

## Issues in measuring innovation

GERBEN VAN DER PANNE

*Delft University of Technology, Faculty of Technology, Policy and Management,  
Economics of Innovation Department, Delft (The Netherlands)*

Innovation research builds on the analysis of micro level data describing innovative behaviour of individual firms. One increasingly popular type of data are Literature-based Innovation Output (LBIO) data. These are compiled by screening specialist trade journals for new-product announcements. Notwithstanding the substantial advantages, the eligibility of LBIO data for innovation research remains controversial. In this paper the merits of LBIO data are examined by means of comparative analysis. A newly built LBIO database is systematically compared with the widely used Community Innovation Survey. It shows that both databases identify similar innovators in terms of firm size, distribution across industries and degree of innovativeness: LBIO data can be considered a fully fledged alternative to traditional innovation data, highly eligible for innovation research.

### Introduction

As early as in 1962, KUZNETS observed that the greatest obstacle in understanding the economic role of technological change is the scholars' inability to adequately measure innovation. In this paper the increasingly popular Literature-based Innovation Output (LBIO) data are tested on their eligibility for innovation research. LBIO data are compiled by screening specialist trade journals for new-product announcements.

---

Received July 17, 2006

*Address for correspondence:*

GERBEN VAN DER PANNE

Delft University of Technology, Faculty of Technology, Policy and Management,  
Department Innovation Systems, P. O. Box 5015, 2600 GA Delft, The Netherlands

E-mail: g.vanderpanne@tudelft.nl

0138–9130/US \$ 20.00

Copyright © 2007 *Akadémiai Kiadó, Budapest*

All rights reserved

The Futures Group (EDWARDS & GORDON, 1984) was the first to compile LBIO data. These have been analyzed by Acs & Audretsch in a series of articles (for a survey see ACS & AUDRETSCH, 1993). In Europe similar databases have been compiled in the Netherlands (KLEINKNECHT et al., 1993), in Ireland (COGAN, 1993), Austria (FLEISSNER et al., 1993), the United Kingdom (COOMBS et al., 1996), Italy (SANTARELLI & PIERGIOVANNI, 1996), Spain (FLOR & OLTRA, 2004) and Finland (SAARINEN, 2005).

LBIO data stand out from traditional innovation data on both theoretical and practical arguments, for reasons to be discussed below. Nonetheless, the eligibility of LBIO data remains controversial. First, LBIO data are based on the market introduction of new products rather than traditional Research and Development (R&D) investments. Second, one cannot apply standard sampling procedures in compiling LBIO databases. This may jeopardize the representativeness of LBIO samples.

In this paper the eligibility of LBIO data is examined. A newly built Dutch LBIO database is systematically compared with the Dutch 1996 Community Innovation Survey (CIS). The latter can be considered representative for total firm population and is taken as frame of reference. If and only if both databases identify similar firms in terms of firm size, industry distribution and degree of innovativeness, one may consider LBIO data a fully fledged alternative to traditional innovation data. If not, future innovation research must take account of the limitations in the practical use of LBIO data.

### **LBIO versus traditional innovation data \***

Innovation research traditionally builds on the R&D indicator, measuring R&D expenses relative to total sales. R&D figures are among the most widely used indicators of innovation, for two reasons. Most importantly, data on firm level R&D figures have been collected continuously for decades. Second, this indicator is standardized, well-defined and allows for inter-firm, inter-industry and international comparisons. Its sophisticated definition distinguishes between process and product innovation on the one hand, and between basic and applied research efforts on the other.

Notwithstanding their subtlety, R&D figures are not comprehensive and miss out on innovation-related investments in design, trial, market testing and fixed assets: It has been estimated that only 25 percent of innovation expenditures relate directly to R&D activities (BROUWER & KLEINKNECHT, 1997). R&D figures rely on accurate accounting practices. Particularly small firms tend to under-report their small-scale and informal R&D activities. Larger firms, by contrast, organize their R&D activities more formally and are more fruitful for accessing R&D data. Systematic bias towards manufacturing industries exists also; R&D figures tend to underestimate innovation in services

---

\* This section is largely based on COOMBS et al. (1996) and KLEINKNECHT et al. (2002).

(KLEINKNECHT et al., 2002). In addition to this limited comprehensiveness, the R&D indicator implicitly presumes that new products are the necessary outcome of innovation activities. Yet R&D inputs are not sufficient so to guarantee that the innovation efforts are taken up to the market introduction of new products.

L BIO data are spared these drawbacks. L BIO data document the market introduction of new products. Only those firms successfully leading their innovation efforts up to new products are identified. In this, L BIO data fully comply with the Oslo Manual definition of the innovating firm: 'A technological product and process innovating firm is one that has *implemented* technologically new or significantly improved products or processes or combinations of products and processes during the period under review' (OECD, 1996: 42).

In the absence of the necessity to survey individual firms, L BIO data are cost efficient. Particularly in the cases of small firm innovation this advantage is capitalized upon. Whereas surveys usually apply minimum firm size restrictions for reasons of efficiency, L BIO data do not discriminate against the large population of small firms.

Without having to burden the firm, the L BIO method allows for analysing the firms' innovation track record by screening preceding or successive volumes of journals, thereby avoiding problems of secrecy and non response. Instead one takes advantage of the firms' incentives to announce their product in publicly available journals: L BIO data can be considered most comprehensive among those using secondary databases (FLOR & OLTRA, 2004).

Unlike R&D figures, L BIO data associate the innovating firm with its innovative output, i.e., its innovations. The product information provided in the announcement allows for classification by type of innovation, degree of technical complexity or newness. One thereby takes advantage of the journal editors' expert opinion, preventing mere product differentiations from being portrayed as new or improved products.

Notwithstanding these advantages, L BIO data remain subject to criticism. One may argue that the counts of new-product announcements increase with numbers of journals screened. Yet the journal selection procedure cannot be standardized and depends on the availability of journals. Particularly small-sized industries may not sustain an appropriate journal and remain under-represented. Moreover, editorial policies regarding new-product announcements may change over time, jeopardizing the consistency of data collection.

Propensities to announce the innovation in a journal need not be equal across industries. Industries that serve narrowly defined and small-sized markets may not be inclined to issue an announcement but rather use more direct communication channels. This has been shown, however, not to severely affect the reliability of L BIO data (KLEINKNECHT et al., 1993).

COOMBS et al. (1996) argue that large firms may be less inclined to announce their products in journals. SANTARELLI & PIERGIOVANNI (1996) add that large firms tend

towards international trade, further reducing the need to announce in domestic journals. Consistent with analyses of American LBIO data (see ACS & AUDRETSCH, 1993) their Italian LBIO data suggest that the LBIO approach sets small firms at the relative advantage. By contrast, Finnish LBIO data do not show any substantive bias regarding firm size (SAARINEN, 2005).

Propensities to announce also differ for product and process innovations. LBIO databases may be less comprehensive in covering the latter since firms have no incentive to announce in-house process innovations.

### **Compiling the LBIO database**

Although the LBIO method of data collection is straightforward, one incurs two issues in compiling LBIO data. The first relates to the journal selection procedure, acting on sample probabilities of individual firms. The second issue concerns the products' origin; counting foreign products compromises the accuracy of LBIO data in measuring domestic innovation efforts.

#### *The journal selection procedure*

Although journal selection procedures cannot be standardized, some general guidelines can be applied. In compiling our database for the Netherlands, all regularly available trade journals were listed such that all major industries are covered. Since the LBIO method excludes advertisements journals are selected on the editorial sections 'new products' and judged on the information provided in the announcements. The announcements were required to reveal the announcing firm as well as a typical feature that distinguishes the new product from preceding versions or substitutes in terms of functionality, versatility or efficiency: cases of mere product differentiation were omitted. These criteria reduced the number of appropriate journals to 43 (see Appendix 1).

During August 2000–September 2002, two volumes were screened for new-product announcements. Every single new-product-announcing firm was sent a questionnaire to document detailed information on the firm, its innovation efforts and the announced product. In total 1585 firms were surveyed, of which 1056 firms or 66.6 percent responded.

#### *Accuracy of LBIO data: spurious counts*

One major advantage of LBIO data over traditional innovation data is cost efficiency. In using publicly available product announcements, innovators are identified

without having to survey individual firms. One implicitly assumes that products announced by domestic firms are of domestic origin. This proposition may, however, not hold. Domestic firms and, particularly, wholesale firms may simply serve as a distribution channel for foreign products. LBIO data may therefore tend to over-estimate domestic innovation. In compiling this particular LBIO database, all announcing firms were surveyed and explicitly asked whether the announced product is developed in-house, within the Netherlands. This information allows for passing judgement on the accuracy of LBIO data in measuring domestic innovation and the necessity to verify the products' origin.

Analysis of our LBIO data confirms this inaccuracy. Drawing a comparison between counts of domestic and foreign innovations, it shows that, in default of origin verification, LBIO data severely over-estimate innovative activity (see Appendix 2, Table 3). Out of 1056 firms, 658 or 62.3 percent report to have imported the announced innovation rather than developing it in-house within the Netherlands. This share of foreign products varies across industries and ranges from zero to 100 percent. Considering only those industries with at least 5 observations in total, transport equipment (SIC 35), research and development (SIC 73) and sewage and refuse disposal (SIC 90) rank lowest. For these industries, LBIO data measure domestic innovation accurately.

By contrast, paper (SIC 21), motor vehicles (SIC 50), wholesale (SIC 51) and financial intermediation (SIC 65) are among those industries dominated by foreign product announcements: in default of origin verification, LBIO data over-estimate innovation in these industries.

This inaccuracy of LBIO data in measuring innovativeness in particular industries may be addressed only if the share of imported innovations were related to trade patterns across industries. This would allow for ex-post correction for spurious innovation counts, thereby circumventing the need for origin verification and saving the cost efficiency argument of LBIO data. Further analysis shows, however, that numbers of imported innovations vary across industries randomly (see Appendix 2, Table 3). This result cancels the cost efficiency argument and implies that origin verification remains indispensable in compiling LBIO databases.

### **Eligibility of LBIO data**

LBIO data differ from traditional innovation data in that these rely on different innovation indicators: LBIO data are based on innovation counts, whereas traditional data build on R&D figures. Moreover, LBIO data involve screening trade journals rather than surveying a stratified sample of firms. To assess the eligibility of LBIO data for innovation research, both the underlying indicator and method of data collection are examined.

Table 1. A comparison between LBIO and CIS data

	Indicator	Method of data collection
LBIO	innovation counts	screening trade journals
CIS	R&D figures	survey stratified sample

*The indicator underlying LBIO data: innovation counts*

To examine the merits of innovation counts underlying LBIO data, innovation counts are compared with the alternative indicator, i.e., R&D figures.

The CIS database documents firm level information on both these indicators and allows for within-sample comparisons, thereby holding the method of data collection constant. \* Any systematic difference between both indicators jeopardizes the unbiasedness of innovation counts and, accordingly, compromises the eligibility of LBIO data.

R&D expenses are not sufficient so to guarantee the successful conclusion of innovation projects leading up to the market launch: Only 54 percent of all firms surveyed in the CIS that engage in R&D activities also reported to have introduced an innovation. Assuming that R&D expenses precede the launch of innovations, inconsistency between both indicators, i.e. R&D figures and innovation counts, can be analyzed by regressing the latter on the former.

The results are summarized in Table 2.

Table 2. Probability to introduce an innovation (CIS database)

R&D intensity	0.13 (0.02)**
R&D permanent <sup>†</sup>	0.75** (0.09)
Firm size dummy <sup>††</sup> :	
>50 employees	-0.08 (0.09)
Industry dummies <sup>†††</sup> :	
machinery	2.25* (1.17)
constant	-1.64 (1.17)
McFadden's R	0.14
N	3016

\*significant at 10% \*\*significant at 5%; standard errors in parentheses

<sup>†</sup> Nature of R&D activities. Dummy takes on value 1 if R&D investments are on a permanent basis, 0 otherwise.

<sup>††</sup> Base category: lower category (10-49 employees).

<sup>†††</sup> Base category: industry with share of firms reporting having introduced new products closest to overall mean (i.e., SIC 16 tobacco). All industries (2-digit SIC level) on which the CIS database documents information on both innovation counts and R&D intensity are included: insignificant industry dummies are omitted.

\* The CIS database documents information on both innovation counts and R&D figures. The innovation count indicator documented in the CIS survey relates to LBIO data in that it documents whether the surveyed firm has introduced an in-house developed product innovation on the market during the 1994–1996 period.

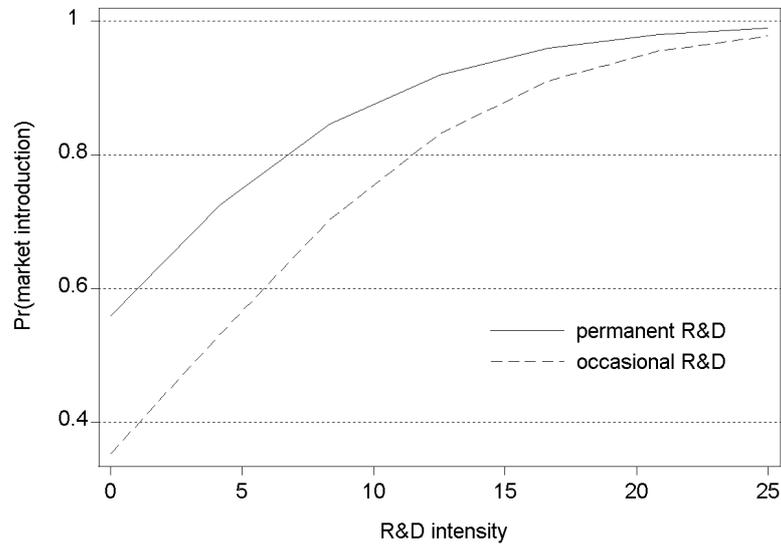


Figure 1. Probability to introduce an innovation (CIS database)

The model shows that the propensity to introduce an innovation on the market increases with R&D intensity and is significantly higher for firms engaging in R&D permanently.\* Setting the cut-off level at 0.5, the odds are that those firms spending less than 3 percent of total sales on occasional R&D activities will *not* introduce an innovation (see Figure 1); these innovators are missed out on by innovation counts and, accordingly, LBIO data. This implies that LBIO data cannot be considered eligible for measuring innovation among occasionally R&D-performing firms.

Most innovators, however, engage in R&D on a permanent basis. Setting the cut-off level at 0.5 again, the odds are that every firm engaging in R&D on a permanent basis will launch an innovation, irrespective of R&D intensity. For permanently innovating firms innovation counts and, accordingly, LBIO data can be considered unbiased towards R&D intensity.

As opposed to results established by ACS & AUDRETSCH (1993), COOMBS et al. (1996) and SANTARELLI and PIERGIOVANNI (1996) the analysis does not suggest that

\* The propensity to introduce an innovation on the market is 54 percent for those firms with R&D intensities larger than 0. This propensity differs significantly between firms engaging in R&D activities permanently and occasionally: 62 and 38 percent, respectively.

firm size influences the propensity to introduce an innovation. Holding all other variables constant, the innovation count indicator remains unbiased towards firm size, adding to the eligibility of LBIO data.

Regarding industry distributions, the propensity to introduce an innovation is significantly higher for firms in one industry, i.e. machinery (SIC 29). For this industry, the innovation count (R&D intensity) indicator over (under) estimates innovativeness. Saving this specific branch, innovation counts remain unbiased across industries.

*The method of data collection underlying LBIO data: screening trade journals*

The CIS database is compiled by means of a stratified survey among a sample of firms stratified by industry and firm size. The LBIO data rely on quite a different method of data collection and draw on specialist trade journals, lacking any stratification. To pass judgement on the use of journals, the innovation counts documented in the LBIO database are compared to those in the CIS database. Holding the indicator (i.e., innovation counts) constant, inconsistency between both databases is to be ascribed to differences in methods of data collection (i.e., screening trade journals versus surveying a stratified sample, see Table 1). If and only if the innovation counts documented in both databases identify similar firms, trade journals may be considered feasible for collecting data on innovation counts.

Comparing both databases on their respective innovation counts, it shows that the survey-based CIS database identifies 23 times as many firms as does the journal-based LBIO database:\* one may argue that trade journals are less effective in identifying innovative firms. This short fall in comprehensiveness does not, however, jeopardize the eligibility of trade journals, provided that the journals remain unbiased towards industries, firm size and degree of innovativeness.

As regards the journals' unbiasedness towards industries, comparative analysis of both databases shows that the use of journals does moderate the identification of innovative firms (see Appendix 3, Figure 2a). Journals miss out on several industries, predominantly natural resources (SIC 11–14), food and beverages (SIC 15) and primary metals (SIC 27). This may typically be due to a lack of appropriate journals. In addition, journals tend to under-estimate several industries, for example furniture textiles (SIC 17), wood (SIC 20) and furniture (SIC 36). Notwithstanding these biases, both the CIS and LBIO database run parallel in terms of industry distributions: Spearman's rank correlation coefficient is 0.7.

As regards firm size, COOMBS et al. (1996) and SANTARELLI & PIERGIOVANNI (1996) assert that journals are biased against large firms. This assertion is, however, not consistent with our comparative analysis: Large firms are far from being under-

---

\* 3517 versus 152 cases for the CIS and LBIO database, respectively.

represented in the L BIO database (see Appendix 3, Figure 2b). Journals do not discriminate against firm size whatsoever: both databases' firm size distributions run parallel (Spearman's rank correlation is 0.8).

Comparing both databases on degree of innovativeness of their respective firms, it shows that journals identify at least as innovative firms as do surveys among stratified samples of firms. L BIO firms spend 8.9 percent of their sales on R&D (median: 5 percent); CIS firms only 7 percent (median: 2.2 percent). The L BIO firms engage in R&D on a permanent basis more often than do CIS firms: 82.2 versus 70 percent. Moreover, shares in total sales generated with (re)new(ed) products are higher for L BIO firms: L BIO firms generate 27 percent of total sales with new products versus only 11 percent for CIS firms.

These comparisons of innovation counts between both databases suggest that the method of data collection underlying L BIO data, i.e. the use of trade journals, does not jeopardize the representativeness of L BIO data.

The above comparisons between the CIS and L BIO database on the underlying indicators and methods of data collection apply to the intersection of both databases' domain, i.e., non service firms larger than 9 employees.\* Service firms and firms smaller than 10 employees are missed out on by the CIS survey. These firms add up to 62 percent of the innovators identified in the L BIO database. Since these firms show as much concern for innovation,\*\* L BIO data prove effective in measuring innovation in service firms and small firms.

### Concluding remarks

In this paper the eligibility of the Literature-based Innovation Output (L BIO) data for innovation research is examined. L BIO data differ from traditional innovation data in that these draw on new-product introductions (rather than R&D figures) and the screening of trade journals (instead of surveys among a stratified sample of firms).

In circumventing the need to survey individual firms, L BIO data are allegedly cost efficient. Analysis of the announced products' origin shows, however, that over 60 percent of the new-product-announcing firms merely served to distribute the announced product and were not involved in developing the innovation. This causes L BIO data to

---

\* Similarity between both databases can be examined for sub-samples only, since the CIS survey applies restrictions in its data collection procedure. First, the survey applies a minimum size restriction and disregards all firms smaller than 10 employees for reasons of efficiency. Second, the CIS survey does not document information on product innovations in the cases of service firms. In order to arrive at accurate comparisons, both these restrictions are to be applied to the L BIO sample as well. The 3517 and 152 firms documented in the CIS and L BIO database, respectively, are at the intersection of both samples' domain.

\*\* Service and non-service firms report R&D intensities (median) of 10 and 2 percent, respectively. Firms smaller than 10 and larger than 9 employees report R&D intensities (median) of 20 and 5 percent, respectively.

severely over-estimate domestic innovation. Moreover, this share of spurious innovations varies across industries randomly and ranges from zero to 100 percent, further compromising the accuracy of LBIO data in measuring domestic innovation. This cancels the efficiency argument: in compiling LBIO data one cannot refrain from verifying the domesticity of every single new-product announcement.

Taking the Community Innovation Survey (CIS) as frame of reference, LBIO data are substantially biased against those innovators engaging in R&D activities only occasionally (as opposed to permanently). Within the group of permanently R&D performing firms, LBIO data are much more comprehensive. Since the vast majority of all innovators report engaging in R&D activities permanently, the bias against the small minority of less dedicated innovators does not jeopardize the eligibility of LBIO data.

Although some industries lack any trade journal in which to announce new products, the distribution of innovators across industries does not systematically differ between LBIO and CIS data. The same holds for distributions across firm size: Comparative analysis shows that LBIO data are unbiased towards firm size. As to innovation efforts, the firms identified in the LBIO database outperform those identified in the CIS database and are at least as innovative in terms of R&D expenses and new-product turnover.

In the absence of any severe biases, consciously compiled LBIO data can be considered a fully fledged alternative to traditional innovation data.

## References

- BROUWER, E., KLEINKNECHT, A. (1997), Measuring the unmeasurable: a country's non-R&D expenditure on product and service innovation, *Research Policy*, 25 : 1235–1242.
- COOMBS, R., NARANDREN, P., RICHARDS, A. (1996), A literature-based innovation output indicator, *Research Policy*, 25 : 403–413.
- FLOR, M. L., OLTRA, M. J. (2004), Identification of innovating firms through technological innovation indicators: an application to the Spanish ceramic tile industry, *Research Policy*, 33 : 323–336.
- KLEINKNECHT, A., MONTFORT, K. VAN, BROUWER, E. (2002), The non-trivial choice between innovation indicators, *Economics of Innovation and New Technology*, 11 : 109–121.
- KLEINKNECHT, A., REIJNEN, J. O. N., SMITS, W. (1993), Collecting literature-based innovation output indicators: the experience in the Netherlands. In: A. KLEINKNECHT, D. BAIN (Eds), *New Concepts in Innovation Output Measurement*. St Martin's Press, pp. 42–84.
- KUZNETS, S. (1962), Inventive activity: Problems of definitions and measurements. In: R. R. NELSON (Ed.), *The Rate and Direction of Inventive Activity*. Princeton University Press, pp. 19–43.
- ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (1996), *Oslo Manual. The measurement of scientific and technological activities*, Eurostat, Paris.
- SAARINEN, J. (2005), *Innovations and Industrial Performance in Finland 1945–1998*. Almqvist & Wiksell International, Stockholm.
- SANTARELLI, E., PIERGIOVANNI, R. (1996), Analyzing literature-based innovation output indicators: The Italian experience, *Research Policy*, 25 : 689–711.

## Appendix 1

### Journals screened for new-product announcements

<i>Aandrijftechniek</i> <i>Aluminium</i> <i>Auto- en Motortechneek</i> <i>Automatie</i> <i>Beveiliging</i> <i>Bouwwereld</i> <i>Chemie</i> <i>Constructeur</i> <i>Electronica</i> <i>Energietechniek</i> <i>Estheticienne</i> <i>Gemeentereiniging en AfvalManagement</i> <i>Glas in Beeld</i> <i>H<sub>2</sub>O</i> <i>Horeca Entree</i> <i>Innovisie</i> <i>Installatie + Sanitair Magazine</i> <i>Keuken- en InterieurMagazine</i> <i>Kunststof en Rubber</i> <i>Laboratorium Magazine</i> <i>Lastechniek</i> <i>Materialen</i>	<i>Metaal en Kunststof</i> <i>Metaal en Techniek</i> <i>Meubel</i> <i>MilieuMagazine</i> <i>Ned. Tijdsch. Voor Tandheelkunde</i> <i>Pakblad</i> <i>Parketblad</i> <i>PetroChem</i> <i>Product</i> <i>PT Industrie</i> <i>Schip en Werf de Zee</i> <i>Technisch Weekblad</i> <i>Technische Revue</i> <i>TextielVisie</i> <i>Transport + Opslag</i> <i>Verfkroniek</i> <i>Verpakken</i> <i>VerpakkingsManagement</i> <i>Verwarming en Ventilatie</i> <i>Vlees &amp; Snack</i> <i>Wegen</i>
---	--

## Appendix 2

Table 3. Counts of new-product announcements across industries (LBIO database)

SIC <sup>†</sup> Industry	Total		Domestic		Foreign		SPIN <sup>††</sup>	SPM <sup>†††</sup>
	Cases	Share	Cases	Share	Cases	Share		
17 textiles	3	0.3	1	0.3	2	0.3	1.1	0.8
18 wearing apparel	1	0.1	1	0.3	0	0.0	3.3	0.6
20 wood and wood products	3	0.3	2	0.5	1	0.2	2.4	0.4
21 pulp and paper	9	0.9	2	0.5	7	1.1	0.6	0.8
22 publishing and printing	3	0.3	2	0.5	1	0.2	2.4	0.9
24 chemicals	24	2.3	13	3.3	11	1.7	1.9	1.4
25 rubber and plastic products	25	2.4	17	4.3	8	1.2	3.3	1.0
26 non-metallic products	14	1.3	7	1.8	7	1.1	1.6	0.8
27 basic metals	2	0.2	0	0.0	2	0.3	0.5	0.9
28 fabricated metal products	34	3.2	19	4.8	15	2.3	2.0	1.0
29 machinery and equipment	132	12.6	75	18.8	57	8.8	2.1	1.0
30 office machinery and computers	4	0.4	2	0.5	2	0.3	1.6	1.0

(continued on next page)

Table 3. (cont.)

SIC <sup>†</sup> Industry	Total		Domestic		Foreign		SPIN <sup>††</sup>	SPM <sup>†††</sup>
	Cases	Share	Cases	Share	Cases	Share		
31 electrical machinery	16	1.5	9	2.3	7	1.1	2.0	0.9
32 communication equipment	8	0.8	5	1.3	3	0.5	2.4	1.0
33 med, prec. instruments	32	3.1	16	4.0	16	2.5	1.6	1.3
34 motor vehicles	3	0.3	3	0.8	0	0.0	6.5	0.7
35 other transport equipment	5	0.5	5	1.3	0	0.0	9.8	1.0
36 furniture	6	0.6	3	0.8	3	0.5	1.6	0.7
37 recycling	1	0.1	1	0.3	0	0.0	3.3	
45 construction	23	2.2	12	3.0	11	1.7	1.8	4.5
50 trade, maintenance vehicles	16	1.5	2	0.5	14	2.2	0.3	
51 wholesale and commission trade	497	47.5	99	24.9	398	61.3	0.4	
52 retail trade	3	0.3	1	0.3	2	0.3	1.1	
55 hotels	1	0.1	0	0.0	1	0.2	0.8	
63 transport	2	0.2	0	0.0	2	0.3	0.5	
65 financial intermediation	21	2.0	0	0.0	21	3.2	0.1	
70 real estate activities	4	0.4	3	0.8	1	0.2	3.3	
71 renting of machinery and equipment	1	0.1	0	0.0	1	0.2	0.8	
72 computer and related activities	52	5.0	31	7.8	21	3.2	2.4	
73 research and development	9	0.9	8	2.0	1	0.2	7.3	
74 other business activities	86	8.2	53	13.3	33	5.1	2.6	
80 education	1	0.1	1	0.3	0	0.0	3.3	
85 health	1	0.1	0	0.0	1	0.2	0.8	
90 sewage and refuse disposal	5	0.5	5	1.3	0	0.0	9.8	
	1056	100	398	100	658	100		

† Standard Industrial Classification

†† The Strength in Product Innovation (SPIN) index is used to distinguish industries dominated by foreign innovations from those with predominantly domestic product announcements. It relates the share of domestically developed innovations in total innovation counts per industry to the national aggregate. The index reads as

$$SPIN_i = \frac{Inn_{di}/Inn_{fi}}{Inn_d/Inn_f}$$

where  $Inn_d(i)$  = counts of domestic innovations (industry i)

$Inn_f(i)$  = counts of foreign innovations (industry i)

Values larger than 1 indicate industries with relatively few foreign product counts.

††† To test trade data as a proxy for shares of foreign announcements, the Strength in Product Markets (SPM) index is constructed as the SPIN counterpart and reads as

$$SPM_i = \frac{X_i/M_i}{X/M}$$

where  $X(i)$  = shipping value of exports (industry i)

$M(i)$  = shipping value of imports (industry i)

The SPM index relates export/import ratios at the industry level to the national aggregate: values smaller than 1 indicate less competitive industries. Comparative analysis does not suggest that the SPM and SPIN indices run parallel (insignificant Spearman's correlation coefficient). Consequently, trade data cannot be used to resolve the bias of LBIO data towards those industries with low SPIN values: origin verification remains indispensable in compiling LBIO databases.

Appendix 3

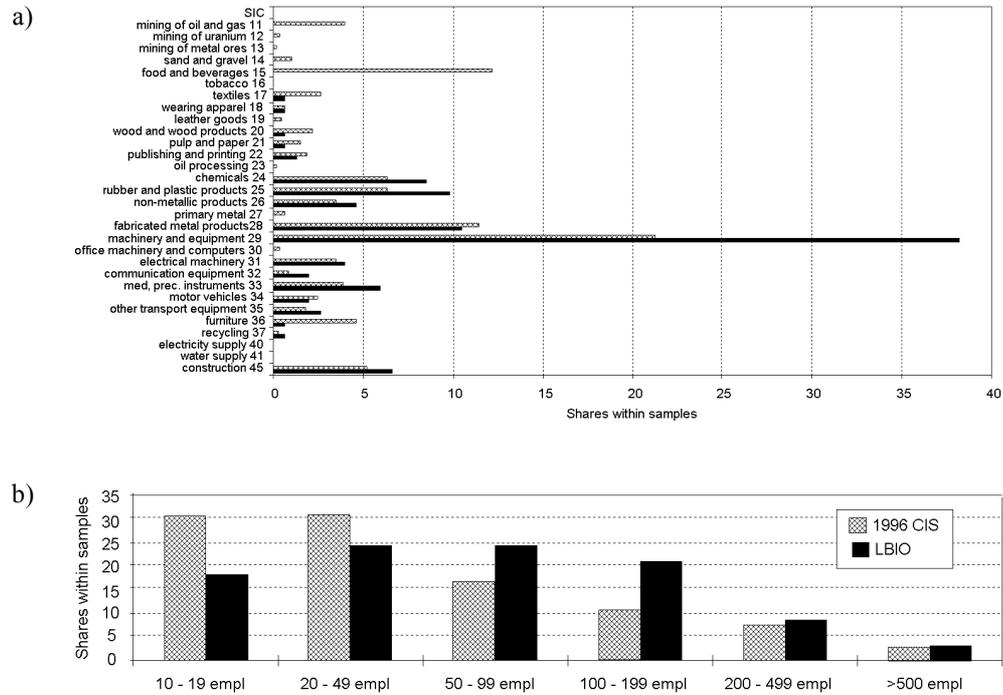


Figure 2. Innovation counts across a) industries and b) firm size: CIS versus LBIO database\*

\* Based on the innovation count indicator, which is documented in both the LBIO and CIS database. With the indicator held constant, differences between both databases in distributions of innovation counts across industries or firm size are to be ascribed to differences in methods of data collection, i.e., screening trade journals versus surveying stratified samples of firms.