



Universiteit
Leiden
The Netherlands

Microdata and Policy Evaluation at CPB

Douven, R.; Geest, L. van; Gerritsen, S.; Jongen, E.L.W.; Lejour, A.; Crato N., Paruolo P.

Citation

Douven, R., Geest, L. van, Gerritsen, S., Jongen, E. L. W., & Lejour, A. (2018). Microdata and Policy Evaluation at CPB. In P. P. Crato N. (Ed.), *Data-Driven Policy Impact Evaluation: How Microdata is Transforming Policy Design* (pp. 203-219). Cham: Springer.
doi:10.1007/978-3-319-78461-8_13

Version: Not Applicable (or Unknown)
License: [Leiden University Non-exclusive license](#)
Downloaded from: <https://hdl.handle.net/1887/67996>

Note: To cite this publication please use the final published version (if applicable).

Microdata and Policy Evaluation at CPB



**Rudy Douven, Laura van Geest, Sander Gerritsen, Egbert Jongen,
and Arjan Lejour**

1 The Activities of CPB Netherlands Bureau for Economic Policy Analysis

CPB Netherlands Bureau for Economic Policy Analysis was established in 1945. CPB is the acronym for the Dutch name *Centraal Planbureau*. CPB is an independent economic think tank, funded by the government, and aims to provide independent, impartial economic analysis that is policy relevant and academically up to standard. CPB's main challenge is to contribute to better economic policies, predominantly through applied empirical research. Theory and good intentions alone are not enough to ensure effective and efficient policy. It is important to establish whether or not these policies will really bear fruit. According to a Dutch saying, 'measurement is the key to knowledge'.

R. Douven

CPB Netherlands Bureau of Economic Policy Analysis, The Hague, The Netherlands

Erasmus University, Rotterdam, The Netherlands

e-mail: R.C.M.H.Douven@cpb.nl

Laura van Geest · Sander Gerritsen

CPB Netherlands Bureau of Economic Policy Analysis, The Hague, The Netherlands

e-mail: L.B.J.van.Geest@cpb.nl; S.B.Gerritsen@cpb.nl

Egbert Jongen

CPB Netherlands Bureau of Economic Policy Analysis, The Hague, The Netherlands

Leiden University, Leiden, The Netherlands

e-mail: E.L.W.Jongen@cpb.nl

Arjan Lejour (✉)

CPB Netherlands Bureau of Economic Policy Analysis, The Hague, The Netherlands

Tilburg University, Tilburg, The Netherlands

e-mail: A.M.Lejour@cpb.nl

© The Author(s) 2019

N. Crato, P. Paruolo (eds.), *Data-Driven Policy Impact Evaluation*,

https://doi.org/10.1007/978-3-319-78461-8_13

Over time, research at CPB has moved from mainly macroeconomic research to a more balanced approach between macro and micro. With the increased accessibility of microdata and the rising power of computers, new research avenues have opened up. At CPB, various routes are explored to enhance evidence-based policies using microdata.

First of all, CPB conducts microeconomic analysis, over a wide range of topics including taxation, education, health care, labour market policies and wealth and income inequality. Academic papers, simulation models and policy briefs are the main outlets. CPB's value added is the focus on the Dutch context, in terms of the data used, the underlying institutions and the choice of research topics.

CPB tends to use microdata that are readily available. Occasionally, CPB collects or compiles its own microdata, undertakes surveys or conducts experiments. Statistics Netherlands, the Dutch Health Care Authority, the Tax and Customs Administration Office and the Dutch Central Bank are CPB's main data providers.

CPB sometimes advises Dutch government ministries on the design of evaluations or experiments undertaken by others: what are viable routes forward? On other occasions CPB draws up guidelines. These guidelines are often used by ministries when they tender evaluations or experiments. In this way, CPB ensures that research projects set out by ministries meet a certain minimum standard.

CPB also assesses new policy proposals, using available international microeconomic research. Governments may want an assessment of policies before their actual implementation. In the absence of reliable Dutch data, it may still be possible to make an assessment by using information from international research. CPB followed this approach, for example, when the government was contemplating various options for an increase in the minimum wage for young people.

Box 1 'Promising Policies'—A Toolbox for Policy-Makers

Aim: the series 'Promising Policies' aims to foster evidence-based policies. Insight into the effectiveness of policies helps policy-makers to make better policy choices. Academic research becomes more accessible for policy-makers in the form of a practical toolbox, drawing from existing literature or models.

Format: a typical report contains (1) an overview of the state of knowledge in a certain policy area, in terms of both outcomes and existing policies, benchmarked against international peers, and (2) a series of policy options. The options are summarized in a table. The text is geared towards policy-makers and the public at large, rather than academics.

Choice of topics: a broad range of topics, on the intersection of political priorities and expertise of the agencies, are involved. The choice of topics should be validated to avoid the appearance of political bias. The series contains reports on the labour market, innovation, education, mobility, science policy and housing.

(continued)

Choice of policies: policy options are distilled from academia, the public debate (politicians, ministries, social partners, civil society) and international peers. Reports provide an overview of options that cater to a wide political spectrum.

Choice of indicators: trade-offs are the rule in economics. The reports provide scores on various quantitative indicators and often also a qualitative indicator, enabling policy-makers to weigh the various pros and cons of a policy.

An example: the reports on labour market policies (CPB 2015, 2016a) include policy options on fiscal policy, social security, employment protection, retirement age and old age pensions, wage formation and active labour market policies. The table reports effects on budgetary impact, employment, productivity, income distribution and miscellaneous factors.

More generally, CPB tries to bridge the gap between academia and policy-makers by making academic research accessible to a broader audience. Therefore CPB uses various channels: policy reports, scientific publications and books and seminars and presentations suited for various audiences. A new series, 'Promising Policies',¹ aims to provide an overview of reform options in the Dutch situation for certain policy areas (see Box 1). This series is aimed mainly at policy-makers and provides a toolbox with many interesting and promising policy options. Most of the reports are presented in a hearing for members of parliament.

This chapter provides a short overview how CPB approaches the issue of 'better data, for better studies for better policies' using microdata. After having discussed CPB's main type of activities in this area, Sect. 2 will briefly stipulate CPB's position in the debate between the use of experimental methods and structural models. The chapter also provides some examples of recent research. It concludes with a summary of the challenges that must be faced while undertaking this type of research and how to overcome them.

2 Policy Evaluations with Microdata at CPB

2.1 *The Pros and Cons of Policy Evaluation Studies and Structural Models*

There are two major strands of literature using large micro datasets. These are policy evaluation studies and structural models. Both approaches have their strengths and weaknesses. The main issues are briefly summarized here.

¹'Promising Policies' (*Kansrijk beleid* in Dutch) is a series produced by CPB and PBL Netherlands Environmental Assessment Agency and SCP Netherlands Institute for Social Research. The publications are in Dutch.

Policy evaluation studies use randomized control experiments or ‘natural experiments’ (typically a policy intervention) as exogenous variation to estimate the impact of a particular policy (for an overview, see Angrist and Pischke 2008; Imbens and Wooldridge 2009). The basic idea is that the policy intervention cannot be influenced by the agents, which generates exogenous variation between the treated group and a control group. Popular techniques of studying policy evaluations are the use of instrumental variables, the differences-in-differences approach and regression discontinuity (Angrist and Pischke 2010). The strengths of these approaches are the identification of the causal impact of an intervention, the possibilities for replication and the relative ease of undertaking robustness analysis (Angrist and Pischke 2010). However, policy evaluation studies also have their weak points. One of these is the external validity of the effects on other environments and other target groups. Moreover, there is often a weak link between economic theory and the estimated effects. The latter point complicates interpretation of the results and often precludes welfare analysis (Heckman 2010; Keane 2010).

Structural models derive the estimating equations from economic theory and estimate policy invariant parameters, such as preference parameters. External validity is one strength of structural models. Another is the close link between theory and the estimated effects (Keane 2010). However, weak points are the identification of the causal relations, that the identifying variation is not always clear or credible and that replication and robustness checks are rather labour intensive (Angrist and Pischke 2010; Heckman, 2010).

Considering the strengths and weaknesses of both approaches, they appear to be the mirror image of each other. Hence, a fruitful way forward seems to combine the best of both worlds (Chetty 2009; Blundell 2010; Heckman 2010). In particular, in the ‘Promising Policies’ series (CPB 2015, 2016b), CPB uses the results of both strands of literature.

There are also two recent empirical approaches that formally integrate both strands of literature, starting from different strands. First, the so-called sufficient statistics approach uses economic theory to provide a theoretical underpinning of the treatment effect in policy evaluation studies and considers counterfactual policy reforms in the ‘vicinity’ of the policy reform on which the effect is estimated (Chetty 2009). The second approach is to ‘validate’ structural model with (natural) experiments (Todd and Wolpin 2006). Specifically, in this approach the authors estimate a structural model and compare the simulated effects of a policy reform with the treatment effect using policy evaluation methods.

In the following sections, this chapter considers four case studies using large micro datasets at CPB. In the first case, the authors validate a structural model for the labour participation of parents with young children, using the policy evaluation method of differences in differences. In the second case, the authors estimate the effect of tax rates on tax evasion by owners of small corporations, using discontinuities in the tax system. In the third case, the authors estimate the effect of teacher experience on children’s outcomes, using the random assignment of twins to different teachers. In the fourth case, the authors evaluate the introduction of performance-based payment schemes in the mental health-care sector.

2.2 *The Impact of Subsidies for Working Parents with Young Children on Labour Supply*

To promote the labour participation of parents with young children, governments employ a number of fiscal instruments. In the Netherlands, working parents with a youngest child up to 12 years of age receive a subsidy per hour of formal childcare and in-work benefits. Since only working parents qualify for these subsidies and benefits, they should promote labour participation of parents with young children. Unfortunately, it is largely unknown which policy works best for employment. Therefore, CPB studies the effectiveness of different fiscal stimuli in a structural empirical model of household labour supply and childcare use and validates this structural model with a policy evaluation study on a large reform in the period 2005–2009.

Bettendorf et al. (2015) use the differences-in-differences approach to estimate the effect of the 2005–2009 reform in childcare subsidies and in-work benefits for working parents. Indeed, over the period 2005–2009, there was a major increase in the childcare subsidy per hour, which halved the effective price for parents, and also a major increase in the in-work benefits for parents with a young child. In total, over the period 2004–2009, childcare expenditure increased from EUR 1 billion to EUR 3 billion, and expenditure on in-work benefits increased from EUR 0.4 billion to EUR 1.4 billion. The treatment group consists of parents with a youngest child up to 12 years of age. The control group consists of parents with a youngest child 12–17 years of age. By comparing the labour market outcomes for the treatment and control after the reform with those before the reform, Bettendorf et al. (2015) can isolate the effect of the policy intervention. They use data from the Labour Force Survey (*Enquete Beroepsbevolking*), from Statistics Netherlands, for the period 1995–2009.

Figure 1 gives a plot of the participation rate of women in the two groups, before and after the reform. The treatment and control group move (more or less) in tandem up to the reform. Considering the control group of women with a young child aged 12–17 years, Bettendorf et al. (2015) observe some convergence between the treatment group and this control group after the reform. Indeed, a regression analysis confirms that the reform had a statistically significant positive effect (an increase of 30,000 persons) on the participation rate of mothers with a young child up to 12 years old. Furthermore, Fig. 1 also shows the participation rate of a potential alternative control group, women without children. Women without children have a trend in the participation rate different from that of the treatment group and are therefore not a valid control group.

De Boer et al. (2015) use a structural model to estimate the preferences over income, leisure and childcare. They use a discrete choice model, which has become the workhorse of structural labour supply modelling (Bargain et al. 2014), and a large and rich administrative panel dataset for the Netherlands, the Labour Market Panel of Statistics Netherlands. The full dataset contains panel data for 1.2 million individuals for the period 1999–2009. The dataset combines information on income

(from various sources, e.g. labour income, profit income and various types of benefits) and working hours from the Social Statistical File (*Sociaal Statistisch Bestand*), data from the municipalities (*Gemeentelijke Basisadministratie*) on demographic characteristics (e.g. gender, age, ethnicity, ages of children, household type), data from the Labour Force Survey (*Enquete Beroepsbevolking*) on the highest completed level of education and data on the price and use of childcare per child (*Kinderopvangtoeslag*). Thanks to the size of the dataset, the authors can estimate preferences for a large number of subgroups. Furthermore, they can account for a large number of observable characteristics and look at a large number of outcomes. The childcare information is available for the period 2006–2009, which is the time span used in the estimations for the structural model. Large-scale reforms in childcare subsidies and in-work benefits during this period benefit the identification of the structural parameters.

De Boer et al. (2015) then simulate the 2005–2009 reform with the structural model and compare the simulated results with the estimated effects of the policy evaluation study for validation. The results are reported in Table 1. The top of the table gives the results for the participation rate and hours worked of mothers and fathers with a youngest child up to 3 years (pre-primary school age). The bottom of the table gives the results for the participation rate and hours worked of mothers and fathers with a youngest child aged 4 to 11 years (primary school age). Table 1 shows that the results for the structural model are in line with the results of the policy evaluation study for mothers. The estimated effect on the participation rate of fathers is again much in line with the prediction from the structural model. For the intensive margin, for fathers with a young child of primary school age, the policy evaluation study suggests a smaller negative effect on hours worked per

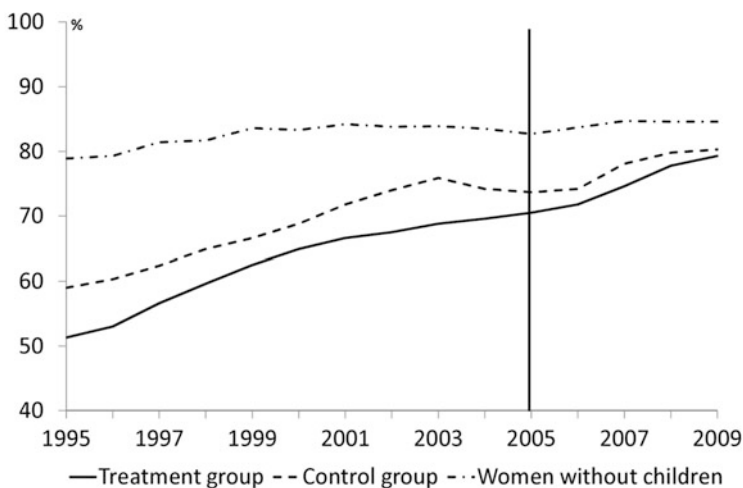


Fig. 1 Labour participation treatment and control group (differences-in-differences analysis). Source: Bettendorf et al. (2015)

Table 1 Comparison prediction structural model with policy evaluation study (differences in differences) for the 2005–2009 reform

Outcome	Structural model	Policy evaluation study (standard error in brackets)
Youngest child 0–3 years old (pre-school)		
Participation rate (mothers)	0.030	0.020 (0.007)
Hours worked per week (mothers)	1.185	1.222 (0.223)
Participation rate (fathers)	0.004	0.006 (0.004)
Hours worked per week (fathers)	0.075	−0.509 (0.237)
Youngest child 4–11 year old (primary school)		
Participation rate (mothers)	0.017	0.022 (0.007)
Hours worked per week (mothers)	0.616	0.750 (0.221)
Participation rate (fathers)	0.001	0.003 (0.004)
Hours worked per week (fathers)	−0.001	−0.180 (0.234)

Source: De Boer et al. (2015)

week than the structural model does, although the coefficients are not significantly different from each other. The only coefficient of the policy evaluation study which differs significantly from the prediction of the structural model is the hours worked response by fathers with a youngest child of pre-primary school age, for which the policy evaluation study suggests a larger negative response than the structural model.

Given that the structural model gives a good prediction of the estimated effect of the policy evaluation study, CPB uses this model to study the effectiveness of a number of counterfactual policy reforms for parents with young children. De Boer et al. (2015) find that an in-work benefit for secondary earners that increases with income is the most effective way to stimulate total hours worked. Childcare subsidies are less effective, as substitution of other nonsubsidized types of care for formal care drives up public expenditures. In-work benefits that target both primary and secondary earners are much less effective, because primary earners are rather unresponsive to financial incentives.

2.3 Tax Shifting by Owners of Small Corporations

In the Netherlands, owners of small corporations determine their own salary and the distribution of profits from their firm. This gives the owners the possibility to shift between tax bases. These owners are managing directors, abbreviated as DGAs (*directeur-groootaandeelhouder* in Dutch). DGAs thus face various types of taxation, such as taxation of corporate income, progressive taxation of labour income and proportional taxation of dividend income. As a consequence, they can exploit several opportunities to minimize the tax burden by shifting income between fiscal partners, between labour and dividend income and, in particular, over time.

From studies in other countries, it is known that the self-employed, like DGAs, are better able to avoid an increase in tax rates, because they face fewer frictions in shifting income to forms that are taxed at a lower rate (le Maire and Schjering 2013; Devereux et al. 2014; Harju and Matikka 2016).

Tax shifting can be observed only using individual tax data of the DGAs and their corporations. Bettendorf et al. (2017) use the individual tax records of DGAs and the firms they own for the years 2007 until 2011 from the Tax and Customs Administration Office. There are about 300,000 DGAs per year and these DGAs own about 200,000 firms. Over time, their numbers are increasing.

Labour and capital incomes are taxed differently according to various boxes (Cnossen and Bovenberg 2001). Labour income is taxed at a progressive rate in the first box. Four tax brackets apply, ranging from 33% for incomes up to EUR 18,000 to the top marginal rate of 52% for incomes beyond EUR 56,000 in 2011. The salary of the DGA is also taxed in the first box. Different rules govern this salary. The rules state that the salary should be at least 70% of what is ‘commonly’ paid to managing directors of similar companies, the so-called reference salary. This reference salary also has a minimum level (EUR 41,000 in 2010). Many DGAs seem to consider this level an absolute minimum, whereas the correct interpretation is that the burden of proof shifts from the tax authority to the DGA at this level.

Distributed profits of DGAs are taxed in the second box. The profits from the sale of the company, or part of its shares, are also taxed in this box at a rate of 25%. The corporate tax rate is 20% for profits up to EUR 200,000 and 25% for profits exceeding EUR 200,000. This implies that the combined corporate and dividend tax rate is typically 40–44%. This hardly differs from the 42% tariff in the second and third tax bracket of the personal tax rate. For higher incomes facing the 52% tariff in the highest tax bracket, it then becomes attractive to shift income from wage income to profit income (see Fig. 2).

Bettendorf et al. (2017) find that taxable labour income, plotted in the panel on the right-hand side of Fig. 2, bunches at the cutoffs of the tax brackets, in particular at the top tax cutoff. The McCrary estimate of the discontinuity shows that the

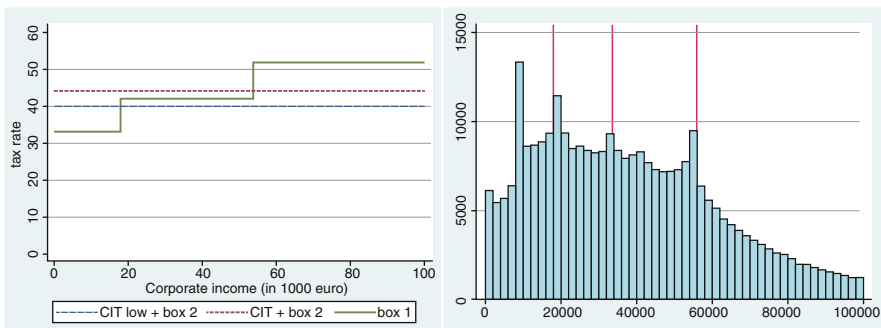


Fig. 2 Tax rate structure and distribution of wage income in the Netherlands, 2010. Source: Bettendorf et al. (2017)

density of gross wage income peaks exactly at this cutoff for all years (McCrary 2008). The elasticity of taxable income ranges from 0.06 to 0.11. Bettendorf et al. (2017) show that bunching at the top tax bracket cutoff is mainly driven by shifting income over time and to a much lesser extent by shifting between wage and profit income within a year.

The modest peak around the labour income tax base at nearly EUR 33,000 in the right-hand side figure may be surprising because, for people below the age of 65, the tax rate increases by only 0.05% moving from the second to the third tax bracket. This tiny increase in the tax rate cannot explain any bunching. However, for those aged 65 and older, the tax rate increases from 17% to 42% between the second and the third tax bracket. Some further analysis shows that a part of this peak is indeed explained by DGAs aged 65 and older. With a peak at EUR 18,000, the first tax cutoff is of nearly the same size as the last cutoff. The excess amount is somewhat smaller for this kink in the tax system than for the kink at the start of the top tax rate.

Bettendorf et al. (2017) conclude that DGAs manipulate their taxable wage and profit incomes to minimize their tax burden. In particular, they avoid the highest tax bracket. The salaries bunch just before the start of the fourth tax bracket. The effect is statistically significant, but on the other hand the economic impact in terms of forgone tax receipts is modest. The strict rules on the salaries prevent many DGAs from fixing their salaries just below the highest tax brackets. Their opportunities to determine their salaries are limited. These rules seem to achieve their goal, at least at the lower end of the wage distribution. The administrative burden is high, however, and discussions between the tax authorities and the DGAs consume time and effort. Moreover, DGAs seem to retain their profits in the firm instead of distributing them. Policy-makers seem to be aware of this situation. One of the loopholes, avoiding paying tax on dividends by emigration, has recently been closed. Other policy proposals are formulated to stimulate DGAs to distribute their profits more evenly over time, by introducing a two-rate tax structure in the second box of the personal income tax.

2.4 Teacher Quality and Student Achievement: Evidence from a Sample of Dutch Twins

The quality of teachers is considered to be a crucial factor for the production of human capital. Understanding the determinants of teacher quality is important for improving the quality of education and therefore a key issue for educational policy. A large literature has investigated the contribution of teachers to educational achievements of students (Hanushek and Rivkin 2006; Staiger and Rockoff 2010). A consistent finding in the literature is that teachers are important for student performance and that there are large differences between teachers in their impacts on achievement. However, the factors that are important for teacher quality remain

unclear. The international literature suggests that the only factor that matters is teacher experience, but evidence is scarce and there are no results for the Netherlands. Gerritsen et al. (2016) investigate the extent to which this result also holds for the Dutch context.

Addressing this question is notoriously difficult because students, teachers and resources are almost never randomly allocated between schools and classrooms. Using nonexperimental methods may therefore yield biased results. For instance, more highly educated parents may select better schools or classrooms because they may be more involved with their children than less educated parents (Clotfelter et al. 2006; Feng 2009).

Gerritsen et al. (2016) try to circumvent these selection issues by examining the effect of teacher quality on student achievement using an experimental method. They use a novel identification strategy that exploits data on pairs of twins who entered the same school but were allocated to different classrooms in an exogenous way. The variation in classroom conditions to which the twins are exposed can be considered exogenous if the assignment of twins to different classes is as good as random. In many Dutch schools, twins are assigned to different classes because an (informal) policy rule dictates that twins are not allowed to attend the same class. As a result, they go to different classrooms. Because twins are more similar than different in early childhood, it seems unlikely that small differences between twins will affect the way they are assigned to different classes. In the empirical analysis, Gerritsen et al. (2016) have tested this assumption and did not find evidence for non-randomness of the assignment.

The research is designed to study classroom quality, as twins go to different classrooms. Classroom quality is a multidimensional concept that includes factors such as peer quality, class size and teacher quality. In the empirical analysis, Gerritsen et al. (2016) focus on the effects of observed teacher characteristics on student outcomes because, in applying this design, teachers seem the most obvious factor differing across classes (Dutch schools equalize other factors such as classroom facilities and class composition across classes).

For the analyses, longitudinal data of a large representative sample of students from Dutch primary education are used. The twins are identified from the population-based sample by using information on their date of birth, family name and school from the biannual PRIMA project. This project consists of a panel of approximately 60,000 pupils in 600 schools. Participation of schools in the project is voluntary. The main sample, which includes approximately 420 schools, is representative of the Dutch student population in primary education. An additional sample includes 180 schools for the oversampling of pupils with a lower socio-economic background (the low-SES sample). Gerritsen et al. (2016) use all six waves of the PRIMA survey, including data on pupils, parents, teachers and schools from the school years 1994–1995, 1996–1997, 1998–1999, 2000–2001, 2002–2003 and 2004–2005. Within each school, pupils in grades 2, 4, 6 and 8 (average age, 6, 8, 10 and 12 years) are tested in reading and math. The scores on these tests are the main dependent variables. Information on teachers and classrooms consists of

variables such as class size and teacher experience measured in number of years working in primary education. These are the explanatory variables.

In line with earlier studies on teacher effects, Gerritsen et al. (2016) find that teacher experience is the only observed teacher characteristic that matters for student performance (Staiger and Rockoff 2010; Chetty et al. 2011; Hanushek 2011). Twins that are assigned to classes with more experienced teachers perform better in reading and math. On average, one extra year of experience raises test scores by approximately 1% of a standard deviation. In the Dutch context, this means that at the end of primary education, a pupil taught by a teacher with 40 years of experience starts on average nearly one track higher in secondary education than a pupil taught by a new teacher. The effects of teacher experience are most pronounced in kindergarten and early grades. Gerritsen et al. (2016) also find that teacher experience matters in later career stages. This finding contradicts ‘the consensus in the literature’ that only initial teacher experience (less than 3 years) matters (Staiger and Rockoff 2010). However, the findings of Gerritsen et al. (2016) are consistent with the results found by Krueger (1999) and Chetty et al. (2011), using data from the STAR experiment, in which students and teachers were randomly assigned to classes, and also with recent findings by Wiswall (2013) and Harris and Sass (2011).

2.5 Evaluation of Performance-Based Payment Schemes in Mental Health Care

In 2008, the Dutch government introduced performance-based payment schemes in Dutch curative mental health care. Since these payment schemes were new, and their impact unclear, the government decided to initially apply the performance-based payment scheme only for a small group of mental health-care providers: the self-employed. The self-employed perform about 10% of all mental health services, while large mental health institutions such as psychiatric hospitals or regional facilities for ambulatory care perform the majority. These large mental health institutions receive an annual budget, and their employees, including psychiatrists, psychologists and mental health-care nurses, receive a fixed salary.

The main idea of policy research by Douven et al. (2015) is to evaluate the performance-based payment schemes for self-employed mental health-care providers using the large mental health institutions as a control group. Douven et al. use administrative data from the Dutch health-care authority covering the years 2008 to 2010. They have information from 1.5 million treatment records, where each record describes the total treatment episode of a patient from start to end. Depending on the severity of the patient’s symptoms, a total treatment episode can take between a few hours and a whole year. Each treatment record contains detailed information about patient, treatment and provider characteristics. The records describe all curative mental health-care treatments that occurred in the Netherlands between 2008 and 2010.

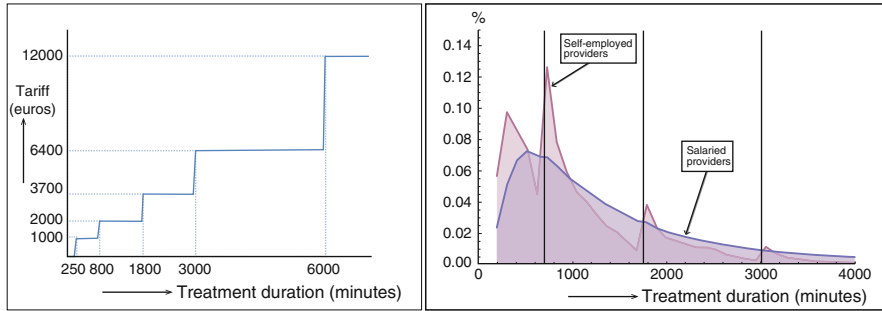


Fig. 3 Performance-based payment scheme and distribution of treatment durations. **(a)** Incremental payment function of self-employed providers (numbers on left axis are rounded off). **(b)** Treatment distribution of self-employed providers and salaried providers in large mental health institutions. Source: Douven et al. (2015)

A mental health-care provider diagnoses a patient and registers the severity of the patient's condition, the face-to-face treatment time with the patient, the type of treatment and daytime activities. The total time of a treatment is measured as a weighted sum of face-to-face time and daytime activities. The performance-based payment scheme depends on total treatment time and follows an incrementally increasing payment function. Figure 3a shows an example of the tariffs that providers receive for depression treatments. The incremental payment function jumps to a higher tariff when total treatment time passes a threshold of 250, 800, 3000 or 6000 min. Douven et al. (2015) distinguish between two types of financial incentives. First, there is an intended incentive. On the flat part of the payment scheme, a provider has no financial incentive at the margin to prolong treatment duration because his or her payment remains the same. Second, however, there is also an unintended incentive. A provider has an incentive to prolong treatment to obtain a higher financial reward. For example, at 2900 min a provider has a strong incentive to prolong treatment by 100 min to obtain a higher tariff. This almost doubles the income from that treatment, from EUR 3700 to EUR 6400.

Figure 3b shows the distribution of treatment durations for self-employed mental health-care providers and salaried providers in large mental health institutions. For the self-employed providers, bunching of treatment durations just after tariff thresholds is observed. For salaried providers, who do not get paid according to the performance-based payment scheme in Fig. 3, no bunching behaviour is visible.

Douven et al. (2015) measure both the intended and unintended incentive effects. For the intended effect, they find that treatment by self-employed providers is 2.6–5.6% shorter than that by salaried providers. However, the unintended effect of bunching around tariff thresholds is also present. Self-employed providers treat mental health-care patients 10–13% longer than salaried providers. Indeed, the unintended effect is stronger than the intended effect. Summing up all effects, Douven et al. find an increase in total costs of 2.5% to 5.3%, which is on average an increase of EUR 50–100 per treatment (where the average price of a treatment

is about EUR 2000). Since self-employed providers treated about 236,000 patients during that period, total costs have increased by EUR 12–24 million.

This research has been useful because it provides a clear demonstration that health-care providers are sensitive to financial incentives. Accounting for the behavioural responses of health-care providers is therefore an important element in the design of a payment system. In the Dutch system of so-called regulated competition, health insurance companies will discipline health-care providers. However, until 2014 health insurance companies did not have information about the exact treatment duration of health-care providers. Thus, insurers had no opportunity to perform the type of analysis conducted in this paper. This is gradually changing; since 2014, health insurers have been able to obtain exact information about treatment durations and are also becoming more financially responsible for mental health-care cost containment.

3 Challenges and Solutions

Academics consider evaluations, experiments and analysis a *sine qua non* for progress in understanding the workings of the economy. However, policies are implemented in the political arena, where different mores apply. Differences in perspective help explain why the road to evidence-based policies is a bumpy one. Here the topic is addressed from three perspectives: policy questions, data and methodology and results and policy implications.

3.1 Policy Questions

Policies are not automatically subjected to evaluation. Politicians and civil servants are not always keen to have their policies evaluated. In general, the public reception of evaluations is asymmetric. Even in the best of times, not all policies will be successful. Failures will be showcased in the media, while good results tend to be ignored. In the short run, evaluations may resemble a game which you can only lose.

Eagerness to undertake experiments varies. Experiments provide a good way to obtain a sense of what works and what does not. The expectation of success is more circumscribed. This does make experiments more attractive. However, experiments take time, and politicians can feel pressured to skip the trial-and-error phase. Experiments are also criticized on moral grounds. If the policy works, it is deemed unethical to deny citizens access. The question is, of course, whether or not this is really the case. Is it ethically responsible to expose the public at large to policies that may prove to be ineffective and sometimes costly? In medical science, experiments are daily business, and the difference in outcomes between receiving or not receiving treatment can be (very) large. Ethical standards have been developed and ethical issues are assessed by specific boards.

There are various ways to enhance the chances of this type of analysis. The law can prescribe evaluations at regular intervals. In the Netherlands, all budgetary outlays need to be reviewed every 5 years. However, rules are no guarantee for success. A box-ticking exercise should be avoided; an evaluation of the process instead of the results is a shortcut, sometimes taken to avoid difficult questions. Educating civil servants can help to overcome reluctance for a proper evaluation of the effects of a policy. Alternatively, independent organizations can be established that are free to undertake this type of research. In the Netherlands, the Court of Auditors not only covers the classic audit questions but is also tasked with an assessment of the effectiveness of budgetary outlays. CPB is a government-funded independent organization that is free to undertake whatever research it deems fit. Good relations and familiarity with ministries through other activities (in CPB's case, the regular forecasts) can also help.

3.2 Data and Methodology

Microeconomic research lives or dies by good data. This is not a hurdle to be underestimated. Data are seldom cheap, access is not always easily ensured, and privacy issues need to be addressed. The approach taken here is a pragmatic one. Data are obtained mainly from the legislative bases where they are made available (e.g. Statistics Netherlands, the Dutch Health Care Authority). This is relatively cheap and helps to overcome privacy issues.

Sometimes, CPB tries to obtain data on a more ad hoc basis. On occasion, it buys or receives data from private companies (e.g. health insurers), which can be more expensive. Legal requirements to collect data in a proper manner, whenever a major policy proposal is implemented, would be a step forward for obtaining data.

The accessibility of microdata, and in particular of administrative data, has increased substantially in the past decade. More and more frequently, Statistics Netherlands uses administrative data instead of surveys. These data are made available via remote access by ministries and academics, for research projects. Because many datasets can be linked, this provides the researchers with a large and rich set of relevant variables, with sometimes millions of observations (or more).

Academic results of policy evaluations are often derived using sophisticated methodologies and have sometimes unexpected outcomes. In particular, if these results are not intuitively clear or desired by policy-makers, it is relatively easy to blame the method or the data. Sometimes, discussions with policy-makers help to explain methodologies and results more clearly, but this is not always the case. Moreover, discussions about the internal and external validity of results (see Sect. 2.1) are frequent. In policy evaluations, it is difficult to defend claims about the general validity of the results. This is not the case with structural models, but here other issues arise. It is often argued that a particular policy instrument is not well modelled or that it misses the specificities of the instrument, for example. This is often used as an argument that the outcomes of the model do not carry over to the

policy instrument. In these discussions, it certainly helps if policy evaluations and structural models deliver the same results, as CPB's experience on the subsidies for working parents with young children show.

3.3 Results and Policy Implications

Negative results may lead to a hostile reception: 'killing the messenger' instead of dealing with the matter at hand. To make the reception of the results more effective, CPB lives by the rule 'no surprises'. Policy-makers may not like bad news, but they definitely hate surprises. A factual as opposed to a normative presentation of the results also helps.

Once evaluations or analyses have been undertaken, there is no guarantee that policies will be changed as a consequence. This frustrates many Dutch academics. What can be done? Naturally, results need to be readily available and easily accessible for policy-makers. This requires short notes, to the point and in layman's terms, with attractive infographics. More generally, a sound communication strategy will help. Moreover, persistence does pay off. The Dutch were at the forefront of analysing the consequences of ageing for public finances. Early on, it was evident that linking the retirement age to rising life expectancy would provide a strong antidote. While it took more than a decade to raise the official retirement age, when it was increased, the indexation followed shortly after. *Frappez, frappez toujours* is an important ingredient for success.

References

- Angrist J, Pischke J-S (2008) Mostly harmless econometrics: an empiricist's companion. Princeton University Press, Princeton
- Angrist J, Pischke J-S (2010) The credibility revolution in empirical economics: how better research design is taking the con out of econometrics. *J Econ Perspect* 24(2):3–30
- Bargain O, Orsini K, Peichl A (2014) Comparing labor supply elasticities in Europe and the United States: new results. *J Hum Resour* 49(3):723–838
- Bettendorf L, Jongen E, Muller P (2015) Childcare subsidies and labour supply: evidence from a large Dutch reform. *Labour Econ* 36:112–123
- Bettendorf L, Lejour A, van 't Riet M (2017) Tax bunching by the owners of small corporations. *De Economist* 165:411–438
- Blundell R (2010) Comments on: Michael P. Keane 'Structural vs. atheoretic approaches to econometrics'. *J Econ* 156:25–26
- Chetty R (2009) Sufficient statistics for welfare analysis: a bridge between structural and reduced-form methods. *Annu Rev Econ* 1:451–488
- Chetty R, Friedman JN, Hilger N et al (2011) How does your kindergarten classroom affect your earnings? Evidence from project STAR. *Q J Econ* 126(4):1593–1660
- Clofelter CT, Ladd HF, Vigdor JL (2006) Teacher–student matching and the assessment of teacher effectiveness. *J Hum Resour* 41(4):778–820

- Cnossen S, Bovenberg L (2001) Fundamental tax reform in the Netherlands. *Int Tax Public Financ* 7:471–484
- CPB (2015) *Kansrijk Arbeidsmarktbeleid (Promising Labor Market Policies)*. CPB Netherlands Bureau for Economic Policy Analysis, The Hague
- CPB (2016a) *Kansrijk Arbeidsmarktbeleid 2 (Promising Labor Market Policies 2)*. CPB Netherlands Bureau for Economic Policy Analysis, The Hague
- CPB (2016b) *Kansrijk Onderwijsbeleid (Promising Education Policies)*. CPB Netherlands Bureau for Economic Policy Analysis, The Hague
- De Boer H-W, Jongen E, Kabatek J (2015) The effectiveness of fiscal stimuli for working parents. IZA Discussion Paper No 9298. Institute for the Study of Labor, Bonn
- Devereux M, Liu L, Loretz S (2014) The elasticity of corporate taxable income: new evidence from UK tax records. *Am Econ J Econ Policy* 6(2):19–53
- Douven R, Remmerswaal M, Mosca I (2015) Unintended effects of reimbursement schedules in mental health care. *J Health Econ* 42:139–150
- Feng L (2009) Opportunity wages, classroom characteristics, and teacher mobility. *South Econ J* 75:1165–1190
- Gerritsen S, Plug E, Webbink D (2016) Teacher quality and student achievement: evidence from a sample of Dutch twins. *J Appl Econ* 32(3):643–660
- Hanushek EA (2011) The economic value of higher teacher quality. *Econ Educ Rev* 30(2):466–479
- Hanushek EA, Rivkin SG (2006) Teacher quality. In: Hanushek E, Welch F (eds) *Handbook of economics of education*, vol 2. Elsevier, Amsterdam, pp 1051–1078
- Harju J, Matikka T (2016) The elasticity of taxable income and income-shifting: what is real and what is not? *Int Tax Public Financ* 23:640–669
- Harris DN, Sass TR (2011) Teacher training, teacher quality and student achievement. *J Public Econ* 95:798–812
- Heckman J (2010) Building bridges between structural and program evaluation approaches to evaluating policy. *J Econ Lit* 48(2):356–398
- Imbens G, Wooldridge J (2009) Recent developments in the econometrics of program evaluation. *J Econ Lit* 47(1):5–86
- Keane M (2010) A structural perspective on the experimentalist school. *J Econ Perspect* 24(2):47–58
- Krueger AB (1999) Experimental estimates of education production functions. *Q J Econ* 114(2):497–532
- le Maire D, Schjerming B (2013) Tax bunching, income shifting and self-employment. *J Public Econ* 107:1–18
- McCrary J (2008) Manipulation of the running variable in the regression discontinuity design: a density test. *J Econ* 142(2):698–714
- Staiger DO, Rockoff JE (2010) Searching for effective teachers with imperfect information. *J Econ Perspect* 24(3):97–117
- Todd P, Wolpin K (2006) Using a social experiment to validate a dynamic behavioral model of child schooling and fertility: assessing the impact of a school subsidy program in Mexico. *Am Econ Rev* 96(5):1384–1417
- Wiswall M (2013) The dynamics of teacher quality. *J Public Econ* 100:61–78

Rudy Douven is a health economist at CPB Netherlands Bureau for Economic Policy Analysis and the Erasmus University Rotterdam. He graduated in mathematics at the University of Eindhoven and received his PhD in econometrics from Tilburg University. During his career he obtained 1-year fellowships at Warwick University and Harvard Medical School. His main research interest is health economics.

Laura van Geest is Director of CPB Netherlands Bureau for Economic Policy Analysis. She graduated both in economics and public administration at Erasmus University Rotterdam. Before joining CPB, she worked at the Netherlands Ministry of Finance, most recently as Director General

for the Budget and Treasurer General. Earlier in her career, she worked for the Dutch Constituency at the International Monetary Fund.

Sander Gerritsen is a researcher at CPB Netherlands Bureau for Economic Policy Analysis. He obtained his Master of Science in Econometrics from the University of Amsterdam and his PhD from Erasmus University Rotterdam. His main research interests are economics of education, labour economics and policy evaluation.

Egbert Jongen is a programme leader in labour markets at CPB Netherlands Bureau for Economic Policy Analysis and part-time associate professor at Leiden University. He obtained his MA from Maastricht University (cum laude) and his PhD from the Free University Amsterdam. His main research interests are the tax-benefit system and labour market policies. He recently headed the MICSIM project, the construction of a behavioural microsimulation model for the analysis of changes in the tax-benefit system in the Netherlands. A key element of the MICSIM project was the empirical analysis of microdata using both structural and quasi-experimental methods.

Arjan Lejour is a programme leader in tax policy and housing market at CPB Netherlands Bureau for Economic Policy Analysis and associate professor at Tilburg University. He obtained his MA from Erasmus University Rotterdam and his PhD from Tilburg University. His current research interests are international taxation of corporate profits and capital income, taxation of firms and self-employed, fiscal treatment of housing and the political economy of fiscal reforms.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

