Netherlands Official Statistics

Volume 13, Spring 1998

Voorburg

Prinses Beatrixlaan 428
P.O. Box 4000
2270 JM Voorburg (Netherlands)

Telephone : . .31 (070) 337 38 00
Fax : . .31 (070) 387 74 29
E-mail: lhka@cbs.nl
Internet: http://www.cbs.nl

Heerlen

Kloosterweg 1
P.O. Box 4481
6401 CZ Heerlen (Netherlands)

Telephone : . .31 (045) 570 60 00
Fax : . .31 (045) 572 74 40

Key figure A-125/1998

Subscription: Dfl. 40.00 per year
Price per copy: Dfl. 20.00

ISSN 0920-2048

Postage will be charged.

Contents

How are we doing? Performance indicators for national statistical systems 5
Willem F.M. de Vries

Imputation the alternative for surveying earning patterns 14
Eric Schulte Nordholt

The use of the calibration estimator in the industrial waste statistics 16
Cees N.J. van Beusekom and José M. Gouweleeuw

Employee training and productivity in Dutch manufacturing firms 19
Martin Boon and Ben van der Eijken

Model for a question bank 25
Jack Claessen

Comparable data from fertility and family surveys 28
Jan J. Latten

How are we doing?

Performance indicators for national statistical systems

Willem F.M. de Vries[1)]

This paper proposes a systematic approach to evaluating the performance of national statistical systems. Its starting points are the so-called Fundamental Principles of Official Statistics, which were adopted by the United Nations some time ago. The aim is to translate the principles into operational terms and concrete questions about 'how we are doing'.

1. Introduction

The rankings (or league tables, as they were called) of national statistical offices, published by *The Economist* newspaper some years ago, caused mild shock waves among official statisticians around the world. The first *Economist* ranking (1991) was primarily based on the timeliness and accuracy of some major statistical series. The second round, in 1993, also took into account judgements of chief government statisticians about the objectivity of statistics (in terms of absence of political interference), reliability of the numbers, the statistical methodology applied and the relevance of the published figures.

The appreciation of the ratings varied of course. The national statistical offices included mentioned in *The Economist*'s list were more or less pleased, depending on their relative position. Offices not in the list wondered why they had not been mentioned. Some offices argued that their rating was questionable or incorrect, because the information used had been incomplete or outdated. However, there was little discussion about the criteria *The Economist* had used, even though there was fairly broad agreement that the assessment had been somewhat superficial.

From *The Economist*'s point of view, as a newspaper primarily voicing the interests of the users of macro-economic statistics, the applied 'objective' criteria (average size of revisions to GDP growth, timeliness, and value for money in terms of number of statisticians per 10,000 population as well as the government statistics budget per head of the population) made good sense. Adding senior statisticians' views to these criteria was perhaps not a bad idea either. However, it was clear to most insiders that the overall ratings at best presented an incomplete picture. This article proposes a more comprehensive, systematic checklist of points to be considered when evaluating a national statistical office or national statistical system[2)].

This 'checklist' is mainly based on the so-called Fundamental Principles of Official Statistics, first adopted by the Economic Commission for Europe during its 47th session in 1992, and subsequently endorsed by the United Nations Statistical Commission (with some minor amendments). These ten principles are a now a universally agreed framework for the mission of national statistical offices and indeed also for the statistical work of official international organisations.

After quoting the official wording of each of the Fundamental Principles of Official Statistics, a brief explanation in simple words of the essence of each Principle will be given. In addition, I have tried to make the principles more operational by raising some questions about them. The answers to these questions should indicate whether and to what extent a principle is adhered to in a given NSI. The article does not discuss all aspects of each of the Principles in any depth. It only raises some points which are thought to be of key interest[3)]. Neither does it discuss measurement issues (in other words: how to 'score' on the questions) in a strictly quantitative sense, although suggestions for a very primitive scoring system are given at the end[4)].

The question has been raised, and rightfully so, whether the approach that I am advocating here ultimately produces real indications about which are 'good' or 'better' statistical systems. A statistical system that scores high on 'my indicators', it is argued, may have a high ethical and professional standard and may do its very best in many ways, but is there any guarantee that it produces good, relevant, timely statistics? The answer to that question would probably be: no, but nevertheless I am convinced that there is a high positive correlation between scoring well on 'my indicators' and being a successful system in terms of output.

However, to accommodate the above views, I have divided the paper into two parts. Part 1 is about the Fundamental Principles, Part 2 is about real statistical output as such.

PART 1

Relevance, impartiality and equal access

1. Official statistics provide an indispensable element in the information system of a society, serving the government, the economy and the public with data about the economic, demographic, social and environmental situation. To this end, official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honour citizens' entitlement to public information.

In other words, Principle 1 means that official statistics should be relevant for society, compiled in an impartial manner, free from political interference and accessible for everyone under equal conditions.

One of the reasons why Britain and the USA were rated relatively low (despite their good performance in other respects) by *The Economist* in 1993 was: 'the lingering suspicion that statistics in America and Britain are subject to political meddling'. Despite recent moves towards more centralisation of official statistics in Britain, a large part of the statistical work is still scattered across some 30 government departments, where the statisticians report directly to ministers. This (wrote *The Economist*) 'allows politicians to take an unhealthy interest in statistics...'.

Several questions can be asked in the context of judging national statistical offices against the background of the principles of relevance, impartiality and equal access.

The ultimate question pertaining to relevance would of course be: to what extent do the users think that the activities (data collections, or ultimately outputs and products) of statistical systems are relevant for them? It is, however, extremely difficult

to express this aspect of 'user satisfaction' in terms of one or a few a simple indicators (which does not mean one should not try to do so). Some users may consider some activities to be very relevant (while others may not), and may be very dissatisfied with other activities (much liked by others). Therefore, I would propose a more general question which has to do more with the general attitude of NSI's in this regard than with concrete indicators or measures. That question is:

1. How well developed are mechanisms to ensure that statistical work programs are relevant for the various user groups?

   In many countries, there is something like a national advisory board for statistics, but whether this works satisfactorily or not is a different matter. In addition, however, there are many other possible mechanisms to foster the relations between users and producers of official statistics. The basic question to be asked here is: are national statistical offices making a real effort to find out what their users need and to adapt their statistical programmes accordingly? And the next question would be: how flexible are they in practice when it comes to tackling 'new' (and probably quite relevant) subject matter areas such as the services sector, the environment, the 'information technology sector' and other matters relating to the economy of the 'intangibles', and last but not least 'the global economy' (including phenomena such as foreign direct investment and correct measurement of the activities of multinationals in general).

   Another, more specific question regarding 'user satisfaction' would be:

2. How well developed are mechanisms to assess user satisfaction with statistical products and their dissemination?

   Apart from statistical programs, which often describe what statistical offices are *doing* or are planning to do in terms of the subject matter areas to be covered, the content and coverage of data collections, and sometimes the methodology to be used and the timing and expected quality of statistical results, there are also the actual *statistical outputs* to consider and how the users appreciate these: news releases, printed publications of various kinds, data in electronic formats, including data bases etc. In other words: do statistical offices have a well developed dissemination system? Are the statistical products what the users want in terms of quality, timeliness, price, distribution modes? Are sales of statistical products increasing or declining? Is there any real, systematic marketing effort?[5)]

   As to impartiality, the question is:

3. How well do national statistical offices adhere to their obligation of impartiality?

   This may sound relatively simple, but in fact rather complex issues are at stake. The complexity largely depends on one's general notion of 'impartiality'. Very orthodox official statisticians may believe that even undertaking a survey at the special request of a ministerial department may affect the impartiality of a national statistical office, especially if this department (usually paying for the extra work to be done) wants to have a say in the methodology of the survey. However, most statisticians may tend to interpret 'impartiality' more loosely as: to avoid taking any partisan view in the choice of definitions or methodology, and, most particularly, to avoid adopting a biased stand as to the release of statistical numbers and commentary on those numbers.
   Most national statistical offices have a strong tradition of not making any non-statistical comments on their figures. Sometimes this principle is adhered to very strictly. In a press release about the latest unemployment numbers, the comment given will then be restricted to something like: 'Compared with the previous quarter, unemployment has decreased by 0.7 percentage points', leaving any additional comments to politicians and others. Nowadays, as many statistical offices wish to improve press coverage of their numbers, some may comment as follows: 'The decrease of unemployment in this quarter was 0.7 percentage points compared with the previous quarter. This is the strongest quarterly decrease since the second quarter of 1982'.

   As a general principle, however, statistical offices should (and indeed most will) avoid making any comments referring to the success or failure of government policy, even if the numbers may seem obvious in revealing this.

   As far as the issue of 'political interference with statistics' is concerned, the pertinent question is:

4. How well are statistical offices shielded from political intervention as to the content and the release of statistical results?

   - Some of the most common forms of unwanted political intervention seem to be:
   - Pressure to change definitions in order to obtain statistics which put government policies in a better light;
   - Tampering with the release of key statistical figures, in order to select a moment for release which is politically favourable or least damaging;
   - Leaking to the media of 'favourable' statistics by politicians before the data are made available for everyone;
   - Pressure to release identifiable micro-data to policy researchers in the case of statistical collections intended for and financed by specific clients (e.g. ministries)[6)];

   Apart from the first category (for which it is hard to formulate general rules of good practice), the highest risk of political interference with statistics therefore occurs at the stage when figures are (about to be) released. To avoid tampering with releases of fresh statistical figures, many countries have now adopted a system of announcing release dates of key statistics well (a month or even a year) in advance. Avoiding leaks may prove to be more difficult. It is the custom in many countries to give ministers a head start with respect to fresh key statistics by supplying them with the figures some time before these are officially released. This may be anything from an hour to several days and the list of recipients of these 'pre-releases' may be quite extensive. There is general agreement among statisticians, however, that it is commendable to restrict both the list and the time lap as much as possible.

   In view of the important role of the media in making statistics available for the general public, it is sometimes argued that supplying information to the media 'under embargo' (i.e. some hours before the official release time), in order to give them a better opportunity to prepare an attractive news item (this applies in particular to television news programmes, where this may take some time), should be possible[7)].

   As for 'equal access' the question is:

5. How well is the principle of 'equal access under equal conditions' adhered to?

   Apart from the political considerations under the previous point, there is also the *general* principle of safeguarding that *all* users are treated equally. Some aspects of this equality

are not trivial. Obviously, for certain figures a head start of a few minutes, for one user over another, may generate a considerable (financial) advantage. Therefore, statistical offices have to find ways to give all users access to new figures at *virtually exactly the same moment.* Apart from recently developed possibilities of simultaneous electronic distribution (e.g. by e-mailing statistical releases to the media), some countries use a system of 'lock-ups' for the release of certain sensitive figures[8].

Another aspect of equality is that, in principle, all users should pay the same prices for the same statistical products and that the number of 'privileged users' who receive the data free of charge (government agencies, members of parliament) should be restricted as much as possible.

A slightly distinct point, which is not covered by the principle of '*equal* access' as such, but which is nevertheless very essential, is the notion that official statistics are (intended as) a public good, which should in principle be freely available for all citizens. Most NSI's put this notion into practice through various means. First of all, as mentioned before, building up good relations with the media is important to serve the general public with basic statistical information. Secondly, it is a generally accepted practice for NSI's to make arrangements for the most important statistics to be freely accessible in their own libraries and in university and public libraries. Thirdly, most NSI's will give free information on the telephone (including follow-ups by sending free copies of tables etc. by mail) or by electronic channels, such as the Internet[9].

Professionalism

2. To retain trust in official statistics, the statistical agencies need to decide according to strictly professional considerations, including scientific principles and professional ethics, on the methods and procedures for the collection, processing, storage and presentation of statistical data.

Principle 2 simply says that official statistics should be compiled by using professional methods and also that statistical results should be presented to the users in a professional manner.

The real issue here is: to what extent is the professional integrity of NSI's safeguarded? Measuring professionalism and the adherence to professional ethics[10], and even more so: comparing these characteristics between national statistical offices, is obviously very difficult. On a subjective level, there may be some agreement among statisticians that national statistical office X or Y is relatively active in terms of methodological innovation in this or that area, but agreeing on some objective measure is an entirely different matter. The number of university graduates and their percentage share in the total staff of a national statistical institute may be an indication of its 'methodological potential', as may the number of research and methodology papers produced and published in respected scientific journals, but few would agree that this is a sound basis for comparisons between different statistical offices. The importance of analysis and research for methodological progress and for increased efficiency and effectiveness of statistical operations is widely recognised. A United Nations report[11], which is still the standard manual on the organisation of official statistics at the national level, underlines the significance of research and analysis for various reasons, including getting a clearer picture of the value of statistics, in particular as to discovering lacks and inconsistencies. An important American report[12] states that 'It became quite clear that it is analysis that holds a statistical system in place, makes possible most communication with decision-makers about their data needs, and informs them of current statistical capability. Analysis is the glue that holds all information systems together'. Lastly, Sir Claus Moser, then director of the Central Statistical Office, said in a speech to the Royal Statistical Society (1979) that 'One more aspect needs mentioning, namely the need for the government statistical service to devote more attention and resources to methodological work....The CSO has much to gain from constantly improving its technical standards, indeed has a duty to do so and to publish its findings'.

Therefore, some general questions may be asked to assess (the focus on) professionalism in national statistical offices.

6. How well is professionalism systematically promoted and shared by such mechanisms as analytical work, circulating and publishing methodological papers, and organising lectures and conferences?[13]
7. Are statistical methods well documented and are methodological improvements made on the basis of scientific criteria?
8. Are decisions about survey design, survey methods and techniques etc. made on the basis of professional considerations (or do other – e.g. political – considerations play a role)?
9. Is training and re-training of professional and other staff a real policy issue for the organisation and is enough effort (e.g. in a percentage of the overall budget) spent on training?
10. Is statistical quality management a real policy issue and are real and systematic efforts (including the promotion of well documented quality management guidelines) made to enhance the quality of statistics?

As for the aspect of 'professional presentation' of statistics, some comments have already been made under 'impartiality'. Some other points will be made under the next paragraph on 'accountability'.

Accountability

3. To facilitate a correct interpretation of the data, the statistical agencies are to present information according to scientific standards on the sources, methods and procedures of the statistics.

Accountability is understood in the sense that statisticians should systematically and thoroughly explain to the users of statistics what the numbers exactly represent and what their quality is.

To some extent this principle may seem trivial, but considering that the issue has long been (and still is) a topic for lively debate among statisticians, some non-trivial aspects are involved as well. The triviality lies in the fact that it is obvious that if you produce and publish figures, you should inform the user in some way what these figures are about. The debate is on how to do this in the best possible manner[14].

In terms of so-called meta-data (information about the data, i.e. definition of the population covered, definition of the variables, description of the data sources used, description of survey methodology, etc.), there is broad agreement that it is essential for the users of statistics to have access to as complete a set of meta-data as possible.
Therefore, national statistical offices should see to it that full descriptions of the complete methodology for all their collections are documented and kept up-to-date. This does not imply, obviously, that all statistical publications must contain a full set of meta-data, as that would be both impractical and user-unfriendly. Statistical databases, however, should preferably contain all the meta-data in some user-friendly form, because it would be a

burden for the users to have to consult separate publications to see what the data are worth[15].

A good example of meta-data are the *Sources and Methods* accompanying the OECD *Short Term Economic Indicators* publications. Also, the initiative taken by the International Monetary Fund in 1996 to set standards[16] (general standards for all countries, plus so-called special standards for the most developed countries) for meta-data about a set of major statistical series, must be mentioned in this respect. A large number of countries have now endorsed these standards.

The question to be asked with regard to meta-data is therefore:

11 How well does a statistical office provide the users with information about what the data really mean and about the methodology used to collect and process them?

Another issue, which is closely related to the previous paragraphs on meta-data, but which is nevertheless slightly different, is how statistical offices inform the users about the quality of the data they produce. Proper meta-data may tell a lot about the quality of statistics (at least for 'professional' users), but they do not give the whole picture. Therefore, though there may be a certain overlap between the two, explicit statements about the quality of statistics are an additional aspect of principle 3. Quality particularly concerns such aspects as sampling and non-sampling error, any biases the data may have, information about non-response and its treatment, about imputations etc. In the eighties, the Conference of European Statisticians of the United Nations Economic Commission for Europe adopted 'Guidelines for quality presentation', which are still very useful and are applied in some form or other, but often not systematically, by quite a few statistical offices. The question is therefore:

12 How well developed and applied is the presentation of the quality of statistics?

Prevention of misuse

4. The statistical agencies are entitled to comment on erroneous interpretation and misuse of statistics.

Principle 4 means simply that statisticians may react to any wrongful use of statistics that they perceive. Although the official wording of the principle is 'entitled', the general understanding of the principle is that statistical agencies indeed have a duty to comment.

There are of course many different ways to define 'erroneous interpretation' and 'misuse', and not all forms of these are equally bad or harmful. Moreover: most instances of misuse will escape the attention of statistical offices. Many users know 'how to lie with statistics', but this need not always be a concern for statistical offices.

However, some kinds of misuse may require corrective actions: in particular misuse by government agencies and by the media. For both categories, it is commendable for statistical offices to undertake immediate corrective actions in whatever way. At Statistics Canada it used to be (and probably still is) standard policy that when any misrepresentation or misinterpretation of official statistical figures in the media was noticed, the Chief Statistician wrote a letter to the editor explaining that a mistake had been made and how the numbers ought to have been correctly presented.
Similar steps were also taken for government misuse. It was felt that this general attitude has had positive effects by 'educating important users of statistics'[17] So, while it may be difficult to prescribe a standard recipe for these situations, the general question that may be asked is:

13 How well and systematically do statistical offices educate their key users in order to promote proper use of statistics and to prevent misuse?

Cost-effectiveness

5. Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies are to choose the sources with regard to quality, timeliness, costs and the burden on respondents.

Principle 5 means that statistical offices must try to be as cost-effective as possible by making the best choice of sources and methods, aiming at improved timeliness and also data quality, at spending tax-money as efficiently as possible and at reducing the response burden.

To some extent, possibilities to achieve cost-effectiveness depend on national circumstances. In countries with good administrative registers which are also available for statistical use, the need to have censuses or indeed traditional sample surveys will be less than in countries where such registers do not exist, are of poor quality or are not put at the disposal of the statisticians.
One of the most eloquent examples of how the national administrative infrastructure affects statistical expenditure very directly is the population census. Whereas in countries which do not have a population register (such as the United States) very costly periodic population censuses remain necessary, other countries (such as the Scandinavian countries and the Netherlands) nowadays produce very much the same statistics that were previously collected through a census by using registers and some additional sample surveys, at a mere fraction of the cost.

In terms of data input, making the best possible, balanced choice of data sources, given national circumstances, should therefore be an important issue for all statistical offices. The general question to be asked is:

14. How well considered is the 'data sources mix' used by statistical offices, and is achieving the best possible mix (also taking cost-effectiveness into account) a subject of systematic improvement effort?

In the different phases of data throughput (the data editing process, aggregation, analysis etc.), there are also many possibilities to increase timeliness, efficiency and/or to improve data quality. There are organisational issues to be considered, as well as methodological and technological aspects and many of these issues and aspects are inter-related. For example: introducing macro-editing instead of the more traditional micro-editing approach is only possible if statisticians are well-trained in this new approach and can make use of advanced information technology (software and hardware). It is impossible to give brief general guidelines, but the central question here seems to be fairly straightforward:

15. How effective and efficient is the data throughput in statistical offices, in terms of organisation, methodology and technology?

And an additional question of perhaps equal importance may be:

16. Is improving timeliness an issue of serious and systematic effort?

The response burden generated by statistical offices is another aspect of their cost-effectiveness, as data collection, apart from costing taxpayers' money, also implies costs for data providers. Therefore, reducing the response burden, in particular for data providers from the private sector, is presently an issue of concern in many countries. There are many different techniques to reduce the response burden[18], some of them fairly simple, others of a more 'high-tech' nature.

Comparison of the level of response burden generated by different statistical offices is very difficult, because the response burden depends on several factors, many of which are related to very specific national conditions and requirements. It is possible, however, to compare the overall development (upwards or downwards) of the response burden, as well as the general attitude of statistical offices with respect to the issue. A general question that could be asked is therefore:

17. How successful has a statistical office been in systematically reducing the response burden it imposes on data providers?

Cost-effectiveness is obviously also a matter of organisation, management and even 'corporate culture'. It is very difficult to measure the 'productivity' of statistical workers and even more so to compare 'productivity' between different statistical offices. Efforts to compare the cost of specific, rather comparable statistical operations (such as the Labour Force Survey or the Consumer Price Index) in a few countries of the European Union have so far been unsuccessful.

Because better standards to measure productivity and cost-effectiveness in statistics do not exist, *The Economist* was probably right in defining a couple of simple indicators to compare these issues between countries. Therefore, I propose to stick with these indicators: number of official statisticians per 10,000 population and the government statistics budget per head of the population[19] For countries which have a decentralised statistical system, the numbers should of course include both the central and the decentralised parts of the system. The problem is, of course, that the question 'how are we doing in this respect' can only be answered if comparable data for other countries are available. Nevertheless, the question must be asked:

18 How cost-effective is a national statistical system (in terms of relative cost indicators such as statisticians per 10,000 population and statistics budget per head of the population)?

Confidentiality

6. Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes.

Again, this seems to be a very simple principle, but it has many ramifications, some of which may involve very complex issues[20]. There is a well known joke, often told in countries which used to have a centrally planned economy, but are now moving towards a more market-oriented system. It is: 'In our country, individual data used to be widely known, while aggregates always were top secret'. This is a clear illustration of how the principle of confidentiality should *not* be interpreted and applied. Unfortunately, it does not say much about how it should.

Various questions can be raised about the concepts 'individual' and 'confidential'. The interpretation of the concepts may also vary from country to country. However, one should first of all consider what the true meaning of the principle is: self- interest of statistical offices. The simple reason why statistical offices must adhere to confidentiality of individual data is that it is the only way to safeguard the trust of the respondents. Respondents must be certain that the information they give is used for statistical purposes only and that they therefore have no interest in supplying anything but true data.

One may look at the issues from various angles. At the general policy level one may take into account what the law (if any) says. In many countries there is legislation about the protection of the privacy of citizens. This often includes provisions for statistics and these provisions may be more or less strict. In the Netherlands, for example, the general 'personal data protection law' makes some exceptions for statistics and research[21]. Equally, the confidentiality of individual business data is often safeguarded legally, be it under a general statistics law or in separate legislation. However, in this respect there may be some more or less essential differences between countries, in particular as far as the legal possibilities for exchange of company data between various government agencies are concerned.

At a more basic and practical level, it seems that most statistical offices have some official policy, or at the very least an accepted practice about how to prevent disclosure of individual data in disseminating their statistical products. A distinction may be made here between disclosure protection in the case of traditional, printed publications, and the more complex issue of disclosure protection with respect to electronic files with micro-data[22]. For printed publications, the rules are in practice often relatively simple, such as (in particular in the case of business statistics) suppressing cells in tables which contain information about just a few (e.g. three or less) individual entities.

For electronic files the rules may be more sophisticated, particularly in the case of so-called micro-data: files containing (anonymous) information about individual entities. In several countries (e.g. in the United States) such files are made generally available for research purposes: so-called public data files. The structure of these files is such that disclosure of individual data is considered to be virtually impossible. A variety of techniques is applied to prevent disclosure. In the Netherlands a distinction is made between such public data files and another type of micro-data: research-files which are not 100% 'disclosure proof', and which are only made available to certain categories of researchers and under very strict legal provisions.

So some general questions can be asked:

19. How well developed and practised are the rules to prevent disclosure of individual data in printed publications?

20. How well developed are techniques and systems to make statistical files available for research purposes, while preventing disclosure in the best possible manner?

Another issue regarding confidentiality is the prevention of non-statistical use of statistical data and guaranteeing administrative immunity of respondent groups. This is a rather complex problem area. When the draft of a Regulation for Community statistics (better known as the 'European Statistical Law') was discussed by the member states of the European Union, prolonged debates took place about the definition of and wording around such concepts as 'statistical data', 'use for statistical purposes' and 'non-statistical use'. For a discussion of the concept of 'administrative immunity', one may consult Begeer et al. (1984).

Yet another issue related to the confidence of citizens in the national statistical office concerns the perception of the public that databases and networks within these offices are in practice secure against external intrusions (by 'hacking' or otherwise). At Statistics Netherlands great care is taken to 'waterproof' the internal systems from the outside world.

I suggest that we do not include all these points, however relevant and even important they may be, in the 'performance indicator system' which is the subject of this article.

Legislation

7. The laws, regulations and measures under which the statistical systems operate are to be made public.

Principle 7 means that the position of statistical offices, including their rights and obligations should be codified in proper, publicly available legislation, in order to show the public what it may expect from the national statistical system.

It is impossible to set out very specific rules for statistical legislation. Much depends on national legal culture and traditions. Many countries have a formal 'general statistics law', but in others the statistical legislation may be scattered over a series of specific laws and various other government documents. Neither situation, however, is a guarantee that official statistics are in good shape, because it is useful to note here, that laws obviously cannot solve all problems. In some countries which do not have a 'general statistics law' (e.g. the United States or the United Kingdom), many of the best possible statistical practices may be adhered to, while other countries may have a statistical law which is perfectly formulated, but in practice is not much more than just another piece of paper.

Nevertheless, it is suggested that statistical legislation and/or other legislation which is also relevant for official statistics, should cover all or most of the following basic points:

- The general position of the national statistical office/system (including points such as who decides on the work programme, who decides on methodological issues, how are data collected, what are the relations between the national statistical office -if any- and other government agencies doing statistical work, what are the relations between the statistical system and government/parliament etc.)
- The position of the head of the national statistical office/system (including points such as who appoints and dismisses, to whom does the 'national statistician' report and about what, does he/she have any specific responsibilities etc.)
- Basic rules of data collection and confidentiality (voluntary and statutory data collection, any penalties for non-compliance with compulsory data collections, general and specific confidentiality rules)

In view of this, the question to be asked about statistical legislation may be:

21. How good is the statistical legislation in a country, in terms of clearly setting out the mission and the competencies of statistical agencies, legal obligations to provide information for statistical purposes and the protection of confidentiality of individual data?

In addition, some *implementation aspects* of statistical legislation or of the principles for good statistical conduct are to be taken into account where the 'performance' of statistical systems is concerned. In particular, it is generally considered to be not more than sensible and decent always to inform respondents properly of the legitimate basis for statistical data collections and other activities of statistical agencies, for instance by briefing them explicitly about the statutory or non-statutory nature of data collections. In the longer run, this is once again a matter of self-interest: 'honesty is the best policy'. A special issue in this regard is 'informed consent' of respondents as to any use of the provided (individual) information for non-statistical or research purposes.

The question to be answered would be:

22. How well developed are the policies and practices of dealing with respondents, in terms of ensuring that they are fully informed of their rights and duties with regard to statistical data collection?

National coordination

8. Coordination among statistical agencies within countries is essential to achieve consistency and efficiency in the statistical system.

In other words, Principle 8 means that in order to prevent inefficiency, undue response burden and the compilation of incomparable statistics, effective mechanisms for national coordination of statistics should be in place.

Statistical coordination has two main aspects: coordination of programmes (in particular with respect to data collections) and coordination of statistical concepts. Coordination of programmes aims at achieving efficiency (avoiding duplication of efforts) and at reducing the response burden (avoiding the same information being collected several times). Coordination of standards (in particular definitions and classifications) also has efficiency and response burden effects, but aims primarily at compilation of comparable statistics.
In this latter respect it is important that the national statistical office is recognised as the 'bureau of standards', standards which are respected and followed by all other agencies which may be active in official statistics.

Obviously, coordination is easier to achieve in countries with a centralised statistical system (such as Canada, Australia, the Netherlands) than in countries where official statistics are highly decentralised (such as the United States, where more than 70 federal agencies are active in statistics) or relatively decentralised (such as the United Kingdom, France or Japan).

Nevertheless, coordination mechanisms in countries with decentralised systems may be well developed and successful, while coordination in countries with a centralised system does not always function perfectly[23]. The question to be asked is therefore:

23. How well developed are national statistical coordination mechanisms and to what extent do they produce the envisaged results?

International coordination

9. The use by statistical agencies in each country of international concepts, classifications and methods promotes the consistency and efficiency of statistical systems at all official levels.

Principle 9 basically means that statistical offices should as much as possible adhere to international statistical standards and best practices, not only in order to produce internationally comparable statistics, but also in order to enhance efficiency of statistical operations and the overall quality of statistics.

There are two different aspects to international statistical coordination.

First of all, it is important that national statistical systems follow international definitions and classifications, in order to achieve cross-country comparability of statistics. This may seem simple

and obvious, but poses considerable problems in practice. International statistical definitions and classifications are by definition the result of a complex process of compromise. The compromise may be such that some countries can live with it better than others. In particular, developing countries may have difficulties in applying the standards fully, because the process of developing the standards is usually dominated by the more advanced countries.

Also, some 'blocks' of countries (e.g. the European Union) may wish to have their own specific standards, which sometimes slightly differ from the world (UN) standards[24]. Therefore, there is general international agreement that international coordination in this respect should be 'flexible', in the sense that countries or groups of countries are entitled to diverge from the world standards, as long as they ensure that the linkage between their standards and the world standards is straightforward and transparent.
The second aspect of international coordination is that countries should benefit as much as possible from methodological, organisational and other practical developments elsewhere. This form of coordination is aimed at improving efficiency and enhancing the quality of statistical products and operations.

Taking both aspects in one stride, the question to be asked with respect to this principle would be:

24. How well does a statistical system adhere to agreed international standards and does it contribute to the best of its abilities to the further development and promulgation of best statistical practices?

International statistical cooperation

10. Bilateral and multilateral cooperation in statistics contributes to the improvement of official statistics in all countries.

Principle 10 means that international cooperation is a prerequisite to enhance the overall, world-wide quality of official statistics. Therefore, national statistical agencies should regard it as part of their core activities to assist other countries to the best of their abilities.

Apart from international meetings of statisticians, where (the improvement of) statistical standards is discussed, quite a lot of other international statistical cooperation is going on. International organisations are trying to promote the use of standards and best practices by issuing handbooks and guidelines in many languages. Some of them also organise and finance technical cooperation programmes for developing countries or countries in transition from a centrally planned economy to a market economy.
A considerable number of training institutions exist, in all continents, where statisticians are trained in statistical methods, techniques and practices. In addition, there is a lot of bilateral cooperation between countries, sometimes financed from international funds, sometimes from national aid programmes.

The efficiency and effectiveness of international technical cooperation in statistics, in terms of avoiding duplication and promoting a systematic, goal-oriented approach, is also a topic of continuous discussion between national statistical agencies and international organisations.

The question to be asked with regard to this principle would be:

25 How actively is a statistical agency involved in international technical assistance?

PART 2

And what about the figures?

(The proof of the pudding is in the eating)

Some users may think that all these noble Fundamental Principles are of course all very well, and that respecting them may certainly help to improve the statistical system in the shorter or longer run, but that they really care more about the bottom line: do national statistical offices produce good statistics? And they have a point. So, in the footsteps of *The Economist* I suggest we also take into account the quality of some key statistics. Without disregard for all other valuable statistics, I propose a list of ten key statistics whose importance is probably undisputed and which are produced, in some form or other, by almost all national statistical systems: annual national accounts, quarterly national accounts, labour statistics (in particular monthly or quarterly unemployment rates), income statistics, basic demographic statistics, external trade statistics, the retail trade index, statistics on the services sector, the industrial production index and the consumer price index. So the questions to be answered are:

1. How good are the annual national accounts?
2. How good are the quarterly national accounts?
3. How good are the labour statistics (unemployment rates)?
4. How good are the statistics on the distribution of income?
5. How good are the basic demographic statistics?
6. How good are the external trade statistics?
7. How good is the retail trade index?
8. How well developed are statistics on the services sector?
9. How good is the industrial production index?
10. How good is the consumer price index?

And finally: may we have your points, please?

As indicated before, the aim of the above checklist was not really to generate 'scores', let alone rankings of statistical offices on the basis of those scores. The primary intention of the list was rather to propose an instrument for systematic 'self-evaluation'.

However, it may be tempting to use the results for some sort of comparison as well. Before discussing this issue, something has to be said about a general point of criticism that may be put forward against the list as such. Some people may rightly maintain that the items in the list are to some extent not entirely independent of each other. For example: the chances are that countries with good statistical legislation will also have good provisions with regard to confidentiality and prevention of political interference. Nevertheless, it is suggested that the inter-dependence of the items is not so strong that scores for individual items are meaningless, and that the overall results will be strongly biased by these inter-dependencies.

If this is accepted, three other measurement questions remain to be solved: the weights of the items, the points to be given and who sets the scores.

Obviously, not all the above issues will be considered as having the same importance. Nevertheless, since it will be impossible to agree on what weight should be given to each individual item, it is proposed to simply use equal weights.

As for points, an equally simple solution is proposed: a five point scale, in which 5 points are given for 'very good', 4 for 'good', 3 for 'fair', 2 for 'poor' and 1 for 'very poor'. With respect to 'who sets the scores', the reality is that only the senior manager(s) of

each national statistical office will be in a position to judge their own agency's performance on each of the criteria[25].
The maximum score to be achieved, then, is 125 points on the Principles and 50 points on the Practice. Assuming that managers set the scores fairly and as objectively as possible, I would suggest that scores of 100 and 40 are perhaps too good to be true. The principal worry of statistical offices in that category should probably be not to become complacent.

For the benefit of those offices interested in finding out what their own score would be, the appendix contains a scoring card[26].

For further information, please contact Willem de Vries at *wvrs@cbs.nl.*

Scoring card

Very good = 5
Good = 4
Fair = 3
Poor = 2
Very poor = 1
Blanks 0 points

Part 1

1. Development of mechanisms to ensure that work programmes are relevant
2. Development of mechanisms to assess user satisfaction with statistical products
3. Adherence to the obligation of impartiality
4. Freedom from political interference with statistical results
5. Adherence to the principle of equal access under equal conditions
6. Systematic promotion and sharing of professionalism
7. Improving methodology on a scientific basis
8. Survey design and methodology based on professional criteria only
9. Systematic efforts to train and re-train staff
10. Systematic promotion of statistical quality management
11. Systematic providing of adequate meta-data
12. Systematic presentation of the quality of statistics
13. Systematic education of key users in order to prevent misuse of statistics
14. Systematic efforts to achieve the best possible 'data sources mix'
15. Systematic efforts to improve cost-effectiveness
16. Systematic efforts to improve timeliness of statistics
17. Systematic efforts to reduce the response burden
18. Cost-effectiveness in terms of statisticians/budget/ population ratios
19. Rules and practices to prevent disclosure from printed publications
20. Development of methods to supply micro data files, preventing disclosure
21. Quality of the statistical legislation
22. Development of practices for honestly dealing with respondents
23. Development of national statistical coordination mechanisms
24. (Flexible) Adherence to international statistical standards
25. Involvement in international statistical cooperation

Part 2

1. Quality of the annual national accounts
2. Quality of the quarterly national accounts
3. Quality of the labour statistics (unemployment rates)
4. Quality of statistics on income distribution
5. Quality of basic demographic statistics
6. Quality of the external trade statistics
7. Quality of the retail trade index
8. Quality of statistics on the services sector
9. Quality of the industrial production index
10. Quality of the consumer price index

Notes

1) Deputy Director-General of Statistics Netherlands. The opinions expressed here are personal and do not necessarily reflect Statistics Netherlands' position or policies. The author thanks Ad Willeboordse, Lidwine Dellaert, Henk van Tuinen, Wouter Keller and Johan Lock for their useful comments on a first draft.
2) Theoretically, there is of course a distinction to be made between 'system' and 'office'. In countries with a decentralised statistical system, the 'system' consists of a collection of 'national statistical offices'. Throughout this article I refer to the systems as a whole, even though I may from time to time use the term 'national statistical office' or 'institute' (NSI for short, this being the commonly used international term). Obviously, measuring the performance of a (decentralised) 'system' may in practice be more complex than measuring the performance of single 'offices', but this article is not so much about the technicalities of measuring.
3) Some Principles (e.g. the one on confidentiality) involve so many complex issues that they may be (and indeed sometimes are) the subject for regular meetings or full-fledged conferences of experts.
4) Statisticians are, naturally, keen on 'how to measure or quantify'. The discussion of measuring techniques, however, is beyond the scope of this paper. The author would hope that this article stimulates the debate about 'how best to measure performance in practice'.
5) One may argue that this is an awkward and tricky question. What if an NSI is very active in marketing and measuring user satisfaction, but gets poor results (low user satisfaction) in return? Does it score high on this issue or not? My assumption is, however, that an NSI which shows this kind of real user orientation, will in the end almost unavoidably improve its performance in this regard.
6) This may be a specific Dutch problem. The policy of Statistics Netherlands is not to give in to such pressure.
7) At Statistics Netherlands, this is still under discussion. It will definitely not be applied to really sensitive statistics, such as the CPI and others.
8) Under this system, members of the press are literally locked up in a room, some time before the moment of official, pre-announced release of the statistics. The journalists are then presented with the statistics to enable them to compose their article or message. The room is equipped with computer facilities and telecommunication equipment. However, telecommunications are of course blocked until a central switch is turned on.
9) Discussions on how far 'free' should go, however, are still inconclusive. Some argue that *all* available statistics should be supplied free of charge on the Internet, others think that only some basic information should be free, while for further details some charge should be paid. The second point of view would be consistent with the most common practices followed for printed material: limited sets of material (e.g. some photocopies) are free, users who need more have to pay the marginal cost of the data carrier plus postage, occasionally even for the extra work involved to compile alternative tabulations etc.

10) The universally agreed standards of professional ethics for statisticians are laid down in the *Declaration of professional ethics* of the International Statistical Institute, 1985.
11) The organization of national statistical services: a review of major issues (New York, 1977)
12) Report of the American President's Reorganization Project for the Federal statistical system (Washington, DC, 198
13) Having units in statistical offices whose main tasks are analytical work and giving methodological advice may not be essential in this respect, but is certainly helpful to promote professionalism.
14) Which of course includes the question: how far and how deep should this information go? Experience shows that some users are very deeply interested in 'what's behind the numbers', while others, to put it bluntly, couldn't care less.
15) A special point of concern is to ensure that the data elements in a time series are consistent and if not, to inform the users clearly about the exact nature of any inconsistencies.
16) This process was initiated with a paper about Development of Standards for Dissemination of Economic and Financial Statistics to the Public by Member Countries, IMF, 1995.
17) It may be argued that the fundamental principle in question is perhaps too defensively worded and that the real issue is that NSI's, more in general, should make an effort to educate and train the users, not so much in order to prevent misuse, but to promote the best possible use.
18) See for example: 'Reducing the response burden; some developments in the Netherlands', by Willem de Vries et al.; International Statistical Review, 2/1996).
19) The Economist used the cost of statistics as such more as a background variable than as a performance indicator in its own right. Performing well at a relatively low cost was of course regarded as an additional positive feature. It may be argued that The Economist's 'formula' is unfair for smaller countries and that something like statistical budget / $\sqrt{}$ population is a more adequate measure.
20) For a comprehensive analysis of some major issues one may wish to read Administration and statistics by W. Begeer et al; Eurostat 1984.
21) In the sense that data files which are kept for statistical or research purposes only, are not subject to the general rule that individuals are entitled to check what is registered about them in the files, as well as to correct this information if they so wish.
22) Statistics Netherlands has developed the so-called ARGUS software to check files on disclosure risks.
23) In the Netherlands statistical activities outside Statistics Netherlands are insignificant in size. Nevertheless, Statistics Netherlands has recently set up a small unit to monitor such activities and to advise other government agencies on how their statistical needs may be fulfilled by them.
24) In the case of the European Union, moreover, these standards are legally imposed on Member States.
25) A better alternative would perhaps be to ask statisticians for their scores about other statistical systems than their own, but the reality is that this would take too much effort, because one would need to collect and study quite some material to do so in a more or less satisfactory manner.
26) Eleven senior managers at Statistics Netherlands scoring in accordance with this list in 1997 produced an average item score of 3.6 points, the extremes being 3.8 on the high end and 3.5 on the low in 1997.

Imputation, the alternative for surveying earning patterns

Eric Schulte Nordholt

In 1997 a research project explored the possibilities of compiling statistics on earning patterns in the Dutch economy by matching data collected in the Annual Survey on Employment and Earnings, records from the Social Security Files and data from the Labour Force Survey (LFS) in order to provide internationally comparable data. The problem of missing values can either be solved by imputing or by weighting methods. In matching records their number proved to be too large to use the random hot deck method. Therefore a sequential hot deck method was opted for, in which the multiple donor problem does not play a major role. Applying state of the art statistical methods enables Statistics Netherlands to get census like information on earnings without having to set up a separate and elaborate survey.

1. Introduction

Information on the structure of earnings is collected every few years. In 1989 payroll data were collected from companies. However, payroll administration data on educational levels and other background characteristics of employees are fairly inexact. To obtain more reliable data on earning patterns and be able to increase the frequency of these statistics, Statistics Netherlands decided to explore the possibilities of obtaining the information needed by matching the records of three main source statistics.

The three sources used in the matching procedure are quite distinct.

- The Annual Survey on Employment and Earnings collects mainly payroll data from the full range of establishments. The public sector is well represented in this business survey.
- The Social Security Files contain an even larger number of records. In this source the private sector is very well represented, but the number of variables is smaller.
- The Labour Force Survey (LFS) is a household survey which collects data on the employment situation, but also on education and occupation.

As the purpose of the project was to give a description of the earnings structure in 1995, only some selected variables from the three sources were considered relevant, although more variables were added for matching, imputation and weighting procedures. Data on 1995 were available from the Annual Survey on Employment and Earnings and the Social Security Files. To obtain reliable data from the Labour Force Survey, data on 1994, 1995 and 1996 were combined. This may have affected some variable scores, but working with cumulative data on three consecutive years was seen as an acceptable compromise between merely relying on 1995 data or opting for data for more than three years. For reasons of expediency, the few survey records with a missing score on one of the relevant variables were dropped. The remaining records were raised to the population totals in the weighting process.

When the records were lined up, the matching process was started. Payroll and social security file micro records were matched with LFS micro data, using linking variables: address, postal code, city, date of birth and sex. Only exact matches were allowed since the aim of the project was a structural analysis of earning patterns. However, this requirement does not prevent mismatches: i.e. records which refer to the same statistical unit according to the sources, but not in real life. Missed matches may also occur: typing errors in postal codes often cause matches which should turn up to be missed.

Analysis showed that mismatches and missed matches notwithstanding, an extensive set of matched records could be obtained from the three sources. This set was reliable enough that no recourse had to be taken to synthetic matching procedures. The payroll data from the business survey were taken as the general framework. Also included were some of the records from the social security files which belong to the target population of the Annual Survey on Employment and Earnings, but were not available from that source due to the sample design or because of non-response. Such social security file records which matched survey data were included in the data set.

The resulting micro data set consisted of a compilation of five subsets which contained the records of matched sources. Table 1 shows these five subsets and the number of records they contain, and also which groups of variables were chosen from which source.

2. Imputation strategy

In subset 5 some variables were missing. This problem of missing values was solved by imputation. Auxiliary variables in this imputation process were the variables available in both the payroll and the registration files.

There are various sequences in which the imputation may take place and the composite variables derived. The first alternative is to impute all relevant variables in the data sources and then to deduce all composite variables. Although this presents hardly any

Table 1
Outline of the record sets in the structure of earning statistics

Subset	Source			Variables from payroll and register files	Variables from payroll	Variables from the Labour Force Survey	Number of records
	pay rolls	registrations	survey				
1	yes	yes	yes	payroll data	payroll data	hh survey data	21,105
2	yes	yes	no	payroll data	payroll data		806,489
3	yes	no	yes	payroll data	payroll data	hh survey data	19,995
4	yes	no	no	payroll data	payroll data		733,084
5	no	yes	yes	register data		hh survey data	84,977

problems, the disadvantage of this approach is that several data sets have to be treated separately, which means that many variables have to be imputed. This would take time, and did not fit the tight time schedule. A second alternative was first to derive all composite variables and finish with the imputation. This approach is quite feasible since all composite variables can be derived for the records in the other subsets and subsequently imputed for subset 5. However, a disadvantage here is that no optimal imputation will result when composite variables are imputed. Therefore a middle course was adopted: first some composite variables were deduced, then the imputation took place and finally the last composite variables were deduced on the basis of imputed data.

Subsets 2 and 4 do not have values for Labour Force Survey variables. This problem was tackled in a weighting procedure as imputation is undesirable here for a number of reasons. In the first place not much auxiliary information is available to base a large scale imputation upon. Secondly, a mass imputation would give dramatic results, if not only the imputed variable is analysed but also the crossing of the imputed variable by other variables not taken into account in the imputation process. Analysing the structure of the data set was the main aim of the project and therefore it was recommendable to limit ourselves to the weighted and imputed data of the subsets 1, 3 and 5. The imputed data set is raised to the population totals in the weighting process.

3. The sequential hot deck method

The missing scores on the payroll variables were imputed using some auxiliary variables that were available from both the Annual Survey on Employment and Earnings and the Social Security Files. Examples of the variables that have to be imputed are gross wages per month, gross wages for overtime per month and the number of holidays. The auxiliary variables for the imputation are sex, type of employment contract, age, gross wages per day and economic sector. All together, 26 payroll variables were imputed using these five auxiliary variables. The classification of the auxiliary variables was chosen in such a way that it resulted in homogeneous groups containing approximately the same number of records. As there is a big difference in the scores on the variables that have to be imputed between different economic sectors, this auxiliary variable was categorised in homogeneous groups that do not all contain approximately the same number of records. The auxiliary variables appeared to be of good quality and did not contain missing values themselves.
As deterministic imputations distort the distribution of the imputed variable and as distributions of variables were of major concern in the statistics we were aiming at, stochastic imputations were necessary in the imputation process. The question was which stochastic imputation method was best suited. An easy choice would have been a stochastic regression imputation, but this does not always lead to feasible imputed values. Therefore, a hot deck method proved a better alternative. As the number of records was too large to use the random hot deck method, the sequential hot deck method was selected as imputation method. A random element is introduced in this method by sorting the non-imputed data set randomly before the imputation process starts. Help arrays with so-called potential donor values were created for 2,240 different combinations of scores on the categorised auxiliary variables.

The first time a missing value from a record is found with that combination of scores on the categorised auxiliary variables, the first score of the relevant help array is copied. The second time, the second score of the help array is copied, and so on. If all records of an array have been used once, a second iteration through the help array starts. The record from which the imputed value is copied is called the donor record. Although we encountered 1,580,673 potential donor records (subsets 1-4) that could be included in the help arrays, we needed to be careful with the number of categories of the auxiliary variables. If too many of these categories were created, the risk would arise of a record having to be imputed with an empty help array, which is obviously not feasible. If the help array contains a few values but many records have to be imputed using this array, there is the problem of multiple donors use which will often lead to underestimation of the variance of the imputed variable. Therefore empty or almost empty help arrays have to be combined with other help arrays. Methodologically this corresponds with the introduction of a priority ordering of the help variables in the random hot deck method. Also, in that case we cannot impute all records using all auxiliary variables categorised in the finest categorisation we have available.

To avoid inconsistencies between related variables as a result of the imputations, record matching was used for the imputation. This means that related variables were imputed simultaneously using the same imputation model. In this way covariance between imputed variables were better preserved, which is important for the analysis.

4. Multiple donors

Having finished the imputation, the problem was obviously to find out how accurate the imputations were. As the real values were not known, this is a difficult problem. Although the performance record of an imputation method in simulation experiments may inspire confidence, it is never certain how accurate it will turn out to be in practice. An important criterion in judging a hot deck imputation is to see whether we encounter the problem of the multiple use of donors. Table 2 presents the distribution of the 83,793 used donor records to impute subset 5 by the number of times that these donor records were used.

Table 2
Used donor records by the number of times that these records were used

Number of times used	Number of donor records
1	83,139
2	502
3	45
4	43
5	22
6	7
7	12
8	5
9	11
>=10	7
total	83,793

From Table 2 we see that some donors are used more than once, but that most used donor records are only used once. The maximum number of times a donor record was used in the imputation process is 16. Therefore we conclude that the multiple donor problem does not play a major role in the current imputation process and this will increase our confidence in the applied strategy.

For further information, please contact Eric Schulte Nordholt at *esle@cbs.nl*

The use of the calibration estimator in industrial waste statistics

Cees N.J. van Beusekom and José M. Gouweleeuw

This article describes the use of the calibration estimator in the compilation of the Dutch industrial waste statistics for 1994. This calibration estimator achieves a consistent estimation of the amount of waste generated per economic activity and at regional level. A comparison is made between the calibration estimator and the separate ratio estimator, which had been used previously for the industrial waste statistics.

1. Introduction

Statistics Netherlands has been compiling waste statistics within the framework of environmental statistics since the early eighties. One particular series of waste statistics is concerned with industrial waste. The methodology for these industrial waste statistics is described in Van Beusekom (1994).

The two-yearly sample survey on which the industrial waste statistics are based has been conducted among large industrial companies since 1978. A stratified sample of around 2,000 companies is drawn from the units in the General Business Register of Statistics Netherlands[1]. The companies are stratified according to economic activity[2] and size class (based on the number of employees). The units included in the sample receive a questionnaires for each local establishment of the company, making it possible to calculate regional figures. The forms comprise questions on the yearly quantities of waste produced. A primary distinction is made between process waste, like organic and inorganic waste, and process-independent waste, for instance packaging waste. In total approximately 130 different kinds of waste are distinguished.

The results are presented for a number of economic activities (SIC) and for each province. In the industrial waste statistics for 1992 and earlier years, the estimate for the total amount of waste for each economic activity and the estimate for the total amount of waste for each province were determined separately, by using the separate ratio estimator. However, as described below, this method contained the possibility of inconsistent estimates. For the 1994 industrial waste statistics a new estimator, the so-called calibration estimator, was used. This estimator ensures that the results for economic activity and the results for the provinces are consistent.

2. The separate ratio estimator

The industrial waste statistics are based on a sample survey. A stratified sample of some 2,000 companies in industry is drawn from the units in the General Business Register. The companies are stratified according to size class (based on the number of employees) and economic activity on a 2-digit level (NACE classification).
It is assumed that the waste generation of companies within one stratum is homogeneous, which means that the waste generated by each company within the stratum is comparable with respect to quality and quantity. If a company is drawn in the sample, each of its local units receives a questionnaire. These local establishments are the responding units. Non-response is assumed to be random, so the total response can be considered as a random sample

In the first survey the population only contained companies with more than 50 employees, but for subsequent surveys the population was expanded to include smaller companies with 10 to 50 employees. The total survey population contains about 10,000 companies.

In order to estimate the total amount of waste from the sample, it is assumed that in each stratum there is a linear relationship between the amount of waste generated by a company and the number of employees of this company. Furthermore it turned out that the number of employees in the population is easy to provide on a stratum level. (The number of employees is the only useful variable for which the population totals are available on a stratum level, therefore it is the only variable that can be used as auxiliary information.) In order to estimate the total amount of waste (the target variable) in a stratum, the separate ratio estimator is used. The estimate for the total amount of waste in a stratum is obtained by computing the average amount of waste per employee in the sample in that stratum and then multiplying this average amount by the number of employees in the population in that stratum. In formula

$$\hat{Y}_h = X_h \frac{y_h}{x_h} = \frac{X_h}{x_h} y_h \quad \text{(the separate ratio estimator)} \qquad (1)$$

in which:
- h is the number of the stratum
- $\hat{Y}_h$ is the estimated amount of waste in the population of stratum h
- X_h is the number of employees in the population in stratum h
- y_h is the amount of waste in the sample in stratum h
- x_h is the number of employees in the sample in stratum h.

The expression after the first equals sign means that first the average amount of waste per employee is computed (y_h/x_h) and then multiplied by the total number of employees (X_h). The expression after the second equals sign means that the amount of waste of every company in the sample of stratum h is multiplied by X_h/x_h (the ratio of the number of employees in the population in stratum h and the number of employees in the sample in stratum h) and after multiplication the results of all companies are added. The number X_h/x_h is then called the weight of the companies in stratum h.

In the industrial waste statistics, estimates of the total amount of waste are given, as well as estimates for the total amount of waste generated per economic activity (according to the SIC) and estimates for the total amount of waste per province. An estimate for the total amount of waste generated in an economic activity can be obtained by adding the estimated amount of waste over all strata that belong to that economic activity. In order to estimate the total amount of waste in a province, the local units are stratified according to province and size class of the local unit. Within each province $\times$ size class of the local unit, the separate ratio estimator is again used, with the number of employees of the local unit as auxiliary information (X_h and x_h in formula (1)). The total amount of generated waste is the sum of the estimated quantity in the different economic activities, or the sum of the estimated quantity in the different provinces of the local units.
These two sums should be equal, but they are not! There is a difference of a few percent each year. This is one of the drawbacks of this estimation method.

Another problem of the separate ratio estimator was caused by the fact that some respondents reported a relatively large value of generated waste in the questionnaire. After estimation, i.e. multiplying the amount of waste by the weight (X_h/x_h), this value would contribute to the estimated amount of waste in a stratum in such a way as to result in unacceptably large amounts of waste, i.e. $\bar{Y}_h$ becomes unacceptably large. These values are corrected for in the estimation process. Respondents with such a large amount of waste are not taken into account in the calculation of $\bar{Y}_h$ in formula (1). The estimation is conducted without these units and afterwards the excluded large amount of waste is added to the stratum. So the weight for units with a large amount of waste is set equal to 1.

This correction method of reducing the influence of unacceptably large values has been used for several years. The disadvantage of this method is that within a stratum different weights (namely X_h/x_h and 1) are generated for one target variable. From a computational point of view, it is better not to change the weight but to change the large value of waste in such a way that multiplied by the calculated weight (i.e. X_h/x_h), the value given by the respondent is returned. The advantage is that the unit is not kept out of the estimation. The risk of this method is that it gives an underestimation if the extreme value is not unique in this stratum. A very good knowledge of the practice of the whole population in the stratum is prerequisite.

In order to overcome the two above-mentioned problems, another estimation method, the so-called calibration method, was used for the 1994 survey. The advantage of this method is that the estimates for the economic activities are consistent with the estimates for the provinces of the local units, i.e. if the estimates are summed over the economic activities then this equals the sum over the provinces. The calibration method is discussed below.

3. Consistent estimation of economic activities and provinces

In the estimation process, the responding unit (local unit) is used twice. Once for the estimation in an economic activity (where the local unit appears as part of a company) and once for the estimation in a province. As stated above, these two estimations are not consistent, due to the method of estimation. Part of the problem is that the data on the population of the economic activities was inconsistent with the data on the population of the provinces, since they are gathered from two different sources at Statistics Netherlands. Furthermore, the units in a size-class of the local units × province do not constitute a random sample.
In order to obtain consistent estimations for provinces and economic activities, the calibration estimator is investigated. This calibration estimator was introduced in Deville and Särndal (1992). The use of the method to obtain consistent estimations is extensively described in Gouweleeuw et al. (1997). When the calibration method is used, the local unit is given the same weight as the company to which it belongs. In formula, the calibration estimator is given by:

$$\hat{Y}_{cal} = \sum_k W_k y_k \qquad (2)$$

in which:

$\hat{Y}_{cal}$ is the calibration estimator for the amount of waste
y_k is the amount of waste of company k
W_k is the calibration weight of company k.

So if the amount of waste in a stratum has to be computed, then first all the companies in that stratum are multiplied by their weight (these weights may all be different). After multiplication, the product is summed over all the companies in the stratum.
The calibration weights W_k have to satisfy the following conditions.
1. They have to be as close as possible (according to a distance function) to the inclusion weights. This inclusion weight is the ratio of the number of companies in population of a stratum and the number of companies in the response of that stratum.
2. They have to satisfy a number of conditions, the so-called calibration equations.

Condition 2 can be used to obtain consistent estimations for economic activities and provinces. The local units are now grouped according to province and cluster of economic activity (five clusters), instead of according to province and size class of the local unit, as information on the total number of employees for province × size class of the local unit could no longer be obtained.
With the calibration equations the following is obtained.

– Every local unit receives the same weight as the company to which it belongs.
– The estimated total number of employees of companies in a size class × economic activity must be equal to the known total number of employees in the population in this size class × economic activity. Similarly the estimated total number of employees of local units in a province × cluster of economic activities must be equal to the known total number of employees in the population in this province × cluster of economic activities. Again, the underlying thought here is that there is a linear relation between the number of employees in a stratum and the total amount of waste produced in that stratum.

Since the known population totals are all consistent with each other, these calibration equations lead to consistent estimations for economic activities and provinces.
For the 1994 statistical data the calibration estimator was used to estimate the total of seven different waste flows. Again, one respondent reported a large amount of waste that was not considered representative. As in previous years, this company was assigned a weight W_k equal to 1.
This could easily be obtained by adding another calibration equation, stating that $W_k = 1$.

Table 1
Results of the different estimators

Waste type	Ratio estimator		Calibration estimator	
	estimate Kton	relative-margin	estimate Kton	relative-margin
inorganic waste	551	18%	545	21%
organic waste	1,928	9%	1,843	7%
paper from indus. proc	412	15%	409	15%
waste paper	155	8%	162	6%
mixed process waste	297	20%	287	12%
mixed non-process	394	7%	391	8%
sludges	1,500	14%	1,590	9%
total	5,237	6%	5,227	5%

In order to get an impression of the quality of the estimates obtained with the calibration estimator, the results are compared with those that would have been obtained using the separate ratio estimator. For the 1994 survey, the total amounts of seven different types of waste were estimated, using both estimators. In addition, margins were determined. For the general mathematical

formulas of the margins, the reader is referred to Deville and Särndal (1992) and Särndal et al. (1992).

The results are presented in Table 1. It is clear that the difference in accuracy for the two estimators is small.

4. Conclusions

The conclusion that can be drawn from this survey is that the calibration estimator used in the industrial waste statistics does not show significant differences in accuracy compared to the ratio estimator, as far as the estimation for total of the seven different waste flows is concerned. No conclusion can be reached about accuracy concerning figures on waste per province or per economic activity (SIC).

The advantage of the calibration method is the generation of consistent figures for the estimations on economic activity level and on regional level. Furthermore, the extreme values of waste are treated in a better way from a statistical point of view.

For further information, please contact Cees van Beusekom at *cbsm@cbs.nl* or José Gouweleeuw at *jgww@cbs.nl.*

References

Van Beusekom, C. (1994) Industrial waste statistics in The Netherlands, Document no.6, Conference of European statisticians, Helsinki, September 1994

Deville, J.-C. & Särndal, C.-E. (1992) Calibration estimators in survey sampling, Journal of the American Statistical Association 87 (418), page 376-382

Gouweleeuw, J.M., Heerschop, M.J. and Van Huis, L.T. (1997) Consistent estimation of kind of activity units and local units using the calibration method. Research paper no. 9702, Department of Statistical Methods, Statistics Netherlands.

Särndal, C.E., Swensson, B. and Wretman, J. (1992) Model assisted survey sampling. Springer Verlag, New York.

Notes

1) Statistics Netherlands General Business Register (Algemeen Bedrijfsregister, or ABR) contains the general information - name and address, economic activity and size class of all companies in the Netherlands, and their local establishments.
2) Since 1993, the Standard Industrial Classification of economic activities (SIC) used by Statistics Netherlands, is based on the European nomenclature of economic activities: NACE version 1 (1990).

Employee training and productivity in Dutch manufacturing firms

Martin Boon and Ben van der Eijken

This paper examines the impact that employee formal training, which contributes to the human capital stock of the firm, has on firm output in the Dutch manufacturing sector, using a production function framework. Our empirical analysis uses linked firm-level data for the years 1990 and 1993. After correction for unobservable firm-specific effects, we find that the output elasticity of human capital is insignificantly different from zero for gross output and 0.07 for value added. From the estimated output elasticities we can derive that - with a depreciation rate of 5 percent (15 percent) for human capital - the private rate of return on human capital is insignificantly different from zero for gross output and 25 percent (56 percent) for value added. The results suggest that investment in human capital have a significant and positive effect on value added for manufacturing firms.

1. Introduction

Increasing competition and technological changes are forcing firms to invest more in the training of their employees, thus contributing to their human capital stock. Investment in training directed to higher product quality is of utmost importance for Dutch firms which have to compete with firms in low-wage countries. It is generally assumed that productivity is higher in firms with a better-trained workforce, the argument being that skilled employees are able to adapt more easily to new production processes and new products.

The issue of productivity growth has attracted much attention from the science of economics. Recent contributions to growth theory emphasise that the concept of capital has to be broadened to include not only physical capital, but human and R&D capital as well. Human capital can be captured in terms of the number of years and type of education and in-company training, while R&D capital concerns knowledge not directly embodied in labour. The effects of education and R&D on productivity growth have already been measured by other researchers. OECD (1994) reviews a number of studies on the influence of initial skills of workers on firm performance. An example of a study on the effects of education (vocational qualifications of the labour force) on productivity growth is O'Mahony and Wagner (1996). Mairesse and Sassenou (1991) have given an overview of econometric studies on the R&D productivity. However, there is little empirical evidence to assess the quantitative impact of further training provided by the employers on productivity performance.

Most empirical work on the link between training and productivity has been based on employee-level surveys to focus on the characteristics of those who have been trained. The majority of these studies, for instance Groot (1994) and Bishop (1994), used a subjective measure of the individual productivity of workers such as the answer to the following question: on a scale of 0-100 how has your productivity changed because of training? Groot (1994) found that after participation in a formal training programme employees are on average 16 per cent more productive than before. Bishop (1994) concludes that formal off-the-job training increases worker productivity by 16 per cent. However, the problem with subjective measures of productivity is that they are not comparable across firms or even within firms over time.
Other studies used objective micro-data on firm characteristics to estimate the magnitude of the yield for firms which train their employees. Using longitudinal firm-level data, Bartel (1991) found that training programmes resulted in productivity increases in the order of 17 per cent. Lynch and Black (1995) found, using cross-sectional firm-level data, that significantly positive effects on firm productivity are associated with certain types of employer-provided training. Tan and Batra (1995) showed, by using cross-sectional data for individual firms in developing countries, that employer investment in formal training has a large and significant impact on value added. The findings of the above-mentioned firm-level studies are subject to some limitations, the most important being that they did not use a quantitative measure of the accumulated stock of human capital over time within a firm. In other words, they did not consider the fact that training expenditure accumulates into a stock of human capital. Both Bartel and Tan and Batra used as a training measure a dummy variable indicating whether or not the firm provided any formal training to its employees. Lynch and Black considered the numbers of workers trained, time spent in training and dummies for specific training types for a given year. Another limitation of both Lynch and Black and Tan and Batra is that their estimates could be biased because of unobservable firm characteristics.

This paper examines the impact that employee formal training has on firm output in the Dutch manufacturing sector, using a production function framework. Our empirical analysis uses linked firm-level data from the training surveys, the production surveys, and the wage and employment surveys for the years 1990 and 1993.

2. Production function framework

First let us describe the framework for analysing the impact of employee training on firm output. Investment in employee training accumulates into a firm's stock of human capital, similar to the way physical capital is formed through investment in fixed assets. The theoretical framework for this article will be a production function with human capital as a separate input.[1)] We adopt the log-linear specification of the Cobb-Douglas production function:

$$q_{it} = \alpha + \beta h_{it} + \gamma c_{it} + \varphi l_{it} + \omega m_{it} + \theta d_t + \varepsilon_{it} \qquad (1)$$

where:
- q_{it} is the (log of) output of firm i in year t
- α is a constant
- h_{it} is the (log of) human capital stock
- c_{it} is the (log of) physical capital
- l_{it} is the (log of) labour
- m_{it} is the (log of) materials
- d_t is a year dummy which is a time-specific indicator of the level of disembodied technology, and
- ε_{it} is a disturbance term.

If output is measured by gross output then the input set consists of the above-mentioned factors, and if it is measured by value added then material input is excluded from this set. The parameters β, γ, φ and ω are the elasticities of output with respect to the inputs.

We can estimate equation (1) on pooled cross-sectional data, under the assumption that the disturbance term ε_{it} has mean zero and a constant variance. However, it is likely that the error term ε_{it}

comprises heterogeneity across firms in their technologies and type of output and this will introduce a firm-specific effect τ_i. In symbols

$$\varepsilon_{it} = \tau_i + \upsilon_{it} \qquad (2)$$

where υ_{it} denotes the remainder disturbance. Under error-component model (2) the pooled estimates of the standard errors of the coefficients are biased. To allow for the firm-specific effects, we use fixed effects and random effects panel estimators. To be precise, in the fixed effects specification the firm-specific effects are assumed to be fixed parameters which have to be estimated, while in the random effects specification the firm-specific effects are assumed to be random variables. It can be shown that if the data consist of only two years the fixed effects approach gives the same result as estimating the model on the first (or long) differences.

It is difficult to measure the accumulated stock of human capital of a firm. The situation for human capital is comparable with R&D capital. Just as training expenditure enhances the stock of human capital, R&D-expenditure accumulates into the stock of R&D capital of a firm. Hall and Mairesse (1995), followed by Bartelsman et al. (1996), applied two alternative methods for measuring R&D capital as a separate input into the production function. We shall use one of these methods, the so-called stock approach, for estimating the production function with human capital as input.

The stock approach calculates the human capital stock of this year as the sum of this year's investment in training and the human capital stock of last year (minus depreciation). In symbols:

$$H_{it} = E_{it} + (1-\delta)H_{i,t-1} \qquad (3)$$

where:

H_{it} is the stock of human capital of firm *i* in year *t,*
E_{it} represents training expenditures and
δ s the rate of depreciation.

This implies that past training continues to have spillover effects on output in the present, although the effect may diminish over time through depreciation. The depreciation is supposed to reflect the obsolescence of skills with age. The content of formal training can vary, from courses related to specific firm's activities (such as machinery operation and quality control) to those related to the firm's general operation (management techniques, accounting, foreign languages, etc.). The magnitude of depreciation differs between the various formal training programmes. Know-how of computer systems, for instance, becomes obsolete faster than knowledge of management techniques. The magnitude of yearly depreciation is usually chosen in the 1 to 12 percent range (see De Mooij (1997) for an overview).

In our data set training expenditure is observed only in the years 1990 and 1993 and no initial human capital stock measure is available. Training expenditure for the intervening years is interpolated using the observed growth rate for each firm. Following Hall and Mairesse (1995), the human capital stock for the year 1990 can be written as:

$$H_{i,90} = \frac{E_{i,90}}{(g+\delta)} \qquad (4)$$

where g is the pre-sample annual growth rate of training expenditure, which is assumed constant across firms. From equation (3) we can derive the following expression for the human capital stock in year 1993:

$$H_{i,93} = (1-\delta)^3 H_{i,90} + \sum_{s=1}^{3}(1+e_i)^s(1-\delta)^{3-s}E_{i,90} \qquad (5)$$

with e_i the annual growth rate of training expenditure for firm *i* in the period 1990-1993.

From the estimate of the output elasticity for human capital (β) we can derive an estimate for the rate of return of human capital (ρ) in the following way. For the Cobb-Douglas specification the marginal product of the human capital stock, ρ, is equal to the output elasticity of human capital times the ratio of output (Q) to the human capital stock (H):

$$\rho \equiv \frac{\partial Q_{it}}{\partial H_{it}} = \beta \frac{Q_{it}}{H_{it}} \qquad (6)$$

The parameter ρ can be interpreted as the amount by which output increases with an increase in training expenditure, i.e. as the private, gross (i.e. including depreciation) rate of return of human capital. In actual estimation it is assumed that for every firm both the output and the human capital stock is equal to the corresponding average over the firms in our data set.

3. Data description

The data used in this study concern information on individual firms in the Dutch manufacturing sector for the years 1990 and 1993. The data were created at Statistics Netherlands (CBS) by linking micro data of the training survey, the production survey and the wages and employment survey.

The training survey asks companies in the private sector with five or more employees to provide information on formal training which is financed wholly or partly by firms. The sampling design for this survey has two phases: first a large sample of firms is surveyed with a limited set of questions about the training activity. In the second phase a subsample is drawn from the responding firms which had been active in this respect. The latter firms receive a comprehensive questionnaire about training expenditure, training participation, number of hours worked of (in-firm) trainers and number of training days (during working hours). Training expenditure is desaggregated into wage costs of lost working time and of training staff, and material costs (which consist of fees of training institutes, compensation of study fees, travel and accommodation expenses). See Slagter (1995) for more details.

In the annual production survey manufacturing companies are asked for detailed information on inputs and outputs. This information contains, among other things, sales, gross output, gross value added (at market prices), wage bill, number of employees, costs of materials, electricity use and capital consumption allowances (depreciation costs). Since 1987 all firms with twenty or more employees are surveyed.

The wages and employment survey gives information on number of hours worked and wages for all companies with employees. This survey is based on a two-stage sample design: first a stratified sample of companies is drawn and then each sampled firm takes a simple random sample of its employees.

Table 1
Summary statistics for the balanced panel of linked data (Panel) and the production survey for total manufacturing (PS), 1990–1993[a]

	Panel		PS	
Year	1990	1993	1990	1993
Gross output[d]	298	252	46	41
Value added[d]	96	91	13	11
Labour productivity[b]	112	127	105	102
Number of hours worked[e]	1,522	1,186		
Number of employees	859	714	127	108
Physical capital per employee[c]	21	20	16	17
Training expenditure[d]	2	2		
Human capital (g=0.05;d=0.05)[d]	24	26		
Human capital (g=0.05;d=0.15)[d]	12	12		
Number of firms	173	173	6,154	6,681

[a] Means.
[b] Value added per employee in thousand 1990 guilders.
[c] Depreciation costs per employee in thousand 1990 guilders.
[d] In million 1990 guilders.
[e] In thousand hours.

The nominal variables in the data set are all deflated to 1990 guilders. Output and materials are deflated by applying 3-digit SIC[2)] product and material price index numbers to all firms within the corresponding industry. Training expenditures are deflated by a composed index of wages of trainers and trainees, and of material prices. The wage changes for trainers and trainees were computed for industry groups at the 2-digit SIC-level as the change in average hourly compensation for employees between 1990 and 1993. Using firm-specific labour and material expenditure shares, the appropriate wage change was averaged with the material price change to construct training expenditure deflators.

The capital input measure required to estimate the production function is proxied by the depreciation costs. Variations in the utilisation of the capital stock can cause differences between the depreciation data and the desired measure of the flow of capital services. When the fixed effects specification of the production function is estimated, changes in the capital inputs are proxied by changes in electricity use. This measure should correct better for fluctuations in the capital usage over time. There are differences between and within firms over time in shares of full-time and part-time employees and in the incidence of shorter working hours and holidays. In order to take into account these differences, the input of labour is measured by total hours actually worked per year instead of total number of employees.

The training expenditure is separated from the other operating expenses of the firm, since the training inputs do not produce current output but are used to increase the stock of human capital. In this way we can avoid the biases in estimation of the production function caused by 'double-counting' of resource inputs (see Mairesse and Sassenou, 1991). In the production function, labour and material input variables are adjusted for the amounts used in training endeavour. This implies that labour input is defined exclusive of hours worked of (in-firm) trainers and of lost working time of trainees and that material inputs contain only non-training inputs. Value added is measured as gross output minus non-training materials.

The individual firms belonging to the cross-sectional data sets for 1990 and 1993 are linked to each other. This link results into a balanced panel consisting of 173 firms. Table 1 presents summary statistics for the balanced panel of linked data. The data have been adjusted for double-counting of training inputs. In 1993 all manufacturing firms together spent 971 million guilders on formal employee training. The firms belonging to the balanced panel contribute to 30 per cent of total manufacturing training expenditure and cover 17 per cent of total manufacturing employment in 1993.

Table 1 examines how representative the balanced panel data are by comparing them with the (original) data of the production survey for the total manufacturing sector. It appears that firms in the balanced panel are larger than in the original sample. The larger average firm size reflects the design for the training survey and the fact that the probability of employer-provided training increases with firm size. We also find that a larger portion of the firms in the balanced panel are in the chemical industry, while other manufacturing sectors account for a smaller portion.

Table 1 shows that in 1993 the average firm in our data has about 700 employees, uses physical capital per employee worth 20 thousand 1990 guilders, and produces 90 million 1990 guilders in value added. It can be inferred that the two years considered here differ somewhat. Employment, gross output and value added decreased in the period 1990-1993. We also calculated human capital for 1990 and 1993 according to (4) and (5) for two different depreciation rates (δ=5 and 15 per cent), using a pre-sample annual growth rate of training expenditure (g) of 5 per cent. The impact of the different depreciation assumptions on human capital is not so very large.

4. Estimation results for the production function

We assessed the effects of investment in training on output by providing estimates of the output elasticity of the human capital stock. We adopted production function (1) in which human capital is measured using the stock approach, and calculated human capital for 1990 and 1993 according to (4) and (5) for two different depreciation rates (δ=5 and 15 per cent), using a pre-sample annual growth rate of training expenditure (g) of 5 per cent. In

Table 2
Estimates of the gross output elasticities, sample of 173 firms, 1990 and 1993 (346 observations)[a]

	(1)	(2)	(3)
Pooled			
Labour	.171 (12.6)	.178 (12.6)	.176 (12.5)
Material inputs	.764 (75.2)	.766 (74.1)	.765 (74.1)
Physical capital	.059 (7.0)	.059 (7.0)	.059 (7.0)
Human capital	.015 (2.5)	.009 (1.4)	.010 (1.6)
R2	.991	.991	.991
Fixed effects			
Labour	.179 (5.1)	.180 (5.1)	.179 (5.1)
Material inputs	.763 (25.5)	.764 (25.4)	.763 (25.5)
Physical capital	.051 (2.3)	.051 (2.3)	.051 (2.3)
Human capital	.004 (0.7)	.019 (0.5)	.012 (0.6)
R2	.998	.998	.998
Random effects			
Labour	.189 (12.4)	.188 (11.6)	.187 (11.5)
Material inputs	.776 (66.6)	.775 (64.4)	.775 (64.6)
Physical capital	.039 (4.6)	.039 (4.6)	.039 (4.6)
Human capital	.009 (1.7)	.010 (1.3)	.012 (1.5)
R2	.997	.997	.997

[a] We included SIC-dummies for four sectors of economic activity (except in the fixed effect specification). Numbers in parentheses are t-values.
(1) Human capital is equal to training expenditures.
(2) Human capital constructed according to (4) and (5) with g=0.05 and δ=0.05.
(3) Human capital constructed according to (4) and (5) with g=0.05 and δ=0.15.

addition, we also used this year's training expenditure as a measure of human capital. We applied different estimation methods depending on the assumption concerning the error term, i.e. the pooled, fixed effects and random effects method. Estimates are based on the balanced panel data which are adjusted for 'double-counting', i.e. labour and material inputs contain only non-training inputs and value added is measured as gross output minus non-training materials. Next to the factor inputs, we have in the specifications dummies for four sectors of economic activity: 1) food, beverages and tobacco, 2) petroleum, chemical industry and allied, 3) metal industries and 4) other industries (textiles, apparel, paper and paper products, and building materials)[3].

Tables 2 and 3 present production function estimates using real gross output and real gross value added respectively as measure of the volume of output. The estimated coefficients of the production function are elasticities, i.e. they denote the percentage rise in output which results from a one percentage rise in the given input factor. The estimates for the elasticities of the factor inputs are reasonably close to the corresponding factor shares in the output value as these should be under the hypothesis of perfect competition.

For both output measures the pooled estimates show that the output elasticity with respect to human capital is positive and statistically significant at the 90 per cent level. A problem with pooled estimates is that under error-component model (2) the standard errors are biased. This problem can be solved by applying the fixed effects estimator. Before we estimated the fixed effects estimator we performed the F-test for the joint significance of the firm-specific effects. The test results indicate the presence of firm-specific effects. We find that the elasticity of human capital becomes insignificant when we control for permanent differences between firms.

Table 3
Estimates of the value added elasticities, sample of 173 firms, 1990 and 1993 (346 observations)[a]

	(1)	(2)	(3)
Pooled			
Labour	.682 (15.9)	.695 (15.5)	.687 (15.3)
Physical capital	.265 (10.2)	.269 (10.3)	.268 (10.2)
Human capital	.071 (3.7)	.057 (2.8)	.064 (3.0)
R2	.898	.896	.896
Fixed effects			
Labour	.712 (5.6)	.709 (5.6)	.710 (5.6)
Physical capital	.255 (3.0)	.256 (3.0)	.256 (3.0)
Human capital	.027 (1.1)	.151 (1.1)	.086 (1.1)
R2	.976	.976	.976
Random effects			
Labour	.757 (16.4)	.742 (14.6)	.736 (14.5)
Physical capital	.215 (7.7)	.213 (7.6)	.213 (7.6)
Human capital	.056 (2.9)	.067 (2.7)	.072 (2.9)
R2	.947	.947	.947

[a] We included SIC-dummies for four sectors of economic activity (except in the fixed effect specification). Numbers in parentheses are t-values.
(1) Human capital is equal to training expenditures.
(2) Human capital constructed according to (5) and (6) with g=0.05 and δ=0.05.
(3) Human capital constructed according to (5) and (6) with g=0.05 and δ=0.15.

In the fixed effects specification a model parameter is estimated for every firm in the balanced panel. Because of the extra parameters the fixed effects estimator results in a large loss of degrees of freedom. Since our sample size of firms and years is rather small, this can easily lead to insignificant regression coefficients. As already mentioned, there is another approach to control for the firm-specific effects, namely the random effects method. The random effects estimate of the value added elasticity of human capital differs significantly from zero. The disadvantage of the random effects model is that its estimates will be biased if the firm-specific effects are correlated with the explanatory variables. By comparing the 95 per cent confidence intervals of the fixed effects and the random effects estimates we can test whether correlated effects exist. We find that the fixed effects estimate of the human capital elasticity does not differ significantly from the random effects estimate. This implies that in our case the firm-specific effects and the human capital variable are not correlated and the random effects estimator is (approximately) unbiased. Then the random effects estimator is the preferred one, because it has a smaller variance than the fixed effects estimator. According to the random effects estimates the output elasticity of human capital is insignificantly different from zero for gross output and 0.06-0.07 for value added.

We turn to the effect of the human capital measure on the estimation results. Using a higher depreciation rate when constructing the human capital variable (compare columns (2) and (3)) makes no difference to the pooled and the random effects estimates, but gives slightly lower coefficients for the fixed effects estimates. To underline the insensitivity of the results to the choice of depreciation rate, we note that the human capital measure based solely on training expenditures gave coefficients with the same order of magnitude as the measure constructed with a depreciation rate of 15 per cent.

Using formula (6) we derive an estimate for the rate of return of human capital (ρ) from the estimate of the output elasticity for human capital (β). It is assumed that for every firm both the output and the human capital stock is equal to the corresponding average over the firms in our data set.[4)] We carry out this exercise only for the random effects estimates but for different human capital measures (except the measure solely based on the training expenditures). Then, we find that with a depreciation rate of 5 per cent (15 per cent) the rate of return is insignificantly different from zero for the gross output specification and 25 per cent (56 per cent) with a standard error of 9 per cent (19 per cent) for the value added specification.

Lastly, we want to point out some limitations of the results presented. We only considered the influence of expenditure on formal training programmes on the output of firms. However, informal training is also an important form of employer-provided training. If informal training is positively correlated with formal training, then the estimated coefficient for human capital will reflect the returns not only on formal but also on informal training. Thus, presumably the estimated rate of return on human capital is upwardly biased. Another source of bias in the estimated coefficients of the production model is the simultaneity between changes in output and investment in human capital, driven either by demand shocks or liquidity shocks. This phenomenon casts doubt on the assumption of exogeneity of the human capital variable. Unfortunately we lack instruments to take endogeneity into account using econometric methods, so we cannot quantify the simultaneity bias.

5. Concluding remarks

We have examined the impact of employer-provided formal training programmes on output using firm-level data for the Dutch manufacturing sector. With different estimation methods and different measures for human capital we estimated a production function. The data are derived from training surveys for 1990 and 1993, which are linked to the production surveys, and the wages and employment surveys.
After correction for unobservable firm-specific effects, we find that the output elasticity of human capital is insignificantly different from zero for gross output and 0.07 for value added. From the estimated output elasticities we can derive that with a depreciation rate of 5 per cent (15 per cent) the private rate of return on human capital is insignificantly different from zero for gross output and 25 per cent (56 per cent) for value added. The empirical results show that investment in human capital has a significant and positive effect on value added for manufacturing firms.
We can further compare our findings for human capital with results recently published for R&D capital for the Netherlands by Bartelsman et al. (1996). They found that the output elasticity of R&D capital is about 0.06 for gross output and 0.08 for value added, while the private rate of return on R&D varies between 12 per cent for gross output and 30 per cent for value added. This means that the rate of return on human capital is of the same order of magnitude as that of R&D capital.

For further information, please contact Martin Boon at *mbon@cbs.nl* or Ben van der Eijken at *bekn@cbs.nl.*

References

Amemiya, T., 1985, Advanced econometrics (Basil Blackwell, Oxford), pp. 385-387.

Bartel, A.P., 1991, Productivity gains from the implementation of employee training programs. Working paper no. 3893 (National Bureau of Economic Research, Cambridge).

Bartelsman, E., G. van Leeuwen, H. Nieuwenhuijsen and K. Zeelenberg, 1996, R&D and productivity growth: evidence from firm-level data for The Netherlands. Netherlands Official Statistics 11 (3), pp. 52-69.

Bishop, J.H., 1994, The impact of previous training on productivity and wages. In: L.M. Lynch (eds.), Training and the private sector: international comparisons (University of Chicago Press, Chicago), pp. 161-199.

Groot, W., 1994, Bedrijfsopleidingen: goed voor produktiviteit en loon (Employer-sponsored training: positive effects on productivity and wages). Economisch-Statistische Berichten (7 December 1994), pp. 1108-1111.

Hall, B.H. and J. Mairesse, 1995, Exploring the relationship between R&D and productivity in French manufacturing firms. Journal of Econometrics 65, pp. 263-293.

Lynch, L.M. and S.E. Black, Beyond the incidence of training: evidence from a national employers survey. Working paper no. 5231 (National Bureau of Economic Research, Cambridge).

Mairesse, J. and M. Sassenou, 1991, R&D and productivity: a survey of econometric studies at the firm-level. Science-Technology-Industry Review 8 (OECD, Paris).

Mooij, R.A. de, 1997, Belastingheffing en de vorming van menselijk kapitaal (Taxation and human capital accumulation), Maandschrift Economie 61, pp. 224-240.

OECD, 1994, The OECD jobs study; evidence and explanations (OECD, Paris), pp. 125-126.

O'Mahoney, M., and K. Wagner, 1996, Anglo German productivity performance: 1960-1989. In: K. Wagner and B. van Ark (eds.), International productivity differences: measurement and explanations (Elsevier, Amsterdam), pp. 143-194.

Slagter, H.C.A., 1995, A survey on employer sponsored training. Netherlands Official Statistics (Summer 1995), pp. 3-27.

Tan, H. and G. Batra, 1995, Enterprise training in developing countries: incidence, productivity effects and policy implications (World Bank, Washington DC).

Notes

1) It can be argued that training is embodied in the particular employees receiving the training. As these workers leave the firm, presumably they take their human capital with them. However, at the short run labour turnover can be neglected.
2) SIC stands for Standard Industrial Classification of Statistics Netherlands. The 3-digit level allocates industrial companies to 122 groups.
3) The fixed effects estimator cannot estimate the effect of any time-invariant variable like the SIC-dummies. These time-invariant variables are wiped out by the 'deviations from means' transformation.
4) We have also calculated the rate of return from the output elasticity for human capital using (6), under the assumption that for every firm the output to human capital ratio is equal to the average over the firms in our data set. However, in this case the calculated rate of return appear to be implausibly high, i.e. 3 times as large as the rate of return under the first-mentioned assumption.

Model for a question bank

Jack Claessen[1)]

This article describes the design of a model for a question bank which can serve as a tool to gain insight into the demand for statistical data and thus contribute to reducing the response burden. The model of this database for questions can be used to co-ordinate the questions posed in the various statistical surveys, and help to develop standardised phrasing for questionnaires. The co-ordination of the various questionnaires already constitutes a reduction in the response burden in itself. The background and set up of the database, the aim of the information system, and the set up and lay out of the model are discussed below.

1. Introduction

Reducing the response burden for enterprises and institutions has been one of the focal points in the policy of Statistics Netherlands for several years. Recently Statistics Netherlands carried out a study on overlap in surveys held among enterprises and institutions, and in the questions asked. One of the conclusions of this study was that a survey registration and management system in the form of a database for questions would improve internal communication and be useful for co-ordinating the questions to be included in surveys, thus improving the relationship with respondents.
The response burden issue has also aroused political interest. In 1995 the Ministry of Economic Affairs formally asked Statistics Netehrlands to specificy a number of projects that could reduce the response burden within the current government period (1994-1998).

2. Aims

The main aims of the question bank are to gain insight into the data required by Statistics Netherlands, and to reduce the response burden for the surveyed companies and institutions.
The database should be able to provide information on surveys, questions, and variables. It would be useful for Statistics Netherlands to have a survey registration and management system, containing data on surveys that have been held and that are planned, as well as data on the survey type (Electronic Data Exchange (EDI), questionnaires/forms), and the questions asked or planned.

The question bank ought to be able to function as a co-ordination and control tool in the data-gathering process and as a co-ordination and integration tool for the questions and variables.

So the question bank should serve both internal and external aims.

The *internal aims* pertain to:
– gaining insight into the data requirements of Statistics Netherlands by listing the number and the types of surveys/questionnaires sent to the respondents;
– gaining insight into the total package of questions and variables by type of respondent so that duplication can be avoided;
– setting up a tool for information, co-ordination and integration of the questionnaires.

The *external aims* pertain to:
– reducing the response burden for companies and institutions;
– providing information to respondents about the type and size of the surveys and the burden involved.

The data from this question bank should provide information on.:
- the questions Statistics Netherlands asks, i.e. the kinds of questions or variables per survey or topic;
- how often the same or similar questions or variables are asked; this gives informtaion on the overlap in the questions and/or variables addressed by the various surveys, or in observation periods (e.g. monthly, quarterly or yearly surveys).
- who has to answer these questions: what is the targeted scope? For instance, activities within the Standard Industrial Classification and a certain size class.

The main users of the question bank will be the survey statisticians, staff involved in data collection methodology, account managers and field workers. The question bank may also turn out to be of use to general policy makers.
For the statisticians who prepare, co-ordinate and integrate the questionnaires it will be very useful to be able to study and select from all previously used questions on a given topic by subject or statistical study. For managers and policy-makers it will be useful to gain insight into the data Statistics Netherlands requires.

Naturally an overall willingness within Statistics Netherlands to build and maintain the database is a sine qua non for the success of the question bank.

3. Steps, methods and co-ordination with other projects

The first step in the development of the question bank model was to list all previous studies carried out on survey overlaps and response burden.
The next step was to devise a model for a question bank pertaining to statistics and questionnaires for enterprises and institutions.
A plan for the further development of this question bank model was presented to potentially interested parties within the Brueau, and on the basis of the reactions to this plan, we decided to go ahead with the construction of a question bank prototype.

In our work on this prototype, we collaborated closely with the developers of the EDI management system, as the two projects, which were developed in the same period, faced similar problems and we wanted to prevent overlap in development and programming.
The EDI system was set up to support EDI observations. The entities incorporated in the data model for the EDI metasystem (part of the EDI management system) on questions, variables etc. are a reflection of the question bank model. The question bank is broader in scope than the EDI metasystem and covers paper questionnaires as well as EDI-specific electronic ones. A team of experts and potential users monitored the progress of the question bank project.

Statistics Netherlands started a service unit for government and other surveys (the CERO project) at the end of 1996. The question bank is expected to be able to play a major role in this project, and this fact has been taken into account in the set up of the model. When the CERO project is implemented in due course, part of the question bank will also be phased in.

4. Question bank model

The question bank has to be a user-friendly computer system, capable of generating outlines and data sets on the basis of various characteristics of surveys, questionnaires, sets of questions, individual questions and variables.

Although we started out with the database for economic statistics in mind, the model is set up in such a way that it can accommodate household and individual surveys as well.

The model is based on the assumption that a statistical study will make use of questionnaire-based surveys. A questionnaire generally consists of questions or sets of questions on a particular topic.
These questions have to be constructed in such a way that the variables pertaining to the topic are filled or can be derived, irrespective of whether the questions are on paper or in electronic format.
A broad definition of the concepts of 'survey' and 'questionnaire' would even allow the inclusion of variables from files used in secondary statistics in the question bank.

We also examined which linking variables would work if we wished to expand the question bank in the future by linking up units at the individual unit level or by linking up with other databases/systems (for instance, financial data on survey costs).

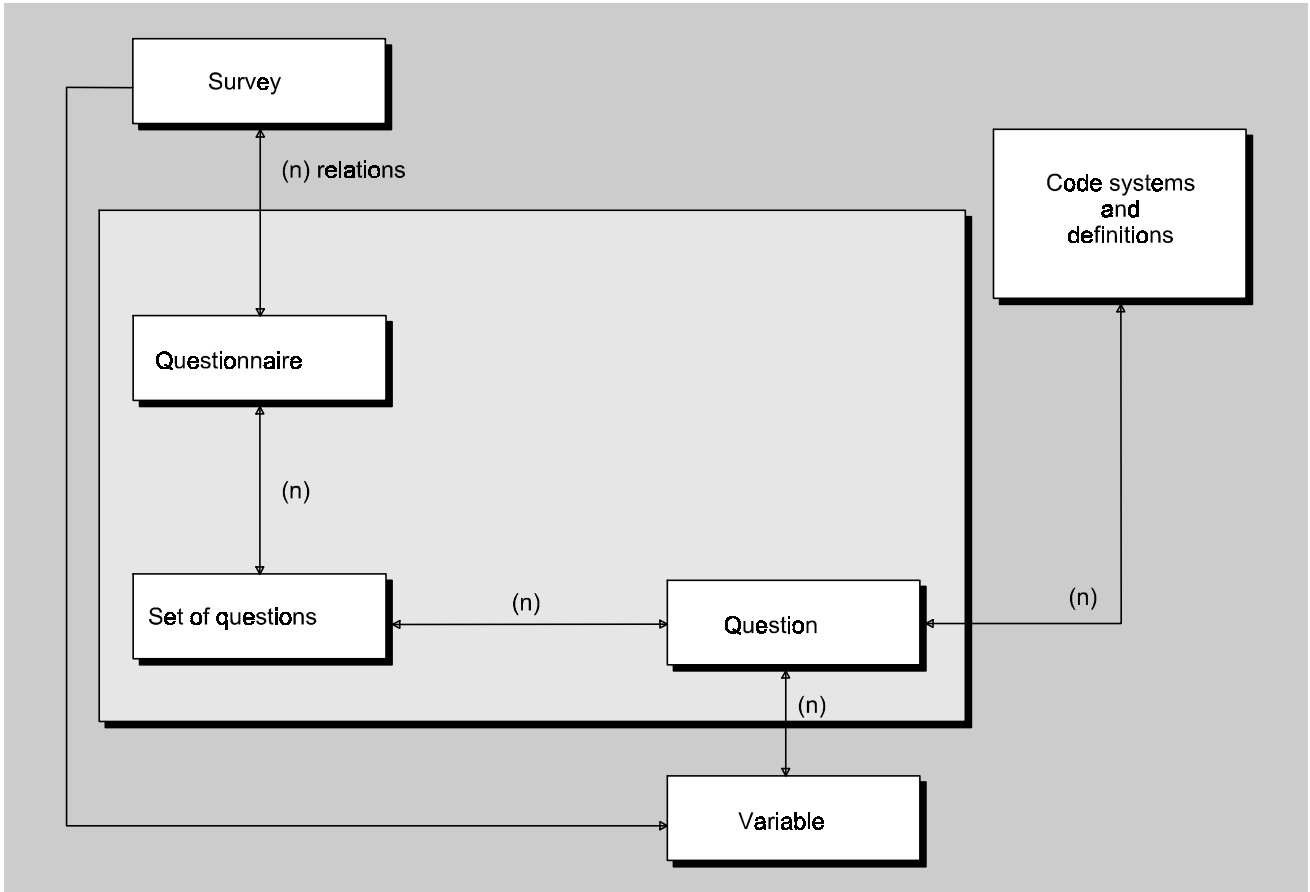
.

The following key concepts are important for the entities included:

- *Survey*: part of the statistical production process aimed at observing and gathering data on the research topics concerned from the survey units (enterprises and institutions, individuals and households).
- *Questionnaires*: the documents, i.e. the paper forms or computer files, which respondents complete and which register the primary information.
- *Set of questions:* a group or set of questions pertaining to a particular subject.
- *Question:* a linguistic formulation aimed at producing a response containing information by the respondent; or any part of a question to which a response is expected.
- *Variables*: information on a certain characteristic of the survey unit, gained directly or indirectly from responses/answers by the respondent.

The definitions of these notions are integrated with those developed for the EDI metasystem, which is being set up at the same time as the question bank, and serves as the electronic data interchange system with respondents. This system also includes a section for questions, variables and coding computer questionnaires.
There has been constant consultation during the development stages of the two system models, so that the registration of electronic and paper questionnaires and the variables, codes etc. can be realised within a single system

Prototype for question bank model



The entity SURVEY must include such things as the name of the survey, the department which carries it out, the codes for costing the survey etc.

The entity QUESTIONNAIRE must include the identification tag, the form (electronic or on paper) the domain of observation (SIC and size classes), intervals, and the time it takes to fill out the questionnaire.

The entity SET OF QUESTIONS contains the topics to which the questions refer.

The entity QUESTION supplies the text of the question, the relationship with the variable under consideration and the text used in the explanatory remarks, if any.

The entity VARIABLE contains the name and definition of the variable, the domain and the relationship with the questions used.

Generating outlines and selections on the basis of a given characteristics requires a systematic arrangement of the questions and variables in question. For instance, all questions and variables pertaining to turnover should be arranged systematically. The CODE SYSTEMS and DEFINITIONS make such arrangements possible. This entity also contains the codes and descriptions of major standard classifications such as the SIC, or size classes, in order to allow access to the data and the creation of outlines according to these lines.

The links between the various entities serve to show interrelated phenomena.

Although adding time aspects should be possible in principle, the current model prototype cannot yet incorporate historical features

5. State of affairs

The question bank model described above was realised in the form of a prototype under the EBS metabase and as such works under the software developed for EBS. In the test phase, at the end of 1996, we studied how the prototype functioned, as well as the possibilities for the filling of entities with the meta-information from questionnaires. We used the questionnaires of a number of key output surveys, including the manufacturing output survey. The development stages for EBS as well as the question bank were completed early in 1997.
Partly as a result of the tests we think that the actual implementation of the question bank system will require a major investment. We agreed to start building a simple set of texts of all the questions in the enterprise surveys, and incorporate this in the question bank at a later stage. The implementation and further development of the question bank will be a step by step process, and will have to take the following points into account:

Classification of questions and variables
The breakdown into questions, variables etc. requires a new classification system. The concept ‘turnover’, for instance, occurs in many different forms, such as net, gross, including VAT, turnover of certain products or groups. The development of such a classification system, as well as the implementation of the database take up a lot of capacity, both in terms of the development and of actually filling it.

Extending the attributes and applications
At the moment we can only accommodate a limited number of attributes per entity, as we described above.
It is possible to expand these with, for example, information on survey techniques. It is also possible to expand the number of possible applications, for instance by making use of the system to design new questionnaires.

Links with administrative and statistical units
As we built the question bank model we felt that it should be possible to link questionnaires at the individual unit level: respondent, observation unit, legal unit, business unit etc. We have not yet established such links between the question bank and the unit registers.

Question bank domain
The domain covered by the question bank has to be clearly defined. Should it be limited to the production statistics or should it cover individual and household surveys as well? The intended first step is to include all business surveys. At first filling the database will require a lot of effort, because all the questions will have to be entered. We have not yet decided on the best methods for this. Efficiency and costs are the main considerations. The methods under consideration are importing data from existing files, or scanning questionnaires. We have studied the possibility of using print files to scan forms or questions, but various technical problems still have to be solved in this respect.

Time aspects
Although we agreed in principle that it should be possible to incorporate time (historical) aspects in the question bank, we have not yet incorporated such a possibility in the prototype.

Organisation and management of the question bank
It is very important to take organisation and management aspects into consideration in the set up stage: should new questions and questionnaires be imported through centralised and co-ordinated procedures, how can the question bank be used in the design of new questionnaires, and the authorisation procedure for its use. Needless to say, there is a permanent need for computer capacity as well as technical support for setting up and maintaining the question bank.

For further inormation, please contact Jack Claessen at *jcsn@cbs.nl.*

Notes

1) With thanks to Hein Gielgens and Jean Ritzen for their contribution to the prototype of the question bank.

Comparable data from Fertility and Family Surveys

Jan J. Latten

In the framework of the UN research programme on changing fertility and family patterns in the ECE region, Statistics Netherlands has just issued its Standard Country Report. This article describes the history of fertility surveys in the Netherlands, how the survey set-up was redesigned in 1993 to fit in with UN requirements, and the results of the 1993 survey in terms of changing family formation behaviour. A brief preview of the 1998 Fertility Survey is also given.

1. Introduction

If findings from national surveys on fertility and family could be effectively compared, this would greatly enhance the understanding of national developments. In the late 1980s this realisation led the Population Activities Unit (PAU) of the United Nations Economic Commission for Europe (UN/ECE) to initiate a long-term sample survey research programme.
This initiative, a programme to focus on fertility and family change, stimulated standardisation in the various European national fertility surveys. National population centres and national statistical offices participate in the programme, and in doing so they benefit from the expertise of existing surveys such as the Fertility Survey of Statistics Netherlands, with twenty years of experience.

The objectives of the programme are to:
- prepare Standard Country Reports;
- conduct comparable Fertility and Family Surveys in about twenty ECE member countries;
- create Standard Recode Files for these surveys and archive these at the Population Activities Unit;
- carry out a programme of cross-country comparative studies.

The Fertility and Family Survey which Statistics Netherlands held in 1993 was one of the coordinated surveys, and recently the Dutch Standard Country Report was published. Up to now other priorities have prevented Statistics Netherlands from participating in the programme of cross-country studies.

2. The origins of the Dutch survey

In the Netherlands, population statistics used to be based on municipal population registers only. The demographic developments in the 1960s and 1970s made the Dutch government increasingly aware of the effect of demographic factors on societal developments, and it installed a special Commission on Population in 1976.
This Commission pleaded for more micro-level research of demographic behaviour to underpin socio-demographic studies which correlate shifts in demographic patterns to social developments. Surveys were to be set up to provide more elaborate data than population statistics based on municipal registers alone.

These new research directives also became manifest in subsequent working programmes of Statistics Netherlands. The first Dutch fertility survey – designed to bring about working hypotheses for the Population Forecasts – was conducted in 1974. Since 1977, fertility surveys have been held every five years to monitor changing demographic patterns.

3. The 1993 redesign

In 1993 the participation of Statistics Netherlands in the UN Population Activities Unit resulted in an extension of the Survey in comparison with the previous editions of 1977, 1982 and 1988. The design was broadened to cover not only the basic data requirements of the Population Forecasts, but also to provide data supporting the construction of quantitative and qualitative hypotheses to be used in the new Household Forecasts. Another aim in redesigning the survey was to collect in-depth data on the formation and dissolution of relationships and families, supplementary to the population data originating from municipal registers.

Compared with earlier surveys the 1993 survey was changed in a number of ways. First of all, the *age brackets* were broadened.
- In earlier surveys the upper limit was 37 years. Due to prevailing fertility delays, it was decided to extend the target population to include women up to the age of 42 years. This would improve the accuracy of the data on births and involuntary childlessness.
- Also in order to obtain more information on new types of households, the same age brackets could prove relevant.
- Lastly, extending the target population would reveal comparative information on the 1950-1954 birth cohorts, which were also included in the previous survey. For these three reasons 42 years was defined as the upper limit. The age of 18 years was kept as the lower age limit.

A second more spectacular aspect of the redesign was the inclusion of men in order to provide information about household positions of men as well. For the first time, this would enable an extensive presentation of demographic phenomena for men as well as women. The target population of the 1993 survey was therefore expanded to include everyone born in the period 1950-1974. These alterations followed the PAU recommendations on fertility and family surveys.

4. Results

In the Netherlands, as in many countries in Europe, the decline in fertility is striking. Dutch women born around 1935, who had their first child around 1960 had an average 2.5 children. The generation born only one decade later had an average 2 children and younger generations are giving birth to even fewer, with an average now well below replacement level. The total period fertility rate fell from well over 3 in 1950 to 1.5 in 1985. In recent years the total fertility rate has been just under 1.6.

A recent, though modest, increase in the total fertility rate may be explained by a catching-up effect, related to delayed motherhood. In the early 1990s, with an average age at first birth of over 28, Dutch women are almost the oldest first-time mothers in the world.

Younger generations differ significantly from the generation of 1945 on levels of childlessness. One in five women born in the 1970s will remain childless, compared with one in ten of those born in 1945.

First marriage rates dropped sharply in the Netherlands in the 1970s and early 1980s. A new phenomenon of paramount importance since the 1970s in family formation is the diffusion of unmarried cohabitation, as a result of which marriage is

postponed. However, most cohabitors marry after some time. The rapid increase in cohabiting – a living arrangement which is more prone to dissolution than marriage – goes hand in hand with an increase in the number of dissolutions of cohabiting couples. Together with the number of divorces, which has been quite stable in recent years, it increases the risk of becoming single again at some point in the life course, at young as well as middle ages. However, the proportion of people wanting to remain single permanently is supposed to have changed relatively little.

Traditionally family building was characterised by the transition from the parental home to marriage. Unmarried cohabitation was not accepted, just as unmarried motherhood. Setting up a family was the natural course of affairs and motherhood was the most important life fulfilment for women. Since then priorities have changed and indeed are still changing. The post-war baby boom generation grew up in a period of industrialisation and growing welfare. Their family building phase coincided with the introduction of new contraceptives, a period of emancipation and prosperity. The traditional pattern was eroded. Women and men could regulate if and when they had children, and even how many. Women invested increasingly in education and oriented themselves towards a career on the labour market.

Some members of these generations – the trend-setters – actively initiated new behaviour. Among other things cohabiting showed a continuous diffusion with each successive cohort. For example, more than half the women aged 35-39 years in 1993 followed the traditional path before they were 25 years old, compared with not even one quarter of women born one decade later (25-29 years old in 1993). The majority has left the traditional path. For the men the same tendency occurs.

Among younger generations cohabiting as a living arrangement is now established behaviour, but mostly seen as a temporary life course phase, demonstrated by the fact that a modest majority of Dutch in their teens and twenties, men as well as women, support the statement that marriage makes a relationship complete. Around one in three disagree. On the other hand about a quarter underline that marriage is merely a formal contract.

Nevertheless, women's mean age at first marriage rose from 22.6 in 1975 to 27.2 years in 1994. The marriage experience is partly compensated by cohabiting experience, though not completely. For example, 86 per cent of 40-42 year-olds had a partnership before the age of 25. For the 25-29 year-olds in 1993 this is true for only 76 per cent. Some of them apparently choose other life-courses during their twenties, such as living alone or staying with their parents. The shift from the traditional to a modern pattern comprises more than just replacement of marriage by unmarried cohabitation. Pluriformity is another symptom.

In addition to the rise in unmarried cohabitation, increasing numbers of children are now born outside marriage. In 1975 2 per cent of children were born out of wedlock, increasing to 4 per cent in 1980 and 13 per cent in 1993, when for first born children the percentage was 17. One third of the first born children were later legitimised by marriage of their parents.

One important factor in this development is the level of education. Each younger cohort of women shows higher percentages participating in education, although always lower than men. Together with the prolonged education period of each younger cohort we see a delay of first entry on the labour market. It is only around the age of 25 that the level of first entry on the labour market for the younger cohorts reaches the levels of the older cohorts.

The growing participation of women in higher education and professional activity confronts them with compatibility problems between a paid job and childcare. In the Netherlands many women seek to solve this problem by working part-time, indeed compared with women working full-time, more part-timers have children.

The level of modernity of family formation is not equal for all groups in Dutch society. Education is an important differentiating background characteristic. Postponement of starting a family, that is the moment the first child is born, for example, is a component of the new pattern that so far seems to be predominantly a phenomenon among the higher educated. People with lower education levels are lagging behind in this development. Nevertheless the trend is visible for them too.

In addition, if we look at the increasing tendency to combine family obligations and labour force participation we see the same pattern. The trend-setters are the higher educated women. The lower the education, the less the chance of combining work and children.

5. The 1998 Survey

The sixth Fertility Survey will be held in 1998. The preparations have been done without a new coordinating initiative from the PAU, but as a result of the experiences with the coordinated 1993 survey, the 1998 survey again will include men. Moreover the age bracket has even been extended to include those up to 52 years old. This extension was made possible partly by financial support from the Netherlands Council for Family Matters (NGR). The first results are expected by the end of 1998.

Fertility and family surveys in countries of the ECE region. Standard country report, the Netherlands. United Nations Economic Commission for Europe in co-operation with Department of population, Statistics Netherlands, New York and Geneva 1997.

Tables from this publication will available on the Internet: http://www.unece.org/deap/pau/f_ home1.htm.

The report is the third issue in the series. Similar reports have been published for Sweden and Norway. Forthcoming editions will focus on Poland, France, Latvia and Finland. For information on other countries, contact Mark Bloch, United Nations Economic Commission for Europe, Palais des Nations, CH-1211 Geneva 10, tel. +41 22 917 3320.

For further information, please contact Jan Latten at *jltn@cbs.nl.*