Coverage of the Business Enterprise Research & Development Survey

Daniel Ker and Matthew Greenaway

Abstract

This information note provides an overview of the UK Business Enterprise Research & Development (BERD) survey design with a focus on the methods and sources used to annually update the sampling frame. This information note is intended to improve understanding of the coverage of the reference list of known UK research and development (R&D) performers that form this sampling frame. A recent UK Statistics Authority assessment of the survey requested a review of the methodology for producing business R&D statistics to identify potential gaps in coverage and suggest how any under-coverage might be addressed. This paper concludes that there is not enough robust information to estimate a reliable level of potential under-coverage at this time. The results presented in this Information Note should therefore be used with caution.

Summary

- The BERD survey is the key source of data on UK businesses' R&D and aims to provide information covering all R&D performed by businesses in the UK.

- The BERD sampling frame is compiled from nine different sources used to identify businesses as R&D performers. Any business identified as an R&D performer remains in the sampling frame unless there is specific information leading ONS to remove it. The largest contribution to the annual update of the sampling frame comes from the ONS Annual Business Survey (ABS). Analysis undertaken to compare the number of R&D performers from the BERD to the number from the Annual Business Survey (ABS) in 2010 suggests that there is good coverage of businesses with 50 or more employees but that there may be under-coverage of businesses with fewer personnel – and particularly of the smallest businesses with fewer than nine employees.

- Some under-coverage may arise because the ABS is itself a sample survey and so only R&D performers that are sampled can enter the BERD sampling frame this way. Furthermore, the ABS does not cover some industries including agriculture, financial activities, and public administration & defence. Other sources are used to identify R&D performers in these industries for the BERD sampling frame.

- At present, there is not enough information to estimate the value of R&D performed by businesses not covered by the BERD and so no robust methodology for adjustment can be identified without gathering further data.

Background

The BERD survey collects UK businesses' total R&D expenditures broken down into different expenditure types (capital, current, etc.) for analytical and policy uses. It distinguishes between expenditures relating to the business' own performance of R&D (intramural) and R&D purchased from outside the business (extramural). It also collects information on sources of funding and the types of resources employed on R&D.

The survey uses the internationally agreed design, methods, and concepts laid out in the Organisation for Economic Cooperation and Development (OECD) 'Frascati Manual of Proposed Standard Practice for Surveys on Research and Experimental Development'[1]. First published in 1963, and last revised in 2002, the Frascati Manual sets out the most widely used definition of R&D as:

> *"Creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture, and society, and the use of this stock of knowledge to devise new applications".*

The BERD survey has been conducted annually in the UK since 1993 in its current form, but Frascati Manual based estimates are available for some earlier periods.

Survey results are published annually as a BERD Statistical Bulletin and also used in the compilation of the UK Gross Domestic Expenditure on R&D Statistical Bulletin (GERD) which also includes R&D expenditure data for the government, higher education and non-profit business sectors. Estimates are used by UK and EU policy makers and for international comparisons. Eurostat receive a direct transmission of results under EU Commission Regulation 753/2004.

Changes made to capitalise R&D in the 2008 revision of the System of National Accounts[2] create another key use for BERD data. BERD estimates of R&D expenditure will be used to estimate R&D output in the National Accounts and will directly affect economic aggregates such as Gross Domestic Product (GDP). This adds to the need for estimates of UK business' R&D expenditures to be complete and robust. The National Accounts aim to cover most UK economic activity to meet policy and analytical needs and R&D capitalisation provides a new driver for ONS to evaluate the completeness of BERD estimates. Similar evaluations are being undertaken by other National Statistics Offices as a result of R&D capitalisation.

Furthermore, in its June 2012 'Assessment of Compliance with the Code of Practice for Official Statistics'[3] the United Kingdom Statistics Authority (UKSA) issued the requirement that the ONS 'publish plans to review the methodology for producing the R&D statistics in order to meet the coverage requirements of the European System of Accounts 2010'. This Information Note goes beyond this requirement by presenting an initial review of the potential for under-coverage in the BERD estimates and suggests possible further work that may be undertaken in the future.

Overview of BERD Survey Design

The target population of the BERD is all businesses performing R&D in the UK. The design of the sampling frame and sample are as follows.

Sampling frame

The sampling frame is a list of around 24,000 businesses *known* to perform R&D. The number of businesses in the sampling frame changes each year as new R&D performers are included and as others cease trading, for example. **Figure 1** shows the various sources that contribute to the BERD sampling frame and their proportional contribution to the net increase of 2,733 businesses in the survey for 2011, which brought the number of businesses covered to 25,511. Because of the method used to compile the BERD sampling frame it is not possible to present proportions of the sampling frame by source. However, the relative contributions to the annual change in 2011 shown may provide an indication - though they can change from year to year.

Figure 1: BERD Sampling Frame: contributions to change by source, 2011

Almost half of the inflow to the sampling frame comes from a question on the ABS asking *"During the next two years does the business plan to carry out any research and development?"* Respondents are referred to a definition of R&D similar to that in the Frascati Manual.

The UK Innovation Survey provides the second greatest inflow as it contains a number of questions relating to R&D, while the classification of businesses to the R&D sector on the Inter-Departmental Business Register (IDBR) also provides a considerable proportion.

Any businesses that cease-trading, merge, or restructure are identified as 'dead'. Each of these is investigated to ensure that only businesses which truly have ceased trading are removed from the sampling frame. Once companies are in the sampling frame they are not removed unless there is sufficient evidence that they have ceased performing R&D. A business will not be removed if it does not perform R&D in one year. As such the BERD sampling frame is built-upon and continuously improved over time rather than being completely re-constructed each year. This ensures that R&D performers remain in the sample once identified.

By way of international comparison, R&D tax credit claimants form the entire sampling frame in the Australian Bureau of Statistics' R&D surveys. Information from Her Majesty's Revenue and Customs (HMRC) has contributed to the BERD sampling frame in the past but these data have not been provided to ONS in recent years. It is hoped that this source will be available again in future.

Sample design

A census is made of around 400 businesses that spent the greatest amount on R&D in either of the previous two years (over £3.3 million in the 2010 survey). These businesses receive a 'long form' questionnaire requesting:
- capital and non-capital expenditure on in-house R&D;
- the description of the type of R&D performed (which industry product groups for example pharmaceuticals);
- the type of research undertaken;
- the source of funds for in-house R&D expenditure;
- how much R&D was purchased from outside the business;
- whether it was purchased in the UK or from overseas;
- the number of employees working on R&D and the full time equivalent (FTE);
- type of employees (i.e. researchers, technicians, support staff);
- post codes which relate to the workplaces where R&D is carried out.

Sampling fractions - varying by level of R&D expenditure, industry product group, and employment with three strata of less than 100, 100 to 399, and 400 or more employees - are used to select around 3,600 other businesses which receive a 'short form' requesting:
- expenditure on in-house R&D;
- whether R&D is for civil and/or defence purposes;
- how much was spent on purchases of R&D;
- the number of employees working on R&D.

These businesses are allocated an industry product group based upon the business' Standard Industrial Classification (SIC) held on the IDBR, which describes the type of activity the business undertakes.

Response rate and imputation

The BERD is a statutory survey with a target response rate for long form businesses of 93%. These businesses undertake approximately 80% of the total UK R&D expenditure recorded. The target for short form responders is 90%, and these account for the remaining 20% of R&D total expenditure along with un-sampled businesses and short form non-responders. Ratio estimation

and a matched pairs methodology with business employment as the auxiliary variable are used to impute estimates for non-respondents.

Estimates are imputed for questions not included on the short form based on the businesses' responses and both long and short form responses from businesses in the same product group and employment strata. Finally, ratio estimation is also used to impute responses for businesses not included in the sample so that the end result is a dataset of all variables for all businesses in the sampling frame.

Potential under-coverage

The BERD is designed in accordance with the guidance in the Frascati Manual and represents an acceptable statistical way to capture the majority of R&D taking place in the UK. The sampling frame design is covered by the standard methods prescribed. However, there are potential sources of under-coverage arising from the design of the sampling frame which lists only businesses *known* to engage in R&D. The question is; 'how many businesses perform R&D *unknown* to the ONS?'

The ABS is the main contributor to the annual updating of the BERD sampling frame. This is also a key source for National Accounts aggregates, which include GDP. However, the ABS is itself a sample survey covering approximately 3.4% (70,000 businesses) of a population of around 1.8 million businesses listed on the Inter-Departmental Business Register (IDBR) which covers 99% of UK economic activity. As a result, businesses' sampling probability on the BERD is in part conditional on their sampling probability in the ABS.

Furthermore, the ABS omits some businesses including those in the financial services and agriculture sectors. Over £448 million of R&D in these sectors was recorded by the BERD in 2010.

As outlined above, supplementary sources are used alongside the ABS when updating the BERD sampling frame and once a business has been identified as R&D performing based on any source it remains in the sampling frame unless there is a specific, verified reason to remove it. This will help ensure comprehensive coverage of R&D performers.

However, there may be R&D performers that have not been identified by any of the sources used to compile the BERD sampling frame and as such could be excluded from the sample and the population which BERD estimates represent.

In any case, there could be a lag between the identification of a business as R&D performing and its inclusion in the BERD sampling frame (and resulting statistics) because the sampling frame is updated annually. As such, even if the sources used to compile the sampling frame do identify all new R&D performers in each period; their first year of R&D activity may not be covered. This could lead to the omission of R&D set-up costs. However, survey validation processes include contacting these businesses to identify these set up costs, and previous years' data will then be revised to take into account this initial omission.

This would also be a particular problem for businesses that do not perform R&D on a continuous, systematic basis but instead perform 'bursts' of R&D. For example, they are engaged in a project in one year and do not perform any R&D in the next year. The [Frascati Manual](#)[1] recognises the

existence of such businesses but notes the difficulties and cost factors associated with identifying them.

Across all businesses, these effects could potentially lead to material under-coverage as some R&D performers may be omitted from the sampling frame and lags in entering the sampling frame may lead to some periods of businesses R&D being omitted.

Comparison with the Annual Business Survey (ABS)

The ABS is the largest ONS business surveys. It gathers data from 70,000 UK businesses and is one of the key sources for National Accounts estimates with results covering two thirds of the UK economy. Although the [ABS](4) is a sample survey, weights are used to produce representative estimates for the areas of the economy covered. It is used for comparison here because of its breadth of coverage as well as the availability of data. An evaluation of information from the UK Innovation Survey (UKIS) showed that the survey does not provide sufficient information to make inferences about BERD coverage, while HMRC information on tax credit registrants may provide insights but has not been available to ONS in recent years.

As the ABS is a sample survey, in any given year it will only capture a sample of all R&D performers. This could result in some under-coverage in the BERD sampling frame unless those R&D performers are identified from some other source or are carried forward from earlier sampling frames.

The simplest way to examine for under-coverage is to use the weights on the ABS to estimate the total number of R&D performers based on the number of positive responses to the question *"During the next two years does the business plan to carry out any research and development?"*

Some ABS respondents may say that they will carry out R&D but then not actually perform any - or may have completed their R&D by the time they enter the BERD sampling frame. This can be adjusted for by estimating the proportion of businesses that respond 'yes' to the ABS question and are subsequently selected for the BERD sample in the following year and provide BERD responses showing that they do perform R&D. This proportion is applied to the ABS estimate of R&D performers.

It should be noted that there may be an opposite effect whereby some businesses did not plan to perform R&D when responding to the ABS question but actually do. It is not possible to estimate the size of effect or the extent of any bias it causes but it should be minimised by the design of the ABS question and the relatively short time span it refers to.

The end result is an estimate of R&D performers adjusted for 'false positive' responses (i.e. businesses which say they will perform R&D on the ABS but do not record any R&D in their BERD responses). This can then be compared to the number of businesses in the BERD sampling frame. However, not all businesses in the BERD sampling frame will perform R&D in the period. Taking only responses from businesses that actually reported performing R&D and applying the BERD design weights yields an estimate of the number of R&D performers based on the BERD.

Subtracting the adjusted estimate of R&D performers based on the ABS from the number of R&D performers estimated by the BERD yields an estimate of the number of R&D performers not covered by the BERD estimates where a negative result would indicate potential under-coverage of the BERD.

In **Table 1** the results have been used to estimate the proportion of R&D performers derived from the ABS that are covered by the BERD. These are broken down by the number of people the businesses employ which is used to stratify both the ABS and the BERD. The BERD shows there to be more R&D performers with 100 or more employees than is estimated from the ABS - coverage is over 100%. Furthermore, near-complete coverage of 93% is estimated for businesses employing 50 to 99 people. According to BERD, over 85% of UK R&D spending was concentrated in these businesses in 2010. As such, the BERD survey's coverage of firms with 50 or more employees appears relatively complete.

Table 1: R&D performing firms covered by BERD as a proportion of R&D performers estimated from the ABS, by employment band

Number of Employees	Estimated Coverage (%)
0 to 9	9.6
10 to 19	54.1
20 to 49	64.1
50 to 99	93.0
100 to 499	125.6
500 to 999	102.8
1,000 and over	109.5

Source: ABS 2009 data, BERD 2010 data

Potential under-coverage is concentrated in businesses employing 49 or fewer people - particularly businesses with fewer than 10 employees for which coverage may be as low as 9.6%. This is unsurprising as the ABS sampling fractions decline with numbers of employees so that fewer of these smaller firms are surveyed. Furthermore, similar methods are used for other surveys so in general larger businesses that are R&D performers are more likely to enter the BERD sampling frame.

The value of expenditures on R&D that these non-covered businesses may be performing is unknown and difficult to estimate. An approximation can be made by inflating the BERD R&D expenditure estimates for each employment band by the proportional under-coverage (i.e. the inverse of the proportional coverage in Table 1). However, this requires the assumption that R&D performers appearing in the BERD are a random sample of all R&D performers when the methods used to compile the BERD sampling frame actually ensure that it better represents R&D performing businesses than a simple random sample. Therefore the result is *maximum estimated under-coverage* of around £2.95 billion in 2010.

Around 75% of this potential under-coverage is concentrated in businesses with fewer than 10 employees, which is estimated by inflating the BERD estimate of R&D spending by businesses with fewer than 10 employees by a factor of 10. This does not seem plausible in practice; the results provide an indication of where potential under-coverage may exist in the BERD but do not provide sufficiently robust information to adjust BERD estimates to account for under-coverage.

Conclusion

The BERD survey follows the framework set out by international guidance in the [Frascati Manual](#)[1]. It aims to estimate all R&D performed by businesses in the UK.

However, there is potential for under-coverage arising from the reliance of the sampling frame on sources which provide an incomplete list of R&D performers. In particular, the main contribution to the sampling frame comes from the ABS which is itself a sample survey taken from a population of over 1.8 million businesses held on the IDBR. Some industries are not covered by the IDBR including agriculture, finance, and public administration and defence. Furthermore, sampling fractions on the ABS are relatively small for businesses employing fewer than 50 people. However, the other sources used will address this to some degree and additional sources such as HMRC data on R&D tax credit registrants should be explored.

An initial comparison of the weighted ABS estimates to the BERD estimates of the number of R&D performers for 2010 showed it is likely that there are businesses with fewer than 50 employees – and in particular with fewer than 10 employees – which do perform R&D but do not appear in the BERD. However, it is difficult to estimate how much R&D expenditure is not being included because of the rules restricting surveying of these smaller businesses. The approach used does mean that it is likely that the businesses with larger numbers of employees that are R&D performers are well represented by the BERD.

There is not sufficient information on *how much money* businesses not covered by the BERD sampling frame spend on R&D to recommend a robust methodology for adjusting for under-coverage. Administrative data held by HMRC on R&D tax credit registrants could provide some insight into the value of under-coverage provided the information can be matched to BERD data and the IDBR. It might also be possible to utilise the UK Innovation Survey but at this time the sample design does not permit such analysis.

The best possible source would be to collect R&D expenditure from an IDBR-based sample of businesses with fewer than 50 employees. While simply asking whether the business plans to perform R&D in the coming two years as on the ABS would allow the BERD sampling frame to be made more complete, it would not provide sufficient insight into the R&D spending of smaller businesses due to the low sampling fractions on the BERD. Any survey made will need to collect the value of these businesses' R&D spending.

References

1   Frascati Manual of Proposed Standard Practice for Surveys on Research and Experimental Development
    http://www.oecd.org/innovation/innovationinsciencetechnologyandindustry/frascatimanualproposedstandardpracticeforsurveysonresearchandexperimentaldevelopment6thedition.htm

2   System of National Accounts 2008
    http://unstats.un.org/unsd/nationalaccount/sna2008.asp

3   Assessment of Compliance with the Code of Practice for Official Statistics: Statistics on Research and Development
    http://www.statisticsauthority.gov.uk/assessment/assessment/assessment-reports/assessment-report-218---statistics-on-research-and-development.pdf

4   Annual Business Survey (ABS)
    http://www.ons.gov.uk/ons/guide-method/method-quality/specific/business-and-energy/annual-business-survey/index.html