RESEARCH TRIANGLE INSTITUTE

May 1979
Contract No. 210-77-0149
RTI Project No. 44U-1535
RTI Report No. RTI/1535/00-12F

FINAL REPORT

NATIONAL OCCUPATIONAL HAZARD SURVEY (NOHS)
DATA BASE DEVELOPMENT AND ANALYSIS

Prepared for

U.S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Public Health Service
Center for Disease Control
National Institute for Occupational Safety and Health
Division of Surveillance, Hazard Evaluations, and Field Studies
Cincinnati, Ohio

Prepared by

R. N. Hendry
K. J. Reeves
R. E. Mason

RESEARCH TRIANGLE PARK, NORTH CAROLINA 27709

ABSTRACT

Beginning in September 1972, the National Occupational Hazard Survey (NOHS) was conducted by the National Institute for Occupational Safety and Health (NIOSH) over approximately two years to gain comprehensive national information regarding exposure of employees to potential hazards across occupational and industrial categories. In an effort of a little over two years, Research Triangle Institute (RTI) created an initial SYSTEM 2000 data base from the resulting survey data and from certain descriptive auxiliary files and established software procedures to utilize effectively the information contained in the data base.

However, due to unforeseen delays in the availability of certain information and due to the cumbersome operation of the retrieval and estimation procedures, the need arose to improve upon some aspects of the initial NOHS data base system and to add some new features. These needs resulted in the current contract with the following specific objectives:

- Develop a system to retrieve information and generate reports in the three basic categories of hazard, occupation, and Standard Industrial Classification (SIC) Code.
- Develop a system to update the NOHS data base by including additional information on primary and secondary Trade Name Products and newly recorded potential hazard codes.
- Develop a system to catalog and store retrieval queries and responses to avoid duplication of prior runs.
- Redesign and/or redevelop the current NOHS algorithm program used to project national statistics.
- Develop a system to permit linkage to NIOSH's Registry of Toxic Effects of Chemical Substances (RTECS) and an internally maintained file on published literature on toxic effects.

Employing the Parklawn Computing Center factilities, RTI largely accomplished these goals, and this effort is summarized here.

This report was submitted in fulfillment of Contract No. 210-77-0149 by the Research Triangle Institute under the sponsorship of the National Institute for Occupational Safety and Health.

NIOSH 00122441

50272-101

REPORT DOCUMENTATION PAGE	1. REPORT NO. 210-77-0149	2. NA	3. Recipient's Accession No.
4. Title and Subtitle: National Occupational Hazard Survey (NOHS) Data Base Development and Analysis			5. Report Date: May 1979
			6. NA
7. Author(s): Hendry, R. N., K. J. Reeves, and B. E. Mason			8. Performing Organization Rept. No. NA
9. Performing Organization Name and Address: Research Triangle Institute, Research triangle Park, North Carolina			10. Project/Task/Work Unit No. NA
			11. Contract(C) or Grant(G) No. (C) 210-77-0149 (G)
12. Sponsoring Organization Name and Address: NIOSH, Cincinnati, Ohio			13. Type of Report & Period Covered: Contract
			14. NA

15. Supplementary Notes

NA

16. Abstract (Limit 200 words)

A system developed by Research Triangle Institute to gain comprehensive national information regarding exposure of employees to potential hazards across occupational and industrial categories is discussed. The objectives of the system developed are identified. The background of the system's development is related. The relationship of the system to the National Occupational Hazard Survey (NOHS) is described. The five major tasks of the system and the findings of the data collection are discussed. The authors recommend that the query logic be expanded to enhance the value of the data base, that the new ALG program be linked to RPTMAIN and the "All Information" report capability, and that the necessary change be made in existing linkage system to form a bilateral linkage between the Registry of Toxic Effects of Chemical Substances (RTECS) and NOHS.

17. Document Analysis a. Descriptors

Information-processing, Data-collection, Health-surveys, Data-analysis, Information-systems

b. Identifiers/Open-Ended Terms

c. COSATI Field/Group

18. Availability Statement	19. Security Class (This Report)	21. No. of Pages
Available to Public	NA	103
	20. Security Class (This Page)	22. Price

(See ANSI-Z39.18) See Instructions on Reverse OPTIONAL FORM 272 (4-77)

CONTENTS

CONTENTS (Continued)

CONTENTS (Continued)

I. INTRODUCTION

The Surveillance Branch, Division of Surveillance, Hazard Evaluations and Field Studies, of the National Institute for Occupational Safety and Health (NIOSH) has the responsibility for development and implementation of a National Surveillance system. This system must be designed to assess continuously the distribution and usage characteristics of occupational hazards as well as the incidence and severity of related illnesses. Obviously, this activity is impossible without properly designed survey models, techniques, and procedures that permit the acquisition and analysis of relevant data on occupational hazards and illness effects from government, commercial, and/or industrial sources.

One of the surveillance mechanisms that is utilized for the collection of information on occupational hazards and the resultant systematic identification of high-risk occupational groups is the National Occupational Hazard Survey (NOHS), which was conducted by NIOSH beginning in 1972. In June 1974, NIOSH let Contract No. CDC-99-74-40 with the Research Triangle Institute (RTI) for professional and technical support of NOHS. The objectives of the contract were to clarify and integrate NIOSH edited data, to analyze the data statistically, to prepare a final report analyzing the survey in total, to prepare a supplemental array of tables delineating specific findings of the survey, and to construct a data base and an associated computer information retrieval system. These objectives were largely met in a period of performance lasting approximately two and one-half years.

However, there still remained several areas of improvement in development and analysis of the NOHS data base that would make it a more efficient and effective tool for NIOSH. Consequently, in September 1977, NIOSH let Contract No. 210-77-0149, "National Occupational Hazard Survey (NOHS) Data Base Development and Analysis," with RTI. The objectives of the work performed under this contract are identified below by the appropriate report section that discusses methodologies used and results obtained in each:

Report Section	Objective
IV.A	Develop a system to retrieve information and generate reports in the three basic categories of hazard, occupation, and Standard Industrial Classification (SIC) Code.
IV.B	Develop a system to update the NOHS data base by including additional information on primary and secondary Trade Name Products and newly recorded potential hazard codes.
IV.C	Develop a system to catalog and store retrieval queries and responses to avoid duplication of prior runs

Report Section	Objective
IV.D	Redesign and/or redevelop the current NOHS algorithm program used to project national statistics.
IV.E	Develop a system to permit linkage to NIOSH's Registry of Toxic Effects of Chemical Substances (RTECS) and an internally maintained file on published literature on toxic effects.

Section II, Background, describes the development of the original NOHS data base and the need for the refinements and modifications described in Section IV, Discussion of Tasks. Section III, Relationship of Current Effort to NOHS System, gives a brief description of the phases or components of the NOHS system and shows the relationship that exists between project tasks and the components of the NOHS system. Section V, Discussion of Findings, summarizes results, while Section VI provides conclusions and recommendations. Section VII lists references cited.

II. BACKGROUND

A. Development of the National Occupational Hazard Survey (NOHS) Data Base

Shortly after the passage of the Occupational Safety and Health Act of 1970 (the Act) created the National Institute for Occupational Safety and Health (NIOSH) within the Department of Health, Education and Welfare, the Department established seven task forces to implement its responsibilities under the Act. One of these task forces, the Task Force on Hazard and Disease Monitoring, foresaw the need for detailed information on the distribution of, use of, and exposure to chemical substances and physical hazards in industries regulated under the Act. The Bureau of Occupational Safety and Health (BOSH), NIOSH's organizational predecessor, had made survey efforts to meet this informational need, primarily through the National Surveillance Network for Occupational Health (NSN), an information gathering system initiated in cooperation with several state industrial hygiene agencies.

After examining the results and procedures of these earlier surveys by BOSH, the Task Force on Hazard and Disease Monitoring outlined a 2-phase information gathering system for NIOSH. The first phase, the National Occupational Hazard Survey (NOHS), was to provide a base of descriptive information on occupational exposures; the second phase was to continue using the NSN system to update information gathered by NOHS. When NSN information indicated that parts of the NOHS information were outdated, NIOSH was to perform new industry specific surveys to replace outdated segments of NOHS.

1. Survey Plan

NIOSH adopted the 2-phase scheme proposed by the Task Force and began the first phase. First, the Bureau of Labor Statistics (BLS) and NIOSH worked together to formulate a general statistical design for implementing NOHS that was feasible. The design resulted in a sample of approximately 5,000 industrial facilities in 67 metropolitan areas and approximately 3,000 facilities in 50 rural counties. Some facilities involved more than one plant, or location. The metropolitan areas included Standard Metropolitan Statistical Areas (SMSAs), groups of SMSAs, and one non-SMSA area (Anchorage, Alaska). Facilities within metropolitan areas were chosen on the basis of Standard Industrial Classification (SIC) and number of employees. Later those facilities from rural counties were dropped from NOHS.

2. Forms Control

Beginning in September 1971, two groups of ten surveyors were called to active duty in the Commissioned Corps of the United States Public Health Service to conduct the NOHS survey, and NIOSH implemented a manual in which survey procedures were specified and a survey form was provided for the surveyors to record the information gained from facility contacts. The NOHS survey actually began in Baltimore, Maryland, in February 1972.

As the NOHS survey reports were sent to Headquarters from the field, the necessity for rigid control over them was immediately apparent, so a procedure involving a desk check (for obvious coding errors) and a Volume Log Book (for organized forms control) was soon developed.

3. Professional and Technical Support

NIOSH made various interagency agreements for various support tasks in relation to NOHS, such as those for microfilming the survey results, coding occupation groups, and aiding the editing and statistical prejection processes. In addition, NIOSH let a major contract for the Trade Name Ingredient Clarification (TNIC)Project to resolve the many trade same products in NOHS into specific exposures.

In June 1974, NIOSH contracted with Research Triangle Institute (RTI) to provide professional and technical support for NOHS. The purposes of the contract (No. CDC-99-74-40) were to clarify and integrate NIOSH edited data, to analyze the data statistically, to prepare a final report analyzing the survey in total, to prepare a supplemental array of tables delineating specific findings of the survey, and to construct a data base and an associated computer information retrieval system. These objectives were largely met in a period of performance lasting approximately two and one-half years.

B. Need for Further NOHS Refinement

While the previous RTI contract objectives were largely met, there were several areas that needed more attention. Specifically, while the computer information retrieval system developed during the previous RTI contract involved a statistical estimation procedure that was versatile and theoretically sound, inherent difficulties were created by certain nonrandom aspects of the NOHS process, the estimation process was extremely slow and expensive to run, and SYSTEM 2000, which was chosen to manage the NOHS data base, was not completely implemented or being used at full effectiveness. Finally, several new ideas were formulated that could not be implemented since no money was available: development of new report formats expected to be needed for future requests, redesigning the projection algorithm to be more computationally efficient, and the development of a query storage and retrieval subsystem to avoid redundant estimate processing. It was during the project on which this report was written that these needed improvements to the NOHS system were implemented.

III. RELATIONSHIP OF CURRENT EFFORT TO NOHS SYSTEM

The NOHS data base management system procedures are comprised of five functional phases or components, which are shown as rectangular boxes in Figure 1. Descriptions of these components as they existed at the time of contract initiation are given in Section A, and the relationship of the current project's tasks to these five components is given in Section B.

A. Description of NOHS Data Base Management System

1. SYSTEM 2000 (S2K) Load/Update

Since the NOHS data base is very large and complex in structure and involves many auxiliary files, it is maintained under a Data Base Management System (DBMS), namely, SYSTEM 2000. Originally, it was thought that the trade name key resolution process would be completed by the time the initial loading of data under SYSTEM 2000, accomplished incrementally, would be finished. Thereafter, only relatively small updates, using the SYSTEM 2000 update capabilities directly, would be needed for keeping the data base current. However, the trade name resolution process has not been completed to date, although an initial data base does exist under SYSTEM 2000. For the large quantities of records involved in each increment of trade name resolution, finalized updating procedures did not exist at the time of current project initiation.

2. Query Analysis and Processing

In the query analysis and processing component, a request for information is received and reviewed in order to determine the specific requirements and the most appropriate solution. A set of query control cards is prepared manually to provide input to the projection process. The query analysis within the first step reads the query control cards, performs a validation operation, and produces output only of valid encoded intermediate requests. An error message report is produced that indicates any discrepancies requiring correction and resubmission.

The query processing phase of the first component accesses the facility-statistical data in accordance with the encoded requests received. The encoded input request, or query, records contain the selection criteria details regarding the desired conditions or specifications from the requests for information. An output of intermediate records is produced that controls operation of the three components that follow query analysis and processing. Those requests for information requiring access to the exposure data create additional intermediate control records identifying facilities that meet the selection criteria provided in the query. The exposure access control records contain data regarding occupation codes and hazard or rank codes and a case code that indicates the mode for processing occupation and hazard data. At least part of the count records input to the projection phase is also produced during query processing.

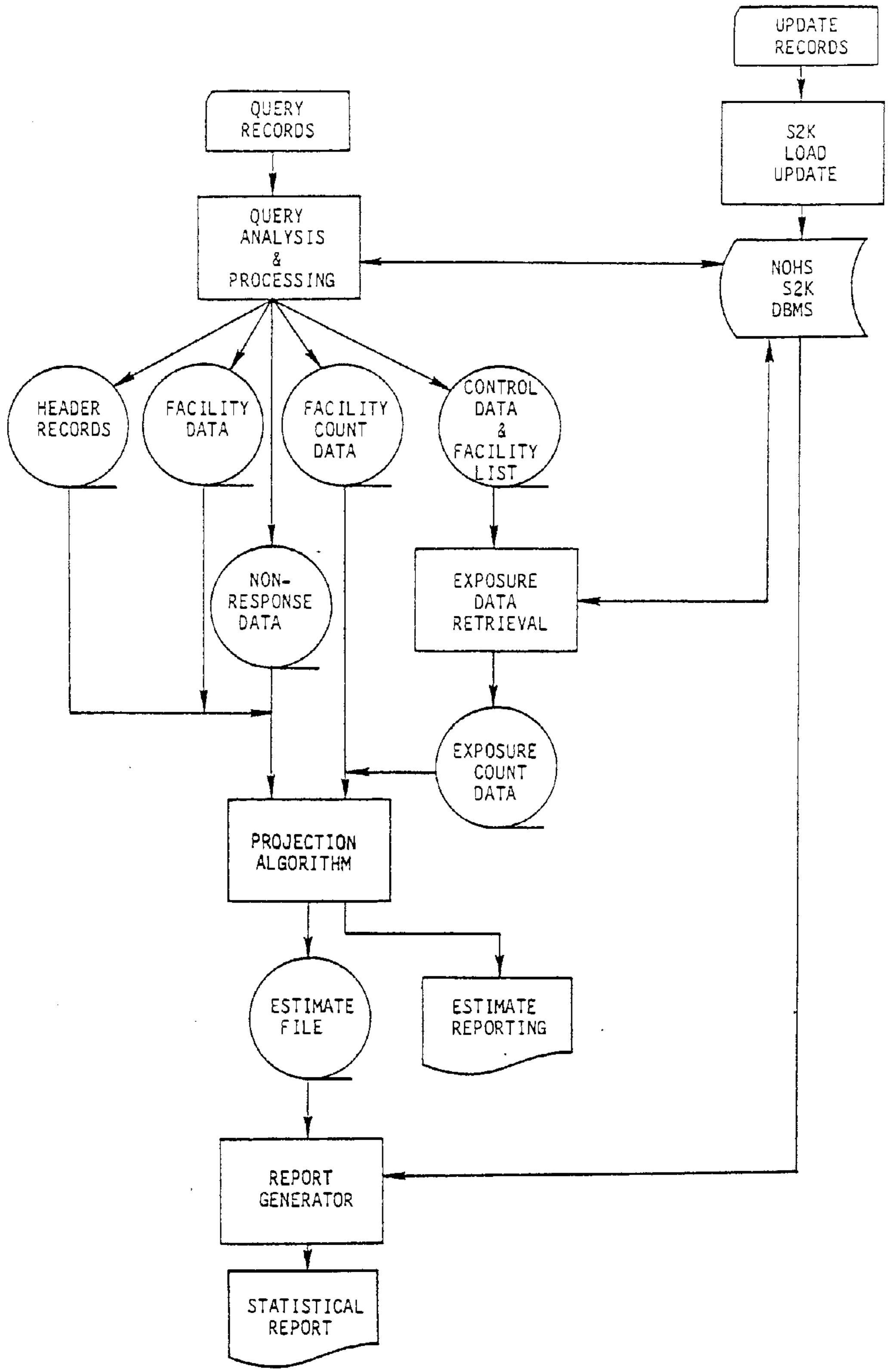


Figure 1. NOHS Data Base Management System

3. Exposure Data Retrieval

The exposure data access phase reads the control records provided by the prior query analysis and processing phase. The control records may indicate that no access to the exposure data is required, in which case the exposure access becomes null; otherwise, the control records provide, for each facility considered in the projections, indicators as to its qualification. Access to the exposure data is provided, first, by the facilities qualifying for a query found through the query processing step. Only those facilities received in the control records as qualifying will be considered for access. Additional access control occurs within each facility in terms of the occupation and hazard/rank limitations. A single specified occupation code will limit the access to a smaller portion of data within a facility range. A single rank code will also address a smaller segment of a given facility range. A request for a specified hazard/rank may have the option of including information regarding obscure exposures related either to trade name data or generic hazard resolution data. This option may expand the request for a single hazard/rank into a list of hazard/rank codes. Output from the exposure data access operation goes into the next step for reduction and summary purposes.

If exposure data have been accessed and a file of retrieved records obtained, a data reduction of retrieved data is needed. The data reduction step resolves duplicate exposures due to obscure exposure retrieval and summarizes the data into a form and sequence for input to the projection phase. Processing case codes for "all" and "every" are differentiated in this phase. Various processing routines handle the different case code requirements. Report number and query position in the report are maintained for control purposes.

4. Projection Algorithm

The projection component receives control and data records from several sources. Basic facility-statistical data required for projection calculations and the necessary query and report control data are received directly from the query analysis and processing phase. A Facility Count Data file may be received from the query analysis and processing phase alone if no reference was specified to exposure data. However, exposure data may be obtained from the retrieval and data reduction steps. Data received from the exposure data retrieval step are sorted by report, pseudo-facility number, and query so as to simplify and expedite projection calculations.

The projection algorithm component takes the above inputs and produces estimate files that become input to the report generator. These files contain the necessary statistical information, including SIC-size classifications, estimates, variances, and response counts, to produce estimate tables such as those found in Volume III of the final report [1] for the previous RTI data analysis contract (No. CDC-99-74-40).

5. Report Generator

The report generator phase receives a combination of control, header, and estimate records. The input records are arranged or sorted into the desired report control records. Various report formats and sequences are provided to permit a wide range of reports.

B. Current Effort on NOHS System Components

As originally conceived (see Figure 1), the NOHS system had five major components: a load/update component, a query analysis and processing component, an exposure data retrieval component, a projection algorithm, and a report generator. Another potential NOHS system component being strongly considered at the beginning of the current project was a query storage and retrieval subsystem. The relationships of these six components to the five tasks in the current contract are summarized in Table I and discussed more fully below.

TABLE I. CONTRACT TASKS BY SYSTEM COMPONENT

NOHS Components	Tasks				
	1	2	3	4	5
Load/update		X			
Query analysis and processing	X				
Exposure data retrieval	X				X
Projection algorithm				X	
Report generator	X				
Query storage and retrieval			X		

Task 1, Develop Retrieval and Reporting Subsystem, had as its objective the development of a system to retrieve information from the NOHS data base and to generate reports beyond those prepared for Volume III of the final report for the previous RTI data analysis contract with NIOSH [1]. The ideal goal was a unified method to present information across any of the three major NOHS stratifications (SIC, OCC, or HAZ) in any of a variety of report formats. So as to facilitate requests for estimates needed for such reports, a more comprehensive, easily understood query language was an important consideration, thus involving the query analysis and processing component as shown in Table I. As also shown in Table I, the expansion of exposure data retrieval capabilities was considered necessary, and the report generator required the addition of new options to meet new report needs.

Task 2, Develop Update Subsystem for NOHS, entailed developing a method of updating the NOHS data base as the trade name resolution process continued. The resolution process had become a longer-term process than originally planned. Thus, since a SYSTEM 2000 data base was loaded that would continue needing massive updates for the foreseeable future, the load/update procedures for the data base received careful consideration. The development of the update subsystem is the central part of implementing a dynamic load/update process.

Task 3, Develop Query Storage and Retrieval Subsystem, concerned the development of a subsystem not originally a part of the NOHS system. The goal of the query storage and retrieval subsystem was to eliminate the need for redundant estimate calculations by storing the queries and their associated estimates in the data base or in some other easily accessible file. Then, when a query has been analyzed and determined to be structurally correct, a search could be made to establish whether or not that query had been previously generated. If so, the calculated estimates would be retrieved or, if not, estimates would be generated and stored. Due to the slow retrieval process and the even slower projection algorithm used previous to the current project, the development of a query storage and retrieval subsystem was anticipated to save much expense by eliminating many costly estimate runs in the future.

Task 4, Redesign Projection Algorithm, was the task carrying the highest urgency for NIOSH. Interest continued to be expressed in specific NOHS estimates after the previous RTI data analysis contract, and the original projection algorithm, although theoretically sound, was very expensive to run. Its inefficiency was caused primarily by computational adjustments for statistical aberrations in the NOHS implementation, the desire for estimates by observed SIC-size (rather than the original BLS SIC-size), and the many extra calculations required for variances and, for some estimates, covariances. The redesign task is shown in Table I to involve primarily a single major component of the NOHS system, the projection algorithm.

Task 5, Develop Subsystem to Link NOHS and RTECS, involved an extension of NOHS retrieval capabilities and was expected to increase the usefulness of both the NOHS and RTECS data bases. The Registry of Toxic Effects of Chemical Substances (RTECS) data base provides comprehensive information on the toxic effects of various chemical substances. The NOHS data base offers very valuable information on the distribution of potentially hazardous exposures to hazards across industrial and occupational categories. As reflected in Table I, the NOHS-RTECS link subsystem was expected to make a major contribution to the exposure data retrieval component of the NOHS system.

The five tasks of the current contract were largely completed, except for the development of the query storage and retrieval subsystem. Completing this subsystem became unnecessary, primarily due to the extent of the improvement made in redesigning the projection algorithm. It should be noted that documentation was included as a major part of several subtasks after contract initiation, due to concerns expressed by NIOSH about the importance of documenting data sets and subsystems developed during the current contract.

IV. DISCUSSION OF TASKS

Under NIOSH Contract No. CDC-99-74-40, RTI was assigned five major tasks, which RTI further divided into subtasks, as noted in Figure 2. Accomplishments made under each of these project tasks are described in detail in the following sections, with subsections identified by the Figure 2 subtasks to which they pertain. A brief description of the relationship of these five tasks to the components of the NOHS system and a discussion of the task objectives are given in Section III, Relationship of Current Effort to NOHS System.

A. Task 1: Develop Retrieval and Reporting Subsystem

Under Task 1, RTI developed a subsystem to retrieve information and generate reports in one or in any combination of the three basic categories of the NOHS data base: hazard, occupation, and Standard Industrial Classification (SIC). This subsystem contains a retrieval subsystem and a report subsystem, each of which is discussed below.

1. Retrieval Subsystem (Subtask 1.1)

The first step in developing the retrieval subsystem was the loading of the large and complex NOHS data base under SYSTEM 2000. This allowed the highly efficient retrieval features of SYSTEM 2000 to be used in designing the retrieval functions needed. However, for retrieval subsystem testing purposes, a small, rather than a large, data base was desirable, so a small data base, which will continue to be loaded incrementally, was initially loaded by January 1978. A description of the current status of the data base may be obtained from the schema and tallies of several key fields in Figure 3.

The retrieval approach utilizes the Procedural Language Interface (PLI) feature of SYSTEM 2000 to retrieve information from the data base through a COBOL program. Efficiency of the retrieval process was disappointing initially but was improved to a very satisfactory level with the aid of a representative from MRI Systems Corporation, the vendor of SYSTEM 2000. The representative's suggestions involved some relatively simple revisions to the retrieval approach to avoid unnecessary use of the SYSTEM 2000 "GET" command, which is very powerful but slow in execution. The program now checks before issuing this command to ensure that the needed information is not already contained in core storage.

The retrieval process produces a file of 47-byte records, which, in turn, are sorted by facility number. These sorted records then become input to the revised projection algorithm so that estimate reports can be produced.

2. Reporting Subsystem (Subtask 1.2)

To develop the reporting subsystem, new estimate table formats were developed that could present reports in one or in any combination of the three basic categories of the NOHS data base: hazard, occupation, and SIC. These report formats supplement those contained in Volume III of the final report on the previous RTI data analysis contract [1]. The development of these new

1. Develop Retrieval and Reporting Subsystem

   1.1 Develop Retrieval Subsystem Functions

   1.2 Develop and Document Reporting Subsystem

2. Develop Update Subsystem for NOHS

   2.1 Determine Current Subsystem Status and Schedule for NIOSH Internal Processing of Trade Name Data

   2.2 Examine S2K Alternatives

   2.3 Determine Explicit NOHS Update Requirements

   2.4 Evaluate and Select Design Options

   2.5 Implement Initial NOHS Update Subsystem

   2.6 Revise "Redbook" Summary Procedure

   2.7 Consider Alternatives to ISAM TRN Update Procedure

   2.8 Inventory and Document Update Procedure

3. Develop Query Storage and Retrieval Subsystem

   3.1 Determine Extent of Existing Available Prior Queries and Responses in Machine Processable Form

   3.2 Examine and Evaluate Storage Media and Methods

   3.3 Determine Viable Design Alternatives Compatible with Task 1

   3.4 Select, Design, Test, and Implement Subsystem

4. Redesign Projection Algorithm

   4.1 Research and Analyze Existing Projections for Comparative Ratios

   4.2 Establish Comparative Evaluation Baseline Files

   4.3 Determine Applicable Approximation Methods

   4.4 Review Dr. Breslin's Algorithm

   4.5 Establish List of Viable Options

   4.6 Select, Test, and Implement Viable Options

   4.7 Evaluate Comparative Run Time and Cost

   4.8 Develop concept for New Projection Algorithm

   4.9 Implement New Projection Algorithm

   4.10 Develop Proof of PROJ4 and ALG Differences

   4.11 Develop a Means for Estimating Variances within the ALG Procedure

   4.12 Inventory and Document Estimation Procedures (PROJ4 and ALG)

5. Develop Subsystem to Link NOHS and RTECS

   5.1 Determine Anticipated Report Requirements

   5.2 Evaluate Compatible Code Relationships

   5.3 Examine Viable Alternatives

   5.4 Test and Implement Link Subsystem

Figure 2. Project Tasks and Subtasks

```
 04/24/79  14:16:14  BEGIN SYSTEM 2000 - VERSION 2.80
---
USER,JKCC:
---
DBN IS NOHSOBS:
 ASSIGNED...NOHSOBS                  1       18     03/27/78     17:01:12
---
CONTROL: PRINT DATA BASE SIZE:
            NOHSOBS                  1       18     03/27/78     17:01:12
             PAGES      PAGES       PAGE
            IN USE     UNUSED       SIZE
NOHSOBS1 -  10          47          4060
  VOL SER-  MOSH02, EXTENTS- 1
NOHSOBS2 -  1166        259         4060
  VOL SER-  MOSH02, EXTENTS- 1
NOHSOBS3 -  1           94          12992
  VOL SER-  MOSH02, EXTENTS- 1
NOHSOBS4 -  4761        3789        4060
  VOL SER-  MOSH02, EXTENTS- 1
NOHSOBS5 -  6599        925         12992
  VOL SER-  MOSH03, EXTENTS- 4
NOHSOBS6 -  11466       219         4060
  VOL SER-  MOSH01, EXTENTS- 7
TOTAL NUMBER OF CHARACTERS = 156403380
---
ACCESS:  DESCRIBE:
SYSTEM RELEASE NUMBER  2.80
DATA BASE NAME IS       NOHSOBS
DEFINITION NUMBER              1
DATA BASE CYCLE NUMBER        18
     1*  NHSOB (NAME X)
   101*  SICDATA (RG)
    1001*  SIC1 (INTEGER NUMBER 9999 IN 101)
    1003*  DES1 (NON-KEY NAME X(32) IN 101)
   201*  FACDES (RG)
    2221*  FSIC (INTEGER NUMBER 99 IN 201)
    2222*  FSIZ (INTEGER NUMBER 9 IN 201)
    2223*  FGRP (INTEGER NUMBER 9 IN 201)
     222*  FBLSSG (INTEGER NUMBER 9(6) IN 201)
     223*  FAC (INTEGER NUMBER 9(5) IN 201)
     224*  FACN (INTEGER NUMBER 9(5) IN 201)
    2051*  FAC50 (NON-KEY TEXT X(50) IN 201)
    2052*  FAC102 (NON-KEY TEXT X(102) IN 201)
    2064*  FACDISP (TEXT X IN 201)
    2053*  FAC27 (NON-KEY TEXT X(26) IN 201)
    2055*  FWGT1 (NON-KEY DECIMAL NUMBER 9(6).99 IN 201)
    2056*  FWGT2 (NON-KEY DECIMAL NUMBER 9(6).99 IN 201)
    2057*  FAC88 (NON-KEY TEXT X(88) IN 201)
   301*  OCCDES (RG)
    3001*  OCC2 (INTEGER NUMBER 9(5) IN 301)
    3002*  DES2 (NON-KEY NAME X(45) IN 301)
   401*  HAZDATA (RG)
    4001*  HAZ4 (TEXT X(5) IN 401)
    4002*  RANK4 (TEXT X(5) IN 401)
     402*  HFORMS (NON-KEY TEXT X(9) IN 401)
    4003*  DES4 (NAME X(140) IN 401)
   350*  EXPDATA (RG)
    3511*  SMA (TEXT XXX IN 350 WITH  FEW FUTURE ADDITIONS )
     351*  EFAC (INTEGER NUMBER 9(6) IN 350 WITH  FEW FUTURE ADDITIONS
           )
     352*  EOCCRG (RG IN 350)
       353*  EOCC (INTEGER NUMBER 9(7) IN 352 WITH  FEW FUTURE ADDITION
             S )
       354*  SUFRG (RG IN 352)
```

(Continued)

Figure 3. Schema and Tallies of Current SYSTEM 2000 Data Base

```
     355*  ESUF (NON-KEY INTEGER NUMBER 99 IN 354)
     356*  REF1 (NON-KEY TEXT X(9) IN 354)
     357*  CNT (NON-KEY INTEGER NUMBER 999 IN 354)
    3501*  HAZRG (RG IN 354)
      3502*  LTAG (NON-KEY TEXT X IN 3501)
      3503*  HAZO (TEXT X(5) IN 3501 WITH  FEW FUTURE ADDITIONS )
      3504*  CONRG (RG IN 3501)
        3505*  CNTLS (NON-KEY INTEGER NUMBER 99 IN 3504)
        3506*  FORM (NON-KEY TEXT X IN 3504)
        3507*  INT (NON-KEY TEXT X IN 3504)
        3508*  DUR (NON-KEY TEXT X IN 3504)
        3509*  CNTL (NON-KEY TEXT XX IN 3504)
        3510*  FUNC (NON-KEY TEXT X IN 3504)
      3515*  KEYRG (RG IN 3501)
        3516*  TKY (TEXT X(7) IN 3515 WITH  FEW FUTURE ADDITIONS )
        3517*  REF3 (TEXT X(15) IN 3515)
      3520*  CMPRG (RG IN 3501)
        3521*  HAZC (TEXT X(5) IN 3520 WITH  FEW FUTURE ADDITIONS )

---
TALLY/ALL/FAC,FACN,HAZ4,RANK4,SIC1,OCC2:
***********************
 ELEMENT-     FAC
***********************
 MINIMUM-        100
-----------------------
 MAXIMUM-      67070
-----------------------
     6330 UNIQUE VALUES
-----------------------
    11498 OCCURRENCES
-----------------------
***********************
 ELEMENT-     FACN
***********************
 MINIMUM-          1
-----------------------
 MAXIMUM-       5749
-----------------------
     5749 UNIQUE VALUES
-----------------------
    11498 OCCURRENCES
-----------------------
***********************
 ELEMENT-     HAZ4
***********************
 MINIMUM-     A1001
-----------------------
 MAXIMUM-     99999
-----------------------
     8513 UNIQUE VALUES
-----------------------
     8536 OCCURRENCES
-----------------------
***********************
 ELEMENT-     RANK4
***********************
 MINIMUM-     00001
-----------------------
 MAXIMUM-     07285
-----------------------
     7266 UNIQUE VALUES
-----------------------
     8536 OCCURRENCES
```

(Continued)

Figure 3. Schema and Tallies of Current SYSTEM 2000 Data Base (Continued)

```
-----------------------
***********************
ELEMENT-     SIC1
***********************
MINIMUM-        10
-----------------------
MAXIMUM-      9490
-----------------------
     1660 UNIQUE VALUES
-----------------------
     1672 OCCURRENCES
-----------------------
***********************
ELEMENT-     OCC2
***********************
MINIMUM-           1
-----------------------
MAXIMUM-         995
-----------------------
      442 UNIQUE VALUES
-----------------------
      442 OCCURRENCES
-----------------------
---
TALLY/ALL/EFAC,EOCC,HAZO,TKY,HAZC:
***********************
ELEMENT-     EFAC
***********************
MINIMUM-        1001
-----------------------
MAXIMUM-       67070
-----------------------
     4496 UNIQUE VALUES
-----------------------
     4496 OCCURRENCES
-----------------------
***********************
ELEMENT-     EOCC
***********************
MINIMUM-           0
-----------------------
MAXIMUM-         964
-----------------------
      345 UNIQUE VALUES
-----------------------
    40123 OCCURRENCES
-----------------------
***********************
ELEMENT-     HAZO
***********************
MINIMUM-     A1001
-----------------------
MAXIMUM-     99999
-----------------------
     3784 UNIQUE VALUES
-----------------------
   481114 OCCURRENCES
-----------------------
***********************
ELEMENT-     TKY
***********************
MINIMUM-     0001801
-----------------------
MAXIMUM-     6001382
-----------------------
     1290 UNIQUE VALUES
-----------------------
     4062 OCCURRENCES
-----------------------
***********************
ELEMENT-     HAZC
***********************
MINIMUM-       58
-----------------------
MAXIMUM-     94220
-----------------------
     4392 UNIQUE VALUES
-----------------------
  4214441 OCCURRENCES
-----------------------
---
EXIT:
  04/24/79  14:49:11  END SYSTEM 2000 - VERSION 2.80
```

Figure 3. Schema and Tallies of Current SYSTEM 2000 Data Base (Continued)

formats was originally to be accomplished by expanding the RPTMAIN concept, developed and used in producing the previous estimate tables. However, the RPTMAIN concept was incomplete at contract initiation, and the new formats were produced, instead, from the creation of the program ALG3, which resulted from the redesign of the projection algorithm and from the subsequent development of an "All Information" Report.

Many formats developed for Volume III, including some of those not incorporated into RPTMAIN, did not seem particularly useful to people making requests of NIOSH. Furthermore, the ALG3 procedure, which will probably be much more utilized than the old projection procedure, produces a report directly rather than writing out a file of estimate records, as did RPTMAIN. The report from ALG3 shows estimates of plants, exposed people in plants, and exposures under hazard by either SIC or occupation.

Specific standard deviation estimates, which would be included in report table formats produced by RPTMAIN, are not feasible under ALG3, but the information from ALG3 seems to meet most requests. Therefore, a major effort to develop RPTMAIN further and possibly to link ALG3 and RPTMAIN into a procedure for producing more versatile reports was not undertaken under the present contract. For examples of ALG3 reports, see Appendix A.

In the development of report capabilities in the three basic categories of the NOHS data base, an alternative to the previous report procedure was the Preliminary Information Retrieval System (PIRS). PIRS was developed as an interim storage and retrieval system for NIOSH to access the NOHS data base. Features of PIRS provided a 3-dimensional approach to the NOHS data base--i.e., by industry, by occupation, or by hazard. A search could be initiated in any one of the three dimensions and result in a cross tabulation of the other two dimensions. National estimates or projections were not available under the PIRS software package. However, certain features of PIRS that met the goals of the contract effort (particularly the availability of 3-dimensional information) are also features of ALG3, which, in addition, incorporates efficiently obtained estimates. Specifically, the ALG3 program helps meet the task goals by allowing estimate tables to be efficiently obtained within hazard by either SIC or occupation.

Another type of report, developed as a result of a special request by RTI personnel on another contract, that should help meet comprehensive information needs similar to those met by PIRS is the "All Information" Report. The "All Information" Report is designed to give facility, people exposed, and exposure estimates by hazard within SIC or occupation or both qualifications, as the user directs, and furthermore gives exposure estimates classified by each value of form, intensity, duration, condition, and function. The utilization of the efficient new algorithm to produce such a report makes possible, in a reasonable time frame, processing a tremendous number of estimates under a very large number of SIC-occupation-hazard combinations to produce an "All Information" Report, an example of which may be seen in Appendix B.

B. Task 2: Develop Update Subsystem For NOHS

Task 2 concerned the development of a subsystem to update the NOHS data base by adding information, including codes for newly discovered potential hazards, on primary and secondary trade name products. A trade name product is a unique product, marketed by a manufacturer under a specific name, composed of certain identifiable chemical compound components. One of the components contained in a trade name product may itself be a trade name product, in which case that component is considered a secondary trade name product, which, in turn, may have as a component another secondary trade name product. A trade name product that contains complete hierarchies of such products, or one that has no secondary trade name product, is a primary trade name product.

All the trade name products, primary or secondary, found in the NOHS data base are resolved into components through basically the same NIOSH processing, which has been a tremendous undertaking. As a result of this processing, additional potential hazards are sometimes identified as trade name components, and new codes for these are incorporated into the NOHS data base. Since the processes (1) of determining that a potential hazard has not been previously incorporated and (2) of subsequently assigning it a unique code are largely manual processes, NIOSH implemented clerical and data base update procedures to perform these functions. However, RTI developed a subsystem that could be used to update the NOHS data base with components for each increment of trade name resolution. This subsystem development also included the revision, into a more useful format, of the "Redbook" summary report, which is meant to present summary information, by hazard, on the status of the updated data base.

1. NIOSH Trade Name Data Processing (Subtask 2.1)

Generating an optimal update subsystem design for the NOHS data base was dependent upon accessing the status of NIOSH internal processing for trade name data. This processing is an activity that has been delayed, partially by the sheer volume of work involved, well beyond original expectations.

In January 1973, a Request for Proposal was let to procure the product formulations for those trade name products found in industry during the NOHS. Approximately 75 percent of the data collected was in the form of a trade name product, and the ingredients were needed to describe accurately the chemical or biological exposures to the employees covered by the NOHS. The contractor proposals to procure the product formulations were evaluated, and the contract was awarded on June 30, 1973. The project was given the title of Trade Name Ingredient Clarification (TNIC).

The TNIC contractor initiated test mailing procedures to receive information from manufacturers regarding certain of their trade name products. From these test procedures, many problems were noted with ambiguous trade names, inadequate responses concerning specific trade name products, and reluctance of many manufacturers to respond to trade name component requests. Also, the impact of secondary trade name products on the amount of processing involved was not realized until responses from the first full-scale mailing in April 1974 were analyzed. Resolution of the processing problems from the test mailings and the extra effort needed for processing secondary trade name products required several major contract modifications to allow time and effort for the TNIC project considerably beyond original estimates. Under the last major modification, a complete mailing first occurred in July 1975.

By June 30, 1976, the contractor received responses for 69,600 trade name products out of a mailing of 92,400 for a response rate of 75 percent. Of the 69,600 responses, 48,800 contained valid chemical information and 20,800 either could not be identified by the manufacturer, were not made by the manufacturer, contained ambiguous information, or were duplicates of another product with a similar name. The total number of trade names in the NIOSH data base was approximately 86,000, and not 92,400, since many duplicates were discovered. The 86,000 trade name products were the products of 10,500 manufacturers.

The formulations for the 22,800 missing products were developed by RTI, which was responsible for the statistical evaluation and the industrial hygiene consistency of the survey data under the previous contract (No. CDC-99-74-40). The formulations were devised through an extensive literature search. NIOSH realized that these formulations were not as good as the exact formula, but they filled a critical gap in the NOHS data. NIOSH is continuing work to obtain the 22,800 missing formulations so that the 1976 publication of the NOHS results can be updated. Prior to the next publication of NOHS results, the NIOSH data base is to be updated with the missing formulations so that answers to inquiries will be as accurate as possible. This updating includes the component resolution process for these missing products, as well as for those products whose initial resolutions contained errors.

The TNIC contractor processed the trade name resolution by manufacturer and released them to NIOSH in groups as they were completed. The NIOSH computer processing involved matching the files to the NIOSH raw-data files and performing various validation processes. These trade name insertion files, which are basically raw-data file records with trade name keys added, have not all been completely processed. The unfinished work is primarily due to staff overloading and questions about the matching process when subsampled facilities are involved, since such facilities are now represented as single facilities in the data base files. The update subsystem for NOHS developed during this contract used a test 73-byte file derived from raw data, with trade name keys inserted, and a test components file extracted from the April 1976 version for systems development.

2. <u>Selection of Update Design (Subtasks 2.2-2.4)</u>

Within SYSTEM 2000, the Data Base Management System being employed for the NOHS data base, there were two modular concepts that could be considered to keep the data base current as new trade name processing is completed: LOAD and UPDATE. After trial runs and discussions with a representative from MRI Systems Corporation (the vendor of SYSTEM 2000), it became clear that the use of UPDATE for the large number of records generally to be involved was inefficient. In fact, LOAD, the method of creating new load tapes and incrementally reloading the data base, would be faster.

In the design of the NOHS update system, several requirements were followed. First, one output file was to be a load file for SYSTEM 2000, and another was to be an input file for a revised "Redbook" summary report procedure. The "Redbook" summary report contains, by hazard, occupation code counts, people estimates, and percentages for exposure source by exposure type (observed exposure, exposure derived from trade name resolution, or exposure derived from generic resolution). Second, in accordance with the schema of the data base (seen in Figure 3 and based on the hierarchical nature of SYSTEM 2000), each trade name key was to be resolved into components where it is found. Third, the method of finding components for keys was not to depend on initial key order in the input file, which was unpredictable, and still was to be reasonably efficient. Fourth, only the first instance of a trade name key within an occupation group and hazard was to be resolved into components. Fifth, generic records were to be eliminated within an occupation group hazard when resolution is successful. And, sixth, the output load file was to be capable of being updated for repeated increments of trade name processing without resolving into keys again those record keys already processed.

Thus, the design option chosen involved taking the previous load file as input, eliminating generics within an occupation group hazard (after resolving into keys each unresolved record key in that group hazard), and incrementally reloading the SYSTEM 2000 data base. As significant progress is made in resolving trade name components, the process should be repeated.

The complete design of the final version of the update subsystem includes two runs of the update program and, as parallel steps, the "Redbook" summary report process and reload process. This design is shown in Figure 4.

3. Implementation of NOHS Update Subsystem (Subtasks 2.5 and 2.7)

The first implementation of the NOHS update subsystem utilized a binary search technique on an internal table to find components for trade name keys when necessary. This idea was very reasonable since only 22 records were contained in the test components file. Two files were produced as output--a revised load file and a 32-byte file that is used as input to the "Redbook" summary report procedure. As these files are created, each new facility is matched against the facility file by facility number. The Inverse Inclusion Probability (IIP) factor from the facility file is multiplied by each count for that facility, and the resulting products can be used directly in forming estimates for the "Redbook" summary reports. Generic records could be eliminated within occupation group hazards when components are successfully found for keys, since the file order is unchanged and the key search reinitialized in each instance. Also, each trade name key is resolved into components within an occupation group hazard only once. Thus, this approach was used successfully to gain counts of output records that could be checked against later implementations from the same test load file. No provision for updating again was made at that time.

The idea of an internal table for trade name component storage by key was impractical for the large component and load files anticipated, both because of core storage limitations and because of time consumed by binary searches reinitialized many times over a large table. The most straightforward modification of the table idea was the use of random access on an Indexed Sequential Access Method (ISAM) trade name components file. This modification was made, and a version of the program resulted that ran nearly as quickly on the small files used for testing as had the internal table approach. This ISAM approach was more appropriate for large files.

The ISAM file concept has been updated by IBM into the idea of Virtual Storage Access Method (VSAM) files. A VSAM file can have alternate keys in addition to the primary one. VSAM is also supposedly more efficient than

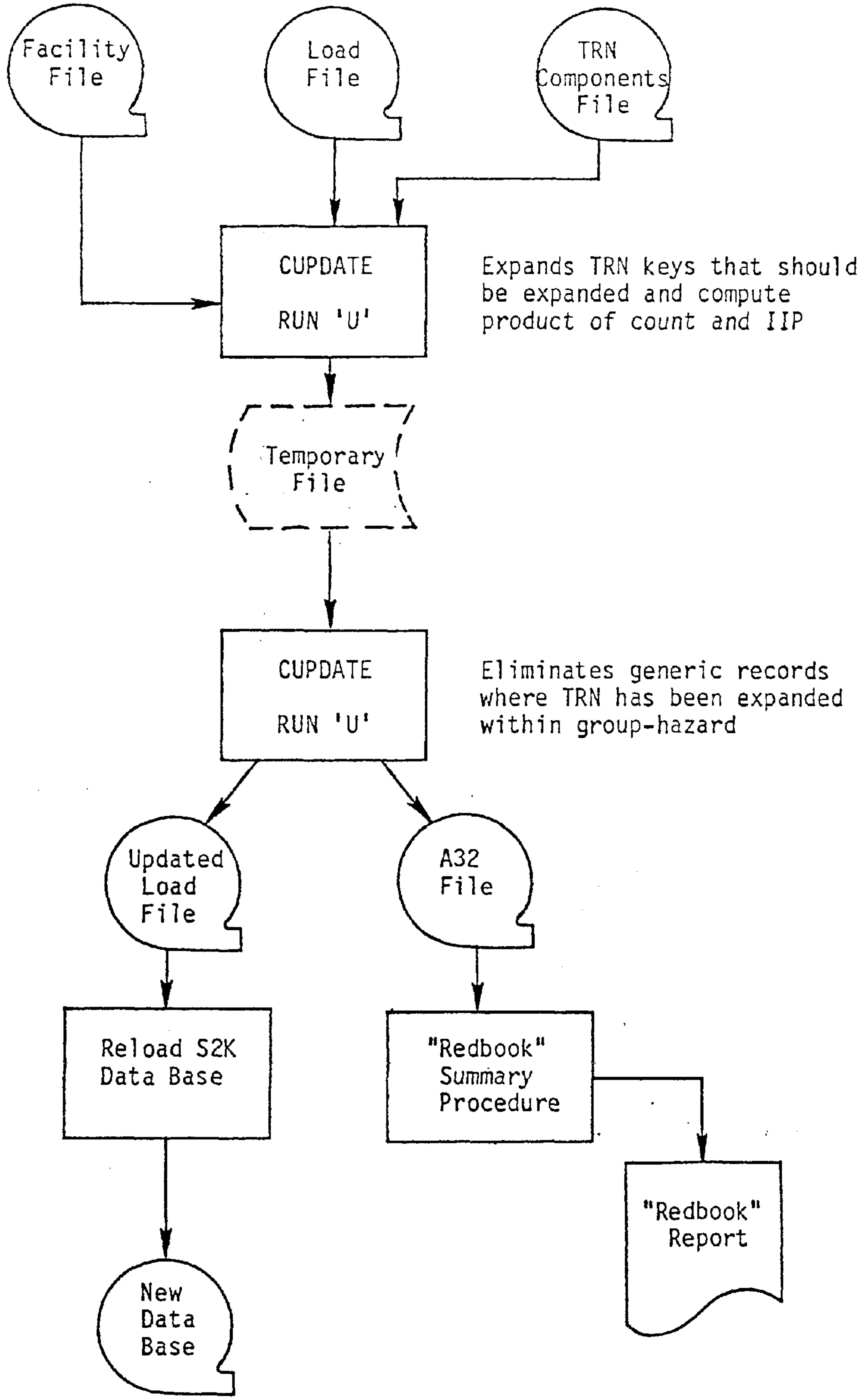


Figure 4. Design of Update Subsystem

ISAM. However, a VSAM file requires more steps to create than an ISAM file: Whereas an ISAM file can be written out directly from a COBOL program, a VSAM file first requires that a catalog for the file and then the file itself be defined using the IDCAMS utility. If alternate indexes are used, these should be built after the file is loaded. Currently, these functions for creating a VSAM file must be coordinated through the Parklawn Computing Center (PCC) Magic Shop, both since the master VSAM catalog is password-protected and since certain naming conventions are being followed. The experience of this project shows VSAM to be slower in retrieval than ISAM on files containing up to several thousand records. This comparison held true for retrieval on the primary key whether or not a VSAM file contained alternate keys. A contributing factor to the slower retrieval time for VSAM is that there is more overhead time (opening and closing operations) with VSAM than there is with ISAM because of allowances made for additional flexibility in file structure. The file size at which VSAM actually becomes more efficient than ISAM cannot be determined from the experience of this project.

The random access of an ISAM file, even so, is not a very fast operation. Using random access, of course, allowed the keys on the load file to be in any order, and therefore file order did not need to be changed. For a relatively small number of key accesses, random access time was not a major consideration and might even be less than extra sorting time. However, a large number of accesses are made in a typical update run. Preliminary results of updating with larger numbers of records and keys involved indicated that less than optimal efficiency was being gained with the ISAM trade name (TRN) components approach. There are approximately 90,000 distinct keys in the NOHS data eventually to be resolved, and most input load files to the update procedure will be very large, 100,000 or more records. Thus, the number of random accesses in a typical update run would be very significant using the ISAM approach; therefore, the search for an alternate procedure, while not mandatory, was certainly a desirable undertaking.

Only one alternate update procedure was tried, and this procedure, contained in the program CUPDATE, was implemented and accepted as the core of the update subsystem for NOHS. In this alternate update procedure, the first step uses a sort of the load file by TRN key. A file of 88-byte records is released to the sort on its input procedure. These records contain all the information from the load file, appended by the IIP-count product needed for the 32-byte file, conditions carried along for insertion in generic and trade name component 32-byte records, and a tag to indicate which trade name keys are to be resolved into components. A tag, XTAG, is also reserved within the first 73 bytes to show in a later update that a record key has been resolved. Thus, provision is made for updating the new load file without again resolving the same keys into components. For the update processing to follow, all generic records have their record tags changed to "Z." On the output procedure of the sort, those records having keys needing resolution are processed through a simple sequential match, and a temporary file containing all the records on the load file plus those formulated using trade name component information (the "T" tag records) is produced.

The second step of the update procedure sorts the temporary file by facility, occupation, occupation group suffix, hazard, XTAG (descending), TRN key, and record type. This order puts the record into the occupation group hazard context, which, originally, was the primary input file ordering. This sort order also allows a switch to be set if any resolution records exist within a group hazard. Subsequently, generic records within that group hazard can be eliminated. From the records that are kept, a new load file is formulated by truncating the last 15 bytes from the sorted 88-byte record, and a 32-byte file becomes input to the "Redbook" summary report procedure. For a system flowchart of the update subsystem, see Figure 4.

4. Revised "Redbook" Summary Procedure (Subtask 2.6)

Since the "Redbook" summary report idea was generated by RTI under the previous data analysis contract (No. CDC-99-74-40) as an aid to NIOSH in understanding the loaded data base content, the "Redbook" summary report is included in the update subsystem as an aid in analyzing the updated data base. External interest was also shown in the type of information presented by the "Redbook" report, and thus a desire was expressed to obtain a more readable and useful format for the report.

RTI formulated a new "Redbook" summary report procedure, rather than revising the old one. The current procedure has two steps. The first step, utilizing the program RBKSUM, sorts the 32-byte file from the update subsystem by hazard code, occupation code, suffix, and facility. The reports that result from the procedure have a summary line for each hazard. Thus, for each hazard-code break occurring in the sorted file, after percentages of exposure source by tag are calculated, the information accumulated for the previous hazard is written out on a summary file. The hazard description is moved to the summary record by a simple sequential match to the ISAM hazard file. Also the Chemical Abstracts Sequence (CAS) number is obtained when it is available and is used to find the Registry of Toxic Effects of Chemical Substances (RTECS) number. For a hazard, occupation, suffix, or facility break, the product of IIP factor and count, contained on the 32-byte file, is added to estimated people exposed, and 1 is added to an occupation count for a hazard or occupation change. For all records returned, the IIP-count product from the sorted 32-byte file is added to the estimated exposures and exposures by tag.

The next step straightforwardly sorts the summary file obtained in the first step into either hazard code or description order and prints the information with appropriate headers. For an example of the "Redbook" report in each order, see Appendix C.

There are many blank spaces for CAS and RTECS numbers because many of the hazard file records have not yet been updated with CAS numbers. The CAS number records ("D" records) from the master RTECS file, current as of October 10, 1978, were used to form an ISAM file of short (24-byte) records composed primarily of CAS number, as primary key, and RTECS number. Thus, when the hazard record is retrieved to obtain the hazard description and CAS number, the ISAM file is searched randomly to get the RTECS number when the CAS number is nonblank. Both CAS and RTECS numbers can then be placed on the summary file; therefore, the CAS number has to be on the hazard file before either number will appear on the report.

C. Task 3: Develop Query Storage and Retrieval Subsystem

So as to catalog and store retrieval queries and responses to avoid duplicate processing for the same information in the future, a query storage and retrieval subsystem needed to be developed for the NOHS system.

As a first step in developing this subsystem, RTI examined the files of responses contained in the estimate tables in Volume III of the final report on RTI's previous data analysis contract with NIOSH (Contract No. CDC-99-74-40) [1].

Some estimate files, containing multiple hazard-occupation combinations in multiple columns, involved so many response records that they were recorded on tape instead of disk. These tapes were released for reuse, and thus these files were not generally available. However, many estimate files contained a reasonably small number of response records and therefore were stored on disk. These files of responses were mostly existent at contract initiation.

The queries, however, which primarily were entered as card images in-stream on estimate table runs, were not permanently stored anywhere and, hence, were generally unavailable.

As a second step, RTI selected a plan for a query storage and retrieval subsystem that involved storing the query and its associated response estimates in the SYSTEM 2000 data base. The query processing phase of the old NOHS retrieval system would then become the controller between several alternate paths. The data base would first be searched for the desired query; if that query were found, the response estimates associated with that query could be retrieved and written out as a file that becomes input to the report generator. If the query were not found, the estimate file would be produced by the retrieval and estimation procedures and added to the data base along with the query. Thus, the system envisioned by RTI was a dynamic one--one that would incorporate new queries as they were formulated.

Several variations of the design outlined above were considered. The use of mass storage (which Parklawn Computing Center [PCC] now strongly encourages because it has greater storage capacity and is more efficient in access than is either disk or tape) was considered as an alternative since it allowed the use of random access. Random access was a feasible method to examine whether or not an estimate file existed for a query and, as more and more queries were run, was certainly the better alternative when compared to sequential search. However, since sequential access was a reasonable alternative for the immediate future, tape was again a possibility along with mass storage.

For any design alternative, the first step was the separation of the query analysis and processing phases into separate programs so that the processing phase could become a control on what happens next. Part of the effort in developing the query storage and retrieval subsystem was to be the simplification of the query language itself. Dividing the query analysis phase into a separate program from the processing phase facilitated this simplification process. The query analysis program produced a file that directed the operation of the query processing program. As a first step in the simplification, the analysis program was modified to generate, by default, the query criteria that are common to most queries.

The query analysis and processing programs were working together satisfactorily, but further development of the control subsystem became a low-priority item for several key reasons. The query control subsystem was centered around the prior NOHS retrieval system with its own query language and was oriented toward the production of tables in Volume III of the final report for RTI's previous data analysis contract with NIOSH [1]. The query language's allowance for versatility and its somewhat abstruse structure made it difficult to use, and some retrieval aspects it was meant to incorporate turned out to be unimportant after the production of the tables for Volume III. Since a SYSTEM 2000 data base became available, retrieval has been made an efficient process, using a simplified query language, with the use of the Procedural Language Interface feature of SYSTEM 2000 in a COBOL program. The main factor in pushing the development of a query storage and retrieval subsystem to a low priority, however, was the concurrent development of the new projection algorithm. This algorithm was efficient enough that the difference between calculating an estimate and retrieving estimate information was not expected to be significant in most cases. Thus, the query control subsystem idea was no longer considered to be necessary and, therefore, was not completed.

D. Task 4: Redesign Projection Algorithm

Because of the expense and length of time associated with it, the old NOHS projection algorithm program needed to be redesigned and/or redeveloped so as to project national statistics more efficiently. Twenty-five percent of the old algorithm's run time (i.e., 25 percent of its run time as defined in Volume III of the final report for NIOSH Contract No. CDC-99-74-40 [1]) was set as a minimum goal of improvement.

1. Testing and Implementation of Viable Options (Subtasks 4.1-4.7)

Since the statistical theory behind the old projection algorithm was sound but had been complicated to develop, the first attempts at increasing efficiency to twenty-five percent of Volume III run time were simply revisions of the old algorithm. The most obvious improvement to be made was the elimination of the many provisions for test displays, which needed to be extensive during program development because of the many possible paths through the algorithm. Specifically, each of the previously developed formats, which were the basis on which the projection algorithm was built, required different control breaks, and an attempt was made to allow some flexiblity, in the output produced and displays shown, even within paths. Many displays were indeed eliminated in a revised version of the projection algorithm, and some efficiency was gained. Other minor variations, including optimized compiles, were tried on the old algorithm, but the eficiency gain attained with these revisions did not approach the project goal.

The next major step was the more essential matter of looking for ways to reduce the number of times frequently performed routines are executed. In this step valuable information on the operating logic of the old algorithm was also provided. Specifically, RTI used a COBOL-provided TRACE routine that aids the process of analyzing program execution and logic by indicating the order of statement performance. However, printing out the TRACE results is only suitable for small programs and operations because of the potential volume of print-out it produces. Hence, a more appropriate technique was to employ the COBOL TRACE routine and capture the TRACE output on either tape or disk for additional analysis. Captured TRACE output was sorted and summarized by paragraph name to produce a frequency count of dynamic program operation. Figure 5 provides an example of a BASE run using this technique and shows frequency counts for several routines from the old projection algorithm.

Table II presents viable options in the projection algorithm based on collapsing the number of attributes for each of the basic classification variables: size, SIC, BLS average misclassification (ratio of Bureau of Labor Statistics [BLS] SIC-size within observed SIC-size), and groups. The top row in Table II, Base, shows classification attributes for the old projection algorithm. The "CONTROL BREAKS" column is derived as the product of the contributing factors listed in the "SIZE," "SIC," "BLS" (misclassification ratio), and "GROUP" columns.

The options illustrated in Table II are concerned primarily with collapsing size and SIC to more general values. As size and SIC are collapsed, the number of control breaks is reduced, and thus efficiency is potentially gained by reducing the frequency with which many old algorithm routines need to be performed. Thus, while Option 7 does not represent a loss in data element definition any greater than that presented in published NOHS data, it, of all the options, represents the greatest reduction in control breaks and, hence, the greatest gain in efficiency.

Several of the projection options identified in Table II were tested with special versions of the facility file so as to determine their efficiency and accuracy. The results of these tests are recorded in Appendix D. From these tests, data were collected for the Base (old projection algorithm) and for Options 1, 6, and 7. Testing of intermediate Options 2, 3, 4, and 5 was deferred and was not expected to provide sufficient additional information beyond the trends presented in Appendix D. Option 7 indicated an improvement in running time of approximately 50 percent. However, this improvement did not meet the project goal of attaining 25 percent of the old algorithm's run time, and all the other options, which involved more control breaks than Option 7, would be expected to produce even less improvement. Hence, there was a need to identify other options in order to attain greater efficiency from the projection algorithm.

In the search for further improvements to the old projection algorithm, the goal was to do as much one-time or advance preprocessing as possible. This would reduce to an absolute minimum the amount of repetitious processing and, thus, would reduce the processing to a single computer program statement to be incorporated as part of the combined retrieval report printing operation. In order to reduce processing to a single computer program statement, a factor, unique to each facility, that can be applied to produce estimates for the entire data base was ultimately found. However, a consistent factor over the entire data base to produce estimates was not feasible in terms of the statistical framework. Another possible approach was to employ a factor that was unique to each 2-digit SIC. This factor could then be adjusted by size within SIC to allow recognition of the omission of large facilities. Furthermore, another adjustment could be applied to each facility to reflect influence of facility nonresponse. It was the search for these simple projection factors that led to the new projection algorithm.

BEGIN-FREQ-ANALYSIS-OF-PGM-OPERATION	
1 =	1
BEGIN-IN-PROC	1
BEGIN-OUT-PROC	1
END-IN-PROC	1
ACCUM-VAR-NUM-LRG	5
ACCUM-VAR-NUM-LRG-END	6
ACCUM-VAR-NUMS	186
ACCUM-VAR-NUMS-END	242
ADD-GRPS	426
ADD-GRPS-END	571
ADD-GRPS-TS	674
ADD-GRPS-TS-END	748
ADD-GRPS-6	1
ADD-GRPS-6-END	2
ADD-GRPS-7	2
ADD-GRPS-7-END	2
ADD-LARGE-FACS	23
ADD-LEVEL2G	2
ADD-LEVEL2G-END	2
ADD-N	452
ADD-N-END	2,069
ADD-NEXT-TS	118
ADD-NEXT-TS-END	192
ADD-NI	533
ADD-QUAN	43
ADD-QUAN-END	55
ADD-QUAN-2	1,158
ADD-QUANX	1,248
ADD-QUANX-END	316
ADD-REC	127
ADD-REC-END	111
ADD-REC-3	248
ADD-REC-4	147
ADD-REC-7	14
AGAIN-STORE	1,994
BC-CHANGED	2
CALC-NEXT-FACS	111
CALC-NEXT-FACS-END	26
CALC-OUT	25
CALC-OUT-END	12
CHECK-SICS	17
CHK-BC	106
CHK-BC-END	701
CHK-BC-2	647
CHK-EST	962
CHK-EST-END	2,384
COL-CNT-SET	26
CONT-IN-PROC	179
FIND-VAR-DIFFS	186
FIND-VAR-DIFFS-END	242
FIND-VAR-DIFFS6	2
FIND-VAR-DIFFS6-END	2
FIND-VAR-DIFFS7	2
FIND-VAR-DIFFS7-END	3
FORM-CNIRL-KEYS	542
FORM-CNTRL-KEYS-END	18
GRAND-TOTAL	1
GRAND-TOTAL	1
GRAND-TOTALX-END	64
IMPUTE-DONE	

Figure 5. Partial TRACE County for Base Run

TABLE II. PROJECTION ALGORITHM OPTIONS AND FACTORS

	Counts by Category				
OPTION	SIZE	SIC	BLS	GROUP	CONTROL BREAKS
Base	7	65	3	6	8,190
1	4	65	3	6	4,680
2	7	65	3	3	4,095
3	7	30	3	6	3,780
4	7	65	1	6	2,730
5	4	30	1	6	720
6	4	30	2	3	720
7	4	30	1	3	360

2. Concept for New Projection Algorithm (Subtask 4.8)

The old projection algorithm involved the use of imputation procedures to compensate for facility nonresponse and, therefore, did not employ projection factors that were unique to each facility and that were applicable for all estimates. The old projection algorithm did employ a product of two Inverse Inclusion Probability (IIP) multipliers for each facility that was later inflated in an attempt to find such projection factors. The new projection algorithm, on the other hand, gained efficiency by further inflating the IIP product and thus formulated a projection factor that was both unique for each facility and appropriate for all estimates. So as to explain the computational efficiencies incorporated into the new algorithm, a brief description of the old estimation procedure is required.

a. Estimation Procedure

Recall that the NOHS sample design is basically a 2-stage design, with stratification imposed at each stage. First stage units are Standard Metropolitan Statistical Areas that were clustered in arbitrary groups; second stage units are systematically defined clusters of facilities that were selected on the basis of SIC and size. Nonprobability subsamples of large facilities were selected during the data collection period. The estimation procedure adopted involved treating the large facilities that were not surveyed as nonresponding facilities, since these facilities were not omitted on the basis of a purposefully selected subsample.

The estimation procedure involves, first, the multiplication of observation variables at the facility level by the first and second stage inverse inclusion probabilities. The resulting quotients are then summed over three dimensions of stratification (called groups, sizes, and SIC codes in earlier documentation). Within a stratification category of group, size, and SIC, nonresponse adjustments consisted of substituting the average value of respondents in place of missing values. If the size by SIC cell itself was missing, group-level averages were substituted. Reporting domains were defined in terms of observed size and SIC classifications, which may be different from the stratum classification. Other reporting domains, such as those defined by a particular hazard, are defined over the observed size by SIC classifications.

Notationally, it is convenient to denote the three dimensions of stratification by a single subscript,

$$\ell = 1, 2, \ldots, L,$$

where L is the total number of stratum cells in the design (i.e., the number of cells formed by classifying facilities by group and by size and SIC categories within groups). A unit (i.e., facility) in the sample is denoted by u. Associated with the uth unit is the value of some observation variable, denoted by Y_u. If there is no nonresponse, an estimated total is given by the expression

$$\hat{T} = \sum_{\ell=1}^{L} \sum_{u \in \ell} W_u Y_u I_u ,$$

where,

$u \in \ell$ implies the unit belongs to the ℓth stratum cell,

W_u is the inverse of the product of the first and second stage inclusions probabilities, and

$I_u = 1$ if the unit belongs to the reporting domain of interest (denoted by $u \in D$), or $I_u = 0$ otherwise.

Denote units that are members of the respondent set by

$$u \in R,$$

and denote units that are members of the nonrespondent set by

$$u \in R^c .$$

The summation over units within a stratum cell is, in fact, over the set of responding units. The number of units belonging to the respondent set and the ℓth stratum cell is denoted by

$$n\{u \in R \cap \ell\} .$$

The value substituted for the missing units is the average,

$$\bar{Y}_\ell = \sum_{u \in R \cap \ell} (Y_u I_u)/n\{u \in R \cap \ell\} .$$

This is the sum of the observation variables for responding units in the cell which are members of the reporting domain, divided by the total number of responding units in the cell. Note that the average reporting domain size in the cell is accounted for in the calculation above.

If

$$n\{u \in R \cap \ell\} = 0,$$

the adjustment is made by computing the average for the size by SIC cell over all groups. Being given the actual whole unit nonresponse patterns experienced, it is not necessary to collapse stratum cells further. That is, collapsing the group dimension always results in at least one responding facility for each size by SIC cell. The representation

$$u \in R \cap \ell$$

will continue to be used, whether or not the group dimension has been collapsed in ℓ.

The estimate of a total then becomes,

$$\hat{T} = \sum_{\ell=1}^{L} \left[\sum_{u\varepsilon R\cap \ell} W_u Y_u I_u + \sum_{u\varepsilon R^c\cap \ell} W_u \sum_{u\varepsilon R\cap \ell} (Y_u I_u)/n\{u\varepsilon R\cap \ell\} \right] . \quad (1)$$

Writing products of the form $Y_u I_u$ in this expression is simply a matter of notational convenience. In actual fact, the determination

$$u\varepsilon D \quad \text{versus} \quad u\notin D$$

is made for each facility, and the summations include only the former. That is,

$$\sum_{u\varepsilon R\cap \ell} W_u Y_u I_u = \sum_{u\varepsilon R\cap \ell\cap D} W_u Y_u \quad .$$

b. Computational Efficiencies

The computational efficiencies used in the modified algorithm are based on precomputing selected values required by the estimation procedure, storing the precomputed values on each facility record, and then computing a total as a simple sum of products. In addition, use is made of the fact that only whole unit nonresponse occurs in the NOHS data set. That is, individual data elements within a facility are never missing; rather, nonresponse always involves a whole facility. With this knowledge, a single multiplier can be computed that incorporates the inclusion probabilities and the nonresponse adjustment. In statistical terms, weighting class adjustments are computed. These adjustments were based on size by SIC classifications within groups, where possible, or they were computed across groups if no facility within a size by SIC classification was a respondent.

As a first step, the values, W_u, are computed for all units (facilities) in the sample. The original algorithm actually employed the separate first and second stage "weights."

Next consider the following equation:

$$\sum_{u\varepsilon R\cap \ell} \left[W_u + \sum_{u\varepsilon R^c\cap \ell} W_u/n\{u\varepsilon R\cap \ell\} \right] Y_u I_u$$

$$= \sum_{u\varepsilon R\cap \ell} W_u Y_u I_u + \sum_{u\varepsilon R^c\cap \ell} W_u \sum_{u\varepsilon R\cap \ell} Y_u I_u/n\{u\varepsilon R\cap \ell\} .$$

Note that the expression on the right hand side of the equal sign is in fact the value in the brackets in Equation (1). Since only whole unit nonresponse is considered, the values

$$M_u = W_u + \sum_{u\varepsilon R^c\cap \ell} W_u/n\{u\varepsilon R\cap \ell\} \quad (2)$$

can be precomputed and stored for each

$$u\varepsilon R\cap \ell \; .$$

The estimate of total is then written,

$$\hat{T} = \sum_{\ell=1}^{L} \sum_{u \in R \cap \ell} M_u Y_u I_u$$

$$= \sum_{\ell=1}^{L} \sum_{u \in R \cap \ell \cap D} M_u Y_u \quad . \qquad (3)$$

The computational efficiencies result from using Equation (3) in place of Equation (1) or, rather, in place of an equation like (1), except that the values W_u are actually products of two parts. The expresions are, however, algebraically equivalent and should not differ except for differences in the numerical accuracies of the two procedures.

3. Implementation of New Projection Algorithm (Subtask 4.9)

As noted in Section D.1, above, none of the revisions of the old projection algorithm offered a sufficient increase in efficiency. Therefore, development of a different approach utilizing the facility factor idea was pursued.

The product of the inverse inclusion probability (IIP) factors at the first and second stages of the NOHS sample design for each facility was placed on the facility file as the first preprocessing step for the old algorithm. Later, a second factor was formed by increasing this product to compensate for the unsurveyed facilities in each BLS stratification category of group, SIC, and size according to the formula described in Equation (2) of the previous subsection. Such categories were understood to be in the context of observed SIC by size cell, where such cell estimates were then added together as necessary to produce the estimate tables for RTI's previous data analysis contract. The restrictions on each domain of interest in the previous section, thus, can also be understood to include being within a given observed cell.

The next logical approach, then, was the use of the inflated IIP factor already on the facility file, multiplied by the BLS size-1 adjustment when necessary, as a simple projection factor. The sum of inflated IIP factors as a projection of plants and the sum of inflated IIP factors multiplied by payroll count as a projection of people yielded estimates very close to those in Table 1 of Volume III [1], where the factors are adjusted for BLS size-1. A facility file was formed in facility number order with the BLS size-1 adjustments already made. A straightforward program, ALG, was developed to utilize the inflated IIP factor concept. The revised facility file is matched to the sorted exposure records from the retrieval process to obtain the product of the count from each exposure record and the inflated IIP factor of the matching facility. This product is added to the exposure records, which are then sorted by SIC or occupation code so that the products can be summed to produce the desired report.

Among the options for developing viable projection alternatives, the ALG procedure could be considered as Option 8, considerably simpler than the other seven (discussed in subsection D.1, above). Option 8 was developed in an attempt to identify a bottom line, or a lowest cost approach, to projections that maintain the statistical integrity of the old projection algorithm. The total time and projections of total plants and people in plants are illustrated in Appendix D.

The computer programs used to implement the old projection algorithm were studied in considerable depth. Several exploratory computer runs were made to detect the impact of redundant or repetitive calculations on run time. Repeated execution of calculation routines with the same data should produce identical results. The approach taken in Option 8 was to accomplish the critical calculation one time and store the results for later use without employing unnecessary repetition.

The ALG computer program was used to calculate projections for asbestos exposure by SIC. The results were compared to a previously published table; these comparisons are shown in Appendix D. Comparisons indicate a mean difference of approximately 1 percent. RTI also produced a universe projection of total plants and total employees in plants and compared the results to a previously published table; the comparative results are illustrated in Appendix D and are within acceptable limits of about 1 percent in most cases. The ALG program and associated JCL were made available to the NIOSH staff for preliminary use.

The cost reduction in computer run time and cost in using ALG exceeded the project specifications by a considerable margin. The first live run of ALG was made in conjunction with a request for NOHS data regarding occupational distribution of asbestos. From the retrieval on asbestos about two hundred occupations were determined to be involved. The old projection algorithm method would have required approximately 45 seconds of computer CPU time for each occupation, or about 9,000 total seconds. The ALG program produced estimates for all occupations in about 5 seconds. NIOSH personnel selected about 20 of the occupations for projection runs in order to conserve time and cost. The projected estimates from the two methods were compared and found to be within acceptable limits, although more improvement was to come.

The development of the ALG program approach to the calculation of projected estimates from the NOHS data base was an evolutionary process in attempting to capitalize on prior experience with the projection algorithm. A more comprehensive version, NEWALG, was developed in response to an urgent need for processing multiple hazard code special requests. NEWALG included the capability of accepting any number of hazard code 47-byte exposure record sets and producing a series of reports.

The totals obtained for estimated people and plants in the universe needed further verification because the differences, though very acceptable, were significant. The original IIP factor on the facility file were not increased when a group cell within SIC-size had no respondents, causing consistent underestimation for certain SICs. A group-level imputation should have been reflected in the facility file by increasing the inflated IIP factor across the group by the average uninflated IIP factor for the respondents when that situation occurred. The old algorithm compensated instead for group cells with all nonrespondents by averaging the summed quantity over respondents within the other groups for that SIC-size and by multiplying that average by the IIP sum over the cell. The facility file that was formed with the BLS size-1 adjustments was updated to reflect this compensation, and the revised file was sorted into facility number order. National universe results were obtained using both the old and new projection algorithms for the three size levels for each 1-digit SIC. A revised version of NEWALG, ALG3, was created adding new runs to aid with this process. The differences were felt to be insignificant and due to round-off in the newly calculated IIP factors. The continuation of the verification process, however, revealed that an error was involved in the adjustment procedure for missing group cells within SIC-size. The results of the old and new algorithm, including the differences before the error was found, are shown in Appendix D. Corresponding results after facility file processing was corrected are also shown in Appendix D. NIOSH personnel were kept aware of all the improvements in the new algorithm as they evolved.

4. <u>Verification of New Projection Algorithm (Subtask 4.10)</u>

Although the new projection algorithm could be theoretically justified in terms of the old projection algorithm, the verification of equivalence by comparison and manual methods was felt to be absolutely necessary. (For a comprehensive explanation of the original algorithm, see Appendix E.) The estimates of total plants and total employees in plants in the national universe was chosen as the basis of comparison between the old and new algorithm for several reasons. These estimates have no special qualifications on them in terms of facility file information. Thus, these national universe estimates are the most straightforward ones to compute and do not require the use of any query statements, allowing ALG3 to remain a stand-alone program without adding new capabilities. Also, the all-inclusiveness of these estimates, which can nevertheless be computed relatively cheaply, should have exaggerated any error in the procedure. <u>County Business Patterns</u> information from the Bureau of the Census was used to verify the national universe results and thus could provide another validation source for such estimates from either algorithm.

A program was developed to display the facility file information used for national universe estimates for a particular SIC. This program was utilized for SIC 21 to allow manual verification of the theory behind the new algorithm. SIC 21 was selected because the number of records in it was relatively small and the difference between the old and new algorithm results appeared to be significant. In carrying out the manual verification, which is shown in Appendix D, RTI found an error in the implementation of IIP factor modification for the new algorithm as then formulated. The two IIP factors on the facility file needed to be checked for equality before adjustment factors were applied for BLS size-1 or before reimputing was performed across BLS SIC-size. If the factors are unequal, no further compensation needs to be made for that facility since its second IIP factor, the one used for estimation, already includes an imputation increment that is spread across records within the group cell. Thus, the second factor should not be added for redistribution across SIC-size in that case. A change to correct this error in implementation of the new algorithm was made, and the resultant run gave essentially the results found with the old algorithm, as also seen in Appendix D.

In making runs requested of it, NIOSH discovered problems in ALG3, though none related essentially to the new projection algorithm itself. Among these problems were differences in estimated plants and people for the same hazard according to whether a report was generated by SIC or occupation, and a similar problem was found on change of hazard. For each descrepancy, a further control break was needed under a particular run-switch. Also, the same hazard name was printed for two different hazard codes, corresponding to two distinct rank codes. However, a check of the hazard files for duplicate names showed the new algorithm to be operating properly. This conclusion was reported to NIOSH, and indications were made that the process of dealing with such duplicate names was continuing.

5. Development of Efficient Variance Estimation (Subtask 4.11)

The variance estimation under the old algorithm was soundly based on the original statistical framework of group within BLS SIC-size within observed SIC-size. (For a complete description of the variance estimator derived for the NOHS data base utilizing the old algorithm, see Appendix E.) This operation was slow, although straightforward, to perform because the calculation was performed for each report column of a previous estimate table across group within BLS SIC-size within observed SIC-size. Furthermore, for the exposure estimate tables, the operations were repeated across all cells within every major control break, such as change of hazard or occupation. The time required was even greater for computing the variance of a ratio estimate because more calculations were involved for covariance at each BLS SIC-size level and for variance calculation at each higher cell level. Also, in order to impute over SIC-size cells for groups containing all nonrespondents, the processing for all estimates was performed in these stages under the old algorithm.

With the new algorithm, a very efficient method of computing estimates has been found. The adaption of the variance estimator from the old to the new algorithm in a manner that would not require the addition of many intermediate calculations to the procedure did not seem feasible. Nor did there seem to be any alternative variance estimator that would fit the original statistical framework and still not significantly affect program efficiency.

Yet there was felt to be a significant need for a statistical measure of estimate reliability. Thus, as an option, the ALG3 program was amended to supply a set of tables at the front of an estimate report that give various example estimates of people in plants and estimates of plants matched to standard deviations. There are large differences in the standard deviations for people estimates when large plants contribute more than one-third of the estimate. This difference is probably due to a large estimate from relatively few observations in that case. The program indicates which deviation to use for relevant estimates, and an explanation of the proper use of the tables is included with them. Particularly important is the fact that the standard deviation tables were derived from what seems to be the average of previous estimate experiences and are by no means exact. The inexactness is such that even interpolation for points between example estimates is numerically unsound. These standard deviation tables, nevertheless, do provide an extremely crude index of reliability without significantly decreasing program efficiency. The tables and the explanation included with them can be found in Appendix F.

E. Task 5: Develop Subsystem to Link NOHS and RTECS

Formerly known as the Toxic Substances List, the Registry of Toxic Effects of Chemical Substances (RTECS) is a collection of published literature and quantitative data on toxic effects with assigned Chemical Abstracts Sequence (CAS) numbers. The NOHS data base is a collection of data indicating the relationship of occupations, industries, and hazards.

Frequently, the operation of one data base can be enhanced by appropriate accessibility and availability of another data base. The independent operation of the NOHS and RTECS data bases was restricted to the information contained in each system. Accessibility of one system to the other, developed by RTI under Task 5, provides a wider scope or range of available information.

1. Compatible Code Relationships Between NOHS and RTECS (Subtasks 5.1 and 5.2)

Access to the NOSH and RTECS data bases implies more than administrative approval to retrieve information from either; it includes a necessary linking or cross-reference method. Chemical substances are identified by CAS numbers in the RTECS data base, and chemical hazards are identified by hazard codes in the NOHS data base. The need for an appropriate cross-reference table of hazard codes and CAS numbers was earlier anticipated by NIOSH and is nearing completion under a separate contract. The entry of CAS numbers by hazard code into the NOHS data base facilitates development of appropriate linkage systems to provide the desired accessibility of one data base to the other.

The linkage of the NOHS and RTECS data bases enlarged the potentials of both data bases so that each becomes a logical extension of the other. Ideally, a suitable linkage system should incorporate bi-directional features. Thus, a search or retrieval in one data base could invoke a secondary retrieval from the other data base. For example, a search in the NOHS data base on occupation or industry could establish a list of hazards (with CAS numbers). The list of hazard codes and CAS numbers could then be used to retrieve toxic effect information from RTECS. Another example might be a search of RTECS that produces a subfile of CAS numbers. The list of CAS numbers could be processed to initiate a NOHS search to indicate the occupational and industrial distribution for the associated chemical substances.

Final implementation of the NOHS-RTECS retrieval link system was dependent on the quality and delivery date of the hazard file (with CAS numbers applied to the hazard file) and had to be tested without benefit of complete files. CAS numbers may not be available for certain generic type hazard codes and may require further special treatment not currently considered.

The code relationships that were found to exist, then, are trilateral, with linkages among hazard code, CAS number, and RTECS ascension number. From use of the RTECS ascension number, the toxicity information from the RTECS master file becomes available, and the CAS number is duplicated both on the CAS file and on the RTECS file. The hazard file, in turn, contains both the CAS number and RTECS ascension number.

2. Implementation of Link Subsystem (Subtasks 5.3 and 5.4)

One preliminary tentative approach to a system to link NOHS and RTECS involved two computer program modules. These modules revolved around the use of the Data Base Management System MARK IV, but the utilization of MARK IV was felt not to offer any significant advantages over program development without the use of a common Data Base Management System (DBMS).

Virtual Storage Access Method (VSAM) was then tested extensively during the development of a link subsystem as well as during other project tasks. VSAM is purported to be much more efficient than Indexed Sequential Access Method (ISAM) and does offer the availability of alternative keys. However, the experience of this project does not show VSAM to be as efficient as ISAM for small files, probably because greater overhead time--extra time before and after actual access--is involved with VSAM. No guidelines for the approximate file size at which VSAM is more efficient than ISAM could be established from the project testing. The logic used for VSAM, then, became the logic adapted to ISAM and formed the basis of the link subsystem actually implemented.

The NOHS-RTECS Information System (NORIS) was designed and implemented with the test data files that became available in April 1978. NORIS was designed to allow information expansion, utilizing ISAM files, from the NOHS hazard file to CAS and RTECS files. Access to the additional files provides information regarding the chemical and toxicity characteristics of a given hazard code.

NORIS consists of three data base segments. The first, or entry, segment is a NOHS hazard file with added fields for CAS and RTECS numbers. The second segment is a copy of the CAS file with a blind suffix number added to the CAS number to create unique CAS Indexed Keys. The third segment is an unmodified copy of a selected portion of the RTECS file.

A preliminary version of NORIS was programmed and data was loaded into the three data base segments. Sample computer test runs were made and printed at both the RTI and Cincinnati NIOSH terminals for demonstration purposes. The Job Control Language (JCL) and instructions for the use of NORIS were also made available. The use of NORIS is recommended in order to determine the applicability of the extended toxicity information to the NOHS data.

The NORIS computer programs and data base segments were located on private mass storage at the Parklawn Computing Center (PCC). Response time to NORIS information requests is in the order of minutes and can be submitted in the most rapid processing operational job class (i.e., class = E, Time = 5 seconds). Access to a Time Sharing Option (TSO) type terminal facilitates response time and bypasses contention encountered with a batch type terminal. However, NORIS is operational under either type of terminal and selection is reduced to personal preference.

Information retrieval is achieved by submitting a single hazard code to NORIS. A random access is performed on the hazard segment to retrieve the modified hazard record with the hazard name or description assigned by NIOSH. If the hazard retrieval receives a nonblank CAS number, a retrieval is attempted to the CAS segment. The CAS segment retrieval will provide information regarding other hazard codes considered as duplicates by Chemical

Abstracts Services and list all synonyms. The special COBOL verb START is used to position the file pointer by a partial key of the CAS number itself, and the file can then be read sequentially until this number changes.

A nonblank RTECS number on the hazard segment retrieval will attempt retrieval from the RTECS data base segment. A successful RTECS retrieval will print all of the data currently stored for the specified RTECS number.

The segmented approach to the NORIS design was selected in order to acheive a cost effective functional operation. The relative cost of data base maintenance (data base loading and updating) needs to be balanced to the anticipated retrieval cost. Combination of data from the three sources into a single hierachical data base possibly would reduce the retrieval cost slightly but would increase the loading and update cost considerably.

The dynamic nature of the RTECS data results in the RTECS master file being updated on a periodic basis. The RTECS update procedure produces a new RTECS file for distribution purposes. Therefore, the NORIS update of the RTECS segment is reduced to a simple file replacement operation without the added cost of a randomized update procedure.

V. DISCUSSION OF FINDINGS

The NOHS data provide potentially valuable information regarding the distribution of possibly hazardous exposures to employees across industrial and occupational categories. This information, of course, must be understood in its proper statistical context. The companies selected for inclusion in the data base were picked from those in each industry with eight or more employees, and the first stage of the sampling framework restricts those selected to certain metropolitan areas. Hence, the survey results can only be used as the basis of national estimates applying to companies of eight or more employees in metropolitan areas as of 1972. The restriction on date must be made because no updating of the original NOHS survey data has occurred through comparative analysis by other surveys and through appropriate industry-specific surveys as originally contemplated. This process has not implemented, partially because NIOSH still is anticipating having a new NOHS undertaken with more efficiency and with much more careful planning and control from a statistical point of view. Such a survey would probably have a stronger update plan included from the beginning also.

There is an unknown amount of bias in the present NOHS results due to the selecting of the sample from 1970 data. This planning was without regard to the fact that some companies might have gone out of business or been relocated between 1970 and 1972. Compensating for the fact that a sample cut was made of large facilities in a nonrandom manner also created statistical difficulties.

Despite these restrictions, useful information from NOHS is obtainable, and, during RTI's two data analysis contracts, the NOHS data have been edited, summarized many different ways, and used to produce estimates. Many of these estimates were collected into tables in Volume III of the final report for the previous RTI contract (No. CDC-99-74-40) [1], and retrieval capabilities have been made abundant, even more so now that the NOHS data base is maintained under SYSTEM 2000. The development of a very efficient projection algorithm and the subsequent development of a more useful "Redbook" summary and the "All Information" Report are probably the major accomplishments of the present contract, particularly in view of the fact that the NOHS survey, from a statistical point of view, was not especially well implemented. The linkage of NOHS and RTECS, as far as it has been pursued, offers a promise of giving comprehensive toxicity and exposure distribution information. The creation of a dynamic, reasonably efficient update subsystem to reflect continued progress in trade name resolution within the data base structure was a vital accomplishment. SYSTEM 2000 seems to be proving quite satisfactory in loading and retrieval processes. Thus, overall, the NOHS system offers a very effective utilization of the data base information, and the system is essentially complete, except for such ongoing processes as trade name resolution.

There are some improvements to the NOHS system that could be considered to increase its usefulness. First, the idea of a comprehensive query language, perhaps coupled with a query control subsystem of some sort, still has merit. The query language developed during the previous RTI data analysis contract offered comprehensive request capabilities, but many users found its algebraic quality abstruse. Furthermore, it was oriented toward facility file criteria and was never clearly explained. Therefore, that query language, designed primarily to meet the needs of estimate requests for the previous data analysis contract, was difficult to use and was never heavily utilized by NIOSH personnel. A more limited language was substituted instead. A powerful, easily usable query language could be a very worthwhile tool to develop and couple with efficient retrieval techniques utilizing SYSTEM 2000.

A second area of improvement would be the further development of RPTMAIN after coupling it with the new projection algorithm. Any formats previously developed and not presently incorporated might be brought into the concept if they are particularly useful. At one time, the use of two basic formats, one having three values for each estimate and the other having eight values for two estimates where the ratio between the two estimates is needed, was contemplated. In the former case, the three values would be the estimate, a very approximate standard deviation, and record count (n). The second format would contain this information for both numerator and denominator and also the ratio and approximate standard deviation. There might still be a possibility that such reports could be a useful development, particularly with completely flexible entry points on hazard, occupation, or industry. The reports developed during this contract, however, met most NIOSH needs, and thus these ideas were not further implemented.

Third, the NOHS-RTECS linkage could have more potential if it were developed to become wholly bilateral. Currently, a matched version of the NOHS hazard file is used to access RTECS data, including the CAS synonyms, through the CAS and RTECS ascension numbers. Accessing a matched NOHS hazard file through CAS or RTECS ascension numbers would allow additional flexibility and could lead to interesting analysis possibilities if the NOHS exposure data were, in turn, utilized. Such a comprehensive NOHS-RTECS link subsytem might well be considered for further study, although the existing link subsystem was felt to be satisfactory for the present.

VI. CONCLUSIONS AND RECOMMENDATIONS

A. Conclusions

The NOHS system offers a fairly complete, efficient analysis method for update, retrieval, summarization, and estimation operations on an edited, congruent data base. This data base offers a valuable source of information regarding the 1972 distribution of potentially hazardous employee exposures across industrial and occupational categories in metropolitan areas throughout the United States. Some general improvements for the current NOHS system are recommended below.

B. Recommendations

First, an expanded query logic would enhance the value of the data base, particularly for research purposes. Therefore, RTI recommends that effort be expended in developing and implementing a query processing subsystem that extends the present retrieval capability.

Second, the new ALG program would be better suited to the expanded query logic if it were linked to RPTMAIN and the "All Information" Report capability. RTI recommends that this expanded report capability be accomplished concurrent with the expanded query capability.

Third, a bilateral linkage between RTECS and NOHS would increase the effectiveness of both data bases; therefore, RTI recommends that the necessary changes be made in the existing linkage system to achieve this capability.

VII. REFERENCES

1. U.S. Department of Health, Education, and Welfare. National Institute for Occupational Safety and Health. National Occupational Hazard Survey. Volume III. Survey Analysis and Supplemental Tables. DHEW (NIOSH) Publication No. 78-114. Washington, D.C.: U.S. Government Printing Office, December 1977.

APPENDIX A

EXAMPLES OF ALG3 REPORTS

APPENDIX A contains an ALG3 estimate report in each of the two possible report orders--SIC code and occupation code. These reports are given within hazard, where more than one hazard may be involved. For the examples given, only one hazard is involved--code 20560, which represents "dimeythl formamide, N, N-." In addition to each SIC or occupation code and description, the reports display estimated plants, people exposed in plants, and exposures, along with grand totals.

EXAMPLE ALG3 REPORT IN INDUSTRY CODE ORDER 04/24/79

HAZ DESCRIPTION

20560 DIMETHYL FORMAMIDE, N,N-

SIC CODE	DESCRIPTION	ESTIMATED PLANTS	ESTIMATED PEOPLE	ESTIMATED EXPOSURES
13	OIL AND GAS EXTRACTION	10	30	40
23	APPAREL AND OTHER TEXTILE PRODUCTS	30	61	122
26	PAPER AND ALLIED PRODUCTS	6	178	475
28	CHEMICALS AND ALLIED PRODUCTS	276	23,356	30,279
29	PETROLEUM AND COAL PRODUCTS	14	42	126
30	RUBBER AND PLASTICS PRODUCTS, NEC	19	1,680	2,469
32	STONE, CLAY, AND GLASS PRODUCTS	30	30	60
34	FABRICATED METAL PRODUCTS	10	610	1,220
35	MACHINERY, EXCEPT ELECTRICAL	59	562	1,080
36	ELECTRICAL EQUIPMENT AND SUPPLIES	94	1,421	2,653
38	INSTRUMENTS AND RELATED PRODUCTS	43	160	343
39	MISCELLANEOUS MANUFACTURING INDUSTRIES	11	22	44
73	MISCELLANEOUS BUSINESS SERVICES	10	127	167
80	MEDICAL AND OTHER HEALTH SERVICES	28	756	756
TOTAL		641	29,035	39,833

Preceding page blank

EXAMPLE ALG3 REPORT IN OCCUPATION CODE ORDER 04/10/79

HAZ	DESCRIPTION
20560	DIMETHYL FORMAMIDE, N,N-

OCC CODE	DESCRIPTION	ESTIMATED PLANTS	ESTIMATED PEOPLE	ESTIMATED EXPOSURES
013	INDUSTRIAL ENGINEERS	10	39	39
023	ENGINEERS, N.E.C.	10	39	39
045	CHEMISTS	463	2,500	5,265
055	OPERATIONS AND SYSTEMS RESEARCHERS AND ANALYS	28	28	28
080	CLINICAL LABORATORY TECHNOLOGISTS AND TECHNIC	33	753	767
150	AGRICULTURE AND BIOLOGICAL TECHNICANS, EXCEPT	33	15,477	15,477
151	CHEMICAL TECHNICANS	233	1,983	4,243
153	ELECTRICAL AND ELECTRONIC ENGINEERING TECHNIC	13	47	95
162	ENGINEERING AND SCIENCE TECHNICIANS, N.E.C.	107	304	617
195	RESEARCH WORKERS, NOT SPECIFIED	6	178	475
233	SALES MANAGERS, EXCEPT RETAIL TRADE	28	56	56
245	MANAGERS AND ADMINISTRATORS, N.E.C.	66	188	244
323	EXPEDITERS AND PRODUCTION CONTROLLERS	10	49	49
381	STOCK CLERKS AND STOREKEEPERS	20	70	130
392	WEIGHERS	10	118	235
415	CARPENTERS	10	10	10
430	ELECTRICIANS	10	39	39
441	FOREMEN, N.E.C.	81	409	674
461	MACHINISTS	10	49	49
475	DATA PROCESSING MACHINE REPAIRMEN	52	519	1,037
495	NOT SPECIFIED MECHANICS AND REPAIRMEN	28	28	28
506	OPTICIANS, AND LENS GRINDERS AND POLISHERS	22	22	43
522	PLUMBERS AND PIPE FITTERS	10	69	69
530	PRESSMEN AND PLATE PRINTERS, PRINTING	10	200	400
575	CRAFTSMEN AND KINDRED WORKERS, N.E.C.	10	10	10
602	ASSEMBLERS	67	430	629
604	BOTTLING AND CANNING OPERATIVES	14	196	588
610	CHECKERS, EXAMINERS, AND INSPECTORS; MANUFACT	37	252	336
612	CUTTING OPERATIVES, N.E.C.	38	385	434
641	MIXING OPERATIVES	46	193	386
642	OILERS AND GREASERS, EXC. AUTO	10	10	10
643	PACKERS AND WRAPPERS, EXCEPT MEAT AND PRODUCE	28	56	56
644	PAINTERS, MANUFACTURED ARTICLES	10	10	20
680	WELDERS AND FLAME-CUTTERS	10	29	49
690	MACHINE OPERATIVES, MISCELLANEOUS SPECIFIED	144	2,117	3,444
692	MACHINE OPERATIVES, NOT SPECIFIED	68	617	1,169
694	MISCELLANEOUS OPERATIVES	38	160	169
695	NOT SPECIFIED OPERATIVES	41	57	86
706	FORK LIFT AND TOW MOTOR OPERATIVES	28	112	112
753	FREIGHT AND MATERIAL HANDLERS	28	84	84
762	STOCK HANDLERS	8	56	112
785	NOT SPECIFIED LABORERS	21	41	41
902	CLEANERS AND CHARWOMEN	28	56	56
903	JANITORS AND SEXTONS	40	992	1,935
TOTAL		*	29,035	39,833

* ESTIMATED PLANTS NOT ADDITIVE BY OCCUPATION

APPENDIX B

EXAMPLE OF "ALL INFORMATION" REPORT

Enclosed is an example of the "All Information" Report, which is based on using the redesigned projection algorithm to produce a large number of estimates necessary to generate the report. The "All Information" Report gives estimates for each new combination of SIC, occupation code, and hazard code if the user specifies "S" in Column 1 of the control card, "O" in Column 2, and "H" in Column 3. Thus, by not specifying one or any combination of these values, the user may choose not to consider SIC, occupation code, or hazard code changes; if Columns 1 through 3 are left blank, only hazard code changes will be considered. The particular example report shown had "SOH" specified so that any new SIC-occupation-hazard combination will produce report estimates. The estimates included are for plants, exposed people in plants, exposures, and source of exposure by each possible value for the condition fields. The latter estimates are expressed as percentages. For the example report, only the hazard "Phthalic Acid Esters," with code M0382, is involved. Columns 11-15 are used to stop processing after SIC 28 and after occupation code 162, respectively. The "All Information" Report should provide useful estimates for a wide range of future NIOSH needs.

Preceding page blank

28 CHEMICALS AND ALLIED PRODUCTS
045 CHEMISTS
M0382 PHTHALIC ACID ESTERS

ESTIMATED PLANTS	ESTIMATED PEOPLE	ESTIMATED EXPOSURES
49	244	607

FORM	% D	% F	% G	% L	% M	% P	% S	% V	% UNK
				40				60	

INTENSITY	% D	% U	% UNK
	90	10	

DURATION	% F	% P	% UNK
		100	

CONTROL	% EP	% EF	% FP	% HG	% HP	% PC	% CR	% FR	% GR	% OR	% SR	% DV	% LV	% NV	% IC	% LT	% OC	% NC	% UNK
												33	26					40	

FUNCTION	% F	% N	% UNK
	33	26	

28 CHEMICALS AND ALLIED PRODUCTS
151 CHEMICAL TECHNICANS
M0382 PHTHALIC ACID ESTERS

ESTIMATED PLANTS	ESTIMATED PEOPLE	ESTIMATED EXPOSURES
49	314	860

FORM	% D	% F	% G	% L	% M	% P	% S	% V	% UNK
				37				63	

INTENSITY	% D	% U	% UNK
	95	5	

DURATION	% F	% P	% UNK
		100	

CONTROL	% EP	% EF	% FP	% HG	% HP	% PC	% CR	% FR	% GR	% OR	% SR	% DV	% LV	% NV	% IC	% LT	% OC	% NC	% UNK
												32	32					37	

FUNCTION	% F	% N	% UNK
	32	32	

CODE DESCRIPTION DATE 04/20/79

28 CHEMICALS AND ALLIED PRODUCTS
441 FOREMEN, N.E.C.
M0382 PHTHALIC ACID ESTERS

ESTIMATED PLANTS	ESTIMATED PEOPLE	ESTIMATED EXPOSURES
28	28	28

FORM	% D	% F	% G	% L	% M	% P	% S	% V	% UNK
								100	

INTENSITY	% D	% U	% UNK
	100		

DURATION	% F	% P	% UNK
		100	

CONTROL	% EP	% EF	% FP	% HG	% HP	% PC	% CR	% FR	% GR	% OR	% SR	% DV	% LV	% NV	% IC	% LT	% OC	% NC	% UNK
														100					

FUNCTION	% F	% N	% UNK
	100		

28 CHEMICALS AND ALLIED PRODUCTS
753 FREIGHT AND MATERIAL HANDLERS
M0382 PHTHALIC ACID ESTERS

ESTIMATED PLANTS	ESTIMATED PEOPLE	ESTIMATED EXPOSURES
28	84	84

FORM	% D	% F	% G	% L	% M	% P	% S	% V	% UNK
								100	

INTENSITY	% D	% U	% UNK
	100		

DURATION	% F	% P	% UNK
		100	

CONTROL	% EP	% EF	% FP	% HG	% HP	% PC	% CR	% FR	% GR	% OR	% SR	% DV	% LV	% NV	% IC	% LT	% OC	% NC	% UNK
														100					

FUNCTION	% F	% N	% UNK
	100		

CODE DESCRIPTION DATE 04/20/79

29 PETROLEUM AND COAL PRODUCTS
472 AUTOMOBILE BODY REPAIRMEN
M0302 PHTHALIC ACID ESTERS

ESTIMATED PLANTS	ESTIMATED PEOPLE	ESTIMATED EXPOSURES
19	19	57

FORM

% D	% F	% G	% L	% M	% P	% S	% V	% UNK
33			33				33	

INTENSITY

% D	% U	% UNK
100		

DURATION

% F	% P	% UNK
	100	

CONTROL

% EP	% EF	% FP	% HG	% HP	% PC	% CR	% FR	% GR	% OR	% SR	% DV	% LV	% NV	% IC	% LT	% OC	% NC	% UNK
							33				33						33	

FUNCTION

% F	% N	% UNK
67		

30 RUBBER AND PLASTICS PRODUCTS, NEC
441 FOREMEN, N.E.C.
M0302 PHTHALIC ACID ESTERS

ESTIMATED PLANTS	ESTIMATED PEOPLE	ESTIMATED EXPOSURES
11	32	32

FORM

% D	% F	% G	% L	% M	% P	% S	% V	% UNK
							100	

INTENSITY

% D	% U	% UNK
100		

DURATION

% F	% P	% UNK
	100	

CONTROL

% EP	% EF	% FP	% HG	% HP	% PC	% CR	% FR	% GR	% OR	% SR	% DV	% LV	% NV	% IC	% LT	% OC	% NC	% UNK
											100							

FUNCTION

% F	% N	% UNK
100		

30 RUBBER AND PLASTICS PRODUCTS, NEC
764 VEHICLE WASHERS AND EQUIPMENT CLEANERS
M0382 PHTHALIC ACID ESTERS

ESTIMATED PLANTS	ESTIMATED PEOPLE	ESTIMATED EXPOSURES
8	16	16

FORM	% D	% F	% G	% L	% M	% P	% S	% V	% UNK
								100	

INTENSITY	% D	% U	% UNK
	100		

DURATION	% F	% P	% UNK
		100	

CONTROL	% EP	% EF	% FP	% HG	% HP	% PC	% CR	% FR	% GR	% OR	% SR	% DV	% LV	% NV	% IC	% LT	% OC	% NC	% UNK
												100							

FUNCTION	% F	% N	% UNK
	100		

30 RUBBER AND PLASTICS PRODUCTS, NEC
785 NOT SPECIFIED LABORERS
M0382 PHTHALIC ACID ESTERS

ESTIMATED PLANTS	ESTIMATED PEOPLE	ESTIMATED EXPOSURES
11	194	194

FORM	% D	% F	% G	% L	% M	% P	% S	% V	% UNK
								100	

INTENSITY	% D	% U	% UNK
	83	17	

DURATION	% F	% P	% UNK
		100	

CONTROL	% EP	% EF	% FP	% HG	% HP	% PC	% CR	% FR	% GR	% OR	% SR	% DV	% LV	% NV	% IC	% LT	% OC	% NC	% UNK
												100							

FUNCTION	% F	% N	% UNK
	100		

30 RUBBER AND PLASTICS PRODUCTS, NEC
481 HEAVY EQUIPMENT MECHANICS, INCL. DIESEL
M0382 PHTHALIC ACID ESTERS

ESTIMATED PLANTS	ESTIMATED PEOPLE	ESTIMATED EXPOSURES
11	76	76

FORM	% D	% F	% G	% L	% M	% P	% S	% V	% UNK
								100	

INTENSITY	% D	% U	% UNK
		100	

DURATION	% F	% P	% UNK
		100	

CONTROL	% CP	% EF	% FP	% HG	% HP	% PC	% CR	% FR	% GR	% OR	% SR	% DV	% LV	% NV	% IC	% LT	% OC	% NC	% UNK
													100						

FUNCTION	% F	% N	% UNK
	100		

30 RUBBER AND PLASTICS PRODUCTS, NEC
690 MACHINE OPERATIVES, MISCELLANEOUS SPECIFIED
M0382 PHTHALIC ACID ESTERS

ESTIMATED PLANTS	ESTIMATED PEOPLE	ESTIMATED EXPOSURES
19	345	345

FORM	% D	% F	% G	% L	% M	% P	% S	% V	% UNK
								100	

INTENSITY	% D	% U	% UNK
	100		

DURATION	% F	% P	% UNK
		100	

CONTROL	% CP	% EF	% FP	% HG	% HP	% PC	% CR	% FR	% GR	% OR	% SR	% DV	% LV	% NV	% IC	% LT	% OC	% NC	% UNK
												100							

FUNCTION	% F	% N	% UNK
	100		

37 TRANSPORTATION EQUIPMENT
680 WELDERS AND FLAME-CUTTERS
M0382 PHTHALIC ACID ESTERS

ESTIMATED PLANTS	ESTIMATED PEOPLE	ESTIMATED EXPOSURES
48	528	1,056

FORM	% D	% F	% G	% L	% M	% P	% S	% V	% UNK
				50				50	

INTENSITY	% D	% U	% UNK
	100		

DURATION	% F	% P	% UNK
		100	

CONTROL	% EP	% EF	% FP	% HG	% HP	% PC	% CR	% FR	% GR	% OR	% SR	% DV	% LV	% NV	% IC	% LT	% OC	% NC	% UNK
														50				50	

FUNCTION	% F	% N	% UNK
	50		

39 MISCELLANEOUS MANUFACTURING INDUSTRIES
162 ENGINEERING AND SCIENCE TECHNICIANS, N.E.C.
M0382 PHTHALIC ACID ESTERS

ESTIMATED PLANTS	ESTIMATED PEOPLE	ESTIMATED EXPOSURES
11	33	33

FORM	% D	% F	% G	% L	% M	% P	% S	% V	% UNK
								100	

INTENSITY	% D	% U	% UNK
	100		

DURATION	% F	% P	% UNK
		100	

CONTROL	% EP	% EF	% FP	% HG	% HP	% PC	% CR	% FR	% GR	% OR	% SR	% DV	% LV	% NV	% IC	% LT	% OC	% NC	% UNK
												100							

FUNCTION	% F	% N	% UNK
	100		

39 MISCELLANEOUS MANUFACTURING INDUSTRIES
624 GRADERS AND SORTERS, MANUFACTURING
M0382 PHTHALIC ACID ESTERS

ESTIMATED PLANTS	ESTIMATED PEOPLE	ESTIMATED EXPOSURES
11	165	495

FORM	% D	% F	% G	% L	% M	% P	% S	% V	% UNK
				33	33			33	

INTENSITY	% D	% U	% UNK
	100		

DURATION	% F	% P	% UNK
	100		

CONTROL	% EP	% EF	% FP	% HG	% HP	% PC	% CR	% FR	% GR	% OR	% SR	% DV	% LV	% NV	% IC	% LT	% OC	% NC	% UNK
												67						33	

FUNCTION	% F	% N	% UNK
	67		

39 MISCELLANEOUS MANUFACTURING INDUSTRIES
643 PACKERS AND WRAPPERS, EXCEPT MEAT AND PRODUCE
M0382 PHTHALIC ACID ESTERS

ESTIMATED PLANTS	ESTIMATED PEOPLE	ESTIMATED EXPOSURES
11	187	561

FORM	% D	% F	% G	% L	% M	% P	% S	% V	% UNK
				33	33			33	

INTENSITY	% D	% U	% UNK
	100		

DURATION	% F	% P	% UNK
	100		

CONTROL	% EP	% EF	% FP	% HG	% HP	% PC	% CR	% FR	% GR	% OR	% SR	% DV	% LV	% NV	% IC	% LT	% OC	% NC	% UNK
												67						33	

FUNCTION	% F	% N	% UNK
	67		

39 MISCELLANEOUS MANUFACTURING INDUSTRIES
656 PUNCH AND STAMPING PRESS OPERATIVES
M0382 PHTHALIC ACID ESTERS

ESTIMATED PLANTS	ESTIMATED PEOPLE	ESTIMATED EXPOSURES
11	33	99

FORM

% D	% F	% G	% L	% M	% P	% S	% V	% UNK
			33	33			33	

INTENSITY

% D	% U	% UNK
100		

DURATION

% F	% P	% UNK
100		

CONTROL

% EP	% EF	% FP	% HG	% HP	% PC	% CR	% FR	% GR	% OR	% SR	% DV	% LV	% NV	% IC	% LT	% OC	% NC	% UNK
											67						33	

FUNCTION

% F	% N	% UNK
67		

39 MISCELLANEOUS MANUFACTURING INDUSTRIES
690 MACHINE OPERATIVES, MISCELLANEOUS SPECIFIED
M0382 PHTHALIC ACID ESTERS

ESTIMATED PLANTS	ESTIMATED PEOPLE	ESTIMATED EXPOSURES
11	440	2,486

FORM

% D	% F	% G	% L	% M	% P	% S	% V	% UNK
			16	40			43	

INTENSITY

% D	% U	% UNK
100		

DURATION

% F	% P	% UNK
91	9	

CONTROL

% EP	% EF	% FP	% HG	% HP	% PC	% CR	% FR	% GR	% OR	% SR	% DV	% LV	% NV	% IC	% LT	% OC	% NC	% UNK
							19				34	31					16	

FUNCTION

% F	% N	% UNK
34	50	

CODE DESCRIPTION DATE 04/20/79

39 MISCELLANEOUS MANUFACTURING INDUSTRIES
706 FORK LIFT AND TOW MOTOR OPERATIVES
M0382 PHTHALIC ACID ESTERS

ESTIMATED PLANTS	ESTIMATED PEOPLE	ESTIMATED EXPOSURES
11	33	99

FORM

% D	% F	% G	% L	% M	% P	% S	% V	% UNK
			33	33			33	

INTENSITY

% D	% U	% UNK
100		

DURATION

% F	% P	% UNK
100		

CONTROL

% EP	% EF	% FP	% HG	% HP	% PC	% CR	% FR	% GR	% OR	% SR	% DV	% LV	% NV	% IC	% LT	% OC	% NC	% UNK
											67						33	

FUNCTION

% F	% N	% UNK
67		

39 MISCELLANEOUS MANUFACTURING INDUSTRIES
753 FREIGHT AND MATERIAL HANDLERS
M0382 PHTHALIC ACID ESTERS

ESTIMATED PLANTS	ESTIMATED PEOPLE	ESTIMATED EXPOSURES
11	451	1,353

FORM

% D	% F	% G	% L	% M	% P	% S	% V	% UNK
			17	17			66	

INTENSITY

% D	% U	% UNK
100		

DURATION

% F	% P	% UNK
100		

CONTROL

% EP	% EF	% FP	% HG	% HP	% PC	% CR	% FR	% GR	% OR	% SR	% DV	% LV	% NV	% IC	% LT	% OC	% NC	% UNK
							16				50	16					17	

FUNCTION

% F	% N	% UNK
50	33	

CODE DESCRIPTION DATE 04/20/79

39 MISCELLANEOUS MANUFACTURING INDUSTRIES
785 NOT SPECIFIED LABORERS
M0382 PHTHALIC ACID ESTERS

ESTIMATED PLANTS	ESTIMATED PEOPLE	ESTIMATED EXPUSURES
11	176	528

FORM	% D	% F	% G	% L	% M	% P	% S	% V	% UNK
				33	33			33	

INTENSITY	% D	% U	% UNK
	100		

DURATION	% F	% P	% UNK
	100		

CONTROL	% EP	% EF	% FP	% HG	% HP	% PC	% CR	% FR	% GR	% OR	% SR	% DV	% LV	% NV	% IC	% LT	% OC	% NC	% UNK
												67						33	

FUNCTION	% F	% N	% UNK
	67		

APPENDIX C

"REDBOOK" REPORT EXAMPLES

Examples of the "Redbook" summary report in both hazard code and hazard description order are contained herein. Each report displays, for each code and description, the CAS and RTECS numbers, estimated people exposed, exposure percentage by each exposure source (observed, trade name component, or generic component), and count of occupation codes. The "Redbook" summary report procedure involves two steps: (1) summarizing the data for the desired quantities to produce an output file, which also contains the necessary descriptive information, and (2) using that summary file to generate the report in either hazard code or hazard description order. The main input file to the first step is the 32-byte records produced by the second CUPDATE step of the update subsystem (see Figure 2 in Section IV.B), which exactly reflects the revised load file produced during the same step. (The 32-byte file used for the enclosed report was derived from the test load file available to RTI, not from an actual load file.) The "Redbook" summary report, then, is the last step in the update subsystem, parallel to the SYSTEM 2000 reload, that gives summary information rpresentative of the reloaded data base contents.

"Red Book" Summary Report
in Hazard Code Order

Preceding page blank

HAZ CODE	CAS NUMBER	RTECS NUMBER	ESTIMATED PEOPLE EXPOSED	SOURCE OF EXP. EST. ACT	TRN	GEN	NO OCCS	CHEMICAL NAME
A1003			659			100	2	TUNGSTEN OXIDES
A1010			659			50	1	VANADIUM OXIDES
A1026			683			34	2	CERAMICS
A1137	000051741		655			25	1	HISTAMINE PHOSPHATE
A1141	000463796		12					CARBONIC ACID
A1187	000053598		655			20	1	TRIPHOSPHOPYRIDINE NUCLEOTIDE
A1211			659			17	1	TITANIUM, OXIDES OF
A1238	008049476		655			14	1	PANCREATIN
A1249			647			12		* * * INVALID-HAZ-CODE * * *
A1297	001319466		12					LEAD CARBONATE, BASIC
A1363	025265752		707			12	1	BUTYLENE GLYCOL
A1433	001345046		683			10	2	ANTIMONY SULFIDE
A1434			683			9	2	RESINOID
A1435			815			10	2	POUR DEPRESSANTS
A1437			647			7		FATTY ACID GLYCERIDES
A1438			647			7		ALKYL PHENOL POLYETHYLENE GLYCOL ETHER
A1440	001759586		647			6		DIMETHYLCYCLOPENTANE, TRANS-1,3-
A1442	009005349		12					AMMONIUM ALGINATE
A1445	000051785		12					AMINOPHENOL HYDROCHLORIDE, PARA-
A1446			12					HYDROGENATED NAPHTHAS
A1448			12					COALESCING AGENTS
A1455			168			2	1	CHLORINATED PHENOLS
A1456			168			2	1	VEGETABLE OIL VARNISH
A1458			707			6	1	AMINE SOAP
A1459			707			6	1	POTASH SOAP
A1460			647			5		BARIUM ALKYL PHENOLATES
A1461			647			5		CALCIUM ALKYL PHENOLATES
A1462			647			5		ORGANIC BORATES
A1463			647			4		TRIARYL PHOSPHATE
A1464	003687227		2,672			16	2	DINITRO(1-METHYLHEPTYL)PHENOL
A1590	000112561		2,672			13	2	BUTOXY-BETA'-THIOCYANODIETHYL ETHER, BETA-
A1591			2,672			12	2	* * * INVALID-HAZ-CODE * * *
A1592	000556229		2,672			19	2	HEPTADECYLIMIDAZOLINE ACETATE, 2-
A1620			647			2		ALKYL ARYL SODIUM SULFONATE
A1624			647			2		ALKYLPHENYL POLYETHOXYETHANOL
A1628			815			3	2	ALKYL STYRENE POLYMERS
A1637	009080175		2,588			8	1	AMMONIUM POLYSULFIDE
A1638	012794955		12					AMMONIUM SILICATE
A1644	008029296		659			2	1	BANDANE
A1650	000741582		659			2	1	BETASAN
A1656			12					BLOOD MEAL
A1660			12					BUTYRONIC ACID
A1662			12					CALCIUM AMMONIUM NITRATE SOLUTION
A1694			659			2	1	* * * INVALID-HAZ-CODE * * *
A1695	000099309		2,588			7	1	DICHLORO-4-NITROANILINE, 2,6-
A1699	000709988		659			2	1	DICHLOROPROPIONANILIDE
A1705	001861321		659			2	1	TEREPHTHALIC ACID, TETRACHLORO-, DIMETHYL ESTER
A1707	002873598		12					DIPHENYL DIDODECYL SILANE
A1713	002514536		659			2	1	ETHYLENEGLYCOL BIS(TRICHLOROACETATE)
A1714	000145733		659			2	1	ENDOTHAL
A1716	000136254		659			2	1	ERBON
A1719	009002908		12					ETHYLENE OXIDE CONDENSATE
A1732			12					FLUOROLUBE OILS

QUATERLY HAZARD SUMMARY REPORT

DATE 04/09/79

HAZ CODE	CAS NUMBER	RTECS NUMBER	ESTIMATED PEOPLE EXPOSED	SOURCE OF EXP. EST. ACT	TRN	GEN	NO OCCS	CHEMICAL NAME
A1749	000080137		12					HALAZONE
A1751			12					HEAVY METAL STEARATES
A1752			671			2	2	HEXACHLOROACETATE
A1755			12					HOOF+HORN MEAL
A1761			12					IODOPHORS
A1776	000330552		659			2	1	LINURON
A1785	012427382		2,588			6	1	MANEB
A1788			12					METALLIC SOAP
A1796	002163806		659			1	1	MONOSODIUM ACID METHANEARSONATE
A1801			12					* * * INVALID-HAZ-CODE * * *
A1822			12					PH INDICATOR
A1825	003478942		2,588			5	1	PIPERALIN
A1828			12					PLASTIC THICKENER
A1840			12					* * * INVALID-HAZ-CODE * * *
A1858	010049088		12					RUTHINIUM CHLORIDE
A1863			12					SKYDROL
A1871	000078488		659			1	1	TRIBUTYL PHOSPHOROTRITHIOATE, S,S,S-
A1923	027896840		12					NITROBENZENEIMIDAZOLE NITRATE, 6-
MMMMM			847		4		3	UNDEFINED CHEMICAL
M0003			36					* * * INVALID-HAZ-CODE * * *
M0063			2,684			5	2	INORGANIC CHROMATES
M0081			683			1	2	* * * INVALID-HAZ-CODE * * *
M0082			36					SODIUM PHOSPHATES
M0097			12					INORGANIC MOLYBDENUM COMPOUNDS
M0099			12					INORGANIC IRON COMPOUNDS
M0104			12					INORGANIC PALLADIUM COMPOUNDS
M0112			12					INORGANIC MAGNESIUM COMPOUNDS
M0125	001317368		12					LEAD MONOXIDE
M0126	001314416		12					LEAD TETROXIDE
M0144			12					INORGANIC BORON COMPOUNDS
M0148			36					INORGANIC ACIDS
M0218			647			1		FATTY ACIDS
M0238	000064175		24				1	ALCOHOL
M0309			12					EPOXIDES
M0327	000133073		2,672			9	2	TRICHLOROMETHYLTHIO)PHTHALIMIDE, N-(-
M0347	000117817		12					ETHYLHEXYL) PHTHALATE, BIS(2-
M0362	014484641		2,660			4	1	FERRIC DIMETHYLDITHIOCARBAMATE
M0377	000062737		2,672			4	2	PHOSPHORIC ACID, 2,2-DICHLOROVINYL DIMETHYL ESTER
M0388			815		1	1	2	ALKANES
M0389			192				1	ALKENES
M0391			192				1	CYCLIC HYDROCARBONS
M0397			851			1	3	* * * INVALID-HAZ-CODE * * *
M0419	000076062		2,672			7	2	TRICHLORONITROMETHANE
M0420	000074839		2,684			4	2	BROMOMETHANE
M0475			2,672			3	2	MERCURY-CONTAINING ORGANOMETALLIC COMPOUNDS
M0497			12					AMMONIUM COMPOUNDS, ORGANIC
M0521	009003296		815			2	2	POLYBUTENE
M0530			12					RUBBERS, SYNTHETIC
M0533			655			1	1	TISSUES AND TISSUE EXTRACTS
M0534			655			1	1	PROTEINS
M0539			707			1	1	DYES AND PIGMENTS
M0542			60					OILS, PETROLEUM-DERIVED
M0598			168				1	ALKYD RESINS

HAZ CODE	CAS NUMBER	RTECS NUMBER	ESTIMATED PEOPLE EXPOSED	SOURCE OF EXP. EST. ACT	TRN	GEN	NO OCCS	CHEMICAL NAME
M0600	008002742		12					PARAFFIN
M0603	008020835		3,519		1	5	7	MINERAL OIL
M0625	000112389		2,588			3	1	UNDECYLENIC ACID
M0626	001314234		683			1	2	ZIRCONIUM OXIDE
M0627			12					GUMS
M0628	008030306		647			1		NAPHTHA
M0658			647			1		ALKYLARYL SULFONATE
M0671			12					TERPENES
M0684	001313979		12					NEODYMIUM OXIDE
M0689	001345057		168				1	C.I. 77115-PIGMENT WHITE 5
M0720			655			1	1	HUMAN SERUM
M0751			12					LEADED ZINC OXIDE
M0779	008040311		12					COCONUT DIETHANOLAMIDE
M0806	000082688		2,588			3	1	PENTACHLORONITROBENZENE
M0829			815			1	2	CALCIUM PETROLEUM SULFONATE
M0841			12					MARBLE
M0894			12					BOILED LINSEED OIL
M1000	007732185		703		3		3	WATER
M1002			659			1	1	SILICONE RESIN
M1030	008001794		707			1	1	CASTOR OIL
M1055	007631905		12,957		11		4	SODIUM BISULFITE
M1107	000121255		12					AMPROLIUM
M1142	000076153		156					CHLOROPENTAFLUOROETHANE
M1198	000142789		647			1		LAUROYL ETHANOLAMIDE
M1266			655			1	1	RABBIT SERUM
M1311			12					* * * INVALID-HAZ-CODE * * *
M1322	009005907		12					TURPENTINE (GUM)
M1340	009035589		655			1	1	THROMBOPLASTIN
M1463			683			1	2	IRON OXIDE, RED
M1596			647			1		SURFACTANTS
M1727	011099119		12					VANADIUM OXIDE
M1737			815			1	2	ZINC DIALKYLDITHIOPHOSPHATE
M1807			12					* * * INVALID-HAZ-CODE * * *
M1830			647			1		SODIUM ALKYL BENZENE SULFONATES
M1882			12					TANKAGE
M1924			36					SEAL OIL
M1966			655			1	1	PHOSPHOLIPIDS
M2370			12					SODIUM LINEAR ALKYLBENZENE SULFONATE
M2417			168				1	STYRENE-BUTADIENE LATEX
M2495			374					SULFATED NONOXYNOL
M2527			2,684			2	2	CALCIUM SOAP
M2569			168				1	ACRYLIC EMULSION RESIN
M2606			374					COCONUT OIL SOAP
M2824			647		1			C6-C9 AROMATIC HYDROCARBONS
M2829			3,547			3	8	PETROLEUM SPIRITS
M2865	000592018		2,672			2	2	CALCIUM CYANIDE
M2927			647					BUTYL- LAURYL- STEARYL- AND DIETHYL AMINOETHYL METHACRYLATES
M2929			815			1	2	BARIUM PETROLEUM SULFONATE
M2963			707			1	1	CASTOR OIL SOAP
M2985	002425061		2,588			2	1	DIFOLATAN
M3241	000144218		659				1	DISODIUM METHYLARSONATE
M3257			12					AMMONIATED SUPERPHOSPHATE
M4139	011112100		12					SODIUM ANTIMONATE

HAZ CODE	CAS NUMBER	RTECS NUMBER	ESTIMATED PEOPLE EXPOSED	SOURCE OF EXP. EST. ACT	TRN	GEN	NO OCCS	CHEMICAL NAME
M4242			192				1	PETROLEUM OIL
M4468	005793840		815			1	2	CALCIUM PHENATE
P0410			671				2	INFRARED RADIATION
P0430			659				1	ULTRAVIOLET RADIATION
P0450	039405306		281					XR
P0610			1,008	1			1	CONTINUOUS NOISE
P0615	008077201		1,135	1			8	PN
P0650			1,505	1				VB
01568	000064197		679				3	ACETIC ACID
02820	000067641		374					ACETONE
03298	000074862		12					ACETYLENE
03800	000107131		2,672			2	2	ACRYLONITRILE
03895			12					ACTIVATED SLUDGE
04280	008039632		374					ALKYLDIMETHYLBENZYLAMMONIUM CHLORIDE
04370	000107186		659				1	ALLYL ALCOHOL
04603			12					* * * INVALID-HAZ-CODE * * *
04605	007446700		24				1	ALUMINUM CHLORIDE
04620	010043013		655				1	ALUMINUM SULFATE
05250	007664417		659				1	AMMONIA
05270	012125029		12					AMMONIUM CHLORIDE
06145	001336216		36				2	AMMONIUM HYDROXIDE
06163	006484522		708	1			1	AMMONIUM NITRATE
06175	007727540		12					AMMONIUM PERSULFATE
06500	008007703		374					ANISE OIL
07310	007440360		695				3	ANTIMONY
07545	007440382		671				2	ARSENIC
07555			168				1	ARSENIC OXIDES
07570	001327533		659				1	ARSENIC TRIOXIDE
08640	000513779		12					BARIUM CARBONATE
09070	000071432		281					BENZENE
09318	000058899		2,672			2	2	LINDANE
10210	000065850		12					BENZOIC ACID
11855			659				1	BERYLLIUM OXIDES
12068			655	1			1	BIOLOGICAL SUBSTANCES
12845	000080057		707				1	BISPHENOL A
12940			3,447			2	8	* * * INVALID-HAZ-CODE * * *
12960	010043353		24				1	BORIC ACID
13025			12					BRASS, OXIDES OF
13455	061789966		12					RUBBER, BUTADIENE-STYRENE
13850	000071363		851				3	BUTANOL
13980	000078933		659				1	BUTANONE, 2-
14380	000123864		12					BUTYL ACETATE
15570	007440439		12					CADMIUM
15630			659			1	1	CADMIUM OXIDES
15730	000156627		671				1	CALCIUM CYANAMIDE
15743	001305620		12					CALCIUM HYDROXIDE
15746	007778543		12					CALCIUM HYPOCHLORITE
15755	001305788		12					CALCIUM OXIDE
17366	007782425		2,684			2	2	GRAPHITE
17367	000124389		815				2	CARBON DIOXIDE
17460	000630080		659				1	CARBON MONOXIDE
17475	012597692		683	1			2	STEEL
17490	000056235		3,319			2	3	CARBON TETRACHLORIDE

HAZ CODE	CAS NUMBER	RTECS NUMBER	ESTIMATED PEOPLE EXPOSED	SOURCE OF EXP. EST. ACT	TRN	GEN	NO OCCS	CHEMICAL NAME
17525	000409212		683				2	SILICON CARBIDE
17685			2,409			1	5	* * * INVALID-HAZ-CODE * * *
18070	000079118		12					CHLOROACETIC ACID
19425			671			1	2	CHROMIUM OXIDES
19680	000077929		60				2	CITRIC ACID
19769			12					COATING
19770	007440484		12					COBALT
20115	007440508		12					COPPER
20170			671				2	COPPER OXIDES
20265	001344281		695				3	ALUMINUM OXIDE
20810	015096523		12					CRYOLITE
21190	000110827		647					CYCLOHEXANE
21560	000108930		192				1	CYCLOHEXANOL
23305			683			1	2	* * * INVALID-HAZ-CUDE * * *
23360	000333415		2,672			1	2	PHOSPHOROTHIOIC ACID, O,O-DIETHYL O-(2-ISOPROPYL-6-METHYL-4-PYRIMIDINYL
23660	000106934		2,672			3	2	DIBROMOETHANE, 1,2-
23880			12					* * * INVALID-HAZ-CODE * * *
24003	000095501		12					DICHLOROBENZENE, ORTHO-
24006	000106467		12					DICHLOROBENZENE, PARA-
24095	000075718		204				2	DICHLORODIFLUOROMETHANE
24130	000107062		3,319			2	3	DICHLOROETHANE, 1,2-
24235	000075434		655				1	DICHLOROFLUOROMETHANE
24270	000094757		731				2	DICHLOROPHENOXYACETIC ACID, 2,4-
24425	001320372		168				1	DICHLOROTETRAFLUOROETHANE
24615	000111422		683				1	DIETHANOLAMINE
25544	000111466		707				1	DIETHYLENE GLYCOL
28855			12					ENAMEL
28880	000115297		2,672			1	2	ENDOSULFAN
29930	000111762		647					ETHANOL, 2-BUTOXY-
31470	000141786		2,253			1	4	ETHYL ACETATE
31500	000064175		3,193			2	7	ETHYL ALCOHOL
32281			659				1	* * * INVALID-HAZ-CODE * * *
32385	000107211		815			1	2	ETHYLENE GLYCOL
32590	000060297		2,241			1	3	ETHYL ETHER
32940	000078104		12					ETHYL SILICATE
33115			12					FELDSPAR
33160	007705080		12					FERRIC CHLORIDE
33565	000075694		847		1		3	TRICHLOROFLUOROMETHANE
33595	014542235		12					FLUORSPAR
33640	000050000		24				1	FORMALDEHYDE
33720	000064186		12					FORMIC ACID
35085	000056815		48				1	GLYCEROL
35507			683	1			2	GRINDING WHEEL DUST
35960	000076448		2,684			1	3	HEPTACHLOR
36060	000142825		647					HEPTANE
36645			1,374				4	HYDROQUINONES
36710	000070304		2,962			1	2	HEXACHLOROPHENE
36955	000110543		647					HEXANE
37510	000108101		12					METHYL-2-PENTANONE, 4-
37630	000107415		707				1	HEXYLENE GLYCOL
38580	007647010		36				2	HYDROGEN CHLORIDE
38605	007722841		12					HYDROGEN PEROXIDE
40030	007553562		12					IODINE

HAZ CODE	CAS NUMBER	RTECS NUMBER	ESTIMATED PEOPLE EXPOSED	SOURCE OF EXP. EST. ACT	TRN	GEN	NO OCCS	CHEMICAL NAME
40297			659				1	IRON OXIDES
40430	000078831		12					ISOBUTYL ALCOHOL
40910	000078591		647					ISOPHORONE
40984	000108214		12					ISOPROPYL ACETATE
40987	000067630		1,326			1	4	ISOPROPYL ALCOHOL
41775	001332587		12,937		5		2	KAOLIN
42355			12					LACQUER THINNER
42410			374					* * * INVALID-HAZ-CODE * * *
42490	007439921		683				2	LEAD
42685			659				1	LEAD OXIDES
43360	007786303		12					MAGNESIUM CHLORIDE
43390			659				1	MAGNESIUM OXIDES
44000	007439965		12					MANGANESE
44025			12					INORGANIC MANGANESE COMPOUNDS
44030	001313139		12					MANGANESE DIOXIDE
44035			659				1	MANGANESE OXIDES
44915			2,588			1	1	INORGANIC MERCURY COMPOUNDS
45315			671				2	MERCURY OXIDES
45930	000067561		317				1	METHANOL
46210	000072435		2,672			1	2	METHOXYCHLOR
46970	000071556		1,302				2	TRICHLOROETHANE, 1,1,1-
47030	000108872		647					METHYLCYCLOHEXANE
47210			2,588			1	1	* * * INVALID-HAZ-CODE * * *
47270	000075092		293				1	DICHLOROMETHANE
47800	000063252		72					NAPHTHYL N-METHYLCARBAMATE, 1-
48628			659				1	MOLYBDENUM OXIDES
50420	007440020		12					NICKEL
50495			659			1	1	NICKEL OXIDES
50742	007697372		12					NITRIC ACID
50865	007727379		168				1	NITROGEN
50875			659				1	NITROGEN OXIDES
51118	010024972		156					NITROUS OXIDE
51705	000111659		647					OCTANE
52138			803	1			1	OIL, LUBE
52141			815	1			2	OIL, MOTOR
52142			72					OIL, OTHER
52143			839	1			2	OIL, PENETRATING
52145			12					OILS, VEGETABLE
52190	000112801		647					OLEIC ACID
52480	000144627		48				1	OXALIC ACID
53900	004685147		659				1	PARAQUAT
54160	000087865		2,672			2	2	PENTACHLOROPHENOL
54790	000127184		647					TETRACHLOROETHYLENE
57740	007803512		659				1	PHOSPHINE
58520	007664382		1,403				4	PHOSPHORIC ACID
59162	001314803		647					PHOSPHORUS PENTASULFIDE
59173			12					PHOTOGRAPHIC PLATE CLEANER
59465			2,409			1	5	PIGMENTS
60125	007440064		12					PLATINUM
60360	007447407		12					POTASSIUM CHLORIDE
60370	007789006		2,588			1	1	POTASSIUM CHROMATE
60420	014075537		683				2	POTASSIUM TETRAFLUOROBORATE
60440	001310583		1,314				3	POTASSIUM HYDROXIDE

HAZ CODE	CAS NUMBER	RTECS NUMBER	ESTIMATED PEOPLE EXPOSED	SOURCE OF EXP. EST. ACT	TRN	GEN	NO OCCS	CHEMICAL NAME
60490	007722647		12					POTASSIUM PERMANGANATE
60540	007778005		695				3	POTASSIUM SULFATE
60570	000333200		12					POTASSIUM THIOCYANATE
60713			803				1	PC-GL
62460	000079094		12					PROPIONIC ACID
63525	000057556		719				2	PROPYLENE GLYCOL
65070	008003347		2,672			1	2	PYRETHRUM
65080	000110861		281					PYRIDINE
66495	014808607		12					QUARTZ
67405	007440166		12					RHODIUM
67530	000083794		2,672			1	2	ROTENONE
67537			12					RUBBERS, NATURAL
67539			12					RUBBER, OTHER
67680	000069727		12					SALICYLIC ACID
68295	007782492		12					SELENIUM
68511			374	1				SHAMPOO
68512	009000593		683				2	SHELLAC
68657	001344098		707				3	SILICIC ACID, DISODIUM SALT
68766			875			1	5	SOAP
68850	000497198		719				4	SODIUM CARBONATE
68870	007775099		659				1	SODIUM CHLORATE
68880	007647145		731				4	SODIUM CHLORIDE
68950	000143339		12					SODIUM CYANIDE
69000	025155300		12					SODIUM DODECYLBENZENESULFONATE
69055	010124568		12					SODIUM HEXAMETAPHOSPHATE
69070	001310732		12					SODIUM HYDROXIDE
69090	007681529		695				2	SODIUM HYPOCHLORITE
69220	007631994		683				3	SODIUM NITRATE
69460	001313822		12					SODIUM SULFIDE
69470	007757837		12					SODIUM SULFITE
69738	009005258		12					STARCH
69855	008052413		192				1	STODDARD SOLVENT
70845	007704349		374					SULFUR
70870	007664939		679				3	SULFURIC ACID
70995			228				4	SURFACTANT
71055	014807966		2,696			1	3	TALC
71640	000078002		647					TETRAETHYL LEAD
71860	000075741		647					TETRAMETHYL LEAD
71900	000137268		2,672			2	2	TETRAMETHYLTHIURAM DISULFIDE
72085			659				1	THALLIUM OXIDES
73075	007440315		12					TIN
73255	007440326		12					TITANIUM
73258			851				3	* * * INVALID-HAZ-CODE * * *
73300	000108883		659				1	TOLUENE
73515			647					TRANSMISSION FLUID
73730	000126738		647					TRIBUTYL PHOSPHATE
73750	033086189		2,672			1	2	DICHLORODIPHENYLTRICHLOROETHANE
73790	000079016		647					ETHYLENE, TRICHLORO-
73860			2,588			1	1	* * * INVALID-HAZ-CODE * * *
73900	000093765		731				2	TRICHLOROPHENOXYACETIC ACID, 2,4,5-
74795	007601549		695				2	SODIUM PHOSPHATE, TRI
74990	008006642		12					TURPENTINE
75158			12					UNCH

HAZ CODE	CAS NUMBER	RTECS NUMBER	ESTIMATED PEOPLE EXPOSED	SOURCE OF EXP. EST. ACT	TRN	GEN	NO OCCS	CHEMICAL NAME
76612			228				4	WATERLESS HAND CLEANER
76614	008063089		647					WAX
76616			659				1	CARBON ARC LAMP GASES
76617			1,399				10	CGAS
76618			659				1	WELDING RODS
76720	001330207		851				3	XYLENE
77115	007440666		12					ZINC
77150	007646857		2,696			1	3	ZINC CHLORIDE
77155			12					* * * INVALID-HAZ-CODE * * *
77190	001314132		168				1	ZINC OXIDE
77195			659				1	ZINC OXIDES
77220	007733020		2,684			1	2	ZINC SULFATE
77265	007440677		12					ZIRCONIUM
80016			655	1			1	AIR FRESHENER
80018			2,241	1			3	CORRECTION FLUID
80019			12					DRAIN OPENER
80022			659				1	FILLER COMPOUND
80027			216				3	AROMATIC HYDROCARBONS
80032			13,795	7			2	SOIL
80037	000637127		2,684			1	2	ALUMINUM STEARATE
80038			12					INORGANIC AMMONIUM COMPOUNDS
80039			374					AJAX
80046	000121755		2,672	1		1	2	DIETHYL MERCAPTOSUCCINATE, O,O- DIMETHYLDITHIOPHOSPHATE OF
80050			659	1			1	HERBICIDE
80051	008006540		228				4	LANOLIN
80053			168				1	OIL, LINSEED
80056	001317335		2,684			1	2	MOLYBDENUM DISULFIDE
80059			36					PETROLEUM SULFONATE
80060			12					INORGANIC PHOSPHATES
80061	008002093		952				3	PINE OIL
80064	007778509		12					POTASSIUM DICHROMATE(VI)
80065			720				2	POTASSIUM OXIDES
80073	007775191		12					SODIUM METABORATE
80076	007758294		659				1	SODIUM TRIPOLYPHOSPHATE
80079	006834920		647					SODIUM METASILICATE
80090			204				2	ALIPHATIC HYDROCARBONS
80094			12					AROMATIC SOLVENT
80096			386				1	OILS, ESSENTIAL
80108			281					* * * INVALID-HAZ-CODE * * *
80109	008009038		374					PETROLATUM
80117			667				2	PHOTOGRAPHIC FIXER
80124			659				1	STEEL, OXIDES OF
80142	007761888		12					SILVER NITRATE
80144	000546930		12					MAGNESIUM CARBONATE
80153	007772987		12					SODIUM THIOSULFATE
80158	007783202		708				1	AMMONIUM SULFATE
80164	008050097		48				1	ROSIN
80165	016389881		12,290		3		1	DOLOMITE
80172			12					PRODUCTS OF ROSIN CORE SOLDER
80177			12					AMINES
80181			655				1	PEPTONE
80182	007601903		12					PERCHLORIC ACID
80183			12					PHENOLIC RESINS

HAZ CODE	CAS NUMBER	RTECS NUMBER	ESTIMATED PEOPLE EXPOSED	SOURCE OF EXP. EST. ACT	TRN	GEN	NO OCCS	CHEMICAL NAME
81876	007758976		160				1	CHROMIC ACID, LEAD(2+) SALT (1:1)
81878			2,684				2	LITHIUM GREASE
81914	000055550		667				2	METHYL-PARA-AMINOPHENOL SULFATE, PARA-
81941			655				1	POLYPEPTONE
81951	000583528		12					POTASSIUM OXALATE
81953	007727211		24				1	POTASSIUM PERSULFATE
81975	001302427		12					SODIUM ALUMINATE
81993	007632044		647					SODIUM PERBORATE
81994	001313606		36					SODIUM PEROXIDE
82006			12					SULFONIC ACID
82009	061789977		2,684				2	TALLOW
82054			2,588				1	* * * INVALID-HAZ-CODE * * *
82105			12					POLYETHOXY POLYPROPOXY POLYETHOXY-IODINE COMPLEX
82156			2,672				2	PYRETHRINS
82172	035471499		659				1	POTASSIUM 2-BENZYL-4-CHLOROPHENATE
82214	001314983		683				2	ZINC SULFIDE
82227			12					GOLD CHLORIDES
82232	010141001		12					CHROMIC POTASSIUM SULFATE
82237	007782992		12					SULFUROUS ACID
82253	013845368		12					POTASSIUM TRIPOLYPHOSPHATE
82272	000062566		12					THIOUREA
82789	001300727		647					SODIUM XYLENESULFONATE
82797			815				2	POLYMETHACRYLATE RESIN
82806	007379284		12					ETHYLENEDIAMINETETRAACETIC ACID, SODIUM SALT
82817	001314201		12					THORIUM OXIDE
82880	001330785		12					PHOSPHORIC ACID, TRITOLYL ESTER
82935	000287923		647					CYCLOPENTANE
82943	000151677		281					HALOTHANE
82953	000096377		647					METHYLCYCLOPENTANE
82963	001312761		12					POTASSIUM SILICATE
82984			647					* * * INVALID-HAZ-CODE * * *
82996	000096140		647					METHYLPENTANE, 3-
83093			655				1	CEPHALIN
83157	000657841		647					SODIUM PARA-TOLUENESULFONATE
83323	008046740		2,684				2	POTASSIUM SOAP
83325			12					ORGANIC ACIDS
83329	008046717		2,684				2	SODIUM SOAP
83334	000120821		12					TRICHLOROBENZENE, 1,2,4-
83451	007778532		374					POTASSIUM PHOSPHATE, TRIBASIC
83475	001338438		36					SORBITAN MONOOLEATE
83520	000591764		647					METHYLHEXANE, 2-
83530	000589344		647					METHYLHEXANE, 3-
83600	000101202		647					TRICLOCARBAN
83643	000139333		647					ETHYLENEDIAMINETETRAACETIC ACID, DISODIUM SALT
83706			12					ALKYL DICHLOROBENZYL DIMETHYLAMMONIUM CHLORIDE
83786	001918021		731				2	AMINO-3,5,6-TRICHLOROPICOLINIC ACID, 4-
83787	012122677		72					ZINC ETHYLENEBIS(DITHIOCARBAMATE)
83788	012001853		2,588				1	ZINC NAPHTHENATE
83793	000115322		2,672				2	BIS(CHLOROPHENYL)-2,2,2-TRICHLORO ETHANOL, 1,1-
83903			12					SUPERPHOSPHATES
84055			293				1	SILICA, AMORPHOUS FUSED
84174	002678413		815				2	BARIUM PHENATE
84182			12					PHENOLIC COMPOUNDS

HAZ CODE	CAS NUMBER	RTECS NUMBER	ESTIMATED PEOPLE EXPOSED	SOURCE OF EXP. EST. ACT	TRN	GEN	NO OCCS	CHEMICAL NAME
80189	025322683		1,081				2	POLYETHYLENE GLYCOL ETHER
80197	007758023		12					POTASSIUM BROMIDE
80218	000051036		2,588			1	1	PIPERONYL BUTOXIDE
80223	000064028		386				1	ETHYLENEDIAMINETETRAACETIC ACID TETRASODIUM SALT
80224	007722885		647					SODIUM PYROPHOSPHATE, TETRA
80231	000057136		12,314		6		3	UREA
80238	025167800		12					CHLOROPHENOL
80244	007778189		647					CALCIUM SULFATE
80247	010124364		2,588			1	1	CADMIUM SULFATE
80252	000598630		12					LEAD CARBONATE
80259			12					AMMONIUM PHOSPHATES
80262	015007611		48				1	ALUMINUM POTASSIUM SULFATE
80270	000087694		36					TARTARIC ACID
80298	001309484		168				1	MAGNESIUM OXIDE
80299	001343880		12					MAGNESIUM SILICATE
80325			12					COTTONSEED
80327			12					FISH MEAL
80329			655				1	HORMONES
80349	009000117		647					CELLULOSE, CARBOXYMETHYL ETHER
80371			12					BONE MEAL
80378	014882189		2,588			1	1	BISMUTH SUBSALICYLATE
80413	001185575		12					FERRIC AMMONIUM CITRATE
80425			655				1	BILE
80434			655				1	PEPTONIZED IRON
80467			655				1	NUCLEIC ACIDS
80488	000866842		12					POTASSIUM CITRATE
80517	001333739		12					SODIUM BORATE
80549	000124414		12					SODIUM METHOXIDE
80595	000050215		12					LACTIC ACID
80602	007320345		647					DIPHOSPHORIC ACID, TETRAPOTASSIUM SALT
80611	007727437		647					BARIUM SULFATE
80675			928				1	PERFUME
80800			281					ANESTHETIC
80830	000314409		659				1	BROMACIL
80860			647					SODIUM PHOSPHATE, TRI, CHLORINATED
80900	009006659		659				1	DIMETHICONE
81070	024980583		168				1	POLYVINYL ACETATE ACRYLIC RESIN
81080	013746662		12					POTASSIUM FERRICYANIDE
81280			386				1	* * * INVALID-HAZ-CODE * * *
81435	009036195		12					POLY(OXY-1,2-ETHANEDIYL), ALPHA-((1,1,3,3-TETRAMETHYLBUTYL)PHENYL)-OMEG
81651	010028156		659				1	OZONE
81663	007757826		671				2	SODIUM SULFATE
81670			12,302	16			2	FERTILIZER
81679	001317391		168				1	CUPROUS OXIDE
81683	007681110		12					POTASSIUM IODIDE
81711			12					DRIER
81713	000123773		374					AZOBISFORMAMIDE
81726	000635654		655				1	BILIRUBIN
81753	010124375		12					CALCIUM NITRATE
81757	012049502		168				1	CALCIUM TITANATE
81769	000056951		655				1	CHLOROHEXIDINE DIACETATE
81830	010138042		12					FERRIC AMMONIUM SULFATE
81855	000110270		655				1	ISOPROPYL MYRISTATE

HAZ CODE	CAS NUMBER	RTECS NUMBER	ESTIMATED PEOPLE EXPOSED	SOURCE OF EXP. EST. ACT	TRN	GEN	NO OCCS	CHEMICAL NAME
84274	007487947		12					MERCURIC CHLORIDE
84328			659				1	COBALT OXIDES
84377			12					* * * INVALID-HAZ-CODE * * *
84428	000557346		2,684				2	ZINC ACETATE
84440	012135761		36					AMMONIUM SULFIDE
84448	007647101		12					PALLADIUM CHLORIDE
84458	000107926		12					BUTYRIC ACID
84465	007784465		659				1	SODIUM ARSENITE
84470	000142596		2,588				1	NABAM
84473	010099748		12					LEAD NITRATE
84478	010108642		2,588				1	CADMIUM CHLORIDE
84534	000060515		72					DIMETHOATE
84542	000563122		72					ETHION
84546	007446142		168				1	LEAD SULFATE
84562			647					POLYALKYLENE GLYCOLS
84581	000093721		659				1	TRICHLOROPHENOXY)PROPIONIC ACID, 2-(2,4,5-
84620	000569642		2,588				1	C.I. 42000-BASIC GREEN 4
84758	001610180		659				1	METHOXY-4,6-BIS(ISOPROPYLAMINO)-SYM- TRIAZINE, 2-
84759			647					* * * INVALID-HAZ-CODE * * *
84783	007803498		12					HYDROXYLAMINE
84786	001317379		683				2	FERROUS SULFIDE
84803	000759944		659				1	ETHYL DI-N,N-PROPYLTHIOCARBAMATE, S-
90310	001332214		12					ASBESTOS
90320	008052424		168				1	ASPHALT
90330			2,684				2	INORGANIC BARIUM COMPOUNDS
90340	001302789		12					BENTONITE
90500			647					CHLORINATED HYDROCARBONS
90590			12					CLAYS
90610	008007452		12					COAL TAR
90630			12					INORGANIC COBALT COMPOUNDS
90640			2,600				2	INORGANIC COPPER COMPOUNDS
90690			647					CYCLOALKANES
90750			659				1	ESTERS
90760			12					ETHERS
90830			659				1	FLUORINE COMPOUNDS
90860			12					FREONS
90880			647					GASOLINE-LEADED
90930			683				2	GLUE-RUBBER BASE
91095	007439896		12					IRON
91115			952				3	OIL, FUEL NO. 1
91120			659				1	KETONES
91160			2,684				2	INORGANIC LEAD COMPOUNDS
92120			707				1	BRAKE FLUID
92135			12					CATALYST
92160			293				1	CC
92220			647					DEGREASER
92255			815				2	DETERGENT
92260			667	1			2	DISINFECTANT
92280			36					DUPLICATOR FLUID
92290			2,696			1	3	DYES
92320			168				1	EMULSIFIER
92450			2,588	1			1	FUNGICIDE
92470			2,684	1			2	GREASE

QUATERLY HAZARD SUMMARY REPORT

DATE 04/09/79

HAZ CODE	CAS NUMBER	RTECS NUMBER	ESTIMATED PEOPLE EXPOSED	SOURCE OF EXP. EST. ACT	TRN	GEN	NO OCCS	CHEMICAL NAME
92500			168				1	OIL, HYDRAULIC
92525			36					INK REMOVER
92545			3,327	2			4	INSECTICIDE
92590			655				1	MEDICINES
92630			168				1	PAINT
92650			12					PAINT THINNER
92670			12					PHOTOGRAPHIC CHEMICAL
92675			667				2	PHOTOGRAPHIC DEVELOPER
92685			2,241				3	PLASTICIZER
92750			12					PRIMER
92760			36					INK, PRINTING
92770			168				1	PROPELLANT-AEROSOL
92845			12					ACCELERATOR, RUBBER
92855			12					RUST INHIBITOR
92910			647					SOLVENT
92980			12					WETTING AGENT
93060			12					ENZYMES
94220			1,152				1	WOODS

"Red Book" Summary Report
in Hazard Description Order

Preceding page blank

HAZ CODE	CAS NUMBER	RTECS NUMBER	ESTIMATED PEOPLE EXPOSED	SOURCE OF EXP. EST. ACT	TRN	GEN	NO OCCS	CHEMICAL NAME
17685			2,409			1	5	* * * INVALID-HAZ-CODE * * *
A1249			647			12		* * * INVALID-HAZ-CODE * * *
A1694			659			2	1	* * * INVALID-HAZ-CODE * * *
A1591			2,672			12	2	* * * INVALID-HAZ-CODE * * *
12940			3,447			2	8	* * * INVALID-HAZ-CODE * * *
04603			12					* * * INVALID-HAZ-CODE * * *
M1807			12					* * * INVALID-HAZ-CODE * * *
M1311			12					* * * INVALID-HAZ-CODE * * *
M0397			851			1	3	* * * INVALID-HAZ-CODE * * *
23305			683			1	2	* * * INVALID-HAZ-CODE * * *
M0081			683			1	2	* * * INVALID-HAZ-CODE * * *
M0003			36					* * * INVALID-HAZ-CODE * * *
23880			12					* * * INVALID-HAZ-CODE * * *
A1840			12					* * * INVALID-HAZ-CODE * * *
A1801			12					* * * INVALID-HAZ-CODE * * *
81280			386				1	* * * INVALID-HAZ-CODE * * *
82054			2,588				1	* * * INVALID-HAZ-CODE * * *
77155			12					* * * INVALID-HAZ-CODE * * *
32281			659				1	* * * INVALID-HAZ-CODE * * *
42410			374					* * * INVALID-HAZ-CODE * * *
84759			647					* * * INVALID-HAZ-CODE * * *
73860			2,588			1	1	* * * INVALID-HAZ-CODE * * *
73258			851				3	* * * INVALID-HAZ-CODE * * *
82984			647					* * * INVALID-HAZ-CODE * * *
84377			12					* * * INVALID-HAZ-CODE * * *
80108			281					* * * INVALID-HAZ-CODE * * *
47210			2,588			1	1	* * * INVALID-HAZ-CODE * * *
92845			12					ACCELERATOR, RUBBER
01568	000064197		679				3	ACETIC ACID
02820	000067641		374					ACETONE
03298	000074862		12					ACETYLENE
M2569			168				1	ACRYLIC EMULSION RESIN
03800	000107131		2,672			2	2	ACRYLONITRILE
03895			12					ACTIVATED SLUDGE
80016			655	1			1	AIR FRESHENER
80039			374					AJAX
M0238	000064175		24				1	ALCOHOL
80090			204				2	ALIPHATIC HYDROCARBONS
M0388			815		1	1	2	ALKANES
M0389			192				1	ALKENES
M0598			168				1	ALKYD RESINS
A1620			647			2		ALKYL ARYL SODIUM SULFONATE
83706			12					ALKYL DICHLOROBENZYL DIMETHYLAMMONIUM CHLORIDE
A1438			647			7		ALKYL PHENOL POLYETHYLENE GLYCOL ETHER
A1628			815			3	2	ALKYL STYRENE POLYMERS
M0658			647			1		ALKYLARYL SULFONATE
04280	008039632		374					ALKYLDIMETHYLBENZYLAMMONIUM CHLORIDE
A1624			647			2		ALKYLPHENYL POLYETHOXYETHANOL
04370	000107186		659				1	ALLYL ALCOHOL
04605	007446700		24				1	ALUMINUM CHLORIDE
20265	001344281		695				3	ALUMINUM OXIDE
80262	015007611		48				1	ALUMINUM POTASSIUM SULFATE
80037	000637127		2,684			1	2	ALUMINUM STEARATE

QUATERLY HAZARD SUMMARY REPORT

DATE 04/09/79

HAZ CODE	CAS NUMBER	RTECS NUMBER	ESTIMATED PEOPLE EXPOSED	SOURCE OF EXP. EST. ACT	TRN	GEN	NO OCCS	CHEMICAL NAME
04620	010043013		655				1	ALUMINUM SULFATE
A1458			707			6	1	AMINE SOAP
80177			12					AMINES
83786	001918021		731				2	AMINO-3,5,6-TRICHLOROPICOLINIC ACID, 4-
A1445	000051785		12					AMINOPHENOL HYDROCHLORIDE, PARA-
05250	007664417		659				1	AMMONIA
M3257			12					AMMONIATED SUPERPHOSPHATE
A1442	009005349		12					AMMONIUM ALGINATE
05270	012125029		12					AMMONIUM CHLORIDE
M0497			12					AMMONIUM COMPOUNDS, ORGANIC
06145	001336216		36				2	AMMONIUM HYDROXIDE
06163	006484522		708	1			1	AMMONIUM NITRATE
06175	007727540		12					AMMONIUM PERSULFATE
80259			12					AMMONIUM PHOSPHATES
A1637	009080175		2,588			8	1	AMMONIUM POLYSULFIDE
A1638	012794955		12					AMMONIUM SILICATE
80158	007783202		708				1	AMMONIUM SULFATE
84440	012135761		36					AMMONIUM SULFIDE
M1107	000121255		12					AMPROLIUM
80800			281					ANESTHETIC
06500	008007703		374					ANISE OIL
07310	007440360		695				3	ANTIMONY
A1433	001345046		683			10	2	ANTIMONY SULFIDE
80027			216				3	AROMATIC HYDROCARBONS
80094			12					AROMATIC SOLVENT
07545	007440382		671				2	ARSENIC
07555			168				1	ARSENIC OXIDES
07570	001327533		659				1	ARSENIC TRIOXIDE
90310	001332214		12					ASBESTOS
90320	008052424		168				1	ASPHALT
81713	000123773		374					AZODISFORMAMIDE
A1644	008029296		659			2	1	BANDANE
A1460			647			5		BARIUM ALKYL PHENOLATES
08640	000513779		12					BARIUM CARBONATE
M2929			815			1	2	BARIUM PETROLEUM SULFONATE
84174	002678413		815				2	BARIUM PHENATE
80611	007727437		647					BARIUM SULFATE
90340	001302789		12					BENTONITE
09070	000071432		281					BENZENE
10210	000065850		12					BENZOIC ACID
11855			659				1	BERYLLIUM OXIDES
A1650	000741582		659			2	1	BETASAN
80425			655				1	BILE
81726	000635654		655				1	BILIRUBIN
12068			655	1			1	BIOLOGICAL SUBSTANCES
83793	000115322		2,672				2	BIS(CHLOROPHENYL)-2,2,2-TRICHLORO ETHANOL, 1,1-
80378	014882189		2,588			1	1	BISMUTH SUBSALICYLATE
12845	000080057		707				1	BISPHENOL A
A1656			12					BLOOD MEAL
M0894			12					BOILED LINSEED OIL
80371			12					BONE MEAL
12960	010043353		24				1	BORIC ACID
92120			707				1	BRAKE FLUID

HAZ CODE	CAS NUMBER	RTECS NUMBER	ESTIMATED PEOPLE EXPOSED	SOURCE OF EXP. EST. ACT	TRN	GEN	NO OCCS	CHEMICAL NAME
13025			12					BRASS, OXIDES OF
80830	000314409		659				1	BROMACIL
M0420	000074839		2,684			4	2	BROMOMETHANE
13850	000071363		851				3	BUTANOL
13980	000078933		659				1	BUTANONE, 2-
A1590	000112561		2,672			13	2	BUTOXY-BETA'-THIOCYANODIETHYL ETHER, BETA-
14380	000123864		12					BUTYL ACETATE
M2927			647					BUTYL- LAURYL- STEARYL- AND DIETHYL AMINOETHYL METHACRYLATES
A1363	025265752		707			12	1	BUTYLENE GLYCOL
84458	000107926		12					BUTYRIC ACID
A1660			12					BUTYRONIC ACID
84620	000569642		2,588				1	C.I. 42000-BASIC GREEN 4
M0689	001345057		168				1	C.I. 77115-PIGMENT WHITE 5
15570	007440439		12					CADMIUM
84478	010108642		2,588				1	CADMIUM CHLORIDE
15630			659			1	1	CADMIUM OXIDES
80247	010124364		2,588			1	1	CADMIUM SULFATE
A1461			647			5		CALCIUM ALKYL PHENOLATES
A1662			12					CALCIUM AMMONIUM NITRATE SOLUTION
15730	000156627		671				1	CALCIUM CYANAMIDE
M2865	000592018		2,672			2	2	CALCIUM CYANIDE
15743	001305620		12					CALCIUM HYDROXIDE
15746	007778543		12					CALCIUM HYPOCHLORITE
81753	010124375		12					CALCIUM NITRATE
15755	001305788		12					CALCIUM OXIDE
M0829			815			1	2	CALCIUM PETROLEUM SULFONATE
M4468	005793840		815			1	2	CALCIUM PHENATE
M2527			2,684			2	2	CALCIUM SOAP
80244	007778189		647					CALCIUM SULFATE
81757	012049502		168				1	CALCIUM TITANATE
76616			659				1	CARBON ARC LAMP GASES
17367	000124389		815				2	CARBON DIOXIDE
17460	000630080		659				1	CARBON MONOXIDE
17490	000056235		3,319			2	3	CARBON TETRACHLORIDE
A1141	000463796		12					CARBONIC ACID
M1030	008001794		707			1	1	CASTOR OIL
M2963			707			1	1	CASTOR OIL SOAP
92135			12					CATALYST
92160			293				1	CC
80349	009000117		647					CELLULOSE, CARBOXYMETHYL ETHER
83093			655				1	CEPHALIN
A1026			683			34	2	CERAMICS
76617			1,399				10	CGAS
90500			647					CHLORINATED HYDROCARBONS
A1455			168			2	1	CHLORINATED PHENOLS
18070	000079118		12					CHLOROACETIC ACID
81769	000056951		655				1	CHLOROHEXIDINE DIACETATE
M1142	000076153		156					CHLOROPENTAFLUOROETHANE
80238	025167800		12					CHLOROPHENOL
81876	007758976		168				1	CHROMIC ACID, LEAD(2+) SALT (1:1)
82232	010141001		12					CHROMIC POTASSIUM SULFATE
19425			671			1	2	CHROMIUM OXIDES
19680	000077929		60				2	CITRIC ACID

HAZ CODE	CAS NUMBER	RTECS NUMBER	ESTIMATED PEOPLE EXPOSED	SOURCE OF EXP. EST. ACT	TRN	GEN	NO OCCS	CHEMICAL NAME
90590			12					CLAYS
90610	008007452		12					COAL TAR
A1448			12					COALESCING AGENTS
19769			12					COATING
19770	007440484		12					COBALT
84328			659				1	COBALT OXIDES
M0779	008040311		12					COCONUT DIETHANOLAMIDE
M2606			374					COCONUT OIL SOAP
P0610			1,008	1			1	CONTINUOUS NOISE
20115	007440508		12					COPPER
20170			671				2	COPPER OXIDES
80018			2,241	1			3	CORRECTION FLUID
80325			12					COTTONSEED
20810	015096523		12					CRYOLITE
81679	001317391		168				1	CUPROUS OXIDE
M0391			192				1	CYCLIC HYDROCARBONS
90690			647					CYCLOALKANES
21190	000110827		647					CYCLOHEXANE
21560	000108930		192				1	CYCLOHEXANOL
82935	000287923		647					CYCLOPENTANE
M2824			647		1			C6-C9 AROMATIC HYDROCARBONS
92220			647					DEGREASER
92255			815				2	DETERGENT
23660	000106934		2,672			3	2	DIBROMOETHANE, 1,2-
A1695	000099309		2,588			7	1	DICHLORO-4-NITROANILINE, 2,6-
24003	000095501		12					DICHLOROBENZENE, ORTHO-
24006	000106467		12					DICHLOROBENZENE, PARA-
24095	000075718		204				2	DICHLORODIFLUOROMETHANE
73750	033086189		2,672			1	2	DICHLORODIPHENYLTRICHLOROETHANE
24130	000107062		3,319			2	3	DICHLOROETHANE, 1,2-
24235	000075434		655				1	DICHLOROFLUOROMETHANE
47270	000075092		293				1	DICHLOROMETHANE
24270	000094757		731				2	DICHLOROPHENOXYACETIC ACID, 2,4-
A1699	000709988		659			2	1	DICHLOROPROPIONANILIDE
24425	001320372		168				1	DICHLOROTETRAFLUOROETHANE
24615	000111422		683				1	DIETHANOLAMINE
80046	000121755		2,672	1		1	2	DIETHYL MERCAPTOSUCCINATE, O,O- DIMETHYLDITHIOPHOSPHATE OF
25544	000111466		707				1	DIETHYLENE GLYCOL
M2985	002425061		2,588			2	1	DIFOLATAN
80900	009006659		659				1	DIMETHICONE
84534	000060515		72					DIMETHOATE
A1440	001759586		647			6		DIMETHYLCYCLOPENTANE, TRANS-1,3-
A1464	003687227		2,672			16	2	DINITRO(1-METHYLHEPTYL)PHENOL
A1707	002873598		12					DIPHENYL DIDODECYL SILANE
80602	007320345		647					DIPHOSPHORIC ACID, TETRAPOTASSIUM SALT
92260			667	1			2	DISINFECTANT
M3241	000144218		659				1	DISODIUM METHYLARSONATE
80165	016389881		12,290		3		1	DOLOMITE
80019			12					DRAIN OPENER
81711			12					DRIER
92280			36					DUPLICATOR FLUID
92290			2,696			1	3	DYES
M0539			707			1	1	DYES AND PIGMENTS

HAZ CODE	CAS NUMBER	RTECS NUMBER	ESTIMATED PEOPLE EXPOSED	SOURCE OF EXP, EST, ACT	TRN	GEN	NO OCCS	CHEMICAL NAME
92320			168				1	EMULSIFIER
28855			12					ENAMEL
28880	000115297		2,672			1	2	ENDOSULFAN
A1714	000145733		659			2	1	ENDOTHAL
93060			12					ENZYMES
M0309			12					EPOXIDES
A1716	000136254		659			2	1	ERBON
90750			659				1	ESTERS
29930	000111762		647					ETHANOL, 2-BUTOXY-
90760			12					ETHERS
84542	000563122		72					ETHION
31470	000141786		2,253			1	4	ETHYL ACETATE
31500	000064175		3,193			2	7	ETHYL ALCOHOL
84803	000759944		659				1	ETHYL DI-N,N-PROPYLTHIOCARBAMATE, S-
32590	000060297		2,241			1	3	ETHYL ETHER
32940	000078104		12					ETHYL SILICATE
32385	000107211		815			1	2	ETHYLENE GLYCOL
A1719	009002908		12					ETHYLENE OXIDE CONDENSATE
73790	000079016		647					ETHYLENE, TRICHLORO-
80223	000064028		386				1	ETHYLENEDIAMINETETRAACETIC ACID TETRASODIUM SALT
83643	000139333		647					ETHYLENEDIAMINETETRAACETIC ACID, DISODIUM SALT
82806	007379284		12					ETHYLENEDIAMINETETRAACETIC ACID, SODIUM SALT
A1713	002514536		659			2	1	ETHYLENEGLYCOL BIS(TRICHLOROACETATE)
M0347	000117817		12					ETHYLHEXYL) PHTHALATE, BIS(2-
A1437			647			7		FATTY ACID GLYCERIDES
M0218			647			1		FATTY ACIDS
33115			12					FELDSPAR
80413	001185575		12					FERRIC AMMONIUM CITRATE
81030	010138042		12					FERRIC AMMONIUM SULFATE
33160	007705080		12					FERRIC CHLORIDE
M0362	014484641		2,660			4	1	FERRIC DIMETHYLDITHIOCARBAMATE
84786	001317379		683				2	FERROUS SULFIDE
81670			12,302	16			2	FERTILIZER
80022			659				1	FILLER COMPOUND
80327			12					FISH MEAL
90830			659				1	FLUORINE COMPOUNDS
A1732			12					FLUOROLUBE OILS
33595	014542235		12					FLUORSPAR
33640	000050000		24				1	FORMALDEHYDE
33720	000064186		12					FORMIC ACID
90860			12					FREONS
92450			2,588	1			1	FUNGICIDE
90880			647					GASOLINE-LEADED
90930			683				2	GLUE-RUBBER BASE
35085	000056815		48				1	GLYCEROL
82227			12					GOLD CHLORIDES
17366	007782425		2,684			2	2	GRAPHITE
92470			2,684	1			2	GREASE
35507			683	1			2	GRINDING WHEEL DUST
M0627			12					GUMS
A1749	000080137		12					HALAZONE
82943	000151677		281					HALOTHANE
A1751			12					HEAVY METAL STEARATES

HAZ CODE	CAS NUMBER	RTECS NUMBER	ESTIMATED PEOPLE EXPOSED	SOURCE OF EXP. EST. ACT	TRN	GEN	NO OCCS	CHEMICAL NAME
35960	000076448		2,684			1	3	HEPTACHLOR
A1592	000556229		2,672			19	2	HEPTADECYLIMIDAZOLINE ACETATE, 2-
36060	000142825		647					HEPTANE
80050			659	1			1	HERBICIDE
A1752			671			2	2	HEXACHLOROACETATE
36710	000070304		2,962			1	2	HEXACHLOROPHENE
36955	000110543		647					HEXANE
37630	000107415		707				1	HEXYLENE GLYCOL
A1137	000051741		655			25	1	HISTAMINE PHOSPHATE
A1755			12					HOOF+HORN MEAL
80329			655				1	HORMONES
M0720			655			1	1	HUMAN SERUM
38580	007647010		36				2	HYDROGEN CHLORIDE
38605	007722841		12					HYDROGEN PEROXIDE
A1446			12					HYDROGENATED NAPHTHAS
36645			1,374				4	HYDROQUINONES
84783	007803498		12					HYDROXYLAMINE
P0410			671				2	INFRARED RADIATION
92525			36					INK REMOVER
92760			36					INK, PRINTING
M0148			36					INORGANIC ACIDS
80038			12					INORGANIC AMMONIUM COMPOUNDS
90330			2,684				2	INORGANIC BARIUM COMPOUNDS
M0144			12					INORGANIC BORON COMPOUNDS
M0063			2,684			5	2	INORGANIC CHROMATES
90630			12					INORGANIC COBALT COMPOUNDS
90640			2,600				2	INORGANIC COPPER COMPOUNDS
M0099			12					INORGANIC IRON COMPOUNDS
91160			2,684				2	INORGANIC LEAD COMPOUNDS
M0112			12					INORGANIC MAGNESIUM COMPOUNDS
44025			12					INORGANIC MANGANESE COMPOUNDS
44915			2,588			1	1	INORGANIC MERCURY COMPOUNDS
M0097			12					INORGANIC MOLYBDENUM COMPOUNDS
M0104			12					INORGANIC PALLADIUM COMPOUNDS
80060			12					INORGANIC PHOSPHATES
92545			3,327	2			4	INSECTICIDE
40030	007553562		12					IODINE
A1761			12					IODOPHORS
91095	007439896		12					IRON
M1463			683			1	2	IRON OXIDE, RED
40297			659				1	IRON OXIDES
40430	000078831		12					ISOBUTYL ALCOHOL
40910	000078591		647					ISOPHORONE
40984	000108214		12					ISOPROPYL ACETATE
40987	000067630		1,326			1	4	ISOPROPYL ALCOHOL
81855	000110270		655				1	ISOPROPYL MYRISTATE
41775	001332587		12,937		5		2	KAOLIN
91120			659				1	KETONES
42355			12					LACQUER THINNER
80595	000050215		12					LACTIC ACID
80051	008006540		228				4	LANOLIN
M1198	000142789		647			1		LAUROYL ETHANOLAMIDE
42490	007439921		683				2	LEAD

HAZ CODE	CAS NUMBER	RTECS NUMBER	ESTIMATED PEOPLE EXPOSED	SOURCE OF EXP. EST. ACT	TRN	GEN	NO OCCS	CHEMICAL NAME
80252	000598630		12					LEAD CARBONATE
A1297	001319466		12					LEAD CARBONATE, BASIC
M0125	001317368		12					LEAD MONOXIDE
84473	010099748		12					LEAD NITRATE
42685			659				1	LEAD OXIDES
84546	007446142		168				1	LEAD SULFATE
M0126	001314416		12					LEAD TETROXIDE
M0751			12					LEADED ZINC OXIDE
09318	000058899		2,672			2	2	LINDANE
A1776	000330552		659			2	1	LINURON
81878			2,684				2	LITHIUM GREASE
80144	000546930		12					MAGNESIUM CARBONATE
43360	007786303		12					MAGNESIUM CHLORIDE
80298	001309484		168				1	MAGNESIUM OXIDE
43390			659				1	MAGNESIUM OXIDES
80299	001343880		12					MAGNESIUM SILICATE
A1785	012427382		2,588			6	1	MANEB
44000	007439965		12					MANGANESE
44030	001313139		12					MANGANESE DIOXIDE
44035			659				1	MANGANESE OXIDES
M0841			12					MARBLE
92590			655				1	MEDICINES
84274	007487947		12					MERCURIC CHLORIDE
45315			671				2	MERCURY OXIDES
M0475			2,672			3	2	MERCURY-CONTAINING ORGANOMETALLIC COMPOUNDS
A1788			12					METALLIC SOAP
45930	000067561		317				1	METHANOL
84758	001610180		659				1	METHOXY-4,6-BIS(ISOPROPYLAMINO)-SYM- TRIAZINE, 2-
46210	000072435		2,672			1	2	METHOXYCHLOR
81914	000055550		667				2	METHYL-PARA-AMINOPHENOL SULFATE, PARA-
37510	000108101		12					METHYL-2-PENTANONE, 4-
47030	000108872		647					METHYLCYCLOHEXANE
82953	000096377		647					METHYLCYCLOPENTANE
83520	000591764		647					METHYLHEXANE, 2-
83530	000589344		647					METHYLHEXANE, 3-
82996	000096140		647					METHYLPENTANE, 3-
M0603	008020835		3,519		1	5	7	MINERAL OIL
80056	001317335		2,684			1	2	MOLYBDENUM DISULFIDE
48628			659				1	MOLYBDENUM OXIDES
A1796	002163806		659			1	1	MONOSODIUM ACID METHANEARSONATE
84470	000142596		2,588				1	NABAM
M0628	008030306		647			1		NAPHTHA
47800	000063252		72					NAPHTHYL N-METHYLCARBAMATE, 1-
M0684	001313979		12					NEODYMIUM OXIDE
50420	007440020		12					NICKEL
50495			659			1	1	NICKEL OXIDES
50742	007697372		12					NITRIC ACID
A1923	027896840		12					NITROBENZENEIMIDAZOLE NITRATE, 6-
50865	007727379		168				1	NITROGEN
50875			659				1	NITROGEN OXIDES
51118	010024972		156					NITROUS OXIDE
80467			655				1	NUCLEIC ACIDS
51705	000111659		647					OCTANE

HAZ CODE	CAS NUMBER	RTECS NUMBER	ESTIMATED PEOPLE EXPOSED	SOURCE OF EXP. EST. ACT	TRN	GEN	NO OCCS	CHEMICAL NAME
91115			952				3	OIL, FUEL NO. 1
92500			168				1	OIL, HYDRAULIC
80053			168				1	OIL, LINSEED
52138			803	1			1	OIL, LUBE
52141			815	1			2	OIL, MOTOR
52142			72					OIL, OTHER
52143			839	1			2	OIL, PENETRATING
80096			386				1	OILS, ESSENTIAL
M0542			60					OILS, PETROLEUM-DERIVED
52145			12					OILS, VEGETABLE
52190	000112801		647					OLEIC ACID
83325			12					ORGANIC ACIDS
A1462			647			5		ORGANIC BORATES
52480	000144627		48				1	OXALIC ACID
81651	010028156		659				1	OZONE
92630			168				1	PAINT
92650			12					PAINT THINNER
84448	007647101		12					PALLADIUM CHLORIDE
A1238	008049476		655			14	1	PANCREATIN
M0600	008002742		12					PARAFFIN
53900	004685147		659				1	PARAQUAT
60713			803				1	PC-GL
M0806	000082688		2,588			3	1	PENTACHLORONITROBENZENE
54160	000087865		2,672			2	2	PENTACHLOROPHENOL
80181			655				1	PEPTONE
80434			655				1	PEPTONIZED IRON
80182	007601903		12					PERCHLORIC ACID
80675			928				1	PERFUME
80109	008009038		374					PETROLATUM
M4242			192				1	PETROLEUM OIL
M2829			3,547			3	8	PETROLEUM SPIRITS
80059			36					PETROLEUM SULFONATE
A1822			12					PH INDICATOR
84182			12					PHENOLIC COMPOUNDS
80183			12					PHENOLIC RESINS
57740	007803512		659				1	PHOSPHINE
M1966			655			1	1	PHOSPHOLIPIDS
58520	007664382		1,403				4	PHOSPHORIC ACID
82880	001330785		12					PHOSPHORIC ACID, TRITOLYL ESTER
M0377	000062737		2,672			4	2	PHOSPHORIC ACID, 2,2-DICHLOROVINYL DIMETHYL ESTER
23360	000333415		2,672			1	2	PHOSPHOROTHIOIC ACID, O,O-DIETHYL O-(2-ISOPROPYL-6-METHYL-4-PYRIMIDINYL
59162	001314803		647					PHOSPHORUS PENTASULFIDE
92670			12					PHOTOGRAPHIC CHEMICAL
92675			667				2	PHOTOGRAPHIC DEVELOPER
80117			667				2	PHOTOGRAPHIC FIXER
59173			12					PHOTOGRAPHIC PLATE CLEANER
59465			2,409			1	5	PIGMENTS
80061	008002093		952				3	PINE OIL
A1825	003478942		2,588			5	1	PIPERALIN
80218	000051036		2,588			1	1	PIPERONYL BUTOXIDE
A1828			12					PLASTIC THICKENER
92685			2,241				3	PLASTICIZER
60125	007440064		12					PLATINUM

HAZ CODE	CAS NUMBER	RTECS NUMBER	ESTIMATED PEOPLE EXPOSED	SOURCE OF EXP. EST. ACT	TRN	GEN	NO OCCS	CHEMICAL NAME
P0615	008077201		1,135	1			8	PN
81435	009036195		12					POLY(OXY-1,2-ETHANEDIYL), ALPHA-((1,1,3,3-TETRAMETHYLBUTYL)PHENYL)-OMEG
84562			647					POLYALKYLENE GLYCOLS
M0521	009003296		815			2	2	POLYBUTENE
82105			12					POLYETHOXY POLYPROPOXY POLYETHOXY-IODINE COMPLEX
80189	025322683		1,081				2	POLYETHYLENE GLYCOL ETHER
82797			815				2	POLYMETHACRYLATE RESIN
81941			655				1	POLYPEPTONE
81070	024980583		168				1	POLYVINYL ACETATE ACRYLIC RESIN
A1459			707			6	1	POTASH SOAP
80197	007758023		12					POTASSIUM BROMIDE
60360	007447407		12					POTASSIUM CHLORIDE
60370	007789006		2,588			1	1	POTASSIUM CHROMATE
80488	000866842		12					POTASSIUM CITRATE
80064	007778509		12					POTASSIUM DICHROMATE(VI)
81080	013746662		12					POTASSIUM FERRICYANIDE
60440	001310583		1,314				3	POTASSIUM HYDROXIDE
81683	007681110		12					POTASSIUM IODIDE
81951	000583528		12					POTASSIUM OXALATE
80065			720				2	POTASSIUM OXIDES
60490	007722647		12					POTASSIUM PERMANGANATE
81953	007727211		24				1	POTASSIUM PERSULFATE
83451	007778532		374					POTASSIUM PHOSPHATE, TRIBASIC
82963	001312761		12					POTASSIUM SILICATE
83323	008046740		2,684				2	POTASSIUM SOAP
60540	007778805		695				3	POTASSIUM SULFATE
60420	014075537		683				2	POTASSIUM TETRAFLUOROBORATE
60570	000333200		12					POTASSIUM THIOCYANATE
82253	013845368		12					POTASSIUM TRIPOLYPHOSPHATE
82172	035471499		659				1	POTASSIUM 2-BENZYL-4-CHLOROPHENATE
A1435			815			10	2	POUR DEPRESSANTS
92750			12					PRIMER
80172			12					PRODUCTS OF ROSIN CORE SOLDER
92770			168				1	PROPELLANT-AEROSOL
62460	000079094		12					PROPIONIC ACID
63525	000057556		719				2	PROPYLENE GLYCOL
M0534			655			1	1	PROTEINS
82156			2,672				2	PYRETHRINS
65070	008003347		2,672			1	2	PYRETHRUM
65080	000110861		281					PYRIDINE
66495	014808607		12					QUARTZ
M1266			655			1	1	RABBIT SERUM
A1434			683			9	2	RESINOID
67405	007440166		12					RHODIUM
80164	008050097		48				1	ROSIN
67530	000083794		2,672			1	2	ROTENONE
13455	061789966		12					RUBBER, BUTADIENE-STYRENE
67539			12					RUBBER, OTHER
67537			12					RUBBERS, NATURAL
M0530			12					RUBBERS, SYNTHETIC
92855			12					RUST INHIBITOR
A1858	010049088		12					RUTHINIUM CHLORIDE
67680	000069727		12					SALICYLIC ACID

HAZ CODE	CAS NUMBER	RTECS NUMBER	ESTIMATED PEOPLE EXPOSED	SOURCE OF EXP. ACT	EST. TRN	GEN	NO OCCS	CHEMICAL NAME
M1924			36					SEAL OIL
68295	007782492		12					SELENIUM
68511			374	1				SHAMPOO
68512	009000593		683				2	SHELLAC
84055			293				1	SILICA, AMORPHOUS FUSED
68657	001344098		707				3	SILICIC ACID, DISODIUM SALT
17525	000409212		683				2	SILICON CARBIDE
M1002			659			1	1	SILICONE RESIN
80142	007761888		12					SILVER NITRATE
A1863			12					SKYDROL
68766			875			1	5	SOAP
M1830			647			1		SODIUM ALKYL BENZENE SULFONATES
81975	001302427		12					SODIUM ALUMINATE
M4139	011112100		12					SODIUM ANTIMONATE
84465	007784465		659				1	SODIUM ARSENITE
M1055	007631905		12,957		11		4	SODIUM BISULFITE
80517	001333739		12					SODIUM BORATE
68850	000497198		719				4	SODIUM CARBONATE
68870	007775099		659				1	SODIUM CHLORATE
68880	007647145		731				4	SODIUM CHLORIDE
68950	000143339		12					SODIUM CYANIDE
69000	025155300		12					SODIUM DODECYLBENZENESULFONATE
69055	010124568		12					SODIUM HEXAMETAPHOSPHATE
69070	001310732		12					SODIUM HYDROXIDE
69090	007681529		695				2	SODIUM HYPOCHLORITE
M2370			12					SODIUM LINEAR ALKYLBENZENE SULFONATE
80073	007775191		12					SODIUM METABORATE
80079	006834920		647					SODIUM METASILICATE
80549	000124414		12					SODIUM METHOXIDE
69220	007631994		683				3	SODIUM NITRATE
83157	000657841		647					SODIUM PARA-TOLUENESULFONATE
81993	007632044		647					SODIUM PERBORATE
81994	001313606		36					SODIUM PEROXIDE
74795	007601549		695				2	SODIUM PHOSPHATE, TRI
80860			647					SODIUM PHOSPHATE, TRI, CHLORINATED
M0082			36					SODIUM PHOSPHATES
80224	007722885		647					SODIUM PYROPHOSPHATE, TETRA
83329	008046717		2,604				2	SODIUM SOAP
81663	007757826		671				2	SODIUM SULFATE
69460	001313822		12					SODIUM SULFIDE
69470	007757837		12					SODIUM SULFITE
80153	007772987		12					SODIUM THIOSULFATE
80076	007758294		659				1	SODIUM TRIPOLYPHOSPHATE
82789	001300727		647					SODIUM XYLENESULFONATE
80032			13,795	7			2	SOIL
92910			647					SOLVENT
83475	001338438		36					SORBITAN MONOOLEATE
69738	009005258		12					STARCH
17475	012597692		683	1			2	STEEL
80124			659				1	STEEL, OXIDES OF
69855	008052413		192				1	STODDARD SOLVENT
M2417			168				1	STYRENE-BUTADIENE LATEX
M2495			374					SULFATED NONOXYNOL

HAZ CODE	CAS NUMBER	RTECS NUMBER	ESTIMATED PEOPLE EXPOSED	SOURCE OF EXP, EST. ACT	TRN	GEN	NO OCCS	CHEMICAL NAME
82006			12					SULFONIC ACID
70845	007704349		374					SULFUR
70870	007664939		679				3	SULFURIC ACID
82237	007782992		12					SULFUROUS ACID
83903			12					SUPERPHOSPHATES
70995			228				4	SURFACTANT
M1596			647			1		SURFACTANTS
71055	014807966		2,696			1	3	TALC
82009	061789977		2,684				2	TALLOW
M1882			12					TANKAGE
80270	000087694		36					TARTARIC ACID
A1705	001861321		659			2	1	TEREPHTHALIC ACID, TETRACHLORO-, DIMETHYL ESTER
M0671			12					TERPENES
54790	000127184		647					TETRACHLOROETHYLENE
71640	000078002		647					TETRAETHYL LEAD
71860	000075741		647					TETRAMETHYL LEAD
71900	000137268		2,672			2	2	TETRAMETHYLTHIURAM DISULFIDE
72085			659				1	THALLIUM OXIDES
82272	000062566		12					THIOUREA
82817	001314201		12					THORIUM OXIDE
M1340	009035589		655			1	1	THROMBOPLASTIN
73075	007440315		12					TIN
M0533			655			1	1	TISSUES AND TISSUE EXTRACTS
73255	007440326		12					TITANIUM
A1211			659			17	1	TITANIUM, OXIDES OF
73300	000108883		659				1	TOLUENE
73515			647					TRANSMISSION FLUID
A1463			647			4		TRIARYL PHOSPHATE
73730	000126738		647					TRIBUTYL PHOSPHATE
A1871	000078488		659			1	1	TRIBUTYL PHOSPHOROTRITHIOATE, S,S,S-
83334	000120821		12					TRICHLOROBENZENE, 1,2,4-
46970	000071556		1,302				2	TRICHLOROETHANE, 1,1,1-
33565	000075694		847		1		3	TRICHLOROFLUOROMETHANE
M0327	000133073		2,672			9	2	TRICHLOROMETHYLTHIO)PHTHALIMIDE, N-(-
M0419	000076062		2,672			7	2	TRICHLORONITROMETHANE
84581	000093721		659				1	TRICHLOROPHENOXY)PROPIONIC ACID, 2-(2,4,5-
73900	000093765		731				2	TRICHLOROPHENOXYACETIC ACID, 2,4,5-
83600	000101202		647					TRICLOCARBAN
A1187	000053598		655			20	1	TRIPHOSPHOPYRIDINE NUCLEOTIDE
A1003			659			100	2	TUNGSTEN OXIDES
74990	008006642		12					TURPENTINE
M1322	009005907		12					TURPENTINE (GUM)
P0430			659				1	ULTRAVIOLET RADIATION
75158			12					UNCH
M0625	000112389		2,588			3	1	UNDECYLENIC ACID
MMMMM			847		4		3	UNDEFINED CHEMICAL
80231	000057136		12,314		6		3	UREA
M1727	011099119		12					VANADIUM OXIDE
A1010			659			50	1	VANADIUM OXIDES
P0650			1,505	1				VH
A1456			168			2	1	VEGETABLE OIL VARNISH
M1000	007732185		703		3		3	WATER
76612			228				4	WATERLESS HAND CLEANER

QUATERLY HAZARD SUMMARY REPORT

DATE 04/09/79

HAZ CODE	CAS NUMBER	RTECS NUMBER	ESTIMATED PEOPLE EXPOSED	SOURCE OF EXP, EST. ACT	TRN	GEN	NO OCCS	CHEMICAL NAME
76614	008063089		647					WAX
76618			659				1	WELDING RODS
92980			12					WETTING AGENT
94220			1,152				1	WOODS
P0450	039405306		281					XR
76720	001330207		851				3	XYLENE
77115	007440666		12					ZINC
84428	000557346		2,684				2	ZINC ACETATE
77150	007646857		2,696			1	3	ZINC CHLORIDE
M1737			815			1	2	ZINC DIALKYLDITHIOPHOSPHATE
83787	012122677		72					ZINC ETHYLENEBIS(DITHIOCARBAMATE)
83788	012001853		2,588				1	ZINC NAPHTHENATE
77190	001314132		168				1	ZINC OXIDE
77195			659				1	ZINC OXIDES
77220	007733020		2,684			1	2	ZINC SULFATE
82214	001314983		683				2	ZINC SULFIDE
77265	007440677		12					ZIRCONIUM
M0626	001314234		683			1	2	ZIRCONIUM OXIDE

APPENDIX D:

OLD VS. NEW ALGORITHM TABLES

The tables contained herein present a series of accuracy and efficiency comparisons of the old algorithm and the major phases in redesigning it. Table D-1 compares the original seven options to the old algorithm; these options are variations of the old algorithm, based on simple collapsing of SIC, size, BLS misclassification ratio, and group categories. The accuracy comparisons are based on national universe estimates for employees. Table D-2 examines efficiency and accuracy of those options tied to the first implementation of the new algorithm in the program ALG. The estimates given are total employees exposed to asbestos when the new algorithm is compared to the old algorithm, and, again, total employee estimates for the national universe are recorded for use in the other comparisons. Table D-3 compares the old and initial new algorithms' results for asbestos, this time by displaying estimates for employees exposed to asbestos by SIC in addition to total estimates. Table D-4 compares facility and employee estimates for the national universe from the old and initial new algorithms, again by SIC and for the totals. Table D-5 examines national universe estimates of facilities by major SIC grouping and size level, as well as totals, for both the old algorithm and the second version of the new one. The second version incorporated a correction of the inflated IIP Factors on the facility file for groups (within BLS SIC-size contained in observed SIC-size categories) that contained all nonresponse. Table D-6 shows the manual verification of the new algorithm concept for SIC 21, which showed an error was made in carrying out the factor modification mentioned above. Table D-7 gives the facility estimates by major SIC grouping, along with the grand totals, for the old algorithm and the final version of the new algorithm, which incorporated a correction for the factor modification error. The tables in this appendix all relate to Section IV.D.

TABLE D-1. OLD PROJECTION ALGORITHM OPTIONS AND FACTORS

OPTION	SIZE	SIC	BLS	GROUP	CONTROL BREAKS	TIME CPU-SECONDS PROGRAM PRJ1A	PROGRAM PRJ4A	REPORT GENERATING	TOTAL	TOTAL PROJECTED PLANTS	TOTAL PROJECTED PEOPLE
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Base*	7	65	3	6	8,190	8.17	43.47	1.19	52.83	739,244	38,262,627
Base**	7	65	3	6	8,190	7.56	42.36	1.14	51.06	1,010,194	42,129,627
1	4	65	3	6	4,680	7.84	36.42	1.23	45.49	739,501	38,264,286
2	7	65	3	3	4,095						
3	7	30	3	6	3,780						
4	7	65	1	6	2,730						
5	4	30	1	6	720						
6**	4	30	2	3	20**	7.59	33.74	1.22	42.55	1,002,270	41,889,239
7**	4	30	1	3	360**	7.48	21.31	1.25	30.04	996,889	41,758,963

* Base (Published, with size-1 adjustment)

** Without size-1 adjustment

Notes: 1. Option 1 should be compared with the Base* estimates (with size-1 adjustment), and Options 6 and 7 should be compared with Base** (without size-1 adjustment). Since size adjustment is accomplished during preprocessing in PRJ1, it should not have any effect on PRJ4 run times reported in Column 10.

2. Size 1 adjustment referes to BLS size-1 and is applicable to the size scale 1-7. The size-1 adjustment compensates for inflated IIP factors for size 1 by SIC. The scan of IIP factors was compared to County Business Patterns information from the Bureau of the Census. It was discovered that size 1 did not correspond due to omission of small facilities with less than a minimum number of employees. The computer programs were written to permit switching the size-1 adjustment on or off. The options that compress the size scale from 1-7 to 1-4 were tested with the size-1 adjustment both present and omitted, and illustrated the differing projection results.

TABLE D-2. PROJECTION ALGORITHM OPTIONS AND FACTORS

						TIME CPU-SECONDS					
OPTION (1)	SIZE (2)	SIC (3)	BLS (4)	GROUP (5)	CONTROL BREAKS (6)	PROGRAM PRJ1A (7)	PROGRAM PRJ4A (8)	REPORT GENERATING (9)	TOTAL (10)	TOTAL PROJECTED PLANTS (11)	TOTAL PROJECTED PEOPLE (12)
Base*	7	65	3	6	8,190	8.17	43.47	1.19	52.83	739,244	38,262,627
Base**	7	65	3	6	8,190	7.56	42.36	1.14	51.06	1,010,194	42,129,627
1	4	65	3	6	4,680	7.84	36.42	1.23	45.49	739,501	38,264,286
6**	4	30	2	3	20**	7.59	33.74	1.22	42.55	1,002,270	41,889,239
7**	4	30	1	3	360**	7.48	21.31	1.25	30.04	996,889	41,758,963
8***	7	65	3	6	NA	NA	NA	NA			
Base (Asbestos)									45.00 (est)	--	1,564,551
Alg-Program									5.00 (est)		1,557,136

*Base (Published, with size-1 adjustment)

**Without size-1 adjustment

***Option 8 was a new computer application program implementation approach to the original projection algorithm. Other options represent variations of old algorithm.

Notes: 1. Option 1 should be compared with the Base* estimates (with size-1 adjustment), and Options 6 and 7 should be compared with Base** (without size-1 adjustement). Since size adjustment is accomplished during preprocessing in PRJ1, it should not have any effect on PRJ4 run times reported in Column 10.

2. Size 1 adjustment refers to BLS size-1 and is applicable to the size scale 1-7. The size-1 adjustment compensates for inflated IIP factors for size 1 by SIC. The scan of IIP factors was compared to County Business Patterns information from the Bureau of the Census. It was discovered that size 1 did not correspond due to omission of small facilities with less than a minimum number of employees. The computer programs were written to permit switching the size-1 adjustment on or off. The options that compress the size scale from 1-7 to 1-4 were tested with the size-1 adjustment both present and omitted, and illustrated the differing projection results.

TABLE D-3. VALIDATION OF ESTIMATED PEOPLE*

SIC-CODES	TABLE 49**	ALG
Total (Asbestos)	1,564,551	1,557,136
07,08,09	1,688	1,687
13	2,700	2,699
15,16,17	430,570	427,839
Total Manufacturing	648,872	644,932
19	988	564
20	7,966	7,969
21	76	75
22	8,256	8,075
23	87,929	87,293
24	6,009	5,956
25	15,043	14,922
26	17,781	17,499
27	55,741	55,507
28	23,314	23,231
29	8,202	8,223
30	11,800	11,800
31	2,089	2,032
32	60,299	60,294
33	61,878	60,785
34	36,643	36,644
35	89,423	89,434
36	38,552	37,803
37	86,195	86,163
38	12,656	12,663
39	18,032	18,000
41-49	112,483	112,216
50-59	256,029	255,734
60-65	8,096	8,096
70-89	104,112	103,933

*Initial Version of ALG.

**In Volume III of final report on NIOSH contract No. CDC-99-74-40 [1].

TABLE D-4. NUMBER OF PLANTS AND EMPLOYEES IN PLANTS IN THE NATIONAL UNIVERSE

	PLANTS		EMPLOYEES IN PLANTS	
SIC-CODES	TABLE 1*	ALG	TABLE 1*	ALG
Total (Asbestos)	739,244	728,405	38,262,627	37,879,686
07,08,09	3,645	3,643	82,311	82,312
13	1,313	1,316	82,098	82,099
15,16,17	68,407	68,370	2,532,382	2,517,014
Total Manufacturing	141,397	140,500	15,220,964	15,116,618
19	113	88	33,438	20,833
20	16,173	16,186	1,403,859	1,403,870
21	196	182	80,758	80,402
22	3,225	3,186	232,122	226,374
23	15,565	15,107	918,752	913,246
24	3,091	3,072	161,332	159,873
25	4,618	4,570	295,253	292,085
26	4,859	4,764	570,792	555,203
27	13,609	13,543	1,239,262	1,236,552
28	6,743	6,645	966,199	965,006
29	993	997	195,698	195,700
30	5,640	5,649	534,789	534,797
31	1,653	1,654	155,028	147,796
32	7,955	7,949	700,104	700,109
33	5,040	4,988	1,347,780	1,334,398
34	19,615	19,600	1,350,047	1,350,051
35	14,190	14,212	1,539,734	1,539,747
36	5,073	5,057	1,501,365	1,468,136
37	3,457	3,461	1,224,441	1,224,446
38	3,812	3,824	386,097	386,100
39	5,776	5,776	384,114	381,894
41-49	36,289	35,656	3,311,290	3,296,920
50-59	281,353	279,412	9,283,788	9,229,845
60-65	52,467	50,862	1,946,338	1,909,425
70-89	154,374	148,646	5,803,454	5,655,454

*In Volume III of the final report on NIOSH Contract No. CDC-99-74-40 [1].

TABLE D-5. COMPARISON OF NEW* AND OLD ALGORITHM RESULTS FOR PLANTS

	Small			Medium			Large			Total			
	New	Old	Diff.	New	Old	Diff.	New	Old	Diff.	New	Old	Diff.	Pct
Agricultural Services, Forestry, Fisheries	3,598	3,599	-1	46	46	-	-	-	-	3,644	3,645	-1	0.0
Mining (Oil and Gas)	1,101	1,101	-	176	176	-	36	36	-	1,313	1,313	-	-
Contract Construction	63,332	62,561	+761	5,986	5,753	+233	94	94	-	69,402	68,407	+906	1.4
Manufacturing	114,926	114,433	+493	22,434	22,392	+ 43	4,581	4,571	+10	141,942	141,399	+545	0.4
Transportation and Other Public Utilities	30,792	30,528	+264	5,056	5,056	-	706	706	-	36,553	36,289	+264	0.7
Wholesale and Retail Trade	267,743	266,434	+1,309	13,928	13,675	+253	1,244	1,244	-	282,915	281,353	+1,562	0.6
Finance, Insurance, and Real Estate	48,622	48,622	-	3,615	3,615	-	230	230	-	52,467	52,467	-	-
Services	144,013	143,489	+524	9,147	9,417	-	1,737	1,737	-	154,898	154,374	+524	0.3
Grand Totals	674,117	670,767	+3,350	60,388	59,859	+529	8,628	8,618	+10	743,134	739,244	+3,890	0.5

*Before final correction in adjusting IIP factors.

TABLE D-6. MANUAL VERIFICATION OF NEW ALGORITHM

Fac- No	SIC1	SIC2	SL	Size	BLS SIC	BLS Size	Grp	Payroll*	BLS IIP	Imputed IIP	Size-Adjt IIP	New** IIP	Payroll X New-IIP
060021	4	21	1	1	21	1†	1	*	20.00	40.00	25.20†	*	*
066019	4	21	1	1	21	1†	5	10	21.60	21.60	13.61†	13.61	136.08
066020	4	21	1	1	21	2	5	18	10.80	10.80		10.80	194.40
066022	4	21	1	2	21	1†	2	21	20.00	40.00	25.20†	25.20	529.20
040013	4	21	1	2	21	2	2	*	15.00	30.00		*	*
060023	4	21	1	2	21	2	2	33	12.00	27.00		37.80	1247.40
047020	4	21	1	2	21	2	3	*	10.80	10.80		*	*
043027	4	21	1	3	21	3	1	85	8.00	8.00		8.00	680.00
047021	4	21	1	3	21	3	3	64	5.40	5.40		5.40	345.60
066021	4	21	1	3	21	3	5	50	5.40	5.40		5.40	270.00
066022	4	21	1	3	21	3	5	55	5.40	5.40		5.40	297.00
066023	4	21	1	3	21	3	5	54	5.40	5.40		5.40	291.60
066024	4	21	1	3	21	3	5	93	5.40	5.40		5.40	502.20
								Size Level 1		Estimated Plants		122.41	
										Estimated Employees			4493.48
033014	4	21	2	4	21	4	2	*	12.00	12.00		*	
047022	4	21	2	5	21	4	3	318	10.80	10.80		10.80	3434.40
060024	4	21	2	5	21	5	2	259	6.00	6.00		6.00	1554.00
060025	4	21	2	5	21	5	2	349	6.00	6.00		6.00	2094.00
047023	4	21	2	5	21	5	3	347	5.40	5.40		5.40	1873.80
066025	4	21	2	5	21	5	5	292	5.40	5.40		5.40	1576.80
								Size Level 2		Estimated Plants		33.60	
										Estimated Employees			10533.00
039053	4	21	3	6	21	4	1	710	13.00	13.00		13.00	9230.00
060026	4	21	3	6	21	6	2	739	2.00	2.00		2.00	1478.00
033016	4	21	3	7	21	7	2	3600	9.42	9.42		9.42	33912.00
029012	4	21	3	7	21	7	3	1400	15.08	15.08		15.08	21112.00
								Size Level 3		Estimated Plants		39.50	
										Estimated Employees			65732.00
								Total		Estimated Plants		195.51	
										Estimated Employees			80758.48

†Adjustment applying to BLS size 1 only.

*Zero indicates a nonsurveyed facility.

**List of those IIP factors used in the estimates (all inclusive).

TABLE D-7. COMPARISON OF NEW* AND OLD ALGORITHM RESULTS FOR PLANTS

	New	Old	Diff.	New	Old	Diff.
Agricultural Services, Forestry, Fisheries	3,644	3,645	-1	82,311	82,311	-
Mining (Oil and Gas)	1,313	1,313	-	82,098	82,098	-
Contract Construction	68,408	68,407	+1	2,532,382	2,532,382	-
Manufacturing	141,397	141,397	-	15,220,968	15,220,964	+4
Transportation and Other Public Utilities	36,289	36,289	-	3,311,291	3,311,290	+1
Wholesale and Retail Trade	281,352	281,353	-1	9,283,790	9,283,788	+2
Finance, Insurance, and Real Estate	52,467	52,467	-	1,946,338	1,946,338	-
Services	154,374	154,374	-	5,803,458	5,803,454	+4
Grand Totals	739,244	739,244	-	38,262,636	38,262,627	+9

*Final version

APPENDIX E

NATIONAL PROJECTION ALGORITHM

APPENDIX E explains in detail the theoretical basis and fundamental organization of the old projection algorithm, including the calculation of variance. A basic understanding of the information contained in this appendix is very helpful in studying the effort made in redesigning the old algorithm. The redesigning process is described in Section IV.D.

APPENDIX E

NATIONAL PROJECTION ALGORITHM

A. Sample Design

The sample design used in the NOHS involved a 2-stage selection procedure. Stratification was imposed at both stages. First stage sampling units were Standard Metropolitan Statistical Areas (SMSAs), groups of SMSAs, and some other organized areas. First stage sampling units were selected with probability proportional to measures of size. Second stage sampling units are clusters of facilities. Facilities were selected systematically across first stage units and within second stage strata.

Replication was not undertaken at either the first or the second stage; consequently, unbiased variance estimates are not available. A total sample size of 67 sampling units was selected in the first stage. The systematic selection employed in the second stage produced observations from a total of 4,775 sample facilities.

The initial sample was selected for NIOSH by the Bureau of Labor Statistics (BLS). Under the initial sampling design, the allocation of the second stage sample was proportional to the relative sizes of the second stage strata, as measured by employee numbers. BLS subsequently modified this procedure, under advice from NIOSH, such that the allocation scheme was inversely proportional to these size measures. During the data collection activities, the larger facilities were further subsampled on a nonprobability basis by NIOSH. Nonprobability subsamples of facility locations were sometimes selected when a single facility was found to occur at multiple locations within an SMSA.

As is the case with any sample survey, inferences from the sample data are restricted to the target population as defined by the sampling frame.

B. Nonresponse

Almost all large-scale surveys are subject to nonsampling errors arising from nonresponse. Nonresponse in this sense includes all missing data items for any unit of observation appearing in the sample, regardless of the circumstances. Insofar as the units of observation for which some nonresponse has occured are not, collectively, a probability subsample of the target population and are likely dissimilar from other sampled observational units in important respects, biases are introduced into the estimates.

C. Other Nonsampling Errors

Other nonsampling errors can also contribute to sources of bias in national estimates. In referring to the definition of the target population, it is obvious that unbiased estimates are not available from the sample data for any other population than that population from which the sample was selected.

Preceding page blank

For example, facilities entering into business after 1970 are not included in the target population. Therefore, estimates computed from the sample data have no inferential ability with respect to these new facilities. Similarly, unbiased estimates are not available for 1970 target facilities that moved from sample first stage units before the field data collection activities but were still in business in other areas. The total number of both types of such facilities can be estimated, but their characteristics are unknown. These facilities, including those that went out of business between 1970 and the time of the data collection activities, can be thought of as contributing bias arising from sampling frame inefficiency. Bias adjustment in these cases is achieved by considering these facilities to have zero-valued observation variables for all data items on the questionnaire, but these facilities are still included in the calculation of the estimates.

Another source of bias arises from the inexactness of the target population definition. Bias may also be due to seasonal facilities and other facilities having regular or irregular fluctuations in their activities.

Note also the possibility of biases introduced by the surveyors themselves. Reference here is to such factors as differential acuity among surveyors. No bias adjustment is made for these effects.

D. Estimation Procedures

The notation used in this section is summarized below:

- The subscript $g = 1, 2, \ldots, 6$ indexes the groupings of first stage sampling units imposed as a dimension of stratification on the second stage frames.
- The subscript h indexes the first stage unit within the gth group, $\max \{h|g\} = 31, 13, 7, 4, 6, 6$, for $g = 1, 2, \ldots, 6$.
- The subscript $i = 1, 2, \ldots, 9$ indexes the size category imposed as a dimension of stratification at the second stage.
- The subscript $j = 1, 2, \ldots, 63$ indexes the Standard Industrial Classification also imposed as a dimension of stratification at the second stage. (Note: i and j are crossed dimensions both nested within g.)
- Class 1 facilities are those facilities not affected by the large facility cut; i.e., those facilities for which $i = 1, 2, \ldots, 5$.
- Class 2 facilities are those facilities in size categories $i = 6, 7, 8, 9$ that were surveyed before the cut date.
- Class 3 facilities are those facilities in size categories $i = 6, 7, 8, 9$ that were surveyed after the cut date.
- $\pi^p (g, h)$ represents the first stage inclusion probabilities.

- $\Pi^{s'}$ (g, i, j) represents the second stage inclusion probabilities for Class 1 facilities.

- n' (g, i, j) is the number of Class 1 sample facilities in the second stage stratum cell indexed by the subscripts g,i,j.

- r' (g, i, j) is the number of Class 1 sample facilities (in the second stage stratum cell) that provided a value, including a zero value, for the observation variable under consideration.

- The subscript k' indexes the Class 1 sample facilities within each systematic sample (i.e., within each second stage stratum cell), k' = 1, 2, ..., r (g, i, j) ≤ n (g, i, j).

- The subscript k" indexes the Class 2 sample facilities, k" = 1, 2, ..., r" ≤ n" = 163.

- The subscript k"' indexes the Class 3 sample facilities, k"' = 1, 2, ..., r"' ≤ n"' = 136.

- $\Pi^{s''}$ (k") represents second stage inclusion probabilities for Class 2 facilities.

- $\Pi^{s'''}$ (k"') represents second stage inclusion probabilities for Class 3 facilities.

- Y' (g, h, i, j, k') equals the value of the observation variable obtained from a Class 1 facility that belongs to the target population or equals 0 for sample facilities not in the target population (e.g., out-of-business facilities and moved facilities, which are included in the count r'(g, i, j)).

- Y" (h, k") and Y"' (h, k"') are similarly defined for Class 2 and 3 facilities.

Each Y-value is assumed to be the value of the observation variable for the sample facility. A nonresponse adjustment procedure was developed by RTI for multiple location facilities; the correct level of Y-values is supplied to the calculations. If individual location level questionnaires are employed, the estimators described in the following sections are not appropriate. Note that a Y-value may denote a classification variable or a continuous variable.

In the following paragraphs, estimators for totals (see Figure E-1) are described. Averages and percentages or proportions are computed in the form of ratio estimates. The estimators, as described, form the basis for the development of computer software to handle the actual calculation of the estimates (projection algorithms).

Variance estimators are described in the following paragraphs. Since replication was not a feature of the sample design at either stage, unbiased variance estimators are not available.

Separate estimates are computed for each of the Class 1, 2, and 3 facilities. The separate class estimates are then summed to obtain the national level estimate.

Note that the subscript $h = 1, 2, \ldots, \max\{h|g\}$ identifies the first stage unit from which the observation was collected. The subscript $k = 1, 2, \ldots, n$ (suitably identified with primes to denote facility class) counts the facilities within the systematic sample at the second stage units. However, since the systematic samples were selected across all first stage units, a summation over the range of k includes all of the h-values represented in the systematic sample for the gth group, ith size category, and jth SIC category. The subscript h is, nonetheless, required for identifying the first stage inclusion probability.

In Figure E-1, the range of summation over k^* implies summation over facilities for which item nonresponse has occurred. The adjustment involves substituting the unweighted average value computed for the particular within-class systematic sample in place of each missing item. The substituted averages are then divided by the first and second stage inclusion probabilities associated with the facility for which the observation variable is missing. The adjustment assumes that an entire systematic sample is never missing. If no nonresponse has occurred, the summation over the k^*-values is defined to be zero.

The above estimator can be summarized in words as follows. Compute the average value of the response variable of the systematic sample separately for each class of facilities. For Class 1 facilities, systematic samples are defined for every value of g, i, and j. There is a single such sample for each facility of Class 2 and 3. Compute the quotient resulting from the division of the value of the response variable, Y, and the product of the first and second stage inclusion probabilities associated with the facility. Substitute the average value for the systematic sample in the place of any missing response variable values. Add the quotients over all of the elements in the systematic sample. For Class 1 facilities, add these sums over the second stage dimensions of stratification. Finally, sum together the separate class estimates.

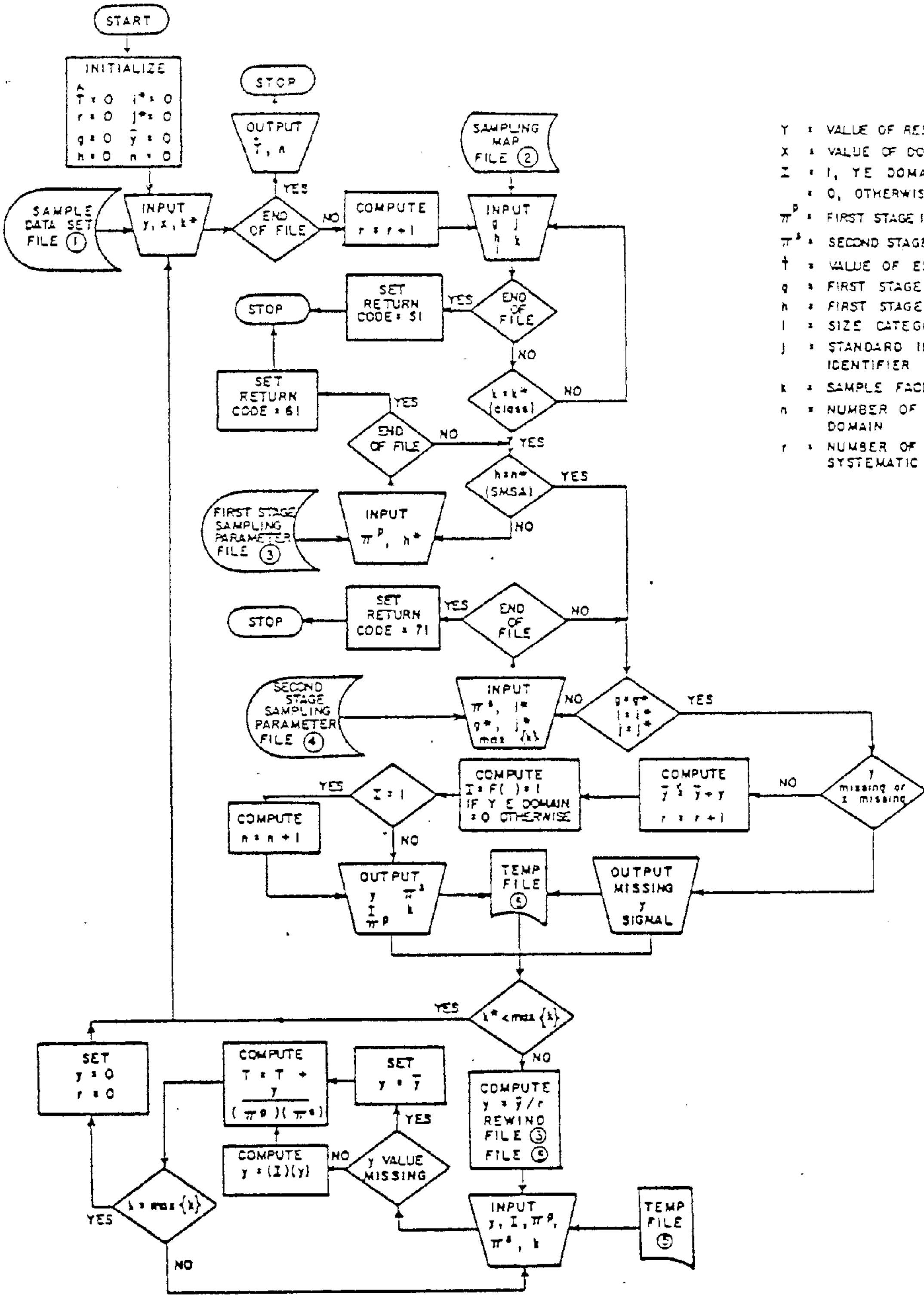
E. Organization of the Computing Algorithm

Figure E-2 provides a summary of the steps required to produce estimates. The figure is based on two assumptions: (1) that there exists a record for every sample facility even if no information was secured for the particular facility, and (2) that there does not exist any systematic sample for which no information was secured, including domain information.

The file labelled "sample data set" is assumed to contain the sample facility identification and the values of the response variables for that facility; as pointed out above, facilities for which no information was collected are assumed to appear, suitably identified, in the sample data set. The fact that this file does not exist as a single entity does not intrinsically influence the computing algorithm.

$$
\begin{aligned}
\hat{T} = & \sum_{g=1}^{6} \sum_{i=1}^{9} \sum_{j=1}^{63} \left\{ \sum_{k=1}^{r'(g,h,i,j)} \frac{Y'(g,h,i,j,k')}{\Pi^{p}(g,h)\ \Pi^{s'}(g,i,j)} \right. \\
& + \left. \sum_{k^{*'}}^{n'(g,i,j)-r'(g,i,j)} \frac{\sum_{k=1}^{r'(g,i,j)} Y'(g,h,i,j,k') \,/\, r'(g,i,j)}{\Pi^{p}(g,h)\ \Pi^{s'}(g,i,j)} \right\} \\
& + \sum_{k''=1}^{r''} \frac{Y''(h,k'')}{\Pi^{p}(g,h)\ \Pi^{s''}(k'')} + \sum_{k^{*''}=1}^{163-r} \frac{\sum_{k''=1}^{r''} Y''(h,k'')/r''}{\Pi^{p}(g,h)\ \Pi^{s''}(k'')} \\
& + \sum_{k'''=1}^{r'''} \frac{Y'''(h,k''')}{\Pi^{p}(g,h)\ \Pi^{s'''}(k''')} + \sum_{k^{*'''}=1}^{136-r'''} \frac{\sum_{k'''=1}^{r'''} Y'''(h,k''')/r'''}{\Pi^{p}(g,h)\ \Pi^{s'''}(k''')} .
\end{aligned}
$$

Figure E-1. Estimators for Totals



Y = VALUE OF RESPONSE VARIABLE

X = VALUE OF DOMAIN VARIABLES

I = 1, Y ∈ DOMAIN

= 0, OTHERWISE

π^p = FIRST STAGE INCLUSION PROBABILITY

π^s = SECOND STAGE INCLUSION PROBABILITY

$\hat{T}$ = VALUE OF ESTIMATE

g = FIRST STAGE GROUP IDENTIFIER

h = FIRST STAGE SAMPLING UNIT IDENTIFIER

i = SIZE CATEGORY IDENTIFIER

j = STANDARD INDUSTRIAL CLASSIFICATION IDENTIFIER

k = SAMPLE FACILITY IDENTIFIER

n = NUMBER OF OBSERVATIONS IN THE DOMAIN

r = NUMBER OF RESPONSES IN THE SYSTEMATIC SAMPLE

Figure E-2. Organization of Computing Algorithm

The file labelled "sampling map" provides the mapping of the sample facility identifiers into the features of the sample design required for the computations. Specifically, these features are the values of g, h, i, and j for the facility and the k-value of the facility class.

The files labeled "sampling parameters" contain the Π-values for the first and second stage units of sampling (for each facility class with the associated identification of the design features), and the maximum values for each of the identifying subscripts.

Each of these four files is assumed to be ordered with respect to the design feature identifiers as indicated by the formulation for totals (see Figure E-1).

F. Variance Estimation Procedure

Subscripting and other notation follow that described earlier in connection with totals estimates. Some simplification of the earlier notation is introduced here.

The estimator for totals, except for the special notation introduced for the precut and postcut large-facility samples and the nonresponse adjustment, is of the form

$$\hat{T}(g, i, j) = \sum_k \frac{Y(g, h, i, j, k)}{\Pi^p(g, h)\ \Pi^s(g, i, j)} .$$

Recall that the subscripts g, i, and j refer to the second stage group by size by SIC stratum cells. The subscript h is used to denote the first stage unit (note that the summation over the range of the subscript k denoting the facilities within the cluster includes all of the h-values represented in the cluster). The values of

$$\Pi^p \text{ and } \Pi^s$$

represent the first and second stage inclusion probabilities. Note that the Π^s values do not depend upon the value of k; however, the Π^p values do depend upon the k-value, since the value of h corresponds to at least one k-value in the range of summation.

Consider the estimate defined by

$$\hat{T} = \sum_g \sum_i \sum_j \left[\sum_k \frac{Y(g, h, i, j, k)}{\Pi^p(g, h)\ \Pi^s(g, i, j)} \right] .$$

This estimate can be thought of as the contribution to the national total provided by the gth group, ith size class and jth SIC category. In these terms, the totals estimator can be rewritten as

$$\hat{T} = \sum_g \sum_i \sum_j \hat{T}(g, i, j) .$$

Let $T(i, j)$ represent the mean estimate of the group estimates within SIC-size (i, j). Then a variance estimator across groups is given by

$$\hat{Var}[\hat{T}] = \sum_i \sum_j \sum_g \frac{(\hat{T}(g, i, j) - \bar{T}(i, j))^2}{g - 1} .$$

Since g = 6 and all group estimates are assumed to be present, including groups with zero estimates because no records occur within the SIC-size-group cell, the above equation becomes

$$\hat{Var}[\hat{T}] = \sum_i \sum_j \sum_{g=1}^{6} \frac{(\hat{T}(g, i, j) - \bar{T}(i, j))^2}{5} .$$

Often the estimate required, at least for Volume III of the final report on RTI's previous data analysis contract (No. CDC-99-74-40) [1], is a ratio estimate; that is,

$$\hat{R} = \frac{\hat{T}_N}{\hat{T}_D} .$$

The variance calculation now involves, in addition to both numerator and denominator estimate variances, a measure of covariance between the two individual estimates comprising the ratio. This covariance is given by

$$\hat{Cov}[\hat{T}_N, \hat{T}_D] = \sum_i \sum_j \left[\sum_{g=1}^{6} \frac{\hat{T}_N(i, j, g) - \bar{T}_N(i, j) \quad \hat{T}_D(i, j, g) - \bar{T}_D(i, j)}{5} \right],$$

and the variance of the ratio is approximated by

$$\hat{Var}[\hat{R}] = \frac{\hat{Var}[\hat{T}_N] + \hat{R}^2 \hat{Var}[\hat{T}_D] - 2\hat{R}\, \hat{Cov}[\hat{T}_N, \hat{T}_D]}{(\hat{T}_D)^2} .$$

An important note is, for Volume III, all estimate and variance calculations are performed within the context of observed SIC-size. If the estimate required is not a ratio estimate, the estimate and variance are computed across group within BLS SIC-size within observed SIC-size, and these totals can then be summed to produce the estimates and standard deviations appearing in Volume III. (The standard deviation is defined as the square root of the variance. Standard deviations are used because of the easier-to-grasp magnitude and because of their more straightforward statistical interpretation.) If a ratio estimate is required, both the numerator and denominator estimates and variances and the covariance are computed across group within BLS SIC-size within observed SIC-size, and these totals are summed to the levels implied by Volume III. The ratio estimate can then be found as a quotient, and the calculation of variance can be performed with the accumulated values of the numerator and denominator estimates, their variances, and the covariance.

APPENDIX F:

STANDARD DEVIATION TABLES

The attached tables provide the NOHS data user with an indication of the reliability associated with estimates of numbers of employees or numbers of facilities. Values in the tables are approximate standard deviations. The values were obtained by regression averaging of a set of 87 standard deviations computed for the 29 major SIC groups and 3 major size classifications used in the original reporting of the NOHS results. The reader is cautioned that the standard deviation for a particular estimate is likely to be different from the averages provided in the table due to differences in the magnitude of the design effect for different reporting variables.

Internal estimates can be obtained by multiplying the standard deviation by two and alternately adding this value to and subtracting it from the estimated number of employees or number of plants.

TABLE F-1. APPROXIMATE STANDARD DEVIATIONS FOR ESTIMATES OF NUMBER OF FACILITIES, NOHS

Number of Facilities	Approximate Standard Deviation
100	220
200	225
400	235
600	245
800	260
1,000	270
2,000	330
4,000	430
6,000	540
8,000	635
10,000	720
20,000	1,340
40,000	2,425
60,000	3,390
80,000	4,270
100,000	5,100
200,000	8,675
400,000	14,365
600,000	19,060
800,000	23,165

Preceding page blank

TABLE F-2. APPROXIMATE STANDARD DEVIATIONS FOR ESTIMATES OF NUMBER OF EMPLOYEES, NOHS

Number of Employees	Approximate Standard Deviation: Less than 1/3 of the Value of the Estimate is Derived from Large Plants	Approximate Standard Deviation: More than 1/3 of the Value of the Estimate is Derived from Large Plants
2,000	735	-
4,000	1,110	-
6,000	1,420	-
8,000	1,700	-
10,000	1,955	-
20,000	3,075	14,900
40,000	4,975	21,900
60,000	6,675	27,300
80,000	8,275	32,000
100,000	9,815	36,100
200,000	16,950	53,000
400,000	30,125	78,000
600,000	42,675	98,200
800,000	54,900	116,000
1,000,000	66,850	132,000
2,000,000	125,500	200,000
4,000,000	238,750	309,000
6,000,000	350,250	403,000
8,000,000	460,750	489,000
10,000,000	570,750	575,000
20,000,000	1,115,000	-
40,000,000	2,190,000	-