Technical Report 07-01
January 2007

Predicting Industry Personnel Security Investigation Requirements

Eric L. Lang
Defense Personnel Security Research Center

Leissa C. Nelson
Northrop Grumman Technical Services

Technical Report 07-01 **January 2007**

Predicting Industry Personnel Security Investigation Requirements

Eric L. Lang, *Defense Personnel Security Research Center*
Leissa C. Nelson, *Nothrop Grumman Technical Services*

Released By – James A. Riedel

BACKGROUND

The 2000 Defense Authorization Bill included a mandate that the Department of Defense (DoD) estimate annual personnel security investigation (PSI) requirements. Accurate PSI predictions are critical so that DoD can develop budgets to cover industry PSI expenses and adjudication workload. The need for this information increased with the 2005 transfer of nearly all DoD PSIs to the Office of Personnel Management (OPM) and its contractors. DSS contacted the Defense Personnel Security Research Center (PERSEREC) for assistance in improving PSI prediction methods. PERSEREC conducted research and developed an adjusted prediction method that could improve prediction accuracy for industry PSI requirements.

HIGHLIGHTS

PERSEREC reviewed data collected by the Defense Security Service (DSS) Survey of Cleared Facilities (SCF). The SCF is administered to approximately 11,000 cleared defense contractor facilities each year to collect information about the number of PSIs anticipated for the current fiscal year and several years into the future. A low response rate and estimate errors have hindered prediction accuracy. PERSEREC used a regression imputation method to estimate missing survey data, and developed a facility-specific method to correct for over and under predictions by facilities. Using several years of SCF data, it was possible to demonstrate that predictions made using a two-stage estimation procedure (the "adjusted prediction method") produced substantially more accurate PSI estimates than those produced using the current DSS prediction method. Additional recommendations for improving PSI predictions included implementing a Web-based SCF (to improve the speed and quality of survey data) and creating policy to encourage SCF participation by all facilities.

Defense Personnel Security Research Center
99 Pacific Street, Suite 455-E
Monterey, CA 93940-2497

REPORT DOCUMENTATION PAGE

REPORT DOCUMENTATION PAGE	**Form Approved OMB No. 0704-0188**

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE: (01-01-2007)	2. REPORT TYPE Technical Report 07-01	3. DATES COVERED (From – To) May 2005 - January 2007
4. Predicting Industry Personnel Security Investigation Requirements	5a. CONTRACT NUMBER:	
	5b. GRANT NUMBER:	
	5c. PROGRAM ELEMENT NUMBER:	
6. AUTHOR(S) Eric L. Lang, Leissa C. Nelson	5d. PROJECT NUMBER:	
	5e. TASK NUMBER:	
	5f. WORK UNIT NUMBER:	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Defense Personnel Security Research Center 99 Pacific Street, Suite 455-E Monterey, CA 93940-2497	8. PERFORMING ORGANIZATION REPORT NUMBER PERSEREC: Technical Report 07-01	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Defense Personnel Security Research Center 99 Pacific Street, Suite 455-E Monterey, CA 93940-2497	10. SPONSORING/MONITOR'S ACRONYM(S): PERSEREC	
	11. SPONSORING/MONITOR'S REPORT NUMBER(S):	

12. DISTRIBUTION/AVAILABILITY STATEMENT:
Approved for Public Distribution

13. SUPPLEMENTARY NOTES:

14. ABSTRACT:

Accurate annual predictions of personnel security investigation (PSI) requirements are important for the Department of Defense (DoD) budgeting process and for ensuring that the necessary resources are available. The Defense Security Service (DSS) currently obtains PSI prediction estimates through an annual survey of cleared industry facilities. However, actual PSI requirements often differ substantially from PSI predictions. At the request of DSS, the Defense Personnel Security Research Center (PERSEREC) reviewed DSS prediction methods, explored supplementary data that might enhance predictions, then developed and tested a new, adjusted prediction method. The new method holds promise for improving PSI prediction accuracy.

15. SUBJECT TERMS: Personnel security investigations, industry surveys

16. SECURITY CLASSIFICATION OF: Approved for Public Distribution			17. LIMITATION OF ABSTRACT:	18. NUMBER OF PAGES: 66	19a. NAME OF RESPONSIBLE PERSON: James A. Riedel, Director
a. REPORT: Unclassified	b. ABSTRACT: Unclassified	c. THIS PAGE: Unclassified			19b. TELEPHONE NUMBER (Include area code): 831-657-3000

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI td. Z39.18

PREFACE

The 2000 Defense Authorization Bill included a mandate that the Department of Defense (DoD) assess personnel security investigation (PSI) requirements. Accurate PSI predictions are critical so that DoD can develop budgets to cover industry PSI expenses and adjudication workload. The need for this information increased with the 2005 transfer of nearly all DoD PSIs to the Office of Personnel Management (OPM) and its contractors. In response to the mandate, the Defense Security Service (DSS) established a Central Requirements Office (CRO) for industry personnel security clearances and instituted an annual survey of cleared facilities to obtain information about the number of PSIs the facilities expected to require in upcoming years.

Because the survey responses have not been sufficiently accurate in predicting actual PSI requirements, DSS contacted the Defense Personnel Security Research Center (PERSEREC) for assistance in improving prediction methods. PERSEREC conducted research and developed an adjusted prediction method that could improve prediction accuracy for industry PSI requirements. Recommended changes in how annual survey data are collected should also enhance prediction accuracy. PERSEREC presented the findings, including prediction algorithms, recommendations and supporting information, to the DSS CRO May 2005. The current report documents the research goals, methods, findings, and recommendations for improving the prediction of industry PSI requirements.

James A. Riedel
Director

ACKNOWLEDGMENTS

The authors appreciate a number of people who helped with this research. Jeremy Peck contributed substantially to early data collection and analysis efforts. Staff at the Defense Security Service (DSS) provided the data from the Survey of Cleared Facilities. In addition, staff at the Federal Procurement Data Center (FPDC) provided a wealth of useful procurement data.

ACKNOWLEDGMENTS

EXECUTIVE SUMMARY

The Defense Security Service (DSS) is responsible for predicting annual industry requirements for personnel security investigations (PSIs).[1] The accuracy of these predictions is important for the Department of Defense (DoD) budgeting process, particularly given that PSI services are now outsourced to the Office of Personnel Management (OPM) and its contractors. Recent Congressional hearings and deliberations regarding requirements in the forthcoming FY07 Defense Authorization Act[2] (HR 5211, 2006) have highlighted the importance of prediction accuracy and the impact it can have, not only on the budgeting process but also on mission accomplishment.

Currently DSS obtains predicted PSI requirements through an annual survey of cleared industry facilities. However, actual PSI requirements often differ substantially from PSI predictions. The Defense Personnel Security Research Center (PERSEREC) reviewed DSS prediction methods, explored supplementary data that might enhance predictions, and then developed and tested a new adjusted prediction method. The new method holds promise for improving PSI prediction accuracy. PERSEREC presented the research strategy, procedures and findings, including technical details for using the adjusted prediction method, and recommendations for improving annual data collections from cleared facilities, to the DSS Central Requirements Office (CRO) in May 2005. The current report documents the research goals, methods, findings and recommendations for improving the prediction of industry PSI requirements.

The DSS Survey of Cleared Facilities (SCF) is administered to approximately 11,000 cleared defense contractor facilities each year to collect information about the number of PSIs anticipated for the current fiscal year and 5 years into the future. Survey participation is voluntary and the number of facilities that respond is typically low (i.e., 50%-52% for the largest facilities, e.g., AA and A, and 10%-12% for the smallest facilities, i.e., E and F). The low response rate and the resulting missing data hinder overall prediction accuracy. To deal with the problem of missing data, PERSEREC used a regression imputation method that capitalized on strong statistical relationships identified in responding facilities to estimate survey data for facilities that did not provide survey responses. Because a review of archival survey data showed that many facilities inaccurately estimate their PSI requirements, PERSEREC developed a facility-specific method for correcting predictions. Using archival SCF data, it was possible to demonstrate that

[1] In response to a 2000 Defense Authorization Bill mandate that DoD assess background investigation clearance requirements, DSS established a Central Requirements Office for industry personnel security clearances.

[2] Section 336 of the House version of the Act requires a report on PSIs that includes "a description of the procedures used by the Secretary of Defense to estimate the number of personnel security clearance investigations to be conducted during a fiscal year" and "the funding requirements of the personnel security clearance investigation program and ability of the Secretary of Defense to fund the program."

predictions made using a two-stage estimation procedure (the "adjusted prediction method") produced substantially more accurate estimates than those produced using the current DSS prediction method.

In addition to the new method for adjusting survey predictions, PERSEREC's review of the process for predicting PSI requirements identified several other ways in which the prediction process could be improved.

RECOMMENDATIONS

1. Develop a secure, user-friendly, Web-based, annual SCF. A Web-based survey would be faster to field, better at automatically identifying and correcting data entry problems, and could quickly build analysis databases "on the fly."

2. Provide feedback, specific to each facility, in order to help facilities improve the PSI out-year estimates they report on the annual DSS surveys. For example:

   - Include predicted numbers of PSIs from each facility's most recent survey for both current and future predictions.

   - Include actual numbers of PSIs required in prior years, so each facility can see the extent to which previous estimates matched actual requirements.

   - Automate checks for errors and anomalies, such as (a) incorrect CAGE codes, (b) when a facility predicts more investigations than its total number of employees cleared at that level, and (c) when a facility's predictions or actual PSIs for a specific PSI type differ greatly from comparable estimates or requirements.

   - Contact facility representatives whose prior-year estimates and/or next-year predictions for a specific PSI type differed from their actual PSI requirements for that year by some threshold amount (e.g., +/- 95% or +/- 100 PSIs). The discussion should identify real changes at the facility, more general trends, or possible errors.

3. Take steps to improve survey response rate.

   - Request that trade associations (e.g., the Aerospace Industries Association, National Defense Industry Association, Industrial Security Memorandum of Understanding Group) urge all cleared industry facilities to participate in the DSS annual surveys of cleared facilities.

   - Explore whether facilities can be required to participate in the annual DSS surveys (e.g., making it a required part of the annual facility inspection or a precondition to DSS processing of PSI requirements).

4. Conduct follow-up tests of the adjusted prediction method using recently available data on actual PSI submissions.

   - For example, use FY05 SCF and JPAS data and the adjusted prediction method to forecast FY06 industry PSIs.

5. Update and automate Form DD254 (Department of Defense Contract Security Classification Specification form) to provide data useful in improving PSI predictions.
   - Revise Form DD254 to require that contractors include estimates of number of cleared personnel required and number of PSIs anticipated.
   - Create an electronic version of Form DD254 so the data can be stored as a DD254 database and can be made available to improve the accuracy of PSI predictions.
6. Explore the use of data from JPAS regarding the number of industry PSIs clearance "conversions" (i.e., transfers of clearances from one organization to another) to assess whether such data can further improve PSI prediction accuracy.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF TABLES IN APPENDICES

LIST OF FIGURES IN APPENDICES

INTRODUCTION

Through the National Industrial Security Program (NISP), the Defense Security Service (DSS) provides advice and oversight to cleared contractor facilities and assists contractors who need to establish and maintain facility security programs. In addition, DSS is responsible for predicting annual industry requirements for personnel security investigations (PSIs).[3] The accuracy of these predictions is important for the Department of Defense (DoD) budgeting process, particularly given that PSI services are now outsourced to the Office of Personnel Management (OPM) and its contractors. Recent Congressional hearings, Government Accountability Office (GAO) reports and deliberations regarding requirements in the forthcoming FY07 Defense Authorization Act have highlighted the importance of prediction accuracy and the impact it can have, not only on the budgeting process but also on mission accomplishment (Low Clearance, 2006; Progress or More Problems, 2006; Government Accountability Office 2006a, 2006b; HR 5211, 2006).

Currently DSS obtains PSI prediction estimates through an annual survey of cleared industry facilities (i.e., facilities involved in NISP). The annual Survey of Cleared Facilities (SCF) was developed by the DSS Central Requirements Office (CRO) to obtain the information necessary for predicting annual PSI requirements and workforce needs. The survey was first administered in 2001 and gathered facility estimates for the number of PSIs the facilities expected to require each year and for several years into the future.

The SCF provided useful data for predicting annual PSI requirements, but DSS believed the process could be further improved. Overall, numbers of PSIs estimated by the DSS/CRO survey respondents have differed greatly from the actual number of PSIs required.[4] The following are possible explanations:

- An unprecedented major event—for example, the attacks of September 11, 2001—resulting in unforeseen increases in numbers and types of PSIs.
- Difficulties among survey respondents in predicting future contract wins.
- Data entry errors.
- Strategic inflation of estimated annual PSI requirements by survey respondents.

At the request of DSS, the Defense Personnel Security Research Center (PERSEREC) reviewed DSS prediction methods, explored supplementary data that might enhance predictions, and then developed and tested a new adjusted prediction method. The new method holds promise for improving PSI prediction accuracy.

---

[3] In response to a 2000 Defense Authorization Bill mandate that DoD assess background investigation clearance requirements, DSS established a Central Requirements Office for industry personnel security clearances.

[4] Although values for predicted and actual industry PSI requirements based on the SCF data appear on page 49 of GAO's report on *DoD Personnel Clearances* (GAO-04-632), documentation on how those values were developed is not available.

The approach taken by PERSEREC involved comparing predicted PSI requirements to actual observed PSI requirements for the same time period, using multiple data sources.[5] Such comparisons allowed for the identification of meaningful patterns and relationships that could be used to impute missing survey data, and to derive prediction algorithms and methods. The databases examined and strategies employed for using observed results to refine predictions of future PSI requirements are described in the next section.

[5] PERSEREC presented the research strategy, procedures, and findings, including technical details for using the adjusted prediction method, and recommendations for improving annual data collections from cleared facilities, to the DSS Central Requirements Office in May 2005.

METHODOLOGY

The overall goal of the project was to increase the accuracy of predictions of annual industry PSI requirements. The strategy for accomplishing this goal involved: (1) identifying potentially useful data elements and data sources for estimating PSI requirements, (2) resolving data quality issues, (3) identifying data for evaluating the new prediction method, and (4) comparing predicted PSI requirements with actual submissions. Each step is discussed in detail in the following sections.

IDENTIFY DATA FOR PREDICTING PSI REQUIREMENTS

The first step towards improving the accuracy of PSI predictions was to identify data elements that could be useful for making those predictions. After potentially useful data elements were identified, PERSEREC contacted database sources, obtained data, and developed project-specific databases.

PSI predictions must take into account two important factors: security clearance requirements and facility category. Clearance requirements refer to the fact that there are several different security clearances, each requiring a different type of PSI. The clearance and corresponding PSI vary depending upon the level of access required and whether the clearance is new or is a reinvestigation (see Table 1). The main clearance requirements included in the current study were: Top Secret (TS), Top Secret-Periodic Reinvestigation (TS-PR), Secret, Secret-PR, Confidential, and Confidential-PR.

Following a request for one of the security clearances listed above, one of three types of investigations is initiated, depending on clearance requirements: (1) Single-Scope Background Investigation (SSBI; for TS), (2) phased SSBI-PR (for TS-PR), and (3) National Agency Check, Local Agency Check, Credit Check (NACLC; for Secret, Secret-PR, Confidential, and Confidential-PR).

Table 1
Clearance Requirements and Investigation Types

Clearance Requirement	Investigation Type
Top Secret	SSBI
Top Secret PR	SSBI-PR
Secret	NACLC
Secret PR	NACLC
Confidential	NACLC
Confidential PR	NACLC

With respect to facility category, cleared facilities are assigned to one of seven categories by DSS based on the complexity of the security requirements the facility must meet in order to hold a classified contract with a government agency. A

number of factors are involved in determining facility category (see Appendix A: DIS Form 162), but generally speaking, larger facilities must meet more complex security requirements than smaller ones. The facility categories include AA, A, B, C, D, E, and F where AA, A, and B facilities must meet the most complex security requirements (and are generally the largest facilities). Categories C through F refer to those facilities that have to meet less complex requirements (and are generally smaller facilities). A single company could have multiple cleared facilities. A facility refers to an organizational unit, and a company may be made up of multiple organizational units.

A number of sources of potential data elements were considered. In particular, the General Services Administration (GSA), DoD Office of the Comptroller, and DSS were identified as promising sources of useful predictive data elements. The GSA data of interest came from the Federal Procurement Data Center (FPDC), which is the central repository for information on federal contracting. The DoD Office of the Comptroller provided the Future Year Defense Planning (FYDP) database which is generated by the process used to forecast defense costs 5 years into the future. The DSS data came from the Survey of Cleared Facilities (SCF) mentioned previously.

In the search for potential data elements, the DoD Contract Security Classification Specification (DD Form 254) was also reviewed as promising. The DD 254 is used to help contractors identify and understand the security requirements they must follow when performing any classified contract work. The DD 254 currently asks contractors to indicate the types of secure or restricted data they will need to access, the types of restricted or classified material or hardware they will generate, and the types of security guidance they will require. The DD 254 does not currently ask for estimates of the required number of contractor PSIs and thus was not useful for this study.

GSA: Federal Procurement Data Center

The FPDC, part of GSA, tracks federal procurement dollars. The Federal Procurement Data System (FPDS), which is now operated and maintained by Global Computer Enterprises as the FPDS-NG (Next Generation), is the central repository of federal contracting information for contract actions over $2,500. FY04 and later data appear in FPDS-NG; data for prior years appear in FPDS.

The FPDS includes information for approximately 50 data elements (see Appendix B). PERSEREC selected four data elements identified as most promising for the purposes of this study. The relevant data elements (i.e., variables) included: (1) the dollar amount of contracts awarded (classified and unclassified), (2) the contractor (facility) name, (3) the product or service for which the contract was awarded, and (4) the government agency funding the contract. Elements three and four were used to help identify relevant contracts. Contracts were aggregated across government agency for each facility of interest for each of 10 years (1993-2002). For example, if there were 150 contracts awarded to Lockheed Martin Corporation by nine different

government agencies in 2001, all 150 contracts were aggregated to one record containing the total sum of contract dollars awarded to Lockheed Martin for that year.

The correlations (i.e., the statistical measure of the relationship between two data elements) between PSI requirements and total contract award amounts (across all AA-F category types) were high, but the data had limited usefulness for prediction purposes. The primary limitation of the FPDS involved the error introduced by aggregating data across large numbers of facilities, many of which changed names and/or merged over time.

DoD: Future Year Defense Plan

The FYDP reflects DoD resource planning by major expense categories for each fiscal year. Data for 5 years (2001-2005) were analyzed to identify any statistical relationship between the FYDP data and PSI requirements. No useful statistical relationships were found, so FYDP data were eliminated from further consideration.

DSS/CRO: Annual Survey of Cleared Facilities

The SCF is administered to approximately 11,000 cleared defense contractor facilities each year to collect information about the number of PSIs each facility expects to require over the subsequent 7 years beginning with the fiscal year (FY) of survey administration (e.g., for the FY02 survey, the predictions are for 2002 through 2008). A single company could have multiple facilities (for example, offices and plants located in different parts of the country). Surveys were sent to each individual cleared facility, even if the facilities were part of the same company. Although the SCF was first administered in 2001, the SCF database used here covered FY02-FY04 only. The FY01 survey excluded smaller facilities and the FY05 data were excluded because corresponding data on actual PSI submissions were not available.

The SCF is brief (see Appendix C). The first section of the survey asks for information about the facility (Company Name, Location, CAGE Code, and Point of Contact information). The second section lists seven fiscal years beginning with the current fiscal year and provides columns for estimating the requirements for different types of PSIs (SSBI, SSBI- PR, Secret, Secret-PR, Confidential, and Confidential-PR) for each of those years. The third section offers space for comments about the predictions for each year.

While the SCF data were the best available for the present research purpose, three problems limited the utility of the data for the current research: (1) the data were available for only 3 years, thus limiting the development and testing of longitudinal prediction methods, (2) the survey response rate was low, and (3) when compared to historical data, those who did respond tended to overestimate the number of PSIs they were likely to require. Both the low response rate and the tendency to overestimate PSI requirements meant that any method for improving predictions

that was based on SCF data would require two separate adjustments: one for missing data and one for estimate errors. Each of these adjustments will be discussed in more detail in subsequent sections.

RESOLVE DATA ISSUES

According to DSS, response rates to the SCF have typically been low, with larger facilities more likely to respond than smaller facilities. In particular, D and E facilities account for over 93% of all facilities, but only 17% of all D facilities and 12% of all E facilities responded to the FY03 survey (see Table 2). The low response rate made it necessary to identify ways to impute missing data for those facilities that did not respond to the survey. DSS had in place one such imputation strategy and PERSEREC evaluated available data to determine whether a more effective strategy could be found.

Table 2
2003 Response Rate and Actual PSI Requirements by Facility and PSI Type

NISP Facility Category	2003 Survey Response Rate	Cleared Facilities as of January 2003		FY03 Actual PSI Numbers					
				TS[1]		TS-PR[2]		NACLC[3]	
		N	%	N	%	N	%	N	%
AA Facilities	50 %	43	0	1,669	11	2,271	16	16,881	23
A Facilities	52 %	87	1	1,389	9	1,931	14	9,667	13
B Facilities	38 %	125	1	1,373	9	1,200	8	5,403	7
C Facilities	26 %	327	3	1,912	13	1,778	13	5,442	7
D Facilities	17 %	4,167	38	5,186	34	4,584	32	17,957	25
E Facilities	12 %	6,104	55	3,738	24	2,350	17	17,090	24
F Facilities	10 %	175	2	11	0	5	0	181	0
Totals		11,028	100	15,278	100	14,119	100	72,621	100

[1]Top Secret; [2]Top Secret-Periodic Reinvestigation; [3]National Agency Check, Local Agency Check and Credit Check. NACLCs are conducted for Secret, Secret-PR, Confidential, and Confidential-PR clearances.

Strategy 1: Mean Imputation and Overall PSI Estimate Correction

DSS used a mean imputation strategy to fill in responses for facilities that did not complete an annual survey. The mean imputation strategy involved computing the average (mean) number of estimated PSIs for facilities in a specific category (e.g., AA, A, B). This was done by summing all PSI survey data within a category and dividing that sum by the number of responding facilities in that category. Then, for each facility category, the facility average was multiplied by the number of facilities in the category that did not respond to the survey. This value was then added to the subtotal for the facilities that did respond in order to arrive at an overall total for all facilities in the category.

As an example, if SCF data showed a total estimate of 1,098 PSIs for Category B facilities, and the number of Category B facilities that responded to the survey was

100, then the estimated average number of PSIs per facility was 11 (1,098 ÷ 100). If 25 other Category B facilities should have responded to the survey, then the estimated average number of PSIs per facility (11) was multiplied by (25), the number of nonresponding facilities, to arrive at an imputed total for nonresponders (25 x 11 = 275). This estimate for nonresponders was then added to the subtotal for responding facilities (275 + 1,098) to arrive (in this example) at a category B grand total of 1,373. The imputation was thus made at the level of the facility category and assumed that the facilities in a given category that did not respond were similar to the facilities in that category that did. Thus, a single average value, based on the mean of responding facilities, was assigned in a "one size fits all" manner to all nonresponding facilities in a given category.

The second adjustment that DSS made to survey responses was to apply a correction for the fact that survey respondents tended to overestimate PSI requirements. DSS applied a 32% reduction factor across all facility categories to the survey estimates (after missing data were imputed). As with the average imputation method, the 32% correction factor was a "one size fits all" adjustment. No documentation of the logic underlying the 32% reduction factor was available.

Strategy 2: Regression Imputation and Facility-Specific Estimate Adjustment

PERSEREC used a two-stage strategy to impute missing data and then adjust survey predictions to account for discrepancies between PSI estimates and actual PSI submissions. The two aspects of the strategy, imputation and discrepancy adjustment, were independent of one another and served entirely separate purposes. The purpose of the imputation strategy was simply to generate a complete data set.

PERSEREC identified a regression imputation strategy as the most effective strategy for filling in missing responses for facilities that did not complete the survey. To adjust for discrepancies in PSI predictions, PERSEREC developed a facility-specific method that took into account discrepancies at the level of the individual facility. The goal was to improve prediction accuracy by first developing a missing data imputation method that did not assume that all facilities within a given category were the same, and then developing a method to adjust for estimate errors that made adjustments at the level of the individual facility, based on characteristics of that facility.

Regression Imputation: Regression analysis investigates how well values on one variable, called the predictor variable, predict values for another variable, called the outcome variable. In regression analysis, the relationship between the predictor and outcome variables is expressed as an equation in which the value for the outcome variable is equal to an intercept value plus the product of a slope value and a predictor variable (outcome = intercept + [slope x predictor]). Regression imputation uses the equation that results from regression analysis to impute values for cases that are missing outcome variable values. Imputation requires complete

data on the predictor variable and a strong relationship between the predictor variable and the outcome variable (the variable that has missing data in some cases).

A number of variables were considered that could logically show a strong relationship (i.e., correlation) with predicted number of PSIs. The variable that proved most useful was the number of employees who already had a given type of clearance at each facility. The number of existing clearances at each facility was obtained from DoD's Industrial Security Facility Database (ISFD). ISFD is a real-time database that includes information on the numbers of cleared employees at all cleared facilities.[6]

The data for the facilities that responded to the SCF were analyzed in conjunction with the ISFD data and a strong relationship was identified between the number of employees at each facility with a given clearance type and facility estimates for annual PSI requirements for that clearance type. For example, facilities with many Top Secret cleared employees usually predict higher annual Top Secret requirements than facilities with fewer Top Secret cleared employees. The correlations between number of cleared employees and number of estimated PSIs were found to be high, in the range of .75 to .90 (correlation values can range from 0.0 to 1.0). The relationships remained strong even when several outliers (extreme data points) were eliminated.

Regression analysis was applied to the ISFD and SCF data. The number of cleared employees from the ISFD data served as the predictor variable, and the estimated number of PSIs from the SCF data served as the outcome variable in each equation. The result was six regression equations, one for each of the six clearance types (Top Secret, Top Secret-PR, Secret, Secret-PR, Confidential, and Confidential-PR). Note that the ISFD database did not distinguish PRs, so the ISFD data were used in the equations for both new investigations and PRs (e.g., the number of Top Secret clearances already existent at each facility were used to predict future Top Secret and Top Secret PRs).

Each regression equation yielded a slope coefficient and an intercept value. Missing survey responses were imputed by entering the ISFD value for each facility into the regression equations. This involved multiplying the ISFD value for each facility by the slope coefficient and then adding the intercept value to impute missing SCF data (i.e., fill in missing PSI estimates) for each facility that did not respond to the survey. Tables summarizing the regression imputation analysis are shown in Appendix D.

Facility-Specific Estimate Adjustment: The second stage in the PERSEREC two-stage adjustment strategy was aimed at adjusting discrepancies between PSI estimates and actual PSI submissions at the level of the individual facility. For each

[6] The ISFD maintains the number of Top Secret, Secret, and Confidential clearances and does not distinguish initial clearances from those requiring reinvestigation.

year and investigation type, PSI estimates were compared to actual PSI submissions for each facility. Next, for each facility, the difference between the estimated number of PSIs and the actual number of PSIs submitted for that year was used to adjust the estimated number of PSIs for the *next* year.

For example, if a facility estimated they would need 100 TS-PRs in year one but only actually required 80 TS-PRs, then the facility overestimated the required TS-PRs by 20 for year one. If the same facility then estimated that it would need 95 TS-PRs for year two, the estimate would be adjusted downward by 20 to account for the facility's past overestimate. Thus, the adjusted TS-PR estimate for year two for that facility would be 75. Similarly, if a facility underestimated the required number of PSIs for a given year, the adjustment strategy would increase the estimate for the following year accordingly. If any correction resulted in a negative adjusted estimate, the adjusted estimate was set to zero. The number of actual PSIs required was obtained from the Case Control Management System (CCMS) that is described in more detail in the next section.

IDENTIFY DATA FOR EVALUATING NEW PREDICTION METHOD

The method used in this study to improve PSI predictions entailed a comparison of estimated PSI requirements and actual PSI requirements; therefore, it was critical to obtain accurate information on actual PSI requirements (i.e., the numbers of PSIs performed each year). Two databases were identified as the most promising sources of information for actual yearly PSI submissions. The first was OPM's "Report M," which included the number of PSIs scheduled by OPM each month. The second was CCMS, which is an electronic store of information from the Questionnaire for National Security Positions (Standard Form 86), the form used to initiate each PSI.

OPM: "Report M"

OPM's "Report M" data were of limited use in this study because they were only available for the first 7 months of FY05. Further, the number of cases appeared unrealistically low and at the time could not be reconciled with OPM's weekly reports on the same data. For these reasons, the OPM "Report M" data were not used in this study to develop the final method for improving predictions.

Recently, OPM has developed other reports (e.g., Report A) that have been analyzed and are regarded as providing accurate estimates of PSIs performed (Nicewander & Richmond, 2006). However, Report A does not provide information specific to industry PSIs alone. Nicewander and Richmond (2006) also report that the Joint Personnel Adjudication System (JPAS) is a reliable source of data and includes additional useful data elements, which suggests that JPAS data are likely to be the best source of actual PSI submissions for use in prediction models and future research.

DSS: Case Control Management System

CCMS was the system used at DSS for PSI case processing and it linked PSI information from various sources. A review of the CCMS data suggested using the date the PSI was opened as the most useful representation of actual PSI requirements. Thus, actual annual PSI requirements (PSIs opened) for FY00 through FY03 were calculated from CCMS. FY04 data were not available because PSIs began going to OPM for processing during that year and were no longer tracked in CCMS.

The CCMS data were linked to the SCF and ISFD data using the data element called the Commercial and Government Entity (CAGE) code, since CAGE codes were supposed to be common across all three databases. CAGE codes are assigned by the Defense Logistics Information Service (DLIS) and identify companies doing, or wishing to do, business with the federal government. CAGE code format requires numeric characters in positions one and five of the code while the second, third and fourth positions may consist of any mixture of alpha and numeric characters, excluding the characters "I" and "O." Unfortunately, CAGE codes in all three databases (SCF, CCMS, and ISFD) included many errors and anomalies (see Table 3 for examples). CAGE code errors were corrected to the extent possible using available reference sources and were then used to link CCMS, SCF, and ISFD.

Table 3
Examples of CAGE Code Data Anomalies

Sample Incorrect CAGE Codes	
OUL13/CAGE	15090`
*05B9	2H9056
#8X519	2Z 880
#8X519/CAGE	33ENNM
)75M7	3BCF5 3BCF5
00000	4.5.5.6
002769	43219-2268
006811389	UNIDYNE
113	CORP
1 FDA 9	UIC
13-16-69034	1D2Q@

OTHER PREDICTION CONSIDERATIONS

In addition to concerns about missing data and errors in estimation, two other important issues were identified that could impact the accuracy of PSI predictions. The first issue concerned the cumulative effect of large discrepancies between the raw (i.e., unadjusted SCF data) and the CCMS data for actual numbers of PSIs

performed. The second issue was the observation that, historically, trends for PSI requirements have been unstable and any predictions that employ historical PSI trend data must be viewed with caution.

Large Discrepancies: Predicted PSIs versus Actual PSIs

Initial analysis of the raw survey data and the CCMS data identified facilities with very large differences between predicted PSI requirements and actual numbers of PSIs required for a given year. While the PERSEREC two-stage adjustment method could counter this to a large extent, the combined effect of large discrepancies would negatively impact the PSI prediction efficacy. Efforts to identify and contact the small number of facilities with large discrepancies before the PSI estimates are finalized could help identify real changes at the facility, more general trends, or possible errors.

To illustrate the problem, the following paragraphs describe large discrepancies observed in the current study data. For Top Secret PSIs, the total number predicted across all facilities for 2003 was 11,326. When the predictions for 2003 were compared to the actual submissions logged in the CCMS database in 2003 and the 10 largest discrepancies noted, those discrepancies accounted for 23% of the total Top Secret PSI estimate (2,624). As a specific example, the facility with the largest discrepancy predicted that it would need 479 Top Secret investigations in FY03. However, the actual number of Top Secret PSIs required in FY03 by that facility was only 42 – a difference of 437.

At the level of Secret clearances, the total number of PSIs predicted by the facilities responding to the survey for FY03 was 45,989. The 10 facilities with the largest discrepancies in predicted versus actual Secret PSIs accounted for 18% (8,227) of the 45,989 predicted PSI requirements. The company with the greatest difference between predicted and actual Secret PSI requirements predicted that it would require 2,640 Secret PSIs in FY03, but actually required only 148.

At the level of Confidential clearances, the total number of PSIs predicted by the facilities responding to the survey for FY03 was 2,569. The 10 facilities with the largest discrepancies in predicted versus actual Confidential PSIs accounted for 34% (874) of the predicted PSI requirements. The company with the greatest difference between predicted and actual Confidential PSI requirements predicted that it would require 440 Confidential PSIs in FY03, but actually required only 20. This facility alone accounted for 50% of the discrepancy across the 10 facilities. See Appendix E for graphs depicting the discrepancies by facility at each clearance level.

For FY03, the prediction discrepancies described above accounted for 18 to 34 % of all estimated PSIs. If taken at face value, such predictions would have a significant negative impact on budgeting and planning processes at DSS. As this section demonstrates, large errors by a very small number of facilities can have a big impact on the overall accuracy of PSI predictions. Efforts to identify and contact the

small number of facilities whose predictions differ from previous years by a threshold number, before finalizing the PSI estimates for a given year, could yield great benefits. It would be possible to identify not only discrepancies that are the result of errors, but also discrepancies due to real changes at the facility level or more general influences (e.g., the consequence of heightened national security concerns).

Historical Trends

Historical trends are another potential source of information for predicting the future. However, examination shows that FY00-FY03 patterns of actual PSI requirements were unstable and inconsistent, making it difficult to use these patterns for prediction. Figure 1 summarizes these trends for all facility categories combined (AA-F). Secret-level requirements appear to sharply increase between FY00 and FY02 and begin to decrease in FY03 whereas Secret-PRs decrease from FY00 to FY01 but then sharply increase between FY01 and FY02. The number of Top Secret and Top Secret-PR requirements also increases but not as sharply as Secret or Confidential PSI requirements. (See Appendix F for line graphs illustrating these trends by separate facility category types.)

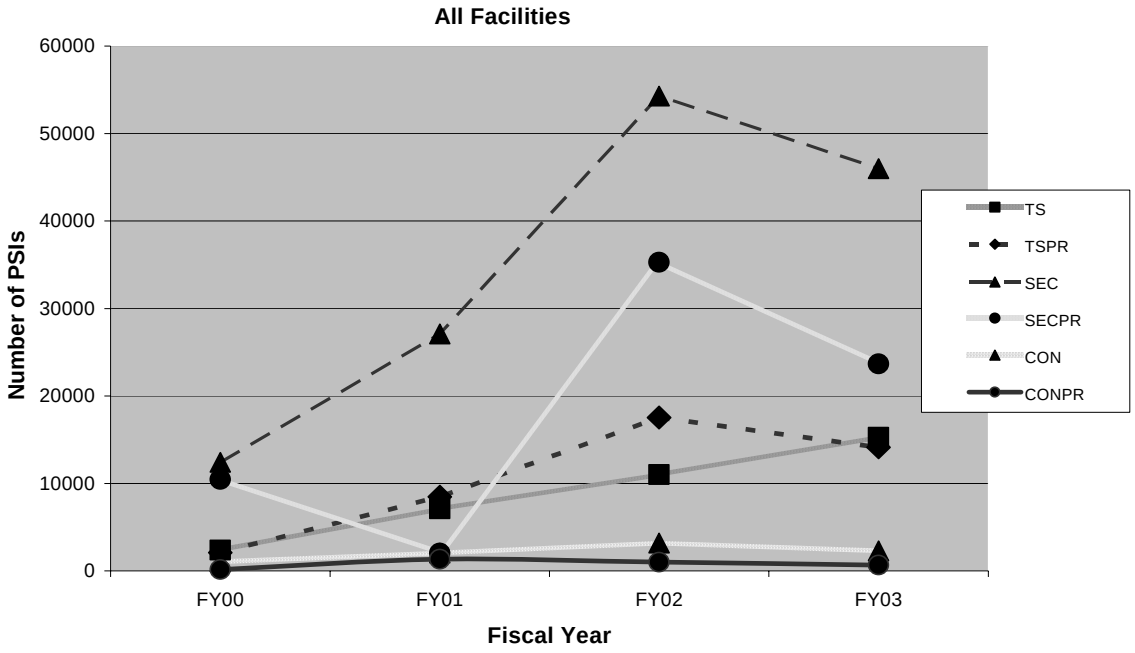


Figure 1 Actual PSI Requirements for all Facility Category Types (AA-F)

METHOD SUMMARY

In summary, the new method for improving annual estimates of PSI requirements involved two adjustments to the data obtained from the SCF: (1) imputing missing data and (2) adjusting for discrepancies between estimated PSI requirements and

actual PSI requirements. The first adjustment accounted for the fact that many facilities did not respond to the SCF. The second adjustment accounted for the fact that estimated requirements and actual requirements did not always correspond.

For the first adjustment, a regression analysis was conducted on a dataset that included the PSI estimates from the SCF and the actual number of cleared personnel from the ISFD for one year for the responding facilities. The slope and intercept values obtained from the regression analysis were applied to the ISFD data for the facilities that did not respond to the survey. The result was a database of PSI estimates. The PSI estimate database consisted of the SCF data from the responding facilities and the imputed data calculated using the slope, intercept, and ISFD values for the nonresponding facilities.

The second adjustment started with a comparison of the PSI estimates and the CCMS data (actual number of PSIs requested) for the *same* year as the PSI estimates. Any observed difference between estimated and actual requirements was added to or subtracted from the estimate for the *next* year, to arrive at a final estimate for each facility. Thus, past-year data is required in order to finalize PSI predictions.

RESULTS

Due to changes in data-keeping systems and the transfer of the investigation function to OPM, only a limited amount of relevant data was available to use to evaluate the new prediction method. In order to conduct a rigorous test of the new prediction method, at least 3 overlapping years of predicted and actual PSI requirements data are necessary. The first 2 years are necessary for method development, and the third year would serve as the actual test of the method. Only 2 years of overlapping data were available (SCF data were available for 2002, 2003, and 2004; complete CCMS data were available through 2003). As a result, only a demonstration of the improved prediction method was possible.

DEMONSTRATE PREDICTION METHOD

Data from FY02 and FY03 were used in the demonstration of the adjusted prediction method. First, PERSEREC imputed missing survey data for the FY02 and FY03 annual surveys using the regression method described earlier. Next, FY02 survey predictions and FY02 actual PSI requirements were used to create an FY02 past-year difference. The FY03 predictions for each facility were adjusted by the FY02 past-year difference. The adjusted FY03 predictions were then compared to actual FY03 PSI requirements to assess the accuracy of the adjusted prediction method.

Results for the demonstration are shown in Figures 2. TS results appear in the first set of bars, TS-PR results appear in the second set of bars, and NACLC results (i.e., all Secret, Secret-PR, Confidential, and Confidential-PR investigations) appear in the third set.

Each set of bars depicts three different numbers of PSI requirements. The left-most bar shows the predicted number of PSI requirements using the DSS adjusted prediction method. The middle bar shows the predicted PSI requirements after application of the PERSEREC adjusted prediction method. The right-most bar shows the actual number of PSIs submitted in FY03.

TS: Using the DSS method, the prediction was 25,537. Using the PERSEREC method, the adjusted prediction for TS PSIs was 12,664. The actual number of TS PSIs (i.e., SSBIs) required in 2003 was 15,278 (i.e., number of “opened” cases according to CCMS). The PERSEREC adjusted prediction method underpredicted TS PSIs by 18%, whereas the DSS method overpredicted by almost 50%.

TS-PR: Using the DSS method, the prediction was 23,544. Using the PERSEREC method, the adjusted prediction for TS-PRs for FY03 was 14,666. The actual number of TS-PRS (i.e., SSBI-PRs) required in 2003 was 14,119. Thus, the PERSEREC adjusted prediction method overpredicted TS-PR requirements by 4% and the DSS method overpredicted by 60%.

NACLC: The DSS method predicted 107,020 NACLCs and the PERSEREC method predicted 85,537 NACLCs. The actual number of NACLCs required for FY03 was 72,621. Thus, the PERSEREC method overpredicted NACLC requirements by 15%, and the DSS method overpredicted by 68%.

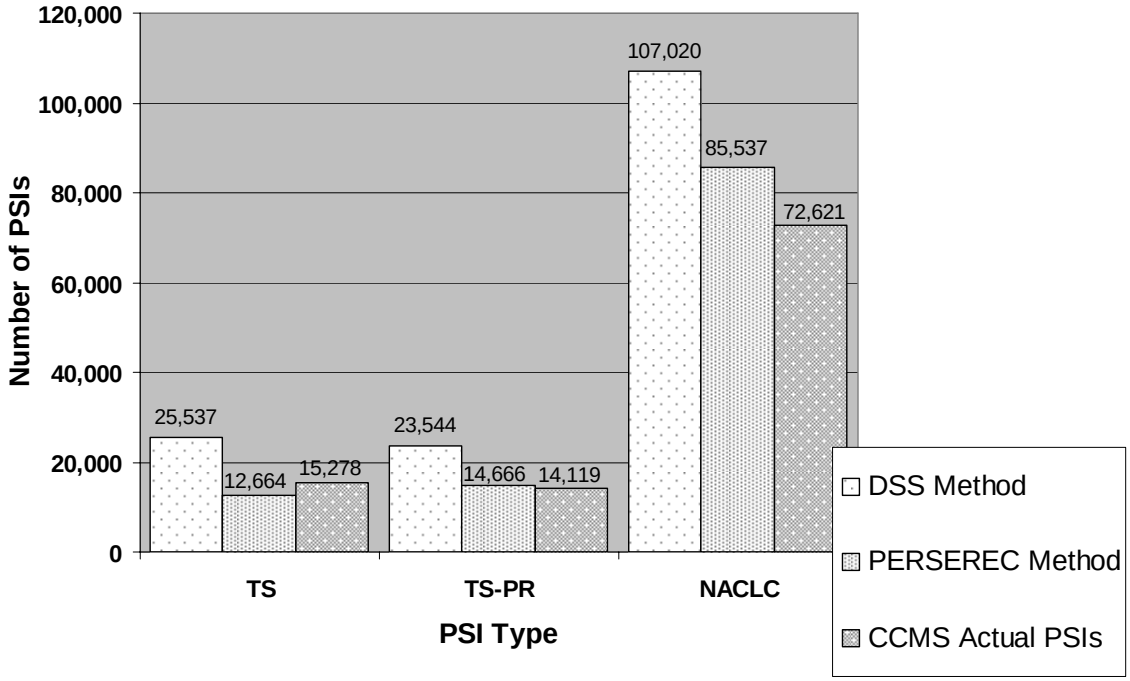


Figure 2 Prediction Method Comparison

SUMMARY

Across all three of the comparisons just discussed, the new prediction adjustment method outperformed the current DSS method by a large percentage. In addition, the new prediction adjustment method has the advantage in that it can handle both over and underprediction by cleared facilities. The current DSS method, which applies a blanket correction by subtracting 32% from the predictions, could have an unfortunate impact if facilities improve their predictions or underpredict their PSI requirements.

CONCLUSION AND RECOMMENDATIONS

Results of the demonstration model indicate that the new adjusted prediction strategy (i.e., regression imputation plus facility-specific adjustments applied to SCF data) could substantially improve predictions of PSI requirements. The new adjusted prediction method, if coupled with feedback to facilities about the accuracy of past predictions, through a Web-based annual survey, should result in a gradual improvement of facility PSI estimates. The proposed strategy for increasing prediction accuracy is designed to accommodate such improvements by making correspondingly smaller adjustments to future-year predictions.

RECOMMENDATIONS

1. Develop a secure, user-friendly, Web-based, annual Survey of Cleared Facilities. A Web-based survey would be faster to field, better at automatically identifying and correcting data entry problems, and could quickly build analysis databases "on the fly."

2. Provide feedback, specific to each facility, in order to help facilities improve the PSI out-year estimates they report on the annual DSS surveys. For example:

   - Include predicted numbers of PSIs from each facility's most recent survey for both current and future predictions.

   - Include actual numbers of PSIs required in prior years, so each facility can see the extent to which previous estimates matched actual requirements.

   - Automate checks for errors and anomalies, such as (a) incorrect CAGE codes, (b) when a facility predicts more investigations than its total number of employees cleared at that level, and (c) when a facility's predictions or actual PSIs for a specific PSI type differ greatly from comparable estimates or requirements.

   - Contact facility representatives whose prior-year estimates and/or next-year predictions for a specific PSI type differed from their actual PSI requirements for that year by some threshold amount (e.g., +/- 95% or +/- 100 PSIs). The discussion should determine identify real changes at the facility, more general trends, or possible errors.

3. Take steps to improve survey response rate.

   - Request that trade associations (e.g., the Aerospace Industries Association, National Defense Industry Association, Industrial Security Memorandum of Understanding Group) urge all cleared industry facilities to participate in the DSS annual surveys of cleared facilities.

   - Explore whether facilities can be required to participate in the annual DSS surveys (e.g., making it a required part of the annual facility inspection or a precondition to DSS processing of PSI requirements).

4. Conduct follow-up tests of the adjusted prediction method using recently available data on actual PSI submissions.
   - For example, use FY05 SCF and JPAS data and the adjusted prediction method to forecast FY06 industry PSIs.
5. Update and automate Form DD254 (Department of Defense Contract Security Classification Specification form) to provide data useful in improving PSI predictions.
   - Revise Form DD254 to require that contractors include estimates of number of cleared personnel required and number of PSIs anticipated.
   - Create an electronic version of Form DD254 so the data can be stored as a DD254 database and can be made available to improve the accuracy of PSI predictions.
6. Explore the use of data from JPAS regarding the number of industry PSIs clearance "conversions" (i.e., transfers of clearances from one organization to another) to assess whether such data can further improve PSI prediction accuracy.

CONCLUSION AND RECOMMENDATIONS

REFERENCES

Government Accountability Office. (2006a). *DoD personnel clearances: Funding challenges and other impediments slow clearances for industry personnel* (GAO-06-747T). Washington, DC: Author.

Government Accountability Office. (2006b). *DoD personnel clearances: New concerns slow processing of clearances for industry personnel* (GAO-06-748T). Washington, DC: Author.

Low Clearance: Why did DOD suddenly stop processing private sector security clearances? Hearing before the House Committee on Government Reform, 109th Cong., (2006).

Nicewander, P.P., & Richmond, D.A. (2006). *DoD personnel security investigation and adjudication timeliness*. Unpublished manuscript. Monterey, CA: Defense Personnel Security Research Center.

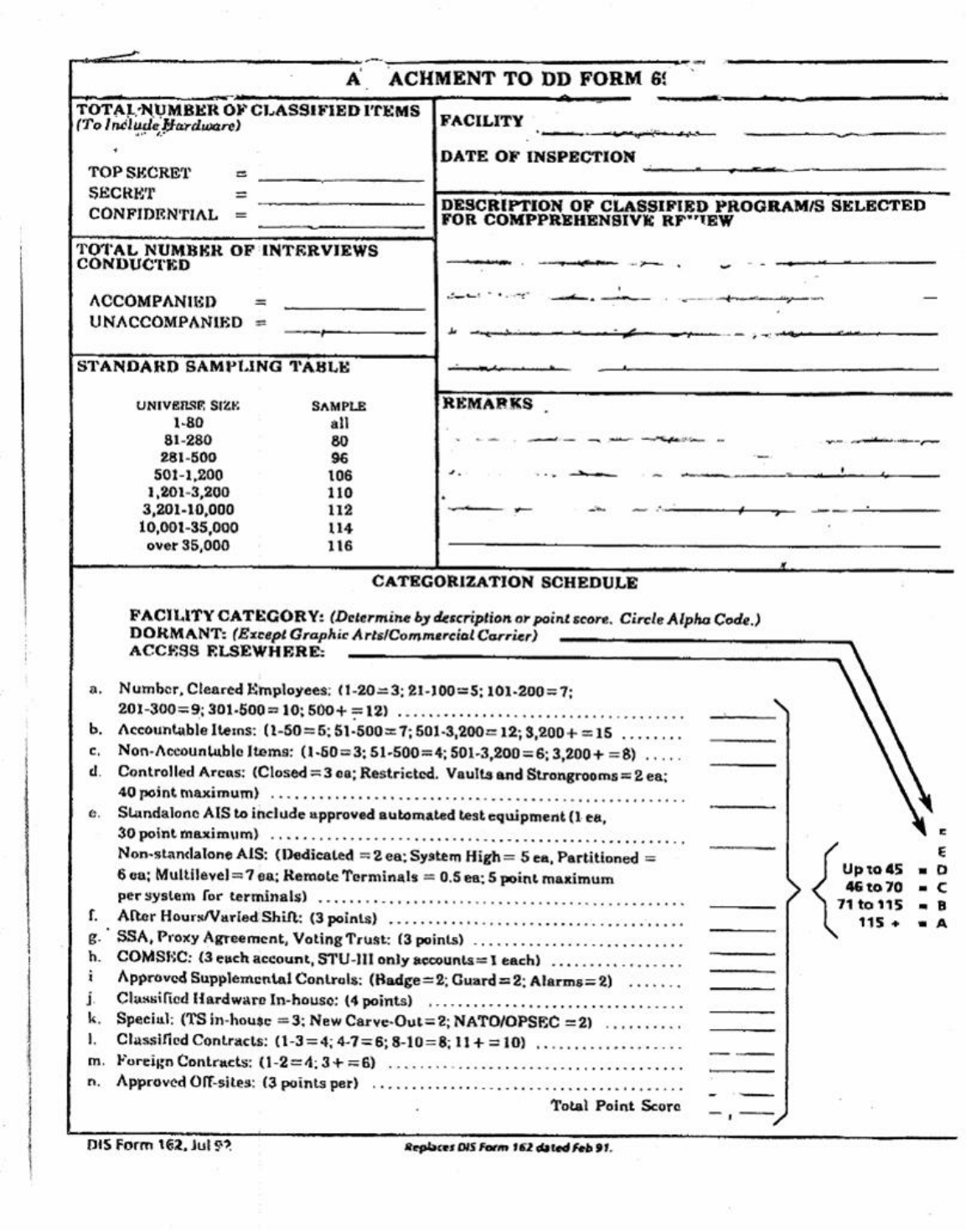
Progress or more problems: Assessing the federal government's security clearance process. Hearing before the Subcommittee on Oversight of Government Management, the Federal Workforce, and the District of Columbia, Senate Committee on Homeland Security and Governmental Affairs, 109th Cong., (2006).

Sonny Montgomery National Defense Authorization Act for Fiscal Year 2007, H.R. 5122, 109th Cong. (2006).

REFERENCES

APPENDIX A

DIS FORM 162 FACILITY CATEGORIZATION SCHEDULE



A ACHMENT TO DD FORM 6

TOTAL NUMBER OF CLASSIFIED ITEMS
(To Include Hardware)

TOP SECRET =
SECRET =
CONFIDENTIAL =

FACILITY

DATE OF INSPECTION

DESCRIPTION OF CLASSIFIED PROGRAM/S SELECTED FOR COMPPREHENSIVE RE''IEW

TOTAL NUMBER OF INTERVIEWS CONDUCTED

ACCOMPANIED =
UNACCOMPANIED =

STANDARD SAMPLING TABLE

UNIVERSE SIZE	SAMPLE
1-80	all
81-280	80
281-500	96
501-1,200	106
1,201-3,200	110
3,201-10,000	112
10,001-35,000	114
over 35,000	116

REMARKS

CATEGORIZATION SCHEDULE

FACILITY CATEGORY: *(Determine by description or point score. Circle Alpha Code.)*
DORMANT: *(Except Graphic Arts/Commercial Carrier)*
ACCESS ELSEWHERE:

a. Number, Cleared Employees: (1-20=3; 21-100=5; 101-200=7; 201-300=9; 301-500=10; 500+=12)
b. Accountable Items: (1-50=5; 51-500=7; 501-3,200=12; 3,200+=15
c. Non-Accountable Items: (1-50=3; 51-500=4; 501-3,200=6; 3,200+=8)
d. Controlled Areas: (Closed=3 ea; Restricted. Vaults and Strongrooms=2 ea; 40 point maximum)
e. Standalone AIS to include approved automated test equipment (1 ea, 30 point maximum)
Non-standalone AIS: (Dedicated=2 ea; System High=5 ea, Partitioned=6 ea; Multilevel=7 ea; Remote Terminals=0.5 ea; 5 point maximum per system for terminals)
f. After Hours/Varied Shift: (3 points)
g. SSA, Proxy Agreement, Voting Trust: (3 points)
h. COMSEC: (3 each account, STU-III only accounts=1 each)
i Approved Supplemental Controls: (Badge=2; Guard=2; Alarms=2)
j. Classified Hardware In-house: (4 points)
k. Special: (TS in-house=3; New Carve-Out=2; NATO/OPSEC=2)
l. Classified Contracts: (1-3=4; 4-7=6; 8-10=8; 11+=10)
m. Foreign Contracts: (1-2=4; 3+=6)
n. Approved Off-sites: (3 points per)
Total Point Score

E
E
Up to 45 = D
46 to 70 = C
71 to 115 = B
115 + = A

DIS Form 162, Jul 9?
Replaces DIS Form 162 dated Feb 91.

Figure A-1 DIS Form 162

APPENDIX B

FEDERAL PROCUREMENT DATA CENTER ELEMENTS

APPENDIX B

(Note: Elements highlighted in grey were used for this study)

Table B-1
Federal Procurement Data System (FPDS) Individual Contract Action Report (ICAR) (SF 279)

REPORTING AGENCY CODE (FIPS 95)
CONTRACT NUMBER
MODIFICATION NUMBER
CONTRACTING OFFICE ORDER NUMBER
CONTRACTING OFFICE CODE
ACTION DATE (YYYYMM)
TYPE OF DATA ENTRY **A** = Original, **B** = Deleting, **C** = Correcting
REPORT PERIOD (YYYYQ)
KIND OF CONTRACT ACTION **A** = Initial Letter Contract, **B** = Definitive Contract Superseding Letter, **C** = New Definitive Contract, **D** = Purchase Orders/BPA Calls Using Simplified Acquisition Procedures, **E** = Order Under Single Award Indefinite Delivery Contract, **F** = Order Under BOA, **G** = Order/ Modification Under Federal Schedule Contract, **H** = Modification, **J** = Termination for Default, **K** = Termination for Convenience, **L** = Order Under Multiple Award Contract, **Z** = Initial Load of Federal Schedule Contract
DOLLARS OBLIGATED OR DEOBLIGATED THIS ACTION (WHOLE DOLLARS)
TYPE OF OBLIGATION **A** = Obligated, **B** = Deobligated
PRINCIPAL PRODUCT OR SERVICE CODE
PRINCIPAL NORTH AMERICAN INDUSTRY CLASSIFICATION SYSTEM
COMMERCIAL ITEM ACQUISITION PROCEDURES **Y** = Yes, **N** = No
CONTRACTOR NAME
CONTRACTOR IDENTIFICATION NUMBER (DUNS)
PRINCIPAL PLACE OF PERFORMANCE (FIPS 55) State City
FOREIGN COUNTRY (FIPS 10)
CONTRACT FOR FOREIGN GOVT. OR INTERNATIONAL ORGANIZATION **Y** = Yes, **N** = No
USE OF EPA DESIGNATED PRODUCTS **A** = EPA-designated product or products were purchased and all contained the required minimum recovered material content, **B** = EPA-designated product or products were purchased without the required minimum recovered material content and a justification was completed based on inability to acquire the products(s) competitively within a reasonable time, **C** = EPA-designated product or products were purchased without the required minimum recovered material content and a justification was completed based on inability to acquire product(s) at a reasonable price, **D** = EPA-designated product or products were purchased without the required minimum recovered material content and a justification was completed based on inability to acquire the product(s) to reasonable performance standards in the specifications, **E** = No EPA-designated product(s) were required
USE OF RECOVERED MATERIAL AND WASTE REDUCTION CLAUSES **A** = Recovered Material and Waste Reduction Clauses, **B** = No Clauses Included
PERFORMANCE-BASED SERVICE CONTRACTING (PBSC) **Y** = Yes, **N** = No
BUNDLING OF CONTRACT REQUIREMENTS **Y** = Yes, **N** = No
COUNTRY OF MANUFACTURE (FIPS 10)
SYNOPSIS OF THIS PROCUREMENT PRIOR TO AWARD **A** = Synopsized Prior to Award, **B** = Not Synopsized Due to Urgency, **C** = Not Synopsized for Other Reason, **D** = Not Synopsized Under the SBA/OFPP Waiver Pilot Program

TYPE OF CONTRACT OR MODIFICATION **A** = Fixed-Price Redetermination, **J** = Fixed-Price, **K** = Fixed-Price with Economic Price Adjustment, **L** = Fixed-Price-Incentive, **R** = Cost-Plus-Award-Fee, **S** = Cost-No Fee, **T** = Cost Sharing, **U** = Cost-Plus-Fixed-Fee, **V** = Cost-Plus-Incentive, **Y** = Time and Materials, **Z** = Labor Hours
CICA APPLICABILITY **A** = CICA Applicable, **B** = Purchase Orders/BPA Calls Using Simplified Acquisition Procedures, **C** = Subject to Statute Other Than CICA, **D** = Pre-CICA, **E** = Commercial Item Acquisition Procedures Under Test Program
SOLICITATION PROCEDURES (Complete only if Item 25 = A) **A** = Full and Open Competition - Sealed Bid, **B** = Full and Open Competition - Competitive Proposal, **C** = Full and Open Competition - Combination, **D** = Architect - Engineer Procedures, **E** = Basic Research, **F** = Multiple Award Schedule, **G** = Alternative Sources, **H** = Reserved, **J** = Reserved, **K** = Set-Aside, **L** = Other Than Full and Open Competition
AUTHORITY FOR OTHER THAN FULL AND OPEN COMPETITION (Complete only if Item 26 = L) **A** = Unique Source, **B** = Follow-on Contract, **C** = Unsolicited Research Proposal, **D** = Patent/ Data Rights, **E** = Utilities, **F** = Standardization, **G** = Only One Source - Other, **H** = Urgency, **J** = Mobilization, Essential R&D Capability or Expert Services, **K** = Reserved, **L** = International Agreement, **M** = Authorized by Statute, **N** = Authorized for Resale, **P** = National Security, **Q** = Public Interest
NUMBER OF OFFERS RECEIVED (Complete Only if Item 25 = A or E) **A** = 1, **B** = 2-5, **C** = 6-10, **D** = 11-15, **E** = 16-20, **F** = 21-50, **G** = Over 50
EXTENT COMPETED **A** = Competed Action, **B** = Not Available for Competition, **C** = Follow-On to Competed Action, **D** = Not Competed
TYPE OF CONTRACTOR **A** = Small Disadvantaged Business, **B** = Other Small Business, **C** = Large Business, **D** = JWOD Nonprofit Agency, **E** = Educational Institution, **F** = Hospital, **G** = Nonprofit Organization, **H** = Reserved, **J** = Reserved, **K** = State/Local Government, **L** = Foreign Contractor, **M** = Domestic Contractor Performing Outside US, **U** = Historically Black College/Universities or Minority Institution (HBCU/MI)
WOMEN-OWNED BUSINESS **Y** = Yes, **N** = No
HUBZONE SMALL BUSINESS CONCERN **Y** = Yes, **N** = No
HUBZONE PROGRAM **A** = HUBZone Sole Source, **B** = HUBZone Set-Aside, **C** = HUBZone Price Evaluation Preference Award, **D** = Combined HUBZone Preference/Small Disadvantaged Business Price Adjustment, **E** = Not Applicable
SMALL DISADVANTAGED BUSINESS PROGRAM **A** = 8(a) Contract Award, **B** = 8(a) with HUBZone Priority, **C** = SDB Set-Aside, **D** = SDB Price Evaluation Adjustment, **E** = SDB Participating Program, **F** = Not Applicable
OTHER PREFERENCE PROGRAMS A = Directed to JWOD Nonprofit Agency, **B** = Small Business Set-Aside, **C** = Buy Indian, **D** = No Preference Program or Not Listed, **E** = Very Small Business Set-Aside
HUBZONE PRICE EVALUATION PREFERENCE PERCENT DIFFERENCE
SMALL DISADVANTAGED BUSINESS PRICE EVALUATON ADJUSTMENT PERCENT DIFFERENCE
SUBCONTRACTING PLAN (Small, Small Disadvantaged, and Women-Owned Small Business) **A** = Required, **B** = Not Required
SUBJECT TO LABOR STATUTES **A** = Walsh-Healey Act, **B** = Reserved, **C** = Service Contract Act, **D** = Davis-Bacon Act, **E** = Not Subject to Walsh-Healey, Service Contract, or Davis-Bacon Acts
ESTIMATED CONTRACT COMPLETION DATE (YYYYMM)
CONTRACTOR'S TIN
COMMON PARENT'S NAME

COMMON PARENT'S TIN
VETERAN-OWNED SMALL BUSINESS (VOSB) **A** = Service Disabled Veteran Owned Small Business, **B** = Veteran Owned Small Business, **C** = Not Veteran Owned Small Business
MULTIPLE AWARD CONTRACT FAIR OPPORTUNITY **A** = Fair Opportunity Process, **B** = Urgency, **C** = One/Unique Source, **D** = Follow-On Contract, **E** = Minimum Guarantee
SMALL BUSINESS COMPETITIVENESS DEMONSTRATION PROGRAM (Applicable to AGR, DOD, DOE, DOI, DOT, EPA, GSA, HHS, NASA, and VA)
DEMONSTRATION PROGRAM **Y** = Yes, **N** = No
EMERGING SMALL BUSINESS **Y** = Yes, **N** = No
EMERGING SMALL BUSINESS RESERVE AWARD **Y** = Yes, **N** = No
SIZE OF SMALL BUSINESS **Number of Employees A =** 50 or less, **B** = 51 - 100, **C =** 101 - 250, **D** = 251 - 500, **E** = 501 - 750, F = 751 - 1,000, **G** = Over 1,000 **OR** **Average Annual Gross Revenue M** = $1,000,000 or less, **N** = $1,000,001 - $2,000,000, **P** = $2,000,001 - $3,500,000, **R** = $3,500,001 - $5,000,000, **S** = $5,000,001 - $10,000,000, **T** = $10,000,001 - $17,000,000, **Z** = Over $17,000,000
FUNDING AGENCY
FUNDING AGENCY – DODAAC
FUNDING AGENCY - COMMERCIAL ITEM CATEGORY **A** = Commercially Available Off-The-Shelf Item, **B** = Other Commercial Item, **C** = Non-developmental Item, **D** = Noncommercial Item, **E** = Commercial Service, **F =** Noncommercial Service
FUNDING AGENCY – REASON FOR PURCHASE **A** = Convenience and Economy, **B** = Expertise, **C** = Specifically Authorized, **D** = Authorized by Executive Order, **E** = Modification or Extension, **F** = Other
FUNDING AGENCY - CLINGER-COHEN ACT Y = Yes, N = No

APPENDIX C

SURVEY OF CLEARED FACILITIES

Table C-1
Sample Survey of Cleared Facilities

Section A	*Column for data entry*
Company Name:	
Facility Category: * (Select AA - F from list box)*	
Cage Code #:	
Location (City):	
Location (State Code):	
Location (Zip Code):	
Contact (Name):	
POC (Email Address):	
POC (Phone Number):	

Section B

Fiscal Year	SSBI	SSBI PR	Secret	Secret PR
2006				
2007				
2008				
2009				
2010				
2011				
2012				

Fiscal Year	Conf	Conf PR	Trust. Invest.
2006			
2007			
2008			
2009			
2010			
2011			
2012			

Section C

Comments concerning FY 2006:	
Comments concerning FY 2007:	
Comments concerning FY 2008:	
Comments concerning FY 2009:	
Comments concerning FY 2010:	
Comments concerning FY 2011:	
Comments concerning FY 2012:	

OMB No.: 0704-0417 Expiration Date: 03/31/08

The public reporting burden for this collection of information is estimated to average 75 minutes per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, completing and reviewing the collection of information.

Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden to Department of Defense, Washington Headquarters Service, Directorate for Information Operations and Reports, (0704-0417), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply if a valid OMB control number is not displayed.

* Facility Categories - acceptable values are "Not Sure", AA,A,B,C,D,E, and F. If your are not sure of the category into which your facility falls, please contact the DSS Industrial Security Representative or Field Office Chief responsible for your facility.

APPENDIX D

SUMMARY OF LINEAR REGRESSION EQUATIONS

SUMMARY OF LINEAR REGRESSION EQUATIONS USED TO IMPUTE PSI ESTIMATES

A large percentage of facilities did not respond to the Survey of Cleared Facilities (SCF) with the requested information, making it necessary to impute values for the missing data. Missing data was imputed using a linear regression strategy. Each regression equation consisted of:

- Predictor variable: number of cleared employees at industry facilities, and
- Outcome variable: number of predicted personnel security investigations (PSI) by type (SSBI (TS), SSBI-PR (TS-PR), Secret, Secret-PR, Confidential, and Confidential-PR) based on the FY02 and FY03 DSS/CRO Survey of Cleared Facilities.

The predictor variable for the regression equation was the number of cleared employees at the Top Secret, Secret, and Confidential levels. The outcome variable was the number of predicted PSI investigative requirements at each clearance level including periodic reinvestigations. Coefficients for the intercept and slope of the regression equations were used to create the value for the imputed predictions as follows:

The unstandardized value for the intercept was added to the product of the slope times the number of cleared employees for a particular clearance level. The resulting value was input into cells for those facilities which did not respond to the DSS/CRO survey.

Table D-1
Using FY02 SCF Data to Predict FY02 PSI Requirements

	Slope	(SE)	Intercept	(SE)	R^2	N
TS	.103**	*(.002)*	1.306**	*(.244)*	.52	2866
TS-PR	.140**	*(.002)*	1.012**	*(.288)*	.58	2866
S	.093**	*(.002)*	5.136**	*(1.062)*	.35	2865
S-PR	.112**	*(.003)*	.081	*(1.200)*	.38	2865
C	.126**	*(.002)*	.446*	*(.193)*	.67	2863
C-PR	.494**	*(.002)*	-.785*	*(.284)*	.93	2863

*p<.05 **p<.001

Table D-2
Using FY02 SCF Data to Predict FY03 PSI Requirements

	Slope	(SE)	Intercept	(SE)	R^2	N
TS	.104**	*(.003)*	1.313**	*(.329)*	.37	2866
TS-PR	.121**	*(.002)*	.696*	*(.240)*	.60	2866
S	.088**	*(.003)*	4.984**	*(1.142)*	.29	2866
S-PR	.060**	*(.001)*	1.087*	*(.427)*	.58	2866
C	.101**	*(.002)*	.448*	*(.175)*	.61	2865
C-PR	.025**	*(.001)*	.152	*(.083)*	.30	2865

*p<.05 **p<.001

Table D-3
Using FY03 SCF Data to Predict FY03 PSI Requirements

	Slope	(SE)	Intercept	(SE)	R^2	N
TS	.108**	*(.001)*	1.393**	*(.056)*	.62	15226
TS-PR	.216**	*(.001)*	-.349**	*(.053)*	.88	15226
S	.090**	*(.001)*	5.750**	*(.131)*	.55	15209
S-PR	.189**	*(.001)*	-1.815**	*(.154)*	.80	15209
C	.015**	*(.000)*	.655**	*(.008)*	.08	15213
C-PR	.058**	*(.001)*	.086**	*(.024)*	.12	15213

*p<.05 **p<.001

APPENDIX E

2003 PSI PREDICTION AND REQUEST COMPARISONS

2003 PSI PREDICTION AND REQUEST COMPARISONS

Figures E-1 through E-6 illustrate the extent to which the actual number of PSIs required in 2003 differed from the number of PSIs predicted for 2003 for the ten facilities that showed the largest discrepancies.

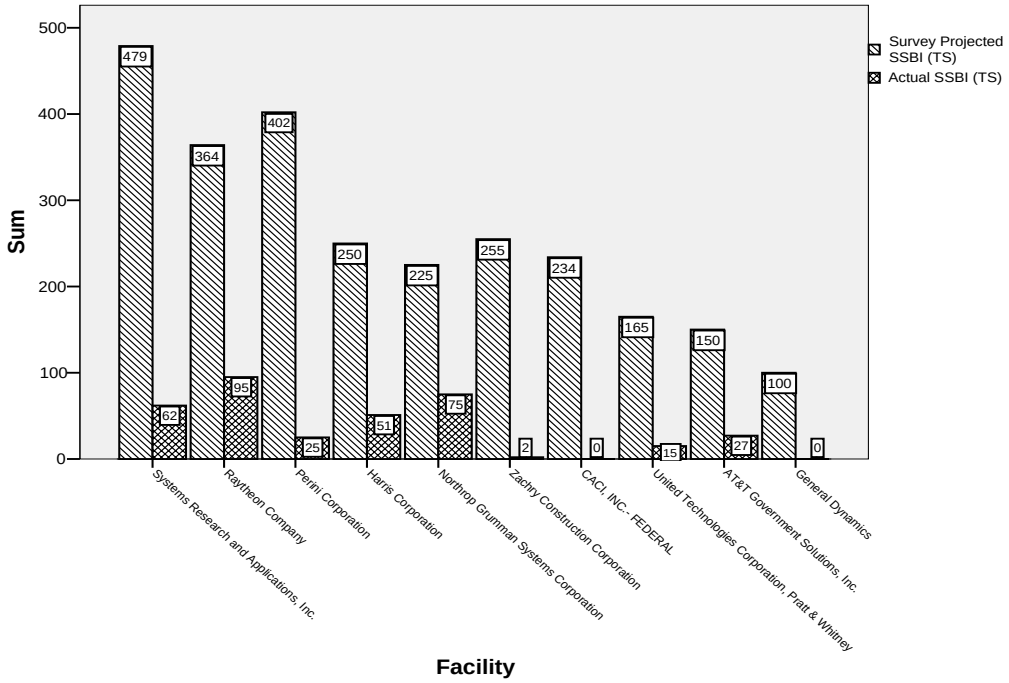


Figure E-1 FY03 Top Secret PSIs: Facilities with Greatest Differences between Survey Predictions and Actual PSI Requirements

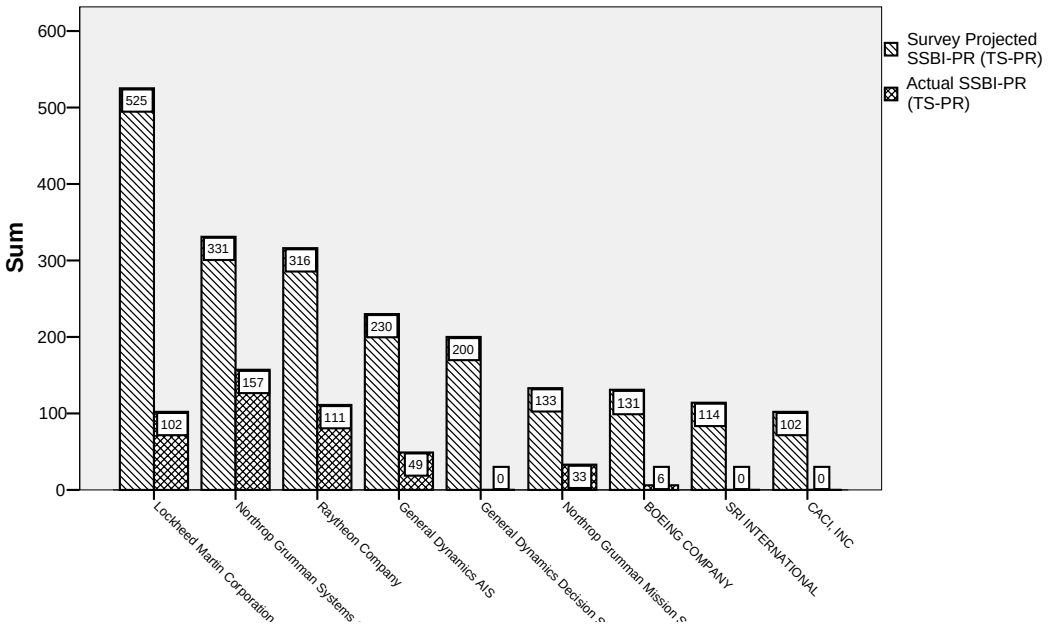


Figure E-2 FY03 Top Secret Periodic Reinvestigations (PRs): Facilities with Greatest Differences between Survey Predictions and Actual PSI Requirements

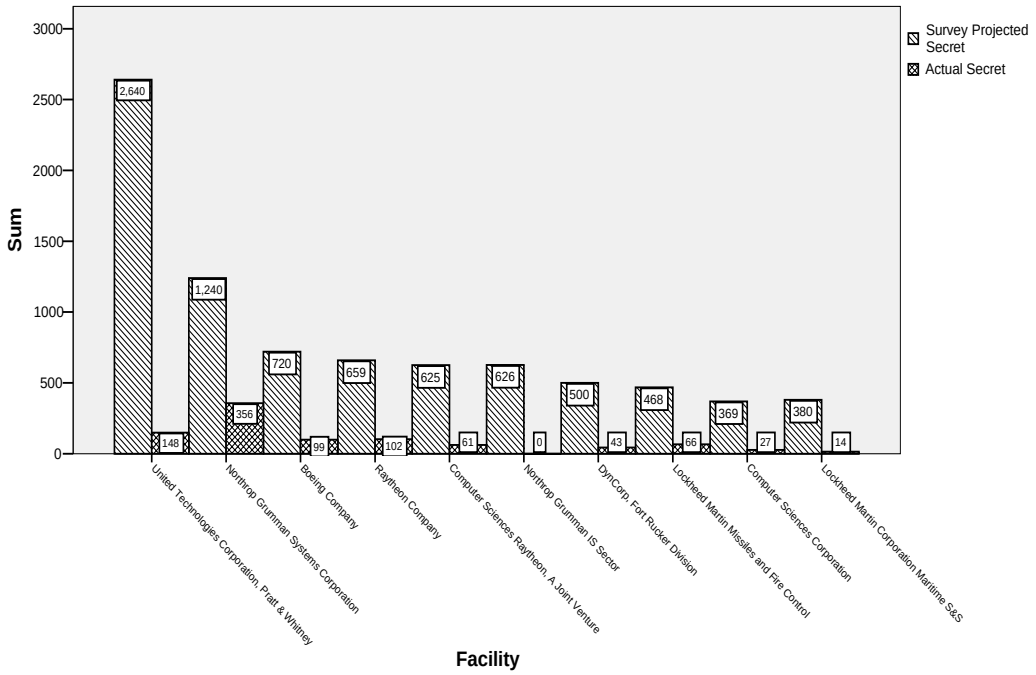


Figure E-3 FY03 Secret PSIs: Facilities with Greatest Differences between Survey Predictions and Actual PSI Requirements

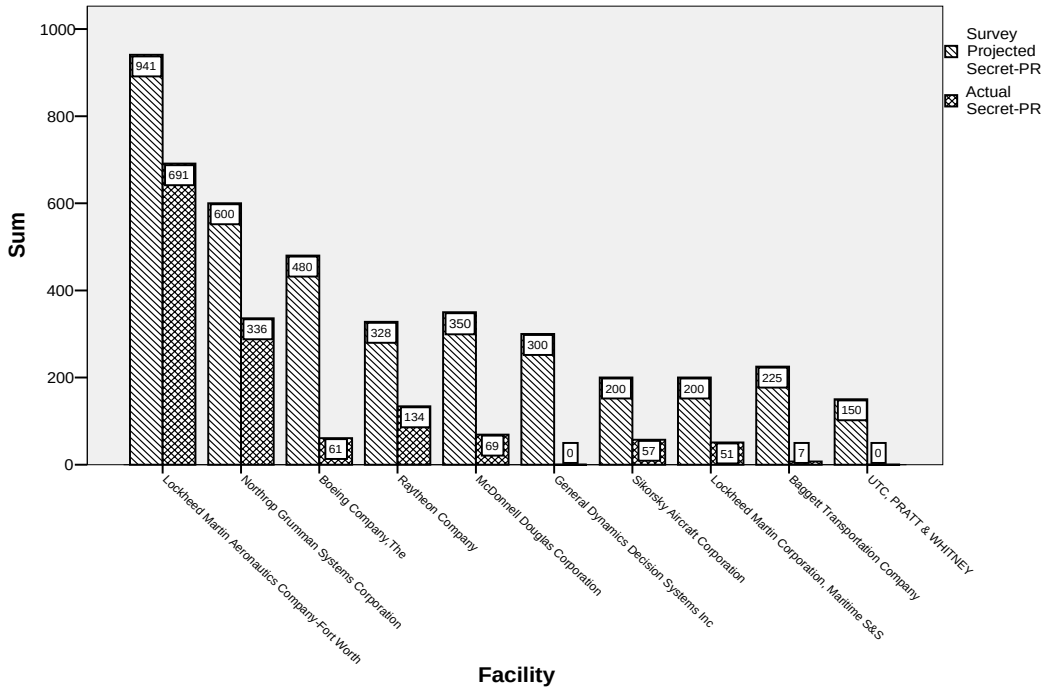


Figure E-4 FY03 Secret PRs: Facilities with Greatest Differences between Survey Predictions and Actual PSI Requirements

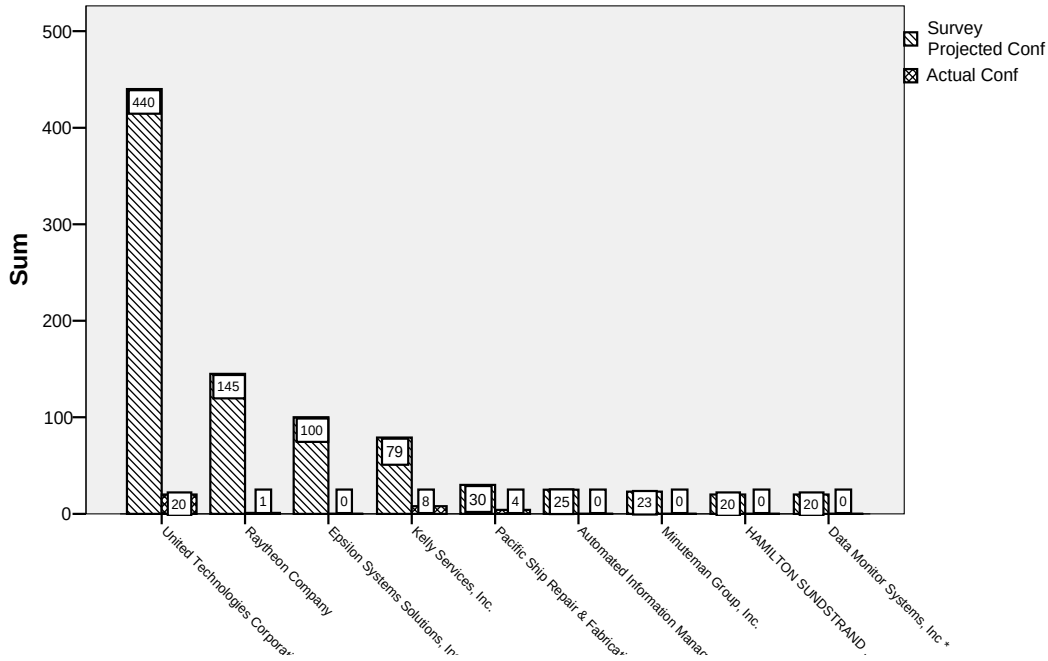


Figure E-5 FY03 Confidential PSIs: Facilities with Greatest Differences between Survey Predictions and Actual PSI Requirements

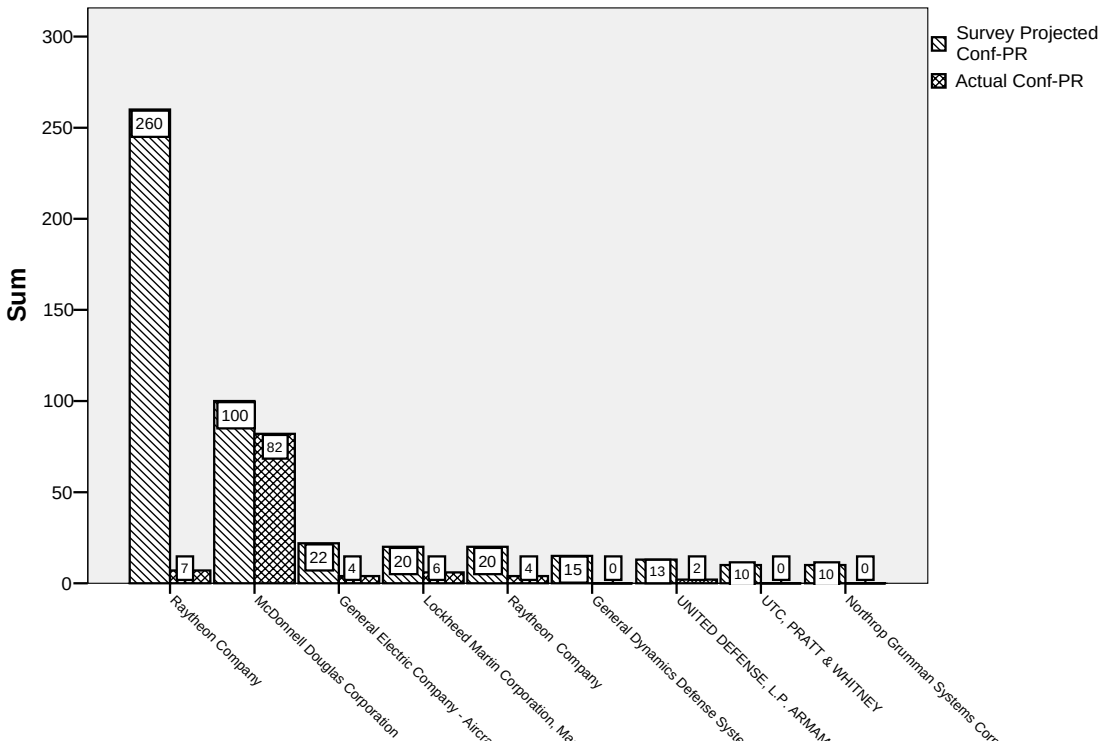


Figure E-6 FY03 Confidential PRs: Facilities with Greatest Differences between Survey Predictions and Actual PSI Requirements

APPENDIX F

PSI REQUIREMENTS BY FACILITY CATEGORY

Figures F-1 through F-7 summarize trends for the actual number of PSI required separately by facility category type. Consistent among each category type is the significant spike of secret and secret-PRs between 2001 and 2002. These numbers begin to decrease in 2003.

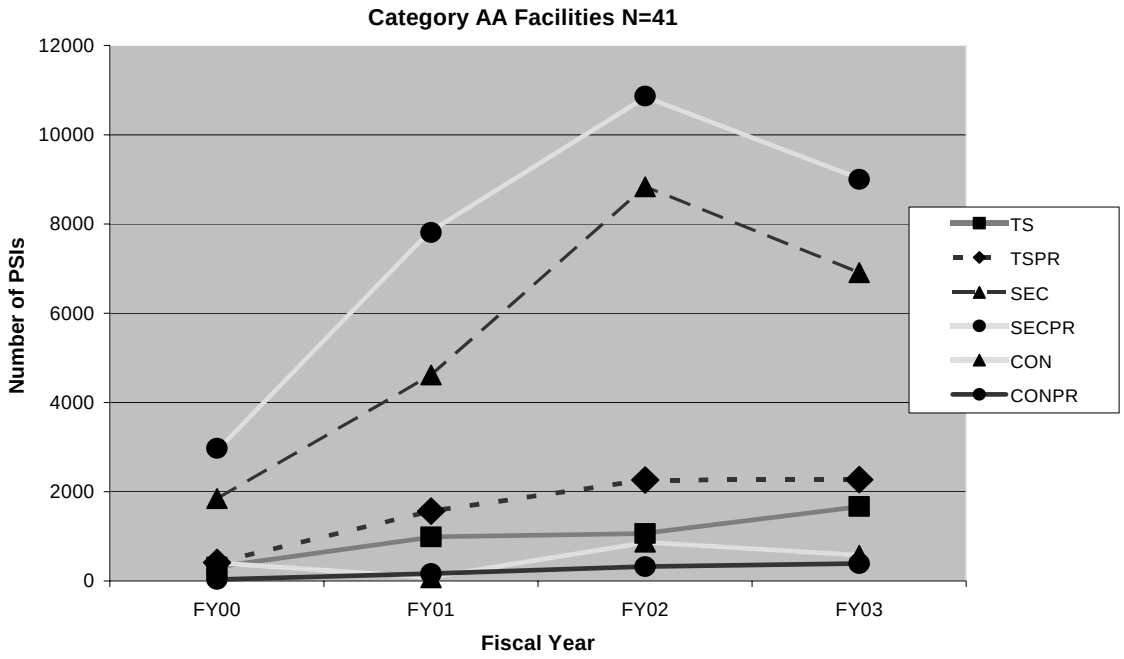


Figure F-1 Actual PSI Requirements for all Facility Category AA

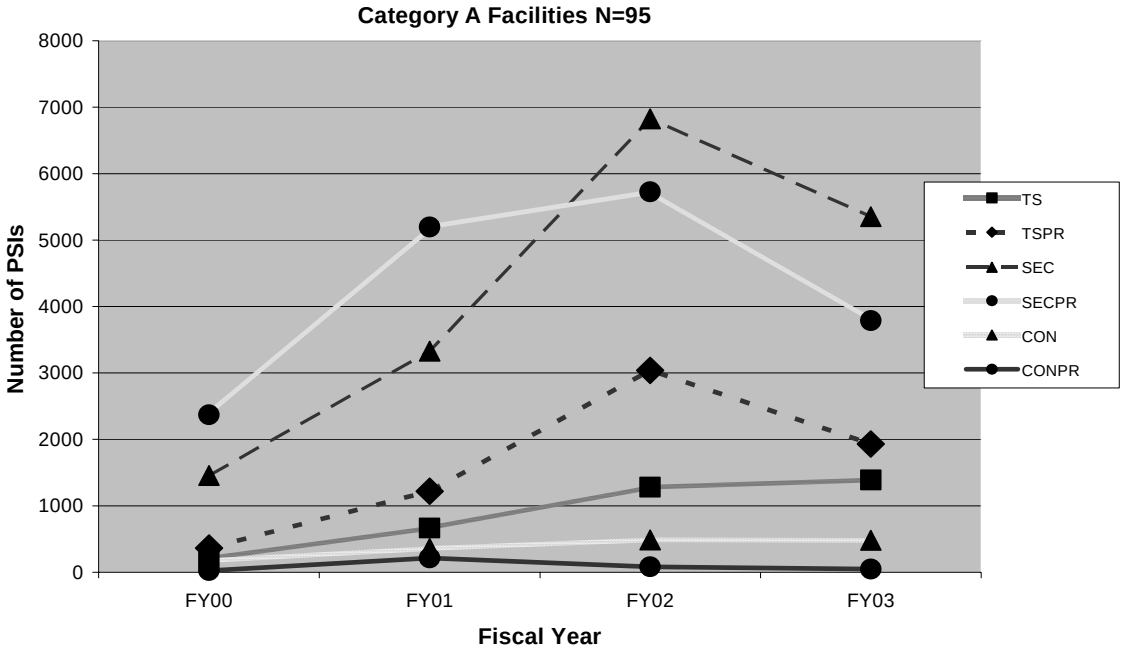


Figure F-2 Actual PSI Requirements for Facility Category A

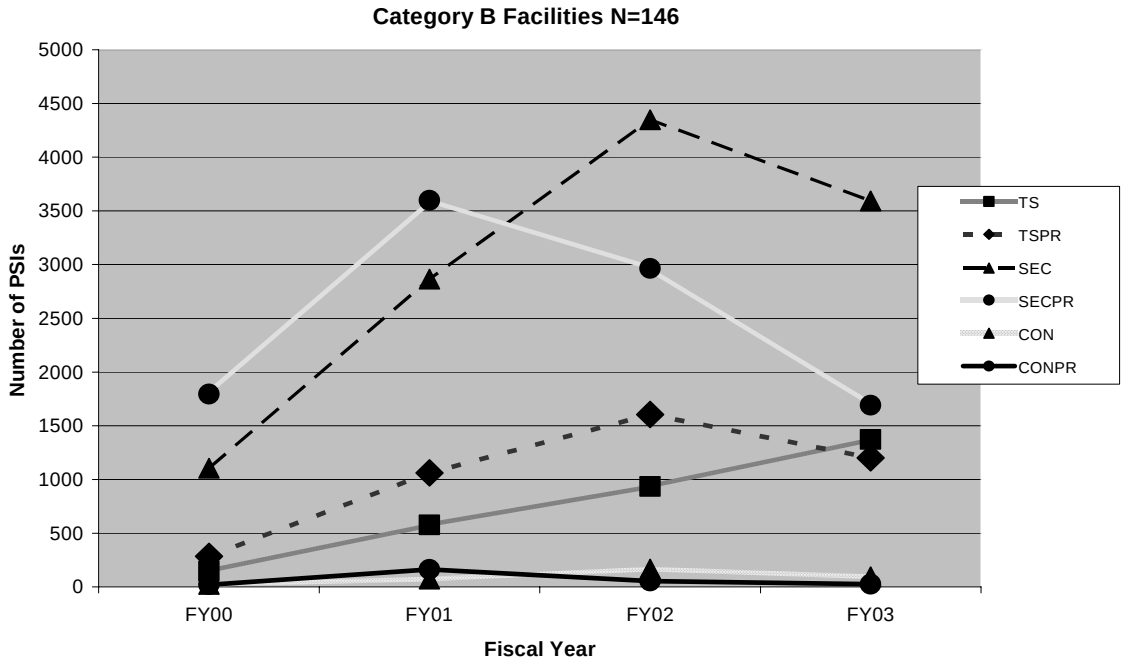


Figure F-3 Actual PSI Requirements for Facility Category B

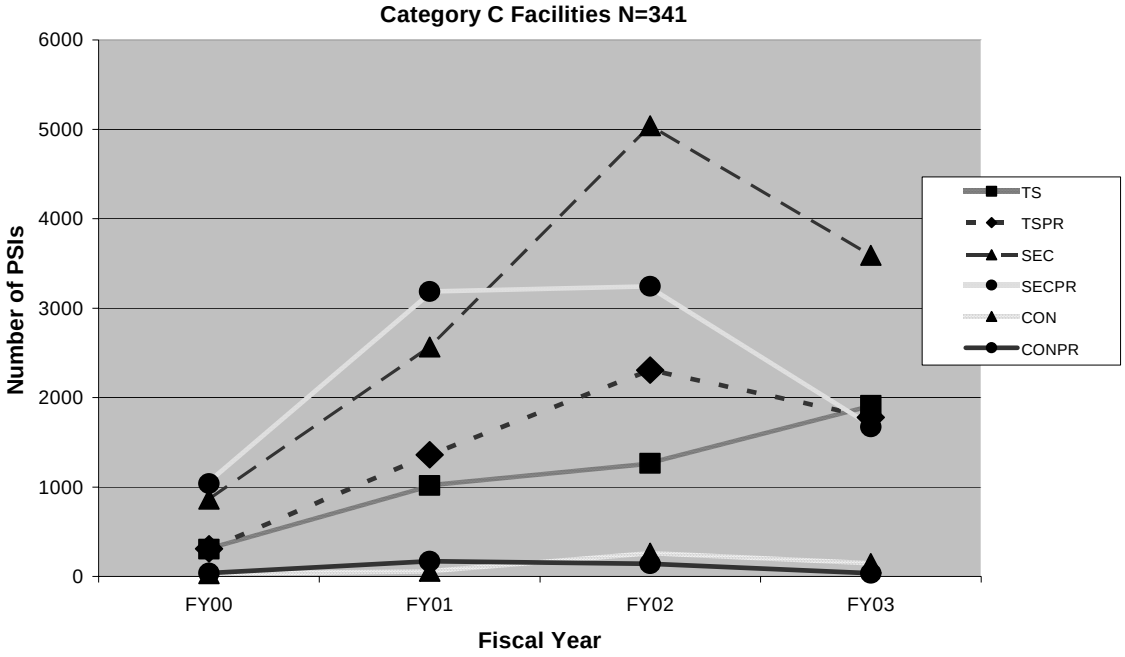


Figure F-4 Actual PSI Requirements for Facility Category C

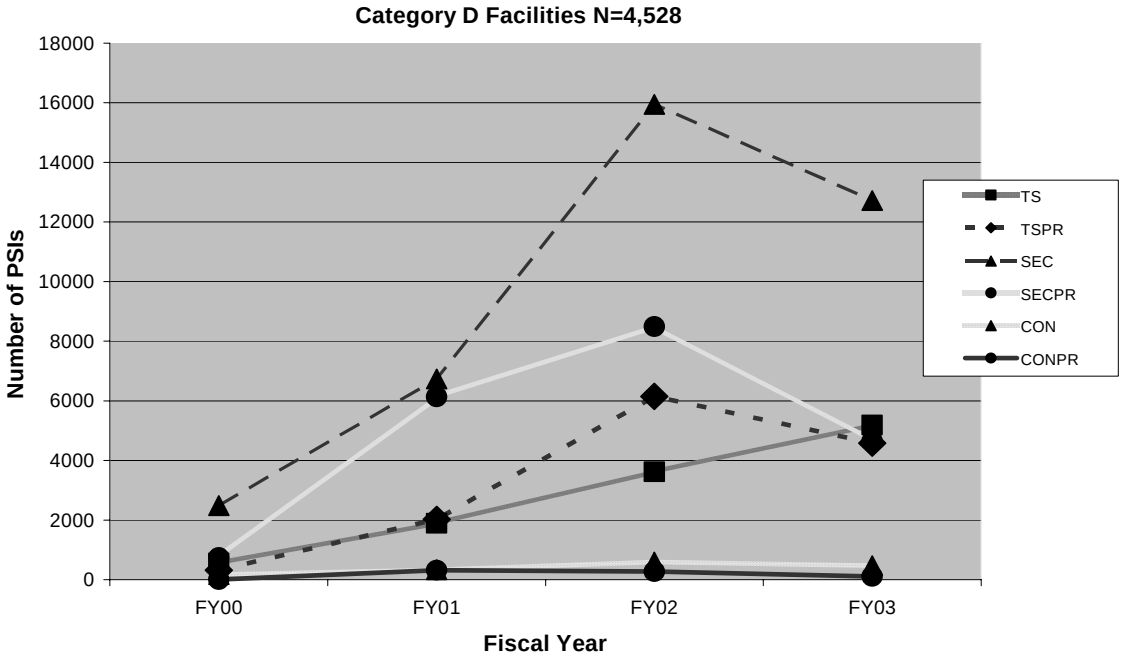


Figure F-5 Actual PSI Requirements for Facility Category D

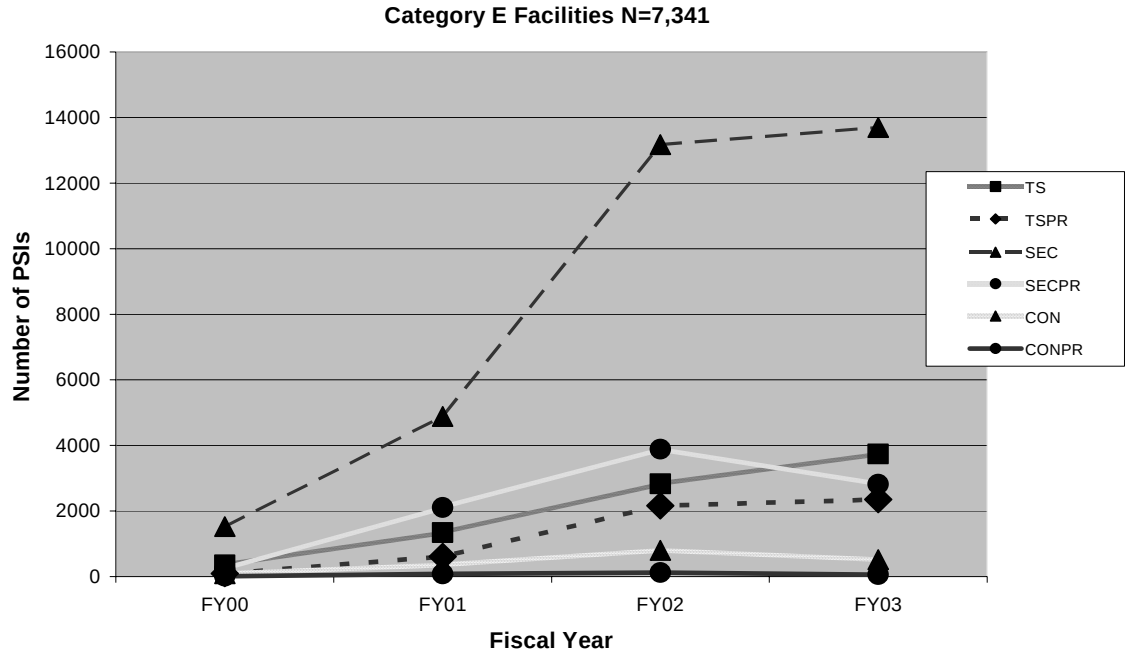


Figure F-6 Actual PSI Requirements for Facility Category E

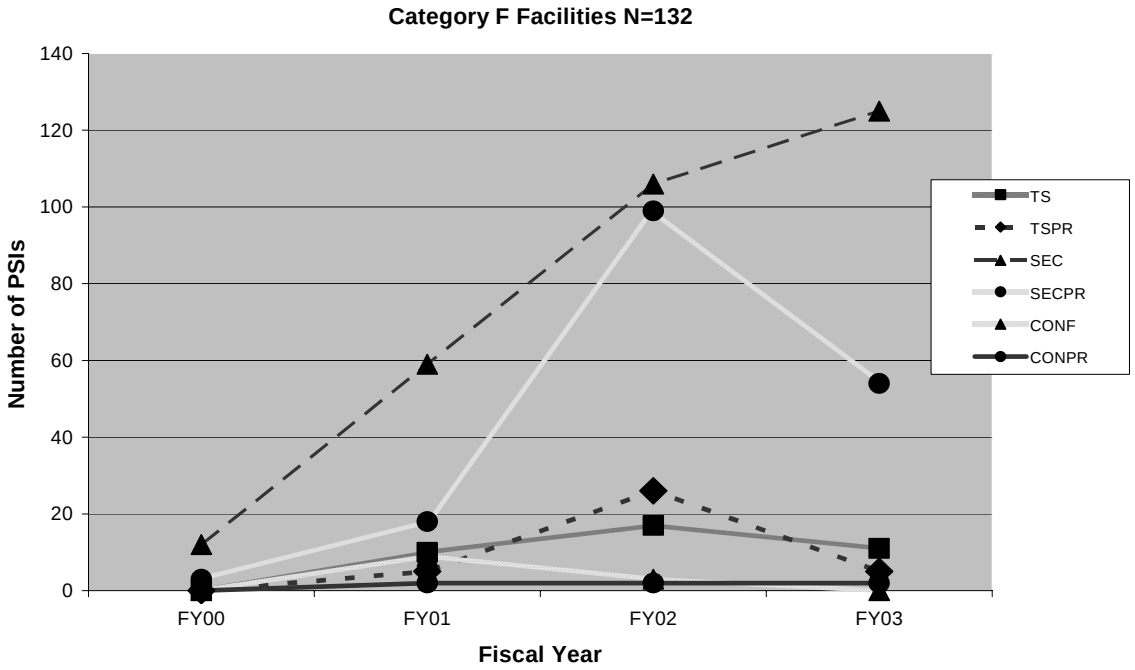


Figure F-7 Actual PSI Requirements for Facility Category F