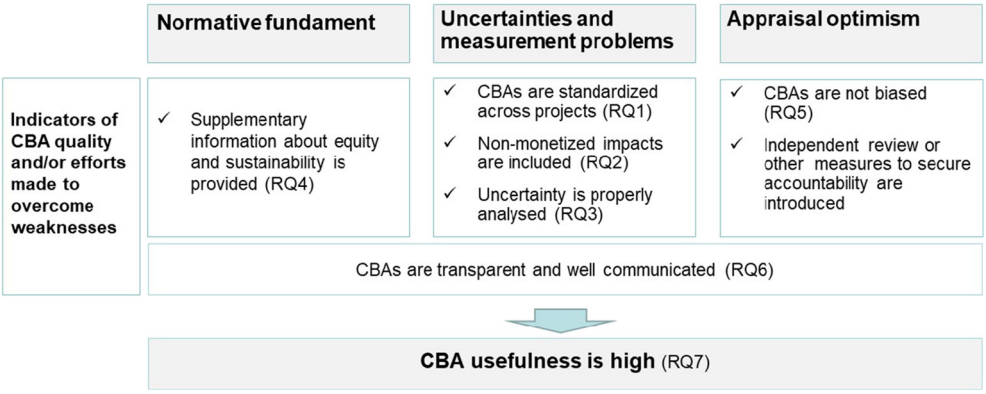

Fig. 2. Framework of analysis.

exist only conceptually. The Cabinet might chose the zero-investment alternative, in which case the project proposal will be rejected. Since very few of the projects have been completed, no information is available that can be used to determine the accuracy of the CBAs.

It should be noted that although the main unit of analysis is the project, we present some of the findings at 'CBA report' level (since most of the projects have two CBAs). At a higher level, one could consider Norway as a case, since all of the projects have been through CBAs in their front-end phase as part of the Norwegian QA scheme. However, this study is not an evaluation of the scheme but rather a study of CBA practice in a relatively large number of case projects, all of which belong to this (supposedly) favourable context.

The seven research questions listed in Section 1 were disaggregated into 25 subquestions that were more specific and contained indicators for the review of documents, as shown in Table 1. Some subquestions may contribute to answering more than one research question (RQ). However, the analysis was also inductive and open for exploring and describing other patterns and relationships that were revealed in the process.

Our main data source was the CA and QA reports for the 58 projects, which constituted the total population of projects that underwent CA and QA in the period 2005–2014, and are thus representative of projects in all of the major sectors that undergo QA in Norway. Currently, the transport sector has the largest number of projects, with most QAs performed on road projects. Other major categories are building construction, defence, and ICT projects.[2] The projects varied in size, complexity, purpose, and stakeholders involved, but in general they were the largest state-funded infrastructure projects in Norway in the period (Table 2).

For five of the projects (three of them within defence), the CA document was exempt from public access. For these projects, we only had access to the QA report and the presentation of the CA results discussed therein. Thus, we had access to a total of 111 CBA reports for our 58 projects.

The CA-QA process is followed by an administrative and political process in government. We established the status of all projects as of 2016, after the choice of project concept had been made by the Cabinet. To do this, we conducted a broad investigation of government documents, with particular focus on White Papers, to establish Parliament's ultimate choice of concept.

Additionally, we held semi-structured interviews with 26 key informants, all of whom were highly experienced within the field of CBA and had been involved in one or more of the studied projects. We considered that the interviews provided us with a deeper understanding, and since they were conducted after the document reviews, we were able to present some key findings and ask the interviewees for comments on them. Ten interviewees were senior ministry officers who commissioned CBAs from agencies, consultants, and quality assurers. They represented the decision-making level in this context. The other 16 interviewees were experts from the agencies and the QA teams and represented the persons who conducted the analyses. The interview guides were structured around the seven research questions, and the interviewees were invited to talk freely, based on their own experiences. It should be noted that the data collected from the interviews did not concern particular

Table 1
Subquestions applied for the review of CBAs.

RQ		Subquestion
RQ1	1	Describe the impacts included.
RQ1, RQ2	2	How are impacts treated (especially monetized or not).
RQ1, RQ6	3	Key assumptions and parameter values used to estimate the NPV (according to a pre-established list).
RQ1, RQ5	4	What is QAs reaction to CBA structure in CA? describe deviations between the two CBAs.
RQ2	5	Analyst's interpretation of the non-monetized impacts ('economic effect' or other).
RQ2	6	Methodology and measurement scale used to assess non-monetized impacts.
RQ2	7	Comprehensive analysis of non-monetized impacts? (pages used in the report)
RQ2	8	Comprehensive analysis of non-monetized impacts? (researcher's judgement)
RQ2	9	Non-monetized impacts – whose judgement? (e.g. experts, stakeholders, decision-maker).
RQ3	10	Type of risk analysis conducted, if any.
RQ3	11	Comprehensive risk analysis (researcher's judgement)? Capital cost, benefits, non-monetized separately.
RQ4	12	Distributional impacts or other considerations included along with the CBA.
RQ4	13	Comprehensive distributional analysis (researcher's judgement)?
RQ4	14	Distributional/other decision criteria clearly separated from CBA (researcher's judgement)?
All	15	Are the CA and QA in agreement on the recommendation?
RQ4, RQ5	16	Sign (and value?) of NPV of recommended alternative.
RQ4, RQ5	17	Is the recommended alternative the one with highest NPV?
RQ5	18	Is the zero option recommended?
RQ6	19	Overall level of transparency (researcher's judgement).
RQ6	20	Are models used to simulate impacts?
RQ6	21	If so, are the models explained? (reference to manuals, model version, etc.)
RQ6	22	Does the report include an executive summary?
RQ6	23	Is the report written in a non-technical language? (researcher's judgement)
RQ7	24	Status of the project as of today.
RQ5, RQ7	25	Whose advice is followed, CA or QA?

Table 2
Projects included in the research.

Projects included (sector)	N = 58
Road	20
Railway	5
Other transport (sea, coast, mixed)	11
Building	8
Defence	5
ICT	4
Sports event	3
Other	2

[2] Some sectors are exempt from the Ministry of Finance's scheme, but have their own, similar schemes, such as the energy and petroleum sector, and the hospital sector. These are not included in the study.

projects, but rather the general practice in central government. Each interview lasted 1–2 h.

A large Excel spreadsheet was applied, in which facts, assessments, notes from the document reviews as well as the interview transcripts were combined in the coding process. A list of the most interesting topics, counts, and possible relationships was continuously revised as we went through the material. The resulting themes and categories were not too different from the initial ones. The findings also included a number of categorizations, counting of occurrences, and cross-tabulation. In particular, the responses to subquestions 15 and 25, about whether the QA approved the CA and whose advice was followed by the decision-makers, were compared with various quality indicators. The results were also cross-tabulated against background variables such as project type.

All of the steps in the coding process gave considerable room for the researcher's own judgement, which might give rise to concerns about subjectivity and potential bias in our results. An important mechanism used to secure reliability and validity was the consultation of reliable sources of information. We used high quality, publicly available documents, as well as interviewees who had first-hand experience of CBA practice. The interviews were transcribed and the interviewees were subsequently given the opportunity to read and comment on the transcription. Furthermore, the use of different sources (i.e. document reviews *and* interviews, and interviewees with different perspectives) to illuminate each RQ, proved useful for revealing any inconsistencies in the data. The coding and analysis were also discussed with fellow researchers.

5. Presentation and discussion of findings

5.1. CBAs are comprehensive and partly standardized (RQ1)

Our overall assessment based on the document review is that most of the CBAs are relatively comprehensive, and that appraisals of similar types of projects generally include the same impact categories. In particular, payable costs, including both the capital cost and the maintenance and operating cost, are thoroughly estimated in most cases. Some benefits are monetized, most notably payable revenues, time savings, other consumer benefits, and in some cases also impacts on health and safety and the environment. Other impacts are treated as non-monetized impacts in the framework. Overall, only about half of the CBAs (45% of CAs, 55% of QAs) monetize all or the most important impacts. The degree of monetization varies across sectors, but even for road projects, less than 80% of the CBAs monetize all or the most important impacts. Thus, non-monetized impacts play a key role in the studied analyses.

Further, the CBAs of road and rail projects are more standardized than the CBAs of other project types. For example, whereas some CBAs of building projects only present and discuss first-order effects for users (e.g. users of a museum, university, or prison), others discuss long-term, wider benefits, such as improved national competitiveness due to better research and education. The interviewees reported that they were often unsure about whether and how to treat indirect, long-term impacts, for which no guidelines exist. Generally, the level of standardization regarding the non-monetized impacts is low. We return to this problem in Section 5.2.

Some quality assurers claim that the CAs are overly 'creative' with regard to the benefits included. This is particularly the case for non-monetized benefits. Table 3 shows the most common changes made by QAs relative to the CAs. The good news is that the largest category of changes is 'No or minor changes'. There are no clear sector differences. It can also be shown that 'No or minor changes' is correlated with QAs approving the final recommendation, cf. subquestion 15.

The calculation of an NPV is normally based on a number of parameters and assumptions, and an overview of some them is given in Table 4. Although it should be possible to vary most parameters due to, for example, local variation in people's WTP, it seems that the observed variation is somewhat higher than expected. For example, there seems to be much confusion about the discount rate and how it should vary according to systematic risk. Similarly, the degree to which real price adjustment is applied seems arbitrary. Some sectors (e.g. transport) have their own CBA guidelines that specify key parameters and values, implying that practice is more consistent in these CBAs. None of the CBAs included independent valuation studies to obtain exact WTPs.

Prior to 2014, hardly any parameters had been fixed as compulsory in the national guidelines issued by the Ministry of Finance, with the exception of the marginal cost of funds. Since 2014, some additional parameters have been fixed, most notably the discount rate and the value of a statistical life. In our view, this has led to a more consistent practice across CBAs, and should have been considered for other parameters too, most notably the social cost of carbon.

5.2. Inconsistent handling of non-monetized impacts (RQ2)

Non-monetized impacts are often essential in the CBAs. However, their interpretation is sometimes unclear and arbitrary, especially in the CAs. Some findings from the document review are presented in Table 5. On the one hand, the ministries and agencies seem to put more efforts into the analysis of non-monetized impacts than do the quality assurers, but on the other hand, they have a less clear understanding of what those impacts actually measure. Many CAs tend to mix economic impacts with goal achievement and other

Table 3
Changes in CBA structure. QA compared with the CA for the same project (most important change registered) (N = 58).

Type of change	Number	%
No change or minor change	17	29
Impact categories removed	13	22
Impact categories added	8	14
More impacts monetized (formerly non-monetized)	3	5
Impossible to compare due to different approach	12	21
No information	5	9
Total	58	100

Table 4
Selected parameters applied in the CBAs (N = 111).

Parameters	Practice observed
Marginal cost of public funds	0.2 (fixed by the Ministry of Finance)
Discount rate	Varies within the range 2–5%, later fixed at 4% and declines over time
Value of a statistical life	Varies in the range NOK 15–35 million, later fixed at NOK 30 million
Value of time	In most cases, average wage is used for business travel, but lower for leisure (in the transport sectors, based on a Norwegian SP study)
Method for calculating residual value	Large variations. Linear depreciation, market valuation, NPV of remaining net benefit flows, or set to 0
Real price adjustment	Large variations. Applied by some sectors, only for some impacts
Social cost of carbon	Varies within the range NOK 110–400 per ton, later an increasing price path is introduced in some sectors

considerations when presenting non-monetized impacts. Political and strategic considerations at various levels (e.g. agency, sector, region, or a stakeholder group) that extend far beyond consumer preferences are frequently brought into the discussion of whether the projects are good value for money. In our view, this is a serious weakness, that may lead to wrong conclusions.

Not only the interpretation, but also the choice of measurement scales varies considerably (e.g. cardinal, ordinal, or purely qualitative). Most CBAs of road projects apply the road agency's recommended framework for assessing five types of negative effects on nature and the environment, which are summarized in terms of 'plusses and minuses' on a scale ranging from −4 to +4. CBAs of other project types have a less systematic approach. Some quality assurers have introduced their own frameworks for analysing non-monetized impacts, but these frameworks are not consistent.

We consider that the documentation of the non-monetized impacts is sufficient in less than half the CBAs (cf. subquestion 8).

Table 5
Selected findings relating to non-monetized impacts in CBAs, sorted by CAs (N = 53) and QAs (N = 58).

Indicator	All (%)	CAs (%)	QAs (%)
Interpretation/perspective (researcher's understanding)			
Economic impact	56	34	77
Goal achievement, mixed or unclear	44	66	23
	100	100	100
Methodology			
Qualitative	22	21	23
'Plusses and minuses'	54	46	64
Other scoring or ranking	24	33	13
	100	100	100
Comprehensiveness			
Average % of CBA (in terms of page numbers)	22	27	17
Well documented (researchers' judgement), % 'yes'	45	53	36

Generally, the data sources used, the people involved, and the principles for valuation, are not well documented. For example, information about whose judgement they are based on is not provided in many cases. Moreover, in general, the development of these impacts over time is not discussed. There are no obvious differences between sectors or project types.

Interestingly, a comprehensive treatment of the non-monetized impacts in the CA is *not* correlated with QAs approving the final recommendation. Only when CAs apply the same interpretation of non-monetized impacts as the QA, they are more likely to agree on the final recommendation, and vice versa. This is supported by the interviews and indicates that quality assurers tend to be suspicious about a thorough discussion of non-monetized impacts that extend beyond an economic interpretation.

Interviewees from ministries and agencies acknowledged that performing the non-monetized part of the CBA is difficult. One interviewee said, 'In our sector [defence] we often discuss the achievement of military goals rather than socio-economic benefits. I guess we need better guidance on how to distinguish between a *multiple-criteria analysis* and a CBA.' By contrast, the quality assurers are more loyal to the economic perspective.

5.3. Uncertainty thoroughly assessed for capital cost, but to a lesser extent for other impacts (RQ3)

Our document review included an assessment of major uncertainties relating to costs and benefits, and how these were assessed and presented. Generally, the studied CBAs were more concerned about risks to the capital cost than risks to benefits and other long-term impacts. The reason probably lies in the QA scheme itself, which requires that stochastic estimation techniques are applied to estimate the capital cost, but there are no such requirements for other impacts. Overall, capital cost uncertainties are well handled in the studied CBAs. Uncertainties relating to other impacts are more varied and often superficial. About 60% of the CBAs (CAs and QAs alike) report sensitivity tests, but such tests are often simple and only focus on one or two parameters. One analyst said, 'We have strict deadlines, and sensitivity testing is just one of the things that we don't have time for.' Uncertainties relating to non-monetized impacts are rarely discussed in the CBAs. In our view, more attention should be paid to uncertainties in all impacts, not just capital cost.

The combination of uncertainties and irreversible investments that gives rise to quasi-option values (Boardman et al., 2011) is discussed briefly and qualitatively in some of the QA reports. Quasi-option values are typically higher in the zero-investment alternative, and in some cases this has been used by quality assurers as an argument for postponing the investment decision.

Overall, we consider that about two-thirds of the CBAs as acceptable with regard to identifying and analysing risk (cf. subquestion 11). QAs perform far better than do CAs (74% acceptable versus 47%). Interestingly, when a CA is in the 'acceptable' category, the QA approves the final

recommendation more often. This indicates that QAs recognize a good uncertainty analysis as a crucial quality indicator of the CBA.

5.4. Other considerations are not clearly distinguished from value for money (RQ4)

Overall, 47% of CAs present other decision criteria (goal achievement, distributional analyses etc.) along with the CBA, whereas only 5% of the QAs do the same (cf. subquestion 14). We do not find any clear sectoral differences. Generally, the discussion of distributional impacts is rather superficial, and in most cases not sufficiently comprehensive. Immediate effects are discussed more often than are long-term distributional effects. For example, impacts on future generations are hardly mentioned in any of the reports. An equally worrying observation is that when such other considerations are included in the report, they are in many cases not clearly separated from the value for money perspective.

As discussed in Section 5.2, benefits for specific groups or regions are often discussed in the CBAs as if they were net economic benefits to the country, although they may be a matter of redistribution. This explains the failure to report distributional impacts in many of the CBAs, particularly the CAs. They are already reported as benefits (but the corresponding negative impacts for other groups are not presented). By contrast, the quality assurers mention that their primary focus is on value for money, and some seem to ignore decision-makers' need for supplementary information altogether. Cross-tabulations show that CAs that present a broad and holistic decision base, correlates with QAs *not* approving their recommendations.

It should be noted that the distinction between wider economic benefits and pure distributional effects (i.e. economic effects that are most likely to be offset elsewhere) is not always clear. Our interviewees confirmed that performing this part of the analysis is challenging, and that more research and better guidance is welcome.

5.5. Appraisal optimism has been avoided for NPV estimation, but may influence the CBA in other ways (RQ5)

Although not always openly stated, there is commonly a preferred project alternative from the agency's perspective. One of the consultants stated: 'Everyone knows which concept the CA is hoping for, and it is always the most expensive one.' This raises the question of whether the CAs are biased in favour of a preferred alternative.

In the absence of ex post data, we compared the CBAs done by agency and quality assurer, in the knowledge that the latter party was independent of the project. It should be noted that the quality assurers may introduce new combined alternatives or adjustments to existing alternatives, for example to make the zero investment alternative more realistic, which implies that the sets of project alternatives assessed in the two reports are not identical. Therefore, instead of pairwise comparisons of alternatives, we studied the characteristics of *each party's highest ranked alternative*.

Generally, the QAs disagree with the CA recommendations, either partly or fully, in the majority of projects (33 out of 58). We have already mentioned that QAs seem to 'reward' CAs for having an appropriate CBA structure and for including a comprehensive uncertainty analysis, but *not* for comprehensive analyses of non-monetized impacts or for presenting a broad decision base. We also found that there are no striking sectoral differences: if anything, there seems to be slightly less disagreement about defence projects. Next, we focus our discussion on the extent to which CAs are systematically more optimistic about the projects' value for money. Specifically, in the knowledge that QAs put much weight on the NPV, one could suspect that the CAs present a biased NPV.

From Table 6, it can be seen that the CAs recommended project alternatives with a negative or zero NPV in 75% of the cases, whereas the corresponding percentage for the QAs is slightly lower (64%). Thus, it is apparent that the ministries and agencies are not concerned about promoting projects with a negative NPV. Rather, these findings may indicate that the NPV is not manipulated to make projects appear more profitable.

It should be noted that in our review of parameters and assumptions (cf. subquestion 3), we also looked for systematic differences between the CAs and the QAs. In this case, too, we did not find any clear indications that the CAs applied more optimistic parameters. Generally, practice seemed to vary as much between different quality assurers as between quality assurers and ministries and agencies.

However, we cannot exclude the possibility that CAs are biased in terms of the non-monetized impacts, or by excluding or systematically downgrading the simplest and less costly alternatives. As shown in the lower part of Table 6, CAs recommend the alternative with the highest or least negative NPV less often than do QAs. CAs hardly ever recommend the zero alternative. One group of projects that attracted our attention is those for which CA recommends an alternative with negative NPV and the QA recommends an alternative with positive NPV (10 projects). In each of these cases, the QA either preferred a less costly alternative, or downscaled the alternative recommended by the CA, thus turning a negative NPV into a positive one.

The findings presented in Table 6 also demonstrate the emphasis that ministries and agencies, and to some extent quality assurers put on the non-monetized impacts, which are

Table 6
Characteristics of the recommended project alternative (N = 58 projects).

Indicator	All (%)	CAs (%)	QAs (%)
Sign of NPV in recommended alternative			
Positive	30	25	36
Negative or zero	70	75	64
	100	100	100
The recommended alternative has the highest/least negative NPV, % of the CBAs	55	44	66
Zero alternative recommended, % of the CBAs	11	3	19

considered to outweigh a negative NPV in the majority of cases. In light of the emphasis put on those impacts, the inconsistent interpretation and treatment of such impacts is worrying (as discussed previously). Furthermore, there are indications that the quality assurers do not scrutinize this part of the CA in the same way as they scrutinize the NPV. One interviewed quality assurer said, 'I guess the agencies realize that any attempts to cheat with numbers will be revealed. It is easier to get away with the qualitative assessments.' The interviewees from the agencies denied that they had manipulated the data. Rather, they accused quality assurers of ignoring important non-monetized benefits. The interviewees who were decision-makers stated that they felt uncertain about how to interpret the reports and which party to believe when the CA and QA differed. First and foremost, they considered it important to be able to trust the quality of the CBAs. Some referred to the QA reports as helpful for determining the quality of the CAs, but one interviewee said he would have liked the QA reports to be 'reviewed by independent experts too'.

5.6. Transparency and communication acceptable, but could be improved (RQ6)

Transparency and clear communication are crucial to ensure CBA usefulness. Overall, we judge the level of transparency as acceptable (cf. subquestion 19) in c.80% of the studied CBAs, meaning that they are documented in sufficient detail, either in the main report or in an appendix. However, many reports could have been improved. Key parameters, such as the discount rate, price level, and period of analysis, are not always explicitly stated; for example, 12% of the CBAs do not include information about the discount rate used. Generally, the QAs are more transparent than are the CAs. There is also a tendency for the more transparent CBAs to have been produced by inexperienced agencies than by, for example, the road and rail agencies, possibly because they lack a standard framework and therefore need to explain every step of their analysis.

Traffic models and impact models are frequently used by the transport agencies, and some consultants have developed their own economic models that produce inputs to the CBAs. These models are not always well explained in the reports, and often appear as black boxes. Even experts in the agencies find the models difficult to understand, as exemplified by one interviewee, who said, 'The result of traffic simulations depends on so many detailed assumptions about the new road, such as curvature, width, velocity, etc. It is impossible to understand everything. You just have to trust the model.' One quality assurer admitted that he often took the traffic estimates from the agencies' models for granted, because it was impossible to verify them. By contrast, interviewees from ministries/agencies accused some consultant of treating their own models as business secrets.

Economic impacts are often presented in an aggregate form in the CBAs. For example, road projects normally generate a range of emissions to air (NOx, CO2, N2O, and local air pollution in the form of particulate matter). These are commonly presented in the reports as 'environmental costs', which obscures their individual impacts.

Furthermore, in all projects, a large number of project-specific assumptions will have to be set by the analyst. These are not always well explained in the CBAs. One example is the assumption made about toll fees on new roads in Norway, which may affect consumer benefits significantly. In two-thirds of the road project CAs, it is assumed there are no user fees, and hardly any of those CAs include an explanation of the reasons behind this assumption. The QA reports are therefore useful because they may question key assumptions. They may agree or disagree with the ministries and agencies, but their discussions will nevertheless add useful information for decision-makers. We only find a slightly positive correlation between the transparency in CAs and the QAs approving the final recommendation.

In many CBAs, technical language is used, and the reports are long: reports with 100 pages or more are common. This is relevant in terms of accessibility because decision-makers normally face constraints in terms of their expertise and time. The majority of CBAs (95% of QAs and 63% of CAs) include a summary. However, most of these summaries are short and rather superficial. In our view, only c.10% of the reports include a sufficiently informative summary that cover all major impacts (whether monetized or not), uncertainties, distributional impacts and/or other considerations, and key assumptions on which the results are based.

The interviewed decision-makers confirmed that they often found it difficult to understand the complexity of CBAs. They also confirmed that they thought summaries should be more comprehensive.

5.7. Decision-makers found CBAs more useful when approved by an independent party (RQ7)

The ultimate test of whether decision-makers' find CBAs useful, is the extent to which they follow the recommendations in the reports. Certainly, other concerns than value for money may affect public investment decisions, and traditionally the CBA has not been very influential in public project decision-making in Norway. However, it is important to note that the CBA follows an assessment of public needs and strategies, implying that the shortlisted alternatives are all considered relevant to these strategies. We therefore expect political decision-makers to follow the ranking based on value for money at least to some extent, given that they have confidence in the analyses.

Overall, in the majority of cases (c.80%), the Cabinet has chosen to go-ahead with either one conceptual alternative or, in a few cases, several conceptual alternatives to be developed further into a major construction project. Only in c.20% of the cases is the zero alternative selected or the project put on hold or withdrawn. There are no clear differences between project types. We did a large number of cross-tabulations to shed light on how CBA quality might have influenced decisions. The following findings are worth mentioning. A low degree of monetization does not seem to reduce adherence. Rather,

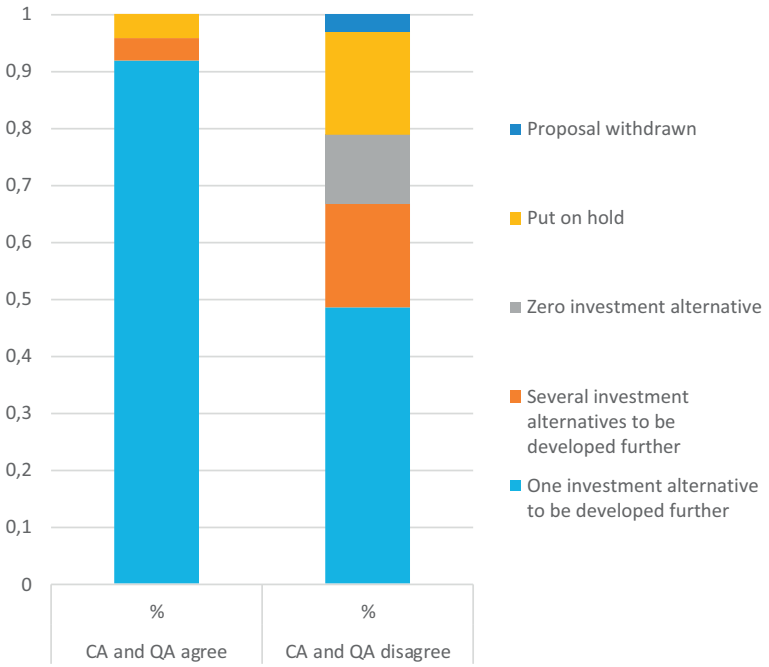

Fig. 3. Cabinet decisions, based on 58 projects, of which for 25 the two CBAs agree and for 33 they disagree with key recommendations. Percentages for each of the two groups.

decision-makers' adherence seem to be higher when the CBAs include comprehensive analyses of the non-monetized impacts and the distributional impacts, and they prefer reports that present a broad decision base that includes more than value for money. There is *no* correlation between adherence and the sign of the NPV in the recommended alternative, which is another indication that decision-makers care about the non-monetized impacts. By contrast, comprehensive risk analysis is not correlated with adherence. This is partly in contrast to the quality indicators that QAs seem to emphasize in their assessment of the CAs.

One finding that attracted our attention was that when CA recommendations (based on both NPV *and* the non-monetized impacts) were approved by the QAs, decision-makers' adherence was substantially higher. The distinction between cases in which ministry/agency and quality assurer agreed on the project ranking and cases in which they disagreed, is shown in Fig. 3. The Cabinet has followed the recommendation in 92% of the cases in which the two CBAs are in agreement. By contrast, when they are not in agreement, the Cabinet has made a clear choice of concept in only 48% of cases, often in line with CA recommendations. In the remaining 52% of cases, the Cabinet has chosen either multiple alternatives or no investment (the latter often in line with QA recommendations) or has put the decision on hold. In one case, a sports event, the proposal was withdrawn following a very critical QA report. These findings suggest that decision-makers care about more than just value for money. They also suggest that CBAs are heeded and that a critical QA can make decision-makers stop and reconsider the case.

Additionally, we asked interviewees to comment explicitly on the perceived usefulness of the CBAs. The majority, especially those who were decision-makers, found the CBAs useful 'given that they are of high quality'. One interviewee said, 'The existence of two CBAs that come to the same conclusion is a strong indicator of quality.' Another interviewee, a consultant, stated that 'In some cases, politicians need an excuse for rejecting a hopeless project, and a critical QA

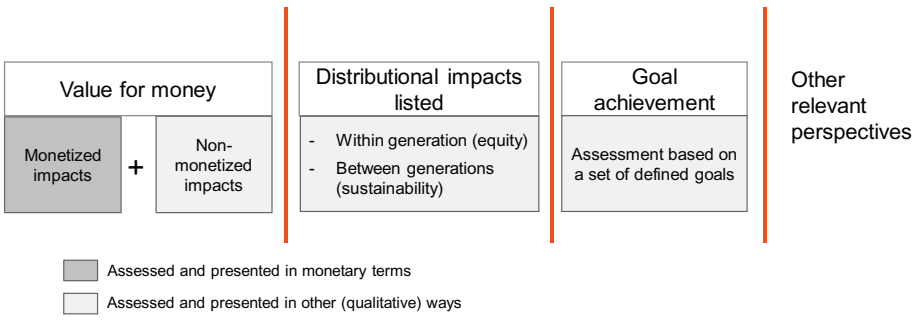

Fig. 4. Suggested early-phase business case – different decision perspectives (not to be added).

report can be that excuse.' Also, one interviewee noted that the QA scheme itself might discourage agencies from coming forward with poor proposals in the first place. However, another interviewee reminded us that decision-makers are not obliged to follow the advice from CBAs, and said, 'It is nice to know a project's value for money, but we cannot make politics only based on that.'

6. Conclusions

A CBA offers a clearly defined interpretation of project success, as may be formally required in relation to public project selection. However, challenges and weaknesses in CBAs may be overlooked, which implies that decision-makers may not find them useful and trustworthy. We have studied the usefulness of CBAs produced as part of compulsory appraisals of major infrastructure projects. Two types of CBAs are done for major public projects in Norway, one by the initiating ministry/agency and one by external quality assurers. Both types of CBAs rank the project alternatives based on their estimated value for money. With a few exceptions, they are openly available to researchers as well as to members of the wider public.

We expected, and found, that the studied CBAs would be and are largely of good quality. In particular, the use of independent quality assurers is normally considered a means to reduce the risk of appraisal optimism. Also, the risk of inconsistent, incomplete, and inaccurate estimates should be limited, given the time and resources spent on the analyses and the considerable expertise involved. Thus, the study of a 'critical case' (Flyvbjerg, 2006) should be useful to explore the potential for overcoming any CBA weaknesses, and to identify weaknesses that are more difficult to avoid than are others.

6.1. CBAs are heeded by decision-makers

A key finding from our research is that decision-makers consider CBAs a vital part of the business case for infrastructure project proposals. This was found through direct measurement (interviews) as well as indirectly (revealed adherence to recommendations). This contrasts with the role of CBAs before the QA scheme was introduced in 2005. In the past, if CBAs were produced at all, they rarely affected public project decision-making in Norway (Nyborg, 1998). Generally, we find that the ministries and agencies invest considerable resources in their CBAs today. This is in line with findings from an earlier study (Volden and Andersen, 2018), which demonstrates that the QA scheme has led to strong efforts in ministries and agencies to strengthen their project competencies and governance models at the agency level. However, we would like to make it clear that we have not proved an effect of the QA scheme as such.

We find that the Cabinet has almost always approved a project proposal if it was recommended as good value for money by the ministry/agency, *and* endorsed by the quality assurer. However, if a project proposal was recommended by the ministry/agency, but not endorsed by the quality assurer, the Cabinet was more likely to have rejected it or reconsidered it. This is a clear indication that the CBAs are heeded by decision-makers. Furthermore, the interviewed decision-makers explicitly stated that they considered the use of two CBAs was a stronger decision base than the use of just one CBA. This finding is in line with literature on appraisal optimism that recommends an external view on the appraisal and planning of a project (Flyvbjerg, 2009; Lovallo and Kahneman, 2003; Mackie and Preston, 1998).

Our findings indicate that appraisal optimism has largely been avoided in NPV estimation (i.e. the third category of weaknesses in CBAs, cf. Fig. 1). Ministries and agencies generally do not estimate NPVs as positive more often than do quality assurers. The fact that an external review will be performed seems to have a disciplining effect on ministries and agencies. However, we cannot exclude the possibility that CAs deliberately downgrade or exclude 'cheap alternatives' in some cases.

Furthermore, the comprehensiveness and consistency of analyses is largely at an acceptable level in the studied CBAs (cf. the second category of weaknesses). This also applies to transparency, which is essential to reveal all three types of weaknesses in CBAs and to increase decision-makers' understanding of the analyses. Thus, the situation in Norway is somewhat more encouraging than that found in the UK by Atkins et al. (2017), where inconsistency, poor transparency, and communication were serious weaknesses in project appraisals. Similarly, Annema (2013) found that transparency in Dutch CBAs was generally poor, despite the introduction of a new CBA guide that had led to other improvements. An explanation may be the requirement that QA reports in Norway should be openly available to the public. Nevertheless, there is potential for improvement in the Norwegian CBAs with regard to consistency and to uncertainty assessments and transparency.

To summarize, the following research questions have all largely been answered with a 'yes' response or at least a 'to an acceptable extent' response: RQ1 about CBA consistency across projects, RQ3 about uncertainties being identified and presented, RQ5 about unbiased estimates, and RQ6 about transparency and clarity. This may, in turn, explain why RQ7 about decision-makers' adherence to CBA recommendations, can also be answered with a conditional affirmative.

6.2. Non-monetized impacts need a clearer definition and more systematic treatment, distinguished from considerations beyond the project's value for money

Two remaining weaknesses in CBAs require attention. First, RQ2 about whether the non-monetized impacts are handled consistently has been answered negatively. Second, the answer to RQ4, about distributional impacts and other considerations, is that such issues are being presented and discussed in CAs, but they are often mixed with the value for money assessments. The former finding is much in line with the findings of Ackerman (2004) and of Mackie and Preston (1998), whereas the latter finding has not been studied systematically, to our knowledge.

It should be noted that the two weaknesses are related. There may be many pros and cons relating to the project beyond value for money. Our findings confirm that decision-makers do care about information beyond value for money assessments. However, when included, such other considerations are often incorrectly referred to as non-monetized impacts and 'added' to the NPV. This creates confusion for decision-makers, who cannot be sure what has been measured (i.e., whether value for money or some other confounded criterion).

One explanation for such observed weaknesses is that the non-monetized part of a CBA is a difficult topic - a fact that is neglected in CBA textbooks and guidelines. However, differences between ministries/agencies and quality assurers may also indicate opportunism. This means that ministries and agencies may deliberately overestimate the non-monetized impacts by including benefits that are not true economic benefits, and they could do this in the knowledge that it would be more difficult for the quality assurers to disprove qualitative arguments than quantitative arguments. If that is the case, the problem of appraisal optimism in CBAs may be present after all, although in another form than expected.

Clearly, methodological improvements as well as guidelines for assessing non-monetized impacts are required. Additionally, quality assurers must take such impacts seriously. Assessments of non-monetized impacts ought to be guided by the question of *whether they are likely to improve or deteriorate the NPV*, not by some other valuation principle (such as whether they are in line with a set of political goals). Admittedly, the distinction between consumer preferences and other perspectives is not easy in practice, but this is also a challenge in monetization (Sager, 2013; Mouter and Chorus, 2016). If we allow for arbitrary interpretations of the non-monetized impacts, the pricing versus non-pricing decision could become an opportunistic one.

As noted by Laursen and Svejvig (2016), the definition of 'value' in projects is often vague and may depend on the perspective taken. The great advantage with value for money as defined by the CBA is the clarity. Therefore, it is important to accept that definition in practice, whether impacts are monetized or not monetized. The great advantage with value for money as defined by the CBA is the clarity. The disadvantage is that only efficiency aspects are covered. We believe the definition of CBA should be accepted, whether impacts are monetized or not monetized. However, with a narrow interpretation of non-monetized impacts, it is even more crucial to balance value for money against other perspectives or interpretations of social value. Not only should each project alternative's distributional impacts be presented as part of the business case, but we suggest that also each project alternative's achievement of relevant goals and strategies is assessed and presented. Goals and strategies may overlap with value for money, which would typically be the case when goals are related to national economic development. In other cases, goals and strategies may be better aligned with distributional considerations, and thus in conflict with value for money considerations. For example, goals could be defined for the well-being of specific groups or regions, environmental sustainability or other considerations not well covered by the CBA (cf. Section 2.1). Basically, goals and strategies could be related to anything that political decision-makers care about.

Admittedly, goal alignment is already checked for the shortlisted alternatives in a CA, but some alternatives will often score higher than do others, which may be relevant for project selection. We think it is important that the three (or more) perspectives are presented separately, as shown in Fig. 4 by the thick lines between them. Thus, it is clear that although the monetized and non-monetized impacts should be added to assess the project's value for money, the different decision perspectives should not be added. Instead, any conflicts between the perspectives should be identified, and the final balancing between them ought to be done by the decision-makers. Should there be no conflicts, this will normally be highly relevant and useful information too. The framework constitutes a holistic business case that can easily be expanded to fit with an early-phase version of the Five Case Model applied in the UK (HM Treasury, 2013) or with the OECD-DAC criteria (Volden, 2018). This topic is worthy of more attention from the research community as well as from governments.

6.3. Recommendations

The findings from our research have provided the basis for a set of practical recommendations to increase CBA usefulness. The target group for these recommendations is project owners and senior officers who are responsible for project governance frameworks. Although the studied projects are public ones, we believe that many of the following recommendations are relevant to private sector organizations too.

1. A number of perspectives beyond value for money may be relevant to decision-makers. We suggest these perspectives are defined by decision-makers in advance and included in the business case. In our study, high-quality CBAs were often presented alone, forming a business case that was too narrow.
2. An important purpose of a CBA is to assess a number of alternative solutions to the problem at hand. Not only large construction projects, but also simple and low-cost solutions. One should be aware that project promoters may not have the right incentives to include the latter type of alternatives.
3. Completeness and consistency are important quality criteria, which comprise, for example, the impact categories included, the extent to which impacts are monetized, and the choice of parameter values. Although all projects are unique, our findings indicate that there is room for more standardization.
4. Possible errors and uncertainties need to be identified and presented as part of the CBA, to the extent that they can affect the ranking and recommendations.
5. The Non-monetized impacts are as relevant as the monetized impacts. They should not be ignored (as some, highly

experienced, analysts tended to do, in this study), nor should they be overvalued or mixed with other perspectives than the value for money perspective.
6. Measures to prevent optimism bias on the part of project promoters are recommended. Relevant measures such as transparency and external quality assurance of reports, seemed to work well in the studied projects.
7. Although not found to be a problem in this study, analyst competence and qualifications are key.
8. Understandability and communication (meaning, for example, the use of simple language and a readily available summary) are important aspects of transparency in reports, and relevant to decision-makers who are not CBA experts.

7. Limitations and further work

The use of case projects from a single country has some limitations. Therefore, broader conclusions cannot be drawn on the basis of our findings. In particular, as highlighted in the governance literature, a project governance scheme ought to be adapted to a specific context. The experiences gained from the application of the Norwegian QA scheme may not be transferable to other countries. An interesting topic for further study would be a systematic comparison of CBA practices in countries that have introduced independent review of CBAs.

Further, for the sake of simplicity, we have assumed that decisions are based on an instrumental decision logic, and we have not considered adverse incentives on the decision-making level. The true potential for improving decisions through better CBAs would probably be moderated by various conditions at the decision-making level. An extended model ought to be established to take this into account.

Additionally, it should be noted that we have studied CBAs in an early project phase. It remains to be seen whether the projects will actually be good value for money after they have been implemented. The selected project alternative needs to be developed further in a detailed planning process before the project is implemented. In that phase, there is a risk of cost escalation, and the realization of intended impacts has to be ensured through active cost and benefits management. An interesting topic for further research would therefore be to follow the projects throughout subsequent phases, and to perform updated CBAs *in medias res* as well as *ex post*, to learn whether the agencies manage to retain their focus on producing value for money.

Funding

The work was supported by the Concept Research Program at the Norwegian University of Science and Technology, which in turn is funded by the Norwegian Ministry of Finance.

Declaration of interest

No conflicts of interest.

Acknowledgements

The author would like to thank her colleagues Ole Jonny Klakegg, Heidi Bull-Berg and Ola Lædre for interesting discussions and useful comments to an earlier draft. Further, she would like to thank the editor and the anonymous reviewers for their contributions to improving the paper in the review process.

References

Ackerman, F., 2004. Priceless benefits, costly mistakes: what's wrong with cost-benefit analysis. Post-autistic Econ. Rev. 25.

Andersen, B.S., Samset, K., Welde, M., 2016. Low estimates – high stakes: underestimation of costs at the front-end of projects. Int. J. Manag. Proj. Bus. 9 (1), 171–193.

Annema, J.A., 2013. The use of CBA in decision-making on mega-projects: Empirical evidence. In: Priemus, H., van Wee, B. (Eds.), International Handbook on Mega-Projects. Edward Elgar, Cheltenham, UK, pp. 291–313.

Association for Project Management, 2018. Resources. https://www.apm.org.uk/resources/ (retrieved April 2018).

Atkins, G., Davies, N., Kidney Bishop, T., 2017. How to Value Infrastructure: Improving Cost Benefit Analysis. Institute for Government, London, UK.

Baccarini, D., 1999. The logical framework method for defining project success. Proj. Manag. J. 30 (4), 25–32.

Bertisen, J., Davis, G.A., 2008. Bias and error in mine project capital cost estimation. Eng. Econ. 53 (2), 118–139.

Boardman, A., Greenberg, D., Vining, A., Weimer, D., 2011. Cost-Benefit Analysis. 4th ed. Pearson.

Breese, R., Jenner, S., Serra, C.E.M., Thorp, J., 2015. Benefits management. Lost and found in translation. Int. J. Proj. Manag. 33, 1438–1451.

Browne, D., Ryan, L., 2011. Comparative analysis of evaluation techniques for transport policies. Environ. Impact Assess. Rev. 31 (3), 226–233.

Dobes, L., Bennett, J., 2009. Multi-criteria analysis: 'good enough' for government work? Agenda 16 (3).

Eisenhardt, K., 1989. Agency theory: an assessment and review. Acad. Manag. Rev. 14 (1), 57–74.

Eliasson, J., Börjesson, M., Odeck, J., Welde, M., 2015. Does benefit–cost efficiency influence transport investment decisions? J. Transport Econ. Pol. 49 (3), 377–396.

Elvik, R., 2017. The Value of Life: The Rise and Fall of a Scientific Research Programme. Doctoral theses at NTNU 2017:340. NTNU, Trondheim, Norway.

Eskerod, P., Huemann, M., 2013. Sustainable development and project stakeholder management: what standards say. Int. J. Manag. Proj. Bus. 6 (1), 36–50.

Finansdepartementet, 2005. Veileder i samfunnsøkonomiske analyser. Regjeringen, Oslo, Norway.

Finansdepartementet, 2014. Prinsipper og krav ved utarbeidelse av samfunnsøkonomiske analyser mv. Rundskriv R-109/14. Regjeringen, Oslo, Norway.

Flyvbjerg, B., 2006. Five misunderstandings about case-study research. Qual. Inq. 12 (2), 219–245.

Flyvbjerg, B., 2009. Survival of the unfittest: why the worst infrastructure gets built—and what we can do about it. Oxf. Rev. Econ. Policy 25 (3), 344–367.

Flyvbjerg, B., Bruzelius, N., Rothengatter, W., 2003. Megaprojects and Risk: An Anatomy of Ambition. Cambridge University Press, Cambridge, UK.

Haavaldsen, T., Lædre, O., Volden, G.H., Lohne, J., 2014. On the concept of sustainability – assessing the sustainability of large public infrastructure investment projects. Int. J. Sustain. Eng. 7 (1), 2–12.

HEATCO, 2006. Deliverable 5: Proposal for Harmonised Guidelines. http://www.kbsz.hu/dokumentumok/20070411_0.2-HEATCO_D5.pdf (retrieved 24 Sept 2018).

Hjelmbrekke, H., Klakegg, O.J., Lohne, J., 2017. Governing value creation in construction project: a new model. Int. J. Manag. Proj. Bus. 10 (1), 60–83.
HM Treasury, 2013. Public Sector Business Cases Using the Five Case Model (Green Book Supplementary Guidance on Developing Public Value from Spending Proposals).
Jenner, S., 2015. Why do projects 'fail' and more to the point what can we do about it? The case for disciplined, 'fast and frugal' decision-making. PM World J. 4 (3).
Kelly, C., Laird, J., Constantini, S., Richards, P., Carbajo, J., Nellthorp, J., 2015. Ex post appraisal: what lessons can be learnt from EU cohesion funded transport projects. Transp. Policy 37, 83–91.
Laursen, M., Svejvig, P., 2016. Taking stock of project value creation: a structured literature review with future directions for research and practice. Int. J. Proj. Manag. 34 (4), 736–747.
Lovallo, D., Kahneman, D., 2003. Delusions of success: how optimism undermines executives' decisions. Harv. Bus. Rev. 81 (7), 56–63.
Mackie, P., Preston, J., 1998. Twenty-one sources of error and bias in transport project appraisal. Transp. Policy 5, 1–7.
Mackie, P., Worsley, T., Eliasson, J., 2014. Transport appraisal revisited. Res. Transp. Econ. 47, 3–18.
Morris, P.W.G., 2013. Reconstructing project management reprised: a knowledge perspective. Proj. Manag. J. 44 (5), 6–23.
Mouter, N., 2017. Dutch politicians' use of cost-benefit analysis. Transportation 44, 1127–1145.
Mouter, N., Chorus, C., 2016. Value of time – a citizen perspective. Transp. Res. A 91, 317–329.
Müller, R., 2009. Project Governance: Fundamentals of Project Management. Gower, New York, NY.
Musawir, A., Serra, C., Zwikael, O., Ali, I., 2017. Project governance, benefit management, and project success: towards a framework for supporting organizational strategy implementation. Int. J. Proj. Manag. 35 (8), 1658–1672.
Næss, P., 2006. Cost-benefit analysis of transportation investments: neither critical nor realistic. J. Crit. Real. 5 (1), 32–60.
Næss, P., Volden, G.H., Odeck, J., Richardson, T., 2017. Neglected and Underestimated Negative Impacts of Transport Investments. Concept Report No. 54. Ex ante Academic Publisher, Trondheim, Norway.
Nicolaisen, M.S., Driscoll, P.A., 2014. Ex-post evaluations of demand forecast accuracy: a literature review. Transp. Rev. 34 (4), 540–557.
Nyborg, K., 1998. Some politicians' use of cost-benefit analysis. Public Choice 95, 381–401.
Nyborg, K., 2014. Project evaluation with democratic decision-making: what does cost-benefit analysis really measure? Ecol. Econ. 196, 124–131.
Office of Government Commerce, 2009. Managing Successful Projects with PRINCE2 (PRINCE2™). 5th ed. TSO (The Stationery Office), London, UK.
Patton, M.Q., 1999. Enhancing the quality and credibility of qualitative analysis. Health Serv. Res. J. 34 (5 Pt 2), 1189–1208.
Pearce, D., Atkinson, G., Mourato, P., 2006. Cost-benefit Analysis and the Environment: Recent Developments. OECD, Paris, France.
Project Management Institute (Ed.), 2017. A Guide to the Project Management Body of Knowledge (PMBOK® Guide), 6th ed Project Management Institute, Newton Square, PA.
Sager, T., 2013. The comprehensiveness dilemma of cost-benefit analysis. Eur. J. Transp. Infrastruct. Res. 13 (3), 169–183.
Sager, T., 2016. Why don't cost-benefit results count for more? The case of Norwegian road investment priorities. Urban Plan. Transp. Res. 4 (1), 101–121.
Samset, K., 2003. Project Evaluation: Making Projects Succeed. Tapir Academic Press. Trondheim, Norway.
Samset, K., Christensen, T., 2017. Ex ante project evaluation and the complexity of early decision-making. Public Org. Rev. 17 (1), 1–17.
Samset, K., Volden, G.H., 2012. The proposal. In: Williams, T., Samset, K. (Eds.), Project Governance: Getting Investments Right. Palgrave Macmillan, Basingstoke, UK, pp. 46–80.
Samset, K., Volden, G.H., 2015. Front-end definition of projects: ten paradoxes and some reflections regarding project management and project governance. Int. J. Proj. Manag. 34 (2), 297–313.
Scriven, M., 2015. Key Evaluation Checklist (KEC). http://michaelscriven.info/papersandpublications.html (retrieved 23rd January 2019).
Serra, C.E.M., Kunc, M., 2015. Benefits realisation management and its influence on project success and on the execution of business strategies. Int. J. Proj. Manag. 33 (1), 53–66.
Shenhar, A.J., Dvir, D., Levy, O., Maltz, A.C., 2001. Project success: a multidimensional strategic concept. Long Range Plan. 34, 699–725.
Small, K., 1999. Project evaluation. In: Gómez-Ibáñez, J., Tye, W.B., Winston, C. (Eds.), Essays in Transport Economics and Policy: A Handbook in Honor of John R. Meyer. Brookings Institution, Washington, DC.
Standing Advisory Committee on Trunk Road Assessment (SACTRA), 1999. Transport and the Economy. TSO, Norwich, UK.
Terlizzi, M.A., Albertin, A.L., de Oliveira, H.R., de Moraes, C., 2017. IT benefits management in financial institutions: practices and barriers. Int. J. Proj. Manag. 35 (5), 763–782.
Venables, A.J., 2007. Evaluating urban transport improvements: cost-benefit analysis in the presence of agglomeration and income taxation. J. Transp. Econ. Pol. 41 (2), 173–188.
Vickerman, R., 2008. Transit investments and economic development. Res. Transp. Econ. 23 (1), 107–115.
Volden, G.H., 2018. Public project success as seen in a broad perspective: lessons from a meta-evaluation of 20 infrastructure projects in Norway. Eval. Prog. Plan. 69, 109–117.
Volden, G.H., Andersen, B., 2018. The hierarchy of public project governance frameworks: an empirical study of principles and practices in Norwegian ministries and agencies. Int. J. Manag. Proj. Bus. 11 (1), 174–198.
Volden, G.H., Samset, K., 2017a. Governance of major public investment projects: principles and practices in six countries. Proj. Manag. J. 48 (3), 90–108.
Volden, G.H., Samset, K., 2017b. Quality assurance in megaproject management: The Norwegian way. In: Flyvbjerg, B. (Ed.), The Oxford Handbook of Megaproject Management. Oxford University Press, Oxford, UK.
Wachs, M., 1989. When planners lie with numbers. J. Am. Plan. Assoc. 55 (4), 476–479.
van Wee, B., 2007. Large infrastructure projects: a review of the quality of demand forecasts and cost estimations. Environ. Plan. B 34, 611–625.
van Wee, B., 2013. Ethics and the ex ante evaluation of mega-projects, in: Priemus, H., van Wee, B. (Eds.), International Handbook on Mega-Projects. Edward Elgar, Cheltenham, UK, 356–381.
van Wee, B., Rietvold, P., 2013. CBA: Ex ante evaluation of mega-projects, in: Priemus, H., van Wee, B. (Eds.), International Handbook on Mega-Projects. Edward Elgar, Cheltenham, UK, 269–291.
Williams, T., Samset, K., 2010. Issues in front-end decision making on projects. Proj. Manag. J. 41 (2), 38–49.
World Bank, 2010. Cost-Benefit Analysis in World Bank Projects. Independent Evaluation Group, World Bank, Washington, DC.
Zwikael, O., Smyrk, J., 2012. A general framework for gauging the performance of initiatives to enhance organizational value. Br. J. Manag. 23, 6–22.