

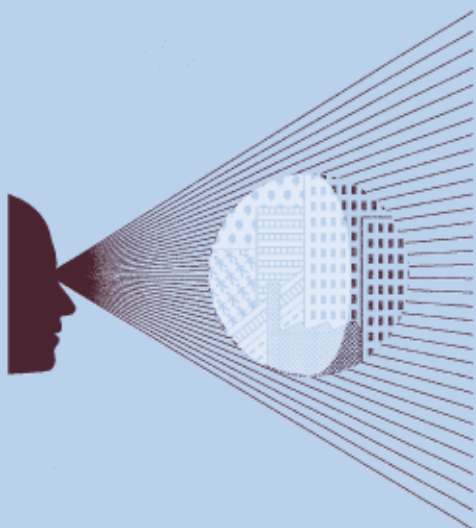
Feasibility study for potential econometric assessment of the impact of R&D tax credits on R&D expenditure

A non-technical summary

**Prepared for
HM Revenue & Customs**

September 2006

HM Revenue & Customs Research Report 19



© Crown copyright 2006.

Copyright in the typographical arrangement and design rests with the crown. This publication may be reproduced free of charge in any format or medium provided that it is reproduced accurately and not used in a misleading context. The material must be acknowledged as Crown copyright with the title and source of the publication specified.

The views expressed in this report are those of the authors and do not necessarily represent those of HM Revenue & Customs.

Published by HM Revenue & Customs.

Oxera Consulting Ltd is registered in England No. 2589629. Registered office at Park Central, 40/41 Park End Street, Oxford OX1 1JD, UK. Although every effort has been made to ensure the accuracy of the material and the integrity of the analysis presented herein, the Company accepts no liability for any actions taken on the basis of its contents.

Oxera Consulting Ltd is not licensed in the conduct of investment business as defined in the Financial Services and Markets Act 2000. Anyone considering a specific investment should consult their own broker or other investment adviser. The Company accepts no liability for any specific investment decision, which must be at the investor's own risk.

Executive summary

The introduction of incentives to encourage R&D aims to redress the failure of the markets to provide a socially optimum level of R&D expenditure. The Finance Bill 2000 introduced the first tax credit scheme in the UK for small and medium-sized enterprises (SMEs) and was supplemented in 2002 with a second scheme targeted at larger companies. The objective of the schemes is to reduce the cost of investing in R&D to firms, and increase the level of R&D closer to the social optimum.

As part of the long-term evaluation of R&D tax credits, Oxera were commissioned by HMRC to explore the feasibility of conducting econometric analysis, in particular:

- to provide HMRC with a clearer understanding of the data currently available for empirical or econometric analysis;
- to construct a matched dataset of identified R&D tax credit data sources, to ascertain if it could be used for an econometric study;
- to undertake a feasibility assessment of whether the dataset is of sufficient quality to support robust empirical analysis.

To create a dataset that includes data on whether firms have an R&D tax credit and the amount of R&D expenditure in that firm, Oxera have matched HMRC's tax return data on companies (CT600) with data from the Office of National Statistics' Business Enterprise R&D Survey (BERD). This combined dataset gives wider information on company characteristics and other relevant variables (eg, company characteristics).

To be able to match the data, the *same* company needs to appear in both databases in the *same* year. As the full impact of R&D tax credits will take time to emerge, even if all companies from BERD in *each* year appeared in the CT600 extract in the *same* year, there is unlikely *at present* to be a sufficiently long time series availability of data.

At present 1,951 observations have been matched, which represents around 10 per cent of all observations from the CT600 extract. For the majority of companies, data is available for only one year, and there are only around 100 companies with data for three consecutive years.

As such, currently there are not a sufficient number of companies over a sufficiently long time period to enable a robust econometric evaluation of R&D tax incentives in the UK to be undertaken. It is not possible, given current legislation, to obtain consecutive data over time for the smaller R&D performing companies. The current level of matching indicates the baseline level of matching that may be expected in the future. It is expected that the number of matched observations may rise over time, particularly as companies reach their six-year limit in which to submit their claims for R&D tax credits.

Contents

1	Introduction	1
1.1	Why are incentives needed?	2
1.2	How does the tax credit work?	2
1.3	Structure of the report	3
2	Assessing the effect of R&D tax incentives on R&D expenditure	4
2.1	Empirical approaches	4
2.2	Data required	6
2.3	Methodological issues	6
3	Data and matching	8
3.1	HMRC's company tax returns	8
3.2	ONS Business Enterprise R&D survey	8
3.3	Matching procedure	9
3.4	Caveat of the matching procedure	11
3.5	Results of the matching	11
4	Conclusions	14

List of tables

Table 3.1	Availability of data over time, as measured by the CRN	10
Table 3.2	Results of matching, by year	11
Table 3.3	Frequency of matched observations	12

List of figures

Figure 3.1	Overlap between the CT600 and BERD databases	9
Figure 3.2	Availability of data over time, number of companies	9

1 Introduction

R&D tax credits are a common policy tool used by governments to redress the failure of markets to provide a socially optimum level of R&D expenditure. By offering tax incentives to firms undertaking R&D, the policy aims to bridge the gap between private costs and the social return.

The Finance Bill 2000 introduced the first tax credit scheme in the UK for small and medium-sized enterprises (SMEs) and was supplemented in 2002 with a second scheme targeted at larger companies.¹ The schemes are volume-based, with higher levels of R&D expenditure leading to greater tax deductions. The objective of the schemes is to reduce the cost of investing in R&D to firms, and increase the level of R&D closer to that of the social optimum.

As part of the long-term evaluation of R&D tax credits², Oxera were commissioned by HMRC to explore the feasibility of conducting econometric analysis, in particular:

- to provide HMRC with a clearer understanding of the data currently available for empirical or econometric analysis;
- to construct a matched dataset of identified R&D tax credit data sources, to ascertain if it could be used for an econometric study;
- to undertake a feasibility assessment of whether the dataset is of sufficient quality to support robust empirical analysis.

Oxera's research comprised two technical reports; firstly, a review of existing literature; and, secondly, an exploration of the data-matching necessary to produce data for econometric analysis. Both reports are purely technical studies to inform future evaluation methods. This summary aims to summarise the main findings of the two technical reports in as non-technical way as possible.

Oxera's (2006) literature review has shown that very few empirical studies have undertaken a full cost–benefit analysis of R&D tax incentives, as a result of the complexities involved in estimating costs and benefits that arise from the introduction of R&D tax incentives.³ Instead, most empirical studies have assessed whether the amount of additional R&D expenditure generated as a result of the scheme is greater than the cost of the scheme.⁴

To assess the effectiveness of R&D tax incentive schemes in the UK, HMRC has gathered data on firms applying for a tax credit. Other data relating to R&D can be found from the Office of National Statistics' (ONS) Business Enterprise R&D Survey (BERD) on the level of R&D expenditure, employees working on R&D and sources of funding of R&D expenditure.

¹ The definition of a SME follows the approach adopted by the European Commission for state aid purposes. A SME has less than 25 per cent per cent of its capital or voting rights owned by an enterprise that is not a SME and has less than 250 employees, and either has an annual turnover that does not exceed € 40m, or an annual balance sheet total that does not exceed € 27m in the current or previous year.

² See, for example, the evaluation survey carried out in 2005 and published on the HMRC website as Research Report 12: <http://www.hmrc.gov.uk/research/index.htm>

³ Oxera (2006), 'Feasibility study for potential econometric assessment of the impact of R&D tax credits on R&D expenditure: A literature Review', report prepared for HM Revenue and Customs, September.

⁴ Hall, B. and Reenen, J., V. (2000), 'How Effective are Fiscal Incentives for R&D? A Review of the Evidence', *Research Policy*, **29**, pp. 449–69.

1.1 Why are incentives needed?

Before the policies can be assessed, it is imperative that the problem they are seeking to resolve is clearly understood. It is commonly considered that, without incentives, less R&D would be undertaken than is socially optimal (ie, a market failure would occur). The main reason for this is that any new product, process or information is a quasi-public good—ie, there is little direct link between consumption and payment. For example, once Mercedes-Benz invented the anti-lock braking system (ABS), other manufacturers were able to copy the technology with relatively little cost; soon ABS was a standard feature on many cars from other manufacturers. Economic theory suggests that a producer is less inclined to invest in R&D if the information to replicate the research is available freely in the final product than if everyone had to pay for it.

New R&D often introduces new production methods, which lower costs or introduce new products, allowing society to produce more outputs for a given level of inputs. If governments do not subsidise such research, this potential growth in productivity may be restrained. In the USA, Mansfield (1984) found that energy firms without public sector support would have funded between 3 per cent per cent and 20 per cent of the energy R&D they performed with public sector funding.⁵

The introduction of incentives to encourage R&D is based on evidence that the market fails to provide a suitable amount of R&D. There is a consensus in the literature that benefits accruing to the company from undertaking R&D—*the private return*—are substantially lower than the benefits to society—*the social return*. Therefore, in the absence of any incentives to encourage R&D, companies are likely to underinvest in R&D compared with the level of R&D investment that is optimal for society.⁶ Griffith (2000) reports evidence that private rates of return to R&D range between 10 per cent and 15 per cent, while social rates of return are often substantially higher.⁷ For example, the author reports evidence that the social rate of return to the manufacturing industry from R&D conducted by firms within the *same* industry has been found to be between 17 per cent and 34 per cent, while the social return attributable to R&D conducted in one industry, but used in another, can range from 41 per cent to 82 per cent. However, there may be a risk that the incentive could encourage non-productive R&D if the incentive scheme is too generous.

1.2 How does the tax credit work?

The R&D tax credit schemes in the UK raise the rate at which R&D expenditure can be deducted when companies calculate their profit for tax purposes from the basic 100 per cent to 150 per cent for SMEs, or 125 per cent for large companies. SMEs that make losses, and thus do not pay tax, can under certain circumstances receive cash from HMRC, known as 'payable credits'.⁸ However, companies can only claim tax credits if their R&D expenditure exceeds £10,000 per year.

⁵ Mansfield, E. (1984), 'R&D and Innovation: Some Empirical Findings', in Zvi Griliches, ed., *R&D, Patents and Productivity*. Chicago: University of Chicago Press.

⁶ Griffith, R. (2000), 'How Important is Business R&D for Economic Growth and should the Government Subsidise it?', October, The Institute for Fiscal Studies, Briefing Note No. 12.

⁷ Griffith, R. (2000), *op. cit.*, pp.3–4.

⁸ The payable tax credit could amount to £24 for every £100 of actual R&D expenditure; it applies only to the SME scheme, and not the large company scheme.

1.3 Structure of the report

This report is a non-technical summary of Oxera's literature review and matching report, which provide a more detailed review of empirical studies and the data on which an econometric evaluation of tax incentives in the UK could be based. This non-technical summary is structured as follows:

- Section 2 summarises the literature that has evaluated the effect of R&D tax incentives.
- Section 3 discusses the available data and the procedure used to construct the dataset.
- Section 4 concludes by discussing the feasibility of undertaking an econometric evaluation, based upon the current availability of data.

2 Assessing the effect of R&D tax incentives on R&D expenditure

To evaluate the effectiveness of any R&D tax credit incentives scheme, the additional R&D expenditure that is generated as a result of the scheme could be compared with the costs of such a scheme. This section summarises Oxera's literature review and discusses the approaches, data and the issues that will need to be considered in any empirical evaluation of R&D tax incentives.

2.1 Empirical approaches

Three common approaches that have been used to evaluate the impact of R&D tax incentives are:

- evaluation of the benefits and the costs of R&D investment;
- estimation of the tax price responsiveness of R&D expenditure;
- estimation of an R&D demand equation with an indicator variable to represent the introduction of R&D tax incentives.

These approaches are discussed in greater detail in the remainder of this section.

2.1.1 Evaluation of the benefits and cost of R&D investment

Theoretically, Hall (1995) suggests that the effectiveness of an R&D policy can be measured by estimating the difference between the social benefits (social return) from expenditure on R&D and the social costs of the investment, which could be measured by the loss in tax revenue.⁹ The social cost of investment in R&D captures the opportunity cost of using money in an alternative way—the projects that are foregone as a result of the government committing funds to finance the R&D tax incentives scheme. The social return represents the benefits to society that arise as a result of companies investing in R&D, such as improvements in productivity over time for the company undertaking the R&D and other companies that adopt the technology.

However, there are a number of difficulties associated with estimating benefits from expenditure on R&D. For example, it is difficult to accurately estimate the productivity gains (and the associated increase in tax revenues) that may arise over time from expenditure on R&D that is undertaken today. As a result of these complications, few studies have adopted such an approach. Instead, researchers have focused on comparing the additional R&D expenditure that arises from the tax incentive schemes with the cost of the schemes, in terms of lost tax revenues. This approach is easier to implement, but only considers the benefits that accrue to the companies that undertake the R&D. As such, this approach does not consider the wider benefits that may arise from investment in R&D, such as improvements in productivity over time for the company undertaking the R&D and other companies that later adopt the technology.

As the tax incentive schemes were introduced to overcome the perceived failure of the market in the UK to provide a socially optimum level of R&D, comparing the additional R&D

⁹ Hall, B. (1995), 'Effectiveness of Research and Experimentation Tax Credits: Critical Literature Review and Research Design.' report prepared for the Office of Technology Assessment, Congress of the USA.

expenditure with the cost of the schemes will not test whether the R&D schemes have had their intended impact on the *wider* economy, such as improvements in productivity. This caveat is also associated with the approaches that are described below in sections 2.1.2 and 2.1.3. It may not be possible to quantify the entire costs and benefits arising from the R&D schemes, particularly as very few empirical studies have sought to undertake a full cost–benefit analysis of R&D incentive schemes. For example, the social cost of R&D should capture the opportunity cost of using money in an alternative way—the benefits arising from projects that are foregone as a result of the government committing funds to finance the R&D tax incentives scheme. Therefore, it would be difficult to quantify both the wider benefits to society from the R&D schemes, as well as their full social cost.

2.1.2 Evaluation of the tax price responsiveness

Under this approach, an equation is estimated that explains R&D demand with the tax price of R&D and other non-tax determinants of R&D, such as companies' turnover. The tax price represents the implicit subsidy given by the tax system to R&D. It is included in the demand equation to elicit the responsiveness of R&D expenditure to the tax subsidy. As the responsiveness of R&D to the tax subsidy is estimated directly, this approach has the potential to achieve a high level of accuracy.

However, there are difficulties associated with the practical application of the tax price approach, which are particularly relevant if the dataset contains data on companies over time. As the cost of R&D varies through time (and across companies), controls need to be made for the impact of inflation on the R&D tax subsidy. Variables—*price deflators*—that can be applied to the tax subsidy to control for inflation only have significant variation across time, and may not be sufficiently advanced to take into account industry-specific responses to R&D tax incentives.

2.1.3 R&D demand equation with an indicator variable

An alternative to including the value of the tax subsidy in the R&D demand equation, which is often not readily available, is to include an indicator variable to represent the introduction of tax incentives in the demand equation. The indicator variable will be set to zero before the introduction of the tax incentives, and thereafter will equal one.¹⁰ As such, the indicator variable will capture the change in the effective price of R&D.

This approach may not accurately estimate the impact of the tax credit incentives, as firms may face different levels of tax credits over time. For example, if the subsidy given by the tax system to R&D increases in the second year of the scheme, this approach will not be able to capture this change. If the availability of the credit varies over time, spurious conclusions about the effectiveness of the tax incentives could be obtained. However, as the indicator variable approach avoids the need to estimate the actual level of the tax credit subsidy for each firm, it is often easier to implement than the other approaches, as a result of data availability.

2.1.4 Summary of the approaches

Very few studies have sought to undertake a full cost–benefit analysis of R&D tax credit incentives, as it is difficult to quantify the entire costs and benefits arising from the scheme. Of the remaining approaches, the tax price responsiveness approach is often preferred on theoretical grounds; however, there are difficulties associated with its practical implementation. In contrast, the indicator variable approach is easier to implement, as it avoids the need to estimate the actual level of the tax subsidy for each firm, although it may

¹⁰ If R&D tax incentives are subsequently withdrawn, the indicator variable will change to equal zero.

be less accurate. All three approaches require that other factors that may influence R&D spending, such as economic growth, are included in the models. If these factors are omitted, estimates of the impact of the tax incentives could be biased.

2.2 Data required

In advising future work, Hall (1995) noted:

the absence of good econometric studies using what appears to be the most appropriate dataset: the individual corporate tax returns used by Altshuler (1989).¹¹

Using tax return data, such as that which is found in the CT600 database of companies' tax returns, enables firm-specific estimation of the effects of the incentive policies, which suggests that the CT600 database is the appropriate starting point for any future analysis.¹²

HMRC's database contains information on applications for R&D tax incentives that were submitted, rather than those applications that were approved. The accuracy of any evaluation of the effectiveness of tax credits would be improved if the dataset identified those applications that were approved. Ideally, an evaluation of R&D tax incentives should be based on consecutive data for the same companies over time—a panel dataset. Panel data econometric techniques could then be applied to evaluate the impact of the tax incentives.

2.3 Methodological issues

There are several specific methodological issues that should be considered in any quantitative analysis of the effect of R&D incentives on R&D expenditure.

2.3.1 Causality

When modelling R&D expenditure, it is not always clear in which direction the causality flows. For example, while the level of turnover may affect the level of R&D expenditure, it is also conceivable that the level of R&D expenditure influences a firm's turnover. Econometric techniques, such as instrumental variables¹³ and Arellano and Bond's (1991) estimator,¹⁴ have been used to mitigate the impact of concerns about the direction of causality. However, restrictions on the availability of data may prevent the use of an instrumental variables approach.

2.3.2 High adjustment costs

Several authors, such as Hall (1992), suggest that firms investing in R&D face high costs when changing the level of their R&D expenditure (such as closing down an R&D department or building).¹⁵ This manifests itself in a low short-term responsiveness of R&D demand to changes in prices. This implies that tax changes, which are seen to be temporary or likely to change, will be unlikely to have a large impact on R&D expenditure.

¹¹ Hall, B. (1995), op.cit.; and Altschuler, R. (1989), 'A Dynamic Analysis of the Research and Experimentation Credit', *National Tax Journal* XLI 1989, pp.453–66. ([http://ntj.tax.org/wwtax/ntjrec.nsf/5A0C0247A873C2838525686C00686D08/\\$FILE/v41n4453.pdf](http://ntj.tax.org/wwtax/ntjrec.nsf/5A0C0247A873C2838525686C00686D08/$FILE/v41n4453.pdf))

¹² The CT600 database comprises information from companies' tax returns, and includes claims made by companies for R&D tax credits. Data on the SME tax credit schemes is published by HMRC, and is used by National Statistics.

¹³ An approach that can be used to overcome problems introduced as a result of causality concerns in an econometric model.

¹⁴ Arellano, M. and Bond, S. (1991) 'Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations', *Review of Economic Studies*, **58**, pp 277–97.

¹⁵ Hall, B. (1992), 'R&D Tax Policy during the Eighties: Success or Failure?', National Bureau of Economic Research Working Paper, No. 4240.

As high adjustment costs are associated with any investment in R&D, this implies that companies cannot adjust their R&D expenditure instantaneously in response to any tax incentives. This suggests that dynamic rather than static econometric models should be estimated, to be able to identify the long-run impact of R&D tax credits. This would require a greater amount of data over time than is currently available in the UK.

2.3.3 The choice of the control group

In any econometric analysis of the effectiveness of R&D tax incentives, the dataset would ideally comprise firms that receive the incentive (sometimes referred to as the treatment group) as well as a group of firms that do not receive the incentive (the control group), but have similar characteristics to those firms that do. The impact of the tax credits will be overestimated if the dataset is composed solely of those firms that receive the tax credit.

Theoretically, the control group can be used to estimate the level of R&D expenditure that would have been undertaken, in the absence of any R&D tax incentives. In practice, it is often difficult to find companies that do not claim R&D tax incentives, but have characteristics that are similar to those companies that claim tax incentives. As the R&D tax incentives schemes in the UK are general schemes from which all R&D companies are able to benefit, clearly defined control groups are not likely to exist by industry, region or other groupings. There are also potential problems associated with using R&D companies that do not claim tax credits as a control group. To ensure the robustness of the modelling, it is important to understand why these companies do not claim tax credits. If the control group comprises firms with different characteristics to the firms that claim R&D tax incentives, this may lead to spurious results being obtained about the effectiveness of the R&D tax incentives.

2.3.4 Relabelling expenditure on R&D

Estimates of the response of R&D to tax credits may be overestimated when a tax relief is introduced for a particular activity, as firms may reallocate expenditure in their accounts, to maximise the amount of qualifying R&D. This issue will be minimised if the impact of the R&D tax incentives is estimated from price elasticities, which do not incorporate a measure of the tax subsidy. If the responsiveness of R&D to tax incentives is estimated through the ordinary price of R&D, rather than the tax price of R&D, the potential for the relabelling of expenditure on R&D to affect the results will be minimised. The selection criteria for approving projects should also limit the possibility for relabelling of expenditure as R&D.

3 Data and matching

To create a dataset that includes data on whether firms have an R&D tax credit and the amount of R&D expenditure in that firm, HMRC's tax return data on companies has been matched with data from ONS BERD. This gives wider information on company characteristics and other relevant variables (eg, company characteristics).

3.1 HMRC's company tax returns

The CT600 database contains returns from companies that submit at least one claim for R&D tax credits or enhanced expenditure during the period covered by the return. Oxera has had access to an extract of the CT600 database, which comprised company tax returns that included a claim for R&D tax credits, and were received by HMRC up to January 2006. This data broadly covers the period between the financial years 1999/2000 and 2005/06.

The extract from the CT600 database contains company-specific financial data, requested as part of the company tax returns, such as turnover, trading profits or losses, profits chargeable to company tax and capital allowances. In addition, the CT600 extract includes data relating to any claim (or claims) for R&D tax credits or enhanced expenditure. This covers four R&D schemes—vaccines research relief, R&D that is sub-contracted to a small to medium-sized enterprise (SME) by a non-SME, the SME scheme and the large companies' scheme.

As it is possible that companies may claim under more than one scheme, the data has been collapsed to a company level, while retaining the number of schemes applied for, before any matching with the ONS BERD data was attempted.

3.2 ONS Business Enterprise R&D survey

The BERD survey is conducted annually by the ONS, and covers expenditure and employment relating to R&D activity undertaken by UK businesses.

Two different forms of the survey are sent to businesses, depending on the amount of R&D that they undertake.

- A *'long' form*—this requests a detailed breakdown of R&D activity. This form is sent to all businesses identified, either through the previous survey or from information gained by the ONS from other sources, as spending more than £3m on R&D. Businesses receiving the long form account for approximately 80 per cent of total R&D expenditure.¹⁶
- A *'short' form*—this requests only R&D expenditure and employment totals from the smaller R&D performers. These companies were identified as those businesses that spend less than £3m on R&D expenditure in the previous survey, or those that responded positively to R&D questions in other ONS inquiries, as well as other identified potential R&D performers.

A crucial difference between the two forms is that for small R&D businesses, which are re-sampled each year, the same business cannot be interviewed more than once every four

¹⁶ Business Monitor MA14 (2005), 'Research and Development in UK Businesses, 2004', National Statistics, January.

years, whereas the long forms that cover large R&D spenders track those businesses through time.

3.3 Matching procedure

Before any matching was undertaken, Oxera assessed the potential degree of overlap between the CT600 and BERD databases. To be able to match the data, the *same* company needs to appear in both databases in the *same* year. Figure 3.1 depicts the matching process.

Figure 3.1 Overlap between the CT600 and BERD databases

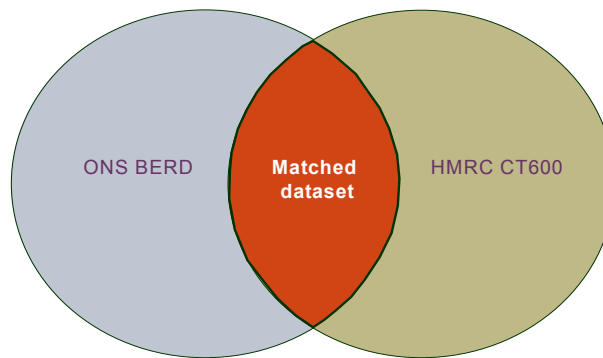
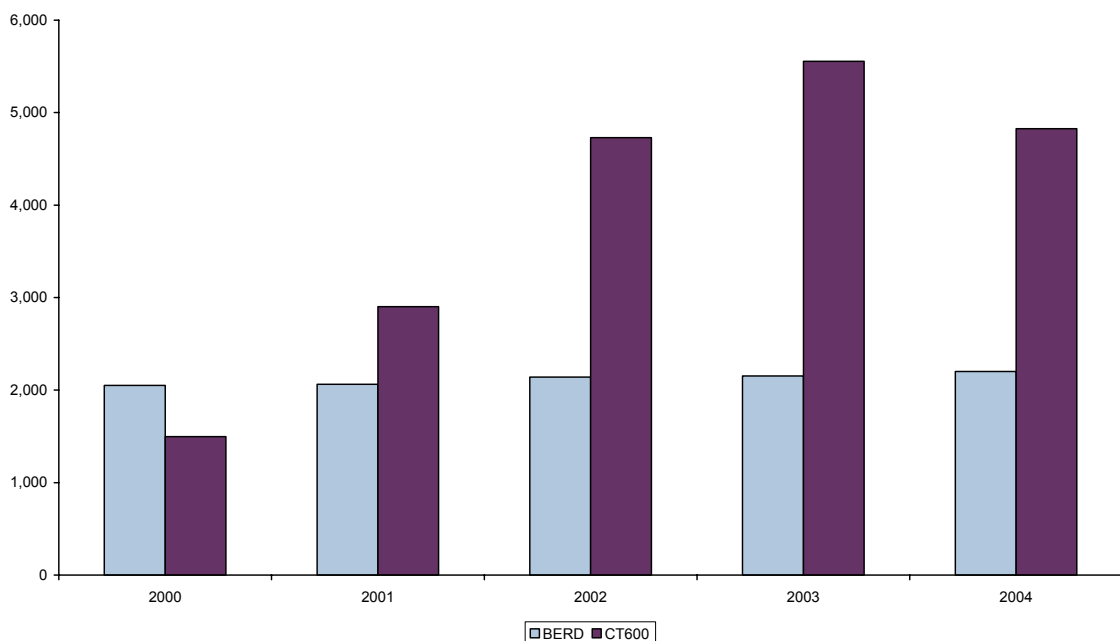


Figure 3.2 shows the general trend in the number of company registration numbers (CRNs) in both the CT600 and the BERD databases over time. The number of companies—as measured by the CRN—recorded in the CT600 extract has increased over time, as R&D companies have become more familiar with the tax incentives schemes. As claims are made retrospectively and companies can, in some circumstances, have up to six years in which to submit and amend their claims for R&D tax credits, the dataset of claims received is not complete for all years, particularly the more recent years. For example, the full set of claims for 2004/05 and 2005/06 may not have been recorded in the CT600 extract, as at January 16th 2006. As a result, it is likely that the matched results may be slightly lower in 2004.

Figure 3.2 Availability of data over time, number of companies



Note: The chart is based on the cleaned datasets. The year variable is defined according to the BERD and CT600 definition, and closely approximates the start of the financial year that includes the accounting period end date. Source: Oxera's calculations, HMRC's CT600 extract and ONS' BERD database.

If all R&D companies claimed R&D tax incentives, all the CRNs each year in the BERD database would also appear in the CT600 extract in the same year. In this ideal state, where all companies from the BERD database are present in the CT600 extract in the *same* year, the final column in Table 3.1 below—total BERD as per cent of CT600—would illustrate the matching rate that would be obtained. This represents the percentage of observations from the CT600 extract that are matched with data from BERD, although the matching rate in 2000 would be capped at 100 per cent.¹⁷ The matching rate—the percentage of observations from the CT600 extract that are matched with BERD—cannot exceed 100 per cent.

However, in practice, the actual matching rate is likely to be lower, as the CT600 database will not include all R&D performing companies in each year. For example, the CT600 database does *not* include:

- companies that have claimed R&D tax credits and/or enhanced expenditure, but have not recorded that claim in their tax returns;
- companies that have not yet claimed R&D tax credits or enhanced expenditure;
- companies that have decided not to make any claims.

Table 3.1 Availability of data over time, as measured by the CRN

Year	Total BERD	CT600	Total BERD as per cent of CT600
2000	2,051	1,497	137
2001	2,063	2,903	71
2002	2,141	4,730	45
2003	2,154	5,554	39
2004	2,200	4,827	46

Note: These results are based on the cleaned data. The year variable is defined according to the BERD and CT600 definition, and closely approximates the start of the financial year that includes the accounting period end date.

Sources: Oxera’s calculations, HMRC’s CT600 extract and ONS’ BERD database.

Choice of unique identifier

To undertake any robust econometric analysis, data from the CT600 extract will need to be matched with data for the *same* company *and* year from BERD. To match observations from the CT600 extract with those from BERD, a variable that uniquely identifies all observations in each database, but is also common across databases, needs to be identified. The CRN, company name and postcode are the only variables that are common across both the CT600 and BERD datasets.

Attempts have been made to identify a matching procedure that ensures the greatest degree of overlap between the CT600 and BERD databases. As a result of the minor differences in the length of the accounting periods across databases, Oxera has experimented with alternative definitions of the time variable, to ensure that the matching procedure is as robust as possible.

The data was matched using CRN and company name. Postcode was not used as the BERD database contained multiple postcodes reflecting the different sites at which a

¹⁷ The final column—total BERD as per cent of CT600—exceeds 100 per cent in 2000 as the number of CRNs in BERD exceeds the number of CRNs in the CT600 extract. In practice, the percentage of observations from the CT600 extract that are matched with BERD cannot exceed 100 per cent.

company operated. In creating the unique identifiers substantial modifying and formatting was undertaken to ensure the best possible match criteria. As an illustrative example:

- the unmodified CRN, *SC001234*, became *0000001234*, while the unmodified CRN, *0023456*, remained as *0023456*;
- the unmodified name, *ALPHA BETA LIMITED*, became *alpha beta*.

3.4 Caveat of the matching procedure

A remaining issue that could affect the robustness of the matching procedure is whether companies are represented at the same reporting level in the CT600 extract and the BERD database. For example, it may be possible that a company is included in the CT600 database at the enterprise level—the smallest group of legal units within a firm. However, the *same* company may be defined in the BERD database by the parent company that owns the enterprise. The enterprise and the parent company may have different CRNs and names, which would lead to a failure to match the observations. Knowledge of the reporting level of the companies would enhance the robustness of any future econometric analysis.

3.5 Results of the matching

Table 3.2 below shows the results of the matching between 2000 and 2004.

Table 3.2 Results of matching, by year

	Matched observations	Total observations in the CT600	Matched observations as per cent of CT600
2000	87	1,497	6
2001	195	2,903	7
2002	528	4,730	11
2003	638	5,554	11
2004	503	4,827	10
Total	1,951	19,511	10

Source: Oxera's calculations, HMRC's CT600 extract and ONS' BERD database.

Table 3.2 shows that the percentage of matched observations from the CT600 extract increases over time, up to 2003. Only a small number of observations were matched in 2000 and 2001, as a result of tax relief for R&D companies only being introduced from 2000. This suggests that the dataset, on which any future econometric analysis would be based, is likely to be particularly unbalanced in the first two years of the sample. Only a small number of companies will be present in the matched database for the full years of the sample, with the number of companies rising substantially in 2002.

Oxera has analysed the matched data to ascertain the number of year's data for each company in the matched dataset. This will help to inform the nature of the dataset that could be expected for future econometric analysis. Table 3.3 shows the availability in the matched dataset of consecutive data over time for the same company.

Table 3.3 Frequency of matched observations

Number of years	Number of CRNs	Percentage of CRNs with data for consecutive years
1	910	n.a.
2	319	74
3	108	93
4	16	94
5	3	100

Sources: Oxera's calculations, HMRC's CT600 extract and ONS' BERD database.

To ensure the robustness and the accuracy of the matching procedure, alternative definitions of the time variable and variations in the CRN (outlined above) have been undertaken. However, it has been found that altering the definition of the time variable (eg. from financial year to calendar year) or the cleaning of the CRNs does not lead to a significant change in the number of matched observations.

At present, 1,951 observations have been matched, which represents around 10 per cent of all observations from the CT600 extract. A higher proportion of observations have been matched from the long form (relative to the total number of long-form respondents), although a greater absolute number of observations have been matched from the short form. The matched observations are concentrated among companies with the highest turnover that operate in the services, manufacturing, electrical machinery and mechanical engineering sectors.

For the majority of companies, data is available for only one year, and there are only around 100 companies with data for three consecutive years. As such, currently there are not a sufficient number of companies over a sufficiently long time period to enable a robust econometric evaluation of R&D tax incentives in the UK to be undertaken, particularly as the majority of the matched observations are from the short-form. As a result, it is not possible given current legislation to obtain consecutive data over time for the smaller R&D-performing companies. Aggregate time series data for small companies is likely to be subject to similar problems in that any change in R&D activity might be due to the sampling rather than the effect of tax incentives. The literature review has also suggested that evaluation work should be based on disaggregated data at the level of the firm, rather than aggregate data at industry level.

The limited level of matching may reflect the following factors.

- The re-sampling of companies from the short form—statistical compliance legislation prevents the same business from being interviewed more than once every four years. As a result, it may be difficult to obtain consecutive data over time for the smaller R&D performing companies.
- The large companies' scheme was not introduced until 2002, and therefore at most only three years of data will be available for the large companies. However, as the CT600 data is not complete for 2004, for the majority of observations, only one or two years will be available.
- Large companies cannot claim payable credit if they are loss-making, although they can carry forward enhanced losses—Oxera does not have exact information on the number of loss-making companies.
- The lack of Northern Ireland data from BERD—although the small number of Northern Ireland forms is unlikely to significantly affect the results of the matching.

- Around 1,000 companies did not undertake R&D during the reporting period, despite being identified as R&D performers by BERD.
- Companies may delay filing their tax returns, causing discontinuities in the availability of data over time. For example, companies may, in certain circumstances, have up to six years in which to submit their claims for R&D tax credits and/or enhanced expenditure. The CT600 database will not include claims for R&D tax credits and/or enhanced expenditure that are not recorded on the company's tax return.
- The number of claims for the SME scheme has taken several years to build up. The lower take-up in the first two years after the scheme was introduced has resulted in much lower matched observations in these years.

Taking the above factors into account, the current level of matching appears reasonable, and indicates the baseline level of matching that may be expected in the future. It is expected that the number of matched observations may rise over time, particularly as companies reach their six-year limit in which to submit their claims for R&D tax credits.

4 Conclusions

The introduction of incentives to encourage R&D aims to redress the failure of the markets to provide a socially optimum level of R&D expenditure. By offering tax incentives to firms undertaking R&D, the policy aims to bridge the gap between private costs and the social return. Oxera has investigated the possible approaches and data sources to assess the feasibility of a robust evaluation of the effectiveness of R&D tax incentives on R&D expenditure in the UK. The empirical literature has indicated the direction any future econometric analysis of R&D tax incentives in the UK should take.

- The empirical literature suggests that tax price elasticity estimation may lead to the most accurate estimates of the impact of any R&D tax credit scheme. However, the indicator variable approach may be easiest to implement in practice. The choice of method for future studies is likely to depend on the available data and its properties.
- Most econometric investigations of tax credit schemes have been based on a panel dataset that includes a minimum of around four years of annual data for the same companies over time.
- Overall, very few studies have been based on data in the UK. The majority of studies focus on data in the USA, reporting the overall evidence of a positive impact from the tax credits. However, given that long-run R&D price elasticities are often at least twice as large as any short-run impacts, it is important to assess the long-term impact of any tax relief for R&D.

The combination of HMRC's CT600 data and ONS' BERD data would appear to fulfil most of the criteria required for a robust quantification of the effects of R&D incentives on R&D expenditure. Oxera has undertaken matching of the two datasets based on the modified CRN, company name and the time variable. There is potential for refinements to be made to the matching process based on the company name; however, it is unlikely that any refinements to the matching process (based on the current datasets) will lead to a significant change in the number of matched observations. As the full impact of R&D tax credits will take time to emerge, even if all companies from BERD in *each* year appeared in the CT600 extract in the *same* year, there is unlikely *at present* to be a sufficiently long time series availability of data over which robust estimates of the impact of R&D credits could be obtained. However, as more data becomes available with time, this would enable a robust econometric assessment of R&D tax credits to be undertaken. That said, the available data is more limited for smaller compared to larger R&D companies, which means that a more restricted range of econometric techniques can be used in assessing the policy effects for these companies.

At present, 1,951 observations have been matched (predominantly from the short form), which represents around 10 per cent of all observations from the CT600 extract. For the majority of the matched companies, data is available for only one year, and there are only around 100 companies with data for three consecutive years. Since the R&D schemes in the UK were not introduced until 2000 at the earliest, currently, there are not a sufficient number of firms over a sufficiently long time period to enable a robust econometric evaluation of the impact of the effectiveness of R&D tax incentives in the UK. To be able to assess the impact of R&D incentive schemes on the level of R&D expenditure, data is needed for a representative group of companies that claim R&D tax incentives in *each industry*, for a time period long enough for the effects of R&D incentives to have increased R&D expenditure. As R&D projects are often long-term and investment is lumpy over time, consecutive data for the same companies over time is required.

Ideally, the evaluation would compare R&D companies claiming R&D tax credits with a representative control group of R&D companies that do not claim tax credits. However, as a result of the inclusive nature of the R&D schemes in the UK, such a control group does not exist.¹⁸ This raises some significant methodological issues; if the evaluation of R&D tax credits were based on a dataset comprised solely of R&D-performing companies claiming tax incentives, the impact of the tax incentives on R&D expenditure is likely to be overestimated.

Therefore, it is important that any evaluation of R&D tax incentives is based on a dataset that includes a control group of R&D-performing companies that do not claim tax incentives. To ensure the accuracy of any future evaluation, the control group should have similar characteristics to companies that claim R&D tax incentives. Possible options for such a control group include the following groups of companies.

- *R&D-performing companies that have not claimed R&D tax credits or enhanced expenditure*—these companies would appear in BERD, but would not be present in HMRC’s tax returns database. As HMRC’s tax returns database will not include companies that have claimed R&D tax credits, but not recorded that claim in their tax return, this approach could lead to companies that have genuinely claimed R&D tax incentives being included in the control group. As companies may, under some circumstances, have up to six years in which to submit and amend their claims, the control group may also comprise companies that intend to submit claims for tax incentives, but have not yet submitted these claims to HMRC. This could lead to an upward bias in estimates of the impact of R&D tax incentives on levels of R&D expenditure.
- *Companies that submit claims for R&D tax incentives, but are not awarded any tax credits and/or enhanced expenditure*—an alternative dataset of companies’ tax returns—HMRC’s amended dataset—contains data for R&D-performing companies whose claims were approved by HMRC. This would enable the companies whose claims for R&D tax incentives are not approved by HMRC to be identified. To ensure that the characteristics of the control group are similar to those firms that are awarded R&D tax credits, it is important to understand why the claims have not been allowed.
- *Companies undertaking R&D before the tax incentives schemes were introduced in 2000*—data for R&D performing companies prior to 2000 could be obtained from the BERD database, which contains information on companies that undertake R&D from 1990 onwards. However, these companies would not be present in HMRC’s tax returns database. The companies would need to be selected to match, as closely as possible, the characteristics of those companies that claim tax credits. As this comparison will cover a different time period, it will be important for the analysis to take account of the different economic conditions and other factors that may have influenced companies’ performance over the two periods. In practice, it will be difficult to control for the influence of other factors, and as such, any estimates of the impact of the tax incentive schemes on levels of R&D expenditure could be subject to bias.

At present, the lack of continuous data for the same companies over time would severely restrict the ability to undertake any econometric analysis. There are a very low number of matched observations in 2000 and 2001. This would mean that the potential dataset on which an econometric analysis could be carried out would be particularly unbalanced in the early years of the sample. In addition, the concentration of the matched observations among those companies with the highest turnover and R&D expenditure (in part due to BERD

¹⁸ The R&D tax incentive schemes in the UK do not contain any clauses restricting companies from particular sectors from claiming R&D tax incentives, provided the qualifying criteria are met.

sampling) would lead to a potential bias in the results from any quantitative assessment of the impact of R&D tax credits or enhanced expenditure.

As statistical compliance legislation prevents companies that spend less than £3m on R&D from being sampled by BERD more than once every four years, it will be difficult to obtain consecutive data over time for the smaller R&D performers. There is only a limited discussion in the literature regarding the approaches that can be adopted if data is restricted in this way. This suggests that any panel data estimation of the impact of tax incentives on R&D expenditure could only be undertaken for the large companies.

Ideally, any econometric evaluation of R&D tax credits would be based on a dataset that includes both the larger and smaller R&D performers. However, for the smaller R&D performers, annual data will only be available every four years for the same company over time, and it is therefore unlikely that any consistent panel data will be available. This is a significant limitation as this will restrict the ability of the econometric analysis to include previous R&D spending as a factor affecting a company's current R&D spending. As the literature has shown that previous R&D spending is a key determinant of current R&D spending, this represents a key issue that any future evaluation will need to consider.

The more limited data available for smaller R&D companies—in particular, the lack of panel or time series data for individual companies—means that possible options for the econometric analysis are much more limited for these companies compared to the larger R&D companies. However, a number of alternative approaches exist that may be able to yield robust results, although each have their own limitations. Given the importance of the evaluation covering as wide a range of R&D companies as possible, this issue merits further consideration.

Empirical studies have been based on a minimum of around four years of data. If consecutive data for the same companies was available in the UK for around four years and the companies were distributed across a wide range of industries, it may be possible to estimate the short-term impacts of the introduction of R&D tax incentives for the larger companies. However, as the restricted availability of time-series data would prevent the dynamics from being fully specified, estimates of the short-run impacts of any R&D tax incentives may be biased. It is likely that around a minimum of ten years of data for the same companies from a range of industries over time would be required to estimate the short and long-run impacts simultaneously. However, the robustness of any procedure would increase, as more time-series data becomes available.

The matching is currently limited by a variety of factors, some which will 'disappear' in time as there is a greater distance from the start of the tax incentive schemes, and as firms approach the six-year limit on amending their tax returns. The above level of matching (10 per cent), in terms of companies and years, is therefore the minimum that could be expected in the future.

Oxera

Park Central
40/41 Park End Street
Oxford OX1 1JD
United Kingdom

Tel: +44 (0) 1865 253 000
Fax: +44 (0) 1865 251 172

www.oxera.com