THE COST AND BENEFITS OF RELIABILITY IN MILITARY EQUIPMENT

Arthur J. Alexander

December 1988

P-7515

The RAND Corporation, 1700 Main Street, P.O. Box 2138, Santa Monica, CA 90406–2138

SUMMARY

This report investigates the costs of achieving greater reliability in military equipment, the benefits of improved reliability in reduced support costs and increased availability, and strategies for attaining reliability goals. Three kinds of evidence were examined: reliability improvement programs, new product developments, and statistical analyses of reliability costs and outcomes in new programs. Literature reviews of both military and commercial experience, and interviews with reliability experts and managers in government and industry, provided additional information to supplement detailed case studies.

We performed detailed case studies of seven systems: F-18 aircraft, CH-47D helicopter modernization, F100 turbine engine, Phalanx Mk15 Close-in Weapon System, LAMPS MKIII helicopter antisubmarine warfare system, Minuteman I inertial navigation system, and the Carousel inertial guidance system. Information produced in other studies was reanalyzed in the context of the present research; these included F-16 aircraft reliability improvements, spacecraft reliability costs, Duane models of reliability growth, and a study of 19 Navy systems. These cases covered several different technologies and were drawn from all three military services. Because the total number of cases on which the conclusions are based is neither large nor homogeneous, these findings cannot be precise. The work is best interpreted as a mosaic; although each piece may suffer from rough edges, the collection as a whole forms a picture more complete than any of the fragments could individually portray. Briefly, the data indicate that reliability improvements are possible, that the greater the improvement the more costly the necessary investment, and that the improvement probably rises proportionally faster than the investment.

THE COST OF RELIABILITY

When reliability was a goal of equipment users and developers, substantial reliability improvements were possible. We observed up to 15-fold reliability improvements (see Table 2 in the text). Original reliability levels did not appear to constrain the size of the improvements. Because costs and improvements varied widely,

we compared the percentage changes in RDT&E investments attributable to reliability with percentage changes in reliability. These data indicate increasing returns to reliability investments: a 10 percent increase in reliability would cost 5 percent more in total RDT&E expenditures, whereas a doubling of reliability would cost 20 percent more; and a five-fold reliability gain would require at least a 50 percent increase in development costs. There was some evidence that programs with serious reliability shortcomings are somewhat less costly to improve than are more "ordinary" programs. Because of a selection bias inherent in the analysis of reliability-improvement programs, these numerical results are overly optimistic when applied to a typical program.

In most cases, unit production cost did not rise with increased reliability. Some possible production cost increases may have been masked by learning-curve effects or by contractor absorption of higher cost, but the consistency of the near-zero cost changes suggests that the bulk of the effects are in nonrecurring investments.

Performance tradeoffs that reduce stress can increase the life of parts and components and reduce the probability of failure. Review of these cases and other relevant experience suggests that reliability is often improved at a cost of reduced performance. Although this route to reliability was not universally applied, it is becoming a more explicit choice among users and developers. In the cases examined, well-chosen reductions or limits to performance often yielded substantial reliability gains.

THE BENEFITS OF RELIABILITY

Reliability improvements have direct effects on maintenance time and manpower, spare parts usage and investments, operational availability, logistics loads, and life-cycle costs (see Table 4 in the text). Availability rates increased by one-fifth to one-half as much as the rate of reliability improvement (a 100 percent reliability gain could yield availability increases of 20–50 percent). An example of the returns to a doubling in reliability is the CH-47D modernization, which generated a 50 percent reduction in unscheduled maintenance, 28 percent reduction in total maintenance, and an estimated savings in labor and parts of about 20 percent. A 200 percent reliability improvement in the Phalanx system was associated with an 80 percent maintenance manhour reduction and 20 percent fall in parts demand. In every case we examined, the financial savings alone justified the investments in reliability, often by substantial margins. However, because reliability-improvements are chosen because they are expected to produce substantial savings, the level of improvements noted here may not be representative of a "random" program. Nevertheless, all of the evidence—when viewed as a whole—suggests that similar benefits are attainable, perhaps at a somewhat greater cost.

RELIABILITY STRATEGIES

Priority for reliability was a necessary condition for the reliability strategies to be effective. The buyer not only has to state a demand for reliability but also has to specify, measure, test, demonstrate, and pay for it.

We considered five approaches for obtaining better reliability levels: (1) improved technology, (2) additional resources directed toward reliability in design and development, (3) performance tradeoffs, (4) higher quality, and (5) time and experience. The evidence does not support a single approach or particular combination as being most effective. The results appear to be specific to time period, knowledge, and technology. In most of our case studies, greater development resources devoted to reliability produced greater reliability than in "standard" programs. However, other experience—most notably Soviet weapons—suggests that performance tradeoffs can yield high reliability without the necessity of spending more in development. In highly complex equipment, especially avionics, high reliability of individual components is not sufficient to guarantee reliability of the system as a whole—primarily because of failure modes introduced by integrating many components and their software. In such cases analysts have called for the use of standard interfaces, the development of reliable building blocks, and a maturational phase for the incorporation of reliable building blocks in an integrated design. This approach has become the norm for aircraft engines.

DATA AND ANALYTICAL NEEDS

The rising demand for more reliable military equipment has generated questions on rates of return and appropriate strategies. Data are not readily available to answer such questions. The military services now have offices with reliability responsibilities. So much information is now routinely passing through these offices, and they are developing an expertise and visibility over all their service's efforts; therefore they are good candidates to perform the data collection task. The cost of reliability has been thoroughly neglected in the past, so collection of such information would help to close this data gap.

ACKNOWLEDGMENTS

Research based on case studies can not be done without the cooperation and participation of many different individuals and organizations. The military services and companies associated with the systems described in the cases provided just such cooperation. Special mention must be made of the courtesy and helpfulness of the staff of Pratt & Whitney, who spent weeks collecting the information used in this report and in educating the authors in their interpretation.

CONTENTS

FIGURES

TABLES

I. INTRODUCTION

In recent years, the U.S. military establishment has been shifting its priorities in systems development: A past emphasis on enhanced operational performance has now been broadened to include higher levels of equipment reliability and supportability. As attention has moved to reliability, decisionmakers have begun to ask the traditional policy questions: How much is enough? How much reliability should be bought? How much does it cost in development, production, and operational performance? What are its benefits in increased operational availability, reduced maintenance demands, and lower spares consumption and inventories? Are there effective strategies for obtaining better reliability? What are the roles of design, technology, testing, performance tradeoffs, and experience? This report seeks to answer these questions, emphasizing the set of issues dealing with the cost of achieving greater reliability.

One type of evidence for a growing consensus that reliability merits increased attention and resources lies in the specifications for new systems. In the early 1960s, the development specification of the F-111 aircraft contained no mention of reliability.[1] Ten years later, the specification document for the F-15 included a qualitative paragraph on reliability. In the mid-1970s, the F-16 development called for several quantitative requirements for reliability at both the system and subsystem level. Such current aircraft programs as the C-17 contain extensive quantitative reliability requirements for which the producer is penalized if the specified levels are not achieved. Perhaps more indicative of a shift in values is the higher priority given to reliability among the different dimensions of performance. In the Navy's F-18 aircraft and F404 engine, the Army's CH-47 helicopter modernization and UH-60 utility helicopter development, and the Air Force requirements for an advanced tactical fighter, reliability has been given first or second priority in the programs' goals.

This concern for reliability is reflected in the attention given by *Aviation Week and Space Technology* magazine in a special issue in which an article notes "the U.S.'s military services now are placing renewed emphasis on the design and engineering disciplines that will give them reliable equipment." The primary reason for this is the

---

[1] These examples are taken from Abell et al., 1988, p. 1.

"stronger emphasis on reliability at top management levels in the Defense Dept. and industry."[2] However, this review of reliability issues goes on to highlight a concern that is the focus of the present study. "Yet to be determined are the costs this may add to future development programs and what changes might have to take place in the Defense budgetary process to shift money forward to the earlier research and development category to pay for increased reliability."[3]

To structure our search for information and its analysis, we emphasize the notion of investments in reliability: expenditures in one time period that produce returns later. The investments required for greater reliability include additional time to gain information, resources for test and evaluation, improved engineering techniques, better knowledge of failure mechanisms, and improved quality assurance techniques in production. In most of the cases we have examined, these investments had small effects on production unit costs. Moreover, the returns were often substantial, as measured by reductions in resource use (labor and parts) in fielded equipment and by increases in equipment availability—equivalent to investments in a larger force size.

ANALYTICAL METHODS

The main obstacle in the path of this study was the absence of data and studies specifically devoted to the cost of reliability. Initially we reviewed the literature on commercial and military equipment, conducted telephone interviews with reliability executives in industry, and visited government offices concerned with reliability in all the military services and the Office of Secretary of Defense. Because of the dearth of empirical treatment, we decided to direct our efforts to the cost of improved reliability through the use of data generated by ourselves and through the reanalysis of earlier studies. Although many previous studies did not directly treat the cost of reliability, it was sometimes possible to extract such information from the published data. We were able to use such diverse sources of information as studies on Soviet aircraft, U.S. Navy engineering analyses, and space satellite reliability data. However, we also concluded that we would have to develop our own data on specific programs with the cooperation of systems program offices in the military services and manufacturers of military

---

[2] "Military Stresses Maintainability, Reliability," *Aviation Week and Space Technology*, October 6, 1980, pp. 42, 43.

[3] Ibid., p. 43.

equipment. Such case studies permitted a more thorough and richer investigation into the cost of reliability than was obtainable from the literature. With the information obtained from published sources as well as from our own case studies, the empirical basis for the present study comprises considerably more than fifty observations.

The underlying conceptual framework of this study is the assumption that tradeoffs are possible among investments, reliability, and product performance. The study attempts to verify the existence of these tradeoffs and to produce some evidence of their dimensions.

The assumed relationship that governs the tradeoffs can be stated simply in terms of an implicit function that defines a tradeoff surface:

$$F(C, X, R) = 0 \ ,$$

where C is total acquisition investments (including development, test, and production costs), R is a measure of product reliability, and X is a vector of other product attributes such as weight, number of parts, or speed. The relationship is deliberately written in an implicit form to emphasize that policymakers can choose which variables will be specified directly and which will be determined by the tradeoff constraints. In a qualitative way, the analysis attempts to isolate the partial derivatives of this function, holding other things constant.

In our own data gathering efforts, we envisioned several information sources for investigating the cost of reliability. These methods involve collection and analysis of data generated by three different kinds of processes: (1) reliability improvement programs for previously developed and deployed equipment, (2) new product development programs that were identifiably different from other programs in their treatment of reliability, and (3) statistical analyses of many programs that can "hold constant" variables that may influence development and production costs.

Reliability improvement programs are usually applied to systems that have already been developed, produced, and deployed. In some cases, demonstrated reliability is thought to be so low that corrective action is necessary to bring the item up to acceptable standards. In other cases, opportunities to improve reliability arise through new technology or improved understanding gained from operating experience. Improvements may also be sought when user values shift to place more weight on reliability than in the past. Among the cases examined in this study, reliability improvement programs were initiated when serious reliability problems emerged in the Phalanx Mk 15 Close-in

Weapon System, the F100 turbine engine, and the early reliability improvement programs on the Minuteman I inertial guidance and Carousel inertial navigation system. Opportunistic reliability improvement programs were undertaken for the CH-47 helicopter modernization and the later stages of the Carousel inertial system as new technology made such progress feasible at relatively low cost.

The analytical benefits of examining reliability improvement programs lie in their control of extraneous variables. Both "before" and "after" reliability measures can be obtained as well as the cost of producing the change. Also, applying the same data collection system and the same definitions reduces the often serious measurement problems associated with evaluating reliability.

Under certain conditions, data generated by reliability improvement programs can lead to biased estimates of the cost of reliability, which may suggest that reliability is cheaper to obtain, in general, than is actually the case. Such a downward bias exists for those programs where field experience demonstrates lower reliability levels than had been planned and where an improvement program is subsequently undertaken. Reliability improvement costs in these cases will reflect actual costs, but they will produce underestimates of projected costs because only those projects with sufficiently low costs or high returns will actually be selected as reliability improvements and thus enter into the database.

Downward biases in cost estimates are also produced by exogenous technological improvements over time that shift the cost-reliability relationship at no expense to the examined project. However, product improvement projects undertaken because of changed values of reliability will produce unbiased estimates.

A different kind of problem with using reliability improvement programs as data is that they may include a disproportionate number of systems whose experience is so poor that the cost of improving them is unrepresentative. Moreover, we do not have firm expectations of whether evidence from such programs over- or underestimate normal experience. Designs may be so bad that improvements are inordinately difficult, or even small changes may produce considerable results. However, the very fact that such programs were undertaken suggests that the expected returns were large enough to justify the program. Indeed, we find that that is the case, although the returns do not appear to be out of line with the returns from other classes of programs.

Another problem is that reliability improvement programs may not be relevant to new systems. We believe that it is more costly to produce an improvement when designs are frozen, as with deployed equipment, than when designs are flexible and choices have yet to be made. If this commonly expressed view is accurate, then the cost of increased reliability in existing systems would be higher than in new developments.

Case study analysis of new development programs must contend with a perplexing question. How can the cost of reliability be segregated from the other development costs? In absolute terms, this question is unanswerable, because almost every aspect of design affects reliability. However, if some feature of an otherwise standard program is different, it may be possible to isolate the cost and effect of that feature. Changing the length or intensity of a test program, for example, or imposing a design constraint such as a high-temperature limit may yield identifiable costs and consequences. The F-18 aircraft case represents such a phenomenon where new reliability development techniques and policies were identified, and where reliability and cost figures could be compared with historical norms. The Duane models described here are also a form of this type of analysis where different levels of testing are associated with changes in reliability. The analysis of spacecraft reliability falls between this method and the statistical method described below because the referenced study attempts to identify all explicit reliability related development expenditures and link these to demonstrated reliability experience. The Navy study of 19 new system developments can also be placed in this analytical category because it attempted to estimate relationships between reliability outcomes and various features of the reliability program.

The third analytical method uses statistical techniques to identify the cost of reliability while accounting for the confounding effects of other variables. From a sample of systems falling within a definable technology class (turbine engines or spacecraft, for example), equations of the type described above could be estimated. This study includes no examples of the pure form of this approach, although the spacecraft study and the study of 19 Navy systems are partially representative.

Neither case studies nor statistical analyses of new programs will generate the kind of downward cost bias noted above for reliability improvement programs. As long as different users exhibit different values for reliability, the data will trace out the cost-reliability frontier (holding other things constant).[4]

---

[4] The task here is to actually hold other things constant across dissimilar classes of products. Otherwise, qualitative evaluation methods establish the limits of precision.

Table 1 classifies those studies from which we could extract numerical estimates of the cost of reliability into the analytical framework described above.[5] Most of our detailed case studies were of product improvement programs. The F-18 was our only example of a new system. However, previous studies provided additional examples of analyses of new systems—the Duane models, the spacecraft sample, and the Navy systems.

DATA SOURCES

The best descriptive term for our case selection technique is "opportunistic." We incorporate the results of cases where sufficient useful information was obtained. We had initiated data collection efforts for several programs that we do not report in this study because some of the required information was unavailable. Case studies of project developments often confront a barrier to thorough analysis: Government project offices or industry contractors often do not keep historical cost data and other information; our

Table 1

DISTRIBUTION OF CASES ON THE COST OF RELIABILITY AMONG ANALYTICAL METHODS

Analytical Method	Cases
I. Product improvement program	
A. Severe reliability problems	F-100 engine Phalanx Close-in Weapon System Minuteman I inertial guidance Carousel inertial navigation (early program phase)
B. Opportunistic improvements	CH-47 helicopter modernization Carousel inertial navigation (late program phases)
II. New systems, case studies	F-18 aircraft Duane models
III. New systems, statistical studies	Spacecraft 19 Navy systems

---

[5] Many other informative studies are referenced in the text. These include: P-3C aircraft (p. 8); Navy electronics equipment (p. 8); electronics components (p. 8); ships' boilers (p. 12); guns and missiles (p. 12); fire-control systems (p. 12); M1 and M60 tanks (pp. 12–13); automobiles (pp. 15–17); tires (p. 17); SR-71 aircraft (p. 22); Soviet fighter aircraft (p. 22); Soviet helicopters (p. 22); F404 engine (p. 23); Soviet R-11 engine (p. 24); U.S. J-79 engine (p. 24); avionics (p. 34).

experience is consistent with such a pattern. We therefore had to rely on cases where we could obtain information within the resource and time constraints of this study. Likewise, the cases drawn from past published and unpublished sources possess much of this same opportunistic character. We used what we could find. Although we do not see any patterns in the data arising from our choice of cases, the sample is certainly not a random selection. The results of our case studies are reported in some detail in the appendixes to this report.

We recognized that most reliability data, as collected by standardized systems, are flawed in several ways. The chief problem lies in the definition of a failure and the implications of different types of failures for mission performance and subsequent resource use in maintenance and repair. We attempted to get around this problem by using, where possible, comparative reliability data that are generated by the same data collection systems according to the same definitions. By focusing on differences rather than on absolute levels of reliability, we reduce the data definition and coverage problems to the extent that they remain stable across our comparisons. For example, the Army gathered CH-47C and CH-47D helicopter reliability data in a special collection effort that was designed to maintain comparability across the two models. We believe that this approach of highlighting net effects gives a more accurate portrayal of the results of investments in reliability than would the use of clearly imperfect absolute values.

Technological change generated outside an analyzed program may distort some comparisons. General knowledge, new materials, or advanced techniques often allow reliability to be increased with little additional project expenditures. The further apart in time that comparisons are made, the more likely that general technological change will play a role. This phenomenon enters into the evaluation of the F-18 aircraft, later stages of the Carousel navigation system, and CH-47D helicopter.

The conclusions drawn from this research cannot be interpreted as having been produced from a fully articulated, statistically estimated model, although a system of interrelationships guided our efforts and established a conceptual framework for fitting in the many details. The work is more of a mosaic, with the placing of odd bits to form a picture more complete than any of the roughly shaped fragments could individually portray. There is a danger in this approach: The composition may be biased, reflecting an unbalanced collection of information and incorrect inferences drawn from it. The possibility of biased results points to the need for additional work along the lines reported here. Our findings cannot be precise; the uncertainty bounds are larger than we would have desired. However, we produce more quantitative results than any other past study and thus see this effort as a needed step in a worthwhile direction.

II. STRATEGIES, COSTS, AND BENEFITS OF IMPROVED RELIABILITY

STRATEGIES FOR IMPROVING RELIABILITY

We identify here several approaches toward better reliability levels: (1) improved technology; (2) additional resources directed toward reliability in design and development; (3) performance tradeoffs; (4) higher quality specifications, materials, and production processes in design and production; and (5) time and experience. These strategies are not pure; they overlap and influence each other. No one of them is sufficient to guarantee desired outcomes; the balance among them is one of the central policy questions that continues to be the subject of considerable research and debate. At a higher strategic level however, an effective reliability program requires that reliability be demanded. Priority for reliability turns out to be a necessary concomitant to successful programs.

Technology

Over the past ten years or so, the greater priority given to reliability in military equipment has encouraged the growth of technology related to failure mechanisms and to materials. For example, the importance of temperature on failure rates of electronic equipment has been vividly demonstrated by analyses showing the effect of environmental temperatures on operating cost and failure rates; application of this knowledge to the Navy's P-3C antisubmarine warfare aircraft showed that a 10° (F) increase in equipment bay ambient temperature from customary levels would decrease mean time before failure (MTBF) by 18 percent (from 4.0 to 3.28 hours) and increase annual operating costs by an estimated $19 million (based on a 200 aircraft fleet). Similarly, a 5° cooling of the equipment would increase MTBF by 11 percent (to 4.5 hours) and save $8.5 million per year in operating costs.[1] At the component level, engineering tests show that reducing avionics junction temperatures from the common 140–150°C to a new Navy standard of 110–120°C would improve component MTBF by

---

[1] "Military Stresses Maintainability, Reliability," *Aviation Week and Space Technology*, October 6, 1980, p. 42.

a factor of 20.[2] Another example of advances in reliability technology is the research on the importance of thermal cyclic stress in turbine engines in the late 1960s that demonstrated the inadequacy of earlier design standards and test regimens. In the materials area, better knowledge of metallic failure mechanisms stimulated the search for methods to produce single-crystal alloys for turbine blades to enhance hot-section reliability of turbine engines.

The importance of reliability technology lies in its general application across models of one type of system, and even across systems. Because the benefits are widespread, the returns to investments in advancing reliability technology are less likely to be captured by the company or organization making the investment than are more narrowly conceived investments in particular system developments. This becomes a prime reason for government investment and incentives in reliability technology; private companies and individual programs are unlikely to make the level of investments that would be desirable from a broader, national level perspective.

Design and Development

Design and development for reliability concentrates the available knowledge and technology during systems development. If reliability has high priority, more design and development resources will be devoted to such things as reducing temperatures of electronics bays. Designers of the F404 turbine engine, which gave reliability high priority, spent considerable effort in reducing the number of parts; as a result, the F404 had about 60 percent fewer parts than the F100 engine designed under a different order of priorities.

Knowledge of the temperature-reliability relationship in electronics enabled the Navy to specify new aircraft design standards and infrared scanning of new equipment to seek out hot spots. These standards and specifications are now part of the design and development process and are not cost free.

In general, an important characteristic of design and development resources applied to reliability is that more reliability requires more resources—for given levels of knowledge, technology, and system performance goals. In our examination of the cost of reliability, increased development expenditures is one of the areas we emphasize.

---

[2] This information was obtained from briefings presented by Willis J. Willoughby, Jr., Deputy Chief of Naval Materiel for Reliability, Maintainability, and Quality Assurance.

Performance Tradeoffs

It is now generally accepted that for given development resources pushing the state of the art in seeking high operational performance will also result in unreliable systems. One method for increasing reliability, therefore, is to back off on performance requirements to reduce component stress. Reduced performance is a price that can be paid for higher reliability and is symmetrical with development and production costs in its potential effects. New technology can ease these tradeoffs; single-crystal alloys, for example, permit higher turbine inlet temperatures with consequent higher thrust-to-weight ratios and improved fuel efficiency, or they can be used to make longer wearing parts. Technology can loosen constraints, but it does not eliminate the need for assigning priorities and considering tradeoffs.

Quality

The quality of specified materials and quality control in manufacturing govern the way in which a design is transformed into a product and the subsequent reliability of that product. Higher quality materials, components, and production processes can drive up production costs unless greater attention is paid to reliability and cost effects in design and development. Here we see one of the tradeoffs that continually beset program managers; with fixed development budgets, they have to choose where to spend their money. They must allocate available resources into improving reliability, lowering production cost, or raising performance.

Time and Experience

Much past research has emphasized the value of time and experience as critical elements in effective and efficient weapons acquisition. Prescriptions emanating from this research have called for greater use of equipment prototypes, separation of subsystem development from platform development, "maturational" development phases, low-rate initial production ("phased acquisition"), and higher priority to upgrading fielded systems rather than to wholly new designs and developments for improved performance.[3] Time and experience are central to attaining reliability goals. It takes time to detect and analyze reliability problems, it takes time to correct the deficiencies through

---

[3] Much of this research is reviewed in Rich and Dews, 1986.

redesign and installation of improved components, and it takes more time to evaluate the changes themselves.

There are three reasons for the necessary use of time and experience in obtaining improved reliability.[4] (1) In the development of a new component, there cannot be complete *a priori* knowledge of that component's failure modes. (2) In the integration of known components into new arrays, new failure modes are introduced that cannot be completely predicted. And (3), these failure modes may be identified only through long test programs or operational experience.

Key policy questions concerning the acquisition of experience are, When in the development-deployment process is the required experience best obtained? and, Are there alternative development strategies that reduce the volume of testing and shorten the time required to produce reliable equipment? The answer to the first question appears to depend on the specifics of a program. Arguments and evidence can be found to support emphasis on almost every stage from initial design to post-deployment. Some claim that it is useless and wasteful to consider reliability issues in early design or prototype tests because the conditions of future use will deviate radically from early conceptions and experience. Others hold that it is much cheaper to obtain and use information early in a program when designs and hardware are flexible and investments are still fairly small.

Answers to the second question on alternative strategies to reduce the need for and cost of testing and experience are clearer. One approach calls for the development and maturation of proven component building blocks for integration into complex systems through standard interfaces.[5] Another strategy would use low-rate initial production to obtain field experience that could be incorporated into designs before full-rate production.[6] The use of system upgrades constrains the number of new and unproven elements in a system and thus reduces the amount of testing needed to generate a desired level of reliability. Reliability improvement programs are essentially system upgrades that focus on reliability rather than on operational performance. Finally, although doubt remains about the use of prototypes to resolve reliability questions, early experience gained through test and operation of prototypes reduces other kinds of technical

---

[4] This analysis is taken from McIver, Robinson, and Shulman, 1974, p. 1. Although the research was based on aircraft electronics, the conclusions are generalizable to other technologies and systems.

[5] Ibid.

[6] Rich and Drezner, 1982; and Lee, 1983.

uncertainty and helps freeze designs at earlier phases of development, thus making subsequent reliability issues easier to deal with.

As we proceed through the analysis of the cases, we shall draw attention to the use of the different approaches toward improving reliability. It is not sufficient to spend the money and expect greater reliability to follow. Resources must be directed in a managed, strategic manner, although the optimum method for any specific program remains a question.

RELATIONSHIPS AMONG RESOURCES AND RELIABILITY

Explicit consideration of the relationships among resource commitments, design decisions, and reliability will help clarify the policies available to decisionmakers as well as the constraints imposed on them by nature, budgets, and organizational processes. Research, development, and test resources can be devoted to various purposes. Resources can be directed at enhancing operational performance, reducing production costs, or improving reliability. Critical choices must be made; priorities must be assigned to the several program goals. Achievable design goals depend on the interacting technical decisions and on development and procurement budgets.

The central issue addressed here concerns the nature of these interactions. However, achieved (versus design) levels of performance and reliability often depend on matters beyond the control—but not beyond the cognizance—of requirements writers and developers. The numbers, quality, training, and experience of operating and maintenance personnel, in conjunction with the design characteristics of the equipment, will determine the actual ability of equipment to perform its missions. Design choices and resource commitments made years earlier will influence operational performance, fielded reliability, availability, and support costs. Decisions during development to emphasize high nominal levels of performance may result in lower achieved levels if the requirements for trained maintenance personnel are unmet. A Navy study found that on ships with complex, high pressure, 1200 pounds per square inch (psi) propulsion plants, downtime varied inversely with the quality of the boiler technicians as measured by standarized tests, training, grade, and years of experience. In contrast, the reliability of less complex, low-pressure, 600 psi boilers did not depend on the crew characteristics, but only on crew size.[7] Similar results were obtained when comparisons were made between guns and missiles, and between more complex and less complex fire-control systems: For the more advanced equipment to operate effectively, higher levels of support crew skills were necessary.

Technology may be able to reduce the importance of crew qualities. The M-1 tank, for example, contains more capable equipment than its M-60 predecessor. The M-1 has a laser rangefinder, solid-state ballistic computer, and stabilization system, to note just a few of the differences. In gunnery tests, the so-called "mental category" (as measured by Armed Forces Qualification Test—AFQT) of the tank commander and gunner had major effects on gunnery scores for the older technology M-60, but much less on the M-1. The smartest (Category I) tankers did 75 percent better in the M-60 than Category IV crews, whereas on the M-1, the Category I crews performed only 19 percent better. Improvements across tank models were equally dramatic. Category IV crews performed 84 percent better on the M-1 than on the M-60; technology improved the scores of the brighter Category 1 crews, however, by only 25 percent.[8] Unfortunately, we do not have the data to determine whether the technological gains attained in gunnery carry through to maintenance and whether the need for highly skilled crews has been transferred from the gunner to the turret mechanic.

To the extent that technology development occurs outside specific programs, its contribution to investment cost will go undocumented in program-oriented analyses and will produce an underestimate of the costs of higher reliability. Design and development resources, however, should be recognized in program costs. Similarly, higher quality materials and more stringent specifications are likely to influence production costs in an identifiable manner.

Measuring the effects on reliability from performance tradeoffs is problematic. Such decisions are often buried in the original specification of a new system. Deliberate, policy-initiated performance reductions in operational equipment are more visible because they represent a change in the status quo. The most difficult kind of effect to observe is the inadvertent operational degradation of performance brought about by the conjunction of unreliable equipment and inadequate maintenance. In some types of

---

[7] Sherman and Horowitz, 1979.

[8] Scribner et al., 1986.

equipment—advanced avionics, for example—this performance tradeoff is insidious and often unmeasured.

Capturing the cost of producing improved reliability through time and experience can involve complex calculations. The essential tradeoff that must be evaluated involves the benefits of acquiring unreliable systems earlier in a program (and the cost of retrofitting changes at a later date) versus more reliable systems later (and a lower cost of incorporating desired changes). An additional complicating factor is that extended testing during development is paid for from the development budget, whereas increased knowledge gained from operational experience is free (to a first approximation). However, undetected failure modes in deployed systems may sideline an entire fleet and cause equipment to become unavailable to perform its mission while imposing burdens on the maintenance and logistics systems. A failure during a development test produces none of these effects. In this study, we capture reliability-related costs to the extent that they occur in a programmatic context.

Because of the interactions among the cost-reliability relationships and their ties to reliability improvement strategies, the systems developers' job is complex, and perhaps even impossible. Consider, for example, the following scenario. Congress establishes development and procurement budget limits. Military commands specify requirements for performance and reliability. Program managers generally, because of the incentives acting on them, place priority on mission performance and direct development resource toward that goal. The services' personnel management and training system, together with the overall ability of the military to attract qualified people, establish the characteristics, quality, and numbers of weapon operators and support crews. Given this array of external or exogenous forces acting on the acquisition and deployment of new equipment, it may not be possible to achieve desired levels of operational availability, reliability, and support costs. Something clearly has to change to obtain these goals: development budgets, production costs, or operational performance—and priorities.

The totality of these relationships imposes limits on the achievements of policies, decisions, budgets, or technology. The degrees of freedom are limited. We began by looking at reliability, but the interactions with other forces imply that reliability is strongly tied to budgets, requirements, the weapon acquisition process, logistics, and the personnel system. The following sections take a look into this system, beginning with the central issue—the cost of reliability.

III. THE COST OF IMPROVING RELIABILITY

THE SEARCH FOR DATA

Information on the cost of reliability was obtained from a literature review of commercial and military analyses and experience, interviews with reliability executives and specialists in industry and the defense establishment, and intensive case studies of several military systems. The literature comprised theoretical articles on the cause and measure of reliability; engineering studies on failure modes; statistical studies on Weibull distributions and other statistical methods of describing, measuring, and accounting for failures; systems analyses for optimizing performance by maximizing the return to investment in reliability; and scores of articles on the need for U.S. industry to meet the Japanese reliability challenge. But there were no data explicitly related to the cost of reliability.

In the meantime, we conducted interviews with Defense Department offices concerned with reliability issues. Although we found little of immediate use, these managers and promoters of improved reliability provided valuable insights into the details of reaching reliability goals.

We next turned to industry executives charged with "meeting the Japanese reliability challenge."[1] Here we received important clues on the functional form of the reliability-cost equation. For example, a conversation with a vice-president responsible for reliability in one of the "big three" U.S. automobile companies produced one such clue. This company had been making a point of increased reliability in its advertising and had also increased the duration and total mileage covered by its warranty. The executive confirmed that reliability had increased, that it was not just an advertising gimmick, and that they were making money on their longer warranty period because reliability had increased more than their guarantees. We then asked him the key question: How much did this improvement cost? He replied, "That's a good question. I wish I knew the answer; my board of directors wishes that I knew the answer. But it was a competitive necessity that we do it, and we had to do it at any price." We were discouraged by this reply but then began to analyze the response. In cars retailing from

---

[1] Review of the commercial literature and interviews with company executives was performed by Ted Shi, a participant in RAND's Graduate Student Summer Program.

$8,000 to $20,000, the cost of reliability certainly could not be as much as $5,000. But could it be as much as $500 or $1,000? The fact that the executive did not know the cost suggested that in this cost-conscious, competitive industry, it could not have been very large. At this point, we again called the executive to probe specific estimates.

"Could reliability improvement have cost $5,000 per car?"

"Absolutely not! That would have driven us out of the automobile market." That dealt with the troubling notion (to an economist) of meeting competition at any price.

"Could it have cost $1,000?"

"No, I wouldn't have my job if it had."

"$500?"

"Unlikely."

"$100?"

"Could be."

"$37?"

"It's possible."

This conversation suggested that considerable improvements in reliability could be produced at quite low cost. A cost-of-reliability curve possessing a flat portion followed by a steeply rising sector is consistent with much of the evidence we have observed. Figure 1 illustrates such a curve. U.S. car manufacturers were probably in the region of point A in Fig. 1, whereas Japanese competitors had discovered that for a modest investment, reliability could be doubled to point B. Competitive forces then required the U.S. companies to move to point B also. The fact that automobile companies do not offer 150,000 mile warranties indicates that moving from B to C would require an investment so large as to make the car's price unattractive to buyers.

The shape of the curve in Fig. 1 is also consistent with information provided by several automobile tire manufacturers. A 40,000 mile tire is only 15 percent more costly to produce than a 20,000 mile tire; and a 20,000 mile tire costs only a few percent more than a 15,000 mile product. However a 75,000 mile tire would be several times the cost of one with a 40,000 mile guarantee, and the companies are not sure they could produce a 100,000 mile tire. The 10,000 to 40,000 mile tires are in the flat part of the reliability-cost curve; the 75,000 and 100,000 models would be in the upper right-hand tail, and the technology to produce a minimal 15,000 mile tire is considerably more expensive than anything with lower performance.

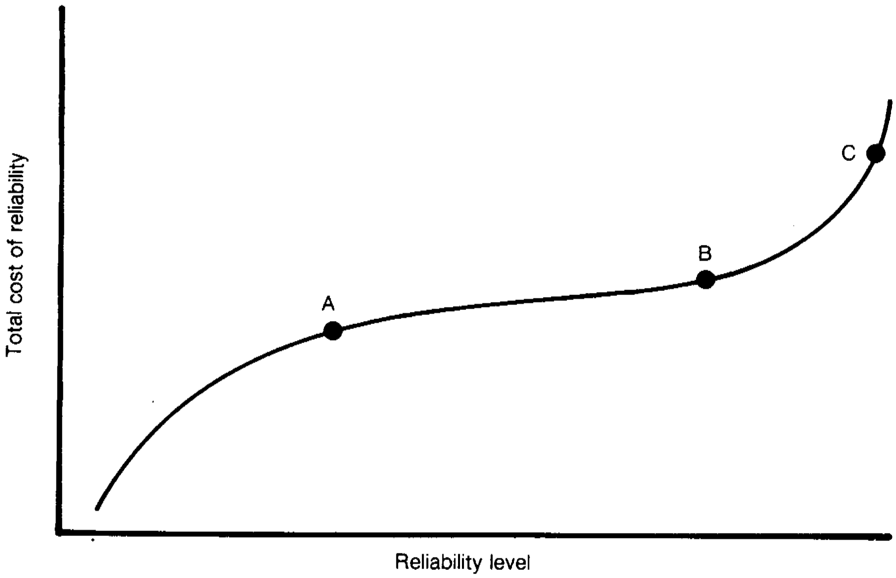

Fig. 1—The general shape of the reliability-cost curve

Although most military systems are considerably more complex than automobiles or tires, we suspect that the reliability-cost curves for military systems possess similar shapes. Because reliability has not generally been accorded high priority by the developers, many military systems are at present in the region of point A in Fig. 1. However, the data show some classes of systems, spacecraft for example, to be in the neighborhood of C—very high reliability levels purchased at very high costs.

This curve is not static; its position and shape during development and over time are affected by program choices, technology, and experience. A demand for greater operational performance, increased system complexity, and greater interactions of subsystems can move the curve up (a given level of reliability will cost more to obtain). Improved technology can lower the curve as it becomes less expensive to attain higher reliability levels. Accumulated system testing and operating experience will shift the curve downward as failure modes become evident. Therefore, a reliability strategy is not simply a matter of choosing an optimum point on this curve, but rather picking from a family of curves and making resource commitments as the curve itself moves with experience.

WEAPON SYSTEMS EXPERIENCE: THE COST OF RELIABILITY

Table 2 presents the reliability increases and their costs as extracted from our case studies.[2] Six of the case studies were performed in the course of the present study, three cases were taken from research performed elsewhere. The top part of the table presents the evidence from the reliability improvement programs, whereas the bottom half shows new systems experience. The cases span the range of technologies and military services: an Army transport helicopter, a Navy fighter aircraft, an Air Force jet engine, a ship gun system, inertial guidance, spacecraft, electronics, and a group of miscellaneous Navy systems. Perhaps the most important point Table 2 makes is that substantial reliability improvements are possible when that was a goal of the users and developers. Across technologies and phases of development and deployment, up to fifteen-fold increases are shown. Except for spacecraft, the original reliability level does not appear to constrain the possible improvements. Total aircraft system MTBF of one hour was doubled to two hours, and inertial navigation system reliability of 1000 hours was also doubled to 2000 hours.

In absolute values, the cost of these improvements varied over a wide range. More than $1 billion (1985 prices) was spent on the F100 engine Component Improvement Program (CIP) and almost a half billion dollars on the Minuteman guidance system. (Both of these systems had been rushed into production despite their immaturity.) In contrast, as little as $15 million was spent on improving the Phalanx Close-in Weapon System.

When we look at spacecraft, which show exceptional reliability relative to the typical military system, there is no discernible relation between investments in reliability and measurable outcomes. In this case, we are far to the upper right on the curve in Fig. 1 with MTBF values of 8,000–40,000 hours. Interestingly, the proportion of total program costs spent on spacecraft reliability is similar to that spent on the early Minuteman guidance system with a reliability level approaching that of spacecraft.

The data on the effect of reliability improvement on unit production costs show that, in most cases, production cost changes were zero. Indeed, in one case examined in detail, the F100 engine CIP, the production unit cost changes were negative—the engine became less costly to produce as a result of the CIP changes. For the Navy's F-18 fighter aircraft, we estimated a small production cost effect of 1.6–2.6 percent on the basis of

---

[2] These cases are described in detail in the appendixes to this report.

Table 2

RELIABILITY IMPROVEMENTS AND COSTS

Case	Reliability Measure	Reliability		Improved Reliability Cost (million 1985$)	
		Original	Improved	R&D	Production
		Reliability Improvement Programs			
CH-47D helicopter conversion	MTBF MTBMAF MTBSOF	2.08 13.9 .80	3.85 26.5 1.71	0 to 242	0 to 1.84
F100 jet engine CIP program	UER(lot) UER(fleet)	12.6 15.1	1.5 5.8	1177	0
F100 CIP total investment for tasks active in 1985	UER(fleet) MTBF Mishaps	4.4 57 2.28	2.5 60 1.79	120	0
Phalanx gun, radar, fire control	MTBF MTBPR A	47 78 .243	137 171 .50	15	0
Minuteman guidance	MTBF	600	9000–10000	427	0
Carousel inertial nav. system	MTBF	100 5–600 1000	5–600 1000 2000	78 38 ?	0 0 ?
		New Systems Developments			
F-18 fighter aircraft	MTBF	.8–.95	1.75–1.90	155[a] 250[b] 405[c]	.31–.48 0
19 misc Navy systems	"Satisfactory" vs. "Unsatisfactory"			47%	
Navy electronic systems	MTBF per part (million hrs.)	1.64	5.55	28[d]	
Duane models[e]	Failure rate @11000 & 14000 test hours	2.53 .39	2.35 .35		
Space craft	Achieved life(yrs)	1–5	1–5		

[a] Airframe costs.
[b] Engine costs.
[c] Total system.
[d] Difference in reliability cost per part (1975$).
[e] Top line is for "low" reliability efforts, bottom line represents "high" reliability effort.

NOTES: MTBF = mean time before failure.
MTBMAF = mean time before mission affecting failure.
MTBSOF = mean time before system operation failure.
UER = unscheduled engine removal per 1000 hours.
Mishaps = mishaps per 100,000 hours.
MTBPR = mean time before parts replacement.
A = operational availability.
Times before failure in hours, except where noted.
Inverse of UER and Mishaps used to calculate reliability improvement.

NOTES: MMHYFH = Maintenance Manhours per Flight Hour.
CIP = Component Improvement Program.

greater weight of the aircraft attributed by the developers to reliability. In some cases, possible production cost increases may have been compensated for by cost reductions arising from learning curve effects, or by contractors absorbing additional costs in reduced profits. Apparently, when reliability is a high-priority design goal—either in a new program or in a post-development reliability improvement program—the bulk of the cost effects are in nonrecurring investments rather than in recurring production costs.

Reliability investments and reliability improvements shown in Table 2 exhibit considerable variability. To address the issue of whether there is any regularity in the cost of reliability, we attempted to homogenize the data in a crude fashion by comparing the percentage changes in RDT&E investments attributable to reliability with the percentage change in reliability itself. These changes are shown in Table 3. The correlation between the logs of these variables is 0.64. A least-squares line drawn through the observations has a slope of 0.52.[3] This slope coefficient indicates increasing

Table 3

RETURNS TO RELIABILITY INVESTMENT
(Percent)

Case	Change in Reliability[a]	Increase in RDT&E Expenditures
Reliability Improvement Programs		
CH-47D	114	0–70
F100	740[b]	40
F100	5[b]	3.5
Phalanx	191	5.4
Minuteman	1400–1500	50
Carousel IV	400–500	180
Carousel IV	67–100	80
New Systems Developments		
F-18	100	12
Naval electronics	228	76

[a] Mean time before failure, unless otherwise noted.

[b] Mean time before engine removal.

---

[3] We considered both "conservative" and "optimistic" figures from Table 2: the conservative estimates used the smallest value of reliability increase and the largest cost increase, and the reverse of this rule was applied for the optimistic estimate (except that an average CH-47 cost increase of 35 percent was used instead of zero). The alternative procedures only affected the third digit of the estimated coefficients.

returns to reliability investments. Thus, a 10 percent increase in reliability would cost 5 percent more in total R&D expenditures, whereas a doubling of reliability (100 percent increase) would cost only 20 percent more in development; a five-fold reliability gain would require a 50 percent increase in original development costs.

We are able to use the data from Table 3 to ask whether systems that exhibited acute reliability problems (because of rushed, immature production) were cheaper or more expensive to improve than other systems. The F100 CIP effort, Phalanx, Minuteman guidance, and Carousel first-phase improvement met the "acute" criterion. Three of these have lower costs than predicted by the regression equation, one has higher costs, and one was as predicted. Three of the nonacute observations have higher costs, one lower, and one is on the line. This evidence provides weak support for the contention that those programs with serious reliability shortcomings are somewhat less costly to improve than are more "ordinary" programs. Compared with the seven product improvements, one of the "new" developments is higher than expected, and one lower.

What can we make of these observations? First we note the downward cost bias existing in the selection process for reliability improvement programs—seven of the nine examples. Second, one of the new systems—the F-18—benefited from the improved technology developed outside of that aircraft program. Third, notwithstanding the above points, the average level of reliability in the sample exhibited almost a five-fold improvement, which required a substantial 50 percent increase in development costs on the average. Suppose that the sample biases doubled the reliability improvement that could be expected from a "random" program. If such were the case, the same investment would only produce approximately a 2.5 times reliability improvment—a significant but less impressive return on investment. However, despite the potential biases, the data suggest that considerable improvements are sometimes possible, that the greater the improvement the more costly the necessary investment, and that the investment appears to rise at a lower proportional rate than the improvements. These data do not allow much greater precision in drawing conclusions.

RELIABILITY-PERFORMANCE TRADEOFFS

Reliability specialists cite the importance of stress reduction on parts, systems, and structures as one of the chief methods for improving reliability. In most cases, reducing stress is equivalent to lowering operational performance. Reducing the temperature, load, or voltage on a part will often increase its life and reduce the probability of failure; increasing the size of a fastener, the thickness of a spar, the number of cooling fins in a radar, or the number of stringers and supports in a fuselage will similarly reduce the stresses on the parts. All of these measures will add weight or otherwise reduce system performance.

An example of the reliability and maintainability effects from the choice of very high performance is the Lockheed SR-71 Mach 3 reconnaissance aircraft. The maintenance resources and skills required to keep this aircraft flying are estimated to be in the hundreds of maintenance manhours per flight hour compared with the 25–75 maintenance manhours per flight hour of other U.S. Air Force and Navy fighter aircraft.[4] The extreme operating environment and the unique structure, propulsion, and subsystems necessary to perform the demanding mission impose burdens on every aspect of maintaining reliability and servicing the aircraft.

Soviet fighter aircraft require only about 10–20 percent of the maintenance of U.S. fighters.[5] The definition of the mission, the "requirement," makes the difference: The typical Soviet aircraft requirement, while just as demanding as U.S. specifications in the main mission, permits performance to decline at those parts of the envelope that are less critical to predicted combat needs. An American helicopter reliability engineer working on a U.S. Army combat helicopter design graphically described to us the different U.S. and Soviet approaches. A problem in helicopter reliability arises from "auto-fretting," whereby rivets under high pressure repeatedly weld themselves to the material they are fastening and then break the weld because of vibrations in the helicopter structure. Eventually, this repeated action leads to failure of the fastener. The U.S. designers were trying to solve this persistent problem by engineering out the vibrations through fine-tuning the frequency modes of the structure. While describing this approach, the reliability engineer noted how the problem had been solved on an advanced Soviet helicopter he had examined. "The Russians simply used fasteners that were one size larger in each application than we had used." When I asked why the U.S. company did not use this same seemingly reasonable and exceedingly simple solution he replied that it

---

[4] "SR-71 Imposes Burden on Maintenance Units," *Aviation Week and Space Technology*, May 18, 1981, p. 105.

[5] Planned times before overhaul on Soviet aircraft and operational lifetimes are typically lower than on comparable U.S. equipment. Overall evaluations are therefore more complex and uncertain than these simple comparisons may imply.

would have added 30 pounds of weight to the structure; the designers were fighting to strip off every pound that they could. The U.S. engineers had chosen to solve their problem through greater investment in development resources, not because they were ignorant of simpler solutions but because their customer had placed top priority on maximizing payload and minimizing aircraft weight.

Shifting priorities in recent years have encouraged weapon users as well as designers to reconsider their preferred solutions. General Alton Slay, Commander of the U.S. Air Force Systems Command, noted in 1979 that among the lessons learned from engine programs of the preceding decade was the necessity to "adjust our sights down from the performance extreme that we set in the F100 engine toward more durability."[6] General Slay noted that he had given instructions on new engine developments to "turn the wick down and get to lower the total stress."[7] Such tradeoffs between performance and reliability have been more clearly recognized as reliability has become more of an issue. The Air Force now is more willing to consider adding weight or otherwise derating equipment. Although perhaps not going as far as the Russians, we have seen definite movement in this direction.

This philosophy was implemented in the Navy's F404 engine developed by General Electric. The engine designers used a set of guidelines to achieve reliability and performance goals while reducing program risks. The guidelines included "backing off from ultimate performance levels and keeping the engine design simple." One of the design tradeoffs was the choice of a single-stage, low pressure turbine rather than a two-stage turbine. "Although the two-stage turbine would have provided somewhat better performance, it would have done so at a substantial penalty to production cost, weight, maintainability, and reliability."[8] The choice contributed greatly to design simplicity. In addition, the turbine inlet temperatures on the F404 engine are lower than on the F100, despite the advances in metallurgy in the years between these two developments.

The users of the Pratt & Whitney F100 engine in the U.S. Air Force Tactical Air Command confronted tradeoffs between performance and reliability after the engine had been fielded. In this case, the Air Force accepted a lower turbine inlet temperature and consequent loss of maximum thrust of about 3 percent to double the life of the turbine

---

[6] U.S. Senate, 1975, p. 5, statement of General Alton D. Slay.
[7] Ibid., p. 63.
[8] Rapp, 1982, p.2.

blades. The using commands maintained these reductions until improved turbine blade materials could be developed and retrofitted into fielded engines several years later.

The Navy's high priority on reliability during F-18 development led to an estimated 500 lb weight increase over earlier design standards. The F-18 reliability in fleet operations was subsequently more than twice that of previous aircraft.

An exceptional example of designing for simplicity and reliability while not giving up critical performance capabilities was the R-11 engine on the Soviet Union's MiG-21 fighter.[9] U.S. engineers conducted a detailed comparison between the Russian engine and the U.S. J-79 fighter engine of the same vintage and gross level of performance. Although the Soviet engine was acknowledged to be an outstanding design, the philosophy and approach on which it was based were quite similar to those of Soviet engines and Soviet weapons more generally. The Soviet engine had 90 percent fewer parts than the American engine (2,500 versus 22,500). Standard gauge materials throughout increased weight but reduced materials costs. Lower turbine inlet temperatures allowed use of conventional materials. The raw material cost per pound was estimated to be 60 percent less than for the U.S. engine. Open clearances reduced manufacturing costs and resulted in some teststand performance degradation, but these levels did not degrade further in operations, as was the case for the more precisely manufactured U.S. engine. Although the Soviet design was highly innovative in aerodynamics design and overall concept, it was conservative in execution. Parts were stressed to about half the U.S. levels. Estimated production costs (using U.S. prices and wages, but duplicating the Soviet manufacturing process) of the MiG-21 engine was roughly one-third that of the American. The analysts judged that the R-11 could have been produced with U.S. technology of the 1930s. Mean time before failure was several times longer, and the Russian engine required only 8 percent of the maintenance manhours per flight hour of the American. However, the time before overhaul for the Soviet engine was only one-third of that specified for its U.S. counterpart. On balance, the Soviet design philosophy ensured highly reliable operation without field repair, but for a shorter time before the equipment had to be sent for depot overhaul. The Russian designers had optimized the engine for performance around the two most probable design points expected in fighter combat. Away from these points, engine performance degraded relative to the U.S. design.

---

[9] Central Intelligence Agency, 1986, p. 25.

Choices and tradeoffs tend to be consistent with the goals of the period and the place. Recently, the Russians have attempted to gain more performance while (perhaps—the evidence is not clear) giving up some reliability. U.S. force planners have come to recognize a new priority for reliability, paying for it with some modest loss of potential operational performance and by greater emphasis and resources in RDT&E.

IV. THE BENEFITS OF RELIABILITY

Reliability is valued not only directly but also because it vitally affects other things that the military services care about: maintenance time and manpower, spares consumption and inventories, the amount of time that a system is operationally available, the logistics load imposed on a military operation, and the life-cycle costs of operating equipment over periods of decades. Table 4 displays the benefits associated with the reliability changes derived from our case studies. Reliability improvement programs are shown in the top part of the table, and other examples in the lower part.

The CH-47D helicopter required 28 percent less maintenance per flying hour than its predecessor model, the CH-47C. Higher reliability and lower maintenance reduced the annual cost of maintenance labor and parts consumption by more than $100,000 per aircraft.

All of the design changes being worked on during a single year's component improvement program for the F100 engine reduced maintenance manhours per flight hour by 15 percent and support cost by 33 percent. The ten-year CIP program reduced total life-cycle cost by $7 billion, equivalent to $600 savings per flight hour.

Operational availability of the Minuteman guidance system climbed dramatically from 28 to 96 percent while yielding a maintenance saving of $1.5 billion (in 1965 prices). The Minuteman case provides a good example of how effective a high-priority, focused reliability improvement program can be. At a time when reliability was generally dominated by operational performance requirements, the Minuteman high reliability program was successful because a crisis in the operability of the U.S. strategic nuclear forces created the high-level attention and resources needed to carry out the job. The Minuteman guidance system effort demonstrated the technical feasibility of setting high standards and obtaining commensurate results, at a cost that was fully half that of the original development.

The Phalanx case is informative because it shows how fairly small expenditures in improving a design can yield large payoffs in reliability. Parts demand fell by 20 percent, maintenance by 82 percent, and operational availability more than doubled.

Since reliability programs of the type just noted were chosen because they were expected to have high returns, the level of improvements may be an over-optimistic

Table 4

BENEFITS OF IMPROVED RELIABILITY

Case	Improvement Category	Original Level	Improvement	Percent Change
	Reliability Improvement Programs			
CH-47D	Maintenance, unscheduled (MMH/FH)	5.13	2.25	44
	Maintenance, scheduled (MMH/FH)	4.25	0.0	0
	Maintenance, total (MMH/FH)	9.38	2.25	24
	Maintenance, labor (1985 $/AC/yr)	272,000	59,000	22
	Parts consumption (1985 $/AC/yr)	291,000	49,000	17
F100: 1-year CIP	Maintenance (MMH/FH)	1.72	.26	15
	Support cost (1985 $/FH)	530	175	33
10-year CIP	Life-cycle cost (1985 $/FH)		600	
Minuteman guidance	Availability (%)	28	68	243
	Parts and maintenance (million 1965 $)		1500	
Carousel inertial navigation	Life-cycle costs			
	MTBF improvement: 100 hrs to 200 hrs			37
	MTBF improvement: 200 hrs to 400 hrs			30
	MTBF improvement: 400 hrs to 500 hrs			5
Phalanx	Parts demand			20
	Maintenance (MMH/FH)	.34	.28	82
	Availability (%)	23	27	117
	Other Methods			
Phalanx	Spares inventory: (million $)[a]			
	double reliability, 20 parts	120	20	17
	double reliability, 30 parts	120	24	20
	double reliability, all parts	120	50	42
	Availability (%):[b]			
	double reliability, 120 parts	70	4	5.7
	double reliability, 130 parts	70	6	8.6
	double reliability, all parts	70	20	28.6
LAMPS MK III	Spares inventory (million $):[c]			
	double reliability, 20 parts	4.0	0.4	10.0
	double reliability, 30 parts	4.0	0.6	12.5
	double reliability, all parts	4.0	1.5	37.5
	Availability (%):[d]			
	double reliability, 20 parts	70	4	5.7
	double reliability, 30 parts	70	6	8.6
	double reliability, all parts	70	14	20.0
F-16	Investment per aircraft (million $):[e]			
double reliability of only engines and fire-control	Engines and engine modules	1.25	.54	43
	Recoverable peacetime operating spares	1.43	.31	22
	Replacements for condemned parts	1.33	.48	36
	Depot level repair of components	1.87	.48	34
	Total	5.88	1.81	34
double reliability of all subsystems	Sortie rate			17
	or			
	Maintenance manpower			9

[a] Assumes two echelons of stocks, 305 ships, 70 percent availability (see App. E).

[b] Assumes two echelons of stocks, 305 ships, spares inventory fixed at $120 million (see App. E).

[c] Assumes single echelon of stocks, ship availability held at 70 percent (see App. F).

[d] Assumes single echelon of stocks, ship spares inventory held at $4 million (see App. F).

[e] Assumes steady-state level of 630 aircraft (see Table B.1).

prediction of what could be expected on a more typical program. Additional examples, however, yield similar results.

Simulations of the Phalanx system showed that improving the reliability of only 20 parts (out of more than 1000) could reduce spares inventories by 17 percent, or $20 million on a fleet-wide basis. The same improvement could alternatively increase Phalanx availability by almost 6 percent.

The same analysis performed for the LAMPS antisubmarine warfare system yielded similar results. Doubling the reliability of the 20 least reliable parts could save 10 percent of shipboard spares or increase operational availability by about 6 percent.

A twofold reliability increase of just the engine and fire-control system of the F-16 could yield estimated savings in spares investment and depot repair costs of $2 million per aircraft—equivalent to saving one-third of the total required spending on these items, or almost 12 percent of the cost of the airplane. A separate calculation indicated that doubling the reliability of the entire aircraft could yield either a 17 percent higher sortie rate or a 9 percent reduction in maintenance manning, without losing the benefits of the savings in spares investment.

We can combine the reliability improvements and the costs from Table 2 with the benefits from Table 4. Consider, for example, the F-18 and F-16 cases.[1] A doubling of aircraft reliability was estimated to produce an increased sortie rate of 17 percent. To a first approximation, this is equivalent to the same size increase in force structure. For the F-18 planned fleet size of 1337 aircraft, the reliability improvement can be considered equal to more than 225 aircraft added to the inventory. At 1984 F-18 costs, this force increment would cost more than $4 billion. The high end of our estimates of the cost of this improvement is a little more than $1 billion. In terms of added military capability, the return is worth the initial investment. If we confine our attention only to the gains from the improvement in reliability calculated for the F-16, we find the savings from reduced investments, consumption, and repair of spares to be $1.1 billion. Noting that F-18 savings in these categories are likely to be larger because of the larger size and complexity of the aircraft, and that the F-16 analysis considered only a limited number of support categories, we suspect that total savings would be several times greater than the $1 billion investment. By either calculation—evaluation of force additions or savings in spares—the returns outweigh the costs by large margins.

---

[1] Because the F-18 actually demonstrated a 100 percent reliability improvement over earlier Navy aircraft—the same value assumed for the F-16 analysis—we amalgamate their results here.

For comparison with the F-18/F-16 analysis in which the program costs were fairly small, we can consider the CH-47D helicopter conversion and look only at the upper bound estimate of the reliability improvement costs. (The lower bound estimate was zero.) The upper bound cost per aircraft was $2.37 million, or $278,000 per year when amortized at a 10 percent discount over the planned 20-year life of the modernized aircraft. The savings in maintenance manpower and parts consumption costs were estimated at about $108,000 per aircraft per year. Spares investment savings, using the F-16 as an analog, could be at least $1 million per aircraft, or $117,000 per year. This figure is probably on the conservative side because only engine and fire-control improvements were analyzed on the F-16. The sum of these conservatively generated estimates produce annual savings of $225,000, which falls short of the upper bound cost estimate by about 20 percent. Consideration of other benefits such as spares investment savings for the total aircraft and increased flying hours available from higher reliability would probably bring the savings up to the cost of reliability improvement even when measured at the high end of the probable range of cost estimates.

Turning from financial returns to the ability to perform a military mission, we find that availability rates increase by about one-sixth to one-half as much as the rate of reliability improvement. The sortie rate for the F-16 was estimated to increase 17 percent of the rate of the reliability increase; the Minuteman guidance availability rose by 16 percent of the reliability growth; on the Phalanx weapon system, availability grew at an estimated 29 percent of the reliability gain of all parts and at an actual rate of 60 percent of measured improvement. These increases in mission capability are perhaps the most important returns to reliability.

For the past decade or so, U.S. military technologists have spoken of "force multipliers" in the guise of advanced technology and performance. A reliability increase that improves availability by 20 percent is also a force multiplier. Even in narrow financial terms, the growth in military capability, as measured by equivalent force size increments, is worth many times the cost of obtaining reliability in all the cases we have experienced. But there are other effects of higher reliability in addition to availability, maintenance manpower, and spares. Although we do not deal with them in this report, they bear mentioning because in many ways they are as important as the things we have been measuring. More reliable forces are more mobile as the volume of equipment, spares, and people required to support the systems is reduced. Soviet aircraft, for example, are notable for their ability to operate on an austere basis for periods of one to two weeks. Several squadrons of Soviet fighter, attack, and transport aircraft can move into an area with very little additional support where they can operate effectively until their logistics catches up with them. As potential enemies acquire improved long-range weapons, the support resources vulnerable to loss raise serious concern about the ability of U.S. forces to maintain effective fighting capabilities in both the short and long run. More reliable equipment is one of the solutions to these military problems.

V. RELIABILITY STRATEGIES

THE NEED FOR PRIORITY

Military equipment, like most other products, can be described by sets of attributes conveniently grouped into familiar categories: performance, cost, reliability, style, etc. For the most part, these are all desirable attributes.[1] Not only must choices be made across the entire list of attributes, but often this list itself will be shortened and constrained by technical limits.[2] The important point is that reliability is technically symmetrical with the other attributes of a product. To achieve more of it, something must be given up elsewhere—in development expenditures, production costs, performance, or time. Moreover, if the available degrees of freedom over desired attributes have already been used up by higher priority choices, it may not even be possible to make the tradeoffs; reliability will be a residual property, predetermined by earlier explicit or implicit choices.

Perhaps our most important observation is that reliability must be demanded to be received. In every case we examined where reliability strategies could be evaluated, priority for reliability was a necessary condition for the various strategies to be effective. Our cases and interviews indicate that if military buyers demonstrate a commitment to their requirement for greater reliability, they can get it. Defense companies in pursuit of their own self-interests are adept at reading the true preferences of weapon buyers. A defense firm's profits depend on its ability to get behind public rhetoric. Therefore the military buyer not only must state a demand for reliability, it has to specify, measure, test, demonstrate, and pay for it. Defense firms have to see that reliability has high priority through buyer commitments to development resources, testing time, and performance tradeoffs.

---

[1] Analytically, it is customary to assume that the measures of negatively valued properties, such as weight, are transformed so as to make their transformed values positively valued.

[2] For a more detailed exposition of the point, see Alexander and Mitchell, 1984, pp. 10–15.

SUMMARIZING THE EVIDENCE

In Sec. II we introduced five approaches for obtaining better reliability levels: (1) improved technology, (2) additional resources directed toward reliability in design and development, (3) performance tradeoffs, (4) higher quality, and (5) time and experience. We have summarized our evidence on the development, production, and tradeoff costs of higher reliability. Our conclusion on the relative efficacies of these different approaches in an overall strategy is easy to state. The evidence that we have reviewed does not support a single approach or combination as being most effective. Rather, the results appear to be specific to time period and technology. Turbine engines, for example, underwent a shift in the relative effectiveness of strategies when better knowledge of failure mechanisms and improved testing techniques permitted greater reliability gains to be made in design and development than in post-development efforts. Compared with earlier strategies, increased engine reliability is now attainable with considerably more development testing and a ten-fold reduction in post-deployment product improvement expenditures.

To show the variety of mechanisms and approaches used in achieving greater reliability, we summarize here the reliability strategies we observed in our case studies. (These cases are examined in greater detail in the appendixes.)

> F-18 Aircraft: Reliability was accorded high priority; the R&D resources devoted to reliability goals were greater than for typical programs; greater knowledge of failure mechanisms in electronics and engines allowed more effective design and test; explicit tradeoffs against performance enhanced reliability; early experience with prototype engines and greater volume of testing greatly contributed to obtaining reliability.

> CH-47D helicopter modernization: Over a ten-year period and more than a million operating hours, reliability tripled for the CH-47 models A, B, and C because of a combination of experience, design changes, product-improvement programs, and greater crew proficiency; in the CH-47D modernization, reliability was a central goal and reliability was doubled; experience, new technology, and better design concepts were used to produce the higher reliability; the evolutionary nature of the program reduced the uncertainties compared with a wholly new design; priority on reliability directed the use of technology, experience, and resources toward reliability rather than operational performance.

F100 engine: The engine was introduced in an immature state as performance and schedules dominated development and deployment; a reliability crisis stimulated Air Force priority, attention, and resources to reliability issues; the Component Improvement Program produced post-deployment reliability over a 10-year period; development of new materials and better knowledge of failure mechanisms were motivated by increased emphasis on reliability and contributed to reliability improvements.

Phalanx Close-in Weapon System: Serious shortcomings in equipment availability led to a high-priority reliability improvement program; fleet experience enabled substantial gains at fairly low cost.

Minuteman inertial guidance system: Minuteman deployment schedules dominated development decisions; the early guidance system was immature and unreliable; an availability crisis stimulated high-priority reliability improvement program; although much of the reliability improvement program could have been undertaken in development, operational experience helped to identify some failure modes.

Carousel inertial navigation system: Low operational reliability induced a high-priority reliability improvement program; although greater effort could have been made in development, project engineers believed that operational fleet experience was invaluable in correcting reliability problems.

Spacecraft: Intense program focus and resources expended on reliability in development produced highly reliable equipment.

Duane models: Increased testing produces greater reliability at an exponential rate varying from 0.3 to 0.6; larger exponents are associated with higher priority on reliability; priority appears to have a greater effect than test hours.

19 Naval systems: Programs that demonstrated "satisfactory" reliability outcomes spent more on reliability in development than "unsatisfactory" programs; "satisfactory" programs performed more reliability directed tasks; reliability activities were undertaken only when Navy buyer emphasized reliability goals and persevered in its demands.

All of the cases with higher levels of reliability demonstrated high priority for reliability. Similarly, in almost all of these cases, greater development resources were devoted to reliability than in some "standard" program. In some cases, experience gained from additional testing or experience was important; but in other cases, it was less relevant. We also found mixed results for performance tradeoffs and technology: Both were used, but not universally. There are no apparent differences in these results between product improvement programs and new developments. Curiously, higher production costs were only rarely related to increased reliability. Discussion of these results with design engineers elicited the universal comment that reliability designs are also often simpler to manufacture, hence less expensive to produce. Supporting this contention is a finding that the cheapest inertial navigation systems were also the most reliable, and that the most expensive systems were the least reliable.[3] Apparently, in the expensive systems, the drive for greater operational performance drove up costs and pushed down reliability.

In other research performed at RAND, for electronic equipment characterized by complex software and subsystem interfaces, high reliability of individual components and subsystems is apparently not sufficient to guarantee reliability of the system as a whole. Therefore, resources devoted to reliability during development need to be supplemented by testing and experience gained in an operational environment. In fact, it is the increased level of hardware reliability in electronics that has helped to unmask this problem, which appears to be growing in severity as electronics devices become more integrated and complex: systems can "fail" without any individual hardware element breaking or failing, at least in the traditional sense. Such problems can stem from several sources.

Software bugs clearly are behind many failures. A subtle kind of software deficiency is an incorrectly specified model by which the software programs attempt to mimic reality. If there is a discrepancy between the model and reality, the system will not perform as required, even though the hardware and the computer programs operate correctly.

A different kind of degradation occurs when the operating characteristics of system components drift from the design point: For example, timing sequences may deviate from their nominal values, or unforeseen signals with complex waveforms may

---

[3] Genet, 1972, pp. 9A–10A.

be generated by interconnected components. Although each of the components may be performing within design limits, their interconnections can create unpredicted system degradation, which may not even be defined as failures by the standard reliability data collection definitions.

Because of the growing incidence of such problems, some analysts are calling for an acquisition strategy that recognizes the difficulty of designing complex digital electronics with high reliability; such an acquisition strategy would plan for sufficient time to discover and correct such problems. However, because neither the calendar time nor the other resources for full maturation of complex electronics systems are usually available during a single weapon system development, an approach has been proposed that includes: (1) the use of standard interfaces between equipment, (2) the development of components that could be treated as reliable building blocks in the design of integrated systems, and (3) a maturational phase during which the reliability of the building blocks could be thoroughly established. In many ways, this strategy is similar to the one that has evolved for aircraft engines. Engines are now known to require more development time than the airframes into which they are placed. Hence, they are started earlier or existing engines are modified. Furthermore, investment in reliability improvements in one application enhances an engine's value to new users. The growing custom of designing aircraft to accept more than one engine type is equivalent to the standard interface being called for in electronics. Thus, the problems seen today in electronics are not different in kind from those seen and dealt with elsewhere.

VI. CONCLUSIONS

GENERAL CONCLUSIONS

The returns to the investments in reliability are both resource-saving and capability-enhancing. It often makes sense to promote reliability on grounds of both economic efficiency and improved war-fighting capabilities. Recognizing the limits on our data, we note the following: For the F-18 aircraft, a 4 percent increment to airframe development costs and 12 percent increase in the total system was associated with a doubling of reliability; 4 percent more of total development costs was the average margin between 19 "satisfactory" and "unsatisfactory" Navy systems; about 5 percent in post-development reliability improvement expenditures on the Phalanx helped to improve reliability by 100–200 percent, depending on the measure used. For other systems, the expenditures were considerably greater and the returns more dramatic. The Minuteman guidance system, for example, demonstrated a 1500 percent reliability improvement for post-development costs that were 50 percent of the original expenditures.

The priority and resources devoted to, for example, a doubling of reliability can also increase operational availability, sortie rate, or other capability measures by at least 15 percent and probably quite a bit more, with no change in recurring manufacturing cost. Reductions in maintenance manpower and spares improve the ability to deploy military force and reduce the vulnerability of the logistics support network.

In the examples we reviewed, increased reliability was obtainable in new developments or product improvements, immature products or veteran systems, or devices with low or high absolute levels of reliability. Only in spacecraft, in which the demand for reliable performance over extended periods is particularly high, do the returns seem to be unrelated to costs.

Over the past 10 years or so, the greater priority given to reliability has encouraged technological change related to reliability in design and engineering. These incentives have produced better knowledge of failure mechanisms, design techniques, testing methods, materials, and modeling of system architecture. Users have also been more willing to consider tradeoffs in performance to reduce component stress. Because of the advances in reliability technology, the cost of achieving greater levels of failure-free performance in the future should not rise as fast as it otherwise would. Although experience will demonstrate the feasibility limits—the knee of the cost-reliability curve—we would conjecture that higher reliability levels could be safely called for in the future than in the past, without jeopardizing other system goals.

The rising demand for more reliable military equipment has also generated questions about the returns to investments in reliability. Such questions can be answered only with adequate data on costs and benefits. This study was hampered by the necessity to develop our own data from original sources. In only one of the cases that we examined, the CH-47D modernization, was there an explicit attempt to collect "before" and "after" reliability information in a comparable format. But even in that case, later phases of data collection have been curtailed. We see a clear necessity to produce information for analytical comparative purposes on both reliability investments and outcomes, especially for those programs in which reliability is a top-priority objective. However, because of its natural self-interest in the results, the project office may not be the appropriate agency to collect such information, especially if the cost comes out of project budgets. Within the services, there are offices with reliability responsibilities. Usually these have been oriented toward developing engineering standards and analyses. Because so much reliability information is now passing through these offices and they are developing an expertise and visibility over all their service's developments, these offices appear to be good candidates to perform the data collection task.

The results of this study also point out the need for more analysis. A fruitful approach may be to examine a few products that are homogeneous within themselves but are representative of different technologies and classes of problems. Turbine engines and inertial navigation systems are two such examples. For each, there are several decades of experience, evolving technologies, shifting priorities and development strategies, and some comparable data. Analysis of such data would help to answer questions on the relative value of investments at different stages of development and deployment, the contributions of testing, the effects of policy changes, and the benefits of improved technology—questions that we could only touch on in this study.

THE CROSS-CUTTING NATURE OF RELIABILITY

Finally we note that investments in reliability tend to come early in a program, whereas the returns flow over a period of decades. Moreover, the investments are the responsibility of one set of organizations and the payoffs accrue to others. And the resources flow from different funding sources—appropriations, Congressional committees, and Defense Department budgets. This split over time and across organizations and budgets raises two problems. The first is the temptation to reduce front-end costs: to "sell" the system; to conserve scarce funds; to make tradeoffs across systems, missions, and services. The second problem is that reliability is often not seen as anyone's special responsibility, because everyone is affected. Unless "everyone" sees reliability as a high-priority item, the necessary commitment is unlikely. It is just such a change in the environment that we suspect is now raising the importance of reliability in defense acquisition.

Appendix A

THE NAVY F-18 AIRCRAFT DEVELOPMENT

The U.S. Navy initiated a package of policy measures intended to improve the reliability and maintainability (R&M) of its weapons and equipment in the early 1970s. These initiatives were part of a broader movement to improve force readiness and capability while making more efficient use of resources. The policy gave high priority—and funding—to improving R&M characteristics from the earliest concept of a new system through development, production, and fielding. The Navy's development and acquisition commands set up special offices to develop engineering, design, and test techniques and manufacturing processes to be applied by Navy system project offices and contractors. These offices were responsible for specifying and managing design and manufacturing standards. Naval aircraft developed under this regime were designated as "New Look" systems. The F-18 fighter and attack aircraft was one of these systems.

Approximately two-thirds of the F-18 subsystems were subject to New Look criteria, the rest having been designed earlier. In addition to giving high priority and attention to R&M the Navy prescribed several design specifications, tests, and technology approaches for the aircraft. For example, sneak circuit analysis, "derating" (lowering stress levels) of components, and high-temperature limits for electronics were called for.

One of the more important subsystems falling under the New Look philosophy of the F-18 was its turbofan engine, the General Electric F404. The priority ranking of engine attributes gave second and third places to reliability and maintainability, with the old favorites of performance and weight taking up the lowest positions. "Operational suitability," a broader measure of system usefulness than the traditional list of performance specifications, was the first priority item. Because of these priorities, tradeoffs were made that reduced thrust and increased fuel consumption somewhat, but that resulted in substantial improvements in cost, complexity, reliability, and maintainability. Examples included choice of a three-stage instead of a four-stage fan, and a single-stage turbine rather than two stages.[1] By such choices, the number of parts in the F404 is about 14,000 rather than the more than 30,000 in the F100, developed according to different priorities six to ten years earlier. Life-cycle cost was treated symmetrically with other performance goals.

The engineers on the F404 project considered that the evolution of the F404 from General Electric's earlier YJ101 prototype, which had already received 1500 hours of factory test and several hundred hours of flight test on the YF-17 prototype aircraft, was a primary advantage in designing a reliable engine.

> The advantage of the prototype program was that it allowed evaluation of a design which was not committed to production. It offered the engineers the opportunity to redesign the engine, to simplify it, to improve reliability and maintainability.... A delay between the YJ101 phase and the start of the F404 development allowed designers to take into account YJ101 operating experience, improvements in manufacturing processes, and materials development.... This delay allowed components to mature before ... full-scale development."[2]

The design and test criteria imposed by the Navy on engine and other subsystem developments, the more stringent and extensive testing that was demanded, and the management, engineering, and design resources committed to R&M were not without cost. The prime contractor of the F-18, McDonnell Douglas, estimated that the *additional* development costs attributable to New Look R&M practices amounted to about $100 million—and to an extra 500 pounds of aircraft structural weight. They derived the $100 million figure by identifying all the organizational and procedural changes emanating from the Navy's new policies.

That is a substantial sum to pay for reliability, and 500 pounds of additional weight can have a considerable influence on an aircraft's cost and performance. In what follows, we shall attempt to assess the costs and match them with the benefits they produce.

The total F-18 development program cost approximately $2.3 billion. Expenditures for increased reliability were therefore about 4 percent of the total development program. (In 1985 inflation-adjusted values, total development was $3.7 billion; and the cost of reliability improvement was $155 million.) At the planned

---

[1] Rapp, 1982, p. 2.

[2] Ibid., p. 4.

procurement quantity of 1377 aircraft, that comes to an investment of $113,000 per aircraft in 1985 dollars. The RDT&E investment in increased reliability represented about six-tenths of 1 percent of the $18.7 million production cost in 1985.

The 500 pound increase in structural weight is also not without cost. We estimated the increased manufacturing cost through the use of cost estimating relationships (CERs).[3] The 500 pound increase is about 2.6 percent of the airframe weight. Applying a CER-derived elasticity of cost with respect to weight of .75 to actual 1985 F-18 airframe costs yields a production cost increase of $310,000. An alternative method of arriving at the cost of the 500 pounds uses average aircraft experience instead of specific F-18 costs. A typical aircraft in the F-18's speed and airframe weight class would cost an average $480,000 more per aircraft (in 1985 dollars) for a 500 pound increase for the first 100 aircraft produced. Increased production costs resulting from the increased weight attributable to higher reliability thus could range from roughly 1.6 to 2.6 percent of total aircraft unit cost.

We estimated the extra cost of reliability in the F404 engine by assuming that the entire YJ101 engine prototype development funding of $79 million ($237 million in 1985 dollars) was applied toward maturing the design and enhancing reliability. The F404 engine underwent 14,000 hours of testing in full-scale development—about 3000 more hours than previous General Electric engines. We estimated that these differences would cost $12 million.[4] The sum of these figures yields a total estimate for increased reliability of about $249 million. F404 development costs through 1982, including the full cost of the YJ101 prototype, was $949 million (in 1985 dollars). The RDT&E cost of reliability was therefore approximately 26 percent of the total development effort.

Both the investment costs and the cost per 1337 aircraft are shown in Table A.1. The extra investment costs for reliability per aircraft amounts to about 3.3 to 4.1 percent of the 1985 F-18 production cost.

These development and production costs have yielded substantial returns in F-18 operation. Compared with earlier aircraft at similar points in their operational fleet experience, the F-18 is demonstrating about twice the reliability, as measured by mean flight hours before failure (MFHBF). Using the same reliability data systems for

---

[3] Large, Campbell, and Cates, 1976, p. 42.

[4] Estimates from an aircraft turbine engine producer suggested that durability testing of a new engine using one test engine would generate about 75 test hours per month, at a monthly cost of approximately $300,000 for an F404 size engine.

Table A.1

THE COST OF IMPROVED RELIABILITY IN THE F-18
FIGHTER AIRCRAFT
(1985 dollars)

Cost Element	Total Investment (million $)	Investment per Aircraft[a]
Total aircraft development[b]	3,700	2,767,400
Marginal cost of improved reliability	155	115,900
Aircraft production cost		18,700,000
Production cost effect of weight increase		310,000–480,000
Total engine development	949	709,800
Engine prototype	237	177,160
Marginal cost of additional engine testing	12	8,975

[a]Costs per aircraft based on procurement of 1337 aircraft.
[b]Does not include engine development.

comparing aircraft, the F-18 is demonstrating a fleet average of about 1.75 to 1.90 MFHBF five years after introduction, compared with about .8 for the F-14 and F-4, and .95 for the A-7E.[5] F-18 reliability is therefore running at about twice the level of aircraft developed before the New Look Philosophy.[6]

When first introduced into service, the F404 experienced an unscheduled engine removal (UER) removal rate of 3.1 per 1000 engine flight hours in the first 2400 hours of fleet experience. This level was about 80 percent lower than engines of a decade or two earlier at similar stages of introduction. Over the next year the UER rate fell even lower to a fleet average of 1.1 at 178,000 cumulative flight hours.[7] At the same level of cumulative flight experience, the UER rate for the F100 engine was about 7.6.[8]

---

[5] MFHBF trend charts for several aircraft were obtained from the Navy's NAVAIR R&M office; F-18 data were current through first quarter of 1985.

[6] We recognize that reliability estimates based on fleet-wide averages can lead to erroneous conclusions because of the changing mix of new and older aircraft at different stages of reliability improvement and aging. However, by using the same data collection and analyses systems and by looking at similar points in their experience profiles, some of the problems of comparing different aircraft programs can be reduced.

[7] Selling, 1983, p. 5.

[8] *F100 PW 100/FF-15 Engine/Module Actuarial History*, Propulsion Management Division Directorate of Materiel Management, Kelly Air Force Base, Texas, 4 May 1984.

Cost of reliability: The identified additional reliability programs during the F-18 development were $155 million (1985 prices), or about 4 percent of total development costs. The airframe weight was 500 pounds heavier because of reliability, inducing an estimated $310,000–$480,000 production cost increase—about 1.6 to 2.6 percent of production unit cost. Engine reliability costs involved $250 million in development—26 percent of the total. Engine prototype costs, which were wholly allocated to reliability, were the major source of engine reliability costs in our estimate.

Reliability improvements and benefits: The F-18 fleet reliability was about double that of its predecessors—1.75 to 1.9 hours compared with about .8 to .95 hours. The F404 engine exhibited a UER rate of 3.1 per 1000 flight hours in its early fleet experience, which fell subsequently to 1.1 as the engine matured. This rate was roughly one-seventh that of preceding generation engines.

Reliability strategies: The F-18 aircraft and F404 engine benefited from several concurrent and reinforcing activities. Reliability was accorded high priority in development. Development attention, resources, and time were devoted to reliability goals beyond the amounts witnessed for typical programs. They also benefited from better reliability technology. More was known about failure mechanisms and how to design and test for them. Materials and electronics components with greater inherent reliability had become available. And tradeoffs were made to give up some performance in exchange for reliability. Although it is not possible here to allocate shares to these different contributing strategies, without priority it is doubtful whether the new knowledge or technology would have been devoted to achieving reliability. Priority was necessary, however, and perhaps even sufficient, but it probably would not have produced the reliability levels permitted by enhanced technology. Thus, the combination of reliability strategies complemented each other in this case.

Appendix B

POTENTIAL BENEFITS FROM F-16 RELIABILITY IMPROVEMENT

A RAND study has examined the hypothetical benefits due to reliability, using planning and simulation models as applied to the U.S. Air Force F-16 fighter aircraft.[1] The author investigated two-fold and four-fold improvements in the reliability of the fire control system and engines to estimate the effects on spare parts use and investment levels (including spare engines and modules), repair costs at depots, maintenance manning, and sortie generation rates. The results of this study may be useful for comparison with the F-18 case described above because the F-18 actually achieved reliability levels twice those of previous Navy aircraft. In the F-16 case description, we therefore examine the results of the two-fold reliability improvement. Note that the F-16's empty weight of about 15,000 pounds is one-third less than that of the F-18, and the F-16's single engine further reduces the complexity, maintenance, and spares requirements from that of the twin-engined F-18. The benefits shown in the F-16 analyses, therefore, would be lower than expected for the F-18.

The effects on life-cycle costs of recoverable spares, engines, and engine modules from a two-fold reliability improvement of the F-16 engine and fire control system are shown in Table B.1. Sortie rates were held at a constant level for this calculation. Although the number of aircraft varied somewhat in the original estimates, we assume a steady-state level of 630. Also, for convenience, we assumed the investments to have been made evenly over a 13-year period (the original estimates show a gradual buildup and decline in annual investments).[2] The total savings per aircraft is $1.8 million, or about 11 percent of the price of a new aircraft of $16.6 million in 1984. The total investment savings of $1140 million would have bought almost 70 aircraft at 1984 prices.

The study also considered the effect on maintenance manning of a 100 percent increase in MTBF for the aircraft as a whole. By the use of a simulation model, the study

---

[1] Abell et al., 1988.

[2] A conservative "churn" value of 0.10 was used. Churn measures the year to year investments required solely by changes in parts characteristics; if the parts lists were completely stable, churn would be zero.

Table B.1

THE BENEFITS OF DOUBLED RELIABILITY OF ENGINES AND FIRE CONTROL SYSTEM ON THE F-16 FIGHTER AIRCRAFT[a]
(Millions 1985 $)

Investment Category	Baseline Investment	Investment for Doubled Reliability	Savings from Improved Reliability	Savings per Aircraft	Savings per Aircraft per Year
Engines and engine modules	788	448	340	.540	.042
Recoverable peacetime operating stocks	901	707	194	.308	.024
Condemnation replacements	837	536	301	.478	.037
Depot-level component repair	1175	870	305	.484	.037
Total	3701	2561	1140	1.810	.129

SOURCE: Abell et al., 1988, Table 5.
[a] Savings per aircraft and per year based on steady-state levels of 630 aircraft and a 10 percent discount rate. Investments based on a "churn" value of .10.

was able to consider tradeoffs between maintenance and sortie rates. A two-fold improvement in reliability could yield either a 9 percent reduction in maintenance manning requirements or a 17 percent increase in sortie production with very little change in savings in spares. The maintenance manpower reduction is equivalent to an annual cost saving of $69,000 per aircraft.[3] The 9 percent manpower savings appears disproportionately among avionics and propulsion technicians, the demand for which would fall by 20 to 30 percent. Such skilled maintenance personnel are more likely to be in short supply than other skilled specialists; the net effect on sortie generation is therefore likely to be greater than suggested by the estimates, which are based on the assumption of no personnel shortages.

To a first approximation, a 17 percent increase in sortie rate is equivalent to having a 17 percent larger force. When applied to eight F-16 wings of 72 aircraft each, the reliability improvement is equal to 98 additional aircraft, which at 1984 contract prices would cost $1.6 billion. Alternatively, current sortie generation capabilities could

---

[3] Abell et al., 1988, pp. 16–17.

be achieved with about 15 percent fewer aircraft, yielding a force size saving of $1.4 billion.

> Reliability improvements and benefits: A two-fold improvement in F-16 reliability, holding sortie rate constant, would yield an estimated savings of over $1 billion, or $1.8 million per aircraft—11 percent of the cost of a new aircraft. Alternatively, the increased reliability could generate a 17 percent increase in sortie rate; manpower savings would be lost, but spares investment would be little affected.

Appendix C

THE CH-47 (CHINOOK) HELICOPTER MODERNIZATION

THE CH-47

The CH-47 (Chinook) is the U.S. Army's only medium-lift helicopter. It was designed in the late 1950s and produced in several succeeding models in the following two decades. Subsystems with 1950s and 1960s technology on the A, B, and C models were modernized when the helicopters were converted to the D model with the replacement of older components by several newly developed components.[1] These include new fiberglass rotor blades, transmissions, hydraulic system, auxiliary power unit, electrical system, flight control system, and a multi-hook cargo load suspension system. An improved engine, introduced previously into the C model, will be placed on all converted models. This modernization program extends the service life of the fleet by up to twenty years. But a major aim of the program is to provide substantial improvements in reliability, availability, maintainability, flight safety, and survivability. Although the modernization also improved operational performance—primarily from replacement of the engines on the CH-47A and B by more powerful engines—emphasis has been on increasing reliability.

CH-47 development began in 1956 when the U.S. Army decided to replace its piston-engine transport helicopters with turbine-powered designs and chose the Vertol Division of Boeing as the system developer. First flight of the new aircraft took place in September 1961. (Table C.1 presents basic data on the different models.) This twin-engine, twin-rotor helicopter quickly became the Army workhorse, and more than 350 of the original A model were produced. Operational experience demonstrated the need for improvements in flying qualities; early in the production cycle of the A model, the original Avco Lycoming engine of 2200 shaft horsepower was replaced by a more powerful 2650 horsepower engine. The B model with a 2850 horsepower engine, redesigned rotor, and other improvements went into production in 1966. The C model, incorporating 3750 horsepower engines, strengthened transmission, greater fuel capacity,

---

[1] Production of new D models was originally contemplated in Army plans, but Congress has authorized conversion of older models only to the D configuration.

Table C.1

CH-47 HELICOPTER CHARACTERISTICS AND PERFORMANCE

Characteristic	Model A	Model B	Model C	Model D
First flight date	9/61	10/66	10/67	5/79
Date into service	10/63	5/67	4/68	2/84
Number produced (U.S. Army)	354	108	270	
Number modernized to D	156	77	203	436 (Total)
Empty weight (lb)	17,913	19,375	20,250	23,100
Maximum takeoff wgt (lb)	33,000	40,000	45,000	50,000
Maximum external payload (lb)	16,000	16,000	20,000	25,000
Maximum cruise speed (mph)[a]	127	178	189	185
Engine model, Avco Lycoming	T55-L-7B	T55-L-7C	T55-L-11A	T55-L-712
Maximum shaft horsepower	2,650	2,850	3,750	3,750[b]

[a] Cruise speed at sea-level, maximum internal load.

[b] Emergency horsepower rating of 4500 available for a total of 30 minutes cumulative time over the overhaul life of the engine.

and other changes was in production by 1968. In the early 1980s, improved reliability engines with an emergency power reserve of 4500 horsepower were retrofitted onto the CH-47C. These same engines are being incorporated in the conversion of A and B models to the D version. Also, fiberglass rotor blades, developed under the modernization program, were retrofitted onto a number of C model aircraft before their full modernization.

The U.S. Army used about 300 CH-47 Chinooks in Vietnam where, among other duties, they recovered more than 5700 disabled aircraft. The U.S. Army procured more than 700 Chinooks. Six other countries have purchased 76 additional aircraft, and it has been licensed for production in Italy. Some of the original airframes produced in the early 1960s will still be flying well into the next century.

THE COST OF IMPROVED RELIABILITY

The CH-47 modernization program produced three major results: extended life, enhanced performance, and improved reliability. Our task here is to estimate the cost of achieving the improved reliability. However, these three outcomes are jointly produced, which introduces an analytical problem: allocation of joint costs to separate outcomes is essentially an arbitrary procedure. Often, joint production yields a lower total cost than if each of the outputs were produced separately. In such cases, if one were to attempt to calculate the production cost of one of the joint outputs (e.g., reliability) by subtracting from the total the costs of the individual outputs (extended life and higher performance) as though they were produced separately, the cost attributed to the residual output (reliability) would be too small (because joint production is less costly than separate production). Reversing the order of the calculation does not solve the problem but results in an estimated cost that is too high.

We deal with this theoretically intractable problem by estimating upper and lower bounds to the cost of reliability. We first treat the cost of improved reliability as a residual, after taking account of the cost of extended life and higher performance. This approach generates a lower bound on the joint cost of reliability. The second approach reverses the order of the calculation; we first estimate the cost of improved reliability, leaving the other outputs as the residual elements, generating an upper bound on the estimate.

New A, B, and C models could (theoretically) be manufactured incorporating the original levels of performance and reliability. The only result of this hypothetical program would be extended life. The cost of such a program would be roughly $1.5 billion (in 1985 values).[2] (See Table C.2 for CH-47 costs.) The modernization program cost roughly $2.9 billion ($242 million RDT&E, $2.70 billion production). The difference of $1.2 billion can be attributed in a simple-minded way to improved performance and reliability.

To calculate the cost of improved performance, we assume that performance of the D model is roughly equivalent to that of late model Cs. The late C models incorporated the product-improved T55-L-11D engines as well as other improved components. We calculate the cost of increased performance as the difference between the late C model cost and the cost of the original A, B, and C models. This calculation produces an estimate of the cost of increased performance of about $1928 million.[3]

Although the C model L-712 engine development and retrofit were not originally a formal part of the modernization program, they enhance the reliability of the helicopter fleet, and C models converted to Ds will not require an engine retrofit. Therefore, to/p

---

[2] This figure was estimated by multiplying the production cost of each model (in 1985 prices) by the quantity entering the modernization program: 156 A, 77 B, and 203 C (Life-extension costs = 156 (2.6) + 77.(2.56) + 203 (4.5) = $1519 million).

[3] Late C model costs were about $7.9 million in 1985 prices. The calculation is as follows: Costs of increased performance = 156(7.9 – 2.6) + 77(7.9 – 2.56) + 203(7.9 – 4.5) = $1928 million.

Table C.2

CH-47 DEVELOPMENT, PRODUCTION, AND MODERNIZATION COSTS
(Millions)

	CH-47A	CH-47B	CH-47C	CH-47D
RDT&E				
Years	1956–64	1963–72		1975–81
Total aircraft: Current$	80	95		140
1985$	339	307		242
Engine: Current$	21	52		25
1985$	101	203		50
Production				
Years	1960–66	1966–68	1968–79	1985
Total aircraft: Current$	.700	.720	.920–5.200	
1985$	2.600	2.560	4.50–7.900	9.700[a]
Engine: Current$	0.068	.063	.115[b]	.450
1985$	.268	.248	.405	.450
Modernization				
Years	1981–92	1981–92	1981–92	
Total aircraft: 1985$	6.900	6.600	5.300	

[a] Boeing contract proposal. This model was never produced.
[b] 1970 production costs.

produce a more accurate estimate of the cost of reliability, we include the cost of the T55-L-712 product improvement program (PIP), and the difference between L-712 and L-11 engine production costs.

With the cost estimates of life extension and increased performance in hand, all that is needed to calculate the residual cost of reliability is to subtract these costs from total CH-47 modernization program costs. Program costs and our estimates for life extension and performance increases are shown in Table C.3, which indicates that if

Table C.3

CH-47 MODERNIZATION PROGRAM COSTS AND ESTIMATES
OF THE COSTS OF LIFE EXTENSION AND
PERFORMANCE IMPROVEMENTS
(Million 1985 $)

Actual Program Costs		Allocated Cost of Benefits	
RDT&E (airframe)	192	Life extension	1519
Conversion	2700	Performance increase	1925
L-712 PIP	50	Total	3444
C model L-712s	18		
Total	2960		

produced separately, their costs would have exceeded the total CH-47 modernization program costs. That implies that reliability improvement costs nothing. Of course, as suggested above, the use of the residual approach in this manner is a lower bound estimate of a joint cost. What the calculations show is that fleet modernization through purchase of new, late-model CH-47C helicopters would have cost more than the program that was actually chosen.

Turning now to the second approach—estimating reliability improvement costs first and leaving the others as residuals—we again assume that the cost difference between the late model CH-47C and a newly produced D model can be attributed largely to reliability improvements. Although Congress never authorized D model production, production contracts between Boeing and the U.S. Army had been in negotiation. We use a 1982 Boeing proposal based on a buy of 150 aircraft and convert to 1985 prices. The production cost of reliability is then the difference between new model D production and late model C. This comes out to $1.8 million per aircraft, or $785 million for the 436 in the modernization program. In Table C.4, this figure is substituted for the total conversion cost shown in Table C.3. We also assume that all of the RDT&E expenditures for the modernization program are allocated to reliability improvements; this assumption will be relaxed below. The total program reliability improvement cost estimate now comes out to $1045 million. This upper bound figure averages out to $2.37 million per aircraft, which is about 30 percent of the price of a late model C and 24 percent of a new model D, if it had been produced. Converting this investment into an annual flow by applying a 10 percent discount factor over a 20-year life yields an annual

Table C.4

RELIABILITY IMPROVEMENT COST,
UPPER BOUND ESTIMATE,
CH-47 MODERNIZATION
(Million 1985 $)

Program Element	Cost
RDT&E (airframe)	192
Production cost of reliability improvement	785
L-712 PIP	50
C model L-712s	18
Total	1045

cost per aircraft of $278,000. (Straight-line depreciation over the 20-year estimated life of the rebuilt helicopters would cut this figure to $118,000).

As noted above, the entire $192 million (1985 prices) RDT&E cost of the modernization program was attributed to reliability improvement. If we make the alternative assumption that the ratio of reliability costs to the costs of the other outcomes are the same in the development phase as in production, RDT&E contribution to reliability is reduced to $56 million.[4] Total reliability improvement costs now fall to $909 million, or $2.08 million per aircraft, about 13 percent smaller than the previous estimate.

CH-47D RELIABILITY

Sufficient operating experience has been generated by the modernized CH-47 fleet to begin to make comparisons to earlier models of this design. Table C.5 presents MTBF data for the A, B, and C models. The chief problem with interpreting these data arises

Table C.5

RELIABILITY GROWTH, CH-47 MODELS

Model	Years	Sample Hours	MTBF (hours)
CH-47A[a]	1963	4,780	.31
CH-47B[b]	1966	108,000	.69
CH-47A,B[c]	1969	724,000	.92
CH-47C[d]	1969–70	4,132	.73
CH-47A,B,C[e]	1972	1,194,582	.96

[a] Includes total cumulative hours of the CH-47A fleet through the end of 1963; Asher et al., 1975, p. BV-4.

[b] Total cumulative CH-47A fleet experience through 1966; MTBF for 1966 only. Asher et al., 1975, p. BV-4.

[c] Total cumulative CH-47A and B experience through 1969; MTBF for 1969 only. Asher et al., 1975, p. BV-4.

[d] Includes hours on three new model C aircraft manufactured in late 1968: The Boeing Company, *CH-47C/L11 Reliability and Maintainability Field Experience*, U.S. Aviation Test Board, Fort Rucker, Alabama, D210-10538-1, 1972.

[e] Total cumulative fleet experience, CH-47A, B, and C, through 1972; MTBF for 1972 only. Asher et al., 1975, p. BV-4.

---

[4] This figure is obtained by multipling the full RDT&E figure by the ratio of the estimate of reliability improvement in production to total conversion costs: (785/2700)192 = $56 million.

from several incomparabilities in sample selection and definitions across samples. Nevertheless, certain overall trends emerge that permit us to make summary judgments about reliability. The first 5000 hours of fleet experience of the new CH-47A helicopter in 1963 demonstrated a low level of reliability, with an average MTBF level of only .31 hours. Over the next several years, reliability doubled and then tripled—reaching a level of .69 hours by 1966 and .92 by 1969. New C models initially were less reliable than the more mature fleet of A and B models. Total fleet-wide figures continued to improve after a temporary decline, so that by 1972 the MTBF for that year of a mixed-model fleet stood at .96 hours.

In 1978, the Army conducted a special reliability data-collection effort with the explicit intention of obtaining information that could later be used for comparison with the CH-47D then under development.[5] These data are presented in Table C.6 with comparable figures for the CH-47D, obtained from the same system of data collection and definition. Mission-affecting failures are those that cause a mission abort. System operational failures result from the inability of a component to perform its function within specification for whatever reason, and requires unscheduled maintenance for correction. Hardware failures are defined as independent, primary malfunctions in the equipment (that is, they are not caused by the failure of other equipment or by operator errors). The mean times between these several types of failures are known as MTBMAF, MTBSOF, and MTBHF. MTBSOF is the measure most directly comparable to the figures in Table C.5. Mission, system, and hardware reliability estimates are defined as the probability of the aircraft completing a mission of a specified number of

Table C.6

ALTERNATIVE RELIABILITY MEASURES, CH-47C AND CH-47D

Reliability Measure	CH-47C	CH-47D
MTBMAF (hours)	13.9	26.5
MTBSOF (hours)	.80	1.71
MTBHF (hours)	2.08	3.85
Mission reliability	.93	.96
System operational reliability	.29	.56
Hardware reliability	.62	.77
Sample hours	2137	4582

SOURCE: Cobro Corp., *CH-47D Helicopter Level II SDC Final Report*, RCS-DRCSM-156, August 1985.

---

[5] Information obtained from U.S. Army Aviation Systems Command.

hours without a designated failure.[6] For each of the TBF values shown in Table C.6, the CH-47D is about twice as reliable as the C. The true differences, moreover, are probably even greater than shown, because the CH-47C aircraft were maintained by a stable and experienced contractor team, whereas the D's maintenance was performed by regular Army crews that were less experienced overall and had less stability as members of a team. Certainly, compared with earlier CH-47 models, the modernized CH-47D is more reliable by considerably more than a factor of two.

One of the payoffs from higher reliability is lower maintenance. Maintenance manhours are determined by several parameters: the probability that a failure requiring maintenance takes place and the frequency of regularly scheduled maintenance (both of these taken together yield mean time between maintenance—MTBM); the mean time that it takes to make a repair (MTTR); and the amount of effort devoted to regular, scheduled maintenance. Maintainability comparisons are shown in Table C.7. Several points emerge from these figures. The number of hours between corrective maintenance (induced by failures) was twice as long for the CH-47D (1.00) as for the C (.48); however, the regularly scheduled preventive maintenance period was only .3 hours. The

Table C.7

MAINTAINABILITY MEASURES, CH-47C AND CH-47D

Maintainability Measure	CH-47C	CH-47D
MTBM[a] (hours): corrective	.48	1.00
preventive	1.85	2.16
total	.40	.68
MTTR[b] (hours)	.63	1.05
MMH/FH[c]: unscheduled	5.13	2.88
scheduled	4.25	4.25
total	9.38	7.13
UER/1000FH[d]	4.63	2.16

SOURCES: MTBM: Cobro Corp., *CH-47D Helicopter Level II SDC Final Report*, RCS-DRCSM-156, August 1985. MMH/FH and UER/1000FH Direct operational unit (AVUM) and intermediate (AVIM) maintenance reported under the Sample Data Collection System through September 30, 1986. U.S. Army Aviation Systems Command.

[a] Mean time between maintenance.

[b] Mean time to repair.

[c] Maintenance manhours per flight hour.

[d] Unscheduled engine removals per 1000 flying hours.

---

[6] This time period is defined as one hour for Table C-6; reliability, viewed as the probability of completing the mission of t hours duration without a failure, is calculated as $R = e^{-t/MTBF}$.

resultant total average time between maintenance actions was consequently not as large as the improved failure rate would suggest—70 percent greater on the D than on the C.

Turning to repair time, the story is reversed. Mean time to accomplish a maintenance action takes two-thirds longer on the D. As a result of all these effects, maintenance manhours per flight hour is 7.13 on the CH-47D, 24 percent less than the 9.38 level on the CH-47C. In the early 1970s, analysts ascribed the higher MTTR on the D to the lack of experience of the maintenance crews on the new model. Also, the C model maintenance under the data-collection program was performed by a stable and experienced contractor work force, which would result in less time to make a repair. Nevertheless, the net result shows the modernized D model to require less maintenance than the older model. Indeed, the actual experience is quite close to that predicted early in the program.

The higher reliability reduces not only the demand for manpower but also the requirements for spare parts. In 1975, the U.S. Army estimated the effects on life-cycle costs of improved CH-47 reliability.[7] Their calculations happened to use estimates of MMH and MTBF that were quite close to the figures coming out of actual operational experience. Those estimates produced cost differences between the CH-47C (with the reliability-improved engine) and the CH-47D; for maintenance labor at both unit and intermediate levels, the cost differences came out to $59,000 per aircraft per year (1985 prices), a 22 percent reduction; the annual cost difference per aircraft attributed to lower consumption of parts on the CH-47D was a 17 percent reduction of $49,000. The estimated savings of $108,000 in the flow of manhours and parts is quite conservative, because it does not account for changes in the stocks arising from reduced flows. As shown above in the F-16 analysis, the savings from lower stocks can be twice as large as the actual consumption of spares.

> Cost of reliability: Because reliability was jointly produced with extended life and improved performance, a precise estimate of the cost of reliability was theoretically not possbile. Instead, high and low bounds delineated a range of possible values. The low end of this range suggested that reliability was essentially free, after the costs of extended life and higher performance were accounted for. Estimates of the high end of the range were $900–$1045 million, or about $2.1 to $2.4 million per aircraft. The RDT&E portion of the $1045 million was $242 million for the aircraft and engine improvement; production costs accounted for $803 million. The

---

[7] U.S. Army, Training and Doctrine Command, 1975, p. O-I-A-96.

high-end RDT&E cost estimate is 37 percent of the total development expenditures on the original A model plus the product improvement costs to develop the B and C models. The high-end estimate of increased production cost per aircraft was $1.84 million ($803 million divided by 436 aircraft): about 19 percent of a new D model (if produced), 23 percent of a late-model C, and 40 percent of an early C model.

Reliability improvements and benefits: In terms of mean times before failure, the CH-47D is roughly twice as reliable as the CH-47C. Early results from field experience show that total maintenance manhours per flight hour were reduced from about 9.4 hours on the CH-47C to 7.1 hours on the converted D models, with all of this improvement showing up in unscheduled maintenance. For a typical flying program, this yields a manpower saving of about 500 hours per year per aircraft. Lower consumption of spares on the CH-47D have been estimated as saving annually about $49,000.

Reliability strategies: From the initial deployment of the Chinook in 1963 to a decade or so later, MTBF gradually rose from .3 hours to about .9 hours. These improvements came about from more than a million operating hours of hardware experience, design changes, new models, product improvement programs, and growth in crew experience and proficiency. Reliability, however, was not a central issue in improving the aircraft. Performance and operational matters dominated the maturing design. In the CH-47D development reliability was raised as the central design consideration in the modernization and life-extension program. Given this requirement and its priority, the developers doubled reliability. The designers were able to take advantage of experience on the equipment and of new technology to meet their goal. But without the reliability requirement and priority, the knowledge and technology probably would have been used to improve performance, as it had been in the past. That is, there is nothing inherent in fiberglass rotor blades or improved turbine materials that restricts their application to reliability improvements. The fact that they could have been used to produce more power, faster speed, or greater lift, but instead were devoted to enhanced reliability, was a matter of choice not technology. Because this program involved improving an already operational aircraft, it considerably constrained the volume of new features to be verified and consequently reduced the uncertainty and difficulty of the effort. Thus, priority and experience were critical ingredients to achieving higher reliability.

Appendix D

PRATT & WHITNEY F100 ENGINE

THE F100 PROGRAM

In 1968, the U.S. Air Force and Navy requirements for new air-superiority fighters called for substantial increases in engine performance—particularly in the ratio of thrust to weight. These engine demands resulted in a joint Air Force-Navy 18-month engineering development program to demonstrate an engine meeting the needs of both services. Pratt & Whitney was chosen to pursue its proposal to develop two engines from a single gas-generator core, with the larger Navy engine obtained by scaling up the fan, afterburner, and other components of the smaller Air Force design. The Air Force engine was designated the F100–100 and the Navy version the F401–400. Full-scale development began in 1970; the F100 was chosen as the powerplant of the McDonnell Douglas F-15, with the F401 planned for the Grumman F-14B. In 1973, the Navy suspended its development of the F401 because of repeated problems and failures during preliminary test and because of funding constraints. The Air Force continued development, and the F100 passed its 150 hour model qualification test (MQT) in October 1973. Subsequently, the engine was chosen for the General Dynamics F-16 aircraft.[1] By the end of 1985, more than 3500 F100 engines had been produced. Pratt & Whitney foresees a total production run of more than 5000, which will take the program well into the 1990s.

As successful as the F100 is today, its early operational experience was marked by several shortcomings, stemming in large part from the Air Force's ambitious performance goals. The new engine, for example, was called on to generate 15 percent more thrust, weigh 25 percent less, and be more fuel-efficient than the TF30 engine in the F-111. The engine met all these requirements. However, its development schedule was rigidly tied to the concurrent airframe development schedule. Because new engines typically take twice as long to develop as new airframes, the F100 designers could not thoroughly assess problems that appeared in the engine test program. As a result, the engine configuration that went into production was immature.

---

[1] Much of the account of the F100 through 1979 is taken from U.S. Senate, 1979.

After several months of operational use a serious "stall-stagnation" problem began to appear, and durability turned out to be much less than expected. Stall-stagnation resulted from a sequence of events: Disturbances in the compressor air flow caused the compressor blades to stall aerodynamically; the engine consequently lost air flow, temperatures increased, and engine speed decayed. When this sequence occurred, the engine often had to be shut down and restarted. The high temperatures produced by stall-stagnation greatly reduced engine life, and the stall itself created obvious safety and operational problems. Other durability and reliability problems, however, were independent of the stall-stagnation effects.

When performance or reliability problems appeared in operations—and they did in most new engines—the Air Force customarily used a Component Improvement Program (CIP) to correct those problems demonstrated by actual experience in the operational environment. CIP is essentially a continuation of the development effort after an engine is already in production and deployed in field units. Its customary use reflected an implicit belief that it was more efficient to identify and correct problems uncovered by the rigors and variety of operations than in test cells or flight tests during full-scale development. CIP was also made necessary by the compressed development phase into which many engines were forced by the policy of concurrent aircraft and engine development programs. However, as performance requirements grew more demanding and as engine technology grew more complex, reliability performance deteriorated under this development strategy.

According to Pratt & Whitney data, the F100 received about 11,000 test hours through MQT. This was a typical figure for the period but was insufficient and of the wrong type to detect many of the problems that were later to affect the design. More recent engines have been put through considerably more test hours to improve reliability; the F404 had 14,000 hours through MQT; the Army's T700 helicopter engine had more than 18,000 hours at qualification and 42,000 hours before delivery of the first production engine.

A quite different problem also surfaced during the mid-1970s. Even if additional testing had been completed, it is problematic whether the test techniques then used would have had much of an effect on reliability. Unexpected durability problems on the F100 and other engines led to the discovery that mission-generated stresses were much more severe than contemplated by users and designers. In the early 1970s, analysts found that engine life and parts durability were more closely linked to the number of thermal cycles and mechanical stress loadings than to the simpler, traditional measure of engine flying hours. Furthermore, specially instrumented flights on several types of aircraft showed that pilots had been putting engines through many times more throttle excursions and cycles per flight hour than expected.[2] However, the F100 designers, believing that they were within the stress allowances established by military and engineering standards, reduced design margins to meet performance goals. Later, on the basis of the new understanding of cyclic stress, Pratt & Whitney devised accelerated testing techniques, including orders of magnitude more thermal cycles. These techniques have since been added to standard engine development practices, but the F100 did not benefit from this new information until later models of the engine were designed.

The stall-stagnation problem, the higher-than-expected thermal cycles, and the reduced design margins led to a severe problem in F-15 engine availability by 1978. High levels of unscheduled engine removals (UERs), maintenance workloads, and spare parts usage resulted in an engine shortfall that reached 90–100 in mid-1978; up to 100 F-15 aircraft were not flying because of engine unavailability. These problems led to an Air Force decision to direct CIP funds and effort toward increased reliability and durability. In what follows, we examine the effects of the F100 CIP in improving reliability.

THE F100 COMPONENT IMPROVEMENT PROGRAM

The F100 CIP began in 1974, after the engine had successfully completed its qualification tests, and overlapped the last two years of full-scale development. From 1974 to 1985, more than $700 million was spent in this program (Table D.1). In inflation-adjusted values, the CIP funding has been about 40 percent of the original development expenditures. As mentioned above, when F100 reliability problems reached their highest levels in 1978, Air Force commanders directed that CIP efforts be devoted to solving these problems.

One important measure of engine reliability is the rate of UERs from the aircraft. For analytical purposes, this variable has several advantages: It is simple to define and measure, especially compared with other reliability parameters such as MTBF or MMH;

---

[2] Birkler, Cote, and Byers, 1977.

Table D.1

DEVELOPMENT AND COMPONENT IMPROVEMENT PROGRAM EXPENDITURES, PRATT & WHITNEY F100 ENGINE
(Millions of $)

Year	Navy/AF Program[a]	Prototype Development	Full-Scale Development	RDT&E Total	RDT&E Total (1985$)	CIP	CIP (1985$)
1969	34.0	34.2		68.2	256		
1970	60.0	28.4	22.7	111.1	391		
1971	63.0		63.3	126.3	421		
1972	79.1		120.1	199.2	638		
1973			123.3	123.3	370		
1974			120.9	120.9	333	106	292
1975			133.7	133.7	317	33.5	79
1976						59.3	129
1977						60.5	120
1978						40.2	73
1979						52.9	86
1980						47.2	66
1981						51.0	63
1982						64.1	72
1983						73.8	79
1984						63.6	66
1985						52.0	52
Total	236.1	62.6	584	882.7	2726	704.1	1177

[a] Comprises 50 percent of the joint Air Force-Navy core development. The engine price index was taken from: Fatkin, 1981.

and an unscheduled removal represents a serious maintenance action that disrupts maintenance and flying operations, absorbs manpower and equipment, and immobilizes an engine until it can be repaired and returned to an aircraft or to stock. We obtained data from Pratt & Whitney on UERs per 1000 flying hours for individual engine production lots. (An engine lot generally represents a year's production under a single contract.) Because engine improvements and design changes tended to be incorporated into a production lot as a whole, tracing out the history of a lot and comparing it with other lots reveals the progress made in reducing UERs. (Data by production lots is preferable to fleet-wide averages, which combine engines of diverse ages and flying histories.) Figure D.1 and Table D.2 show UER rates by engine flying hours for lots 2 to 11, produced from 1974 to 1983. The data clearly indicate remarkable improvements in UER rates, from 13 per 1000 flying hours in lot 2 to a value 10 years later of 1.5.

Because CIP investment was intended, among other things, to reduce UERs, we attempted to see if there was a direct relationship between CIP funding and UER reduction. Plots of year-to-year changes in UER (by manufacturing lot at 600 flight

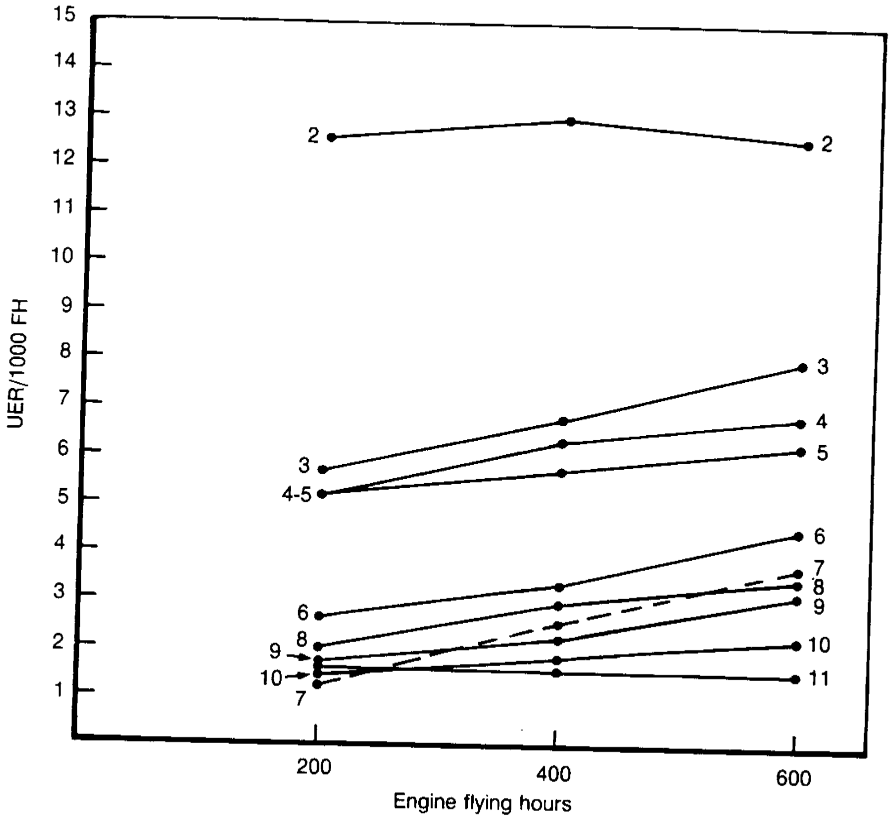

Fig. D.1—F100 unscheduled engine removals by production lot number

hours) show no relationship to annual CIP expenditures; the correlation is close to zero for both absolute and percentage changes (Fig. D.2). Similarly, there is no trend when annual change in UER is plotted against cumulative CIP. These results are consistent with several hypotheses. UER improvements may be to some degree independent of CIP investments, perhaps resulting from better knowledge of failure mechanisms gained from experience, generally available technological advances, improved maintenance methods, and more skilled personnel. Also, because CIP is devoted to goals other than UER, there may not be a one-to-one relationship between the two. Another possibility is that CIP investments in any single improvement is spread over several years, and the results will be observed only with a lag. The cumulative lagged investments of past years may influence the reliability of any specific lot of engines. The absence of a declining trend in UER improvements with cumulative CIP expenditures suggests that the returns to CIP are not falling over time. We shall return to many of these points below.

Table D.2

UNSCHEDULED ENGINE REMOVALS, F100 ENGINE
(per 1000 flying hours)

		UER by Production Lot and Flight Hours[b]			UER F-15, Fleet Average[c]
Year[a]	Lot Number	200 hrs	400 hrs	600 hrs	
1974	2	12.6	13.0	12.6	
1975	3	4.7	6.8	8.0	16.2
1976	4	5.2	6.3	6.8	9.4
1977	5	5.2	5.7	6.3	8.6
1978	6	2.6	3.3	4.5	7.4
1979	7	1.2	2.5	3.7	6.7
1980	8	2.0	2.9	3.5	7.5
1981	9	1.7	2.2	3.1	6.2
1982	10	1.4	1.8	2.2	5.6
1983	11	1.5	1.5	1.5	6.4

[a] Year of manufacture.

[b] Flying hours were accumulated in later years.

[c] F-15 fleet UER is the average experience for the four quarters of the designated year.

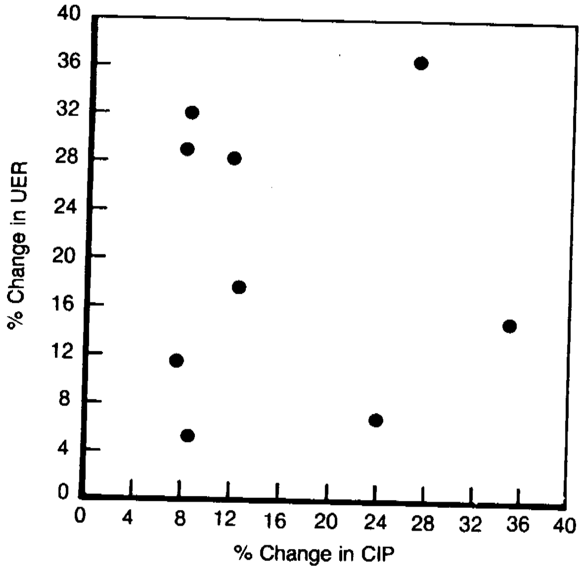

Fig. D.2—Unscheduled engine removal improvements versus CIP investments

LIFE-CYCLE COST SAVINGS

To evaluate proposed engineering changes, Pratt & Whitney has developed a life-cycle cost (LCC) model. The effect of engine design changes on projected life-cycle costs is considered in the decision to pursue a proposed change. Pratt & Whitney analysts have aggregated the individual life-cycle cost savings estimated by the LCC model for all changes funded within annual CIP increments.[3] Included in the life-cycle cost estimation are changes in the manufacturing price of the engine for those design changes incorporated into new production engines, labor and parts costs for modifying operational engines, bookkeeping costs connected with parts control, changes in unscheduled maintenance costs arising from design effects on reliability, and changes in spare parts usage and consumption. Undiscounted LCC savings by annual CIP increment (adjusted to 1985 prices) are shown in Table D.3. (For a 15-year steady-state flow of savings and a 10 percent discount rate, the present value of the LCC savings are about 50 percent of the estimates shown in Table D.3.)

Annual life-cycle cost savings are plotted against annual CIP in Fig. D.3. The relationship between annual values of LCC savings and CIP is random: The correlation

Table D.3

LIFE-CYCLE COST (LCC) SAVINGS
ATTRIBUTED TO CIP EFFORTS
(Million 1985 $)

Year	LCC Savings
1974	164
1975	417
1976	1976
1977	685
1978	364
1979	701
1980	88
1981	769
1982	1183
1983	569
1984	395
Total	7311

---

[3] Aggregate annual life-cycle cost savings estimates were normalized according to the following ground rules: 3388 F-15/F-16 engines, 12.2 million engine flying hours, 1983 prices. *1984 F100 Engine Component Improvement Program*, U.S. Air Force, Aeronautical Systems Division (ASD/YZF), Wright Patterson Air Force Base, Ohio, and Pratt & Whitney, Government Products Division, West Palm Beach, Florida, GP84-443, June 1984, p. 64.

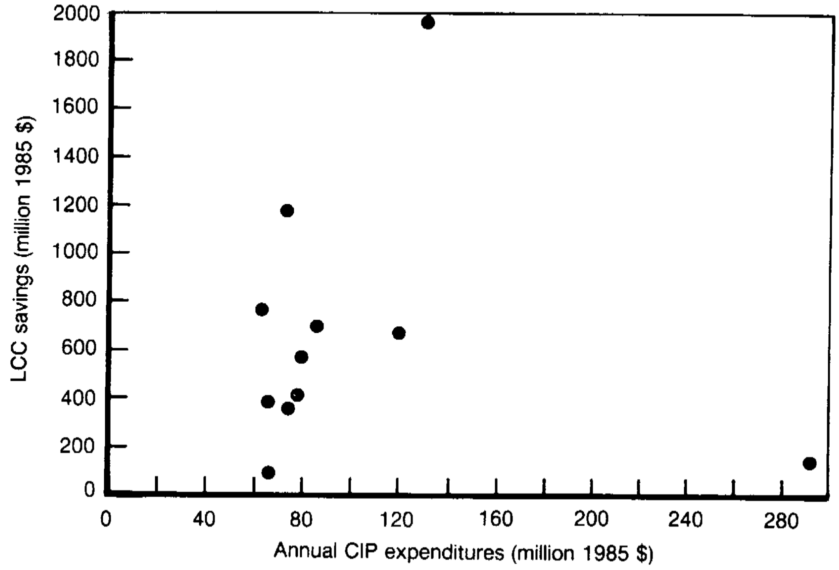

Fig. D.3—The relationship between life-cycle cost savings and CIP expenditures

coefficient between the two variables is -.092.[4] Two points appear to be outliers—the observations for 1974 and 1976. If the 1974 point (at the lower right) is removed, the variables have a correlation of .66 and a strong positive relationship; however, if the 1976 observation (at the top of the graph) is removed, the relationship is weakly negative (r = -.35); and if both are removed, the relationship is even more weakly positive (r = .18).

These results show an average, although highly variable, LCC savings return to CIP investment of approximately 600–900 percent (or about half this rate if a 10 percent discount is applied). Furthermore, there is no evidence that these returns are falling over time. Although it is customary to assume that declining returns will occur eventually at some point during the life of a program, that point apparently has not yet arrived.

Design engineers are often requested to produce lists of proposed changes to a product that will generate benefits in terms of LCC savings, reliability, safety, etc. These changes will also require an investment of RDT&E resources and an implementation cost

---

[4] We do not use the customary aerospace industry approach of plotting cumulative variables against each other because even independent, random variables will show a high correlation when their cumulative values are plotted. (The slope of the plotted curve will be the ratio of the means of the two variables.) Such high correlations will be present whether or not the annual values of the variables are correlated.

for units that are modified (in production or in the field). If each of these potential design changes are arrayed by decreasing returns (that is, by the ratio of benefits to cost) and plotted against cumulative cost, they will by definition trace out a curve with diminishing returns. The shape and position of this curve is affected by the level of technology available at the time, knowledge about the product (e.g., failure mechanisms), the volume of research and testing performed, and the maturity of the product. Such a hypothetical curve is shown as A-A in Fig. D.4. A decision must be made about how many of these design changes to implement. As shown in the figure, program managers chose to invest C_1 in the first time period where the marginal return is at R_1. A central question is whether, in the next period, the return to investment continues to move down the original curve A-A, or whether additional knowledge, experience, and technology shift the curve upward. As depicted in the hypothetical figure, the curve is shifted upward so much that the average return in the second period is as great as in the first. In general, the greater the jump of the curves from period to period, the more that post-development CIP improvements should be used.[5] If there were no advantage to an additional year of experience (if the returns simply continued along the unfunded portion of the original A-

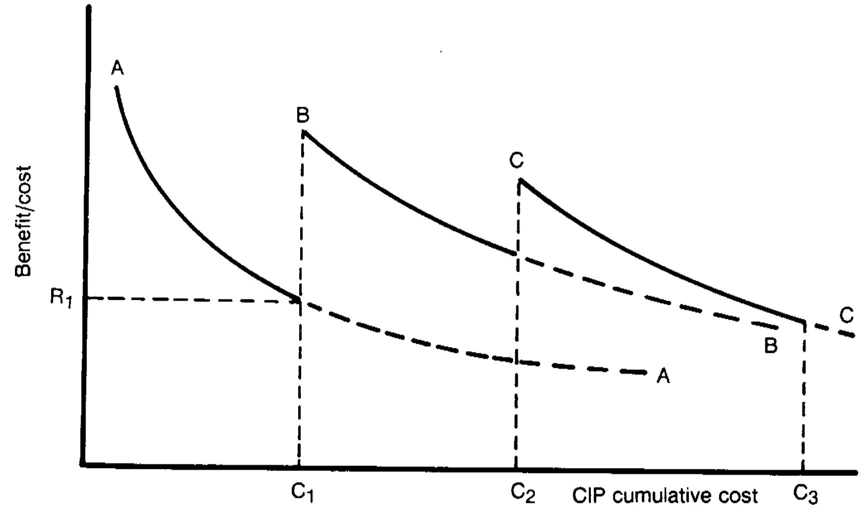

Fig. D.4—Hypothetical CIP-benefits curves over time

---

[5] Of course, other considerations would enter into a CIP decision: for example, the returns to additional testing during development, the cost of uncorrected defects, and the cost of retrofit versus production line changes.

A curve), then planners would do just as well to make all their CIP investments in a single period and immediately reap all the benefits. On the basis of the yearly figures of F100 CIP experience, the gains from an added year of experience apparently match the diminishing returns of any particular year's knowledge base. We now turn to a detailed examination of these and other issues by review of a single year's CIP projects.

CIP TASKS

Pratt & Whitney provided detailed information on the individual engineering design tasks constituting the CIP effort for 1985; because of the multi-year nature of many of these changes, they included tasks to be incorporated in the years 1984 to 1987, with expenditures from 1979 to 1985. For inclusion in our analysis, a task had to have some CIP investment and show a positive effect in at least one of several benefit categories: mishap rate, support cost, UER, maintenance manhours per flying hour, or mean time before failure. Some positive level of CIP funding was found in 82 tasks, and 62 tasks met the additional criteria of having a positive effect in some category. The total CIP investment in the 82 tasks was approximately $120 million (in 1985 prices), spread over the years 1979 to 1985. In addition to CIP investment costs, we also obtained estimates of the costs of incorporating these changes in new production engines and of retrofitting the changes into previously produced engines.

The total estimated effect of these tasks on the F-15 aircraft were substantial: reduction of UER by 1.91 removals per 1000 flying hours (43 percent); reduction of MMH/FH by 258 hours per 1000 flying hours (15 percent); increase of MTBF by 2.87 hours (5 percent); reduction of class A mishaps (on the F-16) by .11 mishaps per 100,000 hours (21 percent); and reduction of support costs by $175 per flying hour (33 percent).

It is generally accepted that incorporating changes into the production line is less expensive than retrofitting them into fielded engines. Pratt & Whitney cost estimates are consistent with this presumption: 41 CIP tasks required either retrofit expenditures or changes in the price of a new engine (or both). The average retrofit cost per engine (parts and labor) was approximately $1800. In contrast, the price effect of incorporating the CIP design change into production engines was a *negative* $50. On the average, production costs were reduced by the CIP change. These cost differences provide an incentive to make design changes early in a program, before a large number of units have already been produced that may have to be retrofitted in the field.

Our CIP task database allows us to plot a curve of CIP benefits against CIP investment based on realistic information, analogous to the hypothetical curves shown in Fig. D.4. To do this, we form the benefit to cost ratio for each task ($\Delta MTBF/CIP$, $\Delta UER/CIP$, etc.), rank the tasks according to this ratio, and plot the values against cumulative CIP. Plots of these ratios are shown in Fig. D.5. A striking feature of these plots is the enormous range of benefit-to-cost ratios and of investment per task. Both variables span four to five or ders of magnitude. A fairly small number of tasks provide substantially higher returns per dollar invested than most of the other changes. Thus, four design changes alone produced reductions of .1 to .6 MMH/FH per thousand dollars of CIP investment; in contrast, 22 changes had effects under .01 MMH/FH, and 38 design changes were estimated as having no effect on maintenance manhours at all. Of the total gains produced in MTBF, 70 percent were obtained from only about 5 percent of the CIP funds, and 99 percent of the gain required 50 percent of the investment. These returns were more broadly based in the area of improved support costs: 50 percent of the money yielded 80 percent of the total benefits. In general, most of the positive benefits are accounted for by a handful of changes and a small percentage of investment resources.[6]

Analyses of single variables only tell a partial story. The simple relationships between CIP investment and changes in MTBF, UER, MMH/FH, and support costs are essentially random. For example, Fig. D.6 plots the reduction in MTBF for each task against that task's CIP investment. The size of the CIP expenditures on each task is not associated with the returns on any single variable; the reasons for proceeding with a particular change are multi-dimensional—safety, reliability, costs, and performance all play a role. Full investigation of these relationships requires multivariate statistical analysis.

MULTIVARIATE CIP TASK ANALYSIS

We tested the assumption that the cost of CIP tasks is related to all of the benefits arising from that task, rather from any single benefit. To do this, we used statistical regression analysis, estimating equations with CIP investment per task as the dependent variable; the independent variables were the several benefits arising from CIP tasks:

---

[6] All of these findings hold when production and retrofit costs are added to CIP investment.

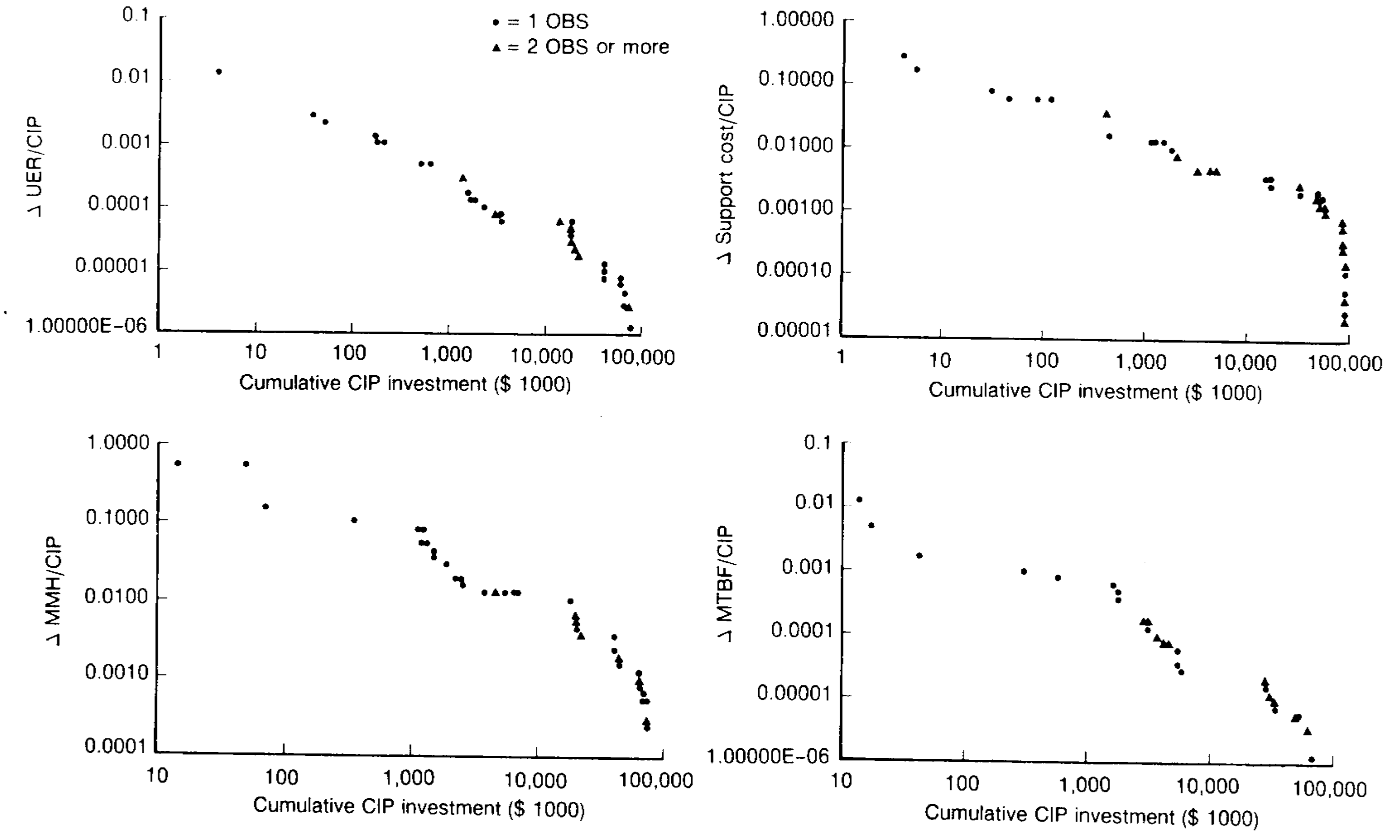

Fig. D.5—CIP benefits per dollar invested versus cumulative investment, for individual CIP tasks ranked in order of decreasing returns per dollar

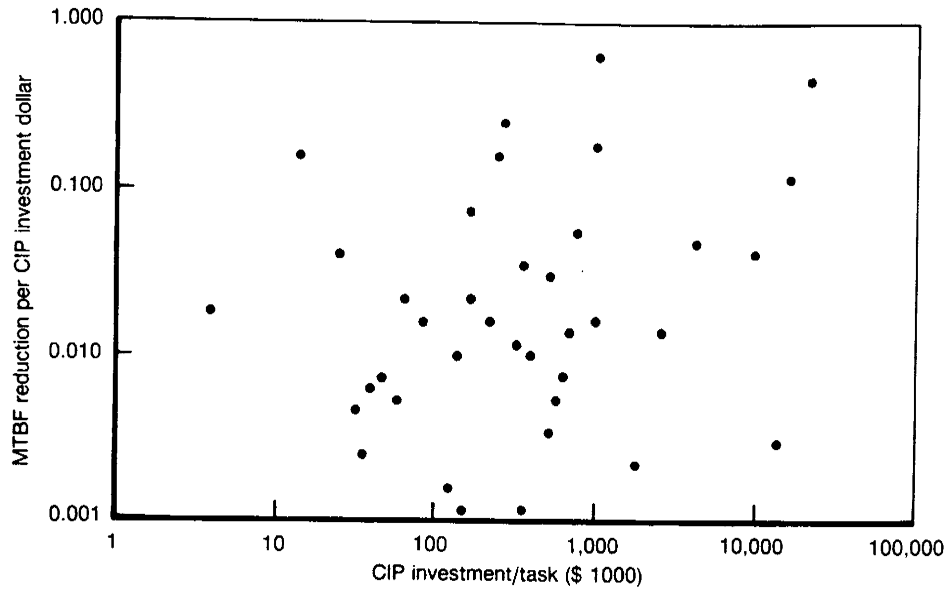

Fig. D.6—CIP benefit (MTBF) versus investment, for individual CIP tasks

changes in support cost, UER, MTBF, MMH/FH, and mishap rate. Testing of alternative functional forms demonstrated that a logarithmic equation with quadratic terms best fit the data. After we eliminated observations whose independent variables all had zero values, 62 observations remained in the sample. However, zero values for UER, MTBF, and mishap rate appeared in a quarter to a third of the remaining observations. To be able to use logarithms, we substituted the lowest observed value of each of these variables for the zero values.

Equation 1, Table D.4, includes the logs and the squares of the logs as independent variables. Support cost and mishap rate are statistically significant, but the coefficients of the other variables are not significant. The probable reason is that support cost is itself a function of UER, MMH/FH, and MTBF. Therefore, once support cost is included in the equation, these other support-related variables add little explanatory power. This is indicated in Eq. (2), where only support cost and mishap rate are included; all the variables are significant, and adjusted R^2 (correcting for differences in degrees of freedom) declines only moderately from .425 to .382, indicating that little explanatory power is lost by the exclusion of the other variables. In Eq. (3), we drop out support cost and retain the other support variables. Several of the support variables are now significant, but others are not. However, because the quadratic terms are highly

Table D.4

EQUATIONS OF LOGARITHM OF CIP COST PER TASK AND BENEFIT VARIABLES

(62 observations; t-statistics in parentheses)

	Equation					
Independent Variable	(1)	(2)	(3)	(4)	(5)	(6)
R^2	.520	.423	.413	.377	.457	.436
Intercept	8.39	9.42	12.04	13.03	9.64	13.13
n Δsupport cost	.576	.506	.488			
	(3.24)	(4.26)			(4.18)	
n Δsupport cost, squared	.132	.121			.115	
	(2.98)	(3.70)			(3.58)	
n ΔMTBF	–.382		–.453			
	(1.03)		(1.13)			
n ΔMTBF squared	–.026		–.031			
	(.89)		(.96)			
n ΔUER	–.151		.060			
	(.29)		(.11)			
n ΔUER squared	.0083		.029	.025		.022
	(.20)		(.65)	(2.89)		(2.50)
n ΔMMH/FH	.306		.510	.451		.453
	(2.14)		(3.67)	(3.85)		(4.02)
n ΔMMH/FH squared	–.063		.049			
	(1.03)		(.87)			
n Δmishap rate	1.27	1.11	2.03	1.92	1.09	1.84
	(1.83)	(1.72)	(2.86)	(3.11)	(1.7)	(3.10)
n Δmishap rate squared	.074	.068	.110	.104	.066	.10
	(1.86)	(1.81)	(2.63)	(2.80)	(1.78)	(2.79)
Time period					–.70	–.95
					(1.88)	(2.40)

CIP = component improvement program task expenditures, thousand 1985 dollars.
Δsupport cost = change in support cost per engine flying hour.
ΔMTBF = change in mean time before hardware failure per 1000 flying hours.
ΔUER = change in unscheduled engine removals per 1000 engine flying hours.
ΔMMH/FH = change in maintenance manhours per 1000 engine flying hours.
Δmishap rate = change in number of class A mishaps per 100,000 engine flying hours.
Time period = dummy variable, equal to one if no CIP expenditures in 1984 and 1985, and zero otherwise.

All change variables defined as positive in the direction of improvement.

correlated with the unsquared linear terms, it is not clear whether the reason for lack of significance is that the basic variable itself is not important, or that the inclusion of both linear and quadratic terms is inappropriate. To test these possibilities, we tried alternative specifications—including and not including each element of the linear-quadratic pair. The best fit was shown by Eq. (4). UER and MMH/FH are the variables with the best statistical properties and the most stable coefficient estimates. (Even with the exclusion of support cost, MTBF continued to be insignificant.) Overall, the statistical properties of the equations show that although the indicated variables—particularly support cost—affect the value of CIP costs, there is still considerable randomness or unexplained effects on CIP.

Concentrating for the moment on Eq. (2), the coefficients indicate that CIP rises with support cost and rises faster the greater the desired change in support cost. Equation (2) indicates that, at the mean value of support cost improvement of about \$3.0 per flying hour, the elasticity of CIP investment with respect to support cost is .75.[7] In this range, CIP cost rises less rapidly than the percentage gain in support costs. However, for support cost reductions of \$10 and \$20 per flying hour, the elasticity rises to 1.06 and 1.23. At these higher levels of support cost reductions, percentage increases in CIP investments rise faster than the percentage reduction in support costs.

To test whether these cost-benefit equations shifted over the years, we used data on annual CIP investments for the years 1979 to 1985. We define a variable with a value of 1 if there were no CIP expenditures in both 1984 and 1985; otherwise, the variable was defined as equal to zero. Thus, the variable equaled one for those tasks completed before 1984. A negative coefficient on this variable would indicate that CIP costs were lower in the earlier period, holding the independent variables constant. The results of inserting the dummy variable into Eqs. (2) and (4) are shown in Eqs. (5) and (6). The statistically significant coefficients on this variable indicate 40–50 percent lower CIP costs before 1984 than in those tasks carried on in the 1984–1985 period.[8] Equation (5) is

---

[7] The elasticity of a relationship is defined as the ratio of the percentage change in the dependent variable to a percentage change in the independent variable. For linear logarithmic equations ($\ell n\, y = a + b\, \ell n\, x$), the elasticity is simply the coefficient of the independent variable (that is, b). In quadratic logarithmic equations ($\ell n\, y = a + b\, \ell n\, x + c\, (\ell n\, x)^2$), the elasticity is $b + 2c\, \ell n\, x$: The elasticity varies with the independent variable. If $\ell n\, y = a + c\, (\ell n\, x)^2$, the elasticity is $2c\, \ell n\, x$.

[8] To find the percentage effect of this variable, we raise the base of the natural logarithm, e, to the power equal to the coefficient of the dummy variable.

plotted in Fig. D.7. We assumed that, in both periods, CIP tasks were undertaken requiring \$100, \$200, \$300, \$400, \$500 and \$1,000 (thousands of 1985 dollars).[9] The ratio of support cost improvements to CIP was calculated from the equation. (The mean value of the mishap rate variable was used to evaluate the equation.) Because of the quadratic functional form of the estimated equations, the curves show sharply declining tails, in contrast to the hypothetical curves pictured in Fig. D.4, but similar to the shape shown in Fig. D.5. The second period curve does not simply continue from the lowest point on the earlier curve; neither does it recover to the original level, but is approximately 40 percent below it. Thus, the statistical results are consistent with the notion of diminishing returns to CIP investment, but with a stimulus provided by experience and technology.[10]

TRIMMING THE WICK

In 1978 and 1979, during the period of maximum reliability problems in the hot section of the F100, one of the solutions implemented was the reduction of maximum turbine inlet gas temperature. Engineering analysis had indicated that a temperature reduction of 20° – 30° F would double the life of turbine air foils while reducing supersonic thrust by 2–3 percent. The Air Force experimented with temperature reductions of up to 80° F. Their estimate in 1979 was that 80° F reduction would save about \$11 million just in parts costs alone and reduce engine removals by about 40 per year; much of the savings were seen to come about from temperature reductions of only 30° F. The Air Force authorized fleet-wide reductions of maximum thrust levels of 3 percent (30° F reduction). A demonstration program at selected Air Force bases was then carried out to test the effect on support costs of increasing thrust from 97.0 to 98.5 percent of maximum. Results showed that considerable additional support resources would be needed to pursue fleet-wide uptrimming to the 98.5 percent level.[11] The Tactical Air Command decided to accept a 30° F reduction and the consequent loss of

---

[9] The arithmetic mean of CIP per task is \$1500; the antilog of the mean of the log of CIP is \$310 (both values in thousands of 1985 dollars).

[10] The selection of cases into two time periods (as was done here) may yield biased results because the subsamples do not represent all the cases funded in these two periods, but rather those that happened to be active in 1984. A more thorough analysis of this subject would require obtaining complete subperiod samples.

[11] Nix and Shelnutt, 1984, p. 72.

plotted in Fig. D.7. We assumed that, in both periods, CIP tasks were undertaken requiring \$100, \$200, \$300, \$400, \$500 and \$1,000 (thousands of 1985 dollars).[9] The ratio of support cost improvements to CIP was calculated from the equation. (The mean value of the mishap rate variable was used to evaluate the equation.) Because of the quadratic functional form of the estimated equations, the curves show sharply declining tails, in contrast to the hypothetical curves pictured in Fig. D.4, but similar to the shape shown in Fig. D.5. The second period curve does not simply continue from the lowest point on the earlier curve; neither does it recover to the original level, but is approximately 40 percent below it. Thus, the statistical results are consistent with the notion of diminishing returns to CIP investment, but with a stimulus provided by experience and technology.[10]

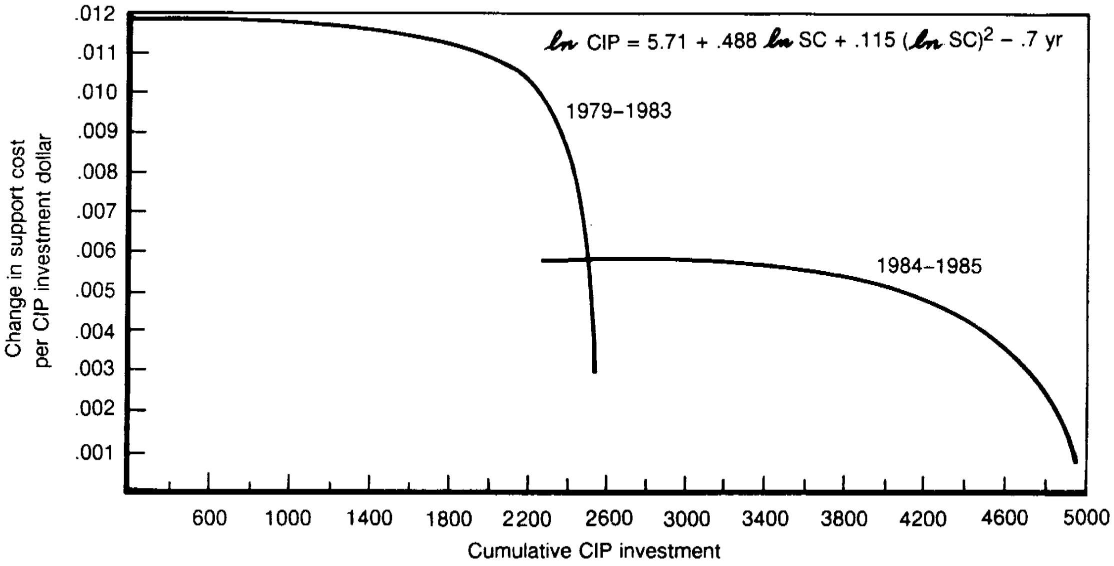

Fig. D.7—Simulated support cost reduction per dollar invested versus cumulative investment, for two time periods, from statistically estimated equation.

---

[9] The arithmetic mean of CIP per task is \$1500; the antilog of the mean of the log of CIP is \$310 (both values in thousands of 1985 dollars).

[10] The selection of cases into two time periods (as was done here) may yield biased results because the subsamples do not represent all the cases funded in these two periods, but rather those that happened to be active in 1984. A more thorough analysis of this subject would require obtaining complete subperiod samples.

TRIMMING THE WICK

In 1978 and 1979, during the period of maximum reliability problems in the hot section of the F100, one of the solutions implemented was the reduction of maximum turbine inlet gas temperature. Engineering analysis had indicated that a temperature reduction of 20° – 30° F would double the life of turbine air foils while reducing supersonic thrust by 2–3 percent. The Air Force experimented with temperature reductions of up to 80° F. Their estimate in 1979 was that 80° F reduction would save about $11 million just in parts costs alone and reduce engine removals by about 40 per year; much of the savings were seen to come about from temperature reductions of only 30° F. The Air Force authorized fleet-wide reductions of maximum thrust levels of 3 percent (30° F reduction). A demonstration program at selected Air Force bases was then carried out to test the effect on support costs of increasing thrust from 97.0 to 98.5 percent of maximum. Results showed that considerable additional support resources would be needed to pursue fleet-wide uptrimming to the 98.5 percent level.[11] The Tactical Air Command decided to accept a 30° F reduction and the consequent loss of maximum thrust. Over the next several years, new materials and an improved-life core were developed and gradually introduced into the fleet. As the hardware fixes were made, the temperature was allowed to move up to the original level.

> The cost of reliability: Total F100 CIP investment from 1974 to 1985 was $1177 million (in 1985 prices). This amounts to about 40 percent of original RDT&E expenditures, or $235,000 per engine (assuming a total Air Force buy of 5000 engines)—equal to 7 percent of the F100 price (in FY84) of $3.2 million. Assuming a 15-year engine life and 10 percent discount rate, this investment is equivalent to an annual cost per engine of about $31,000. Production cost did not change during this period. Maximum thrust was reduced by 3 percent for several years.
>
> Reliability improvements and benefits: For an engine flying program of 200 hours per year, support costs would have to fall by at most $155 per flying hour to justify the CIP investment (if support costs were the only criteria for performing CIP). Air Force and Pratt & Whitney support cost figures indicate that a change of this magnitude was observed in just the three-year period 1982 to 1985. Thus, in support costs alone, CIP investment was a profitable investment. This return to the investment is also reflected in a decline in maintenance manhours per flying hour from 3–4 hours in the mid-1970s to a level of 1.5 a decade later. Similarly,

---

[11] Nix and Shelnutt, 1984, p. 72.

fleet-wide values for unscheduled engine removals fell from average levels of 15 per 1000 flying hours to under four by the end of 1985.

Reliability strategies: The demand for high performance and for early delivery of the F-15 aircraft and its F100 engine precluded the development of a reliable, mature engine at initial operational deployment. Consequently, the Component Improvement Program was used to produce the reliability that had taken lower precedence earlier. Resources, time, experience, and technology contributed to substantial gains in all measures of reliability. However, it took a crisis in operational availability for the Air Force to give priority, attention, and resources to reliability issues.

A central question is not whether CIP was a worthwhile investment (it clearly was in the case of the F100), but whether a higher return can now be obtained from greater expenditures and more time spent in development. Research in the past 15 years on the importance of cyclic stress, and the gains made in restructuring engine tests alter the parameters of the tradeoffs between CIP investments and development, design, testing, and the returns to experience. Engine development programs are now doing more in early development than in post-MQT CIP, relative to the past. This single case provides insufficient evidence to answer the question of how much it pays to invest in reliability earlier in a program.

Appendix E

PHALANX MK15 CLOSE-IN WEAPON SYSTEM

The Phalanx Close-in Weapon System (CIWS) protects against high-speed missiles or aircraft at sea-skimming altitudes. It incorporates the General Electric M16A1 Gatling gun (also used in the Army's Vulcan system), which fires up to 3000 rounds per minute of high-density, 20 mm, depleted uranium projectiles from the gun's six rotating barrels. General Dynamics designed a stand-alone modular package that includes the gun, a search and track radar with a closed-loop fire-control system to track both target and projectiles, and a light-weight structure to protect the automated system. General Dynamics undertook feasibility development and tests for the Navy in the late 1960s; full-scale development ensued in the following years, and the Navy accepted first production deliveries in 1979. Initial installations were completed in 1980. By the end of 1985, more than 300 Phalanx Mk15 systems had been installed in more than 170 U.S. warships. The U.S. Navy plans to fit about 400–500 systems to over 300 ships in 39 ship classes; navies of nine other countries have ordered the Phalanx. From 1969 through 1979, RDT&E investment was $147 million, or approximately $278 million in 1985 prices. Unit cost in 1985 was about $2.5 million.

Within a year after the Phalanx was accepted into the fleet, the Navy became aware of serious reliability problems. Mounted on exposed weather decks, the autonomous system was continually buffeted by salt-water spray, waves, and wind. Its automatic mode of operation did not require constant crew attention; its condition, therefore, was not closely monitored by operating personnel. Moreover, its design did not facilitate maintenance and repair; maintenance personnel, for example, found it difficult to service the electronics equipment under harsh cruise conditions. As reliability problems mounted, the Navy convened a technical review committee to recommend solutions. These were accepted as the so-called "Fowler Fix," named after the chairman of the committee, Admiral Fowler.

The Fowler Fix required design changes for high failure-rate items, structure design changes to provide a better working environment for maintenance crews and better protection for the equipment, and manufacturing quality control improvements including more thorough parts screening, shock temperature tests, burn-in time extension from five hours in 1981 to 20 hours in 1982 to 30 hours by 1983, and an all-up system, high-stress test of 30 hours. These changes were formally incorporated as Ordalt I.[1] Ordalt I changes were retrofitted into fielded systems and incorporated into new production beginning with the 157th item. By the end of 1982, 10–20 percent of the changes were installed; 95 percent of the change was completed by 1983; in 1984, Ordalt I was fully implemented, and a second round of changes, Ordalt II, was initiated.

Ordalt I improvements had a direct and measurable effect on the demand for spare parts. For parts directly affected by design changes, demand fell by 53 percent; the system as a whole experienced an 11 percent reduction in total parts demand. Ordalt II had similar effects: 40 percent reduction in directly affected parts and 9 percent overall. Because of shipboard storage limitations, parts demand has a large effect on system operational availability.

Fleet-wide average MTBF rates are shown in Table E.1. These figures show substantial improvements in all-up system reliability over the period of the design and quality control improvements. Reliability as measured by MTBF increased by more than

Table E.1

RELIABILITY IMPROVEMENT, TOTAL FLEET, PHALANX CIWS

Period[a]	MTBF[b]	MMH[c]	Operational Availability[d]
1981	34.6	.34	.23
1982	54.1	.22	.28
1983	118.5	.12	.41
1984	157.8	.06	.48
1985	137.5	.06	.50

SOURCE: *Nonexpendable Shipboard Equipment Status Log*, NAVSEA 4855, equipment level summary.

[a] Time period is from October 1 to September 30 of designated year.

[b] Mean time before failure (hours).

[c] Corrective maintenance manhours per hour of operating time on longest running unit.

[d] MTBF/(MTBF + downtime) (assumes 100 percent duty cycle).

---

[1] Ordalts (Ordnance Alterations) are the contractual means for backfitting changes to systems in the fleet.

four times from the first installations to the completion of the Fowler Fix. The reduced demands on maintenance crews were just as dramatic, as maintenance manhours per operating hour fell by 80 percent.[2] As a result of longer periods between failures, less maintenance time, and a lower probability of running out of spares, plus such other effects as greater maintenance experience and improved logistics planning, operational availability more than doubled between 1981 and 1985.

The raw fleet-wide averages shown in Table E.1 hide a great deal of variability in the way the Phalanx is used; for some smaller ships, the Phalanx is the only defensive system on board and is in almost continuous operation, while for larger vessels it forms one element of a layered defense and is turned on only when other systems detect potential targets. To more accurately track the effects of the reliability improvement programs across the fleet, the Navy developed standardized MTBF estimates across ship classes. This standardized MTBF experience is shown in Table E.2. These figures tell much the same story as Table E.1. However, in Table E.2, we show an additional measure of the effect of reliability on the logistics systems—the mean time before parts replacement (MTBPR). As noted above, the total demand for parts over the three year period fell by about 20 percent as the average time interval between replacements doubled.

Table E.2

STANDARDIZED MTBF ACROSS SHIP CLASSES, PHALANX CIWS

Period	% Ordalt I Implemented	MTBF[a]	MTBPR[b]
1980–1982	10–20	47	78
1983	95	112	148
1984	100	137	171

SOURCE: Phalanx Program Office.
[a]MTBF based on all-up systems operations, all failures.
[a]MTBPR = mean time before parts replacement.

---

[2] Different parts of the Phalanx system operate for different periods of time. The search radar may operate whenever the ship is out of port, fire-control only when a target is detected, etc. More than 12 clocks or meters record the energized time on major units. MTBF and MMH are based on the time recorded on the longest running meter.

From 1981, about $3 million per year was spent on product improvement developments and other nonrecurring engineering. Phalanx Program Office personnel estimated half of this amount as directly related to reliability issues; the other half of the development and engineering funds was devoted to enhanced maintainability, which indirectly affected reliability and availability. The annual $3 million product improvement costs represented about 60 percent of the total investment in reliability; the contractor absorbed the rest.[3] A rough estimate of reliability improvement investment thus comes to $15 million.[4] This investment is 5.4 percent of the original development cost; spread over 400 systems, it represents an investment of $37,500 per unit, or roughly 1.5 percent of unit costs. Assuming a 10 percent discount rate and 15 year system life, the investment is equal to an annual expenditure flow of $4930.

Total production costs per unit of $2.33 million, which includes the prime contract with General Dynamics and the gun from General Electric, did not change during the period in which the Ordalt I improvements were incorporated. (However, these figures did not include the costs of retrofitting the changes into installed equipment.)

In addition to the investments in nonrecurring development and engineering, the Phalanx Program Office noted other sources of improved reliability, maintainability, and availability for which costs have not been assessed. These include improved training, better parts availability through improved management techniques, countless minor hardware and software modifications, and other management improvements. "These changes are all part of the initial increase of knowledge and growth from an infant weapon system to present."[5] Thus, the system benefitted from a maturation process embedded in a management environment that emphasized reliability issues, where operating experience was incorporated into an ongoing reliability improvement effort.

Increased reliability can drastically influence the inventory stock levels required to support a system's operational availability. To estimate the effect of reliability on stock levels, we have applied a Navy inventory model normally used to plan the parts types and quantities to store aboard ship and at supply points. For their usual inventory

---

[3] The contractor had been experiencing cost underruns, and this cost absorption reduced the amount of the underrun.

[4] This figure was obtained by summing $3 million over the three years 1981–1983, and adding in the 40 percent of the investment incurred by the contractor.

[5] This comment was taken from a Program Office review of an earlier draft of this case study.

planning runs, stock managers specify an inventory value or shipboard volume constraint, and the model calculates an optimum inventory parts list and resulting level of operational availability based on parts reliability, repair times, and supply response times from depots or supply ships. We have run the same model, permitting inventory levels to vary from zero to more than $250 million, with the baseline parts information as of February 1986. We then ran three additional calculations based on assumptions that the 20 least reliable parts had twice the specified reliability levels (half the given failure rate), and did the same thing for the 30 worst parts, and for all parts.[6] The results of the runs are displayed in Table E.3 and Fig. E.1. At an operational availability level of 70 percent, improving the reliability of only 20 parts (out of more than 1000) would save $20 million in inventory investments; doubling the reliability of every part would save $70 million.[7] At a 10 percent discount rate, 15-year system life, and 530 Phalanx systems (as assumed in the model), the inventory savings from doubling the reliability of 20 parts equals almost $5000 per Phalanx system per year. An alternative evaluation, maintaining

Table E.3

OPERATIONAL AVAILABILITY AT ALTERNATIVE STOCK LEVELS AND PARTS RELIABILITIES, PHALANX CIWS, TWO ECHELON STOCK SYSTEM
(Million $)

	Inventory stock level			
Operational Availability (%)	Baseline	20 Parts, 50% Failure Rate	30 Parts, 50% Failure Rate	All Parts, 50% Failure Rate
20	22	18	17	6
30	41	35	32	13
40	62	50	46	22
50	80	66	61	30
60	94	80	75	40
70	120	100	96	49
80	160	140	130	70
90	250	230	210	110

---

[6] The Navy's Availability Centered Inventory Model (ACIM) was developed by CACI, Inc. CACI performed these analyses under our direction. The assumptions of the runs were as follows: 530 systems on 305 ships, mean time to repair of .092 days, supply response time of 8 days, two echelons of stock, repair turnaround time of 180 days; and procurement lead time of 900 days.

[7] In a simple system with no redundancy and independent failures, a doubling of reliability for every part will double system reliability.

the same inventory and allowing availability to change, shows that doubling the reliability of 20 parts would permit availability to increase from 70 percent to 74 percent; about one-quarter of the savings gained by doubling the reliability of every part is achieved by only the 20 worst actors. These returns to improved reliability can be evaluated in several ways. Holding operational availability constant at, for example, 70 percent would permit the $20 million inventory savings obtained from improving only 20 parts to be invested in five new Phalanx systems. Alternatively, the gain in operational availability of four percentage points (an almost 6 percent increase over the baseline value) is equivalent to 30 more Phalanx systems in the assumed case of 530 units. Thus, reliability improvements can be converted directly into increases in military capability at very favorable rates of exchange.

> Cost of reliability: The $15 million reliability improvement program was 5.4 percent of original development cost. Contracted unit production cost did not increase. No performance tradeoffs were necessary to obtain higher reliability.
>
> Reliability improvements and benefits: Measured MTBF increased from 47 hours (in standardized, fleet-wide estimates) to 137 hours. MTBPR more than doubled from 78 hours to 171 hours. Operational availability also more than doubled from 23 percent to 50 percent. Inventory model simulations showed that doubling the reliability of the 20 most unreliable parts could save $20 million in inventory, while holding availability constant at 70 percent; or it could increase availability from 70 to 74 percent at the same inventory level.
>
> Reliability strategies: A reliability improvement program based on fleet experience permitted substantial gains in reliability at fairly low cost. However, it took a near crisis in equipment availability to make reliability a priority for attention and resources.

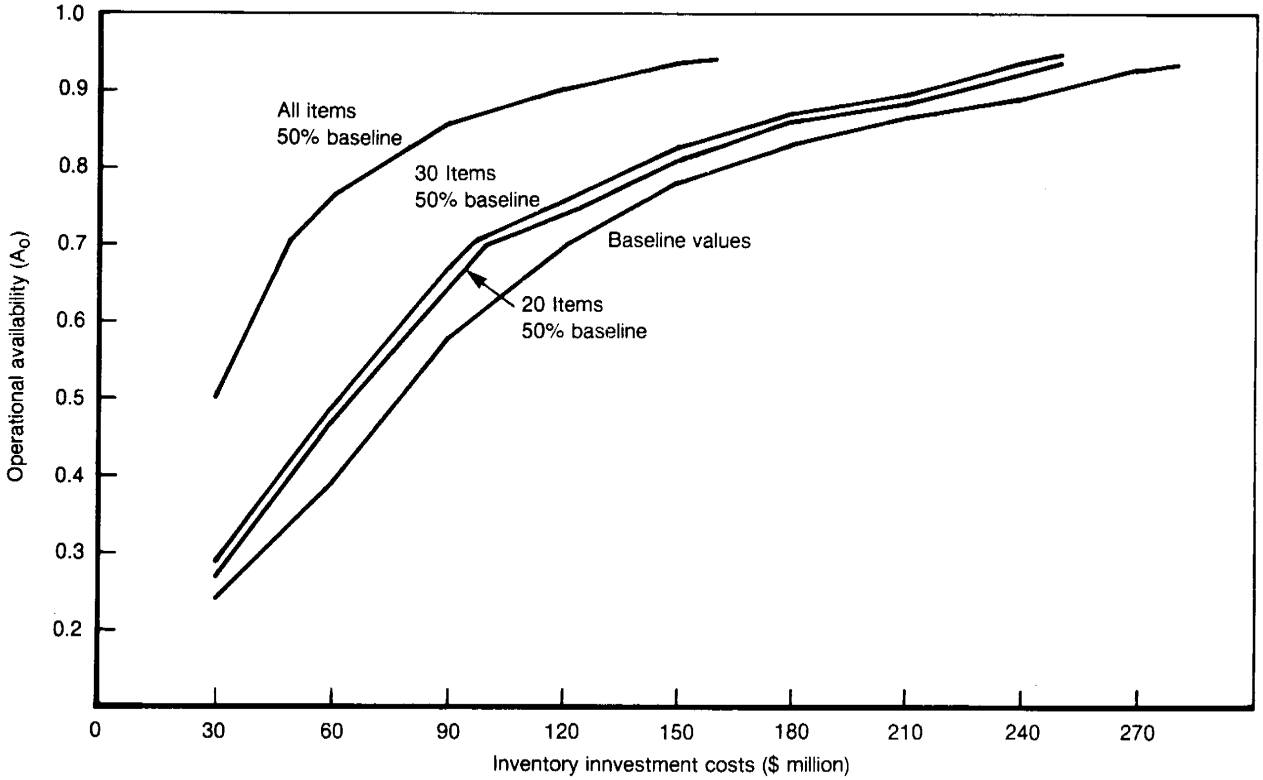

Fig. E.1—Phalanx CIWS operational availability versus inventories for alternative parts reliability factors

Appendix F

RELIABILITY AND REDUCED INVENTORIES FOR THE NAVY LAMPS MK III

The U.S. Navy's Light Airborne Multipurpose System (LAMPS) MK III is an antisubmarine warfare system comprising a Sikorsky SH-60B helicopter (which provides a remote platform for deployment of sonobuoys and torpedos, sensors, and processors) and several shipboard electronic systems. We estimated the effect of parts reliability on shipboard inventories and operational availability by using the same inventory optimizing model that we described above for the Phalanx.[1] Ships stock approximately $4 million of spares for the LAMPS MK III, covering more than 317 items. Increased parts reliability could be used either to reduce spares investment or to increase system availability. As in the Phalanx case, we calculated the effects of 50 percent reliability improvements of the 20 least reliable parts, the 30 least reliable, and all parts. In this case, we look only at the shipboard level of sparing for a single ship. Tables F.1 and F.2 show the results of this calculation.

Improving reliability of just 20 parts of the more than 300 considered in the inventory model allows inventories to be reduced by 7.5 percent, or $300,000, while maintaining equipment availability at 70 percent. A 30-part improvement saves a half million dollars, and doubling the reliability of every part could allow inventories to be reduced by 35 percent. Equivalently, we could preserve the same inventory value at, for example, $4 million and take the gains from the 20-part reliability gain in an increase of operational availability from 70 percent to 74 percent; the 30-part improvement would yield an operational availability of 76 percent. These reliability-related increases in availability are roughly equivalent to a 6 percent and 8.6 percent increase in force structure.

> Reliability improvements and benefits: Doubling the reliability of the 20 most unreliable parts (out of more than 300) would allow shipboard inventories to be reduced by 7.5 percent ($300,000), while maintaining availability at 70 percent; or availability could be increased from 70 to 74 percent, while holding inventories constant.

---

[1] Calculations based on the ACIM were performed for us by CACI, Inc. The inventory was optimized to support the aircraft, the SQQ-28 and SRQ-4 ship electronic subsystems. Mean time to repair was assumed to be .038 days. Mean resupply time was 8.75 days for aircraft parts and 17.5 days for electronics parts.

Table F.1

OPERATION AVAILABILITY AT ALTERNATIVE STOCK LEVELS AND PARTS RELIABILITIES, LAMPS MK III, SINGLE ECHELON STOCK SYSTEM
(Million $)

	Inventory stock level			
Operational Availability	Baseline	20 Parts, 50% Failure Rate	30 Parts, 50% Failure Rate	All Parts, 50% Failure Rate
30	.9	.8	.6	.3
40	1.6	1.4	1.1	.6
50	2.5	2.2	1.8	1.0
60	3.3	2.9	2.6	1.8
70	4.0	3.7	3.5	2.6
80	4.8	4.5	4.4	3.6
90	6.0	5.7	5.5	4.7

Table F.2

SPARES COST AT ALTERNATIVE OPERATIONAL AVAILABILITY LEVELS AND PARTS RELIABILITIES

	Operational Availability (percent)			
Spares Cost ($ millions)	Baseline	20 Parts, 50% Failure Rate	30 Parts, 50% Failure Rate	All Parts, 50% Failure Rate
2	45	47	53	63
3	57	61	65	74
4	70	74	76	84
5	81	84	86	92
6	91	91	92.5	96.5
7	93	94	94.5	97.
8	95.1	96	96.3	97.5

Appendix G

MINUTEMAN I INERTIAL GUIDANCE SYSTEM

When the Minuteman I ICBM was fielded in the early 1960s the MTBF of its inertial guidance system was 600 hours.[1] Because the guidance system was in continual operation, this level of reliability led to an average of 15 repairs per year per silo. Each repair required exchanging guidance units and took seven days for the changeover, during which time the missile was out of service. The 105 days of missile availability lost each year just because of guidance reliability problems were unacceptable to the U.S. Air Force, which launched a major corrective program in 1963.

The program to improve the Minuteman's inertial guidance reliability cost $150 million, or about 50 percent of the original development cost of the system. Reliability rose dramatically as a result of the program: MTBF went from 600 hours to the 9000–10,000 hour level, averaging more than a year without a failure. Availability climbed from 28 percent to 96 percent. Discounted cost savings in parts and crew were estimated at more than $1.5 billion (1965 prices) over the life of the guidance system.

When the bulk of the Minuteman guidance computer was later used on the Titan ICBM, the computer's early MTBF was as high as 27,000 hours, equal to the experience on the mature Minuteman design. The use of previously tested high-reliability parts and designs demonstrated the benefits of evolutionary changes to mature systems that constrained the novelty and uncertainty of a new application.

Despite these substantial returns to the investment in reliability, the investment itself was large relative to the development cost and to the price of the individual unit. If the $150 million reliability program investment were spread over the 1300–1400 guidance units that were produced, the cost per unit would be about $110,000, or roughly 25–30 percent of its $300–400,000 cost of production. Because the depot unit repair cost by itself was more than $30,000 in 1962, the elimination of 15 repairs per year saved more than the original price of the unit in just one year. Moreover, the 15-fold MTBF growth is one of the largest we have observed and the benefits in improved military capabilities alone justified the investment.

---

[1] Information on the Minuteman guidance system was obtained from the files of RAND colleague Hyman L. Shulman.

Cost of reliability: The reliability improvement program was 50 percent of the original guidance system development, or $150 million in 1965 prices ($427 million in 1985 prices). Production costs did not change. Reliability improvement investment costs per unit were 25–30 percent of the production cost.

Reliability improvements and benefits: MTBF increased from 600 hours to 9000–10,000 hours. Availability rose from 28 percent to 96 percent. Repairs fell from 15 to one per year.

Reliability strategies: The Minuteman ICBM development possessed the highest national priority. The goal to deploy it as quickly as possible dominated other program goals. A missile in a silo was believed to possess great military value, even if some of its components proved to be unreliable. As a result, the system that was deployed was immature. Insufficient time and resources had been devoted to reliability. Once deployed, however, the high failure rate of the guidance system generated a readiness crisis, but this crisis was over a fleet of deployed missiles with strategic capabilities. The capability was present, albeit costly to maintain.

The subsequent reliability improvement program was given highest Air Force priority; it was subject to intense scrutiny, which was backed by activities that were carried out in this post-deployment phase could have been undertaken in development. However, the experience of failure modes produced in the operational environment helped to identify problems that may not have been caught in development testing.

Appendix H

CAROUSEL INERTIAL NAVIGATION SYSTEM

When Boeing chose the Carousel inertial navigation system (INS) for its 747 airliner in the mid-1960s, the Carousel became the first certificated commercial INS used in nonmilitary applications.[1] The system was warranted by its producer, the AC Division (now Delco) of General Motors, for 1500 hours of failure-free operation; this guaranteed level was set by contract to rise to 2500 hours four years after introduction.[2] The airline customers based their spares on the warranty assumptions. The manufacturer was responsible for providing any additional spares beyond the warranty levels. Actual experience with the early Carousels demonstrated MTBF figures of about 100 hours. Faced with the cost of supporting the low-reliability equipment, Delco initiated a program of reliability analyses and fault-isolation tests leading to improved designs and manufacturing quality control. The output of this effort, costing $30 million in the late 1960s, was an improved MTBF of 500–600 hours. This reliability improvement program cost twice as much as the U.S. Air Force-sponsored $15 million development effort for the original system.[3] A second reliability improvement program a few years later, costing $15 million, further raised MTBF to 1500 hours. This new model was improved once again in the mid-1970s to yield a 2000 hour MTBF. These data are summarized in Table H.1. Recent reports list the MTBF of the Carousel IV at 2150 hours for military aircraft and more than 4000 hours on commercial types.[4] The high reliability of the Delco Carousel IV stimulated scores of commercial users to specify its use in other Boeing and McDonnell Douglas aircraft; the U.S. Air Force and other defense users have chosen the Carousel IV for the C-5B, C-141, and KC-135 aircraft, the Titan III missile, and almost 30 other military programs. Six thousand Carousel INS sets have been delivered to users.

---

[1] Much of the data for this case was obtained from the personal files of RAND colleague Hyman L. Shulman.

[2] "Airline Navigator Sales Battle Intensifies," *Aviation Week and Space Technology*, November 11, 1968, p. 79.

[3] This $15 million development cost does not include the additional expense of commercializing the original military design.

[4] *Jane's Avionics, 1984–1985*, Jane's Publishing Co., London, New York, 1984.

Table H.1

DEVELOPMENT COSTS AND RELIABILITY IMPROVEMENTS FOR CAROUSEL INERTIAL NAVIGATION SYSTEM

Phase	Year	Cost (million$)	Cost (million 1985 $)	MTBF (hours)
Original development	1960	15	45.0	100
Reliability improvement:				
First phase	1966	30	77.9	500–600
Second phase	1970	15	33.7	1500
Later phases	1975	NA	NA	2000
	1980s	NA	NA	4000 (civil) 2150 (military)

The early price of the system was $110,000 per unit. Three units typically made up a single navigation ship set; but as the system became more accurate and reliable, aircraft operators reduced the number of units in a set to two, and in some cases to a single unit. When the $112 million investment (1985$) in reliability is spread across the 6000 sets, the reliability cost per unit is $18,600. When compared with the original development, the reliability improvements required a total of 248 percent more in price-adjusted resources.

Although we do not have direct evidence of the support cost savings from reliability improvements on the Carousel, a 1974 study estimated such costs through the use of a system support-cost analysis model.[5] This simulation was based on a system with the Carousel's characteristics as used in military aircraft.[6] Doubling MTBF from 100 to 200 hours was estimated to reduce discounted life-cycle costs by $400 million, or by 37 percent. A further doubling of MTBF to 400 hours saved an additional $210 million (30 percent). Increasing MTBF to 500 hours ran into diminishing returns in life-cycle cost savings ($10 million) mainly because of the low number of annual flying hours assumed in the simulation. Total savings in going from 100 hour MTBF to 500 hours was $660 million, or 60 percent of the baseline support costs.

---

[5] McIver, Robinson, and Shulman, 1974, pp. 20–21.

[6] The model parameters were: 2500 aircraft; 35 flying hours per month; $100,000 unit cost of INS system; 10-year life; 10 percent discount rate; repair only at depot, costing $11,000; 24 day repair time. Costs can be doubled to convert to 1985 prices.

Cost of reliability: In the multi-phased reliability improvement effort, the first phase cost $30 million, or 1.73 times the original development. The second reliability effort was about 75 percent of the original, in 1985 prices. Production costs did not change.

Reliability improvements and benefits: MTBF increased by five times in the first effort and then was tripled in the second phase. By the 1980s, reliability again more than doubled. As reliability improved, the commercial demand for this navigation system expanded across aircraft types, commercial airline users, and military applications, especially as the higher reliability levels permitted fewer units to be included in a ship set.

Reliability strategies: Although the Carousel was not driven through development with the same priority as the Minuteman guidance, it had to deal with other complicating factors: The environment of an aircraft was harsher and more unpredictable than the stable, buffered world of a missile in an underground silo; airline reliability requirements were greater than for navigation systems in military aircraft; and price was an important design attribute. Many of the quality-enhancing techniques open to a designer of military equipment were therefore closed to an organization producing for a commercial market. The Carousel designers believed that operational experience was critical to their achieving high reliability. The sheer volume of fleet experience would generate more information in comparatively short periods than could be produced in years of testing. Moreover, testing would not have been able to duplicate the operating environment. Therefore, obtaining and evaluating fleet experience was not an inappropriate reliability strategy in this case, although the level of unreliability early in the program was considerably greater than program personnel had counted on.

Appendix I

SPACECRAFT RELIABILITY COSTS

A RAND study of 23 spacecraft acquisition programs examined the statistical relationship between spacecraft program investments in reliability and spacecraft life.[1] In 30 different equations, no statistically significant relations were found among the observations. The reliability measures used were design life and achieved life. The other variables considered included reliability investment, total program cost, the ratio of reliability costs to total cost, costs per pound, calendar date of program, and recurring and nonrecurring costs. Linear and logarithmic equations were tested.

Several hypotheses—substantive and methodological—can be advanced to account for the total absence of any discernible relationship. The primary explanation is that spacecraft are extremely reliable systems. Their design life varied from one to five years, and the average achieved lifetimes in the analyzed sample ranged from one to eight years. One year of operation without a failure represents an MTBF of 8760 hours; five years equals an MTBF value of over 40,000 hours. System reliabilities this large are rarely observed in conventional military equipment. The amounts spent on reliability were commensurately large: 12 to 30 percent of total program cost (development plus procurement) was invested in reliability. At such high levels of reliability, marginal differences in investment across programs could be overwhelmed by features peculiar to each program. In addition, the number of items produced in any program was quite small (1–10) compared with those of other types of systems; there may be an incentive to invest more in reliability for a smaller program because there is little possibility of substituting quantity for reliability. This reasoning may be especially relevant for space probes where the loss of a single spacecraft could doom the mission.

Methodological problems also may have led to the observed statistical outcomes, although the direction of any bias or effect is not known. The achieved spacecraft life times are truncated at the upper end because a substantial number were still operating at the time of the study. Their observed life times at that point were used as proxies for

---

[1] Personal communication from RAND colleagues, P. Konoske Dey and David J. Dreyfuss.

actual life. Also average program variables were used instead of individual spacecraft information. The use of program averages reduces the sample variance. In addition, only simple, one-variable (or, in a handful of cases, two-variable) models were examined. Some more complex relationships therefore were unexamined. Nevertheless, the simple finding must stand: Out of 30 reasonable equations that were tested, not a single nonrandom relationship was found.

> The cost of reliability: 12 to 30 percent of total program cost was invested in reliability.
>
> Reliability improvements: Achieved MTBF was one to eight years.
>
> Reliability strategies: Large investments during development and production assured highly reliable operations, although no systematic relationship was found between reliability investments and outcomes.

Appendix J

DUANE MODELS

In the early 1960s, J. T. Duane of General Electric put forth a model relating cumulative failures during test or operation of a type of equipment to cumulative operating hours of experience.[1] The central theoretical construct is that a given number of failure mechanisms are inherent in a design and that a fixed proportion of the undiscovered failures are disclosed in proportion to new experience. Duane derived a simple exponential model:

$$\Sigma F / \Sigma H = K(\Sigma H)^{a} \text{ , 0 and}$$

$$\Delta F / \Delta H = (1 + a) K (\Sigma H)^{a} \text{ ,}$$

where ΣF is the cumulative number of failures, ΣH is the cumulative hours of experience at the point of the observation, ΔF is the number of failures within a time interval ΔH, K is a constant, and a is a growth-of-reliability parameter. Duane showed examples of five different programs that fitted this model very closely, with exponents of about –0.5. Because MTBF is the inverse of the failure rate, the marginal or current value of MTBF will rise with cumulative experience according to the positive value of the exponent.

Most of the statistical studies in the aerospace industry based on the Duane model involve estimation of the cumulative version of the model rather than the marginal because "most of them fit so well that it looks as if the data was rigged."[2] As noted above in the F100 case, one must be especially cautious about analyses based on cumulative variables; statistically independent observations will show a high correlation when their cumulative values are plotted against each other. Even if there is a true relationship between the variables, the estimated value of "a" be unreliable when the cumulative values are plotted against each other; this lack of reliability is masked by high correlation coefficients.

---

[1] This was originally published as a General Electric technical report: Duane, 1962.
[2] Codier, 1968.

To test the robustness of Duane's results, we extracted the original data from four of his five plots and analyzed the marginal values of failure rate rather than cumulative values. The marginal correlations were generally not as strong as the original cumulative plots, and the reliability growth parameter deviated, sometimes significantly, from the originally calculated value.[3] Because the marginal estimates do not suffer from the estimating problems inherent in the aggregate estimates, the marginals should be the preferred technique for predicting reliability growth.

Despite this statistical problem, the results of Duane and those who have taken his analytical lead strongly suggest a regular relationship between experience and failure rates. When Duane's original work is combined with several others, analysts have noted an intriguing result: The reliability growth exponent takes on values of –.1 to –.3 for programs with "low" reliability efforts, and values of –.5 to –.6 for "high" reliability efforts.[4] That is, failure rates fell two to three times faster in programs where reliability was a concern than where little attention and effort were given to reliability. For the "high-effort" programs, failures fell (or MTBF rose) at a rate somewhat greater than the square root of test and operational hours (if test hours were increased four times, MTBF doubled). Unfortunately, the several referenced studies do not define or provide measures for high and low reliability efforts. However, the results shown in Appendix K suggest that such program elements as failure mode analyses are found in "successful" reliability programs.

We can use the results of Duane model analyses to consider the differences between high and low reliability programs with different amounts of testing. The F100 and F404 engines can serve as examples to motivate such a comparison. The F404 failure rate (as measured by unscheduled engine removals) was only 20 percent that of the F100 at a similiar early stage of deployment. The F404 had 14,000 hours of testing at model qualification, and the F100 had about 11,000 hours. We assume that the constant term *K* in the Duane model was the same for both engines, and that the F404 had a growth parameter "a" of –0.5, which appears to be typical for a high reliability effort program. These assumptions imply a growth parameter of –0.34 for the F100, which is in the range estimated for programs with less emphasis on reliability. If we push these hypothetical calculations further, we could consider the reliability that may come out of a

---

[3] The average estimate of a was –0.5; the range of deviations was 0.2.

[4] These results were brought to the authors's attention by RAND colleagues M. R. Davis, M. Kamins, and W. E. Mooz.

program like the T700 engine, which had 18,000 test hours at qualification and 42,000 hours at first deployment. Using the same reliability growth parameter as assumed for the F404 engine, the additional testing before qualification would produce an MTBF 27 percent longer; the additional pre-deployment test program would extend that period by another 50 percent.

Table J.1 summarizes the above calculations across three testing programs and three reliability growth parameters, representing low, "standard," and high reliability priority programs. The estimates are normalized to a typical engine program of the 1960–1970 period. According to these estimates, reliability gains of 15–20 percent can be obtained from sharply increased testing programs; but much more substantial gains of 600–700 percent are possible from vigorous, high priority programs emphasizing reliability.

> Cost of reliability: Empirical estimates of Duane models indicate that reliability increases with testing and experience exponentially, with exponent values of 0.2 to 0.6. Within a class of products—turbine engines, for example—the range between programs with low and high priority on reliability appears to be about 0.3 to 0.5.

> Reliability strategies: For parameter estimates derived from turbine engine programs, the priority devoted to reliability has greater effect than the number of test hours. These estimates do not assess the cost of priority compared with the cost of testing.

Table J.1

FAILURE RATE: RATIO TO A STANDARD PROGRAM[a]

	Reliability Growth Parameter		
Test hours	.3	.4	.5
11,000	2.53	1.00	.39
14,000	2.36	.91	.35
18,000	2.19	.82	.31

[a] Ratios calculated from Duane model with the value of the constant, K, assumed equal for different estimates.

Appendix K

RELIABILITY COSTS IN NAVAL SYSTEMS

The Naval Sea Systems Command sponsored a study in the mid-1970s to investigate the costs and programmatic efforts associated with reliability, maintainability, and manufacturing quality assurance in 22 naval systems.[1] Questionnaires and follow-up interviews lasting several days developed detailed cost information on activities related to reliability, maintainability, and quality on 19 of these systems; in addition, descriptions of detailed development efforts were obtained in order to determine whether specific actions were associated with successful reliability programs. The studied systems included sonars, guns, radars, missiles, fire-control systems, mines, torpedos, and engines.

Reliability programs during development were classified as being either successful or unsuccessful. Successful systems achieved "satisfactory" reliability growth during development and early production without extensive redesign. Unsuccessful systems failed a reliability demonstration test, required additional funding for extensive redesign, or did not achieve reliability growth meeting specified requirements.

One of the main conclusions of the report was that those systems categorized in the satisfactory reliability group "invariably had higher expenditures on reliabilty, maintainability, and quality activities than unsatisfactory systems in the sample; the higher the expenditures, the higher the achieved levels and the more rapid the reliability growth."[2] The ten satisfactory programs spent an average of 12 percent of total program development costs on reliability, maintainability, and quality activities, compared with only 8 percent for the nine systems in the unsatisfactory category. More telling than these simple averages is the fact that there was virtually no overlap in the spending figures: Only one satisfactory system spent less than the average of the unsatisfactory systems, and only a single unsatisfactory system spent more than the average for the other group.

---

[1] Evaluation Associates, Inc., 1978.

[2] Ibid., p. 3.

Additional analyses were performed on homogeneous subsamples. In particular, expenditures per part in the eleven electronics systems were related to the demonstrated failure rate. In the satisfactory group of electronics systems (with a failure rate per part of 0.18 failures per million hours), $65 per part was spent on reliability and quality, versus only $37 in the unsatisfactory systems, whose failure rate was 0.61. A semi-logarithmic equation was fit to these data:

$$n(F/N) = 7.0 - .0214(R + Q)/N \quad ,$$

where F is failure rate per 10^9 hours, N is the number of parts, and $(R + Q)$ are the expenditures on reliability and quality.[3] This equation generates predictions of reliability for different levels of expenditures, as shown in Table K.1. The equation shows an increasing proportional return to investment in reliability, although the absolute gains per dollar decline.

The Navy study produced important qualitative conclusions, in addition to the quantitative results. Deficiencies in reliability programs in the areas of design, parts screening, and tests were found in all program areas among the unsatisfactory systems. For example, failure mode analyses were performed in all but one of the satisfactory systems, and in only one of the unsatisfactory.[4] In addition, the satisfactory programs performed more fault tree analysis, design reviews, stress analysis, maintenance of approved vendors and parts lists, control of nonstandard parts, failure budget analysis,

Table K.1

RELIABILITY AS A FUNCTION OF INVESTMENT IN ELECTRONIC SYSTEMS

Investment per Part in Reliability and Quality ($)	Failures per 10^9 hours[a]
10	900
20	700
40	500
60	300
80	200

[a] Failure rate is rounded to the nearest hundred.

---

[3] Although statistical measures of goodness of fit were not given in the report, we calculated a value for R^2 of .80 from the original data.

[4] Evaluation Associates, Inc., p. 69.

and other reliability-oriented management. Thus, the additional expenditures on reliable systems were directed toward program activities designed to enhance reliability outcomes.

When the Navy emphasized reliability, it was able to influence costs and performance by contractual requirements, particularly through the demand for reliability demonstration tests. In general, systems with satisfactory reliability had demonstration tests; the unsatisfactory systems did not. Most important, the customers' attention to reliability from the beginning of the program demonstrated their real priorities. The report noted, "The histories of the systems studied make it clear that specified reliability is frequently traded away during contract negotiations in favor of reduced costs or more rapid delivery."[5] When the buyers worked out clear and meaningful reliability goals, specified them in contracts, and emphasized them during development, the developers performed various engineering activities related to reliability, performed demonstration tests, spent more resources, and achieved better reliability results.

> The cost of reliability: Programs with "satisfactory" reliability outcomes spent on average 50 percent more on reliability, quality, and maintainability activities than "unsatisfactory": 12 percent of total development efforts versus 8 percent. For electronic systems, failure rates per part fell by more than two-thirds when reliability expenditures per part increased by 75 percent from \$37 to \$65; converting failure rates to MTBF shows an increase of MTBF per part from 1.64 million hours to 5.55 million hours.
>
> Reliability improvements and benefits: Satisfactory programs, by definition, had good reliability growth during development and early operations without extensive redesign. Unsatisfactory programs did not meet requirements and additional funding and extensive redesign were necessary.
>
> Reliability strategies: Not only did the satisfactory programs spend more on reliability, they performed many more tasks specifically directed toward improved reliability than did the unsatisfactory programs. But these activities were undertaken only when the Navy emphasized reliability by contractual requirements for both reliability goals and specified activities and tests. Moreover, the customer had to demand adherence to these requirements throughout the development effort.

---

[5] Ibid., p. 4.

BIBLIOGRAPHY

Abell, John B., Thomas F. Kirkwood, Robert L. Petruschell, Giles K. Smith, *The Cost and Performance Implications of Reliability Improvements in the F-16A/B Aircraft*, The RAND Corporation, N-2499-AF, March 1988.

Alexander, Arthur J., and Bridger M. Mitchell, *Measuring Technological Change of Heterogeneous Products*, The RAND Corporation, R-3107-NSF, May 1984.

Asher, Norman J. et al., *Changes in Helicopter Reliability-Maintainability Characteristics over Time*, Vol. III, Institute for Defense Analyses, AD-4014469, Study S-451, March 1975.

Birkler, J. L., S. M. Cote, and J. L. Byers, *The Propulsion System Duty Cycle: The Navy's New Look*, AIAA Paper 77-883, July 1977.

Central Intelligence Agency, *The Soviet Weapons Industry: An Overview*, DI86-10016, September 1986.

Codier, Ernest O., "Reliability Growth in Real Life," *Proceedings of the Annual Symposium on Reliability*, IEEE, 1968.

Duane, J. T., *Reliability Monitoring for Development Projects*, G.E. Technical Information Series, DF62MD300, February 1, 1962. Subsequently issued as *IEEE Transactions on Aerospace*, Vol. 2, No. 2, April 2, 1964.

Evaluation Associates, Inc., *Reliability, Maintainability, and Quality Engineering Costs in Naval Systems Development* (published by Naval Sea Systems Command), T0800-AA-GIP-010, June 1, 1978.

Fatkin, Allan, *Generic Inflation Indexes for Weapon Systems*, Air Force Systems Command, Cost Analysis Improvement Group, Report NR1, July 1981.

Genet, Russell M., "Generalized Conclusions Concerning Life Cycle of Air Force Inertial Navigation Systems," *Sixth Data Exchange for Inertial Systems*, Aerospace Guidance and Metrology Center, Newark Air Center, Tobyhanna, Pa., October 1972.

Large, Joseph P., Harry G. Campbell, and David Cates, *Parametric Equations for Estimating Aircraft Airframe Costs*, The RAND Corporation, R-1693-1-PA&E, February 1976.

Lee, Allan D., "A Strategy to Improve the Early Production Phase in Air Force Acquisition Programs," The RAND Corporation, P-6941-RGI, December 1983.

McIver, D. W., A. I. Robinson, H. L. Shulman, with W. H. Ware, *A Proposed Strategy for the Acquisition of Avionics Equipment*, The RAND Corporation, R-1499-PR, December 1974.

Nix, John, and Riley Shelnutt, "Behind the Alternate Fighter Engine Competition," *Aerospace America*, May 1984.

Rapp, G., *The F404 Development Program—A New Approach*, AIAA-82-1180, 18th Joint Propulsion Conference, Cleveland, Ohio, June 1982.

Rich, Michael, and Edmund Dews, with C. L. Batten, *Improving the Military Acquisition Process*, The RAND Corporation, R-3373-AF/RC, February 1986.

Rich, Michael, and S. M. Drezner, *An Integrated View on Improving Combat Readiness*, The RAND Corporation, N-1797-AF, February 1982.

Scribner, B. L., D. A. Smith, R. H. Baldwin, and R. L. Phillips, "Are Smarter Tankers Better? AFQT and Military Productivity," *Armed Forces and Society*, Vol. 12, No. 2, Winter 1986.

Selling, Albert L., *F404 Engine Reliability Program*, PSAE Technical Paper Series No. 83159, October 1983.

Sherman, Allen, and Stanley Horowitz, "Maintenance Costs of Complex Equipment," *Naval Engineers Journal*, December 1979.

U.S. Senate, Committee on Armed Services, *Hearings on F-15 and F-16 Engine Problems*, 96th Cong., 1st Sess., November 27, 1979.

U.S. Army Training and Doctrine Command, *Medium Life Helicopter CH-47 Modernization Program, Concept Formulation Package*, COEA, Vol. V, Appendix O (Cost Analysis), 1975.