# CPB Document

**No 48b**
April, 2004

## Towards evidence based policy

**M.F. Cornet and H.D. Webbink**

CPB Netherlands Bureau for Economic Policy Analysis
Van Stolkweg 14
P.O. Box 80510
2508 GM  The Hague, the Netherlands

Telephone +31 70 338 33 80
Telefax +31 70 338 33 50
Internet www.cpb.nl

ISBN 90-5833-156-3

# Contents

# Preface

The interest in evidence-based policy has been growing steadily in recent years. Controlled and natural experiments with policy instruments, with the associated evaluations, can provide a strong empirical foundation for existing and new policy instruments. This report gives an overview of national and international experiences with policy experiments and considers possible applications in the Netherlands.

The report has been written by Maarten Cornet and Dinand Webbink at the request of the Ministry of Economic Affairs, in preparation for a collection of papers which will offer an interdisciplinary perspective on market regulation.

F.J.H. Don
Director, Netherlands Bureau for Economic Policy Analysis (CPB)

# Summary

The effects of policy are often not known. This is true for both long-term policies and newly implemented policies. The costs of pursuing and continuing ineffective policies can be very high. Experiments can yield convincing evidence of the effects of policy. The benefits of experiments can therefore be very high. The opportunities to gather knowledge with the help of policy experiments have as yet not been widely exploited in the Netherlands.

Two types of experiments can be distinguished: controlled and natural experiments. In a controlled experiment a researcher creates a randomly selected experimental group and a control group. The experimental group is treated with a certain policy intervention, the control group is not. The difference in results between the two groups can only be attributed to the policy intervention. In a natural experiment the experimental group and the control group are formed by a coincidental situation in reality, for instance elements in regulations or changes in populations.

Controlled experiments often encounter ethical objections in the Netherlands, because they involve unequal treatment of people. On the other hand, all future generations can benefit from the knowledge acquired in a controlled experiment. The alternative of 'not experimenting' carries the risk that ineffective or harmful policies will be continued. Controlled experiments have been used in the development of new medical drugs for many years.

Controlled experiments can be applied in many areas. They can yield advantages especially in recurrent policy issues, i.e. those policy issues where the effects of policy instruments have been debated for a long time. Examples are financial incentives to encourage participation in university-level technical courses, incentives for the reintegration of unemployed people into the labour market, R&D-subsidies, and rules concerning market operation and innovation.

Natural experiments are a relatively cheap source of knowledge about the effects of policy. A welcome side effect is that the results become available relatively quickly. However, natural experiments are based on coincidental circumstances which do not occur in all policy areas.

The design of policy can be a first step towards the evaluation of policy. For now, various opportunities remain unused, however. Thus greater use could be made of the random allocation of subsidies, for instance in situations where the demand for subsidies exceeds the available budget. By allocating subsidies randomly, an experimental group and a control group are formed as the basis for future evaluation. Another possibility is the deliberate incorporation of discontinuities in subsidy schemes. For instance, in the Netherlands in recent years additional funds have been allocated for staff and computers to schools with at least 70% pupils from disadvantaged backgrounds. Schools just below this limit do not receive an additional subsidy. A cut-off point of this kind makes an evaluation possible. By comparing schools around this cut-off point, a real experiment can be imitated. By incorporating such discontinuities into schemes more frequently, opportunities arise to determine the effects of policy.

Pilot studies are widely used in the Netherlands. Pilots usually do not have a control group and thus differ from policy experiments. The lack of control group makes it difficult to determine the effects. In the case of pilots, paying systematic attention to looking for a randomly selected control group can yield more convincing evidence.

# 1 Introduction

What are the effects of policy? This obvious question is often difficult to answer. This is true for both long-term policies and newly implemented policies. Thus, in various policy areas large amounts of money have been spent for some time without the effects being known. An example is policies aimed at overcoming underachievement in education. The Netherlands Court of Audit (Algemene Rekenkamer) has concluded on the basis of 35 evaluation studies that no convincing evidence was available on the correlation between the performances of pupils from disadvantaged backgrounds and specific policies.[1] Other examples are subsidies and tax breaks for innovation, technology, the environment, energy or exports, computers in education and training for the unemployed. Recently less encouraging experiences have also been gained with large-scale policy changes, for instance the deregulation of the taxi market or the introduction of the concept of independent study (*studiehuis*). Of course these changes were not introduced without preparation, but there was no strong empirical foundation for these policies either.

Against this background it is understandable that the interest in evidence-based policy has been growing steadily in recent years. The introduction of the 'budgetary policy to budgetary accountability' (VBTB) process can be seen as an illustration of this. The Rekenkamer's recent report on the implementation of policy underlines that much remains to be done in the area of policy evaluation.[2] Furthermore, several reports have been published in the Netherlands in recent years which call for experimentation with policy.[3] For example, CPB argues in a recent study for the conduct of policy experiments as a means of acquiring knowledge about the effects of knowledge policy.[4]

Knowledge about the effects of a policy will strengthen that policy. Policies that do not work or actually do harm can be stopped. Policies that work can be introduced across the board. Because the social costs of ineffective policies can be very high, the social benefits of knowledge about the effects of policies can also be very high.

Policy experiments can yield convincing knowledge about the effects of policy. Hence the social benefits of policy experiments are potentially great.

In this contribution we will consider the advantages and disadvantages of experimenting with policies. Section 2 discusses the policy evaluation problem: why experiment? This section also defines two types of experiment: the controlled policy experiment and the natural policy experiment. Section 3 discusses the experiences with controlled policy experiments; section 4 those with natural policy experiments. Section 5 compares the advantages and disadvantages of controlled policy experiments, natural policy experiments and non-experimental evaluation methods. Section 6 describes the opportunities for applying experiments in the Dutch context. Section 7 sets out conclusions.

[1] Algemene Rekenkamer (2001).
[2] Algemene Rekenkamer (2003).
[3] See e.g. WRR (2002)
[4] CPB (2002).

# 2 Policy experiments

## 2.1 Why: the policy evaluation problem

Determining the effects of policy is often difficult because many factors may have an impact. Comparisons over time, before and after the introduction of the policy measure, are often inadequate because many factors can change. And a comparison between groups that are treated with the measure and those that are not usually does not immediately yield a reliable estimate of the policy effect. Researchers often only take account of a limited number of factors. The upshot is that unobserved factors can distort the results. An example of this very common evaluation problem is the determination of the effects of training for the unemployed. To determine the effect of training, the group of participants in the programme must be compared with a group of non-participants. Differences between the two groups have to be taken into account. The crucial problem then is that participants select themselves for the training. Or to put it differently, the group which applies and is willing to follow the training programme differs from the group which does not apply. A researcher can take account of some of these differences. But this is difficult for some differences, such as motivation or certain skills, because it is difficult to observe these differences. In that case it is impossible to determine whether certain outcomes after the training period, for instance finding a job more quickly, are due to the training or to differences in motivation or skills.

In formal terms, this evaluation problem can be represented as follows:

$$Y_i = \alpha + \beta X_i + \delta T + \varepsilon_i$$

where $Y$ is the outcome variable, for instance pay or period of unemployment, $X$ are personal characteristics of participant $i$ observed by the researchers, $T$ represents undergoing the intervention or otherwise, for instance participation in the training programme, $\varepsilon$ are unobserved factors and $\alpha$, $\beta$ and $\delta$ parameters to be estimated. The most interesting parameter is $\delta$, because this represents the effect of intervention. The main problem is that unobserved characteristics may be correlated with the intervention: $(\text{cov}(\varepsilon_i, T) \neq 0)$. For instance, participants in a training programme are more highly motivated than non-participants, but the researcher fails to notice this. In that case an estimate of $\delta$ will pick up not only the effect of the training but also the effect of the motivation.

## 2.2 What: controlled and natural policy experiments

In principle this evaluation problem can be solved by taking account of more factors (expanding $X$). In practice this is often not possible, however, because some factors cannot be observed (among those that spring to mind are motivation or aptitude), and because of restrictions in

terms of time and money. Econometrists have proposed several sophisticated methods for dealing with the selection problem in evaluations, such as the Heckman model[5] and propensity score matching.[6] But the problem with these methods is that they require additional assumptions. It remains difficult to prove that the observed effect is indeed the result of the intervention, and not the result of unobserved characteristics.

This explains the growing interest in the recent evaluation literature in experimental approaches, in the form of controlled experiments and above all in the form of natural experiments.

**Definition of 'controlled policy experiment'**

The researcher creates an experimental group and a control group randomly. The experimental group undergoes a certain policy intervention ('treatment'). Evaluation of the policy consists of measuring the difference in results between the two groups. This difference can only be attributed to the intervention. The difference is the effect of the policy intervention.

**Definition of 'natural policy experiment'**

A coincidental situation in reality creates an experimental group and a control group. Only the experimental group undergoes a certain policy intervention. Evaluation of the policy intervention consists of measuring the difference in results between the two groups. The difference can only be attributed to the intervention. The difference is the effect of the policy intervention.

[5] Heckman (1979).

[6] See e.g. Dehejia and Wahba (2002).

# 3 Controlled policy experiments

## 3.1 The controlled experiment

In the natural sciences and in the medical world the controlled experiment is considered the ideal method for gathering knowledge. In a medical experiment, a group of patients, for instance, is randomly assigned to two groups:

- the experimental group which receives the drug;
- the control group receives a placebo instead of the drug.

Allocating participants randomly to an experimental group and a control group sharply reduces the chances of other factors besides the intended intervention influencing the results. The difference in health between the two groups can thus be attributed with a large degree of certainty to the drug. To further increase this certainty, the experiment is often conducted 'double blind'. This means that neither the participants nor the researchers know which group is the experimental group and which the control group.

In the social sciences, including economics, experiments can also be an important instrument for gathering knowledge. Thus a controlled experiment can solve the selection problem for skills training programmes outlined above. By allocating participants randomly (by drawing lots, for instance) to the training group and to the control group, there is no reason to assume that the groups will differ systematically in terms of unobserved factors. Or to put it differently, the chances of a correlation between $e$ and $T$ are very small. Hence the effect of $T$ can be determined convincingly.

#### Pilots are not controlled experiments

The random allocation of participants to the experimental group and the control group is the foundation of the evidence. This means that there is a crucial difference between controlled experiments and what are called 'pilots', which are very common in the policy sphere. The evidence obtained from pilots is less convincing because it is not clear to what extent the results are caused by the intervention. It is possible, for instance, that the results of pilots are not representative because the participants are more than averagely motivated. It is also possible that participants in a pilot change spontaneously relative to the broad population as they gain more experience.

## 3.2 Advantages of controlled policy experiments

#### Controlled experiments produce convincing evidence

The major advantage of controlled experiments is that they offer convincing empirical evidence about what works and what does not. Ultimately scientists and policy makers believe that evidence obtained by experimental means is more convincing than evidence obtained in a different way. A welcome side effect is that this evidence can often be clearly communicated.

#### The benefits of experimenting can be very high

The pursuit of a policy whose outcomes are not clear can be very costly. This even more so if this policy is continued for many years. Experimenting yields knowledge about what works and what does not. Clearly the benefits of this can be very high.

## 3.3 Disadvantages of controlled policy experiments

#### Ethics: experimenting with people

A widely cited disadvantage of experimenting is the ethical aspect. A controlled experiment by definition involves unequal treatment of people. After all, a successful intervention may be denied to a group of people. This inequality in experimenting can put the political feasibility under pressure. Against this it can be argued that experiments with random assignment are fully accepted in the medical sciences. Why then should this be unethical in socio-economic policy making? Moreover, sometimes it is not certain in advance whether a particular intervention will actually generate the intended benefits. If the effect has been observed with a great degree of certainty, the coming generations can benefit from it, either by applying a successful intervention on a large scale or by not continuing an existing harmful policy. The alternative of ‘not experimenting’ carries the risk that ineffective or harmful policies will be continued. The objection of unequal treatment can also be mitigated by giving the experimental group additional resources and the control group what it would be entitled to under normal circumstances. Policy makers can also make participation in the experiment voluntary. But in that case it is less the effect of the policy instrument that is being studied, and more the effect of offering the policy instrument. And finally, policy makers can circumvent any political objections by encouraging third parties, for instance academic researchers, to devise controlled policy experiments rather than initiate such experiments themselves.

#### Cost and duration

A disadvantage of experimenting is that setting up, carrying out, following up and evaluating a good experiment is not straightforward and may entail considerable costs, including the costs of the measure itself. Moreover, in some cases the results will only be known after several years. Experimenting can be a long-term process, while the time horizon of many politicians is often

limited to a single government term (i.e. four years). On the other hand, however, the benefits can be very high.

#### Limited duration of experiments

Experiments have a limited duration. The direct consequence of this may be that effects which occur after the end of the experiment will not be observed. Participants may also react differently to a one-off intervention than to a permanent intervention. Incidentally, when the experiment is being devised, the experimenter can vary the duration of the intervention to evaluate its impact on the outcome variable.

#### Some questions cannot be answered properly with experiments

Experiments are suited above all to answering questions at the micro level, for instance about the effect of an income tax cut on the labour supply. How these changes pan out at the macro level, that is to say, what general equilibrium effects will occur, is difficult to determine with an experiment. In some cases it also not easy or meaningful to experiment. For instance, in practice it is not feasible to randomly allocate an increase in the minimum wage to a particular group of businesses.

#### Scaling up small-scale experiments

A recent point of criticism has been that the results of experiments do not give a reliable indication of what will happen when the intervention is applied on a larger scale. It is possible, for instance, that the target population will change when the experiment is scaled up. If skills training for people on welfare benefit were made national policy, then certain people, faced with the prospect of having to participate in this training programme, may decide to opt out of welfare support. The upshot may be that the population of the experiment diverges from the population when the experiment is scaled up. This problem also arises in all non-experimental techniques, incidentally.

#### Not all persons in the experimental group participate in the intervention

In many experiments not all people in the experimental group are prepared to undergo the intervention, for instance participation in a training programme. Consequently the experiment will not measure the effect of participating in the training but the effect of offering the training. In that case additional assumptions are often necessary to determine the effect of participation in the training. That does not need to be a serious problem, however, because policy makers are often interested precisely in the effect of offering an intervention.

#### Spill-overs: the experimental group influences the control group

For many reasons, spill-over effects may occur between the experimental group and the control group. Thus in the evaluation of the additional training allowance for employees aged 40 years

or older, it was found that the control group of 39-year-olds postponed participation in training.[7] Another example concerns the prevention of HIV infection.[8] In this case the infection in the control group may depend on which part of the population is being considered. It is also known that spill-over effects in the context of R&D subsidies can distort the results of the control group. A catch-22 situation may even occur here, because generating spill-overs is often the policy objective. When there is little difference in the output between the experimental group and the control group, this could mean that policy is not successful and does not generate much innovation. But it could also mean that policy is very successful and is generating much innovation with major knock-on effects.[9] The problem of spill-over effects is also found in non-experimental approaches, incidentally. It should also be noted that techniques have been developed to isolate the direct effect of the experiment from the spill-over effect.[10]

#### External validity

To what extent is an experimental result measured at time t and location A also valid elsewhere? And can an experimental result be extrapolated when the policy intensity is increased or reduced? The external validity of the results depends on the comparability of the situations. In general, results from other countries or previous periods will not automatically be valid for the Netherlands today or in the future. This problem also applies for results which have been obtained in a non-experimental way.

## 3.4 US experiences with controlled policy experiments

Across the world there are innumerable examples of policy changes and evaluations of policy changes. But there are only few examples of controlled socio-economic policy experiments. Greenberg et al. (1999) counted a total of 143 controlled experiments in the United States between 1962 and 1996, in which a total of 293 policy interventions were tested. However, the number of experiment starts per year is tending upwards. These are some of the striking aspects:

- Most experiments were aimed at socio-economically weaker groups and aim to learn about the income and employment effects of education policy and on-the-job and other training, of government help in seeking and keeping work, and of financial and material assistance. Increasingly, the effects found through an experiment are embedded in a social cost-benefit analysis. Only very few experiments were concerned with middle and high income groups or with businesses.
- Participation in experiments was broadly as frequently on a compulsory as on a voluntary basis.

[7] Leuven and Oosterbeek (2003).
[8] Philipson (2000).
[9] Klette et al. (2000).
[10] See e.g. Philipson (2000).

- The majority of experiments in the 1960s and 70s concerned new policy instruments. In the 1980s and 90s the emphasis was on experiments with changes in existing policy instruments. Few experiments were intended to evaluate existing policies. This might be due to the power of vested interests and ethical problems related to withholding existing policies from some participants.
- The costs of experiments – the costs of additional policy plus the operational costs – has been declining over the years. This is due to the growing use of administrative databases rather than experiment-specific surveys, the shorter duration of experiments or the stronger focus on short-term effects, and the reduction in the number of policy interventions per experiment. The last two explanations in particular may have the disadvantage that the experiment generates less knowledge.
- The experiments were usually commissioned by a public authority, sometimes by a non-governmental organisation and occasionally by a business.

A number of US experiments have made a demonstrable contribution to the fundamental knowledge about economic behaviour, and in some cases even to major policy changes (see table 3.1). Mandell et al. (2001) examined five controlled experiments to see how and to what extent the results of controlled experiments contributed to the policy process. They conclude that in all cases the outcomes played a role in policy decisions, sometimes as a convincing confirmation of a consensus view, and sometimes with a very heavy weight. However, the decision to deploy a policy instrument widely or not generally depends on many factors, including the results of the experiment. Furthermore, experiments inspire new experiments, incremental or otherwise. A major barrier to the utilisation of the results of experiments seems to be the time necessarily required for their preparation, execution and evaluation. Experiments are often initiated when the policy issue is 'hot' by the time the results become available, the policy agenda has often changed. This suggests that experiments are primarily of value to policy makers when they give answers to recurrent issues. An experiment may of course put an issue on the policy agenda; but in fact that did not apply in any of the five cases. Policy makers often regard the results of experiments as 'final', even when they are still heavily disputed within scientific circles. An intensive communication policy is required to inform policy makers about the contents and applicability of the main outcomes of an experiment, especially when the size of the experiment is limited.

No controlled experiment is fully controlled. The actual execution of an experiment always falls short of the ideal design. Krueger (1999) shows that the STAR experiment with class-size reduction in the US state of Tennessee was plagued by a less than wholly random distribution of pupils across experimental and control groups. When the researcher does not adjust for this, the analysis may produce a distorted picture of the actual outcome of the experiment. If the researcher does adjust for this, using the observed deviations from the ideal experimental design, then he or she can report not only the actual outcome of the experiment, but also the

effects of unintended but policy-relevant deviations from the design. Krueger uses, for instance, the switch of pupils from control to experimental groups to estimate the effect of the stability of a class population on learning performances.

**Table 3.1 Examples of experiments and their policy impact**

| Experiment | Result | Impact | Reference |
|---|---|---|---|
| Health insurance experiment | precise knowledge of the price elasticity of demand for care | policy makers can assess the health effects of changes in cover | Brook et al. (1983) Manning et al. (1987) |
| Job Training Partnership experiment | policy has positive effect on participation and income when focused on school-age children | policy makers reduce funds targeted for non-school-age children | Bloom et al. (1993) |
| Work-welfare experiments | job-oriented training and job-seeking assistance is effective | substantial impact on Family Support Act of 1998 | Manpower Demonstration Research Corp.; Gueron and Pauly (1991) |

Source: Burtless (1995), p. 69

# 4 Natural policy experiments

Setting up controlled policy experiments is sometimes difficult and expensive in terms of time and money. That is why in recent years researchers have increasingly looked for natural experiments. These are coincidental situations in reality which are comparable to a controlled experiment. In natural experiments people are also allocated randomly to an experimental group and a control group. The difference with a controlled experiment is that in this case the allocation is not made by the researcher, but is the consequence of coincidental elements in the regulations or of other political or social causes.

### Examples of natural experiments

In the Netherlands two studies have been conducted recently which use coincidental elements in the regulations. The first study evaluated the effect of the additional tax allowance for training employees aged 40 years or older.[11] Employees just below this age limit were not eligible for the tax allowance, and they formed the control group. Employees aged 40 or older were eligible for the tax allowance, and they formed the experimental group. By comparing the participation in training programmes of employees around the age of 40, an estimate could be made of the effect of the measure.

The second study evaluated a recent measure under which additional funding was allocated to schools with at least 70% pupils from disadvantaged backgrounds.[12] Schools just below this limit did not receive any additional subsidy, while schools at or above the 70% limit did. By comparing schools around the 70% limit (and taking into account the percentage of pupils from disadvantaged backgrounds), a controlled experiment could be imitated. Schools with at least 70% pupils from disadvantaged backgrounds formed the experimental group, while schools just below the 70% limit formed the control group. The effect of the additional funding on pupils' performances was analysed by using this 'regression discontinuity', as it is called.

In recent years studies using natural experiments have been conducted in many policy spheres, for instance education, health, safety and the labour market. By way of illustration, the table below shows a number of examples of applications of natural experiments.

[11] See Leuven and Oosterbeek (2003).

[12] See Leuven et al. (2003).

**Table 4.1 Examples of studies using natural experiments**

| The causal effect of | On | Natural experiment | Reference |
|---|---|---|---|
| Occupational disability insurance | Labour supply | Variations in regulations over time and by region | Gruber (2000) |
| Police deployment | Crime | Election cycles | Levitt (1997) |
| Unemployment insurance contributions | Pay/employment | Legislation in states | Anderson and Meyer (2000) |
| Heart operations | Health | Distance to hospital | McClellan et al. (1994) |
| Length of prison sentence | Employment and income | Random allocation of federal judges | Kling (1999) |
| Smoking by mother | Birth weight | Excise duty on tobacco | Evans and Ringel (1999) |
| Education | Income | Birth quarter | Angrist and Krueger (1991) |
| Education | Income | Distance to school | Card (1995) |
| Education | Income | Variations in construction of schools over time and by region | Duflo (2001) |
| Class size | Educational performance | Regulations on maximum class sizes | Angrist and Lavy (1999) |
| Student grant | Participation in higher education | Regulations on student grants | Van der Klauw (2001) |

Source: Angrist and Krueger (2002), p. 82

# 5 Comparison of controlled and natural experiments and non-experimental analyses

### Natural experiments versus controlled experiments

An advantage of natural experiments over controlled experiments is that the former often take place unnoticed by participants, supervisors or the media. In controlled experiments participants may behave differently owing to the fact that they have been selected for an experimental group (Hawthorne effects), or precisely because they have not been selected (John Henry effects; for instance, the teacher of a class which has not been allocated additional resources works even harder to show that he or she can overcome this setback (Krueger, 1999)). This can distort the results. The same applies for experiments which receive close attention in de media. This is not an issue in natural experiments. Another advantage of natural experiments is that political or ethical objections often do not play a role because the random allocation is not made by the researcher but is the consequence of other factors. The disadvantage of natural experiments is that the researcher has no control over the intervention or the target group. And sometimes additional assumptions are required. In a regression discontinuity, for instance, it is assumed that unobserved factors do not play a role with regard to the discontinuity.

### Controlled and natural experiments versus not experimenting

The key advantages and disadvantages of the various approaches are summarised in table 5.1. Controlled and natural experiments can provide convincing evidence about the effects of policy. In non-experimental analyses there is always a possibility of distortion by unobserved factors. The benefits of convincing evidence about the effects of policy can be very high. Ineffective policy can be stopped, and effective policy can be continued. Conducting a controlled experiment can be expensive and take a long time. Evaluations on the basis of natural experiments are cheaper and take less time. Evaluations on the basis of non-experimental analyses are often also cheaper and faster. In controlled experiments there is scope to direct the study. This is not the case with natural experiments because they depend on accidental circumstances. The direction of non-experimental analyses also depends on the scope for observing certain interventions and gathering data on them. The availability of data is most certain in controlled experiments.

**Table 5.1 Advantages and disadvantages of different evaluation approaches**

| Experiment | Controlled experiment | Natural experiment | Non-experimental analysis |
|---|---|---|---|
| Evidence | convincing | convincing | scope for distortion |
| Benefits | high | high | not clear |
| Costs | low/high | low | low/average |
| Duration | short/long | short | short/long |
| Adjustability | extensive | minimal | average/minimal |
| Availability of data | good | uncertain | uncertain |

# 6 Application of policy experiments in the Netherlands

Many policy questions lend themselves to experiments, but not all policy spheres are equally suitable. Experiments are suitable in particular when the focus of the study is narrow, for instance aimed at determining the effect of a clearly demarcated intervention on one or more clearly identified target variables.

To decide whether an experiment should be conducted, ideally the costs and benefits of the experiment have to be compared with alternatives. Often this will come down to assessing the question whether the greater reliability of experimental results offsets the additional cost. In the case of recurrent policy issues in particular, the value of convincing evidence from a controlled experiment seems to be considerable in relation to the costs of such an experiment.

## 6.1 Controlled experiments for recurrent policy issues

Controlled experiments are highly suitable for many policy measures in order to acquire knowledge about what works and what does not. Examples are subsidy instruments in the areas of the environment, innovation, working conditions, regional development, exports and training. Experiments are also suitable to determine the effects of regulation, such as location requirements or rules for specific professional groups such as estate agents or lawyers. But controlled experiments seem to be most suited for obtaining evidence in recurrent policy issues. By way of illustration, we will consider four such issues.

For many years now all kinds of policies have been tried to raise student participation in technical courses. Time and again the question arises to what extent additional grants for students taking technical courses can boost participation. This question can be answered by conducting a controlled experiment in which a randomly selected group of final-year secondary school pupils is offered additional student grants. Then an analysis is made of the extent to which participation in technical courses in the experimental group varies from participation in the control group.

A second example of a recurrent policy issue is what is known as the ‘poverty trap’. To what extent can financial incentives contribute to the reintegration of unemployed people into the labour market? Various experiments have been conducted in this area in the United States. As yet there have been none in the Netherlands or in other European countries.

A third example of a recurrent policy issue in which controlled experiments can bring to light significant knowledge is the additionality of R&D subsidy schemes and tax credits, such as the Dutch WBSO: how much additional R&D does an R&D subsidy induce? This question can be answered with a controlled experiment in which a randomly selected group of businesses is offered a higher R&D subsidy rate than usual. A concrete implementation of this approach can unfold as follows. The WBSO provides for three different subsidy regimes: as a company’s R&D activities increase, the marginal subsidy rate falls. In a controlled experiment a randomly

selected group of businesses which fall just inside the next, lower, subsidy regime can be offered the higher subsidy rate from the previous regime.

A fourth recurrent policy issue exists in the area of market regulation. Does lowering or raising the entry barriers improve the operation of market forces, and does it encourage innovation? Experiments with rules concerning enterprise, such as location requirements with regard to quality and environment, can throw light on this question. In an experiment these rules can be waived for a randomly selected group of regions. The experimental group and the control group can then be compared in terms of competition and innovation indicators.

## 6.2 Opportunities for evaluation through sophisticated policy design

The design of policy is a first step to determining the effects of policy. Accidental elements in regulations may create random control groups. By deliberately incorporating 'accidental elements' into schemes, it may be possible to conduct a convincing evaluation. Or to put it differently, a sophisticated design of policy will make it possible, after a period of time, to obtain hard evidence on the effects of this policy. Firstly by allocating subsidies randomly in certain cases, and secondly by deliberately incorporating differences (discontinuities) into subsidy schemes.

### 6.2.1 Random allocation of resources

As subsidies are granted, many people or organisations will not receive a subsidy. Often a control group can be formed from these groups. An example is offered by the Head Start project in the United States. This project makes resources for pre-school education available to certain target groups, especially children from poor families. Recently an advisory committee evaluating Head Start recommended that if not enough resources were available for all participants, the allocation of children should proceed randomly. This will allow a convincing evaluation of the effect of this programme at some point in the future.

Another example can be found in the area of tenders. During the design of tender regulations, a controlled experiment can be constructed. Research proposals which are just barely or not quite good enough for a subsidy can be subsidised randomly.[13] As it becomes more difficult to decide objectively which research proposals are of sufficient quality and which are not, it would also make sense to recognise this and allocate subsidies randomly.[14]

### 6.2.2 Deliberate incorporation of discontinuities into schemes

A second option is to deliberately incorporate differences (discontinuities) into subsidy schemes, in analogy with the example of the additional subsidy for schools with at least 70%

[13] See also CPB (1999) and Jaffe (2002).

[14] When numbers are small, a random allocation to the experimental group and the control group may still produce differences between the two groups. In such a case a degree of managed allocation may yield better results.

pupils from disadvantaged backgrounds mentioned above. These opportunities may arise, for instance, in different Dutch subsidies for businesses and projects,[15] including the WBSO scheme. Other opportunities may arise in the context of ICT in the education system. Major investments have been made in this area in recent years. Thus far nothing is known about the effects on pupils' performances. These effects are difficult to determine because there is no real control group. By incorporating some differences in terms of making resources available, opportunities arise to determine the effects. An example in this area can be found in a recent study by Goolsbee and Guryan (2002).

## 6.3 From pilots to controlled experiments

Pilots are used in many policy areas. The problem with them is that usually no control group is formed, and certainly no randomly selected control group. Consequently the results of pilots are less convincing. Opportunities are thus unexploited. The step from a pilot to a controlled experiment consists of looking for a randomly selected control group. By systematically paying attention to control group selection, significant gains can be made in the direction of evidence-based policy.

[15] See www.subsidieshop.nl.

# 7 Conclusions

The analysis leads to the following conclusions:

- The costs of pursuing and continuing ineffective policies can be very high.
- Controlled and natural policy experiments can yield convincing evidence about the effects of policy. The benefits of policy experiments may therefore be very high.
- Controlled experiments can yield advantages especially if they are applied to recurrent policy issues.
- Natural experiments are a relatively cheap source of evaluations. And evaluation results become available on a relatively short notice.
- The design of policy can be a first step towards the evaluation of policy. Opportunities which seem to remain unexploited at the moment include the following:
  - the random allocation of resources when demand for subsidies is too high;
  - the deliberate incorporation of discontinuities into subsidy schemes.
- In the case of pilots, paying systematic attention to looking for a randomly selected control group can yield more convincing evidence.

# References


Algemene Rekenkamer, 2001, *Bestrijding van onderwijsachterstanden*, The Hague.

Algemene Rekenkamer, 2003, *Tussen beleid en uitvoering, lessen uit recent onderzoek*, The Hague.

Angrist, L. and A.B. Krueger, 2002, Instrumental variables and the search for identification: from supply and demand to natural experiments, *Journal of Economic Perspectives*, vol. 15, no. 5, p. 69-87.

Burtless, G., 1995, The case for randomized field trials in economic and policy research, *Journal of Economic Perspectives*, Spring 1995, p. 63-84.

Brown, M.A., T.R. Curlee and S.R. Elliot, 1995, Evaluating technology innovation programs: the use of comparison groups to identify impacts, *Research Policy*, vol. 24(5), p. 669-684.

CPB, 1999, 'Het vijfde kaderprogramma voor R&D van de Europese Unie', werkdocument 111, juli 1999, Centraal Planbureau, The Hague.

CPB, 2002, *De pijlers onder de kenniseconomie: opties voor institutionele vernieuwing*, Centraal Planbureau, The Hague.

Dehejia, R.H. and S. Wahba, 2002, Propensity score matching methods for non-experimental causal studies, *The Review of Economics and Statistics*, vol. 84(1), p. 151-161.

Goolsbee, A. and J. Guryan, 2002, The impact of Internet subsidies in public schools, NBER Working Paper nr. 9090.

Greenberg D., Shroder, M. and M.Onstott, 1999, The social experiment market , *Journal of Economic Perspectives*, Summer 1999, p. 157-172.

Fraker, T. and R. Maynard, 1987, The adequacy of comparison group design for evaluations of employment-related programs, *Journal of Human Resources*, vol. 22, p. 194-227.

Heckman, J.J., 1979, Sample selection bias as a specification error, *Econometrica*, vol. 47, p. 153-161.

Jaffe, A., 2002, Building programme evaluation into the design of public research-support programmes, *Oxford Review of Economic Policy*, Spring 2002, vol. 18(1), p. 22-34.

Klette, T., J. Moen and Z. Grilliches,, 2000, Do subsidies to commercial R&D reduce market failures? Microeconomic evaluation studies, *Research Policy*, vol. 29, p. 471-495.

Leuven, E. and H. Oosterbeek, 2003, Evaluationg the effect of tax deductions on training, *Journal of Labor Economics*, forthcoming.

Leuven, E., M. Lindahl, H. Oosterbeek and D. Webbink, 2003, The effect of extra funding for disadvantaged students on achievement, University of Amsterdam, SCHOLAR working paper 39/03.

Krueger, A. B., 1999, Experimental estimates of education production functions, *Quarterly Journal of Economics*, vol. 114(2), p. 497-532.

Majone, G., 1989, *Evidence, argument, and persuasion in the policy process*, Yale University Press, New Haven.

Mandell, Greenberg, Linkz, 2001, The politics of evidence: the use of knowledge from social experiments in U.S. policymaking, working paper presented at the 2002 Annual Meeting of the American Political Science Association.

Philipson, T., 2000, External treatment effects and program implementation bias, Technical Working Paper Series T0250, NBER.

WRR, 2002, *Van oude kennis en nieuwe kennis, De gevolgen van ICT voor het kennisbeleid*, The Hague.

# Abstract

The effects of policy interventions are often unknown. Policy experiments offer opportunities to find convincing evidence about the impact of policy. This paper studies the pros and cons of controlled and natural policy experiments in the search for evidence of policy effectiveness. It concludes that controlled policy experiments are appropriate to deliver evidence on policy matters that have dragged on for a long time. Furthermore, it argues that many opportunities to provoke natural experiments in standard policy making exist, and that capture of these opportunities will increase our knowledge about policy effects importantly.