

Postprint Version	1.0
Journal website	http://dx.doi.org
Pubmed link	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=10958990&query_hl=52&itool=pubmed_docsum
DOI	10.1016/S0168-8510(00)00088-9

* Corresponding author. Tel.: +31-30-2729652; fax: +31-30-2729729. E-mail address:
w.boerma@nivel.nl (W.G.W. Boerma).

0168-8510/00/\$ - see front matter © 2000 Elsevier Science Ireland Ltd. All rights reserved. PII: S0168-8510(00)00088-9

Science in practice: can health care reform projects in central and eastern Europe be evaluated systematically?

SASKIA J. GRIELEN^A, WIENKE G.W. BOERMA^{A,*}, PETER P. GROENEWEGEN^{A,B}

a NIVEL, Netherlands Institute of Primary Health Care, P.O. Box 1568, 3500 BN Utrecht, The Netherlands

b Department of Sociology and Department of Human Geography, Utrecht University, Utrecht, The Netherlands

ABSTRACT

Since the beginning of the 1990s, health care reform projects have taken place in many of the former Communist countries, but these projects are rarely evaluated systematically. Evaluation, however, is an important tool for increasing their rationality and continuity. The aim of this paper is to identify the difficulties in the efforts towards systematic evaluation and draw lessons for the future. For this aim, the requirements for a more rigorous, controlled evaluation are compared with our experiences of evaluating a health care reform project in the Slovak republic. From this comparison a number of discrepancies arise: it was difficult to set clear and realistic goals at the start of the project; the outcomes of the project could not always be measured, nor could 'the process' always be distinguished from the outcomes. Systematic evaluation was further hampered by an insufficient degree of structuration of the project, in advance and during the implementation, and by the absence of a tradition and infrastructure for data collection. On the basis of the experiences and relevant literature, recommendations for future evaluations are formulated. The main lesson is that, given the context, often it will not be possible to use an ambitious evaluation design, and concessions need to be made. At the same time, continuous efforts towards more systematic evaluation procedures should be made, but it is wise and more sustainable to do this in an incremental way.

1. INTRODUCTION

The collapse of the Communist regimes in central and eastern Europe heralded an era with profound reforms in the economic, political and social field [1,2]. One of the fields of change was the health care system. The reasons behind the reforms were many: the health status of the population was bad and the life expectancy worsening. Moreover, there was widespread popular dissatisfaction with the health services and the system was considered inadequate, which was reflected in inefficiency, financial shortages and low quality of the available services [3-5].

Before the fall of Communism, the health care systems in the countries of central and eastern Europe (CEE) were highly centralised and financed from the state budget, with physicians as salaried state employees. In a number of countries, since 1989 a social insurance system has been developed and privatisation in primary care has started. A common aim in the reforms is strengthening primary care and (re)introducing the family physician. The aim of this policy is to reduce the use of specialist care, in order to contain health care costs [3,5].

Since the beginning of the 1990s, with foreign support, health care reform projects have taken place in most of the former communist countries. The aim of these projects is to provide support for the design and implementation of health care reforms. The projects are carried out with money and expertise from the western industrialised countries. Generally, they deal with the financing and organisation of health services. An area of special concern is the development of primary care and general practice [3,5,6]. Initially, running these projects was a new experience and the necessary knowledge had to be built up: there was no historical experience of project implementation in central and eastern Europe, where the difficulties of systemwide transformation were aggravated by a severe economic crisis [7,8].

Nowadays, health care reform projects have become more established, but until now few attempts have been made to evaluate them systematically. According to Gross, this is also true of systemwide health care reforms in other countries: only few governments have initiated structured, planned evaluations of their reform efforts. Structured evaluation of health care reforms is important, however, because it provides information on the effectiveness of the reforms as well as evidence about the need for mid-course modifications [9]. In the same way, evaluation can greatly increase the rationality and continuity of health care reform projects in several ways. It is suggested, for example, that in eastern Europe, for a number of reasons, an evolutionary approach might be more appropriate for the implementation of health care reform. Basically, this is the idea that changes are tested locally, and one step at a time, before they are introduced nationwide [5,6]. Another recommendation is that a scheme of continuous follow-up and evaluation of the results of the changes should be introduced for the sound formulation of health policy [7]. This could begin with the systematic evaluation of the health care reform projects that are taking place. At the project level, evaluation can increase the rationality in two ways. Firstly, the projects can be improved by the formulation of goals which are better attuned to one another and by better relations between goals and means. Secondly, they can be improved by enhancing the use of (the results of) past projects by sorting out the successful projects from the failures.

The issue, therefore, that we would like to explore in this paper is: to what extent can health care reform projects in CEE countries be evaluated by means of more systematic evaluation procedures? We shall answer this question on the basis of our experience in evaluating a small-scale reform project in the Slovak republic. The authors of this paper were involved in the project in the role of project assistant and project leaders at Netherlands Institute of Primary Health Care (NIVEL). During the implementation phase of the project, NIVEL, in its role of research institute, decided to take the opportunity to attach evaluation research to this project, in order to investigate the feasibility of project evaluation in eastern Europe.

The aim of this paper is to identify the difficulties experienced in the efforts towards a systematic evaluation of a reform project in eastern Europe and draw lessons from this experience for the future. The paper will be of use for consultants concerned with project implementation in central and eastern Europe, as well as for academics who want to evaluate the health care reforms. We shall start the paper with a brief outline of the health care reform projects of the Phare programme, the large EU programme for health reform in eastern Europe. This is followed by an overview of approaches of evaluation research, which culminates in a set of requirements for more rigorous, controlled evaluation. In the next section, the discrepancies between these requirements and our experiences of evaluating a health care reform project in the Slovak republic will be examined. On the basis of our experiences, and based on relevant evaluation literature, recommendations for future evaluations will be formulated.

2. EUROPEAN UNION (EU)-FUNDED HEALTH CARE REFORM PROJECTS: THE PHARE PROGRAMME

The EU has created two major programmes for aid and development for the countries of central and eastern Europe: Phare and Tacis. The Tacis programme is directed to the countries of the former Soviet Union; the Phare programme to the Baltic states, Poland, Czech republic, the Slovak republic, Hungary, Rumania, Bulgaria and Albania. In addition, there are funds and programmes established by the individual EU member states. Here we shall limit ourselves to the aims and procedures of the Phare programme, which aims to help these countries to rejoin the mainstream of European development and build closer political and economic ties with the EU.

The Phare programme was set up by the European Union in 1989, in order to support the changes taking place in Poland and Hungary. In the years that followed, more and more countries of central and eastern Europe came into the programme. Between 1990 and 1995, the EU spent ECU 5416.9 million on the Phare programme. Health care restructuring is only one of the priorities of the programme; other sectors include public administration, energy, transport and agriculture. The procedure of the Phare programme is that the EU and the partner country agree on an overall plan, in which it is decided which sectors have priority and how funds are to be spent and in which the actions for each of the sectors are outlined. This plan is called an indicative programme. Each receiving country is then responsible for running its own programmes.

The implementation of a Phare project is largely decentralised and managed by the Phare Programme Management Units (PMU), usually staffed by civil servants of the relevant Ministry in the receiving country and supported by external experts (also sponsored by Phare). PMUs are responsible for project preparation (including the tender procedure), implementation and monitoring. This includes the production of the Terms of Reference (TOR), which is a document specifying activities, outcomes and budget of a project. TORs are usually jointly produced by PMU staff and external experts, who can be hired for this purpose.

The Phare projects are usually implemented by contracted firms by means of a competitive tender. For the tender procedure normally a list is compiled of firms which have expressed their interest (on a call in the official EU Journal) and have a provable track-record in the area in question. From this long list a short list is made of firms which are invited to tender. The firm offering the best performance: price ratio will get the contract. The project is jointly carried out by a number of local experts, selected and appointed by the PMU, supported by the external experts as mentioned in the tender proposal, under the supervision of the PMU. They are together responsible for the development and implementation of the programme. The team of external experts is usually responsible for the preparation of an initial and final report. The initial report, which should be produced shortly after the start of the project, is also called inception report and outlines a detailed plan of action for the project. In the final report the activities of the project are summarised and evaluated and recommendations for the future are made. The relevant Ministry is usually the client of the project, and is then called the commissioner of the project.

Phare projects may take place in a political context which may be hard to understand and predict. The need for change can be very urgent, for instance because of sharply growing costs of health care. Under such conditions no time is allowed for well-planned actions. Furthermore, continuity of policy may be severely threatened by frequent changes among politicians and civil servants. It may be difficult to win the necessary commitment of the relevant authorities. These circumstances can hardly be anticipated in the phase of writing the tender proposal. Hence, implementing Phare health care projects includes a great deal of improvisation.

3. PRINCIPLES OF EVALUATION RESEARCH

Evaluation research has become an established subdiscipline of applied social research, with an ever-growing amount of literature and dozens of advocated approaches [11–15]. The term evaluation expresses the notion of both a description and a judgment. Social interventions, also called ‘planned human action’, are often the object of evaluation research [11,16,17]. These interventions can take the form of projects, programmes or policy measures, e.g. in the field of health, education or social welfare. In order to be able to make a judgment, criteria are needed. The traditional evaluation model takes the goals of an intervention as criteria for success.

For our evaluation, we used the conceptual model of Stake [18], an example of a goal-based approach. In this model a programme consists of a cohesive set of goals to be achieved and the means applied to reach these goals. The model also takes into account intervening conditions that might hamper the relation between the means and the goals. It contains two essential elements:

- the unravelling of a programme into goals, means and conditions
- the systematic comparison between the planned and the realised programme

In this model, there is room for in-between changes in the means that are actually realised and in the conditions that occur in practice. This dynamic element made it suitable for evaluating health care reform projects, because these are not always implemented in a rational and planned way and the conditions may be hard to predict. In short, this approach fits in with a projectwise approach to changes, but few clues for the appropriate research design are given. The distinctive evaluation approaches all advocate a specific design and methodology [11–14]. One way to order these approaches is according to the extent of control evaluation researchers have over the intervention that is being evaluated.

On the one extreme, where evaluation researchers have no control whatsoever over the conditions under which an intervention takes place, is what is commonly known as the field inspection. The other extreme is the Community Intervention Trial (CIT), which is modelled after the RCT, where evaluation researchers control the intervention to a large extent. In between are different kinds of evaluation that have in common that evaluation researchers — either by necessity or by choice — accept that the way the intervention develops is part of the process of evaluation.

The field inspection is a form of evaluation that often takes place in developing countries. It is mostly used in order to assess whether development activities have been implemented successfully and whether the planned results have been achieved. The field inspection has traits of the naturalistic evaluation model, in which the evaluator seeks understanding of the programme. For this purpose, the project is evaluated in its natural context. It is a more or less intuitive, implicit evaluation model: few elements of the evaluation design and procedures are formalised and the evaluation criteria are often not explicit. Most often, qualitative methods are used, like document study and site visits, which are characterised by observation and interviewing. The main disadvantages of the field inspection are the absence of an explicit and coherent methodology and the lack of statistical material. Also, a zero measurement, in order to assess the situation at the start of a project, is often lacking. During a field visit, time is limited and there is usually no opportunity for the structured collection of primary data. Nevertheless, the field inspection is an important tool for gaining insight into the causes of success or failure of a project and it has a distinct role in the accountability of spending public money [12,15,19].

At the other extreme figures the CIT, which is becoming an increasingly accepted paradigm for evaluating the effect of health interventions, e.g. health promotion programmes. Evaluations of these interventions are usually conducted with a high degree of control [20,21]. The CIT sets a number of evaluation requirements. In operational terms: firstly, the criteria for the evaluation should be set beforehand. If the goals of an intervention are used as criteria, they should also be clearly established before the start of the intervention. Secondly, the focus of the evaluation should be on the outcomes or the effects of an intervention. The effects should be measurable, hence the preference for quantitative methods. To a limited extent, attention should be paid to the process, in order to assess whether the intervention was carried out according to plan and whether any intervening circumstances occurred. Thirdly, in order to trace the causal relation between an intervention and its effects, i.e. to minimise any other possible explanations, a (quasi) experimental design is preferred. This involves the use of an intervention (experimental) group and a control group, and of a pretest and posttest measurement. Finally, the evaluation should be conducted by an independent outsider [20–22]. We shall call this the rigorous evaluation approach, because of its aim towards methodologically sound research.

We are aware that the rigorous evaluation approach is by no means uncontroversial. Design and methodology issues often lead to a (unnecessary) philosophy of science paradigms debate of the ‘hard’ sciences paradigm versus the more naturalistic one [15,17,22]. Still, we start from the rigorous evaluation model, because of its methodologically sound foundations: these will give more persuasive power and more guarantees that the evaluation findings can — to some degree — be generalised. This seems especially important in view of the utilisation of evaluation results in a wider context, e.g. for decisions about (implementation of) future reforms or for the formulation of health policy. Therefore,

although we realise that transition projects are not controlled situations, we want to find out to what extent a more rigorous evaluation approach than a field inspection is feasible in such a context. For this purpose, we shall compare, in the next section, the rigorous evaluation requirements with the evaluation practice experienced in the Slovak republic.

4. THE PHARE PROJECT 'PRIMARY HEALTH CARE FINANCING IN THE SLOVAK REPUBLIC'

The experiences on which we base this article come from the Phare project 'primary health care financing in the Slovak republic', which was carried out from January until November 1997 by a multidisciplinary team of Dutch (external) experts, directed by NIVEL, and in close collaboration with a team of Slovak (local) experts. The aim of this project was to contribute to the quality and cost-effectiveness of the primary health care services (the focus was on the general practitioners). In order to achieve this, a number of goals were set in four project areas: remuneration, pharmaceutical cost containment, quality of care and the information system. An overview of the main problems and the operational goals of each of the four project areas is given in box 1.

Box 1: The Phare project 'primary health care financing in the Slovak republic' problems, project areas and operational goals

The general aim of the project

Contributing to the improvement of cost-effectiveness and quality of primary health care services in the Slovak Republic.

The problems, project areas and operational goals of the project

Area 1 :	Remuneration system and financial control
Main problems:	Low income of primary care doctors Lack of positive incentives for quality of care
Goal 1:	Changing the doctors' remuneration system in order to provide proper incentives for quality of care and cost-effectiveness
Area 2 :	Pharmaceutical cost containment
Main problems:	Explosive growth of pharmaceutical prescriptions The high proportion of the health budget spent on drugs
Goal 2:	Providing incentives and other conditions for doctors in order to stimulate cost-effective prescribing of drugs that contributes to cost containment in pharmaceutical care
Area 3 :	Quality of care
Main problems:	Former structure for continuing medical education no longer adequate Need for modern methods for quality improvement
Goal 3:	Introducing modern methods of peer review and guideline development and making recommendations for an infrastructure for quality assurance
Area 4 :	Information system
Main problems:	Lack of information on process and outcome in primary care The data that are collected at the moment are rarely fed back to the doctors themselves
Goal 4:	Improving the availability of collected practice data for the benefit of the project activities and formulating future options for a primary care information system
Source:	[10]

The first three of these areas presuppose the availability of information or generate information themselves. It was just this information that could also be used in evaluating the project. The new remuneration system was to contain incentives for cost-effective general practitioner's care. These incentives should be based on targets for specified medical services, e.g. referrals and prescriptions. Data were needed to provide baseline information for these targets. For this purpose, a data collection was initiated in the practices of the members of the peer review group (see below), in different parts of the country. They provided insight into the professional behaviour of GPs and provided information about contacts, diagnoses, prescriptions, referrals and orders for laboratory tests. The data were also

used for the other project areas: for pharmaceutical cost containment, they provided evidence about possible savings in the prescriptions of pharmaceuticals; for quality of care, they gave feedback to the peer review group on their professional behaviour. In the area quality of care, a peer review group of eight GPs was started. Coached by two external experts, participants learned to reflect on quality by making explicit their practice routines and to give each other feedback in a structured way. As a way of 'learning by doing', the group developed a Slovak version of a professional guideline. For more detailed information about the project, we refer to an earlier published report, which provides an overview of the process of implementation and the outcomes of the project. In a separate chapter a systematic evaluation of the project is reported [10].

5. DISCREPANCIES BETWEEN IDEAL AND PRACTICE: THE CASE OF PHARE SLOVAK REPUBLIC

5.1. Criteria for success

When the goals of an intervention are used as criteria for success, they should be set before the start of the project. In our project, the Terms of Reference could be used for the determination of the goals, but it was troublesome to establish clear goals in advance. In this respect, three points can be raised.

First, it was difficult to formulate final, specific goals at the beginning of the project. Planned goals and activities were formulated at the start, but the content of these activities was still rather general: it referred more to the process aspects inherent in any process of change, e.g. reviewing the current system, formulating recommendations, implementation and evaluation, than to specific and detailed measures. The reason was that it was difficult to identify, at the outset, the exact content of the measures to be taken. Sometimes the commissioner of the project does not know in detail what it wants to be done in the project. In fact, finding out which specific solutions should be taken in the Slovak republic — a joint task for local and external experts — was the aim of the project; in other words, the project was largely programme development.

Secondly, it can be questioned to what extent the formulated goals were realistic. It was clear for the expert team from the moment the Terms of Reference were issued that the project goals were formulated too ambitiously: in the way they were formulated and in view of the time span available for implementation of the project, the goals could not be attained. Nevertheless, the local and external experts agreed that — for political reasons — it was necessary to formulate the goals in this way, although the Slovak Ministry of Health, as commissioner of the Phare project, had the final say about the project goals. It appeared as if, by putting forward ambitious goals, the Slovak government wanted to show its commitment to the realisation of changes in the health care sector.

Finally, a health care reform project is not a static situation, but typically evolves. This also applies to the goals of the project: the goals of a project often change over time, when they are adapted to changed circumstances, e.g. when new knowledge becomes available or when impossible barriers occur. In the project, this was readily visible in the area of pharmaceutical cost containment, where the initial focus was on the reduction of pharmaceutical prescribing by GPs. One of the project activities was reviewing the current system of distribution, pricing and prescribing of drugs and it turned out that many more factors than the prescription behaviour of GPs contribute to the high pharmaceutical expenditure. In particular the pharmacists had, owing to inadequate legislation, an interest in delivering the most expensive drug. The goal was subsequently adapted to address this situation, but these problems were eventually considered to be beyond the scope of the project. It was then decided to limit the project to the GP's role in pharmaceutical cost containment, but this had obvious implications for the original goal, so the goal was changed again.

5.2. Effect evaluation

In the rigorous evaluation approach, the focus is on the measurable effects of the intervention. Regarding this requirement, two issues need to be considered. Firstly, process and outcome cannot always be distinguished. Very often, something which is regarded as 'process' in the rigorous approach is also part of the outcome. For example, in the project the process of learning the method of peer review was part of the process, but at the same time it was also an important outcome in the field of quality of care: peer review as a decentralised, bottom-up activity, was a completely new phenomenon in Slovak health care.

Secondly, the results of the project could not always easily be measured. The project had many less tangible outcomes, which was related to the nature of the goals, which concerned adjustment or development rather than 'simply' reduction or increase. For example, the result of 'raising awareness' was an important outcome in the working areas of information system and pharmaceutical cost containment: awareness of the importance of a carefully planned information system in running efficient primary health care services and in strengthening the position of general practice; and gaining insight into the complexity of the problem of pharmaceutical cost containment and of the failing current price mechanism in the Slovak republic. Awareness is a necessary precondition for successful future changes, but is hard to measure in a clear way or even make tangible.

5.3. (Quasi) experimental design

In the rigorous evaluation approach, a (quasi) experimental design should be used. The (quasi) experimental design, with its focus on measurable effects, is often accompanied by quantitative methods. Quantitative methods, however, need a high degree of structuring in advance and in the case of our project, and probably more projects in the CCEE, this was hard to achieve, because of the turbulent circumstances that surrounded it and the unexpected and unplanned activities that took place.

In former days, the old Communist regimes in eastern Europe used to withhold all sorts of information, and in many fields there is still a lack of reliable data. A tradition and infrastructure for data collection is largely absent, which is a major hindrance, not only for the evaluation, but also for the implementation of these projects. In our project, it meant that the infrastructure needed for an early and immediate zero measurement was lacking. In fact, one of the aims of the project was to help to establish an infrastructure for proper data collection. The 'basis' for data collection had to be prepared first, so it was not until a few months after the start of the project that some data could be collected.

Even 'small-scale' data collection was difficult and hard to achieve, because of logistical and practical problems. For example, the description of variables and of the codings of pharmaceuticals that was available in the data set (among other things needed for the area of pharmaceutical cost containment), was too detailed and had to be adapted. The Slovak pharmacist that was needed to do this could not be found and the job was finally done in the Netherlands. Moreover, there was little experience on the Slovak side with the obstacles of data collection and analysis.

5.4. External evaluator

Preferably, the evaluator is external. In the practice of project evaluation in foreign countries, this usually means that the evaluation is conducted by an expert from one of the donor countries, which was also the case in our project. In general, an external evaluation poses certain requirements for the organisation, reporting and management of a project: it needs to be structured and made explicit much more beforehand than we experienced in the project described in this paper. This applies for example to the (progress) reports, memos and work plans produced by the local and external experts. In the implementation process, many things were arranged in an informal way and decisions did not always seem to be the result of a rational process. In the project in the Slovak republic, for example, the GPs were selected for participation in the peer review group on the basis of the practice information system they used, rather than on the basis of their affinity with quality issues.

Another point was the nature of the implementation process of the project, and indeed of many health care reform projects in eastern Europe: the external experts, as well as the evaluator, were visiting the country (only) every once in a while, whereas project implementation was a more continuous process in the Slovak republic; the daily activities of the project were managed by local experts. Thus, it was not always possible to gain a good insight into the processes that led to the outcomes of the project. The evaluator could not always determine whether, and how, the planned means of the project were put into action. Insight into the process was further hampered by cultural and language obstacles, which were obvious barriers to complete understanding of the process. It thus remains to be seen whether the external evaluator is the best approach for a meaningful evaluation, in which justice is done to the (complex) reality that underlies such projects.

5.5. Political context

The political context is not a requirement for a rigorous evaluation approach, but nevertheless is a factor of considerable importance, so we shall touch upon it briefly. There are many political aspects inherent in health care reform projects in eastern Europe (and surely not only there). Policy decisions

have many political components and are hardly ever based on a 'rational' judgment alone; the same seems to apply to the judgment about (the success of) our project: an 'objective' evaluation judgment revealed that not all planned goals had been achieved, but the 'client', the Ministry of Health, was nevertheless satisfied with the results. It appears that a positive judgment of the project is not always directly related to the results. The commissioner of the project might have its own (political) unstated goals, which are not necessarily in conformity with the officially formulated goals. Evaluating these projects is a 'sensitive matter' in another sense too: the evaluation report usually exists alongside the final report of the project itself. The final report, apart from giving an account of the activities and outcomes of the project, almost always serves the second aim of maintaining good relations with the commissioner of the project. The conclusions of the evaluation report might very well be irreconcilable with the final report and in that case expressing the evaluation judgment is hardly appreciated. In short, political and relational motives often interfere with developmental or evaluative objectives.

The conclusion from this section is probably clear by now: it is worth striving for a rigorous, methodologically sound evaluation design, but given the nature of these projects and the context in which they take place, many times such a design will not be feasible (and not desirable!). What to do then? Taking a pragmatist point of view, we think it is better to work with an imperfect evaluation than to have no evaluation at all, which is the current situation in most health care development projects in the CCEE. At the same time, it is extremely important to make continuous efforts to accomplish a sounder research design. Therefore, we shall continue with recommendations for future evaluations in the next section.

6. LESSONS LEARNT: RECOMMENDATIONS FOR FUTURE EVALUATIONS

6.1. Criteria for success

When a project concerns intervention or programme development, as is usually the case with the reform projects in eastern Europe, there should be an explicit moment of goal determination and/or clarification. The official goals of the project are a good point of departure, but these should be used as a guideline. To improve the utilisation of evaluation findings, it is important that the criteria are credible to the stakeholders. It is thus particularly important to involve the relevant parties in this process, e.g. by holding focus groups under the guidance of the evaluator [15,23].

Posavac and Carey stress the need for multiple sources: both objective and subjective measures should be included in the criteria for success [15]. Whenever possible, after the process of goals clarification, objective outcome measures should be set. Subjective measures, like opinions of the persons involved, are also relevant, because in their judgment the external influences of the project will be taken into account. On the other hand, these opinions are obviously subjective. Paradoxically, the sources that know the programme best may be more biased, whereas those who have the least self-interest might know too little about the project. Combining objective outcome measures with subjective assessments has the advantage that these do not share the same biases; the disadvantage is that discrepancies between the objective and subjective measures might arise [15].

This discrepancy illustrates exactly the political loadedness of the whole process: different parties with different interests judge differently about the success of the project. The same phenomenon can occur when the goals are being determined, e.g. when the Ministry continues to aim at too ambitious goals. In the case of the goals, the negotiation and other strategic skills of the evaluator are vital in order to come up with an acceptable compromise; in the end, however, when no compromise can be reached, the evaluator should decide [15,23]. In the case of discrepant judgments, the situation is more complicated, but, in any case, it points very much to the relevance of establishing objective outcome measures. On the other hand, the evaluator should not necessarily try to 'reconcile' the different judgments: part of the evaluation is also to analyse these differences and explain why they exist.

6.2. Effect evaluation

The primary focus of the evaluation should be on the effects of the project, but thorough investigation of the process is equally needed, especially when the process is part of the outcome. Moreover, these projects are not implemented in a vacuum, but, on the contrary, are often highly influenced by the turbulent economic, political and social circumstances in which they take place.

Insight into the implementation process is necessary to find out which impeding circumstances influenced the outcomes of the project, and in which way. Naturally, this is always important in view of lessons learnt for future projects, but it seems even more urgent, since, as Sabbath points out, very little is in fact known about the factors affecting project implementation in the CCEE [8]. One way to include the implementation process more systematically in the evaluation is to establish, in addition to outcome measures, a set of process measures too.

Related to the focus on the process, is the choice of qualitative methods. Qualitative methods, like observation, document analysis and (semi) structured interviews are better suited to outcomes for which no standardised, validated measurement instrument is available as well as to less tangible outcomes [23–25]. The difficulties that might arise during the quantitative data collection also prompt the use of supplementary qualitative methods. We should like to stress that the ideal is not to eliminate either approach, but to integrate quantitative and qualitative methodologies [13–15].

6.3. (Quasi) experimental design

Ideally, the aim should be a research design as solid as possible, implying the use of an intervention group and a control group and pretests and posttests. A more rigorous design, however, requires a logistically more complicated data collection. Our experiences with the small scale-data collection of the project show that the everyday practice can be fraught with difficulties. Therefore, thorough reflection about the feasibility of an ambitious research design is needed. In the difficult, sometimes rather unstructured reality, it is perhaps better to start with not too many demands at a time. This means beginning with more observational designs, and gradually trying to supplement or substitute them, in the course of one or more projects, with characteristics of a more experimental design, like pretests and control groups. In this way it is possible to proceed step by step in the direction of a more systematic design.

Data collection should be an area of special concern. Great effort needs to be put into obtaining the necessary data. One solution could be to appoint an on-site field work coordinator, whose main occupation is the smooth running of data collection and analysis. Given the relevance of the availability of reliable data, it can well be defended that the project budget should include the expenses of such a coordinator. Moreover, a field work coordinator in the recipient country can not only facilitate data collection, but is also in a position to strengthen the role of monitoring in the project. More explicit monitoring is not only beneficial to the evaluability of these projects, but also has a distinct role in improving their implementation and structuration [15,19,23].

6.4. External evaluator

The question whether the evaluator should be internal or external remains a dilemma. On the one hand, we would be inclined to say that, when the requirements for organisation, management and reporting can be met, an independent and objective outsider, in the form of an external evaluator, is the best alternative. On the other hand, there are also good arguments for a more internal type of evaluator. For one thing, an internal evaluator is in a better position to gain knowledge about the actual workings of a programme [15]. This is even more so when, as was the case in the Phare project in the Slovak republic, the implementation process is characterised by many unplanned and unformalised activities.

There is another reason why a strict separation between implementation and evaluation, as the rigorous evaluation approach demands, is perhaps not so desirable. When the evaluator works together with the project team in the implementation of the project, he can play a distinct role in improving the structuration of the project, among other things by creating explicit monitoring mechanisms and keeping track of them. This is only possible, however, when the evaluator collaborates closely in the project. In this way, blurring the boundaries between the roles of participant and evaluator could increase opportunities for more systematic and controlled evaluations.

In this respect, local evaluators should play a more important role in relation to foreign experts. Not only will the local evaluator have a better knowledge of language, context and nuances, also the necessary knowledge of health systems and health services research in general, and of evaluating the effects of changes in the health care system in particular, will be built up in the country itself; both of them are very much needed in most CEE countries. In the future, non-governmental partners, like

research institutes or professional organisations, might be particularly suitable for these roles, because they are less bound by formal political restrictions and can take a more neutral position [7,8,19].

To conclude with, the recommendations given in this section are briefly summarised in box 2.

Box 2: Checklist with recommendations for future evaluations

Criteria for success

- arrange an explicit moment of goals clarification and determination
- involve the relevant parties in this process, e.g. by holding focus groups under the guidance of the evaluator
- set objective outcome measures as evaluation criteria at the start of the project, but also take into account subjective assessments

Focus of the evaluation

- pay equal attention to the results of the project and to the implementation process
- establish, in addition to outcome measures, also a set of process measures

Research design

- make a pragmatic start with a more observational design
- supplement or substitute this design gradually (in the course of one or more projects) with characteristics of a more experimental design, like pretests and control groups

Methods

- put considerable effort into setting up a (quantitative) data collection and obtaining the necessary data
- supplement quantitative data with qualitative methods, like observation, document analysis and (semi) structured interviews

Position evaluator

- consider the option of a semi-internal evaluator, who collaborates closely with the project team
- give evaluators from the recipient country a more prominent role

Everyday practice

- appoint an on-site field work coordinator for the smooth running of the data collection and analysis
- create more explicit monitoring mechanisms and keep track of them (this could be a task for either the field work coordinator or the semi-internal evaluator)

7. CONCLUSIONS

In this paper we have argued in favour of evaluation of health care development projects in the CCEE. Up till now evaluation is not a priority in these projects. However, evaluation is an important instrument for increasing the continuity and rationality of health care reforms in general and of reform projects in particular. If it is acknowledged that evaluation should have a higher priority, it is important to realise that methodologically rigorous evaluation is often not (yet) possible. Although the more rigorous approaches to evaluation are preferable from a scientific point of view, in the practical situation (economical, political, etc.) of the CCEE, they are usually many steps too far. This paper argues first that this is no reason to refrain from evaluation at all and secondly, that one should gradually aim at methodologically better evaluation methods. The experiences of Gross in the evaluation of the Israeli health care reform show that systematic evaluation is possible and that a structured evaluation can actually contribute to the shaping of the reforms [9].

In this paper we have examined the discrepancies between the requirements of a formalised and controlled evaluation approach and our experience of evaluating a health care reform project in the Slovak republic. Our experiences show that evaluation necessarily implies tradeoffs: the challenge in these projects for the evaluator is to find an acceptable compromise between the methodological and scientific requirements of an evaluation, understood as systematic research into the effects of an intervention, and the dynamic practice. When evaluating health care reform projects in eastern Europe, it is essential to take the context in which these projects take place into account: often essential prerequisites for success, like sound preparation, efficient working time and good structuration in

advance are lacking. This implies that — most of the times — it will not be possible to use an ambitious research design, and concessions need to be made. At the same time, it is vital that continuous efforts towards more systematic evaluation procedures are made. Just as it is suggested to implement health care reforms in central and eastern Europe in an incremental way, for evaluation too a step-by-step approach could be the wisest option. Given the context, it is probably more realistic *and* sustainable to create or extend opportunities for systematic evaluation gradually. Only by integrating evaluation and monitoring better in the implementation of health care reform projects is it possible to truly ‘learn from doing’.

In concluding, we should like to discuss the role of the European Union and the spin-off health care reform projects may have. As already mentioned, apart from the goals of an individual project, Phare projects usually also have the broader aim of strengthening the political and economic bonds between the European Union and the former Communist countries. This implies that there are several levels at which one can judge the success of a project. It could mean, for example, that a project has failed at the project level, in the sense of not having achieved its goals, but at the same time, it might have been successful at this ‘higher’ level. The EU will certainly also be interested in these ‘political’ outcomes. Another form of spin-off is the development of a research culture in the countries of central and eastern Europe. The EU can help to start this process by earmarking funds for evaluation within the budget of individual projects and/or by making the evaluation of a project a prerequisite for financing it. In this way, as a direct effect, it will be possible to learn from the success and failures of very different projects, and as an indirect effect, local research expertise will be built up. There is certainly an important role in this area for the European Union.

REFERENCES

- [1] Ringen S, Wallace C, editors. *Societies in Transition: East-Central Europe Today*. Aldershot: Avebury, 1994.
- [2] Ganzeboom HBG, editor. *Workshop Transformation Processes in Eastern Europe*. The Hague: ESR, 1996.
- [3] Marrée J, Groenewegen PP. *Back to Bismarck: Eastern European Health Care Systems in Transition*. Aldershot: Avebury, 1997.
- [4] McKee M. Health services in central and eastern Europe: past problems and future prospects. *Journal of Epidemiology and Community Health* 1991;45:260–5.
- [5] Saltman R, Figueras J. *European Health Care Reform: Analysis of Current Strategies*. Copenhagen: WHO/Regional Office for Europe, 1997.
- [6] Boerma WGW, Fleming DM. *The Role of General Practice in Primary Health Care*. St Crispins: The Stationery Office/WHO Regional Office for Europe, 1998.
- [7] Zarkovic G, Satzinger W. Politics and foreign involvement in reforming the health care systems of former socialist countries. In: Altenstetter C, Björkman JW, editors. *Health Policy Reform, National Variations and Globalization*. Basingstoke: Macmillan Press, 1997.
- [8] Sabbat J. International assistance and health care reform in Poland: barriers to project development and implementation. *Health Policy* 1997;41:207–27.
- [9] Gross R, Rosen B, Chinitz D. Evaluating the Israeli health care reform: strategy, challenges and lessons. *Health Policy* 1998;45:99–117.
- [10] Boerma WGW, Grielen SJ. *Groundwork for Primary Care in Slovakia: Report from an EC/Phare Project*. Utrecht: NIVEL, 1998.
- [11] Rossi PH, Freeman HE. *Evaluation: A Systematic Approach*, 5th edition. London: Sage, 1993:1979.
- [12] House ER. *Evaluating with Validity*. London: Sage, 1980.
- [13] Herman JL, Lyons Morris L, Taylor Fitz-Gibbon C. *Evaluators Handbook*. Beverly Hills, CA: Sage, 1987.
- [14] Stecher BM, Davis WA. *How to Focus an Evaluation*. Newbury Park, London: Sage, 1987.
- [15] Posavac EJ, Carey RG. *Program Evaluation: Methods and Case Studies*, 5th edition. Upper Saddle River, NJ: Prentice-Hall, 1997.
- [16] Berk RA, Rossi PH. *Thinking About Program Evaluation*. Newbury Park, London: Sage, 1990.
- [17] Shadish WR, Cook TD, Leviton LC. *Foundations of Program Evaluation: Theories of Practice*. Newbury Park, London: Sage, 1991.
- [18] Stake RE. The countenance of educational evaluation. *Teachers College Record* 1967;68(7):523–40.

- [19] Netherlands Development Cooperation, Operations Review Unit. Evaluation and Monitoring: The Role of Project Evaluation and Monitoring in Netherlands Bilateral Aid. The Hague: Ministry of Foreign Affairs, DGIS, 1995.
- [20] Mackenbach J, Gunning-Scheepers L. Het evalueren van interventies gericht op het terugdringen van sociaal-economische gezondheidsverschillen (Evaluating Interventions Aimed at Reducing Socio-Economic Health Differences). The Hague: Zorg Onderzoek Nederland, 1995.
- [21] Koepsell TD, et al. Selected methodological issues in evaluating community-based health promotion and disease prevention programs. Annual Review of Public Health 1992;13:31–57.
- [22] Taylor Fitz-Gibbon C, Lyons Morris L. How to Design a Program Evaluation. Newbury Park, London: Sage, 1987.
- [23] Patton MQ. Practical Evaluation. Newbury Park, London: Sage, 1982.
- [24] Patton MQ. Qualitative Evaluation and Research Methods. Newbury Park, London: Sage, 1990.
- [25] Guba EG, Lincoln YS. Effective Evaluation: Improving the Usefulness of Evaluation Results through Responsive and Naturalistic Approaches. San Francisco: Jossey-Bass, 1981.