REVIEW AND ASSESSMENT OF THE OTAF CONCORDANCE BETWEEN THE U.S. PATENT CLASSIFICATION AND THE STANDARD INDUSTRIAL CLASSIFICATION SYSTEMS: FINAL REPORT

Prepared by the Office of Technology Assessment and Forecast, Patent and Trademark Office, U.S. Department of Commerce

January 1985

This material is based upon work supported by the Science Indicators Unit of the National Science Foundation under Interagency Agreement No. SRS-8315600. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

EXECUTIVE SUMMARY

In the spring of 1983 the Office of Technology Assessment and Forecast (OTAF) of the Patent and Trademark Office (PTO), in cooperation with the Science Indicators Unit of the National Science Foundation (NSF), began a review and assessment of the Concordance between the U.S. Patent Classification (USPC) system and the Standard Industrial Classification (SIC) system. The project consisted of an internal OTAF study followed by a Workshop of Concordance users convened to comment on the study results and make recommendations concerning the Concordance.

The Concordance is a means of relating subclasses of the USPC to the SIC. In 1981 the Concordance was used to relate about 1.4 million patents in about 100,000 patent subclasses to one or more of 55 "product fields," which are groupings of SIC categories drawn primarily from the manufacturing section of the SIC. The purpose of the Concordance was originally to provide an indicator of inventive activity in U.S. industry for the Science Indicators report series; its development and subsequent updating have been funded by the NSF. The Concordance has also been of interest to researchers desiring to relate patent activity to R&D and economic data at the industry level.

The Concordance was developed in 1974 by manually assigning patent subclasses to the SIC product fields. In the period from 1974 to 1983 the Concordance underwent biennial updates, which were necessitated by the continual reclassification of the USPC, but was not seriously scrutinized for error or analytical utility.

Problems inherent in using patent data to measure industrial inventive activity have been long recognized. Not all inventions are patented; patent counts do not distinguish between inventions of different value; and patent counts are affected by extraneous factors such as cost and administrative procedures, in addition to the level of inventive activity. Despite these shortcomings, the long time series available, the broad coverage of technologies, and the increased accessibility of patent data through automated data services have made patents an increasingly attractive source of data for analyses of issues related to industrial innovation and international competitiveness.

A problem specific to the use of patents to measure inventive activity by industry is the organization of the data. The USPC is designed to help patent examiners search the prior art; therefore it is based upon the technological features of inventions that make them patentable. The basis for classification is largely functional and therefore the classes and subclasses do not always correspond closely to industries or technologies as commonly defined. On the one hand, for example, windmills and egg beaters are both found in USPC class 416, "Impellers." On the other hand, subclasses, and even individual patents, in USPC class 424, "Drugs, Bioaffecting and Body Treating Compositions," include both pharmaceuticals and pesticides, so that they cannot be separated from one another.

Because the technology contained in a patent subclass does not always relate one-to-one with an SIC product field, the practice was adopted in the Concordance of assigning subclasses to all the SIC product fields to which they pertained. The rationale was that the invention would thus be counted in all the product fields to which it applied. Most patent subclasses are assigned to only one product field. In the 1981 Concordance, only 28% of the subclasses were assigned to more than one product field. Nonetheless, the 1981 Concordance resulted in an overall number of patent counts that was 41% higher than the actual number of patents in the data base. The rate of multiple counting was higher in some product fields than others. It was highest in the transportation equipment product fields, because many transportation-related patent subclasses, such as brakes, do not specify the mode of transportation to which they apply, e.g. automobiles or airplanes. While there is no a prior theoretical reason not to assign multiple product fields to patent subclasses, the practice has led to potentially serious, misleading analytical results. As shown by Soete at the Workshop, analysis of Concordance data on aircraft and aircraft equipment showed a dramatic rise in patenting by Japan and declines in patenting by the United States and the United Kingdom. Further analysis, however, showed that the apparent rise in Japanese aircraft patenting was in fact due to the multiple assignment of patents from the Japanese automobile industry.

A major problem addressed by the OTAF study was the error level in the Concordance. Error in the Concordance may be viewed in two ways. First, error is possible in the assignment of patent subclasses to product fields, because some subclasses may not be assigned according to the decision rules. Second, error is possible at the level of individual patents because whole subclasses are assigned; that is, individual patents may be incorrectly assigned, even if the subclass is correctly assigned.

OTAF investigated the first type of error by examining the assignments of a random sample of 110 subclasses. The criterion for correct assignment was very strict. The Concordance assignment had to be entirely correct, that is, it had to include all the correct product fields and not include any incorrect product fields. Using this criterion, the overall error rate in the Concordance was estimated to be 32.5%. The error rate at more aggregated levels, i.e., at the level of 2-digit SIC product fields, was somewhat lower -- 19.8%. A strong relationship was found between error in assignment and the number of product fields to which a subclass was assigned. Subclasses with one assignment had a 18.9% error rate while subclasses with four assignments had a 92.9% error rate. All subclasses with more than four assignments were found to be incorrect. The error rates also varied by product field groupings; the rate of error was highest for nonelectrical machinery product fields and was lowest for rubber and miscellaneous plastic product fields.

OTAF investigated error in the assignment of individual patents by examining the Concordance assignments of a set of patents that had been assigned SIC codes manually in a similar analysis by Scherer. Scherer had found that the Concordance assignment agreed with the individual patent assignment at both the 2- and 3-digit SIC levels for 50 of 99 patents.

Looking at the remaining 49 patents, and using the same criterion for percent correct that Scherer used, OTAF determined that 42 were in agreement. Hence the overall Concordance error rate in the assignment of individual patents was estimated to be 7% (i.e., 100 - 42+50/99). No definite conclusion can be drawn from this analysis because of the small sample of patents examined. Moreover, because the criterion for correct was different from that used in the analysis of error at the subclass level, the two error rates cannot be compared.

Additional problems were investigated in the course of the OTAF review and assessment. U.S. patents are generally assigned to more than one USPC subclass. One of these classifications is designated as the original classification (OR); all others are designated cross-reference classifications (XR's). The OR is not necessarily the most important or best classification. A general question pertaining to patent counts as measures of inventive activity is whether the patent should be counted only in its OR subclass or in both its OR and XR subclasses. Theoretically it is better to count both ORs and XRs because if one counts only OR's, one will miss inventive activity in the other, XR, fields. Early reports based on the Concordance counted both OR's and XR's. The disadvantage of counting both is that it results in a yet higher level of multiple counting and increased multiple counting is related to higher error levels. For this reason, recent versions of the Concordance have been based on OR's only. OTAF's investigation showed that the use of only OR's leads to different analytical results than the use of OR's and XR's, but the magnitude of the difference remains unknown pending further study.

Concordance users have expressed the desire for additional and more disaggregated product fields. Researchers would like additional manufacturing groups, as well as nonmanufacturing groups. They would also like additional detail in some areas. OTAF identified eight SIC major manufacturing groups that could be broken out of the existing "All Other" category. OTAF also identified several nonmanufacturing SICs that might usefully be added to the Concordance, although the addition of nonmanufacturing SIC's would require a complete reworking of the Concordance. The need for additional SICs did not arise as a major issue at the Workshop and as a result none were added in the 1984 update.

As a result of the OTAF review and the Workshop discussion, several changes were made in the Concordance. These changes were made during the course of the 1984 biennial update. A team of three PTO Classifiers, each expert in one of the technology disciplines -- e.g., chemical, electrical, and mechanical -- assigned SIC's to the new subclasses created since the last update. The decision rules were modified slightly and additional guidelines were established in order to assure that the assignments were made according to the philosophy of the Concordance. Although no absolute limit was imposed on the number of product field assignments, classifiers were instructed to use a reasonable interpretation based on the subclass content -- i.e., not to think of unusual or extreme cases. All USPC classes were examined to correct subclass assignment errors, resulting in about 35,000 changes to subclass assignments. Two reports were prepared on the updated data, the traditional report with multiple counting and a

report that eliminates multiple patent counts through a system of fractional counting. No new product fields were added, nor was the scope of the product fields changed. The reports were based on OR's only, as have been other recent Concordance reports.

With these changes, some of the main criticisms of the Concordance have been addressed. The level of error in subclass assignments should now be substantially lower than in the previous Concordance. The new, fractionalized report should correct for some misleading biases due to multiple counting. The modified decision rules and additional guidelines, along with institutional changes in the update process should assure that more accurate assignments are made.

The essential nature of the Concordance remains unchanged, however. It is still an assignment of SIC product fields to patent subclasses and therefore reports based on it are liable to the inherent shortcomings of patent data and the USPC for measuring the inventive activity of industry. The issue of whether OR's or both ORs' and XR's should be counted remains unresolved. Subclasses whose proper SIC assignments are ambiguous are still included. The Concordance still covers a limited number of fairly aggregated manufacturing SICs. For these reasons, use of the Concordance for any particular analytical project should be considered on a case-by-case basis.

The main conclusion arising from the project is that the Concordance should be continued until a superior method of assigning patents to industry categories can be developed and implemented. With the improvements instituted in the 1984 update, the Concordance should meet the need to relate patents to aggregate-level industry categories fairly well.

The primary alternative to the Concordance discussed at the Workshop was patent-by-patent assignment of SICs, on the model of the Canadian system. The Canadian Patent Office assigns SIC's for both industry of manufacture and industry of use to each patent that issues. It is widely presumed that patent-by-patent assignment of SIC's would solve some of the problems associated with the Concordance. Among the benefits attributed to patent-by-patent assignment are more accurate assignments, avoidance of periodic updates, and the ability to assign SIC's of use as well as manufacture. The opportunity to assign SIC's to patents as they issue may arise as part of the review of patent classification issues that has been engendered by the automation effort in the PTO. Administrator for Documentation Lawson indicated that the PTO is seriously interested in such a system but that the costs and benefits of such a move must be compared to alternative patent information dissemination and retrieval mechanisms that might be initiated. Moreover, the presumed advantages of patent-by-patent assignment should be verified before a commitment is made to the approach.

There could be drawbacks to implementing a system of assigning SIC's to patents before the SIC Manual is revised, however. The revision, which is scheduled for July 1, 1986, will likely require extensive changes in some areas such as instruments, electronics, and nonmanufacturing. Ideally any new system based on SIC's would not be implemented until the revised

Manual is available. Nevertheless, if the need for a more detailed means of relating patents to industries exists and resources become available, it may be wise to act sooner. If patent-by-patent assignment is successfully implemented, both the patent-by-patent system and the Concordance should be continued in parallel for some period of time. Then, based on experience with both systems, a decision can be made whether to discontinue the Concordance.

Patent-by-patent assignment of SIC's is not the only alternative to the Concordance. Other methods of relating patents to industrial or technological categories that could be explored include:

- Use of a variety of types of patent data to characterize technologies, similar to the efforts being undertaken by Computer Horizons, Inc., and Battelle Pacific Northwest Laboratories
- Use of the International Patent Classification system either to determine weightings for USPC subclasses that are counted more than once in the Concordance or to develop an international Concordance
- Use of an industry-oriented patent classification system such as Derwent

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES (cont.)

LIST OF FIGURES

INTRODUCTION

As the importance of science and technology to the American economy and international competitiveness has been increasingly recognized, efforts to develop measures or indicators of scientific and technological progress and their impacts have also increased. Patent data are of particular importance because patents are a result of research and development (R&D) and to some extent reflect commercial interest. Also, they are available in long time series and cover most technologies. A key objective in developing patent data as science and technology indicators has been to organize them in a way that: (1) is suitable for studying levels of invention in various industries or product fields and (2) allows them to be related to other data organized on a similar basis. The OTAF Concordance between the U.S. Patent Classification (USPC) and the Standard Industrial Classification (SIC) systems is the main attempt to date to organize patent data in such a way.

This report summarizes a project aimed at improving the OTAF Concordance. The purpose of the project was: (1) to review the Concordance thoroughly in light of recent criticism, (2) to assess the magnitude and import of any problems found and the feasibility of correcting those problems, and (3) to propose corrective measures. The project was begun in June 1983 with the aim of being completed in time for the scheduled update of the Concordance in April 1984.

The project consisted of three phases: an in-depth study conducted by OTAF staff, a workshop held to review the study, and changes made in the 1984 update. Part I of this report presents the OTAF review. Part II summarizes the discussion at the workshop. Part III presents the recommendations that arose from the project and Part IV describes changes made in the 1984 update. The rest of the Introduction describes the OTAF Concordance and the background of the project.

The OTAF Concordance

The OTAF Concordance is a system for relating U.S. patents, which are organized according to the USPC system, to the SIC system. It is important to note at the outset that the Concordance does not assign individual patents to the SIC. Rather, it assigns patent subclasses to SIC-based "product fields."

The Concordance was established in 1974 through the cooperative effort of OTAF and the Science Indicators Unit of the National Science Foundation. OTAF constructed the original Concordance, with NSF funding, by manually assigning more than 85,000 patent subclasses to one or more of 36 product fields based on aggregations of SIC codes at the 2- and 3-digit levels. The subsequent Concordances resulted from a series of biennial revisions to the original Concordance necessitated by the continuing evolution of the USPC system. The version of the Concordance which was the focus of this study was the 1981 Concordance, covering the years 1963

(the first year the PTO picked up any bibliographic data on patents other than patent number and classification in machine-readable form) through 1981. Unless otherwise noted, that is the version of the Concordance to which this report refers.

Today's Concordance covers 55 SIC-based product fields. (See Table 1.) Patent subclasses which relate to technology not included in any of product fields 1 through 55 are assigned to product field 99, "All Other." Product field 99 accounts for only about 5.3% of patents.

Of the 55 SIC product fields, there are 12 major groupings which correspond generally to the 2-digit SIC codes. Five of these--product fields 3, 18, 22, 33 and 44--are disaggregated into product fields that correspond to 3-digit SIC codes and combinations of 3- and 4-digit SIC codes.

Further disaggregation within product fields is shown by indentation of the titles. Product fields 3, 4, 5, 10, 18, 22, 28, 33, 34, 37, 41, 44, 45, and 48 are combinations or "roll-ups" of the product fields indented under them. The remaining product fields are referred to as unique product fields.

Table 2 shows a page from the computer printout that is the closest thing to a physical embodiment of the Concordance. The full document is 583 pages long. It consists of a listing in numerical order of the approximately 100,000* subclasses in the USPC system along with their product field assignments. The key to Table 2 shows how to read the subclass numbers. After each Subclass there is room for up to nine product field assignments. Another document, called the "Reverse Concordance," shows for each SIC product field the patent subclasses which have been assigned to it. (See Table 3.)

The Concordance makes it possible to extract U.S. patents from the computerized OTAF data base** and classify them into SIC product fields. This capability allows one to show the distribution of U.S. patent activity across the 55 SIC product fields. Computer reports can be prepared for each SIC product field that show the total level of U.S. patenting for each year, as well as the level of patenting by country of origin and by ownership. The data can be distributed by date of patent grant and by date of patent application. Both numbers of patents and percentages can be shown. (See Appendix B.)

---

*The number of subclasses contained in the OTAF Concordance increases with each update because of reclassification within the USPC. The number of subclasses in the OTAF data base as of June 30, 1983, was 101,104.

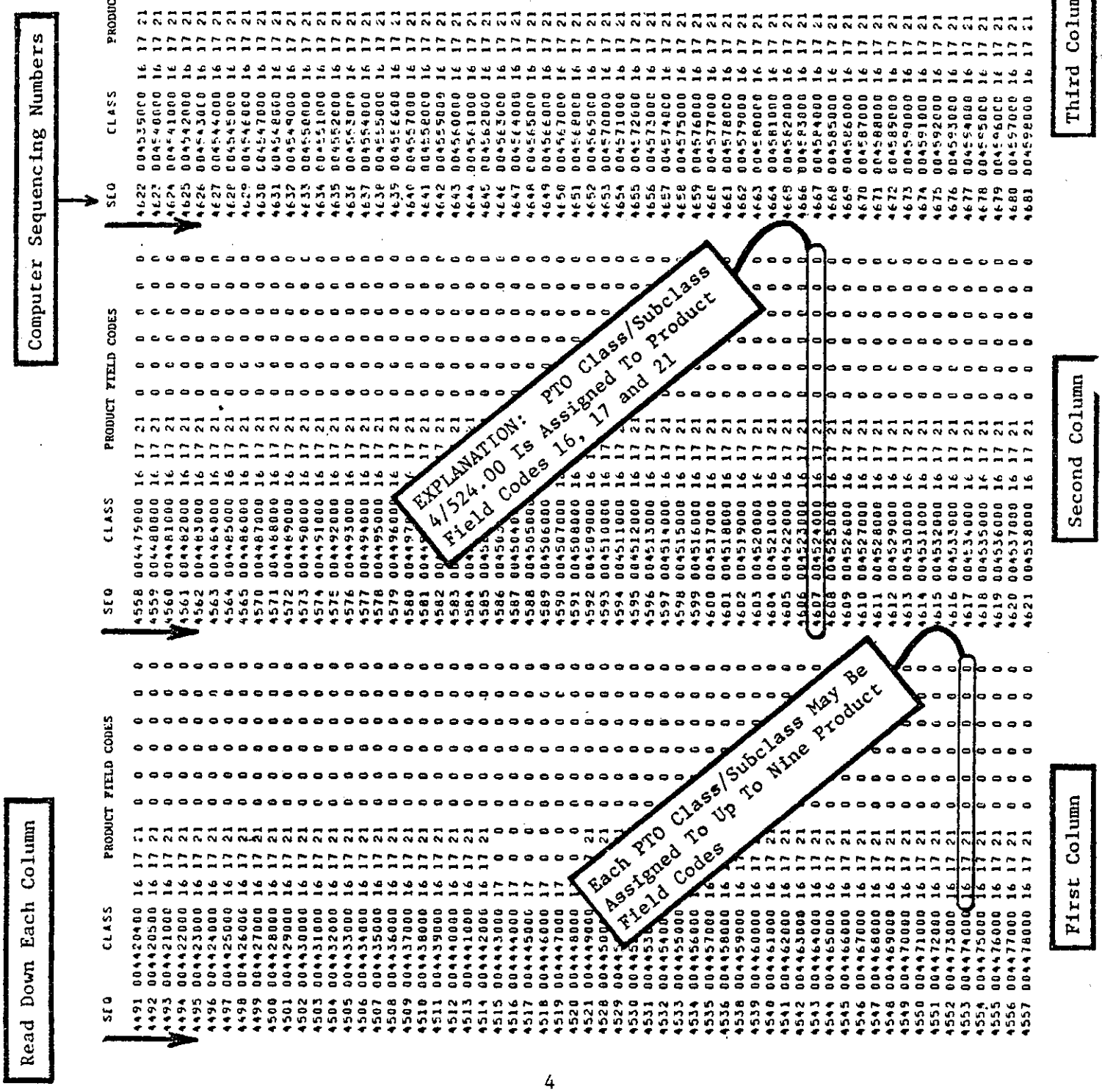
**See Appendix A for a description of the OTAF data base.

TABLE I

STANDARD INDUSTRIAL CLASSIFICATION PRODUCT FIELDS
(AS OF 12/31/83)

CODE	Product Fields	SIC Code
1	FOOD AND KINDRED PRODUCTS	20
2	TEXTILE MILL PRODUCTS	22
3	CHEMICALS AND ALLIED PRODUCTS	28
4	Chemicals, Except Drugs & Medicines	281, 282, 284-289
5	Basic Industrial Inorganic & Organic Chemistry	281, 286
6	Industrial Inorganic Chemistry	281
7	Industrial Organic Chemistry	286
8	Plastics Materials & Synthetic Resins	282
9	Agricultural Chemicals	287
10	All Other Chemicals	284, 285, 289
11	Soaps, Detergents, Cleaners, Perfumes, Cosmetics & Toiletries	284
12	Paints, Varnishes, Lacquers, Enamels, & Allied Products	285
13	Miscellaneous Chemical Products	289
14	Drugs and Medicines	283
15	PETROLEUM & NATURAL GAS EXTRACTION & REFINING	13, 29
16	RUBBER & MISCELLANEOUS PLASTICS PRODUCTS	30
17	STONE, CLAY, GLASS AND CONCRETE PRODUCTS	32
18	PRIMARY METALS	33, 3462, 3463
19	Primary Ferrous Products	331, 332, 3399, 3462
20	Primary & Secondary Non-Ferrous Metals	333-336, 339 (except 3399), 3463
21	FABRICATED METAL PRODUCTS	34 (except 3462, 3463, 348)
22	MACHINERY, EXCEPT ELECTRICAL	35
23	Engines & Turbines	351
24	Farm & Garden Machinery & Equipment	352
25	Construction, Mining & Material Handling Machinery & Equipment	353
26	Metal Working Machinery & Equipment	354
27	Office Computing & Accounting Machines	357
28	Other Machinery, Except Electrical	355, 356, 358, 359
29	Special Industry Machinery, Except Metal Working Machinery	355
30	General Industrial Machinery & Equipment	356
31	Refrigeration & Service Industry Machinery	358
32	Miscellaneous Machinery, Except Electrical	359
33	ELECTRICAL AND ELECTRONIC MACHINERY, EQUIPMENT AND SUPPLIES	36, 3825
34	Electrical Equipment, Except Communication Equipment	361-364, 369, 3825
35	Electrical Transmission & Distribution Equipment	361, 3825
36	Electrical Industrial Apparatus	362
37	Other Electrical Machinery, Equipment & Supplies	363, 364, 369
38	Household Appliances	363
39	Electrical Lighting & Wiring Equipment	364
40	Miscellaneous Electrical Machinery, Equipment & Supplies	369
41	Communication Equipment & Electronic Components	365-367
42	Radio & Television Receiving Equipment Except Communication Types	365
43	Electronic Components & Accessories & Communication Equipment	366-367
44	TRANSPORTATION EQUIPMENT	37, 348
45	Motor Vehicles & Other Transportation Equipment, Except Aircraft	348, 371, 373-376, 379
46	Motor Vehicles & Motor Vehicle Equipment	371
47	Guided Missiles & Space Vehicles & Parts	376
48	Other Transportation Equipment	373-375, 379 (except 3795)
49	Ship & Boat Building & Repairing	373
50	Railroad Equipment	374
51	Motorcycles, Bicycles & Parts	375
52	Miscellaneous Transportation Equipment	379 (except 3795)
53	Ordnance Except Missiles	348, 3795
54	Aircraft & Parts	372
55	PROFESSIONAL & SCIENTIFIC INSTRUMENTS	38 (except 3825)
99	ALL OTHER SIC's	

TABLE 2
A PAGE FROM THE CONCORDANCE PRINTOUT



Read Across Each Row

TABLE 3
A PAGE FROM THE "REVERSE CONCORDANCE"

CLASSIFICATIONS WITH PRODUCT FIELD CODE 1

All PTO Subclasses Assigned To Product Field Code 1 Are Listed In Order

EXAMPLE: PTO Class/Subclass 426/99.00 Is Assigned To Product Field Code 1

017001000A	017001000A	017001000B	017001000C	017001000D	017001000E	017001000F	017001000	017001000R	017001000S
017001000T	017011000	017011100A	017011100R	017011200	017011300	017012	000	017014000	017015000
017016000	017017000	017018000	017019000	017020000	017021000	017023000		017042100	017043000
017044000	017044100	017044200	017044300	017044400	017045000	017046000		017048000	017049000
017050000	017051000	017052000	017053000	017054000	017055000	017056000		17058000	017059000
017060000	017061000	017062000	017063000	017064000	017065000	017066000	0170	8000	017069000
017070000	017071000	017072000	017073000	017074000	017075000	099DI8011	099DI80	13	099DI8014
099DI8015	127029000	127030000	127031000	127032000	127033000	127034000	127036000		127038000
127039000	127040000	127041000	127042000	127043000	127044000	127045000	127046100		127046300
127047000	127048000	127049000	127050000	127051000	127052000	127053000	127054000	12	127056000
127057000	127058000	59000	127060000	127061000	127062000	127063000	127064000	127065000	127066000
127067000	12706800	00	127070000	127071000	204138000	308DI8014	308DI8015	426001000	426002000
426003000	426004		426006000	426007000	426008000	426009000	426010000	426011000	426012000
426013000	426014000		426016000	426017000	426018000	426019000	426020000	426021000	426022000
426023000	426024000		026000	426027000	426028000	426029000	426030000	426031000	426032000
426033000	426034000	42	000	426037000	426038000	426039000	426040000	426041000	426042000
426043000	426044000	4260	0	426047000	426048000	426049000	426050000	426051000	426052000
426053000	426054000	42605500		426057000	426058000	426059000	426060000	426061000	426062000
426063000	426064000	426066000		426068000	426069000	426071000	426072000	426073000	426074000
426075000	426076000	426077000		426079000	426080000	426081000	426082000	426083000	426084000
426085000	426086000	426087000	42	426089000	426090000	426091000	426092000	426093000	426094000
426095000	426096000	426097000	426098000	426099000	426100000	426101000	426102000	426103000	426104000
426105000	426106000	426107000	426108000	426109000	426110000	426111000	426112000	426113000	426114000
426115000	426116000	426117000	426118000	426119000	426120000	426121000	426122000	426123000	426124000
426125000	426126000	426127000	426128000	426129000	426130000	426131000	426132000	426133000	426134000
426135000	426138000	426139000	426140000	426143000	426144000	426231000	426232000	426233000	426234000
426235000	426236000	426237000	426238000	426239000	426240000	426241000	426242000	426243000	426244000
426245000	426246000	426247000	426248000	426249000	426250000	426251000	426252000	426253000	426254000
426255000	426256000	426257000	426258000	426259000	426260000	426261000	426262000	426263000	426264000
426265000	426266000	426267000	426268000	426269000	426270000	426271000	426272000	426273000	426274000
426275000	426276000	426277000	426278000	426279000	426280000	426281000	426282000	426283000	426284000

Since 1974, Science Indicators has used data based on OTAF's Concordance to indicate the output of U.S. industrial R&D and to compare U.S. R&D output with that of other countries.* Researchers have also used the Concordance to relate patent activity to R&D and economic data at the industry or product line level.**

Criticisms of the Concordance

Despite the evident need for the functions performed by the OTAF Concordance -- i.e., monitoring invention by industry and relating patent data to other R&D and economic data -- there have been questions about the adequacy of the Concordance since its inception. The history in Part I will show that certain limitations of the Concordance have been recognized since it was originally constructed. These limitations have been reemphasized and additional problems have been recognized over time, as the Concordance has seen wider use. Perhaps best known are the recent criticisms by F.M. Scherer, then of Northwestern University, who attempted unsuccessfully to use the Concordance to relate patent data to R&D expenditures at the Federal Trade Commission's line-of-business level.*** Luc Soete of the University of Sussex, Zvi Griliches of Harvard University, and Mark Carpenter of Computer Horizons, Inc., have also had problems in using the Concordance.

The following are the main criticisms of the Concordance:

Error Level in the Concordance. A fundamental criticism has been made of the OTAF approach of assigning patent subclasses to SIC product fields. Professor Scherer has reported that -- for his purpose of linking patents to the industry where the originating R&D was done -- he obtained better results by assigning SIC codes to patents on an individual basis than through the OTAF Concordance. His criticism raised two primary issues: (1) OTAF's error level in assigning patent subclasses to SIC product fields, and (2) the error level in assignments resulting at the patent level when assignments are made at the subclass level.

---

*U.S. National Science Board. Science Indicators 1974 (Washington, D.C.: U.S. Govt. Print. Off., 1975). Also Science Indicators 1976, 1978, 1980, and 1982.

**See, for example: Mark P. Carpenter, "Patent Citations as Indicators of Scientific and Technological Linkages," paper presented at the Annual Meeting of the American Association for the Advancement of Science, Detroit, MI, May 30, 1983; Keith Pavitt and Luc Soete, "Innovative Activities and Export Shares: Some Comparisons between Industries and Countries," in Technical Innovation and British Economic Performance, ed. by Keith Pavitt (London: Macmillan Press, 1980).

***F.M. Scherer, "The Office of Technology Assessment and Forecast Industry Concordance as a Means of Identifying Industry Technology Origins," World Patent Information, v. 4 (1982): 12 - 17.

Multiple Assignments. The Concordance assigns patent subclasses to all the SIC product fields to which they pertain. Prior to this project, the patents in a subclass which was multiply assigned were counted more than once. Thus in the 1981 Concordance, the total number of patents granted during the period is less than the sum of the patent totals of all unique product fields (See Table 4*) since the same patent could be counted in more than one product field. Likewise, the number given for each combination product field is less than the sum of the patent totals for its components because patents appearing in more than one of the components were counted only once in forming the combination field.

Table 4 also lists the number of subclasses in each of the unique product fields. Since a subclass can be concorded to more than one product field, the sum of the numbers given is greater than the actual number of subclasses (about 101,000) in the Concordance.

The practice of multiple counting is said to introduce bias, especially when patents in an SIC product field with a high propensity to patent are also assigned to an SIC product field with a low propensity to patent. For example, Soete claims that while the Concordance appears to show that Japan and West Germany have high levels of U.S. patenting activity in aircraft, this is in fact due to the multiple assignment of patents for internal combustion engines and other inventions for use in automobiles. Moreover, Soete reports that the level of multiple assignments is increasing over time and is higher in the newest technology areas, in the most rapidly growing patent subclasses, and for foreign-origin U.S. patenting activity.

Coverage. Users have expressed the desire for a patent concordance to additional SIC product fields not currently covered -- such as agriculture, construction, and additional manufacturing groups. Some of the SIC groups specifically requested are:

SIC 21 Tobacco products
SIC 23 Apparel, other textile products
SIC 24 Lumber and wood products
SIC 25 Furniture and fixtures
SIC 26 Paper and allied products
SIC 27 Printing and publishing
SIC 31 Leather and leather products
SIC 39 Miscellaneous manufacturing industries

---

*Tables based on the Concordance are generally constructed using only the original classifications (ORs) of the patents. U.S. patents also have cross reference classifications (XRs) to other patent subclasses, but these are not involved in the data discussed here.

TABLE 4

Number of Patents and Number of Subclasses
by SIC Product Field: 1981 Concordance

SEQ. NO.	Product Fields	#Patents 1963-1981	# Subclasses*
1	FOOD AND KINDRED PRODUCTS	14,105	1,036
2	TEXTILE MILL PRODUCTS	12,753	957
3	CHEMICALS AND ALLIED PRODUCTS	204,445	---
4	Chemicals, Except Drugs & Medicines	201,786	---
5	Basic Industrial Inorganic & Organic Chemistry	113,548	---
6	Industrial Inorganic Chemistry	25,005	1,304
7	Industrial Organic Chemistry	96,918	5,559
8	Plastics Materials & Synthetic Resins	47,333	2,363
9	Agricultural Chemicals	24,560	894
10	All Other Chemicals	19,331	---
11	Soaps, Detergents, Cleaners, Perfumes, Cosmetics & Toiletries	6,780	279
12	Paints, Varnishes, Lacquers, Enamels, & Allied Products	1,138	121
13	Miscellaneous Chemical Products	12,158	763
14	Drugs and Medicines	30,945	923
15	PETROLEUM & NATURAL GAS EXTRACTION & REFINING	17,098	878
16	RUBBER & MISCELLANEOUS PLASTICS PRODUCTS	64,353	5,049
17	STONE, CLAY, GLASS AND CONCRETE PRODUCTS	29,825	2,261
18	PRIMARY METALS	16,307	---
19	Primary Ferrous Products	11,649	1,177
20	Primary & Secondary Non-Ferrous Metals	9,693	587
21	FABRICATED METAL PRODUCTS	157,200	17,176
22	MACHINERY, EXCEPT ELECTRICAL	385,768	---
23	Engines & Turbines	32,731	3,124
24	Farm & Garden Machinery & Equipment	38,762	5,219
25	Construction, Mining & Material Handling Machinery & Equipment	64,893	7,014
26	Metal Working Machinery & Equipment	37,738	4,348
27	Office Computing & Accounting Machines	41,396	4,187
28	Other Machinery, Except Electrical	239,342	---
29	Special Industry Machinery, Except Metal Working Machinery	98,579	10,740
30	General Industrial Machinery & Equipment	121,843	12,105
31	Refrigeration & Service Industry Machinery	29,761	3,255
32	Miscellaneous Machinery, Except Electrical	18,754	1,704
33	ELECTRICAL AND ELECTRONIC MACHINERY, EQUIPMENT AND SUPPLIES	251,272	---
34	Electrical Equipment, Except Communication Equipment	133,272	---
35	Electrical Transmission & Distribution Equipment	44,092	2,686
36	Electrical Industrial Apparatus	40,009	2,985
37	Other Electrical Machinery, Equipment & Supplies	66,240	---
38	Household Appliances	22,042	2,046
39	Electrical Lighting & Wiring Equipment	16,274	1,703
40	Miscellaneous Electrical Machinery, Equipment & Supplies	27,713	1,776
41	Communication Equipment & Electronic Components	143,156	---
42	Radio & Television Receiving Equipment Except Communication Type	24,999	1,709
43	Electronic Components & Accessories & Communication Equipment	140,142	8,701
44	TRANSPORTATION EQUIPMENT	84,131	---
45	Motor Vehicles & Other Transportation Equipment, Except Aircraft	78,246	---
46	Motor Vehicles & Motor Vehicle Equipment	46,243	4,600
47	Guided Missiles & Space Vehicles & Parts	8,263	466
48	Other Transportation Equipment	26,303	---
49	Ship & Boat Building & Repairing	7,673	633
50	Railroad Equipment	13,788	1,909
51	Motorcycles, Bicycles & Parts	2,918	243
52	Miscellaneous Transportation Equipment	14,598	1,246
53	Ordnance Except Missiles	8,580	925
54	Aircraft & Parts	27,201	2,790
55	PROFESSIONAL & SCIENTIFIC INSTRUMENTS	149,427	9,829
99	ALL OTHER	100,799	10,428

*Active subclasses in the OTAF data base as of December 31, 1981. Number of subclasses not recorded for "roll-up" categories.

Disaggregation. Users have also requested additional disaggregation of portions of the existing Concordance. Currently, 7 of the 12 major product fields are concorded only to the 2-digit SIC level. Increased detail would allow more precision. Even where the existing Concordance is to the 3-digit SIC level, additional detail could be useful. For example, OTAF has been asked to consider further disaggregating SIC 355 -- Special Industry Machinery -- which includes Food Machinery, Textile Machinery, Woodworking Machinery, Paper Industries Machinery, Printing Machinery, and Other Special Machinery not elsewhere classified.

Need for Better Documentation. The construction of the original Concordance was not well documented. The lack of documentation has made it difficult to ensure consistency in the biennial updates. It also makes it difficult to locate possible errors in assignment. Professor Griliches has suggested that improved documentation would also be helpful to researchers using the Concordance.

In view of the above criticisms OTAF proposed, and the Science Indicators Unit agreed to fund, a thorough review and assessment of the Concordance. That project is the subject of this report.

PART I

REVIEW AND ASSESSMENT

Introduction

This part of the report describes OTAF's internal review and assessment of the USPC-SIC Concordance. The study effort consisted of the following elements which are discussed in order:

- History of the Concordance
- Analysis of differences between trends based on OR's only and on OR's plus XR's
- Decision rules
- Analysis of the error rate in the Concordance's assignments of subclasses to product fields
- Analysis of the error rate in the assignment of individual patents to product fields based on the Concordance
- Analysis of the extent of multiple assignments
- Identification of possible additional SIC product fields
- Identification of opportunities to disaggregate further SIC product fields already in the Concordance
- Use of the Concordance

History of the Concordance

Many of the issues being addressed with respect to the OTAF Concordance today have been recognized throughout its life-span of nearly a decade. This section briefly describes the early Concordance efforts and the process of biennial updates since then. It also describes an alternative approach that was attempted unsuccessfully. This section is based primarily on internal OTAF and NSF documents.

The Original Concordance

In 1973 the Science Indicators Unit of the National Science Foundation began considering the development of patent data for use in its Science Indicators report series. It was believed that patent statistics could be of value in the development of output-oriented indices. While NSF recognized that patent statistics were far from being a perfect output indicator, it felt that they could be helpful in determining overall trends in science and technology, or at least as a fairly reliable index of "inventive activity."

The difficulties of using patent statistics as technology indicators were recognized even at that early time. Of particular concern were (1) certain "inherent difficulties" and (2) the "organization" of patent data. "Inherent difficulties" included, for example, the inability of a straight patent count to distinguish patents of different scientific and/or economic value; the unknown proportion of inventions that are not patented; and factors such as rising cost and pendency (i.e., the time that lapses between the filing date of a patent application and its grant date) that affect the rate at which inventions are patented.

Nonetheless, it was felt that the advantages of patent statistics as prospective indicators outweighed the disadvantages for several reasons. First, most patented inventions have passed the test of technological and economic feasibility. Second, patent statistics were felt to be a more practical and reliable assessment of innovation than listings of important innovations, based on the findings of Schmookler and others in the 1960's. Third, the dearth of tangible output statistics of any kind necessitated using patent data.

The USPC. The problem of the "organization" of patent data pertained to the nature of the U.S. Patent Classification (USPC) system. The basic problem facing the developers of the original Concordance remains today and can be illustrated using the 1981 USPC system. The 1,234,650 U.S. patents that were issued from 1963-1981 are organized into 352 classes and about 101,000 subclasses (excluding design and plant patents). The purpose of the USPC is to facilitate the search procedure carried out by patent examiners; therefore, it is based on the technological features of inventions that make them patentable. The basis for classification is largely functional and, to a lesser extent, structural. As seen in Table 5, classes bear such titles as "Dispensing" (Class 222) and "Fluent Material Handling" (Class 141).

The classes and subclasses of the USPC often do not correspond to "technologies" or "industries" as they are commonly defined. For example, such different inventions as windmills and egg beaters are found in the same class, "Impellers" (Class 416). Another example, Class 424, "Drugs, Bio-affecting and Body Treating Compositions," includes both pharmaceuticals and pesticides. It is impossible to separate pharmaceuticals from pesticides at the subclass level. In fact some patents may specify that a composition is both a drug and a pesticide. Therefore, a person interested in measuring inventive activity in pharmaceuticals cannot separate it from that in pesticides. Other examples abound.

The other side of this problem is that some technologies or industries are spread across many classes and subclasses. For example, a recent OTAF report on Industrial Robots required the manual searching of more than 100 subclasses in 13 classes and arrived at a total of 212 patents. These patents were distributed over more than 64 subclasses which each contained 3 or more robot patents. For these reasons, many analysts believe that the USPC is not suitable for measuring the growth of technologies or industries.

TABLE 5

UTILITY PATENT CLASS TITLES IN CLASS NUMBER ORDER

Class	Title
D 1	Foodstuffs
D 2	Apparel and Haberdashery
D 3	Travel Goods, Cases, Umbrellas and Personal Belongings not Elsewhere Specified
D 4	Brushware
D 6	Furnishings
D 7	Household Articles, Not Elsewhere Specified
D 8	Tools and Hardware
D 9	Packages and Containers for the Transport or Handling of Goods
D10	Measuring, Testing or Signaling Instruments
D11	Butchering
D12	Transportation or Hoisting
D13	Equipment for Production, Distribution or Transformation of Electricity
D14	Recording, Communication or Information Retrieval Equipment
D15	Machines, Not Elsewhere Specified
D16	Photographic, Cinematographic and Optical Equipments
D17	Musical Instruments
D18	Printing and Office Machinery and Equipment, Not Elsewhere Specified
D19	Stationery, Artists and Teachers Materials, and Office Equipment, Not Elsewhere Specified
D20	Sales and Advertising Equipment and Signs
D21	Games, Toys, Tents and Sports Goods
D22	Arms, Pyrotechnic Articles, Articles for Hunting, Fishing and Pest Killing
D23	Fluid Distribution Equipment, Sanitary, Heating, Ventilation and Air Conditioning Equipment, Solid Fuel
D24	Medical and Laboratory Equipment
D25	Building Units and Construction Elements
D26	Lighting
D27	Tobacco and Smokers Supplies
D28	Pharmaceutical and Cosmetic Products, Toilet Articles and Apparatus
D29	Devices and Equipment Against Fire Hazards, for Accident Prevention and for Rescue
D30	Care and Handling of Animals
D32	Washing, Cleaning or Drying Machine
D33	Embroidery and Trimmings
D34	Material or Article Handling Equipment
D47	Knitting and Netting
D59	Paper Manufacturers
D64	Printing
D92	Weaving, Sheet Material and Linoleum
D99	Miscellaneous
Plt	Plant
2	Apparel
3	Artificial Body Members
4	Baths, Closets, Sinks and Spittoons
5	Beds
6	Bee Culture
7	Compound Tools
8	Bleaching and Dyeing; Fluid Treatment and Chemical Modification of Textiles and Fibers
9	Boats, Buoys and Aquatic Devices
10	Bolt, Nail, Nut, Rivet and Screw Making
11	Books, Making
12	Boot and Shoe Making
13	Electric Furnaces
14	Bridges
15	Brushing, Scrubbing and General Cleaning
16	Miscellaneous Hardware
17	Butchering
19	Textiles, Fiber Preparation
23	Chemistry, Analytical and Physical Processes
24	Buckles, Buttons, Clasps, etc.
26	Textiles, Cloth Finishing
27	Undertaking
28	Textiles, Manufacturing
29	Metal Working
30	Cutlery
33	Geometrical Instruments
34	Drying and Gas or Vapor Contact with Solids
35	Education
36	Boots, Shoes and Leggings
37	Excavating
38	Textiles, Ironing or Smoothing
40	Card, Picture and Sign Exhibiting
42	Firearms
43	Fishing, Trapping and Vermin Destroying
44	Fuel and Igniting Devices
46	Amusement Devices, Toys
47	Plant Husbandry
48	Gas, Heating and Illuminating
49	Movable or Removable Closures
51	Abrading
52	Static Structures, e.g., Buildings
53	Package Making
54	Harness
55	Gas Separation
56	Harvesters
57	Textiles, Spinning, Twisting and Twining
59	Chain, Staple and Horseshoe Making
60	Power Plants
62	Refrigeration
63	Jewelry
64	Machine Elements, Shafting and Flexible Shaft Couplings
65	Glass Manufacturing
66	Textiles, Knitting
68	Textiles, Fluid Treating Apparatus
69	Leather Manufactures
70	Locks
71	Chemistry, Fertilizers
72	Metal Deforming
73	Measuring and Testing

TABLE 5 (continued)

Class	Title
74	Machine Elements and Mechanisms
75	Metallurgy
76	Metal Tools and Implements, Making
79	Button Making
81	Tools
82	Turning
83	Cutting
84	Music
86	Ammunition and Explosive-Charge Making
87	Textiles, Braiding, Netting and Lace Making
89	Ordnance
91	Motors, Expansible Chamber Type
92	Expansible Chamber Devices
98	Ventilation
99	Foods and Beverages: Apparatus
100	Presses
101	Printing
102	Ammunition and Explosive Devices
104	Railways
105	Railway Rolling Stock
106	Compositions, Coating or Plastic
108	Horizontally Supported Planar Surfaces
109	Safes, Bank Protection and Related Devices
110	Furnaces
111	Planting
112	Sewing
113	Sheet Metal Ware, Making
114	Ships
116	Signals and Indicators
118	Coating Apparatus
119	Animal Husbandry
122	Liquid Heaters and Vaporizers
123	Internal-Combustion Engines
124	Mechanical Guns and Projectors
125	Stone Working
126	Stoves and Furnaces
127	Sugar, Starch and Carbohydrates
128	Surgery
130	Threshing
131	Tobacco
132	Toilet
133	Coin Handling
134	Cleaning and Liquid Contact with Solids
135	Tents, Canopies, Umbrellas and Canes
136	Batteries, Thermoelectric and Photoelectric
137	Fluid Handling
138	Pipes and Tubular Conduits
139	Textiles, Weaving
140	Wireworking
141	Fluent Material Handling, with Receiver or Receiver Coating Means
142	Wood Turning
144	Woodworking
145	Woodworking Tools
147	Coppering
148	Metal Treatment
149	Explosive and Thermic Compositions or Charges
150	Cloth, Leather and Rubber Receptacles
152	Resilient Tires and Wheels
156	Adhesive Bonding and Miscellaneous Chemical Manufacture
157	Wheelwright Machines
159	Concentrating Evaporators
160	Closures, Partitions and Panels, Flexible Portable
162	Paper Making and Fiber Liberation
163	Needle and Pin Making
164	Metal Founding
165	Heat Exchange
166	Wells
168	Farriery
169	Fire-Extinguishers
171	Unearthing Plants or Buried Objects
172	Earth Working
173	Tool Driving or Impacting
174	Electricity, Conductors and Insulators
175	Boring or Penetrating the Earth
176	Nuclear Reactions and Systems
177	Weighing Scales
178	Telegraphy
179	Telephony
180	Motor Vehicles
181	Acoustics
182	Fire Escapes, Ladders, Scaffolds
184	Lubrication
185	Motors, Spring, Weight and Animal Powered
186	Store Service
187	Elevators
188	Brakes
190	Baggage
191	Electricity, Transmission to Vehicles
192	Clutches and Power-Stop Control
193	Conveyers, Chutes, Skids, Guides and Ways
194	Check-Controlled Apparatus
196	Mineral Oils: Apparatus
198	Conveyers, Power-Driven
199	Type Casting
200	Electricity, Circuit Makers and Breakers
201	Distillation: Processes, Thermolytic
202	Distillation: Apparatus
203	Distillation: Processes, Separatory
204	Chemistry, Electrical and Wave Energy
206	Special Receptacle or Package
208	Mineral Oils: Processes and Products
209	Classifying, Separating and Assorting Solids
210	Liquid Purification or Separation
211	Supports, Racks
212	Traversing Hoists
213	Railway Draft Appliances
215	Bottles and Jars
217	Wooden Receptacles
219	Electric Heating
220	Metallic Receptacles
221	Article Dispensing
222	Dispensing
223	Apparel Apparatus
224	Package and Article Carriers
225	Severing by Tearing or Breaking
226	Advancing Material of Indeterminate-Length
227	Elongated-Member-Driving Apparatus
228	Metal Fusion Bonding
229	Paper Receptacles
231	Whips and Whip Apparatus
232	Deposit and Collection Receptacles

TABLE 5 (continued)

Class	Title
233	Centrifugal-Bowl Separators
234	Selective Cutting (e.g., Punching)
235	Registers
236	Automatic Temperature and Humidity Regulation
237	Heating Systems
238	Railways, Surface Track
239	Fluid Sprinkling, Spraying and Diffusing
241	Solid Material Comminution or Disintegration
242	Winding and Reeling
244	Aeronautics
245	Wire Fabrics and Structure
246	Railway Switches and Signals
248	Supports
249	Static Molds
250	Radiant Energy
251	Valves and Valve Actuation
252	Compositions
254	Pushing and Pulling Implements
256	Fences
258	Railway Mail Delivery
260	Chemistry, Carbon Compounds
261	Gas and Liquid Contact Apparatus
264	Plastic and Nonmetallic Article Shaping or Treating: Processes
266	Metallurgical Apparatus
267	Spring Devices
269	Work Holders
270	Sheet-Material Associating or Folding
271	Sheet Feeding or Delivering
272	Amusement and Exercising Devices
273	Amusement Devices, Games
274	Sound Recording and Reproducing
276	Type Setting
277	Joint Packing
278	Land Vehicles, Animal Draft Appliances
279	Chucks or Sockets
280	Land Vehicles
281	Books, Strips and Leaves
282	Manifolding
283	Printed Matter
285	Pipe Joints or Couplings
289	Knots and Knot Tying
290	Prime-Mover Dynamo Plants
291	Track Sanders
292	Closure Fasteners
293	Vehicle Fenders
294	Handling, Hand and Hoist-Line Implements
295	Railway Wheels and Axles
296	Land Vehicles, Bodies and Tops
297	Chairs and Seats
298	Land Vehicles, Dumping
299	Mining or In Situ Disintegration of Hard Material
300	Brush, Broom and Mop Making
301	Land Vehicles, Wheels and Axles
303	Fluid-Pressure Brake and Analogous Systems
305	Wheel Substitutes for Land Vehicles
307	Electrical Transmission or Interconnection Systems
308	Machine Elements, Bearings and Guides
310	Electrical Generator or Motor Structure
312	Supports, Cabinet Structures
313	Electric Lamp and Discharge Devices
314	Electric Lamp and Discharge Devices, Consumable Electrodes
315	Electric Lamp and Discharge Devices, Systems
316	Electric Lamp and Discharge Devices, Manufacture and Repair
318	Electricity, Motive Power Systems
320	Electricity, Battery and Condenser Charging and Discharging
322	Electricity, Single Generator Systems
323	Electricity, Power Supply, Regulation
324	Electricity, Measuring and Testing
325	Modulated Carrier Wave Communication Systems
328	Miscellaneous Electron Space Discharge Device Systems
329	Demodulators and Detectors
330	Amplifiers
331	Oscillators
332	Modulators
333	Wave Transmission Lines and Networks
334	Tuners
335	Electricity, Magnetically Operated Switches, Magnets and Electromagnets
336	Inductor Devices
337	Electricity, Electrothermally or Thermally Actuated Switches
338	Electrical Resistors
339	Electrical Connectors
340	Communications, Electrical
343	Communications, Radio Wave
346	Recorders
350	Optics, Systems and Elements
351	Optics, Eye Examining, Vision Testing and Correcting
352	Optics, Motion Pictures
353	Optics, Image Projectors
354	Photography
355	Photocopying
356	Optics, Measuring and Testing
357	Active Solid State Devices, e.g., Transistors, Solid State Diodes
358	Pictorial Communication: Television
360	Dynamic Magnetic Information Storage or Retrieval
361	Electricity, Electrical Systems and Devices
362	Illumination
363	Electric Power Conversion Systems
364	Electrical Computers and Data Processing Systems
365	Static Information Storage and Retrieval
366	Agitating
367	Communication, Electrical-Acoustic Wave Systems and Devices
368	Horology-Time Measuring Systems or Devices
369	Dynamic Information Storage or Retrieval
370	Mulitplex Communications
371	Error Detection/Correction and Fault Detection/Recovery
375	Pulse or Digital Communications

TABLE 5 (continued)

Class	Title
376	Induced Nuclear Reactions, Systems and Elements
400	Typewriting Machines
401	Coating Implements with Material Supply
402	Binder Device Releasably Engaging Aperture or Notch of Sheet
403	Joints and Connections
404	Road Structure, Process and Apparatus
405	Hydraulic and Earth Engineering
406	Conveyors, Fluid Current
407	Cutters, for Shaping
408	Cutting by Use of Rotating Axially Moving Tool
409	Gear Cutting, Milling, or Planing
410	Freight Accommodation on Freight Carrier
411	Expanded, Threaded, Headed, and Driven Fasteners, and Locked or Coupled Bolts or Nuts
414	Material or Article Handling
415	Rotary Kinetic Fluid Motors or Pumps
416	Fluid Reaction Surfaces (i.e., Impellers)
417	Pumps
418	Rotary Expansible Chamber Devices
422	Process Disinfecting, Deodorizing, Preserving or Sterilizing, and Chemical Apparatus
423	Chemistry, Inorganic
424	Drug, Bio-Affecting and Body Treating Compositions
425	Plastic Article or Earthenware Shaping or Treating: Apparatus
426	Food or Edible Material: Processes, Compositions and Products
427	Coating Processes
428	Stock Material or Miscellaneous Article
429	Chemistry, Electrical Current Producing Apparatus, Product and Process
430	Radiation Imagery Chemistry-Process, Composition of Product
431	Combustion
432	Heating
433	Denistry
434	Education, Demonstration, and Cryptography
435	Chemistry-Molecular Biology and Microbiology
440	Marine Propulsion
455	Telecommunications
474	Endless Belt Power Transmission Systems and Components
493	Manufacturing Container or Tube from Paper: or other Manufacturing from a Sheet or Web
501	Compositions: Ceramic
518	Chemistry, Processes which include A Fischer-Tropsch Reaction; or Purification, Recovery or Conversion of such Processes
521	Part of the Class 520 Series—Synthetic Resins
525	Part of the Class 520 Series—Synthetic Resins
526	Part of the Class 520 Series—Synthetic Resins
528	Part of the Class 520 Series—Synthetic Resins
536	Part of the Class 530–570 Series—Organic Compounds
542	Part of the Class 530–570 Series—Organic Compounds
544	Part of the Class 530–570 Series—Organic Compounds
546	Part of the Class 530–570 Series—Organic Compounds
548	Part of the Class 530–570 Series—Organic Compounds
549	Part of the Class 530–570 Series—Organic Compounds
556	Part of the Class 530–570 Series—Organic Compounds
560	Part of the Class 530–570 Series—Organic Compounds
562	Part of the Class 530–570 Series—Organic Compounds
564	Part of the Class 530–570 Series—Organic Compounds
568	Part of the Class 530–570 Series—Organic Compounds
570	Part of the Class 530–570 Series—Organic Compounds
585	Chemistry, Hydrocarbons

Therefore, the NSF decided to try to develop a system of patent organization along industry lines. Specifically, it sought a form of organization based on the Standard Industrial Classification (SIC), a classification on which much U.S. economic data are directly or indirectly based. If organized in this way, patent data could also be related to other economic data (e.g., R&D expenditures, manpower data, and import/export data).

The SIC. The SIC defines industries in accordance with the composition and structure of the U.S. economy and covers the entire field of economic activities. Its purpose is to promote the comparability of statistics describing various facets of the nation's economy. The SIC classifies establishments, as opposed to enterprises, commodities, products, or occupations. An establishment is "an economic unit, generally at a single physical location where business is conducted or where services or industrial operations are performed,"--e.g., a factory, mine, farm, bank, or hotel. Establishments are classified by the type of activity in which they are engaged.*

The general principles of SIC classification are:

(1) the classification should conform to the existing structure of American industry,

(2) each establishment is to be classified according to its primary activity, and

(3) to be recognized as an industry, the group of establishments constituting the proposed classification must be statistically significant in the number of persons employed, the volume of business done, and other measures of economic activity.

The SIC is structured into Divisions, 2-digit, 3-digit, and 4-digit industry codes at increasing levels of industrial detail. The Divisions and Major Groups (2-digit) are shown in Table 6. The SIC Manual is revised periodically to reflect the changing industrial composition of the economy. The last major revision was in 1972; the next major revision is planned for 1987. In addition to the 4-digit codes found in the SIC Manual, the Bureau of the Census in the Department of Commerce provides 5- and 7-digit codes for manufacturing and mining industries. These are revised and published every five years in connection with the census of manufactures.**

---

*Executive Office of the President, Office of Management and Budget, Standard Industrial Classification Manual, 1972 (Washington, D.C.: U.S. Government Printing Office, n.d.)

**U.S. Department of Commerce. Bureau of the Census. Numerical List of Manufactured Products (Washington, D.C.: U.S. Government Printing Office, 1977.)

TABLE 6

DIVISIONS AND MAJOR GROUPS OF THE STANDARD INDUSTRIAL CLASSIFICATION:

TABLE OF CONTENTS FROM SIC MANUAL 1972

TABLE 6 (continued)

STANDARD INDUSTRIAL CLASSIFICATION

CONTENTS

The NSF Effort. The Science Indicators Unit undertook a project in 1973 to develop indicators based on SIC-oriented patent data for Science Indicators 1974. The project was to consist of four major components: organization of patents by SIC codes, extraction of information from patents, development of indicators, and correlation and analysis with other data.

The Science Indicators Unit initially tried to conduct the project internally. That effort established some of the basic parameters for the subsequent OTAF effort. The project consisted of assigning patent subclasses to SIC-related product areas, signalling a decision not to assign individual patents and to assign patents primarily to the product-related portions of the SIC. The decision was made to place products with the industry that manufactures them rather than with industries that use them, a decision rule that has remained in effect. The NSF effort differed from the OTAF efforts somewhat, however, in that the decision rule in cases of subclasses where the correct assignment was ambiguous was to omit them, rather than to assign them to all possible SICs. Also, processes were assigned to the industry that manufactures the associated apparatus, on the theory that processes are "normally associated intimately with the apparatus which performs the process ...."

Based on this internal project, the Science Indicators Unit concluded that the procedure had "proved inadequate due to the ambiguity involved in assigning certain subclasses to particular product areas." In fact, "there appears to be a relatively large number of patent subclasses which pose assignment problems."

The next approach considered by NSF consisted of examining a sample of patents from ambiguous subclasses in order to determine the SIC code or codes to which the subclass should be assigned. A sample of patents would be selected from each patent subclass and the companies to which those patents were assigned would be matched to the SIC codes published in Dun & Bradstreet for those companies. Based on the frequency of the SIC codes occurring in the sample, the patents in the subclass would be allotted to one or more of the SIC codes.

The OTAF Effort. At that point NSF requested OTAF's comments or suggestions on the project. After discussions between staffs of the two offices, OTAF submitted a proposal to NSF in March 1974 to establish a "concordance" between the SIC and the USPC. This proposal went through some revisions and resulted in an interagency agreement between NSF and the PTO dated June 14, 1974, and amended July 19, 1974. OTAF's proposal described two reports that would be prepared for NSF. One would "provide detailed and comprehensive concordant relationships between the SIC and the POC [USPC], covering the entire range of technologies and going to the most detailed level of the SIC permitted by available data." That report would use computer matching and statistical techniques, and would be completed with the assistance of patent examiners and other experts within the Patent Office and the Department of Commerce. It would use the Patent Office's data base of 11,000,000 patent documents classified

into 85,000 subclasses; data relating the SIC to specific corporations; data relating patent ownership to specific corporations; and other available data as necessary.

Apparently due to the time pressures associated with publication of Science Indicators 1974, however, OTAF proposed to provide an "initial," "limited" concordance, which would be prepared "manually." The final version of the interagency agreement called only for the preparation of the manual concordance, leaving the possibility open for future development of the detailed, computer-based concordance.

The SIC product fields covered in the original Concordance were developed from lists supplied by NSF of "product fields" and "industries" used in NSF's publication Research and Development in Industry, 1971. These lists are shown in Table 7. The product fields were defined primarily at the 2- and 3-digit SIC level and occasionally the 4-digit level. They were (and, with some revisions, still are) used in NSF's survey to report industrial applied research and development costs by product field. This information is filled in by the firms receiving the survey questionnaire, according to the following instructions supplied by NSF:

> Costs should be entered in the field or product group in which the research and development project was actually carried on regardless of the classification of the field of manufacturing in which the results are to be used. For example, research on an electrical component for a farm machine should be reported as research on electrical machinery. Also, research on refractory bricks to be used by the steel industry should be reported as research on stone, clay, glass, and concrete products rather than primary ferrous metals, whether performed in the steel industry or the stone, clay, glass, and concrete industry. Research and development work on an automotive head lamp would be classified in Group 25, [Other Electrical Machinery Equipment and Supplies (balance of SIC 36)] regardless of whether performed by an automotive or electrical company.

The product field codes differed somewhat from the industry codes, which are based on the SIC assigned by the Bureau of the Census to the primary economic activity of an "establishment." The final list of product fields chosen was drawn primarily from the product field list, with some additions from the industry list, and other changes. The final list of 36 product fields covered in the original concordance is shown in Table 8.

OTAF agreed to develop the following data for each year over the 1963 - 1973 period for each of the product fields for which a USPC-SIC concordance was established:

1)	Total Patents	#	100%
	Originating in the United States	#	%
	Originating in Other Countries	#	%

TABLE 7

List of Product Fields and Industries Supplied to OTAF by NSF

PRODUCT FIELD

Product field	SIC code
Total	
Atomic energy devices[1]	
Ordnance, except guided missiles	19,except192
Guided missiles and spacecraft	192
Food and kindred products	20
Textile mill products	22
Chemicals, except drugs and medicines	28,except 283
Industrial inorganic and organic chemicals	281
Plastics materials and synthetic resins, rubber, and fibers	282
Agricultural chemicals	287
Other chemicals	284-89
Drugs and medicines	283
Petroleum refining and extraction	29,13
Rubber and miscellaneous plastics products	30
Stone, clay, and glass products	32
Primary metals	33
Ferrous metals and products	331-32,3391,3399
Nonferrous metals and products	333-36,3392
Fabricated metal products	34
Machinery	35
Engines and turbines	351
Farm machinery and equipment	352
Construction, mining, and materials handling machinery	353
Metalworking machinery and equipment	354
Office, computing, and accounting machines	357
Other machinery, except electrical	balance of 35
Electrical equipment, except communication	36,except365-67
Electric transmission and distribution equipment[3]	369
Electrical industrial apparatus[4]	362
Other electrical equipment and supplies	363-64,369
Communication equipment and electronic components	365-67
Motor vehicles and other transportation equipment	37,except 372
Motor vehicles and equipment	371
Other transportation equipment	373-75,379
Aircraft and parts	372,19
Professional and scientific instruments	38
Other product fields, not elsewhere classified	

INDUSTRY

Industry	SIC code
Total	
Food and kindred products	20
Textiles and apparel	22,23
Lumber, wood products, and furniture	24,25
Paper and allied products	26
Chemicals and allied products	28
Industrial chemicals	281-82
Drugs and medicines	283
Other chemicals	284-89
Petroleum refining and extraction	29,13
Rubber products	30
Stone, clay, and glass products	32
Primary metals	33
Ferrous metals and products	331-32,3391,3399
Nonferrous metals and products	333-36,3392
Fabricated metal products	34
Machinery	35
Electrical equipment and communication	36,48
Radio and TV receiving equipment	365
Communication equipment and electronic components	366-67,48
Other electrical equipment	361-64,369
Motor vehicles and other transportation equipment	371,373-75,379
Aircraft and missiles	372,19
Professional and scientific instruments	38
Scientific and mechanical measuring instruments	381-82
Optical, surgical, photographic, and other instruments	383-87
Other manufacturing industries	21,27,31,39
Nonmanufacturing industries	07-12,14-17,41-47, 49-67,739,807,891

Source: RESEARCH AND DEVELOPMENT IN INDUSTRY: 1971 (NSF 73-305)

TABLE 8

Product Fields Included in Original 1974 Concordance

Product Field

- Food and Kindred Products
- Textile Mill Products
- Chemicals, Except Drugs and Medicines
  - Basic Industrial Inorganic and Organic Chemicals
  - Plastics Materials and Synthetic Resins
  - Agricultural Chemicals
  - All Other Chemicals
- Drugs and Medicines
- Petroleum and Natural Gas Extraction and Petroleum Refining
- Rubber and Miscellaneous Plastics Products
- Stone, Clay, Glass and Concrete Products
- Primary Metals
  - Primary Ferrous Products
  - Primary and Secondary Non-Ferrous Metals
- Fabricated Metal Products
- Machinery, Except Electrical
  - Engines and Turbines
  - Farm and Garden Machinery and Equipment
  - Construction, Mining and Material Handling Machinery and Equipment
  - Metalworking Machinery and Equipment
  - Office, Computing and Accounting Machines
  - Other Machinery, except Electrical
- Electrical Equipment, Except Communication Equipment
  - Electrical Transmission and Distribution Equipment
  - Electrical Industrial Apparatus
  - Other Electrical Machinery Equipment and Supplies
- Communication Equipment and Electronic Components
  - Radio and Television Receiving Equipment, except Communication Types
  - Electronic Components and Accessories and Communication Equipment
- Motor Vehicles and Other Transportation Equipment, Except Aircraft
  - Motor Vehicles and Motor Vehicle Equipment
  - Guided Missiles and Space Vehicles and Parts
  - Other Transportation Equipment
  - Ordnance, except Missiles
- Aircraft and Parts
- Professional and Scientific Instruments

2) Originating in the United States	#	100%
Owned by U.S. Corporations	#	%
Owned by Individuals	#	%
Owned by U.S. Government	#	%
Foreign-Owned	#	%
3) Originating in Other Countries	#	100%
U.S.-Owned	#	%
Foreign-Owned	#	%

The actual process and decision rules used in establishing the original Concordance are difficult to reconstruct. Most of the work of building the Concordance was apparently done over three or four months in the spring and summer of 1974. Perhaps the key decision made during this phase was to place subclasses in all the product fields to which they pertained. This was done on the theory that the technology disclosed in the patents in those classifications was pertinent to each of the several product fields. (Some of the problems caused by multiple assignment are discussed later in the report.)

The basic tools used in relating patent classifications to SIC product fields were the U.S. Patent Classification Manual updated through June 30, 1973, and the 1972 Standard Industrial Classification Manual. The 15-volume set of definitions of U.S. Patent Classifications was also used in those instances where class and subclass titles were not sufficiently descriptive to permit assignment to product fields.

As reported by OTAF:

> As the first phase of creating the concordance each of the 308 major classes of the USPC was considered. For each class it was determined if the subject matter of the class was sufficiently homogeneous so that it could readily be assigned in whole or in major portion to the product fields relevant to it. More than one-half of the classes were susceptible to assignment in this manner. In those cases where it was determined that such homogeneity did not exist, the classes were put aside for later analysis and assignment of the subclasses making up the class.
>
> The detailed analysis of the non-homogeneous classes involved scanning the subclass blocs within each class to determine the extent to which they were amenable to assignment as a whole to the relevant product fields. If they were not, then each subclass was considered individually until all were assigned.*

---

*Office of Technology Assessment and Forecast, "Indicators of the Patent Output of U.S. Industry," August 21, 1974, p. 3.

According to one team member, most of the assignments were made through a group decision process. The assignments were apparently heavily subjective and not believed to be scientifically reproducible. The nonreproducible nature of this "initial" Concordance was noted in correspondence between OTAF and NSF at that time.

The precise decision rules used in making the assignments in the original Concordance are not known. At least three different versions of the decision rules exist as shown in Table 9. The version of the decision rules that appeared in the Fifth OTAF Report (August 1975) has until recently been regarded as the authoritative statement of the decision rules. The earliest version, which was found in a file with OTAF's working papers from the 1974 effort, is apparently the closest to the decision rules actually used by the team. Subsequent versions incorporated wording changes that result in different assignments.

Once the Concordance was completed it was possible to access the OTAF data base, extract all U.S. patents for the period 1963 through June 30, 1973, and, through the Concordance, classify these patents into the specified product fields. This procedure resulted in the extraction of about 5.1 million patent records. As explained elsewhere the number of patent records exceeds the number of patents actually granted because the Concordance assigns subclasses to more than one product field.* To provide a more accurate picture of patent activity by product field, a procedure was devised to eliminate multiple counting of patents within a single product field. Also, patents appearing in more than one of the component product fields were counted only once in forming combination fields.

This file was then summarized on the basis of year of issue of patents, U.S. or foreign residence of inventor, and type of ownership of patents for each product field. The summary file served as the input to the programs that produced statistical tables and graphs for each product field. Figure 1 diagrams the process flow that produced the original Concordance, from the manual assignment of subclasses to product fields, to the generation of statistical reports on each product field.

First OTAF Concordance Report. The first OTAF report on the Concordance, entitled "Indicators of the Patent Output of U.S. Industry," was delivered to NSF on August 21, 1974. It described the methodology used in developing the Concordance and included statistical reports on the 36 product fields covered. For each product field, a table and series of graphs were developed. The primary differences between the first reports and more recent reports were that the former did not break out foreign-origin patents by individual country and did not show patents distributed by application date.

---

*The level of multiple counting was especially high in the first Concordance reports because they were based on both original classification references (OR's) and cross references (XR's). OR's and XR's are discussed in more detail in the next section.

TABLE 9

VERSIONS OF DECISION RULES USED IN ORIGINAL CONCORDANCE

Version 1 (Source: File of OTAF working papers on 1974 SIC project)

1. Follow P.O. classifier broad characterization (i.e., in class title or outdented subs) as "product", "process" or "apparatus".

2. Place "product" subclass groups in SIC category that best fits that product. Place "apparatus" subclass groups in SIC machinery category (which is based on use). Scan "process" subclass groups and place in SIC category for relevant product or machinery depending on which it appears more closely related to (especially applicable in mechanical areas).

3. Where there is no P.O. characterization of the class or major portions of it, scan each major subclass bloc and assign it to SIC category as in step 2 based upon any characterization that can be made as to "product", "apparatus" or "process".

4. Where substantial doubt exists as to the characterization of a subclass bloc as to "product" or "apparatus" or to "product-or apparatus-related process", place it in both the relevant SIC product and machinery categories.

Version 2 (Source: First SIC Report, "Indicators of the Patent Output of U.S. Industry," August 21, 1974)

Because of the nature of each classification system the basic question to be resolved in relating patent classifications to SIC categories was: "What type of establishment would be engaged in producing the product or apparatus encompassing the structural or functional features represented by that patent classification or in carrying out the process steps included in the patent classification?" In those cases where this question could not be clearly answered and ambiguity existed as to which of several possible SIC groups to assign a patent classification, it was placed in all of the possible groups.

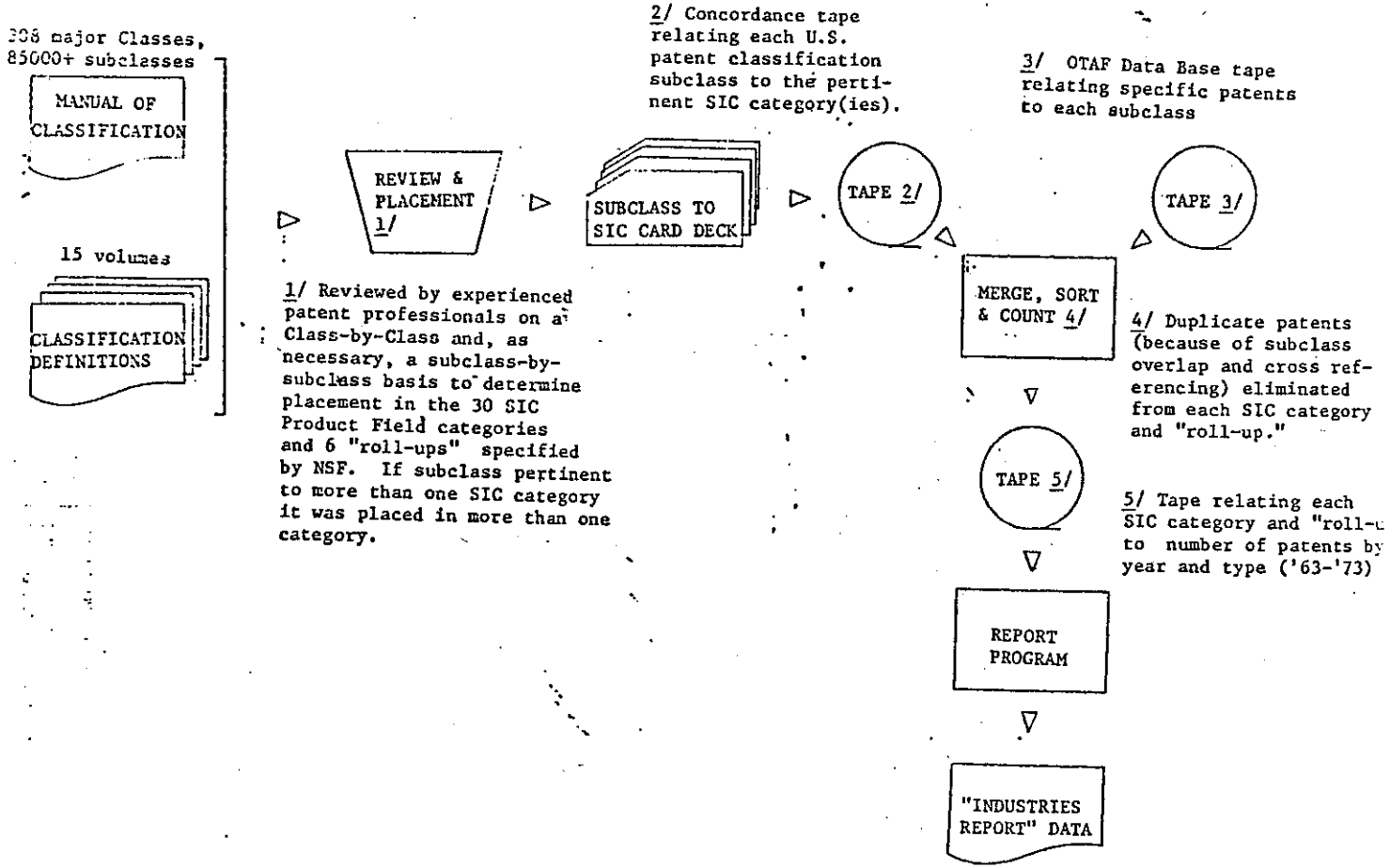
Version 3 (Source: Fifth Technology Assessment and Forecast Report, August 1975)

Because of the nature of each classification system, the basic questions asked in relating patent classification to SIC categories were: "What type of establishment would be engaged in producing the products or apparatus having the structural or functional features encompassed by that patent classification? or "What type of establishment, in producing products or apparatus, would be carrying out the process steps included in the patent classification?"

In those cases where these questions could not be clearly answered and ambiguity existed as to which of several possible SIC groups to assign a patent classification, it was placed in all of the possible groups.

FIGURE 1

PROCESS FLOW DIAGRAM FOR ORIGINAL CONCORDANCE



In transmitting the report to NSF, the OTAF director wrote "I believe that you share with me the understanding that in the project just completed we have left the starting point and made it past the first hurdle but that we still have a long way to go to reach our goal of an accurate, reliable detailed concordance." The report contained suggestions for developing such a detailed SIC-USPC concordance.

Excerpts from that suggestion follow:

> While the concordance between the SIC and USPC developed under this contract is adequate for the presentation of patent data for large industrial groupings, it must be recognized that it is only a gross concordance. Because of this, it is not possible to take advantage of the extremely fine detail of patent data. When over 85,000 technological categories are assigned to only thirty different industry groupings [excluding the combination fields], of necessity, much is lost. This need not be the case however. The Standard Industrial Classification system also contains very specific industry categories. These several thousand four-digit industry classifications can be used as the units to which patent classifications are related....
>
> ....Obviously it would not be feasible to develop a detailed SIC-USPC concordance using the same manual procedures utilized in building the current concordance. To assure a satisfactory level of accuracy, a better "handle" than a patent classification should be assigned. Fortunately such a "handle" does exist in Patent Office data files which provide the names of patent assignees. The Office of Technology Assessment and Forecast has done preliminary design work on a system to build a detailed concordance by relating the four-digit SIC codes assigned to corporations owning patents with the classifications of the patents they own. Such a procedure would harmonize well with the establishment based nature of the SIC system. It would relate patent classifications to SIC categories on the basis of the SIC codes associated with the establishments which are in fact performing work in those patent classifications.
>
> In addition to greatly facilitating the relating of patent and economic data, this detailed concordance would also represent a quantum advance in terms of accuracy. This enhanced accuracy would be attained as a result of the use of the SIC codes associated with corporate patent owners as the basis for associating patent classifications with industries. Thus even when rolled up to gross industry levels, the new concordance would represent a major improvement over the one created under this contract.

A feasibility study of a detailed concordance was undertaken in the third OTAF report on the Concordance.

The second report, entitled "Indicators of the Patent Output of U.S. Industry, II," was dated July 1976. It updated the figures to cover the period 1963 - 1975 and provided data on patent activity in 52 product

fields, including 16 new product fields requested by NSF. Table 10 shows the product fields covered in the second report; the 16 new fields are checked. The second report included a report based on OR data only, as well as one based on OR and XR data together. These figures are also shown in Table 10.

Feasibility Study of the "Detailed Concordance"

As early as January 1976, OTAF and NSF were negotiating the development of a "detailed" SIC-USPC Concordance. The project was attrative because it promised to provide "a detailed reproducible, reliable, and flexible concordance between the two systems [SIC and USPC]." For the Science Indicators Unit it represented a sophisticated basis for the patent indicators in the Science Indicators report and would allow much more analysis. It was also believed to be capable of providing "the kind of concordance that many economists and patent investigators have desired for over a decade."

An interagency agreement was signed November 1, 1976. The resulting report, entitled "Establishment of a Detailed Concordance between the Patent Office Classification and the Standard Industrial Classification Systems: Phase 1, Feasibility Test Final Report," was issued in December 1977. This effort is particularly interesting because it bears some similarity to the approach suggested earlier by the NSF of using information on patent assignees to determine SIC assignment. Reliance on information about assignees was also a feature of Scherer's recent work and has been suggested as an alternative to the present Concordance.

The feasibility study of the "detailed" concordance was undertaken specifically "to overcome the lack of scientific reproducibility of the manual concordance and to achieve a greater breadth and depth of detail...." The technique studied was that of "computer mapping." The general approach was to:

- code the companies in OTAF's patent files with Dun and Bradstreet (D&B) codes to enable access to D&B SIC data relating to company activities (these data were at the establishment, or plant, level):
- construct technologically homogenous groupings of patents by reviewing the USPC and identifying hierarchical groupings of subclasses which clearly pertained to a single, dominant 4-digit SIC category.
- sort, run, and rank by most frequently occurring, the SIC's assigned by D&B to the companies owning patents in a grouping;
- use this ranking as representative of the proportional distribution, by SIC, of the patent activity in the grouping -- in other words, as the USPC/SIC concordance between the grouping and the SIC.

TABLE 10
TOTAL PATENTING TABLE
FROM SECOND CONCORDANCE REPORT*

SIC CATEGORY	1963-1975 PATENTS	
	OR's ONLY	OR's & XR's
Food and Kindred Products	9532	13172
Textile Mill Products	9181	23249
Chemicals, Except Drugs & Medicines	129243	167884
Basic Industrial Inorganic & Organic Chemistry	76792	107238
✓Industrial Inorganic Chemistry	16458	32557
✓Industrial Organic Chemistry	66374	94073
Plastics Materials & Synthetic Resins	31114	53196
Agricultural Chemicals	12905	34140
All Other Chemicals	11560	32675
✓Soap, Detergents, and Cleaning Preparation, Perfumes, Cosmetics & Other Toilet Preparations	3562	9069
Paints Varnishes, Lacquers, ✓Enamels, and Allied Products	774	3876
✓Miscellaneous Chemical Products	8513	25983
Drugs and Medicines	16767	38969
Petroleum & Natural Gas Extraction & Petroleum Refining	11572	21327
Rubber & Miscellaneous Plastics Products	42825	85630
Stone, Clay, Glass and Concrete Products	19244	43253
Primary Metals	10980	20934
Primary Ferrous Products	7865	16650
Primary & Secondary Non-Ferrous Metals	6317	12521
Fabricated Metal Products	110955	193673
Machinery, Except Electrical	273237	363325
Engines & Turbines	18902	27933
Farm & Garden Machinery & Equipment	26368	38892
Construction, Mining & Material Handling Machinery & Equipment	46041	69176
Metal Working Machinery & Equipment	25653	41160
Office Computing & Accounting Machines	31507	49832
Other Machinery, Except Electrical	170893	252904
✓Special Industry Machinery, Except Metal Working Machinery	69786	114056
✓General Industrial Machinery & Equipment	86988	145961
✓Refrigeration & Service Industry Machinery	21660	37802
✓Miscellaneous Machinery, Except Electrical	13671	30189

---

*Checks indicate product fields included for the first time in the second Concordance report.

TABLE 10 (continued)
TOTAL PATENTING TABLE
FROM SECOND CONCORDANCE REPORT

SIC CATEGORY	1963-1975 PATENTS OR's ONLY	OR's & XR's
Electrical Equipment, Except Communication Equipment	95734	146583
Electrical Transmission & Distribution Equipment	30376	54675
Electrical Industrial Apparatus	30413	51349
Other Electrical Machinery, Equipment & Supplies	47330	81136
✓Household Appliances	15835	28151
✓Electrical Lighting & Wiring Equipment	12411	26401
✓Miscellaneous Electrical Machinery, Equipment & Supplies	18939	33465
Communication Equipment & Electronic Components	101182	138765
Radio & Television Receiving Equipment, Except Communication Types	16083	28505
Electronic Components & Accessories & Communication Equipment	99716	136836
Motor Vehicles & Other Transportation Equipment, Except Aircraft	50533	73077
Motor Vehicles & Motor Vehicle Equipment	27287	43630
Guided Missiles & Space Vehicles & Parts	5946	9357
Other Transportation Equipment	18317	29115
✓Ship & Boat Building & Repairing	5659	9847
✓Railroad Equipment	9851	16785
✓Motorcycles, Bicycles & Parts	1731	3862
✓Miscellaneous Transportation Equipment	9761	16633
Ordinance Except Missiles	5970	8501
Aircraft & Parts	15077	23569
Professional & Scientific Instruments	90777	136418
All Other		

OTAF had its file of about 50,000 patent-owning companies coded by D&B, a process that resulted in only a 55% "match" rate between the OTAF file and the D&B file. The reason hypothesized for the low match rate was the great number of very small companies involved in U.S. patenting, companies so small or so short-lived that they were not captured on the D&B data base.

The consequent decrease in company SIC data had a marked effect on the results of the feasibility test. Tests were done both using all six D&B-assigned SIC's and using only the first, dominant D&B-assigned SIC. Correct SIC's (i.e., matches with the 4-digit SIC manually identified by PTO) were obtained in 50% or less of the cases. Further, there was little or no consistency in the pattern of results.

OTAF concluded that the SIC codes assigned by D&B to companies did not correspond to the SIC codes descriptive of the technology disclosed in the patents those companies own. Suggested reasons for this were: (1) in limiting to six the number of SIC codes assigned to companies, D&B ignores many activities of the company; and (2) companies do R&D and obtain patents in areas outside their commercial activities, for example in branching into new areas.

On the basis of the feasibility study, OTAF concluded that further work was unlikely to be fruitful and should not be undertaken.

The Update Process

Since the feasibility study of the "detailed" concordance, the primary concordance effort has been a biennial update of the original,"manual" Concordance. Three "roll-up" product fields have been added -- "Chemicals and Allied Products," "Electrical and Electronic Machinery, Equipment and Supplies," and "Transportation Equipment" -- making the 55 product fields currently covered.

The updates are required to keep up with the continuous expansion and reclassification of the patent file. While it is impossible to estimate with precision the proportion of the Concordance that is "new" -- i.e., has been redone since the original Concordance -- it should be similar to the proportion of the total patent file that has been reclassified over the same period of time. Between 1973 and 1981 about 28% of the OR's in the patent file were reclassified.

In a reclassification project the PTO takes all the patents in an existing subclass and puts them in new subclasses that did not exist before. Typically, a class or subclass is reclassified because it has gotten too large to be searched effectively by examiners. Reclassification may also occur when technologies converge or diverge. Reclassification changes patent data for all years, because all patents in the old subclasses are reassigned. To fit the new subclasses into the Concordance, OTAF assigns them to SIC product fields.

The update process can result in inconsistencies in data for two reasons. First, USPC reclassification can cause shifts in relationships within the SIC system. This may occur, for example, when patents are moved into subclasses with different product field assignments than the old.

Second, the subjective process of assigning new subclasses to SIC product fields can yield decisions inconsistent with those made in initial Concordance development. The updates have not traditionally received the same concerted effort by a team of coders as the original Concordance, but rather have been the responsibility of one person who generally has had no previous experience with the Concordance. Moreover, the updates have been done on a relatively tight schedule. The schedule is constrained at one end by the time at which OTAF receives the updated patent data from PTO and at the other end by the deadline for providing data for Science Indicators. In the past, OTAF has waited until it received the updated patent data from the PTO to begin getting the new subclasses from the Reclassification Orders. Then the basic OTAF file must be updated and the new subclasses assigned to product fields. These conditions, combined with a lack of detailed, step-by-step decision rules, have created a situation where error is likely to occur.

This brief history of the Concordance provides background for the following sections of the report, which describe OTAF's analysis of the 1981 Concordance.

Analysis of Patterns Based on OR's Only vs. OR's Plus XR's

U.S. patents are categorized or classified into the USPC system by the patent examiner at the time he or she makes the decision to issue the patent. In classifying the patent, the examiner places a copy in all subclasses which pertain to the disclosed and/or claimed subject matter of the application. If, to use a hypothetical example, a patent recites a chair invention and a table invention, the examiner places a copy in the chair subclass and the table subclass.

For each patent with multiple classifications, one and only one may be designated as the original classification (OR). All additional copies placed in the file are designated as cross-reference classifications (XR's). The rules which determine which of the multiple classifications is designated as the OR are often esoteric. They can depend on the wording of the claims of the patent or the particulars of the USPC system, and, for the purpose of this study, are unimportant. The important point here is, that if one looks at the OR only there is a chance that some pertinent technology revealed in the patent will not be counted.

In the example of the chair/table patent, assume that the chair subclass is the OR and the table subclass is the XR. If you want to count all patents which contain table technology and you consider only OR's in the table subclass, you will not count this patent since the OR is in the chair subclass.

If, on the other hand, you want to count all patents which contain "furniture" technology, you may do this by counting all patents in the chair subclass and the table subclass. If only OR's are considered, the chair/table patent will be counted as an OR in the chair subclass. If OR's plus XR's are considered the patent is counted in the chair subclass and the table subclass, but computer techniques can be used to ensure that it is only counted once. The chair/table patent is counted as one "furniture" patent in either case--if only OR's are considered or if OR's plus XR's are considered. At this higher level of aggregation, the effect of the OR vs. OR plus XR distinction is diminished. At the ultimate level of aggregation-- all patents in the data base--the results of considering OR's as opposed to OR's plus XR's are the same as long as duplicate counts are eliminated. Even at the high levels of aggregation which occur in the 55 SIC product fields of the Concordance, however, some patent counts pertinent to a particular field are missed when only OR's are considered.

Against this disadvantage is the advantage of reducing the amount of multiple counting (and error*) which results when subclasses are assigned to more than one SIC product field. If the OR classification of a patent

---

*OTAF found a strong relationship between multiple assignments and error level, as discussed later in the "Analysis of Error in Subclass Assignment."

is assigned to two SIC product fields and the XR to two additional product fields, the patent will be counted four times if OR's and XR's are considered, but only twice if OR's only are considered.

The first report produced from the OTAF Concordance in 1974 considered both the OR and XR classifications of each patent. Because this procedure resulted in an unacceptably high level of multiple counting, the second (1975) and third (1977) updates considered the OR only classification, as well as OR's plus XR's. Subsequent updates have been based on OR's only.

The issue of whether OR's only or OR's plus XR's should be used in counting patents needs to be investigated further. In part, it is a theoretical question of the definition of technological change and which method -- patent counts based on OR's only or OR's plus XR's -- is a better measure of technological change. In part, it is a practical question of additional assignment error introduced by using XR's.

It should be pointed out that the results obtained from the two methods differ, but we do not know precisely how they differ. Two hypothetical examples are given.

FIGURE 2

Hypothetical Example of Patent Activity Time Series in a Single Product Field Using OR's only and OR's plus XR's

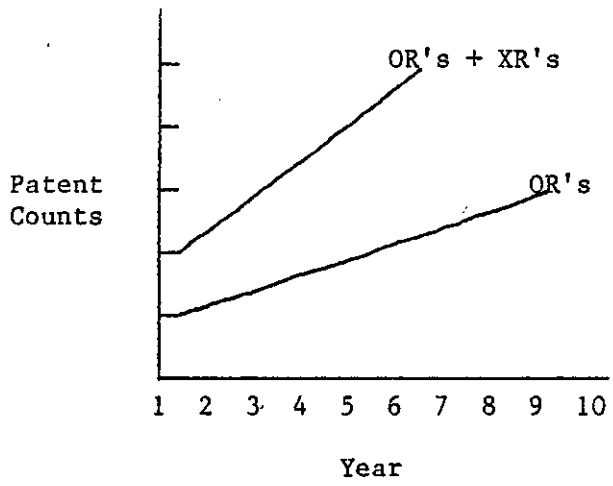
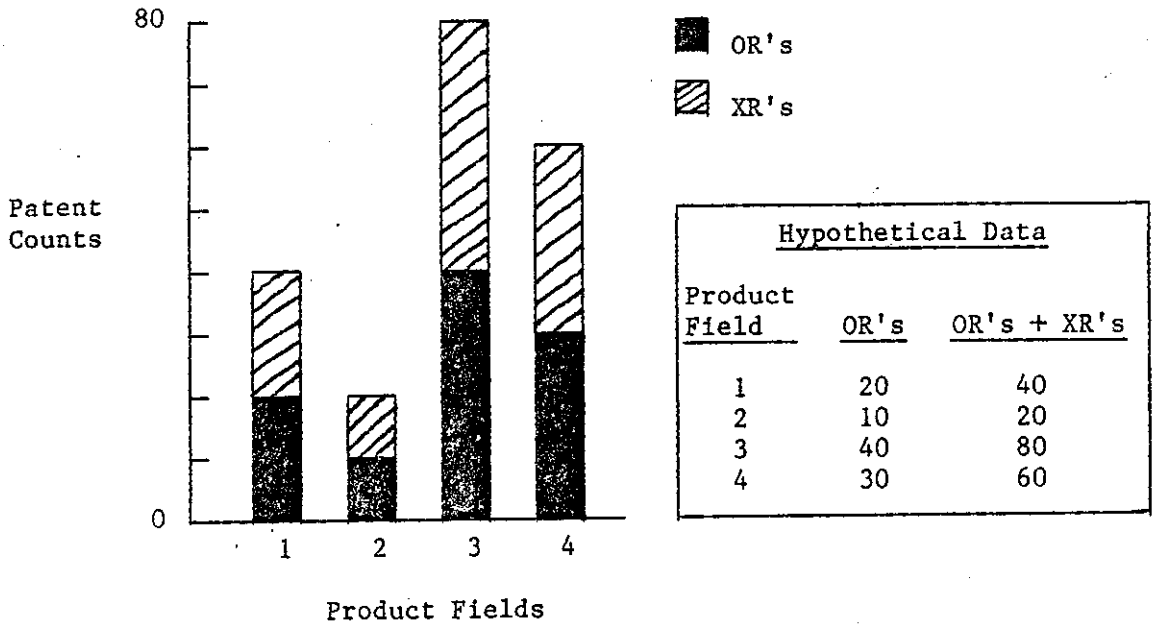


Figure 2 presents a hypothetical example of a patent activity time series in a single product field, using both OR's and OR's plus XR's. This corresponds to asking the rate of technological change in a product field over time. The relationship of the number of XR's to the number of OR's is unknown. But if one assumes that the number of XR's varies proportionately with the number of OR's, as in Figure 2, then it is obvious that the slope, or rate of change, of OR's plus XR's will always

be greater than that of OR's only, but the magnitude of the difference will be unknown. Therefore, estimates of the rate of technological change based on OR's only will be lower by some unknown amount than those based on OR's plus XR's.

FIGURE 3

Hypothetical Example of Patent Activity in Four Product Fields in One Year, Using OR's only and OR's plus XR's



Hypothetical Data

Product Field	OR's	OR's + XR's
1	20	40
2	10	20
3	40	80
4	30	60

Figure 3 presents a hypothetical example of patent activity in four product fields in a single year, using OR's only and OR's plus XR's. The figure has been constructed on the assumption that XR's = OR's; that is, as OR's increase (decrease), XR's increase (decrease) by an equal amount. Therefore, if OR's increase by a factor of 2, OR's plus XR's also increase by a factor of 2. (See Hypothetical Data for Figure 3.)

It is obvious from Figure 3 that the choice of OR's rather than OR's plus XR's reduces the magnitude of differences among the product fields. In our hypothetical example it does not affect the ratio of the differences, which is 1 to 2, but that is because of our assumption that XR's = OR's. Since the precise relationship of XR's to OR's is not known, the effect of choosing OR's only on the ratio of the differences is unknown. In fact, the effect may vary if the relationship between XR's and OR's varies across product fields. Hence, estimates of differences in the level of technological activity in different product fields will be reduced by some unknown amount if based on OR's only rather than if based on OR's plus XR's.

Decision Rules and Process

Decision Rules

The USPC subclasses are assigned to SIC product fields based on certain "decision rules." These decision rules assign subclasses based on the industry of manufacture and allow subclasses to be assigned to more than one industry.

Table 9, shown earlier, exhibits three different versions of the decision rules. They differ in two major respects -- how to assign process subclasses and what to do in ambiguous cases. Since there is no evidence of a conscious effort to alter the decision rules, it appears that the different versions evolved from wording changes which were intended to clarify the rules and not to affect their substance. Because assignments may differ depending on which version is followed, however, the decision rules must be interpreted and used consistently if the correct assignment of subclasses is to be determined.

As noted previously, Version 3, which appeared in the Fifth OTAF Report, was previously regarded as the authoritative statement of the decision rules. However, in the course of this project OTAF found that the original assignments of some subclasses had been made by following Version 1 decision rules.

This section of the review and assessment discusses the decision rules, how they differ and how they are used when determining the "correct" subclass assignment. Only after the correct subclass assignment has been determined, can the level of error in the Concordance's assignments be studied.

Versions 2 and 3

Versions 2 and 3 of the decision rules are nearly identical; Version 3 appears to be a rewording of Version 2. In both versions, the rules state that if an ambiguity exists as to which of several possible SIC groups to assign a patent classification, it is placed in all of the possible groups. Both versions ask the question, "What type of establishment would be engaged in producing the product or apparatus ... in the patent classification?" For patent classifications (subclasses) which include process steps, they ask what type of establishment would carry out the steps. In Version 2, the question reads, "What type of establishment would be engaged ... in carrying out the process steps included in the patent classification?" Version 3 reads, "What type of establishment, in producing products or apparatus, would be carrying out the process steps included in the patent classification?" (Emphasis added.) It may be that the phrase "in producing products or apparatus" was inserted to clarify the "process" situation, possibly to limit the establishments to manufacturing establishments. On its face, this difference between Versions 2 and 3 appears insignificant. When Version 3 is compared to Version 1, however, the difference may have some meaning.

Version 1

Version 1 is very different from the later versions. While it provides for multiple assignments, these are not the same as those provided for in the "ambiguity rule" of Versions 2 and 3. Version 1 first determines if a USPC is product, apparatus or process. If a product, the rule says to place it in the SIC that best fits that product. It does not say what to do if the "best fit" cannot be determined. It places apparatus in "SIC machinery category" and again says nothing about ambiguous cases. Process subclasses are placed with the product or with the apparatus "depending on which it appears to be more closely related to" -- a determination that may be difficult to make in practice.

Multiple assignments are provided for when doubt exists as to whether the subclass is a product or apparatus or whether the process is more closely related to process or apparatus. In such cases the subclass is placed in "both the relevant SIC product and machinery categories."

Although Version 1 appears to state a "best fit" rule, i.e., one where ambiguity is resolved in favor of the single SIC that is most appropriate, there is no evidence that such a rule was ever followed. The "ambiguity rule" which provides for multiple assignments is clearly stated in Version 2 which accompanied the First SIC Report of 1974. The history of the Concordance and examination of original assignments indicate that the ambiguity rule has generally been followed, although in practice the number of assignments was limited to a "reasonable" number.

"Process" subclasses include methods of doing, methods of making and methods of using. They create a problem since they do not relate to a manufactured product or machine and the OTAF Concordance relates subclasses to the industry of manufacture. Version 1 places process subclasses with the relevant product or machinery depending on which it is more related to or with both if substantial doubt exists. This approach makes sense and it appears to be the one followed by the Concordance developers. Thus, a method of sorting bottles using a specific apparatus goes with the sorting apparatus. The decision rules of Versions 2 and 3 ask the assignor to speculate what establishment would carry out the process of sorting bottles. Thus, the question becomes, "who uses the sorting apparatus?" The answer is not clear. It could be used by bottle makers, bottle fillers, bottle packagers or service industries.

To determine the correct assignment of subclasses for error analysis, OTAF essentially used Version 3 of the decision rules except in instances of process-type patent classifications. In such cases, Version 1 was used. In addition, OTAF limited the number of unique SIC assignments to six.

The resulting decision rule is shown on in Table 11. OTAF suggested the "New Decision Rule" for future use and comments on the rule were solicited at the Workshop.

TABLE 11

NEW DECISION RULE

1. Determine if the subclass is characterized as "product," "apparatus," and/or "process."

2. For products, place the subclass by determining "What type of establishment would be engaged in producing the products having the structural or functional features represented by that patent classification?"

3. For apparatus, place the subclass by determining "What type of establishment would be engaged in producing the apparatus having the structural or functional features represented by that patent classification?"

4. If the subclass is a process, determine whether the process is more closely related to the product of that process or the apparatus used in the process. If more closely related to the product, place in the relevant SIC product code. If more closely related to the apparatus, place in the relevant SIC machinery code. Where substantial doubt exists as to which it is more closely related to, place it in both the relevant SIC product and apparatus categories.

5. In cases where these questions cannot be answered clearly and ambiguity exists as to which of several SIC categories to assign a subclass, it is placed in all possible groups, the maximum being six unique SIC categories.

Decision Process

In the assignment of USPC subclasses to SIC product fields, the ability to make the correct assignment depends on a clear understanding and proper use of the decision rules. Before discussing error analysis, we discuss the procedure used to determine the correct assignment and some of the problems which we encountered.

The following steps were used to determine the correct SIC assignment:

1) Consider the subclass title in its full context to determine the subject matter of the patents in the subclass. (This involves reading and interpreting the Patent Classification Manual.)

2) If the subject matter cannot be identified from the title, look at the subclass definition in the U.S. Patent Classification Definitions.

3) If the subject matter is still in doubt, look at specific patents within the subclass and/or talk to the patent examiner or classifier who handles the subclass.

4) Following the decision rules and using the SIC Manual (and in some cases the 1977 "Numerical List of Manufactured Products") determine the correct product field(s).

To illustrate some of the problems OTAF encountered using this process, specific examples of class/subclasses are given with their titles:

81/129	: A wrench with slidable jaw adjustments.
222/503	: A dispensing apparatus with interconnected, relatively movable sectional flow controllers or closures.
192/57	: Fluent material and mechanical clutches.
340/870.05	: Continuously variable, calibrated electrical indicating device, e.g., telemetering.
29/1.11	: An ordnance-making process or apparatus.

For USPC 81/129, the subject matter appears clear from the subclass title. Applying the decision rules, we ask, "What type of establishment would be engaged in producing wrenches?" Consulting the SIC Manual, we find wrenches in SIC 3423, provided they are hand tools and not power driven. In this case, it is necessary to look at the other subclasses in USPC class 81 to determine if subclass 129 can include both hand and power tools. Finding that other subclasses specifically provide for machine tools and power tools, we conclude that SIC 3423, in product field 21, is the correct assignment.

USPC 222/503 is an example of a subclass which OTAF was unable to assign correctly. The Concordance assigned it to SIC 352, "Farm and Garden Machinery and Equipment;" SIC 353, "Construction, Mining and Material Handing Machinery and Equipment;" and SIC 356, "General Industrial Machinery and Equipment." OTAF examined both the subclass definition and selected patents included in the subclass. Of the five patents granted in this subclass since 1971, the abstracts disclosed them to be for dispensing scrap metal, particulate matter, a product within a bin, grain and molten metal. OTAF concluded that this subclass should be assigned to product fields within the "Machinery, Except Electrical" grouping, SIC 35, but at the disaggregated, 3-digit SIC level, the correct assignments could not be determined. While "Farm and Garden Machinery & Equipment" is correct, it is not clear what other product fields should also be included since the machinery of this subclass is generic to many industries.

USPC 192/57 relates to fluent material and mechanical clutches. Here the subclass relates to a machine part, but does not specify the machine. Nonvehicle clutches are placed in SIC 3568. For vehicle clutches, placement depends on the type of vehicle. In addition to product field 30 (SIC 3568), this subclass is properly assigned to nearly all product fields indented under "Transportation Equipment." The Concordance assigned it to SIC 356, "General Industrial Machinery and Equipment;" SIC 371, "Motor Vehicles and Motor Vehicle Equipment;" SIC 374, "Railroad Equipment;" and SIC 379, "Miscellaneous Transportation Equipment."

USPC 340/870.05 presents a situation where the SIC Manual is unclear. Electronic telemetering equipment goes in SIC 3662, while industrial process type telemetering instruments go in SIC 3823. Here, correct placement depends on an interpretation of the SIC categories. In this case, OTAF reasoned that SIC 3823 includes devices which measure and control and thus does not include the subclass subject matter which measures only. The Concordance assigned it to SIC 366-367, "Electronic Components and Accessories and Communication Equipment."

USPC 29/1.11 relates to metal working, specifically an ordnance-making method or apparatus. Proper placement of the apparatus was considered to be in SIC 355, product field 29, "Special Industry Machinery, Not Elsewhere Classified." Placement of the method depends on whether it is more closely related to the product or to the apparatus. In this instance there is uncertainty, so in addition to SIC 355 for the apparatus, it is also placed in SIC 348, "Ordnance and Accessories" for the product. If methods were not included in the subclass, the coder would have to resist the temptation to place the (apparatus) subclass with the product oriented SIC -- a fairly common error. The Concordance assigned this subclass to SIC 348 only.

In some cases, the SIC Manual does not specifically disclose the part specified in the USPC. In such cases, the part was assigned based on the final product. For example, fan blades were assigned to the product field which includes fans.

Analysis of Error in Subclass Assignment

This section addresses the level of error in the Concordance's assignments of USPC subclasses to SIC product fields.

Causes of Error

Assignment errors may be the result of one or more of the factors listed below:

- Coding errors
- Inadequate knowledge of the USPC or SIC
- Misunderstanding of decision rules
- Ambiguity of the decision rules
- Lack of compatibility between the SIC and USPC

Some coding errors can be identified using computer techniques. One such method, described below as "Computer-Identified Coding Error" identifies subclasses which are assigned only to roll-up product fields. Other types of coding errors occur because of the overlap between the product field numbers and the SIC codes. For example, product field 38 is "Household Appliances" and SIC 38 is "Professional and Scientific Instruments." While the extent of this type of error was not determined, OTAF instituted a new numbering system for product fields to prevent confusion between product field numbers and SIC codes.

Errors also result due to the assigner's lack of knowledge of the USPC and SIC systems. Coders are not experts in all technical areas of the USPC. Subclass titles and/or definitions often do not give a clear picture of the subclass contents. Likewise SIC category descriptions are not all-inclusive nor always clearly delineated.

In many cases, errors result because of ambiguities which exist in the decision rules, or ambiguities which arise because of the difference in the structure and purpose of the USPC and SIC systems. In instances where there is little compatibility between the USPC and SIC systems, the "correct" assignment is difficult to determine. Error rates are high in areas where the USPC fails to specify an industry or product, or where the SIC is silent about the specific subject matter of the subclass. The examples in the discussion of "Decision Process" illustrated how easily error may result from these factors.

OTAF conducted two error checks. The first looked at computer-identified coding errors. The second examined a random sample of subclasses within the Concordance in order to estimate the overall amount of error. Both analyses used data based on OR's only. These projects are discussed below.

Computer-identified Coding Error

Fourteen of the 55 SIC product field categories are "roll-ups" or combinations of the product fields indented under them. A USPC subclass is properly assigned to one of these roll-up product field only when it

is assigned to one or more of its less aggregated components. If a subclass is assigned to a roll-up only, without assignment to a component product field, the assignment is a clear error.

A computer check of the data base identified 343 subclasses, less than 0.4% of the total, which were incorrectly assigned to a roll-up only.

Random Sample Study

This study consisted of comparing the Concordance assignments for a sample of USPC subclasses to the correct assignments as determined by a team of three OTAF analysts. A stratified random sample of 112 subclasses was generated which had an equal distribution of subclasses with one, two, three and four SIC product field assignments. The sampling strategy was intended to take advantage of the expectation that the error rate would be related to the number of assignments per subclass. This expectation was based on a preliminary examination of all the subclasses with four, five, or six assignments. All subclasses with five or six assignments have assignment error. Of those with four assignments, nearly all have assignment errors.

Table 12 shows the distribution of active USPC subclasses according to the number of unique SIC product fields to which they are assigned. Active subclasses are those which received at least one patent as an original classification during the time period 1963-1981. The number of active subclasses is 83,438. Unique SIC product fields are those which are not roll-ups of less aggregated, finer divisions. Forty-one of the 55 SIC product fields are unique. The "All Other" category, product field 99, is also included in the counts of Table 12.

TABLE 12

DISTRIBUTION OF ACTIVE SUBCLASSES BY NUMBER OF UNIQUE SIC PRODUCT FIELDS TO WHICH THEY ARE ASSIGNED

Number of unique SIC assignments	1	2	3	4	5	6
Number of active subclasses	59,745	14,985	6,566	1,835	284	23
Percent of active subclasses	71.6	18.0	7.9	2.2	0.3	< 0.1

Subclasses with five and six assignments were not included in the sample since they all were known to have errors. Representing less than 0.5% of the total, they have little effect on overall error rate. Sampling equally from subclasses with one, two, three, or four assignments produced a sample which was heavily weighted towards subclasses with

multiple assignments. That is, 28 of the 59,745 subclasses having one SIC assignment were selected, while 28 of the 1,835 subclasses having four assignments were selected.

The selected subclasses were given SIC product field assignments, de novo, by three OTAF analysts. The OTAF staff then compared the results to the assignments of the Concordance, and decided through group discussion on the "correct assignment." In many cases, the decision as to what was the "correct assignment" was difficult. For two subclasses, the correct assignment could not be determined.

The Concordance assignments of the subclasses were given one of three designations:

C = Correct. The Concordance assignment was exactly the same as the correct assignment. It contained no errors of inclusion or exclusion.

I = Inclusion Error. The Concordance assignment included one or more product fields not included in the correct assignment. If it included errors of inclusion and errors of exclusion, it was designated I.

E = Exclusion Error. The Concordance assignment contained no inclusion errors, but excluded one or more product fields contained in the correct assignment.

Using this method, a subclass assignment was correct only if it was completely correct -- the strictest definition possible. Note also that C+I+E is equal to 110, the total number subclasses for which the correct assignment was determined. The results of this analysis are shown in Table 13.

Table 13 shows that, based on this study of a sample of 110 subclasses, the overall percentage of correct subclass assignments in the Concordance was 67.5%. Conversely, the overall error rate was 32.5%. As previously noted, this analysis applies the most rigid standard in the tabulation of correct assignments since subclasses which were partly correct were not counted as correct. It should be noted that this is not the error in the assignment of individual patents; at the level of individual patents, it is possible that there may be compensating errors. Also, this analysis does not account for the number of patents in the subclasses.

TABLE 13

SUBCLASS ERROR ANALYSIS

Number of SIC Assignments	1	2	3	4
Percent of Total Subclasses	71.6	18.0	7.9	2.2
Number of Sample Subclasses	28	28	26*	28
Percent of Sample Subclasses	25.4	25.4	23.6	25.4
Weighting Factor**	2.8	0.7	0.3	0.1
Number Correct, Actual	23	12	4	2
Number Correct, Weighted	64.4	8.4	1.2	0.2

Total Correct, Weighted = 74.2
Percent Correct, Weighted = 67.5%

One would expect the error rate to be lower for assignments to product fields at more aggregated levels. Therefore, the error rate of assignments to product fields at the 2-digit SIC level was checked. Assignments were tabulated as correct at the 2-digit SIC level if placement within the 13 major 2-digit product fields was entirely correct. The results are shown in Table 14.

As expected, the percentage correct increased to 81.2% when only assignments at the 2-digit product field level were considered.

OTAF further analyzed the results of this study to answer the following questions:

- Does the error rate in fact vary by the number of assignments?
- Does the error rate vary by SIC product field?

---

*Two subclasses could not be assigned.
**Weighting factor = Percent of Total Subclasses/Percent of Sample Subclasses

TABLE 14

SUBCLASS ERROR ANALYSIS AT THE 2-DIGIT PRODUCT FIELD LEVEL

Number of SIC assignments	1	2	3	4
Percent of Total Subclasses	71.6	18.0	7.9	2.2
Number of Sample Subclasses	28	28	28**	28
Percent of Sample Subclasses	25	25	25	25
Weighting Factor**	2.9	0.7	0.3	0.1
Number Correct, Actual	26	15	13	12
Number Correct, Weighted	75.4	10.5	3.9	1.2

Total Correct, Weighted = 91
Percent Correct, Weighted = 81.2%

Relationship Between Error Rate and Number of Assignments

There is a strong, unidirectional, inverse relationship between number of assignments and the percent of subclasses correctly assigned. (See Table 15.) Of subclasses with only one assignment, 82.1% were correctly assigned -- that is, had no errors of any kind. Of subclasses with 4 assignments, only 7.1% were correctly assigned.

There is a strong positive relationship between number of assignments and percent of subclasses with errors of inclusion. Of subclasses with one assignment, only 10.7% had errors of inclusion, while 78.6% of subclasses with four assignments had such errors. The relationship between number of assignments and percent of subclasses with errors of exclusion is more erratic.

The relationship between number of assignments and error rate, as tested by chi square, is statistically significant at the .001 level.

---

*Weighting factor = Percent of Total Subclasses/Percent of Sample Subclasses

**At the 2-digit level, correct assignment could be determined for all 112 subclasses.

TABLE 15

PERCENTAGE DISTRIBUTION OF TYPE OF SUBCLASS ERROR BY NUMBER OF PRODUCT FIELD ASSIGNMENTS

TYPE OF ERROR*	NUMBER OF PRODUCT FIELD ASSIGNMENTS			
	1	2	3**	4
C	82.1%	42.9%	15.4%	7.1%
I	10.7	53.6	80.8	78.6
E	7.1	3.6	3.8	14.3
TOTAL	99.9%	100.1%	100.0%	100.0%

Chi square = 44.189

Relationship Between Error Rate and SIC Product Field

In this analysis the units of analysis were the individual product fields to which a subclass is assigned, whether that assignment was correct, incorrectly included, or incorrectly excluded. The "N" for this analysis was thus the total number of product field assignments made to the 110 subclasses, or 274. For instance, the following table shows the tabulated data for product field 16.

	#C (%C)	#I (%I)	#E (%E)
Product Field 16	11 (85%)	2 (15%)	4 (27%)

The frequencies indicate that the Concordance assigned 11 subclasses correctly to product field 16; it assigned 2 subclasses to product field 16 that should not have been so assigned; and it did not assign 4 subclasses to product field 16 that should have been so assigned. The percentages are calculated and interpreted as follows:

$$\%C = \frac{\#C}{\#C+I} = \frac{11}{13} = 85\%$$

In 85% of the cases in which the Concordance assigned a subclass to product field 16, the assignment was correct.

---

*The types of error are defined in the text.
**Two subclasses out of the sample of 112 could not be assigned.

$$\%I = \frac{\#I}{\#C+I} = \frac{2}{13} = 15\%$$ In 15% of the cases in which the Concordance assigned a subclass to product field 16, the assignment was incorrect.

$$\%E = \frac{\#E}{\#C+E} = \frac{4}{15} = 27\%$$ In 27% of the cases in which the Concordance should have assigned a subclass to product field 16, it did not.

NOTE: C and I add to 100%.

The frequencies and percentages of C's, I's, and E's were tabulated for all product fields. Then the data were condensed to include only those product fields or related groups of product fields with 10 or more assignments. The percentages for the condensed data appear in Table 16. It should be recalled that the percentages do not correspond to those in earlier parts of this analysis, because they have been calculated differently. Also unlike the earlier analysis, these figures have not been weighted to correspond to the population of subclasses; they represent only the sample of 110 subclasses. If these figures had been weighted to correct for the oversampling of subclasses with multiple assignments, the percentage error would probably decrease. Also, it should be recalled that these are not errors at the level of individual patents. The figures in Table 16 should be used only to compare error levels across major product fields.

Table 16 shows that there is a relationship between error level and product field. A chi square test revealed that the relationship was significant at less than the .001 level. The first column shows that the product fields range in "percent correct" from a low of 28% (product fields 23-27) to a high of 86% (product field 99). Other product fields with low "percent correct" are 29-32 (31%) and 54 (31%). Other product fields with high "percent correct" are 16 (85%) and 35-43 (77%). Because of the method of calculation, the product fields are ordered inversely with respect to "percent I."

The third column of Table 16 shows that the product fields range in "percent error of exclusion" ("percent E") from a low of 11% (product fields 6-14) to a high of 49% (product fields 29-32). Other product fields with low "percent E's" were 35-43 (19%) and 23-27 (21%). Other product fields with high "percent E's" were 99 (46%) and 55 (38%).

Table 17 shows the product field groups ranked by "percent correct" and by "percent error of exclusion." Both "percent C" and "percent E" have been ranked so that the more correct product fields are at the top. The "percent C" ranking shows that the assignments made by the Concordance are correct substantially more often for product fields at the top of the list -- e.g., "all other SIC's," "Rubber and Miscellaneous Plastic Products," and "Electrical Machinery" -- than they are for those product fields at the bottom of the list -- e.g., "Nonelectrical Machinery, except Other," "Other Nonelectrical Machinery," and "Aircraft and Parts."

The high rate of error in "Nonelectrical Machinery" product fields appears, on the basis of our sample analysis, to be caused by the large proportion of function-based classes -- e.g., Dispensing and Fluent Material Handling -- that are assigned to these product fields. In other

TABLE 16

PERCENTAGE DISTRIBUTION OF TYPE OF SUBCLASS ERROR BY PRODUCT FIELD

PRODUCT FIELD	TYPE OF ERROR*		
	%C	%I	%E
6 - 14	62%	38%	11%
16	85%	15%	27%
21	65%	35%	32%
23 - 27	28%	72%	21%
29 - 32	31%	69%	49%
35 - 43	77%	23%	19%
46 - 53	59%	41%	24%
54	31%	69%	29%
55	71%	29%	38%
99	86%	14%	46%
Avg. of above product fields	46%	53%	31%

*Defined in the text.

words there is a basic mismatch between the USPC and the SIC systems in this area. The high error level in "Transportation Equipment" product fields also appears to be caused by a mismatch between the USPC and the SIC. Most classes in this area designate a function or structure, such as brakes or internal combustion engines, but not the product -- e.g., automobiles or aircraft. The SIC on the other hand is based on the distinctions between products. The problem of mismatches is discussed in more detail in the section on multiple assignments, below.

There also appears to be a tendency for product fields with no disaggregation -- e.g., 99, 16, 55, and 21 -- to have a higher percentage of correct assignments than disaggregated product fields. This makes sense intuitively, since assignment in those cases does not involve such fine distinctions as in disaggregated product fields.

The second ranking in Table 17 tells us that the product fields near the top of the list contain substantially fewer errors of exclusion -- i.e., cases where the Concordance incorrectly omitted a subclass -- than do the product fields near the bottom of the list. For example, "Chemical and Allied Products" has fewer errors of exclusion than "Other Nonelectrical Machinery." The causes of differences in the level of "percent E" are less obvious than the causes of differences in levels of "percent C."

TABLE 17

PRODUCT FIELD GROUPS RANKED BY PERCENT CORRECT AND PERCENT ERROR OF EXCLUSION

PERCENT C			PERCENT E		
Product Field Code	Title	Percent C	Product Field Code	Title	Percent E
99	All Other SIC's	86%	6-14	Chemical & Allied Prods.	11%
16	Rubber & Misc. Plastic Prods.	85%	35-43	Electrical Machinery	19%
35-43	Electrical Machinery	77%	23-27	Nonelectrical Machinery, except "Other"	21%
55	Prof'l & Sci. Instrs.	71%	46-53	Transportation Equip., except Aircraft	24%
21	Fabricated Metal Prods.	65%	16	Rubber & Misc. Plastic Prods.	27%
6-14	Chemical & Allied Prods.	62%	54	Aircraft & Parts	29%
46-53	Transportation Equip., except Aircraft	59%	21	Fabricated Metal Prods.	32%
54	Aircraft & Parts	31%	55	Prof'l & Sci. Instrs.	38%
29-32	Other Nonelectrical Machinery	31%	99	All Other SIC's	46%
23-27	Nonelectrical Machinery, except "Other"	28%	29-32	"Other" Nonelectrical Machinery	49%

Patent-by-Patent Error Analysis

This section addresses errors which occur at the individual patent level as a result of assigning entire USPC subclasses to SIC product field categories.

The purpose of such an analysis is to provide an indication of the accuracy with which the Concordance assigns patents to product fields.

In carrying out this analysis, OTAF examined the set of patents which were used in a similar analysis done by Professor F.M. Scherer and reported upon in "The Office of Technology Assessment and Forecast Industry Concordance as a Means of Identifying Industry Technology Origins," World Patent Information, Vol. 4, No. 1, pp. 12-17, 1982. Professor Scherer selected 99 patents from a larger random sample of 15,112, and compared SIC codes assigned through the OTAF Concordance to codes assigned using a manual process which coded individual patents. Scherer found that the two assignment methods produced agreement at both the 2- and 3-digit SIC levels for 50 of 99 patents when both the original and cross-reference classifications were considered. Although the criteria used by Scherer in assigning the patents to product fields differ somewhat from those used by OTAF, the 50 which were found to agree were not further examined.

For the remaining 49 patents, two OTAF analysts considered the technology revealed in the patents and determined the SIC assignment or assignments in accordance with the OTAF criteria. In this study the procedure of the "New Decision Rule" (Table 11) was used. This de novo assignment was then compared to the assignment which resulted by applying the Concordance to the USPC classifications of the patent.

The Concordance assignment was considered to be "OR Correct" if the de novo assignment matched all or part of the assignments produced by concording the original classification of the patent. This standard is higher than that used by Scherer since it requires all de novo assignments to be produced by the Concordance and since it only considered the original classification of the patent.

The Concordance assignment was considered "OR-XR Correct" if at least one of the de novo assignments matched at least one of the assignments produced by concording both original and cross-reference classifications of the patent. This standard corresponds to Scherer's "Agreement at Both the 2- and 3-digit Levels."

Those assignments which were neither OR Correct nor OR-XR Correct were further examined to determine if they were correct at the "2-Digit SIC Level."

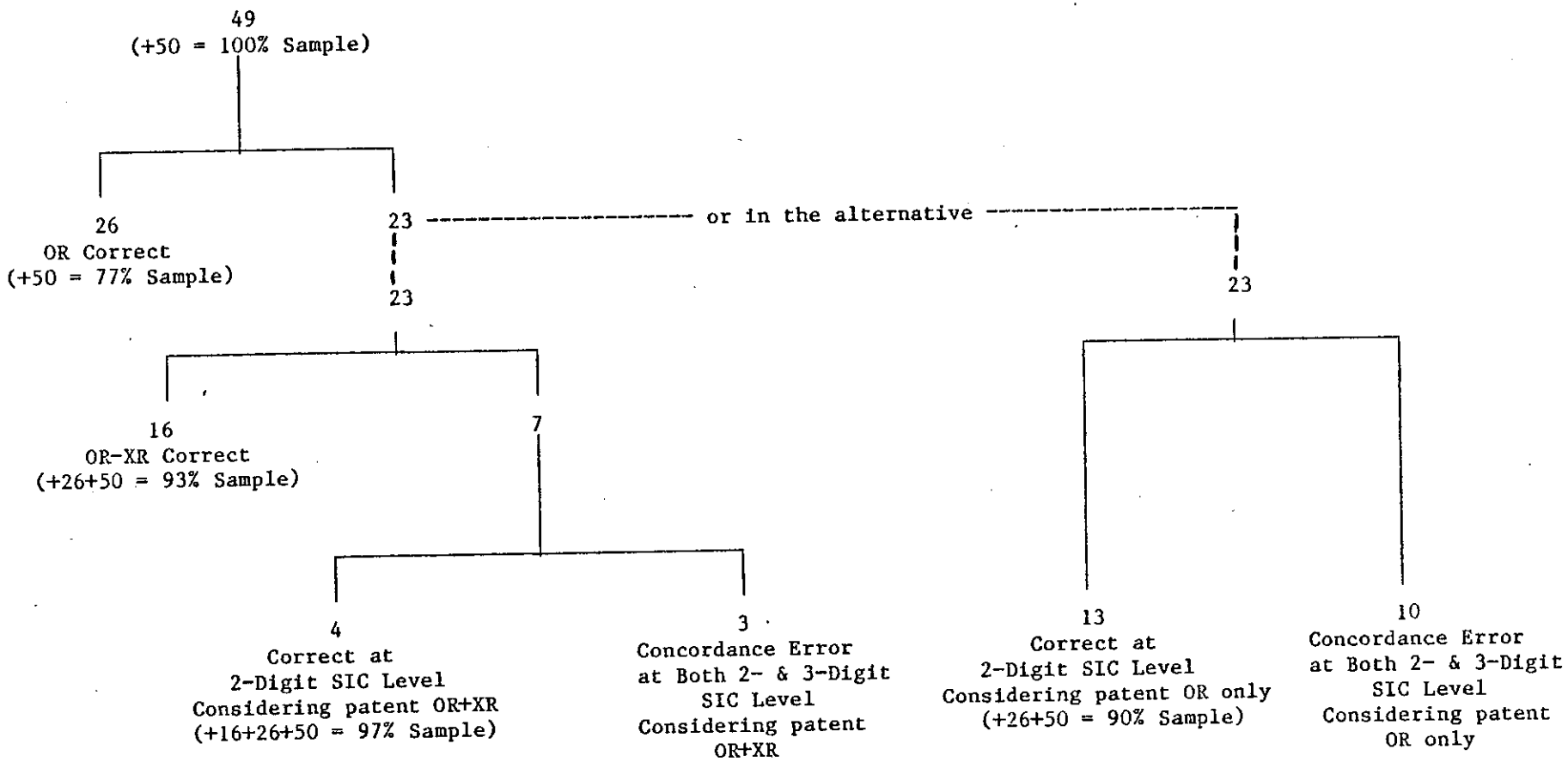
The results are shown in Figure 4. Of the 49 patents considered, 26 were OR Correct. Sixteen of the remaining 23 were OR-XR Correct. Adding these to the previous 26 and the 50 correct from Scherer's original sample, we have 92 correct out of 99 (93%). If we also include the four

which are correct at the 2-Digit Level we are left with only a 3% error rate. Of the 23 patents which were not OR correct, an additional 13 were found to be OR correct at the 2-Digit Level, resulting in a 10% error rate.

No definite conclusions can be drawn from this analysis, because of the very small number of patents studied. To obtain a precise estimate of Concordance error at the level of individual patents, it would be necessary to study a larger sample of patents. It would be useful to get agreement on the appropriate standard of accuracy that should be applied in such an analysis.

FIGURE 4

RESULTS OF PATENT-BY-PATENT ANALYSIS OF ERROR



OR Correct = Concording patent OR yields de novo assignment
OR-XR Correct = Concording patent OR and XR yields at least one de novo assignment
Correct at 2-Digit Level = Concording patent classification yields at least one de novo 2-digit SIC assignment

Analysis of the Extent of Multiple Assignments

The OTAF Concordance assigns approximately 101,000 subclasses of the U.S. Patent Classification (USPC) system to 55 product fields of the Standard Industrial Classification (SIC) system. Because of differences which exist between the USPC and the SIC, a one-to-one concordance at the subclass level is often not possible. Realizing this, the developers of the Concordance assigned patent subclasses to all of the possible SIC product fields to which they were pertinent. Therefore, if a subclass contained patents which related to more than one field of industrial activity, it was assigned to more than one SIC product field.

This procedure resulted in "multiple counting" of subclasses in more than one of the 55 SIC product fields. As a result, the sum of patents counted in all of the 55 product fields was greater than the actual number of patents granted.

For example, the USPC system classifies bearings based on structure regardless of end use. In the SIC, placement of bearings depends on whether they are used in motor vehicles or other types of machines or if they are made of plastic or metal. The result was that virtually all of the patents in the USPC subclasses which include bearings were triple-counted, i.e., they were concorded to SIC's relating to motor vehicle parts, general industry machinery and plastic products. Clearly not all of these patents pertain to all three SIC product fields, but some might. At the USPC subclass level, it is not possible to determine the proper distribution of patents into each category.

The questions which OTAF wanted to answer in this portion of the Concordance assessment were:

- How much multiple counting is there and what causes it?
- How does multiple counting affect the use of the Concordance?
- Should the placement decision rules be changed to minimize or eliminate multiple counting?
- Are there acceptable alternatives to multiple counting?

The Amount of Multiple Counting

SIC product fields of the OTAF Concordance are used to classify manufacturing sectors, while the USPC subclasses are used to classify inventions. Considering the differences in scope and purpose of the USPC and SIC, the amount of multiple counting might have been expected to be higher than it was.

Table 18 shows the distribution of active USPC subclasses according to the number of unique SIC product fields to which they were concorded. Active subclasses are those which received at least one patent as an original classification during the time period 1963-1981. The number of

active subclasses is 83,438. Unique SIC product fields are those which are not roll-ups or summaries of less aggregated, finer divisions. Forty-one of the 55 SIC product fields are unique. The "All Other" category, product field 99, is also included in the counts of Table 18.

As shown in Table 18, 71.6% of the active USPC subclasses were concorded to only 1 unique SIC category and less than 3% were concorded to 4 or more categories.

To determine the level of multiple counting for patents, a multiple count ratio was calculated by dividing the number of patents counted in unique SIC product fields by the actual number of distinct patents granted.* For the years 1963-1981, the sum of the patents in the unique SIC product fields was 1,740,736. The actual number of patents granted during that time period was 1,234,650. The multiple count ratio was therefore 1.41. That is, the sum of the patents counted in these categories was 1.41 times (or 41%) higher than the actual number of patents granted.

In the product fields which correspond to 2-digit SICs, the ratio dropped to 1.20. In other words, the sum of the patents counted in the 13 major product field groupings (including product field 99) exceeded the actual number of patents granted by about 20%. Less multiple counting is expected at this more aggregated level, since the technology in a patent subclass is less likely to pertain to very different product fields -- e.g., "Chemicals and Allied Products" (SIC 28) and "Nonelectrical Machinery" (SIC 35) -- than it is to more similar product fields -- e.g., "Motor Vehicles" (SIC 371) and "Motorcycles" (SIC 375). As shown in Table 19, the multiple counting ratio at this level was relatively stable over time.

When the 3- and 4-digit SIC product fields within the 2-digit SIC levels were considered -- i.e., those that are indented under product fields 3, 18, 22, 33, and 44 -- the multiple counting ratio varied by product field. (See Table 20.) The multiple counting ratios in Table 20 were calculated by dividing the sum of the number of patent counts in the unique product fields under each 2-digit product field by the patent count in the 2-digit product field (which does not contain duplicates). The level of multiple counting within the 2-digit SIC levels may have been underestimated by this method because it did not account for multiple counts that cross 2-digit SICs -- e.g., SICs 35 and 37.

As shown in Table 20, the multiple count ratio varied from 1.20 in "Chemicals and Allied Products" to 1.54 in "Transportation Equipment." Differences were not a function of the number of unique categories contained in the 2-digit code; for example, product field code 18, a roll-up of only 2 more disaggregated codes, had the second highest

---

*Only unique SIC product fields were considered, since roll-ups by definition recount patents in their subdivisions.

TABLE 18

DISTRIBUTION OF ACTIVE USPC SUBCLASSES BY NUMBER OF UNIQUE SIC PRODUCT FIELDS TO WHICH THEY ARE ASSIGNED

Number of unique SIC assignments	1	2	3	4	5	6
Number of active subclasses	59,745	14,985	6,566	1,835	284	23
Percent of active subclasses	71.6	18.0	7.9	2.2	0.3	< 0.1

TABLE 19

YEARLY MULTIPLE COUNT RATIO AMONG 2-DIGIT PRODUCT FIELDS

YEAR	RATIO
1967	1.17
1968	1.18
1969	1.21
1970	1.21
1971	1.21
1972	1.21
1973	1.21
1974	1.22
1975	1.19
1976	1.22
1977	1.22
1978	1.20
1979	1.19
1980	1.20
1981	1.21

TABLE 20

MULTIPLE COUNT RATIOS FOR 3- AND 4-DIGIT SIC LEVEL PRODUCT FIELDS WITHIN 2-DIGIT SIC LEVEL PRODUCT FIELDS BASED ON 1963-1981 TOTALS

2-Digit Product Field Code & Title (number of patents)	Unique 3- or 4-digit Product Field Codes within 2-Digit Code (number of patent counts)	Multiple Count Ratio
3 - Chemicals & Allied Products (204,445)	6-9, 11-14 (244,837)	1.20
18 - Primary Metals (16,307)	19, 20 (21,342)	1.31
22 - Machinery, Except Electrical (385,768)	23-27, 29-32 (484,458)	1.26
33 - Electrical Machinery (251,272)	35, 36, 38-40, 42, 43 (315,273)	1.25
44 - Transportation Equipment (84,131)	46, 47, 49-54 (129,264)	1.54

ratio. OTAF codes with the highest and lowest ratios each contained eight subdivisions. This suggested that the differences were specific to product fields.*

For two of the five 3- and 4-digit product fields, multiple counting appeared to be affected by time and ownership. As shown in Table 21, the multiple counting ratio for product fields under "Transportation Equipment," increased from 1.49 in 1969 to 1.65 in 1981. For foreign-origin patents the ratio was higher than the ratio for all patents, and increased from 1.64 in 1969 to 1.81 in 1981. Similar, although not as marked, trends appeared in "Chemical and Allied Products." Multiple counting within the other three product fields did not follow this trend.

The observed increase in multiple counting over time in certain product fields is difficult to explain. Possibly, it is an artifact of the data. If real, it could not be caused by an increase in the number of subclasses with more, rather than fewer, assignments because of the nature of the Concordance update process. Each update of the Concordance results in the reorganization of all the patents granted during the period covered by the OTAF data base into the subclasses contained in the current USPC classification. Therefore the increase in multiple counting must have been caused by changes in patent activity. Thus the proximate explanation is that patenting in subclasses with more assignments grew faster than patenting in subclasses with fewer assignments.

Possible explanations for these patterns of patent activity were offered at the workshop. Luc Soete suggested that multiple allocation was higher in new patenting areas because the specific industry was difficult to identify. Moreover, foreign patenting tended to be more in those new patenting areas. Ned Ellis suggested that new patents were going into

---

*With respect to product field 18, "Primary Metals," which has the second highest level of multiple counting, the following points are noted. (See Table 1 for product field numbers.) The distinction between the two unique product fields within it is unclear in some cases. For example, SIC 3399, "Primary Metal Products, Not Elsewhere Classified," is placed in product field 19, which is entitled "Primary Ferrous Products," but in fact it includes nonferrous products as well. Moreover, while product field 18 is labeled "Primary Metals" after SIC 33, which is "Primary Metal Industries," both include secondary metals as well -- i.e., product field 20, "Primary and Secondary Nonferrous Metal," and SIC 334, "Secondary Smelting and Refining of Nonferrous Metals."

No change was made with respect to these product fields in the 1983 Concordance, but future changes should be considered. The product field titles could be reworded to reflect more accurately the subject matter of the SIC categories included. Alternatively, in view of the relatively low number of patents in product field 18 (about 16,000 out of about 1,200,000), disaggregation may not be necessary and could be eliminated.

TABLE 21

MULTIPLE COUNT RATIOS FOR 3- AND 4-DIGIT SIC LEVEL PRODUCT FIELDS WITHIN 2-DIGIT SIC LEVEL PRODUCT FIELDS FOR SELECTED YEARS

Product Field	1969	1971	1973	1975	1977	1979	1981
Chemicals and Allied Products							
Total Patents	1.16	1.17	1.17	1.19	1.23	1.25	1.25
Foreign Origin	1.20	1.19	1.19	1.20	1.26	1.29	1.28
Primary Metals							
Total Patents	1.33	1.28	1.37	1.32	1.32	1.28	1.29
Foreign Origin	1.35	1.33	1.42	1.33	1.26	1.29	1.27
Machinery Except Electrical							
Total Patents	1.25	1.27	1.26	1.26	1.23	1.26	1.24
Foreign Origin	1.24	1.23	1.23	1.23	1.20	1.23	1.21
Electrical Machinery							
Total Patents	1.28	1.27	1.25	1.26	1.25	1.24	1.25
Foreign Origin	1.29	1.28	1.28	1.29	1.28	1.28	1.28
Transportation							
Total Patents	1.49	1.48	1.54	1.54	1.53	1.58	1.65
Foreign Origin	1.64	1.61	1.73	1.72	1.80	1.72	1.81

areas that were not covered by the SIC, which is due for an update. Richard Campbell suggested that the "generality" of technology, that is, its range of applicability, was increasing.

It remains to be determined which of these explanations, if any, is correct or whether the observed trends are even real. [OTAF Note: The issue has become somewhat moot, because the trends are much reduced in the 1983 concordance.]

The Cause of Multiple Counting

The underlying source of multiple counting in the Concordance was the decision rule to assign a subclass to all product fields to which it is pertinent. Because of the differing natures of the two classification systems it was impossible to assign some patent subclasses to a single product field without being arbitrary. For example, in the case of a subclass covering "tubs" which does not specify the material the tub is made of, it would be highly arbitrary to assign it to "Fabricated Metal Products" rather than "Miscellaneous Plastic Products" both of which contain "tubs." The level of multiple counting was higher in areas where the "match" between the USPC and SIC systems is not good. (Some of these areas are discussed below.)

A certain amount of multiple counting was also caused by human error in making product field assignments. That is, coders sometimes assigned subclasses to product fields to which they were not pertinent. The level of "human" error was, of course, heightened by the multiple-assignment decision rule and the ambiguous relationship that exists between some portions of the USPC and SIC systems.

Table 22 lists all 27 USPC classes which had multiple counting ratios above 2. Of these, seven classes (62, 124, 125, 126, 331, 536, and 68) had errors which appeared to be primarily the result of human errors--they were simply assigned to product fields to which they were not pertinent. These errors could be easily corrected to reduce the number of multiple counts. One class -- 261 -- has errors that were a result of both human error and a bad match with the SIC.

The multiple counting in the other 19 classes resulted primarily from a "bad match" between the USPC and SIC systems. Here there was no easy solution. The result was excessive multiple counting and, concomitantly, the increased likelihood of errors in assignment.*

These "problem" classes were of two general types:

1) Machine parts where the USPC does not specify the machine, but where the SIC places the parts with a specific machine, e.g., brakes, clutches, internal combustion engines, bearings, impellers.

---

*The relationship between multiple counting and error was discussed in the section on subclass error analysis.

TABLE 22

USPC CLASSES WITH MULTIPLE COUNT RATIOS GREATER THAN 2

Class Number	Subject Matter	Multiple Count Ratio*	Primary Cause of Multiple Counting	
			Bad Match	Human Error
303	Fluid Pressure Brakes	4	X	
241	Solid Material Comminution	4	X	
192	Clutches	4	X	
138	Pipes & Conduits	4	X	
125	Stone Working	4		X
124	Mechanical Guns	4		X
416	Impellers	3.5	X	
239	Fluid Sprinkling	3.5	X	
188	Brakes	3.2	X	
126	Stoves & Furnaces	3.2		X
406	Fluid Conveyors	3	X	
331	Oscillators	3		X
308	Bearings	3	X	
226	Advancing Materials	3	X	
209	Sorting Solids	3	X	
141	Fluent Material Handling	3	X	
134	Cleaning & Liquid Contact	3	X	
123	Internal Combustion Engines	3	X	
360	Dynamic Magnetic Information Storage & Retrieval	2.8	X	
261	Gas and Liquid Contact Apparatus	2.8	X	X
222	Dispensing	2.7	X	
536	Carbohydrates	2.5		X
221	Article Dispensing	2.5	X	
68	Fluent Treating Textile Apparatus	2.5		X
62	Refrigeration	2.5		X
4	Baths	2.5	X	
206	Special Receptacle	2.3	X	

*approximate

2) General purpose machinery and methods where the USPC does not specify the industry but where the SIC placement depends on the industry, e.g., machines and methods for separating, advancing, conveying, dispensing, cleaning and sprinkling.

Classes of the first type are those which, in general, fall into "Transportation Equipment" and explain the high multiple count ratio (1.54) which occurs within that product field. Since the majority of subclasses do not specify the type of transportation equipment in which the part is used, these subclasses were assigned to all possible product fields to which they pertained. Even though the majority of the patents in a particular subclass relate to motor vehicles, the subclass may also contain patents which pertain to other transportation and general industrial equipment.

The second type of problem area relates for the most part to Product Field 22, "Machinery, Except Electrical." Here the problem of multiple counting was compounded by the inability to determine the correct SIC product field. This problem was discussed in more detail in the section on error analysis. A specific example is USPC Class 222, Dispensing. Its subclasses generally describe the structure of the dispensing machine, but recite neither the industry using the machine nor the article dispensed. A sample of patents contained in the subclasses will show some specific uses, but other possible uses are not excluded. The proper SIC product fields can only be speculated upon.

The Effect on Concordance Use

The high multiple counting ratios and the increase in that ratio over time in 3- and 4-digit "Transportation Equipment" product fields indicated that the Concordance might be of limited use for analysis seeking to distinguish among these product fields. This came as no surprise to users of the Concordance who had found patents to Japanese auto makers counted together with those of the aircraft industry. Because multiple counting increased over time, time series data were also distorted. Under the then existing placement rules, disaggregation of "Transportation Equipment" produced results of limited utility.

Trends in "Chemical and Allied Products" also indicated an increase over time, such that care should be taken when drawing conclusions based on these data. It was possible that some of the multiple assignments could be eliminated through error correction. Further, since the multiple counting ratio in "Chemical and Allied Products" was less than the ratio in "Transportation Equipment" (1.20 vs. 1.54) the problem was of a lesser magnitude.

OTAF concluded from its analysis that it might be possible to use the Concordance without severe distortion of the results if use was restricted to the major groupings, i.e., the product fields that correspond to the 2-digit SIC codes, where multiple counting was only about 20%.

Several possible approaches to ameliorating the multiple counting problem were discussed at the workshop. That discussion is summarized in section II of this paper, "Workshop Summary."

Possible Additional SIC Categories for the Concordance

Currently the OTAF Concordance assigns patent subclasses to 13 2-digit SIC codes (i.e., Major Groups) plus an "all other SIC" category. The 13 Major Groups are:

Existing SIC Major Groups

SIC 20	Food and Kindred Products
22	Textile Mill Products
28	Chemicals and Allied Products
13	Oil and Gas Extraction
29	Petroleum Refining and Related Industries
30	Rubber and Miscellaneous Plastics Products
32	Stone, Clay, Glass, and Concrete Products
33	Primary Metal Industries
34	Fabricated Metal Products
35	Machinery, Except Electrical
36	Electrical and Electronic Machinery, Equipment and Supplies
37	Transportation Equipment
38	Measuring, Analyzing, and Controlling Instruments; Photographic, Medical, and Optical Goods; Watches and Clocks

All are manufacturing industries, with the exception of SIC Major Group 13, which is combined with SIC Major Group 29 to form OTAF Product Field 15 "Petroleum and Natural Gas Extraction and Refining."

Prior to this project OTAF had received a specific request from the University of Sussex to consider adding the following manufacturing Major Groups to the Concordance:

Requested Additional SIC Major Groups

SIC 21	Tobacco Products
23	Apparel, Other Textile Products
24	Lumber and Wood Products
25	Furniture and Fixtures
26	Paper and Allied Products
27	Printing and Publishing
31	Leather and Leather Products
39	Miscellaneous Manufacturing Industries

Other users and potential users had expressed interest in a patent concordance to nonmanufacturing SIC codes.

OTAF investigated the feasibility of adding manufacturing and nonmanufacturing SIC codes to the Concordance by looking at (1) OTAF Product Field 99 ("All Other SIC's") and (2) the Canadian SICs assigned to patents by the Canadian Department of Consumer and Corporate Affairs.

Manufacturing SICs

Patent subclasses which disclose technology pertaining to a Major Group not in the Concordance are assigned to OTAF product field 99. Some subclasses are counted only in product field 99, while other are counted both in product field 99 and in one or more of the other product fields. Product field 99 contains about 5.3%-5.8% (depending on how the figure is calculated) of the patents in the Concordance. Although this is a relatively small share of the patents, OTAF believed it would be possible to break out some additional SIC's from product field 99.

To identify SIC codes that could be broken out of product field 99, OTAF reviewed 48 of the 91 USPC classes* assigned to that product field. Twenty-two of the 48 classes are assigned entirely to product field 99. (See Table 23.) Twenty-six of the classes have a significant number of patents in product field 99 -- i.e., 1,000 or more patents or 30% or more of all their patents. (See Table 24.)

OTAF made preliminary assignments of some of the 48 classes to SIC's not already in the Concordance by using the class schedules in the Manual of Classification and the 1972 Standard Industrial Classification Manual. These were only preliminary assignments and were based on the technology in the class as a whole, rather than on the specific subclasses that were assigned to product field 99. Many, but not all, classes could be assigned to SIC categories using this technique.

Table 25 shows the SIC Major Groups which apparently can be broken out of product field 99 as separate product fields in the Concordance. It should be noted that this list includes all the Major Groups requested by the University of Sussex. Table 25 also includes data on 1980 corporate R&D funding and net sales, data which may be used to determine priorities for adding SIC codes to the Concordance. The R&D data indicate that SIC 26, 39 and 21 may be active in patenting because they have relatively high R&D expenditures. Based on net sales, SIC 24 and 27 could also be given high priority for being added to the Concordance.

Nonmanufacturing SICs

As noted previously, the OTAF Concordance includes only one nonmanufacturing SIC code--Major Group 13, Oil and Gas Extraction--which is combined with SIC 29 to form Product Field 15--Petroleum and Natural Gas Extraction and Refining. There are 63 nonmanufacturing SIC Major Groups which are not in the Concordance. The questions are: (1) whether it is feasible to include any more nonmanufacturing SIC codes in the Concordance and (2) which ones?

---

*Note that this analysis was done at the class level, rather than the subclass level. Thus some or all of the subclasses in 91 classes are assigned to Product Field 99. If the decision is made to add new SICs, a more thorough attempt must be made to identify all relevant subclasses.

TABLE 23

USPC Classes Assigned Entirely to Product Field 99

Class Number	Class Title	Number of Patents in Product Field 99
5	Beds	2,880
6	Bee Culture	104
14	Bridges	460
27	Undertaking	165
46	Amusement Devices, Toys	3,668
54	Harness	203
63	Jewelry	313
79	Button Making	18
84	Music	3,821
132	Toilet	1,660
163	Needle and Pin Making	11
190	Baggage	395
211	Supports, Racks	2,859
217	Wooden Receptacles	261
229	Paper Receptacles	4,732
273	Amusement Devices, Games	8,023
281	Books, Strips and Leaves	303
282	Manifolding	384
283	Printed Matter	206
312	Supports, Cabinet Structures	3,443
401	Coating Implements with Material Supply	1,785
402	Binder Device Releasably Engaging Aperture or Notch of Sheet	249

TABLE 24

USPC Classes with a Significant Number of Patents in Product Field 99*

Class Number	Class Title	Number of Patents in Product Field 99
2	Apparel	3,285
9	Boats, Buoys and Aquatic Devices	676
12	Boot and Shoe Making	355
15	Brushing, Scrubbing and General Cleaning	1,509
24	Buckles, Button, Clasps, etc.	1,131
36	Boots, Shoes and Leggings	1,518
40	Card, Picture and Sign Exhibiting	2,893
43	Fishing, Trapping and Vermin Destroying	2,841
52	Static Structures, e.g, Buildings	8,363
108	Horizontally Supported Planar Surfaces	1,541
124	Mechanical Guns and Projectors	935
131	Tobacco	1,536
135	Tents, Canopies, Umbrellas and Canes	707
150	Cloth, Leather and Rubber Receptacles	573
162	Paper Making and Fiber Liberation	2,212
206	Special Receptacle or Package	4,954
224	Package and Article Carriers	834
231	Whips and Whip Apparatus	10
272	Amusement and Exercising Devices	1,847
280	Land Vehicles	2,292
297	Chairs and Seats	4,009
404	Road Structure, Process and Apparatus	808
405	Hydraulic and Earth Engineering	3,442
427	Coating Processes	1,696
428	Stock Material or Miscellaneous Article	5,933
434	Education, Demonstration, and Cryptography	2,944

---

*Classes which have 30% or more of their patents assigned to product field 99 or 1,000 or more patents.

TABLE 25

ADDITIONAL SIC MAJOR GROUPS WHICH CAN BE BROKEN OUT OF PRODUCT FIELD 99

	SIC MAJOR GROUP	1980 Corporate R&D Funding (In $Millions)	1980 Corporate Net Sales (In $Millions)
21	Tobacco manufactures	141	24,275
23	Apparel & other finished products made from fabrics & similar materials	19	6,075
24	Lumber & wood products, except furniture	97	15,612
25	Furniture & fixtures	68	6,203
26	Paper & allied products	508	48,289
27	Printing, publishing, & allied industries	89	12,866
31	Leather & leather products	18	3,771
39	Miscellaneous manufacturing industries	162	7,164

---

*National Science Foundation, Research and Development in Industry, 1980. Detailed Statistical Tables. (Washington, D.C., NSF 82-317.) Table B-2, Selected Data for Companies Performing Research and Development by Industry: 1979-1980.

In most cases, the inventive activity takes place in the industry that manufactures a product or apparatus, even if that product or apparatus is subsequently used in a nonmanufacturing industry. Recent research by Scherer has noted this general tendency in interindustry technology flows.* Hence, many patents pertaining to agriculture, construction, and other nonmanufacturing industries are assigned to manufacturing SICs such as "Farm and Garden Machinery and Equipment" (SIC 352) or "Construction, Mining, and Materials Handling Machinery and Equipment" (SIC 353).

Some inventive activity does occur in nonmanufacturing industry, however. The NSF collects data on R&D funding in four categories for nonmanufacturing industries:

	SIC	1980 R&D Funding (In $Millions)
"Electric, Gas, & Sanitary Services"	49	220
"Miscellaneous Business Services"	73	514
"Miscellaneous Services"	89	195
"Nonmanufacturing Industries"	07-17, 41-48, 50-67 807	640

Source: Table B-2, "Research and Development in Industry, 1980," NSF.

Combined, these industries accounted for 3.6% of total industrial R&D funding in 1980. Other nonmanufacturing industries not in this list presumably account for such a small proportion of R&D funding that it cannot be captured in the NSF survey.

Among the USPC classes in product field 99, some pertain to nonmanufacturing industries--e.g., education, undertaking, animal husbandry, and bridges. By its nature, however, the inventive activity in these classes would often take place in some manufacturing industry--e.g., educational equipment or chemicals.

The Canadian Department of Consumer and Corporate Affairs assigns each patent to appropriate categories in the Canadian SIC. Each patent can be given both an SIC of Manufacture and an SIC of Use. To get an indication of which nonmanufacturing SIC categories have patent activity, OTAF analyzed Canadian data on SIC of Use from 1978 and 1979.** OTAF analyzed data for: (1) all patents and (2) U.S.-origin patents. (See Table 26.) According to this analysis approximately one-third of the 1978 and 1979 patents disclose a category of use which is in a nonmanufacturing SIC.

---

*F.M. Scherer,"Interindustry Technology Flows in the United States" Research Policy, v. 11 (1982): 227-245.

**These data are from a paper by Mr. E.D. Ellis entitled "Canadian Patent Data Base." (nd., np.)

If U.S.-origin or total Canadian patent activity is an indicator of patent activity in the United States, then the nonmanufacturing areas which might usefully be added to the U.S. SIC Concordance are those which correspond to the following Canadian SICs:

- Service Industries

  Electric Power, Gas and Water Utilities
  Health and Welfare Services
  Other Services

- Nonindustrial Use

  Consumer's Use

- Construction Industries

  General Contractors
  Special-Trade Contractors

- Mineral Fuels

It must be remembered, however, that the above list is based on SIC of Use, whereas the OTAF Concordance in general does not place subclasses with the industry of use. To do so would require making new assignments for most USPC subclasses, a formidable task. Also, some technology, especially electronics or information-based technology, is so broadly used in the nonmanufacturing sector that it would be difficult if not impossible to determine all possible assignments or even the major assignments. Perhaps the Canadian experience can be enlightening in this respect. Another possible approach to adding nonmanufacturing SICs to the Concordance may be to establish a special product field for nonmanufacturing SICs.

Comments or suggestions were invited from Workshop participants on the addition of new SICs to the Concordance. The subject did not engender much discussion, however, and no additions were made to the subsequent update of the Concordance.

TABLE 26

CANADIAN PATENTS CATEGORIZED BY SIC OF USE FOR NONMANUFACTURING INDUSTRIES 1978-1979*

	TOTAL		U.S.-ORIGIN**	
	Number	Percent	Number	Percent
Farms	689	1.5	401	1.5
Services to agriculture	159	0.4	95	0.4
Forestry	135	0.3	69	0.3
Fishing & trapping	85	0.2	45	0.2
Metal mines	156	0.3	58	0.2
Mineral fuels	708	1.6	510	2.0
Non-metal mines	26	0.1	10	-
Quarries & sand pits	11	-	6	-
Services to mining	395	0.9	231	0.9
Construction industries				
General contractors	1,884	4.2	1,015	3.9
Special-trade contractors	865	1.9	539	2.1
Services industries				
Electric power, gas, water utilities	1,177	2.6	692	2.6
Health & welfare services	1,612	3.6	1,078	4.1
Other services	3,365	7.4	2,120	8.1
Non-industrial use				
Consumer's use	2,886	6.4	1,766	6.8
SUM OF ABOVE	14,153	31.4	8,635	33.1
MANUFACTURING	31,111	68.6	17,392	6.9
GRAND TOTAL	45,264	100.0	26,027	100.0

*Details may not add to totals because of rounding error.

**The numbers in this column were calculated from Ellis' Table 1 as the number of patents in particular SIC of Use times the percentage which were U.S. origin. The percentages in this column were calculated from the numbers in the column immediately to their left.

Source: Ellis, E.D., Canadian Patent Data Base, Research and International Affairs Branch, Consumer and Corporate Affairs, Canada, Table 1.

Use of the Concordance

One of the issues involving the Concordance is its utility -- e.g., the extent to which it is used, the ways in which it is used, and the increase in utility that might come from various suggested changes or improvements. Ultimately the continuation and possible enhancement of the Concordance depend on OTAF and NSF being convinced that it is serving a real need.

OTAF did not conduct a thorough assessment of the use of the Concordance as part of this project. The user Workshop was intended as the main effort to obtain information on use. OTAF did analyze its logbook entries to determine the proportion of inquiries it receives that pertain to the Concordance, however.

The logbooks for the period September 1981 through October 1983 were examined. (Logbooks were not kept prior to September 1981.) These logbooks record each inquiry OTAF receives regarding patent-related data and information. Not all inquiries result in OTAF providing data or information. In many cases the inquiring party decides not to pursue the matter further.

Each "inquiry" regarding Concordance based data was recorded, including the date, the inquiring party, and whether data were actually provided. Table 27 summarizes the findings.

TABLE 27

NUMBER OF INQUIRIES REGARDING THE OTAF CONCORDANCE

Inquiring Party	Nature of Inquiry		Total Inquiries
	Inquiry Only	Data Provided	
Business or Industry	6	7	13
University or Nonprofit Research Organization	5	8	13
Government	5	5	10
Individual	1	2	3
Total	17	22	39
Percent of total OTAF inquiries (estimated)	1.6%	3.9%	2.4%

During the period examined, OTAF received 39 inquiries regarding Concordance-based data, about 2.4% of the total number of inquiries received by OTAF. Of the 39 inquiries, 22 or 56% resulted in the actual provision of data. This compares to 35% for all OTAF inquiries. However, the data provided in Concordance-related inquiries is almost always either a photocopy or microfiche of an existing report, not a new report, which many OTAF requests do involve.

Inquiries about Concordance-based data came primarily from business organizations (13), universities (13), and government organizations (10). Of those inquiries which resulted in the provision of data, most came from universities (8), then business (7), and government (5).

The analysis shows that inquiries regarding Concordance-based data constitute a small proportion of the inquiries received by OTAF. This should not be interpreted to mean the Concordance is of little value to patent data users. Many businesses, in particular, find that patent data organized by USPC or patent owner meets their needs for patent searches or examination of competitors' technological position. Also, data based on the Concordance is just one of many services provided by OTAF and has not been emphasized over the others. Moreover, about 1.5 Concordance inquiries were received per month over the period studied and almost one Concordance data report was provided per month.

OTAF concluded tentatively on the basis of this analysis and more subjective information that the Concordance meets a significant, although specialized need and may be particularly useful to universities and government organizations. This conclusion was confirmed at the workshop.

PART II

WORKSHOP SUMMARY

Introduction

As part of the project to review and assess the OTAF Concordance, a workshop was held November 18-19, 1983, at the Hospitality House Motor Inn in Arlington, Virginia. The participants in the workshop are listed in Appendix C of this report.

At the workshop, OTAF, NSF and users of the Concordance discussed the issues raised in the draft report and debated alternative courses of action. This allowed the major parties interested in the Concordance to participate in the assessment of its utility and validity and to make decisions about its future.

Workshop Organization

The Concordance Workshop was divided into four sections:

- the OTAF review and assessment, which was presented by OTAF staff;
- three presentations on experience in using the Concordance and desired improvements, given by Griliches, Carpenter, and Soete;
- two presentations on possible improvements, given by Evenson and Campbell; and
- three presentations on the use of patent-by-patent assignments of SIC's, given by Scherer, Ellis and Kirsh, and Lawson.

There was considerable discussion of each presentation, as well as summary sessions at the close of both days.

Issues Addressed by the Workshop

The three main issues addressed by the Workshop were:

- Should the Concordance be continued?
- What changes or improvements should be considered?
- What alternatives to the Concordance should be considered?

The discussion of these issues was guided by the following subsidiary questions or topics:

- Should the Concordance be continued?

  - Does it address an important need?
  - Does it meet that need adequately?
  - Is the error level acceptable?
  - Are the decision rules appropriate?

- Can the Concordance be made acceptable, if it is not?
- Is the Concordance based on valid conceptual grounds?
- Are there more attractive, feasible alternatives to the Concordance?

• What changes or improvements should be made in the Concordance?

- Reduce the error level
- Address the multiple counting problem
- Revise the decision rules
- Change the update process
- Establish additional SIC fields
- Dissaggregate product fields further
- Use additional information to make assignments
- Replace the SIC with some other classification

• What alternatives should be considered?

- Switch to patent-by-patent assignment
- Use additional information to characterize technologies
- Use only the USPC to characterize technologies
- Use the Derwent patent classification system as basis for Concordance

The following sections of the report summarize the Workshop discussion of the above issues and identify those points on which agreement was reached. In a few places information has been inserted that was not brought up at the Workshop; these are labeled as "OTAF notes."

Should the OTAF Concordance Be Continued?

Summary

The fundamental issue addressed by the Workshop was whether the OTAF Concordance should be continued. The Workshop arrived at the following general points of agreement. The Concordance should be continued on at least a temporary basis, while the feasibility of implementing a patent-by-patent system of assigning SICs is evaluated by the PTO. Certain immediate improvements are needed in the Concordance, including correction of obvious errors in assignment and proportional weighting of subclasses with multiple assignments. If the Concordance continues in the longer term, more fundamental changes such as revision of the decision rules may be desirable.

Does It Address an Important Need?

The general need for patent data organized by industry, technology, or product is obviously important, as evidenced by the attendance at the Workshop and the many approaches taken or suggested. The OTAF Concordance has proven useful to the Science Indicators Unit of the NSF,

which uses the resulting patent data as indicators of the output of R&D, and to researchers at the Science Policy Research Unit (SPRU) at Sussex University, who use them as indicators of innovativeness, among others.

The specific needs of many researchers are not met by the Concordance, however. Participants Griliches, Scherer, and Carpenter testified to difficulties experienced in trying to use the Concordance. Moreover, because the Concordance is based on the USPC, it does not solve all of the problems of that Classification with respect to industry analyses. Alternative systems of classifying patents by industry make possible different types of analysis that may be especially useful for policy decisions, as pointed out by Scherer, Ellis, and Kirsh.

Does It Meet That Need Adequately?

Most researchers who have tried to match patent data with R&D and/or productivity data have concluded that the Concordance is not adequate for that purpose. The discovery of a high level of error and multiple counting bias in some parts of the Concordance raised serious questions about its adequacy for the purpose of Science Indicators, also. These problems are discussed in more detail later.

Is the Error Level Acceptable?

The error level in the assignment of subclasses to product fields, as reported by OTAF, was 32.5% at the level of the 41 unique product fields and 18.8% at the level of 2-digit product fields. These estimates are probably high because of the strict criterion of "correct" that was used. An assignment was counted as correct only if the subclass was assigned to all the correct product fields and was not assigned to any incorrect product fields. Nonetheless, the general opinion seemed to be that the error level was unacceptable. Concern was expressed that such error levels could seriously mislead policy makers and confound research results.

Are Decision Rules Appropriate?

Problems with the existing decision rules were discussed, as well as possible alternatives. This discussion is described in more detail later. The main conclusion was that the decision rules should not be changed for the 1984 update, but the possibility of future changes was left open.

Can the Concordance Be Made Acceptable?

It appears that the Concordance can be made acceptable for those who want highly aggregated data -- i.e., Science Indicators and the SPRU research -- by correcting obvious errors and changing to proportional weighting. For other purposes, however, such as research that relates patenting to R&D expenditures and/or to productivity increases at detailed levels, the Concordance probably will be less acceptable.

Is the Concordance Based on
Valid Conceptual Grounds?

Several conceptual problems in the Concordance were discussed. The underlying relation between patents and invention is obscure not only because of the legal vagaries of patenting but also because the concept of invention is obscure.

There are also problems in assigning patents (or inventions) to industries. First a distinction must be made whether the patent is to be assigned to the industry that does the R&D that originates the patent, the industry that manufactures the patented product, or the industry that uses the patented product or process. With respect to industry of use, there is the problem that many patents have application in multiple industries, some have what might be called ubiquitous use, and some are used solely by consumers.

While the decision rules will be discussed later, it should be noted that the current decision rules are somewhat ambiguous with respect to the distinctions above and therefore are less useful than they could be. Process patents and patents for apparatus combined with a method pose a special problem in terms of industry assignment.

The conceptual problems above appear to be resolvable. No conclusion was reached with respect to the overall conceptual validity, but participants seemed to agree that the Concordance is a reasonable approach to monitoring levels of inventive activity.

Are There More Attractive, Feasible
Alternatives to the Concordance?

Alternatives to the Concordance, which were discussed in some depth, are described in more detail later. In summary, the favored alternative was patent-by-patent assignment of SICs, on the model of the Canadian system. It was believed that patent-by-patent assignment could solve some of the conceptual problems described above. PTO Administrator for Documentation Bill Lawson indicated that there is serious interest in the PTO in a patent-by-patent assignment system. [OTAF Note: This possibility is now being considered as part of an overall review of patent categorization issues related to the PTO's implementation of an automated patent system.]

Questions that need to be answered before possible PTO implementation include: 1) whether Canadian assignments are more accurate than Concordance assignments and 2) whether Canadian assignments are really independent or simply reflect an implicit Concordance. Other questions which would affect the decision include the contribution of SIC categorization to increased effectiveness in patent information dissemination (a PTO mission) and the cost-effectiveness of such a program compared to other steps that could be taken to improve patent information dissemination.

If a patent-by-patent assignment system is implemented, it was agreed that the Concordance should be continued until that system has been operational for a while.

What Changes or Improvements Should Be Made in the Concordance?

Reduce Error Level

Workshop participants agreed that the error level in assignments of subclasses must be reduced if the Concordance is to continue. The favored approach was to identify obvious errors and troublesome subclasses and to correct those. Manually redoing the entire Concordance was not recommended because of the size of the task and the inherent subjectivity of many of the assignments. [OTAF Note: Identifying obvious errors and troublesome subclasses proved to be much more difficult in practice than in concept. Hence the corrections made by the PTO in the 1984 update essentially amounted to redoing most of the Concordance.]

Address the Multiple Counting Problem

Under the then-current decision rules, if a patent subclass was relevant to more than one product field it was assigned to all of them. Thus the patents in a subclass that was assigned to three product fields were counted three times. The rationale for multiple assignment was that for purposes of tracing invention by industry it is important to capture all inventive activity relevant to an industry.

OTAF's review found that the overall multiple counting ratio was 1.41 -- that is, the total patent counts in the unique product fields were 1.41 times the actual number of patents. At the 2-digit level, the multiple count ratio was somewhat lower -- 1.20. For 3- and 4-digit product fields within 2-digit product fields, the ratio varied across product fields, being highest for transportation equipment. For two of five product fields at this level, the ratio was higher for foreign-origin patents than for U.S.-origin patents, and it was increasing over time.

Many inventions are genuinely relevant to multiple product fields. For example, organic molecules synthesized by a chemical firm may have pharmaceutical, herbicidal, and pesticidal effects. Therefore one must build into the Concordance, or other system, the possibility of multiple assignments.

The then-current way of doing multiple assignment, which will be called "multiple counting," however, resulted in several problems, some of which were quite serious. The basic effect was that the data presented at the 3-digit SIC level were not substantially independent. For example, Zvi Griliches reported that only 10% of the patents in the drug industry (product field code 14) were classified only in the drug industry; he reported a similar pattern for the farm and garden machinery industry (product field code 24). He believed it was false for OTAF to publish two data series (pharmaceuticals and agricultural chemicals) which overlap 90% and claim that they are two separate series. He questioned whether industries in which less than half of the time series variance is accounted for by patents that are classified solely into that industry, should be a separate category. Mark Carpenter also noted the problem of overlap of product fields in the research of Computer Horizons, Inc. (CHI).

Dependent data make it difficult to establish statistical relationships between patenting by industry and R&D funding, productivity, or other phenomena of interest. Griliches reported that he was unable to find a relationship between corporate patenting and productivity or R&D funding until he controlled for multiple counting. Mark Carpenter reported that he was unable to get correlations between "science-connectedness" and growth of technologies using either the Concordance or USPC classes.

Multiple counting also appears to lead to a significant overestimation of Japanese and, to a lesser extent, German patenting activity in a number of sectors, particularly in the transportation area. Luc Soete illustrated this problem in the aircraft industry -- SIC 372 or OTAF product field code 54--which is the probably the worst case of "multiple allocation bias." This is the bias introduced when patents are assigned to two or more industries, some of which have a relatively high propensity to patent, such as fabricated metal products, machinery, and engines, and some of which have a very low propensity to patent, such as aircraft. In this case, the latter sector quickly becomes biased in terms of the sheer numbers of the patents from the high-propensity sector.

SPRU's analysis of OTAF code 54 revealed a dramatic rise in aircraft patenting by Japan, a corresponding relative decline by the United States, and a substantial decline by the United Kingdom. Further analysis, however, revealed that these results were very misleading and were directly the result of the assignment of a number of patent subclasses to both motor vehicles and aircraft. The apparently dramatic increase in Japanese aircraft patenting was really due to the Japanese automobile industry. In fact the major companies included in OTAF code 54 are heavily motor vehicle firms; the major aircraft firms come only well after the motor vehicle firms in terms of numbers of patents. The danger of such spurious results is that public policy might be made on the basis of erroneous information.

Another perverse result of the multiple counting is that it tends to give the most weight to those patent subclasses that are most ambiguous in terms of their proper assignment. This is because the decision rule states that in cases where it is unclear to which of several product fields a subclass should be assigned, it should be assigned to all of them. Thus, whereas a subclass that is clearly assignable to a single product field is counted only once, a subclass that is ambiguous may be counted many times.

Various solutions to the multiple counting problem were discussed. Luc Soete described several alternative ways that his groups looked at aircraft patents after discovering the problems with OTAF code 54. The approaches combined weighting with eliminating peripheral or potentially biasing patent subclasses.

First they looked only at those patents that were uniquely assigned to aircraft--i.e., the "core" patents. This showed a drastically different trend than OTAF code 54 as a whole: the significant predominance of the United States continued, the United Kingdom continued as a major foreign patenting country, and Japan was only in sixth place.

Next they weighted all patents in OTAF code 54 by the number of product fields to which they had been assigned. This made only minor changes in the ranking of countries and their growth or decline in importance. Foreign patenting was still high and Japan was still the top foreign country for recent years. This picture was judged to still be distorted.

Then, using the same weights, they left out patent class 123--internal combustion engines--which is relatively less important to the aircraft sector. This time foreign patenting declined, the United Kingdom was the top foreign patenting country, and Japan was in fourth place.

Last, using the same weights, they left out all patent classes which had also been assigned to the motor vehicle field. The resulting picture was somewhere between that with core patents only and that with the weights. Repeating the above series of exercises at the company level resulted in increasing importance of the aircraft firms as compared to motor vehicle or motor engine firms, in the total patent ranking of firms.

In SPRU's view, the last two methods probably produce the most accurate picture of the trends in aircraft patenting behavior over the period considered. In particular the Japanese level is more consistent with what is known about the Japanese aircraft sector.

Another weighting scheme was proposed by Zvi Griliches. He suggested that for subclasses that contain patents relevant to more than one product field -- e.g., drugs and agricultural chemicals -- one should take a sample of the subclass (about 50 patents) and determine what percent are drugs and what percent are agricultural chemicals. If 80% of the sample of patents are drugs and 20% are agricultural chemicals, then the weights would be .8 and .2, respectively. Another proposal, which is discussed later, would use the International Patent Classifications (IPC's) to establish weights for ambiguous subclasses. Under both of these schemes, weights could change over time. [OTAF Note: Since the average subclass size has historically been about 40 patents, a sample of 50 patents may be too large.]

Two alternative ways of getting around the multiple counting problem, which will be discussed in more detail later, were coding into more disaggregated SIC codes and leaving out ambiguous patent subclasses.

By the end of the Workshop there seemed to be agreement that the Concordance should move to fractional assignments in the next update--i.e., a subclass with 2 assignments is weighted by 1/2. Although such a system of weights is essentially arbitrary, it would be a substantial improvement, because it gives more weight to uniquely allocated subclasses and less to ambiguous subclasses. This should result in better indication of the technological performance and

performers within sectors and should help correct for country-of-origin bias, as shown by Soete. The other proposals were not rejected, but since they would involve more drastic alterations of the Concordance, they should probably be delayed for consideration until the future of the Concordance is better known.

Revise the Decision Rules

The decision rules for assigning patent subclasses to "industries" run to the basic purpose for the Concordance. Perhaps the main point made at the Workshop was that the choice of decision rules depends on what one wants to do with the data. Different purposes generally require different decision rules. The two main types of purposes for which the Concordance has been used are to follow trends in inventive activity by "product field" and to correlate with other data series, such as R&D funding and productivity changes.

The original purpose of the Concordance was to help the Science Indicators Unit of NSF trace trends in inventive activity by "product field," a classification used by NSF to report industrial applied research and development funding. The product field essentially classifies product technologies in terms of SIC codes. For example, if a steel company does applied research on refractory bricks, that research is supposed to be reported under the field for refractory bricks, not steel. Likewise, if a patent subclass contains patents for refractory bricks, it would be classified under the product field for refractory bricks, regardless of where the patent originated.

Those criticizing the Concordance decision rules have generally been pursuing other purposes. For example, Zvi Griliches is interested in relating corporate patenting to productivity and growth. He noted that, while the drug companies in his data base have about 5,000 patents in the drug SIC, they have about 6,500 patents in organic chemicals. Therefore the patents in the drug SIC cannot be thought of as having the primary impact on the productivity or sales of the drug industry.

Similarly for the computer firms in their sample the bulk of patents was in instruments (about 5,000), about 4,000 were in electronic components, and about 2,500 in computers. [OTAF Note: Instruments are in OTAF code 55; electronic components are in OTAF code 43; and computers are in OTAF code 27.] The Concordance does not classify the bulk of their patents in the computer industry, which leads to a problem of interpretation from Professor Griliches' point-of-view. The problem of a company's patents being in several product fields results from the broad range of technical activity that is often undertaken by a company, for example the refractory brick work in the steel company above. This problem is exacerbated by firms that are in many lines of business, such as General Electric, which has patents in 245 patent classes.

Mike Scherer used patent data primarily as a device to trace the productivity improvements resulting from R&D, based on the assumption that productivity improvements accrue to those who use the new

technology. He needed to link patents backward to the R&D funding that gave rise to them (industry-of-origin) and forward to the productivity improvements in the companies using them (industry-of-use).

Scherer's industry-of-origin is closer to, but not identical with, OTAF's product field assignment. Because his R&D data were on the Federal Trade Commission (FTC) line-of-business level, his industry-of-origin classification placed patents with the line-of-business under which the company would report the R&D to the FTC. Thus, whereas the OTAF decision rules would classify patents for a machine that welds jet engine turbine blades with welding apparatus, Scherer would classify it with aircraft and aircraft equipment, if the company that held the patent was in that line-of-business. Scherer would also assign codes for industry of use -- in this case, presumably also aircraft and aircraft equipment.

It was generally agreed that it is desirable to classify both by industry of origin and by industry of use. It was also acknowledged that the product field is really neither, but some kind of intermediate thing. There appears to a significant amount of noncompliance with NSF's applied research and development product field survey. One suggestion was to clean up the NSF product field data. Another was to drop the product field as the basis for the Concordance and go to an industry of origin/industry of use basis. Some of the difficulties that would probably be encountered in doing so will be discussed under patent-by-patent assignment.

The OTAF review found that the official decision rules for constructing and updating the Concordance had changed somewhat over the years and were not necessarily the ones actually used. OTAF therefore proposed a "new decision rule" --actually a restatement of the current rules -- that it believed 1) could be applied in a consistent manner and 2) corresponded to what was actually used in previous years. (See Table 11.) This proposal provided the focus for discussion of the current decision rules.

The main subject of discussion pertained to difficult or ambiguous cases. Much of this has been described already under the multiple counting problem. Two of the major recommendations were to use fractional allocation (either based simply on the number of assignments or on a sampling of the patents in the subclass) or the IPCs.

A major portion of the difficult cases are subclasses for processes and for methods which are essentially methods of using apparatus. Based on Mike Scherer's sample, process or method patents account for about 25% of patents. They also appear to be concentrated in particular areas, such as the mechanical subclasses. The different versions of the current decision rule treat process patents differently. One suggestion was to group processes and apparatus into one group. A related suggestion was to make a category that is not assigned a SIC code, on the philosophy that OTAF is better off not pretending to be able to assign subclasses when the odds are no greater than chance that it is correct. No clear conclusion was reached.

The Concordance decision rules ask what kind of establishment would be engaged in producing a product with the structural and functional characteristics represented by the patent subclass. This is based on the idea that the SIC is an establishment classification, as stated in the introduction to the SIC Manual. Don Buzzelli pointed out, however, that in reality, the portion of the SIC Manual we are using is simply a classification of industries based on a classification of products. Thus to ask what kind of establishment would produce the product is an unnecessary step. The decision rule could simply be what kind of product in the SIC Manual embodies the structural and functional characteristics represented by the patent subclass?

One person noted that if R&D funding data were available on an establishment basis, and it were possible to present patent data in a compatible way, that might be the best solution. The patent classification system obliterates information about establishments, however.

The main conclusion with respect to decision rules was that no change should be made in the 1984 update. It was reasoned that by the next update, we may know whether we can implement a patent-by-patent assignment system. If we can, it would be wise not to make a large investment in redoing the Concordance. If not, it may be desirable to reconsider appropriate decision rules for a long-term Concordance. No clear conclusion was reached as to which of the various forms of the current decision rules should be applied to the 1984 update. [OTAF Note: The decision rule actually used in the clean-up and update was the "new decision rule" discussed above, without any limit on the number of SIC assignments.]

Change the Update Process

The possibility was raised that PTO/Documentation might take over the update function, because it has a much broader range of technical expertise than OTAF and is responsible for reclassification. Such a transfer of responsibilities would not occur before the 1984 update. If a patent-by-patent assignment system is adopted and replaces the Concordance, the update will no longer be an issue. If the Concordance is continued, however, with significant changes such as fractional allocations, nonassigned subclasses, and different decision rules, it is possible that the update process might become more complicated.

Establish Additional SIC Fields

OTAF reported on SICs that could possibly be added to the Concordance, but this did not become a major topic of discussion. The issue arose primarily in the context of industry-of-use classification. An industry of use classification necessarily includes nonmanufacturing industries, as evidenced by the assignments of both Scherer and the Canadian system. Inclusion of nonmanufacturing industries is desirable, since North American economies are becoming more heavily based on nonmanufacturing in

terms of GNP, employment, and trade. It also allows more complete tracing of technology flows. [OTAF Note: No SICs were added to the Concordance in the 1984 update.]

Disaggregate Product Fields Further

Luc Soete recommended that at least in certain cases the patent subclasses should be concorded into 4-digit SIC groups. For example, patent subclasses related to motors and engines such as internal combustion engines, power plants, and others could be assigned to an SIC-group defined as SICs 3714, 3724, 3519, 3621, and 3694. In Soete's view, forcing the OTAF data into NSF's 2- and 3-digit R&D codes gives a misrepresentative picture because of the multiple counting problem. He believed that getting down to 4-digit SIC groups is the only way a significant improvement in the Concordance will be made. Each patent class or subclass would be related to 4-digit SICs. Then researchers and others could aggregate the 4-digit SIC groups to make them consistent with the data with which they wish to correlate them. Disaggregation might lead to less multiple counting if it split up internal combustion engines relevant to motor vehicles and those relevant to aircraft, for example. [OTAF Note: The problem of multiple counting is most serious for subclasses where the end use of the invention is not specified and for processes. For such subclasses further disaggregation would likely increase multiple counting.] More disaggregated SICs in the special industrial machinery area would help get at process inventions in the industries that use that machinery.

There seemed to be agreement that to concord into 4-digit SICs across-the-board would be too big a job. No firm conclusion was reached regarding more disaggregated SICs in certain areas. [OTAF Note: No further disaggregation of SICs was attempted in the 1984 update.]

Use Additional Information to Make Assignments

It was widely recognized that the patent classification system is severely limited as a starting point for an industry-based classification. The extent of cross-referencing within the patent classification system indicates the difficulty of assigning patents to a single technology. And as mentioned before, the classification obliterates any information on the origin of the invention. There thus seems to be a need to use information in addition to the USPC to make SIC assignments. Such additional information might be put into the data base and printed on the patent face.

The main suggestion was for the use of the International Patent Classification (IPC) in conjunction with the USPC. U.S. patents have both the USPC and IPC on the front page. Hypothetically, this method would eliminate the necessity of allocating subclasses fractionally. Rather one would allocate patent subclasses crossed with IPC subclasses and they would fractionalize naturally. That is, patents in ambiguous U.S. subclasses would be assigned to product fields on the basis of their IPC assignment.

Bob Evenson reported on a small experiment using USPC 424/300. USPC 424/300 is a chemical subclass pertaining to compositions containing an ester of carbamic or thiocarbamic acid. It contains pharmaceuticals, agricultural chemicals, and other chemicals. Using the Pergamon Patsearch online patent data base, he found that about 60% of the patents in USPC 424/300 were in IPC A01N (agricultural chemicals), 25% were in A61K (pharmaceuticals), and 15% were in C07C (general chemical compositions). When A01N/20, the most frequently occurring IPC subclass, was mapped back into the USPC, 50% mapped back into class 424, 30% into 71 (chemistry, fertilizers), and 20% into 260 (chemistry, carbon compounds). Therefore this experiment was inconclusive as to whether the IPCs can help divide a U.S. subclass into pharmaceuticals, pesticides, and herbicides.

Certain problems were raised with the IPC classification. IPC assignments on U.S. patents are not believed to be highly accurate. Moreover, Derwent has noted that different countries may put different IPC's on the same equivalent patent document.

The question is whether investing in the IPC approach is more effective than going to a sampling basis for weighting or to a patent-by-patent assignment system. It may be worthwhile to work with more subclasses, perhaps the rest of class 424.

Replace the SIC with Some Other Classification

There was some discussion of replacing the SIC with another industrial classification. The SIC has the advantage of widespread use, but it has limitations as well. It is revised infrequently and thus can become out-of-date with respect to new technologies or products. The level of detail available is sometimes not adequate to enable confident placement of patent subclasses. Moreover, the SIC Manual is not always internally consistent.

The FTC line-of-business categories are more detailed than the SIC, and generally map into the SIC. Unfortunately, it is unlikely that R&D data will continue to be collected on this basis.

The possibility of relating patents to an international classification was also discussed. Luc Soete reported that the European Economic Community (EEC) has tried to work out a Concordance between the NACE (Nomenclature Generale des Activites Economiques dans les Communautes Europeennes -- General Industrial Classification of Economic Activities within the European Communities) and the European Patent Classification (EPC). It is not a detailed concordance of patent subclasses into industrial classes, however. An international industrial classification system such as the International Standard Industrial Classification (ISIC) would allow one to make comparisons with the Organization for Economic Cooperation and Development (OECD) industrial R&D data or an international trade classification system such as the Standard International Trade Classification (SITC).

There was no strong support for replacing the SIC at this time.

Other

There was considerable discussion of the suggestion that only certain patent classes or subclasses should be concorded. Zvi Griliches stated that given the amount of uncertainty that will remain, it is not clear that the best way to go is to try to classify all patents into the SIC, where they do not necessarily fit very well. Maybe one should use "indicator" classes which are reasonably good, interesting, and clear. The others could be put somewhere else. Luc Soete's analysis of "core" patents and exclusion of internal combustion engines from aircraft are examples of getting better results by not including everything.

For indicators of inventive activity that are well related to an industrial measure classification, it may be that the smaller number of better fitting facets may provide more information than trying to homogenize the whole thing. This comes back to the point made under multiple counting that there is little point in producing data for 3-digit SIC's if in fact less than 50% of the patents are unique to that classification.

Some participants suggested that for very problematic areas a special category might be created that is not assigned to the SIC. Zvi Griliches argued that OTAF is better off not pretending to be able to allocate subclasses when the odds are no better than chance that it is correct. He suggested that OTAF leave those subclasses in a special category and if the researcher wants to merge or allocate them, let him or her do it. There could be categories like general industrial processes, chemical processes, materials handling, etc. This may be the best way to handle the process problem also -- simply set up a process group with no SIC classification. If this were done there should be an explanation of what the non-SIC matched groupings are. Luc Soete agreed that the above would be a very good system. Researchers-users would likely want to go one step further by using company data to allocate the unassigned segments. That would be their responsibility. [OTAF Note: Creation of special, nonallocated groupings would cause a substantial change from the numbers previously used by NSF. Therefore it is not being recommended for implementation at this time.]

Zvi Griliches called for more warnings about problems like the overlap between drugs and organic chemicals. OTAF could suggest that researchers use different data sets, or combinations of data sets.

What Alternatives Should Be Considered?

Switch to Patent-by-Patent Assignment

The main alternative to the Concordance that was discussed at the Workshop was patent-by-patent assignment of SICs at the time of issue, such as the system developed by the Canadian Patent Office.

Such a system is purported to have several advantages:

- It is less likely to assign patents to SICs where they do not belong. Therefore, technology profiles and other standard analyses can be conducted with more confidence that the patents included are in fact relevant to the SICs of interest.

- An SIC assigned to an individual patent is presumably an independent piece of information and thus provides an additional search or research tool.

- Patent-by-patent assignment avoids the need for periodic updates to keep up with patent reclassification (although SIC revisions can still occur).

- More than one decision rule can be applied -- e.g. industry of manufacture and industry of use

- The system makes possible new or improved patent information products, analyses, and services including:

  - technology flow charts -- i.e., industry-of-origin/ industry-of-use matrices
  - search aids
  - industrial competitiveness studies
  - supplements to the Official Gazette based on SIC
  - lists of abandoned patents by SIC

The main issue discussed with respect to the patent-by-patent system was that of implementation. In the past the feasibility of implementation in the U.S. PTO has been regarded as low because of the additional resources required. Thus, one of the options discussed was that of having the applicant assign SICs. That would require approval from the Office of Management and Budget (OMB), however.

The preferred alternative appeared to be having the PTO assign SICs. As mentioned earlier, Bill Lawson, PTO Administrator for Documentation, indicated that there is serious interest in the PTO in assigning SICs to the patents. The timing is fortunate, he stated, due to the PTO automation effort and this Administration's desire to make the patent system more responsive to industry.

It will not be as easy as it is for the Canadians, who deal with a much smaller number of patents and have a different processing system. The PTO's principal justification for the additional expenditure would be as an information identification and retrieval mechanism. It is important to be able to ascribe categories to patents that will make the information in those patents more available and useful to the public. The SIC can do this because it is used for so many purposes.

The main costs of implementing a patent-by-patent assignment system would be the staff time necessary to make the assignments. There would also be costs related to initial training and disruptions in the paper flow. Another problem would be discontinuity in the data series. Mr. Lawson suggested that both systems should continue in parallel for a while to compare the results of the two. A patent-by-patent assignment would probably not be done retrospectively because of the cost.

Another set of problems that might arise deals with the making of assignments. There was concern that without a quality check the assignments would be inaccurate. Several conceptual issues about SIC assignments would also have to be answered. Should it be SIC of origin,

SIC of use, or both? Should the SIC of origin be based on a company, establishment, product field, or line-of-business basis? Should the SIC of use be the proximate user or the ultimate user? How would processes be classified? There seemed to be agreement that these problems could be worked out.

Mike Scherer described some of the difficulties in making industry of use and industry of origin assignments. He recommended making assignments at a line-of-business level to correspond with the FTC R&D data. Two problems may make that impractical, however. First, it is unlikely that the FTC data will continue to be collected. Second, to link patents to the line-of-business from which they stem would in many cases require additional information on the way firms are organized or keep their books, information the collection of which would require OMB approval.

Although discussion at the Workshop focused primarily on the question of whether we can implement a patent-by-patent system, some questions remain on the wisdom of doing so. As the Canadians themselves admitted, their assignments have not been independently verified. Therefore it is not known with certainty that their assignments are more accurate than a Concordance approach. There is also some question about whether the Canadian assignment is truly independent or whether it reflects an implicit concordance that exists in the heads of those who make the assignments. Some research should probably be done to answer these questions before a final commitment is made to patent-by-patent assignment in the United States.

Use Additional Information to Characterize Technologies

Ways of using additional information to define techologies in terms of patents were described by Mark Carpenter of Computer Horizons, Inc. (CHI). CHI has developed a series of programs that use several different sources of information in addition to USPC subclasses. These include cross-references, key words and key word pairs, examiner citations, and assignee information. The final decision on what is included in the technology is based on a scan of the abstract. By using all this information, CHI appears to be able to identify all the patents relevant to a technology -- e.g., CAT scanner technology -- and to eliminate irrelevant patents.

The method is not particularly effective at picking up process inventions, however. For example, a process invention related to the manufacture of the CAT scanner would probably not be picked up unless the abstract of the patent mentioned that the application was for CAT scanners.

CHI is developing another method of grouping patents into technologies that is called "technology mapping." The initial attempt took the USPC patent classes and clustered them on the basis of cross-references. (See Table 28.) The decision rule for clustering was that within any group of patent classes, there is at least one class that has no patent cross-referenced to any class within another group. For example, in

TABLE 28

CHI's INITIAL RESULT OF CLUSTERING USPC CLASSES USING ORIGINAL AND CROSS-REFERENCE CLASSES

215	BOTTLES AND JARS
150	CLOTH, LEATH, & RUBB. RECEPTACLE
220	METALLIC RECEPTACLES
206	SPECIAL RECEPT, OR PACKAGE
229	PAPER RECEPTACLES
231	WHIPS & WHIP APPARATUS
133	COIN HANDLING
194	CHECK-CONTR. APPARATUS
109	SAFES, BANK PROTECT. & REC DEV
232	DEPOSIT & COLLECT. RECEPT.
140	WIREWORKING
245	WIRE FABRICS & STRUCTURE
191	ELECT., TRANSMIS. TO VEHICLES
104	RAILWAYS
105	RAILWAY ROLLING STOCK
238	RAILWAYS, SURFACE TRACK
246	RAILWAY SWITCHES & SIGNALS
14	BRIDGES
182	FIRE ESCAPES, LADDERS, SCAFFOLDS
187	ELEVATORS
212	TRAVERSING HOISTS
254	PUSH. & PULL. IMPLEMENTS
16	MISCELLANEOUS HARDWARE
49	MOVABLE OR REMOVABLE CLOSURES
160	CLOSURES, PARTIT. & PANELS ...
181	ACOUSTICS
256	FENCES
258	RAILWAY MAIL DELIVERY
13	ELECTRIC FURNACES
75	METALLURGY
266	METALLURGICAL APPARATUS
11	BOOKS, MAKING
270	SHEET-MAT. ASSOC. OR FOLD.

Source: Figure 13 in "Use of the SIC-USPC Concordance at CHI Research," by Mark P. Carpenter, delivered at the SIC-USPC Concordance Workshop, November 18, 1983.

Table 28 the group that includes railways has at least one class that does not have a patent that is cross-referenced to any patent in the next group -- i.e., bridges, fire escapes -- or any of the other groups. The resulting set of groups has a certain amount of face validity, although there are several anomalies.

During discussion of technology mapping, the question was raised of whether the decision rule was capable of defining a unique set of groupings. The question was also raised of how one would use technology groupings based on aggregated patent classes such as this. Luc Soete, whose group has developed a similar system, commented that they are not useful for economic analysis but primarily from an engineering perspective.

There seemed to be agreement that the methods described above should be pursued further, although they will probably not serve the purposes of Science Indicators. The technology mapping method may be particularly promising if some basic questions regarding validity can be resolved.

Use Only the USPC to Characterize Technologies

An alternative posed by the Workshop organizers, chiefly as a "straw man," was to use the USPC alone to characterize technologies for Science Indicators. Most of the comments on this alternative were negative.

There seemed to be agreement that the patent classification system as presently constituted is not useful for somebody interested in industries or products. Zvi Griliches noted that patent classes vary with respect to the scope of technology covered and "uniqueness," in the sense of extent of cross-referencing. The ratio of cross-references to original references differs widely from subclass to subclass. The ultimate uncertainty is at the subclass level, where the connection to industries or products is not straightforward.

CHI reported that relatively well-defined technologies consisting of about 200 patents were often spread across several patent classes and many subclasses. At the patent class level, the categories are too broad. Horology, for example, includes two very different components -- the old, stagnant mechanical timepieces and the rapidly growing, foreign-dominated digital electronic watches.

This results in an inability to identify patents relevant to industries or products in an unambiguous manner. Zvi Griliches reported that one of his students found that about 26% of the patents in a PTO report on CAT scanners had nothing to do with CAT scanners, and that the report missed about 30% of the CAT scanner patents found in one year's issues of the Official Gazette.

There was some discussion of the possibility of revising the USPC. Zvi Griliches commented that in the long run, the right place for the PTO to invest its effort is in improving the classification system and putting more information on patents. He thought that this might be useful even

for meeting the needs of PTO examiners, noting that there are some broad areas of the USPC that could be subdivided and are not. [OTAF Note: Reclassification is a large and continuing PTO activity that addresses this problem. As noted earlier, patent categorization issues are receiving a major review as part of PTO implementation of an automated patent system.]

Use the Derwent Patent Classification System
as a Basis for the Concordance

The possibility of using the Derwent patent classification system as the basis for a Concordance was also considered at the Workshop. The claim was made that the Derwent system is more industry-oriented than other patent classification systems. This is presumably because their clients, major industrial patenting organizations, demand to see all patent documents relevant to their areas of interest and do not want to see irrelevant documents.

The Derwent classification system has 280 categories. For example, one section deals with the petroleum industry and has classes that start with exploration and go through extraction to downstream processing. The classification rule is "Who among our clients would be interested in this?" Thus a machine for welding turbine blades would go in with welding machines and maybe with impellers. If it discloses such a wide range of use that it cannot be put in all specific areas, it is put in the more general area.

The Derwent classification is assigned by Derwent staff who look at the documents themselves. Thus patents from all over the world are looked at by one group of technically competent people.

Derwent, Inc., has indicated that it is willing to talk about the possibility of using its classification system as the basis for a Concordance. There are some questions about the terms of availability, however. The system is intended for retrieving specific documents, rather than statistical analysis, and contains items of proprietary information. One workshop participant commented that it probably has many of the same problems as other patent classifications. An important advantage may be its coverage of international patent "families."

Other Issues Addressed by the Workshop

Researchers pointed to the importance of keeping track of whether a patent is still in force now that renewal fees must be paid. The PTO will collect data on an ongoing basis on who pays their maintenance fees and who does not. Data on maintenance fees will be stored on PTO computer files and should be available to the public. Commercial vendors of patent information are also expected to pick it up. This will require an update process on the patent bibliographic records, but it will not present any particular problem.

Other new information is being collected. Since October 1, 1982, data have been collected on whether the assignee is a small business (defined as having fewer than 500 employees), a nonprofit organization, or an individual.

Researchers stated that it would be helpful if OTAF would keep the same company code number for the same company with each update of the file. Currently, each update changes the numbers of the companies so researchers must redo their tapes from scratch. OTAF reported that there is a possibility of changing the company code procedure to something like the Dun and Bradstreet system, which permits assignment of a unique number that can stay the same, even if companies on both sides of it change. [OTAF Note: This suggestion is currently under evaluation. No change was made in the company code procedure for the 1984 update.] There was also a request for consolidated company code numbers that would help identify subsidiaries and mergers.

Researchers also reported problems with disclosures of government support. Mike Scherer found that patents contained such disclosures in only one of every three or four cases where they should have.

Attention was called to the upcoming revision of the SIC. There was agreement that we should avoid making a large investment in a new or revised Concordance, only to have to do it again for a new SIC manual. The sentiment was expressed that NSF's Science Resources Studies Division should be represented in the SIC revision effort. [OTAF Note: Publication of a revised SIC manual is scheduled for July 1, 1986. It is expected that certain industries such as instruments, electronics, and nonmanufacturing will be fairly extensively changed.]

The need was expressed for a federal statistical policy or standardized cross-federal statistical program. The main policy today is to reduce the reporting burden on the private sector.

PART III

RECOMMENDATIONS

The following recommendations are based on OTAF's review and assessment of the Concordance and the discussion at the Workshop.

General Recommendations

The Concordance should be continued at least until a superior system of linking patents to industries is developed and implemented.

The possible inefficiencies of acting before the SIC Manual is revised need to be weighed in any decision to make major changes in the Concordance or implement alternative systems.

Discussion

The Concordance should be continued for the time being because it provides the only detailed assignments of U.S. patents to product fields that is continuous, covers all technologies, and has a long time series. Users such as Science Indicators and SPRU are dependent on the Concordance in the absence of anything better.

There is considerable uncertainty concerning the upcoming SIC revision. A revised SIC Manual is scheduled to be published on July 1, 1986. Substantial changes are expected in some areas such as instruments, electronics, and nonmanufacturing. These changes, the exact nature of which cannot be predicted, may require extensive changes in patent assignments. In the past, however, revisions to the SIC have been delayed. Moreover, if an opportunity presents itself to implement a significant improvement in the assignment of SIC's to patents, that in itself is a strong argument to act now.

The Concordance in the Immediate Future

Immediate improvements needed in the Concordance include correction of obvious errors in assignment and proportional weighting of subclasses that are assigned to more than one product field.

The update process should be transferred from OTAF to PTO/Documentation. New or revised subclasses should be assigned to product fields at time of reclassification and updated SIC tapes prepared annually.

Discussion

OTAF has already completed a major effort to correct errors in assignment, which is reflected in the 1984 update. Proportional weights have been assigned according to the number of product fields to which a patent subclass is assigned -- i.e., a subclass with assignments to two unique product fields is counted one-half in both.

The update function is in the process of being transferred from OTAF to PTO/Documentation. Documentation, which has a larger staff with a broad range of technical expertise, can conduct the update in conjunction with its reclassification function. This has the advantage of using the classifier's intimate knowledge of the subclasses being reclassified and of making the new product field assignments in a more timely fashion.

Patent-by-Patent Assignment of SICs

The PTO should seriously consider the possibility of implementing a patent-by-patent system of assigning SICs, modeled after the Canadian system. Both industry-of-origin and industry-of-use assignments should be made.

Principal responsibility for implementing patent-by-patent assignment should rest with PTO/Documentation.

The Canadian system should be evaluated prior to PTO implementation. Important questions include the relative accuracy of Canadian assignments compared to a Concordance approach and the extent to which they are independent of the patent classification system. This evaluation should be conducted by an outside group.

If patent-by-patent assignment is successfully implemented, both systems (patent-by-patent and the Concordance) should run in parallel for some period of time. Then, based on experience with both systems, a decision can be made whether to discontinue the Concordance.

Discussion

Assignment of SICs to patents is currently being considered by the PTO as part of a review of patent categorization issues instigated by the implementation of an automated patent system. Because the utility of such data depends largely on long time series, it is desirable to act as soon as possible. As noted above, however, implementing patent-by-patent assignment before the SIC is revised may involve some inefficiencies.

In addition to a better evaluation of the Canadian system, it is necessary to consider the costs and benefits of SIC assignments compared to other possible PTO programs to improve the availability and utility of patent data. Arguments in favor of SIC assignments would be strengthened if evidence can be found that the SIC makes patent data significantly more useful to the public.

The Concordance in the Longer Term

If the Concordance is continued in the longer term, more fundamental changes may be desirable.

Discussion

In the period while the long-term continued existence of the Concordance is uncertain but still a possibility, it would be useful for the concerned community to investigate more fundamental changes to the Concordance or alternative methods of relating patents to industrial, economic, or technological data. Some of the subjects that could be investigated include:

- Alternative decision rules for the Concordance
- Effect of using OR's versus OR's plus XR's
- Use of the IPC to determine weightings for patent subclasses that are counted more than once
- "Technology mapping" -- i.e., using a variety of patent information to characterize technologies
- Use of the Derwent patent classification system
- Eliminating troublesome subclasses or creating special groups that are not allocated to the SIC
- Development of an international patent-industrial classification Concordance

PART IV

The 1984 Update

Previous Updates

In previous updates the OTAF staff identified the subject matter covered in the new subclasses using the Manual of Classification and class definitions. Classification orders, the documentation for reclassification projects, were also consulted since they usually indicate the abolished subclasses which are the sources of the established (new) subclasses. In many cases the update process involved determining the product fields assigned to the abolished subclass and making the same assignment to the newly established subclass.

However, classification orders do not always show a one-to-one correspondence between abolished and established subclasses. The source of the established subclass may be several abolished subclasses with several SIC product field assignments. In such cases, the proper product field assignment may not be obvious from the classification order.

In other situations, the updater may decide that the product field assignment of the abolished subclass was not correct for the established subclass, either because of error or because the technology has been redefined. Decisions are complicated if the updater is not familiar with the technology encompassed by the established subclass.

The 1984 Concordance Update

The primary purpose of the 1984 Concordance update was to add to the Concordance those subclasses which had been established in 1982 and 1983. The process entailed identifying subclasses in existence as of the end of 1983 which had no SIC product fields associated with them. Since the update performed in 1982 had assigned product fields to all subclasses in existence at the end of 1981, this procedure provided a listing of those subclasses which had been established during 1982 and 1983.

As a first step in upgrading the process, OTAF used three Classifiers, one from each technology discipline (Chemical, Electrical and Mechanical), to do the 1984 update. These Classifiers, from the PTO's Documentation Branch, have experience in reclassification procedures which span their particular disciplines. The newly established subclasses were divided by discipline, and each Classifier updated subclasses in his discipline. In those cases where the Classifier was unfamiliar with the technology, a more knowledgeable Classifier was consulted.

The Decision Rules

The decision rules used to make subclass assignments were modified slightly from those previously published and additional guidelines were instituted. However, the "philosophy" adopted was essentially the same as that used during previous projects.

Following the decision rules for assigning SIC product fields to USPC subclasses is, in the majority of cases, straightforward. For subclasses which identify a particular product or apparatus (and most do) the assignment is the same regardless of which published version of the decision rules is followed. Problems arise, however, when 1) ambiguity exists as to which of several possible SIC groups to assign to a patent classification, and 2) the subclasses identify a process or method of use.

In the first situation, the "ambiguity rule" which appears to have been followed since the Concordance began was observed. Thus, if ambiguity existed, a subclass was assigned all SIC groups possible. However, in this update and review, the number of SIC groups was limited by reasonable interpretation of the subclass content. Assignors were instructed not to think of extreme or unusual examples of products or apparatus falling within the scope of the subclass. No definite limit was set on the number of possible assignments. However, only very few subclasses were assigned more than four SIC product fields.

It appears that process subclasses have been previously assigned SIC product fields by determining whether they were more closely related to a corresponding product or apparatus. In reviewing process subclasses for errors, the three guidelines listed below were followed to determine "more closely related." The theory behind these guidelines is: when subclasses encompass both process and product, the process is more closely related to the product than the apparatus; when subclasses encompass both process and apparatus, the process is more closely related to the apparatus.

1) If a subclass includes process only or process and product place it a) with the product, if known, b) with the apparatus if the product is not known, or c) in the "All Other" category if neither product nor apparatus is known.

2) If a subclass includes both process and apparatus, place it a) with the apparatus if known, b) with the product if the apparatus is not known, or c) in "All Other" if neither the apparatus or product is known.

3) In the rare case that a subclass includes process and apparatus but there is a specific subclass elsewhere for the apparatus, treat the subclass as "process only."

Error Correction

In addition to assigning SIC product fields to new subclasses, the 1984 update involved correction of errors in previous SIC assignments. The corrections were done by the three Classifiers who performed the update

in conjunction with the OTAF staff. The first approach was to locate areas of clear error by examining lists of subclass ranges within product fields and correcting obvious errors.

This approach was abandoned for several reasons. First, errors were not obvious. In many cases, subclass definitions had to be consulted to determine their content because class and subclass titles were not sufficient. Secondly, since product fields include subclasses of different disciplines, they could not be distributed among the three Classifiers. More importantly, classes often appeared in many product fields, requiring multiple looks at the same class definitions.

It was discovered that once the Classifiers had gone into the class definitions it was best to go ahead and look at all the subclasses within that class. As a result, it was decided to look for subclass assignment errors on a class-by-class basis, the classes being divided by discipline and examined by the Classifiers on that basis.

All of the approximately 350 USPC classes were examined to correct subclass assignment errors. Because of the large number of mechanical classes, not all of these were examined with the same detail as others. Those where there were suspected problems, however, were examined first.

The update and error correction procedures resulted in about 35,000 changes to subclass assignments.

Code Changes

OTAF's review of the Concordance showed assignment errors that occurred because of coding mistakes. That is, the person assigning SIC's to subclasses recorded the wrong OTAF code. In some cases, subclasses were assigned roll-up codes without being assigned unique codes within the roll-up. In other cases, errors were caused because of similarities between the OTAF codes (1-55) and SIC product field numbers (20, 22, 28, etc.)

To eliminate these problems, the product fields have been recoded so that the codes used by those making subclass assignment correspond to SIC product field numbers. New codes are shown on Table 29. Former codes 1-55 are not used in making subclass assignments, although they are still used in the computer programs as sequence numbers. Roll-ups are designated by R1-R14 but actually have no OTAF code. Subclasses cannot be assigned these roll-up designations. In the traditional version of the report, subclasses are counted in the roll-up product fields by computerized function which combines all subclasses of the unique product fields within them and eliminates duplication.

TABLE 29

STANDARD INDUSTRIAL CLASSIFICATION PRODUCT FIELDS
1984 UPDATE

SEQUENCE NUMBER	PRODUCT FIELD	Product Fields	SIC Code
1	20	FOOD AND KINDRED PRODUCTS	20
2	22	TEXTILE MILL PRODUCTS	22
3	R1	CHEMICALS AND ALLIED PRODUCTS	28
4	R2	Chemicals, Except Drugs & Medicines	281, 282, 284-289
5	R3	Basic Industrial Inorganic & Organic Chemistry	281, 286
6	281	Industrial Inorganic Chemistry	281
7	286	Industrial Organic Chemistry	286
8	282	Plastics Materials & Synthetic Resins	282
9	287	Agricultural Chemicals	287
10	R4	All Other Chemicals	284, 285, 289
11	284	Soaps, Detergents, Cleaners, Perfumes, Cosmetics & Toiletries	284
12	285	Paints, Varnishes, Lacquers, Enamels, & Allied Products	285
13	289	Miscellaneous Chemical Products	289
14	283	Drugs and Medicines	283
15	1329	PETROLEUM & NATURAL GAS EXTRACTION & REFINING	13, 29
16	30	RUBBER & MISCELLANEOUS PLASTICS PRODUCTS	30
17	32	STONE, CLAY, GLASS AND CONCRETE PRODUCTS	32
18	R5	PRIMARY METALS	33, 3462, 3463
19	331+	Primary Ferrous Products	331, 332, 3399, 3462
20	333+	Primary & Secondary Non-Ferrous Metals	333-336, 339 (except 3399), 3463
21	34-	FABRICATED METAL PRODUCTS	34 (except 3462, 3463, 348)
22	R6	MACHINERY, EXCEPT ELECTRICAL	35
23	351	Engines & Turbines	351
24	352	Farm & Garden Machinery & Equipment	352
25	353	Construction, Mining & Material Handling Machinery & Equipment	353
26	354	Metal Working Machinery & Equipment	354
27	357	Office Computing & Accounting Machines	357
28	R7	Other Machinery, Except Electrical	355, 356, 358, 359
29	355	Special Industry Machinery, Except Metal Working Machinery	355
30	356	General Industrial Machinery & Equipment	356
31	358	Refrigeration & Service Industry Machinery	358
32	359	Miscellaneous Machinery, Except Electrical	359
33	R8	ELECTRICAL AND ELECTRONIC MACHINERY, EQUIPMENT AND SUPPLIES	36, 3825
34	R9	Electrical Equipment, Except Communication Equipment	361-364, 369, 3825
35	361+	Electrical Transmission & Distribution Equipment	361, 3825
36	362	Electrical Industrial Apparatus	362
37	R10	Other Electrical Machinery, Equipment & Supplies	363, 364, 369
38	363	Household Appliances	363
39	364	Electrical Lighting & Wiring Equipment	364
40	369	Miscellaneous Electrical Machinery, Equipment & Supplies	369
41	R11	Communication Equipment & Electronic Components	365-367
42	365	Radio & Television Receiving Equipment Except Communication Types	365
43	366+	Electronic Components & Accessories & Communication Equipment	366-367
44	R12	TRANSPORTATION EQUIPMENT	37, 348
45	R13	Motor Vehicles & Other Transportation Equipment, Except Aircraft	348, 371, 373-376, 379
46	371	Motor Vehicles & Motor Vehicle Equipment	371
47	376	Guided Missiles & Space Vehicles & Parts	376
48	R14	Other Transportation Equipment	373-375, 379 (except 3795)
49	373	Ship & Boat Building & Repairing	373
50	374	Railroad Equipment	374
51	375	Motorcycles, Bicycles & Parts	375
52	379-	Miscellaneous Transportation Equipment	379 (except 3795)
53	348+	Ordnance Except Missiles	348, 3795
54	372	Aircraft & Parts	372
55	38-	PROFESSIONAL & SCIENTIFIC INSTRUMENTS	38 (except 3825)
56	99	ALL OTHER SIC's	
57	R15	ALL INDUSTRIES	

The Update Report

Using information from the 1984 Update, OTAF produced "Indicators of the Patent Output of U.S. Industry, 1963-1983," referred to as the 1983 Indicators Report. This report reflects coding changes made during the update and adds data for 1982 and 1983. Like the previous Indicators Report, which contained data through 1981, this report considers only the original classification of the patent and distributes data into 55 SIC product fields. No new product fields have been added, nor has the scope of the product fields been changed.

OTAF produced two versions of the 1983 Indicators Report. One version was generated like previous reports with patents counted in every product field assigned to its original classification. In this case, multiple counting among product fields occurs.

The other version eliminates multiple patent counts among product fields through a system of fractional counting. This system divides each patent count by the number of product fields to which it is assigned. Then only the resulting fraction is counted in each product field. Except for minor differences due to rounding error, the counts in the combined product fields (roll-ups) equal the total of their subcomponents, and unique product fields add to the total number of patents included in the report.

Table 30 shows the number of 1963-1983 patents in each product field for both report versions. In the fractional report, the number of patents in each of the 41 unique product fields adds to a total of 1,345,248, which is the actual number of patents granted between 1963 and 1983 minus those patents for which the database has no original classification information. In the report which is not fractionalized, the number of patents in the unique product fields adds to 1,926,940 which, as a result of multiple counting, is 43% higher than the actual number of patents.

Future Updates

PTO's Documentation Branch will be responsible for future updates. SIC product fields will be assigned after each reclassification project which establishes new subclasses. They will be made by the Classifiers who worked on the 1984 update in consultation with the project Classifier who has just completed the reclassification project. The project Classifier will serve as the expert in the technology. The Classifier who performed the 1984 update will serve as the expert in SIC philosophy and previous assignment practice.

TABLE 30

Number of Patents and Number of Subclasses
by SIC Product Field: 1983 Concordance

SEQ. NO.	Product Fields	#Patents 1963-1983 Fractional	#Patents 1963-1983 Not Fractional	# Subclasses*
1	FOOD AND KINDRED PRODUCTS	10,493	10,610	597
2	TEXTILE MILL PRODUCTS	8,937	13,375	875
3	CHEMICALS AND ALLIED PRODUCTS	200,443	218,535	---
4	Chemicals, Except Drugs & Medicines	182,152	211,168	---
5	Basic Industrial Inorganic & Organic Chemistry	117,205	123,619	---
6	Industrial Inorganic Chemistry	22,701	26,963	1,492
7	Industrial Organic Chemistry	94,505	101,094	5,879
8	Plastics Materials & Synthetic Resins	21,618	26,546	1,604
9	Agricultural Chemicals	14,291	24,272	490
10	All Other Chemicals	29,038	38,065	---
11	Soaps, Detergents, Cleaners, Perfumes, Cosmetics & Toiletries	6,769	7,124	264
12	Paints, Varnishes, Lacquers, Enamels, & Allied Products	9,232	16,287	1,137
13	Miscellaneous Chemical Products	13,038	14,679	882
14	Drugs and Medicines	18,290	31,853	1,019
15	PETROLEUM & NATURAL GAS EXTRACTION & REFINING	16,183	17,127	731
16	RUBBER & MISCELLANEOUS PLASTICS PRODUCTS	53,499	102,635	7,256
17	STONE, CLAY, GLASS AND CONCRETE PRODUCTS	24,626	41,859	2,786
18	PRIMARY METALS	14,149	16,900	---
19	Primary Ferrous Products	7,135	11,746	1,047
20	Primary & Secondary Non-Ferrous Metals	7,015	9,582	662
21	FABRICATED METAL PRODUCTS	115,414	167,520	16,618
22	MACHINERY, EXCEPT ELECTRICAL	333,620	394,058	---
23	Engines & Turbines	12,172	30,612	2,435
24	Farm & Garden Machinery & Equipment	20,042	41,303	4,954
25	Construction, Mining & Material Handling Machinery & Equipment	43,114	67,989	7,164
26	Metal Working Machinery & Equipment	41,996	51,735	5,697
27	Office Computing & Accounting Machines	30,488	36,239	3,359
28	Other Machinery, Except Electrical	185,809	248,207	---
29	Special Industry Machinery, Except Metal Working Machinery	75,653	96,382	10,943
30	General Industrial Machinery & Equipment	78,816	143,233	12,830
31	Refrigeration & Service Industry Machinery	22,164	49,739	4,371
32	Miscellaneous Machinery, Except Electrical	9,177	24,508	2,058
33	ELECTRICAL AND ELECTRONIC MACHINERY, EQUIPMENT AND SUPPLIES	231,032	255,251	---
34	Electrical Equipment, Except Communication Equipment	96,791	126,108	---
35	Electrical Transmission & Distribution Equipment	30,500	43,109	2,566
36	Electrical Industrial Apparatus	33,660	44,986	3,327
37	Other Electrical Machinery, Equipment & Supplies	32,631	48,626	---
38	Household Appliances	11,779	24,679	2,271
39	Electrical Lighting & Wiring Equipment	11,364	12,846	1,426
40	Miscellaneous Electrical Machinery, Equipment & Supplies	9,489	11,108	869
41	Communication Equipment & Electronic Components	134,239	151,227	---
42	Radio & Television Receiving Equipment Except Communication Type	11,687	20,450	1,423
43	Electronic Components & Accessories & Communication Equipment	122,553	146,958	8,949
44	TRANSPORTATION EQUIPMENT	69,622	220,891	---
45	Motor Vehicles & Other Transportation Equipment, Except Aircraft	54,807	84,534	---
46	Motor Vehicles & Motor Vehicle Equipment	27,241	58,380	5,059
47	Guided Missiles & Space Vehicles & Parts	1,443	2,921	108
48	Other Transportation Equipment	18,208	26,548	---
49	Ship & Boat Building & Repairing	5,254	5,767	426
50	Railroad Equipment	4,580	9,557	1,521
51	Motorcycles, Bicycles & Parts	2,006	8,594	745
52	Miscellaneous Transportation Equipment	6,369	16,096	1,540
53	Ordnance Except Missiles	7,916	8,398	815
54	Aircraft & Parts	14,814	40,297	2,916
55	PROFESSIONAL & SCIENTIFIC INSTRUMENTS	154,632	170,732	10,866
99	ALL OTHER	112,593	137,050	14,115

*Active subclasses in the OTAF data base as of December 31, 1983. Number of subclasses not recorded for "roll-up" categories.

APPENDIX A

THE PATENT DATA BASE OF THE OFFICE OF TECHNOLOGY ASSESSMENT AND FORECAST

The computerized base of data relating to the U.S. patent file includes, at present:

- all subclasses of the U.S. Patent Classification (USPC) System, and the classification within this System of all U.S. patents, including utility and design patents
- the relationship of all utility subclasses in the U.S. Patent Classification System to 55 Product Fields and combinations of Product Fields in the Standard Industrial Classification (SIC) System
- the category of ownership at time of issue, e.g., U.S. Government, foreign government, U.S. corporation, foreign corporation, U.S. individual, foreign individual (for utility patents issued since 1963 and for design patents issued since 1977)
- the country or state of residence of the inventor (for utility patents issued since 1963 and for design patents issued since 1977)
- the date the application for patent was filed in the United States (for utility patents issued since 1967 and for design patents issued since 1977)
- the specific (i.e., named) ownership of all patents which, at time of issue, were owned by an organization (for utility patents issued since 1969 and for design patents issued since 1977)
- the patent title (for utility patents issued since 1969 and for design patents issued since 1977)
- the name and address of inventors of unassigned patents (for utility patents issued since 1975 and for design patents issued since 1977)
- the field of search and references cited in the examination leading to the patent grant (for utility patents issued since 1975 and for design patents issued since 1977)

8/26/83

APPENDIX B

SAMPLE PROFILE REPORT OF PRODUCT FIELD 1 (SIC 20) FOOD AND KINDRED PRODUCTS

This profile is from the 1981 SIC Report which the Office of Technology Assessment and Forecast prepared for the National Science Foundation. It is the first of the 55 reports, one for each product field, contained in the 1981 SIC Report.

SIC 20. FOOD + KINDRED PRODUCTS

1979-81 AVERAGES	
PERCENT GROWTH	26.41
PERCENT FOREIGN	39.40
PERCENT CORPORATE OWNED	80.61
PERCENT GOVERNMENT OWNED	2.43
PERCENT U.S. OWNED OF FOREIGN	8.81
1979-81 XR/OR	0.00

DEFINITIONS

PERCENT GROWTH = 1979-81 PATENTS/1972-81 PATENTS X 100.

PERCENT FOREIGN = 1979-81 U.S. PATENTS WITH A FOREIGN RESIDENT INVENTOR/1979-81 PATENTS X 100

PERCENT CORPORATE OWNED = 1979-81 U.S. PATENTS ASSIGNED TO CORPORATIONS/1979-81 PATENTS X 100.

PERCENT GOVERNMENT OWNED = 1979-81 U.S. PATENTS ASSIGNED TO THE U.S. AND FOREIGN GOVERNMENTS/1979-81 PATENTS X 100.

PERCENT U.S. OWNED OF FOREIGN = 1979-81 U.S. PATENTS WITH A FOREIGN RESIDENT INVENTOR THAT ARE ASSIGNED TO A U.S. ORGANIZATION/1979-81 U.S. PATENTS WITH A FOREIGN RESIDENT INVENTOR X 100.

1979-81 XR/OR = 1979-81 U.S. PATENTS WITH A CROSS REFERENCE DESIGNATION/1979-81 U.S. PATENTS WITH AN ORIGINAL PRIMARY DESIGNATION.

SIC 20. FOOD + KINDRED PRODUCTS

PATENT ACTIVITY (1/63-12/81) BY DATE OF PATENT GRANT

NUMBER OF PATENTS

	63-67	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	TOTAL
TOTAL	2778	577	697	681	974	982	844	917	1031	966	793	647	531	783	904	14105
U.S. ORIGIN	2207	459	539	509	688	703	596	584	635	590	489	382	337	465	542	9725
FOREIGN ORIGIN	571	118	158	172	286	279	248	333	396	376	304	265	194	318	362	4380
JAPAN	153	34	49	41	112	125	128	140	156	101	92	86	64	126	127	1534
WEST GERMANY	93	13	23	26	39	35	16	36	51	56	35	36	33	42	55	589
UNITED KINGDOM	68	9	15	21	26	24	23	39	55	39	37	33	18	29	27	463
FRANCE	42	8	12	14	19	18	11	24	22	32	26	15	12	24	26	305
CANADA	50	10	11	17	16	12	19	18	8	15	20	13	10	15	14	248
SWITZERLAND	26	5	7	8	6	7	8	17	26	33	20	22	15	21	24	245
NETHERLANDS	28	13	9	13	22	13	9	11	22	17	19	15	10	12	19	232
SWEDEN	19	4	1	9	6	8	4	7	11	24	4	2	8	11	10	128
ITALY	18	4	3	1	12	7	4	8	9	15	7	5	4	3	16	116
DENMARK	10	3	2	3	5	4	4	7	2	8	6	7	5	6	9	81
U.S.S.R.	2	1	1	1	2	2	3	6	4	3	6	5	3	6	9	54
AUSTRALIA	6	2	1	3	3	1	3	2	6	6	2	6		2	5	48
BELGIUM	3		3	4	2	7	2	1	4	1	5			4	1	37
AUSTRIA	7	1	3	1		1	2	4	4	3	2		1	2	4	35
ISRAEL	5			2	1	5	1	1	1	3	2	4		4	1	30
MEXICO	10	4	1	1	3	1				3		1	1	1	2	28
CZECHOSLOVAKIA	3	1	5	1	1					3	1	2	2	1	1	21
NORWAY	2	1	2	1	4	1	1	1	2	1	2	2				20
HUNGARY	1		1	1		1		2	3	3	2		1		3	18
SPAIN	2					1		1	4	3	1	1	1	1	1	16
FINLAND						2		3	1	1	3	1		1		12
NEW ZEALAND	3								1	1	1	1			4	11
ARGENTINA	2		1	1				1		1			2	2	1	11
POLAND			2						1	1		1	3	1		9
S. AFRICA		1			1	1	1	2			1	1				8
IRELAND	1			1							1	1		1	1	6
BULGARIA							2				2		1	1		6
PORTUGAL	1	2	1			1									1	6
CHILE					2				1					1		4
GUATAMELA	2						1				1					4
INDIA			1							1	1	1				4
CHINA (MAINLAND)	1					1	2									4
HONDURAS			3													3
BRAZIL				1		1		1								3
PHILLIPINES									1		1	1				3
OTHER CNTRIES (22)	13	2	1	1	4		4	1	1	2	4	3		1	1	38
U.S. ORIGIN	2207	459	539	509	688	703	596	584	635	590	489	382	337	465	542	9725
U.S. CORP. OWNED	1639	342	425	375	551	541	438	454	490	466	383	297	246	368	419	7434
U.S. GOVT. OWNED	84	20	18	17	23	35	34	25	34	22	16	19	13	9	16	385
U.S. INDIV. OWNED	475	94	94	110	109	116	113	98	104	91	82	64	73	83	103	1809
FOREIGN OWNED	9	3	2	7	5	11	11	7	7	11	8	2	5	5	4	97
FOREIGN ORIGIN	571	118	158	172	286	279	248	333	396	376	304	265	194	318	362	4380
U.S. OWNED	66	17	21	21	39	26	17	35	47	37	32	22	14	29	34	457
FOREIGN OWNED	505	101	137	151	247	253	231	298	349	339	272	243	180	289	328	3923
FOREIGN CORP.	384	86	105	119	213	204	195	247	307	285	227	199	155	244	267	3237
FOREIGN GOVT.	2			2		1			3	5	1	3	1	8	7	33
FOREIGN INDIV.	119	15	32	30	34	48	36	51	39	49	44	41	24	37	54	653

SIC 20. FOOD + KINDRED PRODUCTS

PATENT ACTIVITY (1/63-12/81) BY DATE OF PATENT GRANT

	PERCENT OF PATENTS															
	63-67	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	TOTAL
TOTAL	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
U.S. ORIGIN	79	80	77	75	71	72	71	64	62	61	62	59	63	59	60	69
FOREIGN ORIGIN	21	20	23	25	29	28	29	36	38	39	38	41	37	41	40	31
JAPAN	6	6	7	6	11	13	15	15	15	10	12	13	12	16	14	11
WEST GERMANY	3	2	3	4	4	4	2	4	5	6	4	6	6	5	6	4
UNITED KINGDOM	2	2	2	3	3	2	3	4	5	4	5	5	3	4	3	3
FRANCE	2	1	2	2	2	2	1	3	2	3	3	2	2	3	3	2
CANADA	2	2	2	2	2	1	2	2	1	2	3	2	2	2	2	2
SWITZERLAND	1	1	1	1	1	1	1	2	3	3	3	3	3	3	3	2
NETHERLANDS	1	2	1	2	2	1	1	1	2	2	2	2	2	2	2	2
SWEDEN	1	1		1	1	1		1	1	2	1		2	1	1	1
ITALY	1	1			1	1		1	1	2	1	1	1		2	1
DENMARK		1			1			1		1	1	1	1	1	1	1
U.S.S.R.								1			1	1	1	1	1	
AUSTRALIA									1	1		1			1	
BELGIUM				1		1					1			1		
AUSTRIA																
ISRAEL						1						1		1		
MEXICO		1														
CZECHOSLOVAKIA			1													
NORWAY																
HUNGARY																
SPAIN																
FINLAND																
NEW ZEALAND																
ARGENTINA																
POLAND													1			
S. AFRICA																
IRELAND																
BULGARIA																
PORTUGAL																
CHILE																
GUATAMELA																
INDIA																
CHINA (MAINLAND)																
HONDURAS																
BRAZIL																
PHILLIPINES																
OTHER CNTRIES (22)											1					
U.S. ORIGIN	79	80	77	75	71	72	71	64	62	61	62	59	63	59	60	69
U.S. CORP. OWNED	59	59	61	55	57	55	52	50	48	48	49	46	46	47	46	53
U.S. GOVT. OWNED	3	3	3	2	2	4	4	3	3	2	2	3	2	1	2	3
U.S. INDIV. OWNED	17	16	13	16	11	12	13	11	10	9	10	10	14	11	11	13
FOREIGN OWNED		1		1	1	1	1	1	1	1	1		1	1		1
FOREIGN ORIGIN	21	20	23	25	29	28	29	36	38	39	38	41	37	41	40	31
U.S. OWNED	2	3	3	3	4	3	2	4	5	4	4	3	3	4	4	3
FOREIGN OWNED	18	18	20	22	25	26	27	32	34	35	34	38	34	37	36	28
FOREIGN CORP.	14	15	15	17	22	21	23	27	30	30	29	31	29	31	30	23
FOREIGN GOVT.										1				1	1	
FOREIGN INDIV.	4	3	5	4	3	5	4	6	4	5	6	6	5	5	6	5

SIC 20. FOOD + KINDRED PRODUCTS

PATENT ACTIVITY (PATENTS GRANTED 1/67-12/81) BY DATE OF PATENT APPLICATION

	NUMBER OF PATENTED APPLICATIONS															
	PRE 68	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	TOTAL
TOTAL	2446	740	759	877	783	709	778	896	799	756	697	708	689	236	3	11876
U.S. ORIGIN	1877	485	535	654	553	462	476	538	509	450	402	423	421	162	1	7948
FOREIGN ORIGIN	569	255	224	223	230	247	302	358	290	306	295	285	268	74	2	3928
JAPAN	166	114	100	91	109	93	103	114	84	101	109	106	99	33	1	1423
WEST GERMANY	86	26	28	24	20	30	40	49	32	48	40	38	43	11		515
UNITED KINGDOM	49	32	22	19	21	35	40	41	33	35	36	17	22	5		407
FRANCE	34	17	17	12	16	17	19	27	28	23	15	25	13	3		266
CANADA	50	9	12	22	10	13	9	11	15	16	14	14	11	3		209
SWITZERLAND	30	2	3	8	12	14	17	36	21	23	16	23	15	7		227
NETHERLANDS	35	15	15	12	9	14	12	14	16	19	19	13	13	1		207
SWEDEN	19	7	4	7	8	1	14	14	11	5	6	12	8	1		117
ITALY	12	10	6	3	3	6	11	12	6	8	6	6	7	3		99
DENMARK	17	6	2	1	3	6	5	5	4	6	10	2	10			77
U.S.S.R.	6		2	2	4	1	5	2	7	4	6	5	6	3		53
AUSTRALIA	9	2		2	2	1	7	3	7	3		6	2			44
BELGIUM	6	2	6	2	2	2	2	3	3	1		2	2	1		34
AUSTRIA	6			2	4	3	2	3	2		3	2	2			29
ISRAEL	2	4		3		2	2	2	4	2		1	4			26
MEXICO	8	1		1				2	1	1	1	1		2		18
CZECHOSLOVAKIA	7	1						1	3	2	1	2	1			18
NORWAY	4	4	2	1			1	2	1	2	1					18
HUNGARY	2		1		1	2	3	3	1		1	2	1			17
SPAIN			1		1	1	3	4	1		1	2				14
FINLAND				2	1	2		2	2	1	1		1			12
NEW ZEALAND								2	1		1	1	2	1		8
ARGENTINA	1	1			1	1					1	3			1	9
POLAND	2						1	1			4	1				9
S. AFRICA	2			1	1		2		1		1					8
IRELAND	1							1		1			2			5
BULGARIA				2					2		1	1				5
PORTUGAL	4												1			5
CHILE		1	1					1					1			4
GUATAMELA					1				1							2
INDIA	1						1			2						4
CHINA (MAINLAND)				3												3
HONDURAS	3															3
BRAZIL	1			1			1									3
PHILLIPINES							1		1	1						3
OTHER CNTRIES (22)	6	1	2	2	1	3	1	3	2	2	1		2			26
U.S. ORIGIN	1877	485	535	654	553	462	476	538	509	450	402	423	421	162	1	7948
U.S. CORP. OWNED	1436	388	417	498	408	351	367	414	404	345	311	321	335	122	1	6118
U.S. GOVT. OWNED	73	19	20	31	32	22	14	30	16	19	18	14	11	4		323
U.S. INDIV. OWNED	353	73	96	109	106	84	89	85	80	84	67	84	72	34		1416
FOREIGN OWNED	15	5	2	16	7	5	6	9	9	2	6	4	3	2		91
FOREIGN ORIGIN	569	255	224	223	230	247	302	358	290	306	295	285	268	74	2	3928
U.S. OWNED	74	24	27	18	24	27	37	36	31	27	22	23	30	4		404
FOREIGN OWNED	495	231	197	205	206	220	265	322	259	279	273	262	238	70	2	3524
FOREIGN CORP.	399	192	168	167	176	183	222	285	204	236	237	220	186	63	1	2939
FOREIGN GOVT.	2	1	1				2	4	4	4	2	7	5			32
FOREIGN INDIV.	94	38	28	38	30	37	41	33	51	39	34	35	47	7	1	553

SIC 20. FOOD + KINDRED PRODUCTS

PATENT ACTIVITY (PATENTS GRANTED 1/67-12/81) BY DATE OF PATENT APPLICATION

PERCENT OF PATENTED APPLICATIONS

	PRE 68	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	TOTAL
TOTAL	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
U.S. ORIGIN	77	66	70	75	71	65	61	60	64	60	58	60	61	69	33	67
FOREIGN ORIGIN	23	34	30	25	29	35	39	40	36	40	42	40	39	31	67	33
JAPAN	7	15	13	10	14	13	13	13	11	13	16	15	14	14	33	12
WEST GERMANY	4	4	4	3	3	4	5	5	4	6	6	5	6	5		4
UNITED KINGDOM	2	4	3	2	3	5	5	5	4	5	5	2	3	2		3
FRANCE	1	2	2	1	2	2	2	3	4	3	2	4	2	1		2
CANADA	2	1	2	3	1	2	1	1	2	2	2	2	2	1		2
SWITZERLAND	1			1	2	2	2	4	3	3	2	3	2	3		2
NETHERLANDS	1	2	2	1	1	2	2	2	2	3	3	2	2			2
SWEDEN	1	1	1	1	1		2	2	1	1	1	2	1			1
ITALY		1	1			1	1	1	1	1	1	1	1	1		1
DENMARK	1	1				1	1	1	1	1	1		1			1
U.S.S.R.					1		1		1	1	1	1	1	1		
AUSTRALIA							1		1			1				
BELGIUM			1													
AUSTRIA					1											
ISRAEL		1							1				1			
MEXICO														1		
CZECHOSLOVAKIA																
NORWAY		1														
HUNGARY																
SPAIN																
FINLAND																
NEW ZEALAND																
ARGENTINA															33	
POLAND											1					
S. AFRICA																
IRELAND																
BULGARIA																
PORTUGAL																
CHILE																
GUATAMELA																
INDIA																
CHINA (MAINLAND)																
HONDURAS																
BRAZIL																
PHILLIPINES																
OTHER CNTRIES (22)																
U.S. ORIGIN	77	66	70	75	71	65	61	60	64	60	58	60	61	69	33	67
U.S. CORP. OWNED	59	52	55	57	52	50	47	46	51	46	45	45	49	52	33	52
U.S. GOVT. OWNED	3	3	3	4	4	3	2	3	2	3	3	2	2	2		3
U.S. INDIV. OWNED	14	10	13	12	14	12	11	9	10	11	10	12	10	14		12
FOREIGN OWNED	1	1		2	1	1	1	1	1		1	1		1		1
FOREIGN ORIGIN	23	34	30	25	29	35	39	40	36	40	42	40	39	31	67	33
U.S. OWNED	3	3	4	2	3	4	5	4	4	4	3	3	4	2		3
FOREIGN OWNED	20	31	26	23	26	31	34	36	32	37	39	37	35	30	67	30
FOREIGN CORP.	16	26	22	19	22	26	29	32	26	31	34	31	27	27	33	25
FOREIGN GOVT.									1	1		1	1			
FOREIGN INDIV.	4	5	4	4	4	5	5	4	6	5	5	5	7	3	33	5

APPENDIX C

WORKSHOP PARTICIPANTS

Mr. Gary Auton - Patent and Trademark Office
Ms. Jennifer Bond - National Science Foundation
Mrs. Deneise Boyd - OTAF
Dr. Donald Buzzelli - National Science Foundation
Dr. Richard Campbell - Battelle Pacific Northwest Laboratories
Mr. Mark Carpenter - Computer Horizons, Inc.
Mr. Jason Christian - National Science Foundation
Mr. E.D. Ellis - Consumer and Corporate Affairs, Canada
Professor Robert Evenson - Yale University
Mr. Jeffrey Forman - Derwent, Inc.
Mr. Jeffrey Franklin - OTAF
Professor Zvi Griliches - Harvard University and National Bureau of Economic Research, Inc.
Mrs. Bronwyn Hall - Stanford University and National Bureau of Economic Research, Inc.
Mr. L.B. Kirsh - Consumer and Corporate Affairs, Canada
Dr. Carole Kitti - Office of Management and Budget
Mr. William Lawson - Patent and Trademark Office
Dr. Mary Ellen Mogee - OTAF
Mrs. Jane Myers - OTAF
Mr. Fran Narin - Computer Horizons, Inc.
Mr. James Peterson - OTAF
Ms. Susan B. Rifkin - OTAF
Professor F.M. Scherer - Swarthmore College
Dr. Theodore Schlie - Office of Competitive Assessment, Department of Commerce
Dr. Luc Soete - Sussex University
Dr. John Terapane - OTAF
Dr. Nestor Terleckyj - National Planning Association