THE MX MISSILE TEST PROGRAM

Staff Working Paper

January 1986

The Congress of the United States
Congressional Budget Office

PREFACE

Most of the MX missiles the Air Force plans to buy are earmarked for the test program, which establishes and monitors system capability and reliability over the system life. Thus, though the Congress halved the planned deployment of MX, the total system's size--and hence its cost--has not fallen in proportion. As requested by the Senate Committee on Armed Services, this paper analyzes the basis for the planned test program--determined largely by statistical guidelines established by the Joint Chiefs of Staff. In accordance with the Congressional Budget Office's mandate to provide objective analysis, the paper makes no recommendations.

Bonita J. Dombey of CBO's National Security Division prepared the study, under the general supervision of Robert F. Hale and John D. Mayer, Jr. Jeffrey A. Merkley of CBO assisted with statistical analysis; Joel Slackman also provided assistance. Johanna Zacharias edited the manuscript.

Rudolph G. Penner
Director

January 1986

NOTE

Unless otherwise specified, all costs are expressed in current dollars. All dates refer to calendar years.

SECTION I. OVERVIEW

The MX, a highly accurate intercontinental ballistic missile (ICBM) carrying ten nuclear warheads, has long been a source of contention. Debate centers on the 195,000-pound missile's production, basing, and cost. These issues have been put to rest, at least temporarily, with fiscal year 1986 Congressional action specifying that no more than 50 MX missiles be deployed (in Minuteman silos) at any one time. Twelve MX missiles were authorized for fiscal year 1986, bringing the total number authorized thus far (including 20 research and development test missiles) to 74--enough to complete the deployment.

Though the planned deployment of MX has been numerically halved, from 100 to 50, the total purchase of missiles--and hence the system's cost--has not been cut by a proportionate amount. The reason for this disparity is that the program for testing the MX missile is not subject to change as the planned deployment decreases. Of the total 193 MX missiles that have been or will be bought (including Research and Development missiles), 143 are designated for testing. All 119 missiles still to be bought are earmarked for the test program. Specifically, the Congress provided that from 12 to 21 missiles be procured for this purpose in fiscal year 1987.

In this paper, the Congressional Budget Office examines the basis for that test program, which will consume about three-fourths of all MX missiles purchased. The study focuses on Operational Testing and Evaluation (OT&E), which constitutes the largest segment of the test program and is to proceed in two phases:

- Phase I of OT&E follows Research and Development (R&D) testing and is designed to establish an estimate--with a high degree of confidence--of the accuracy and reliability of the system. This baseline estimate is important both for operational and target planning, and for monitoring the MX system's capability over time. The Air Force program calls for 24 missiles to be tested over three years (1988-1990) for Phase I.

- Phase II of OT&E, designed to monitor degradation in system performance and reliability, then tests on a yearly basis over the remaining system life. A total of 84 missiles, or about seven per year, are planned to be tested over the 12 years (1991-2003) of Phase II.

Analysis of Current Plans

The Joint Chiefs of Staff have established general guidelines for weapons systems included in the strategic nuclear war plan--also known as the Single Integrated Operational Plan (SIOP)--designed to ensure an acceptable level of testing. These guidelines form the foundation for the sizing of a ballistic missile test program. They specify annual testing to maintain the statistical level of confidence required in each service's estimate of weapon system reliability. However, the guidelines leave the statistical analysis and details of the test program to the individual services. The services often prefer to test to more strict criteria than specified by the JCS, especially in Phase II where the guidelines are less stringent.

The CBO analysis finds that the 24 missiles allocated by the Air Force for Phase I testing might not suffice to meet the statistical guidelines set out by the JCS and the Air Force's Strategic Air Command (SAC). Assessment based on the Air Force's assumptions and Classical statistical techniques indicates that at least 36 missiles would be needed. The Air Force rationale for using only 24 missiles for Phase I testing, however, rests on experience with land-based missiles and a judgment that, in light of constraints on the total test program, additional test missiles should be shifted to Phase II of the program.

The seven missiles per year allocated to Phase II slightly exceed numbers necessary to meet JCS guidelines as interpreted by the Air Force: about six missiles per year would suffice. The Air Force wants more Phase II missiles, because it believes JCS guidelines for Phase II are not demanding enough, and also because the Air Force wants flexibility to meet diverse needs. Extra test missiles would be needed, for instance, if the MX remained in service longer than the 15 years currently planned--as has happened with past generations of U.S. land-based missiles--or if the MX at some point requires a major modification. Minuteman II missiles have been deployed for about 20 years, and although there are only 26 test missiles remaining, the Minuteman is likely to remain deployed through the end of this century.

Given the objectives of a ballistic missile test program, when Phases I and II are considered together, the MX test program seems modest in size. Phase I is actually too small as judged by Classical statistical techniques; the extra missiles in Phase II might be needed if the MX stays in the inventory for more than 15 years. The MX test program is also modest in size compared to test programs for other U.S. ballistic missiles.

SECTION II. WEAPONS SYSTEM TESTING

Why Test?

Any weapons sytem will undergo certain expected and other unexpected degradation over time. The general objective of a test program is to establish and monitor a system's capability and reliability--that is, its ability to perform with the precision, speed, and aim that its operators count on in planning. To assure a weapon's continued reliability, continued monitoring is necessary. Ancillary benefits of a system's test program include crew training and however much deterrent effect might result from public demonstrations of confidence in the system's effectiveness.

A test program is organized in segments or phases. Research and Development testing generally consists of controlled testing of specific components and capabilities. Operational Testing and Evaluation, the largest segment, involves flight tests that closely simulate the operational environment; these test the viability of the overall system. Aging and Surveillance testing involves extensive ground testing to detect stress and age-related defects before they might impede performance.

None of the objectives of a test program is affected by the number of missiles to be deployed. Under each of the formal plans for MX deployments--ranging from as many as 200 missiles to as few as 50--the size of the test program has remained constant.

How Much Testing is Enough?

There is no objectively "best" amount of testing. Significant costs are associated with testing. And ballistic missile testing is in some sense paradoxical. Since a ballistic missile is destroyed in the testing process, the missiles that remain operational are obviously not those that have been tested. Thus no amount of testing can assure how a system would perform. Testing, however, should be sufficient to give planners confidence in the statistically predicted reliability and performance of an overall system.

To make this process less vague, the Joint Chiefs of Staff establish minimum guidelines for operational testing of strategic systems employed in a nuclear war plan, also known as the Single Integrated Operational Plan (SIOP). These guidelines, explained in more detail in the next section, are designed to ensure that the services maintain an acceptable level of statistical confidence in their estimate of system reliability. Thus, they

form the foundation for the sizing of a test program. However, the guidelines leave the statistical analysis and details of the test program to the individual services. Furthermore, these guidelines are described as lower bounds for confidence and reliability; the services might impose more stringent criteria for testing than specified by the JCS.

SECTION III. THE AIR FORCE'S MX TEST PROGRAM

Each of the three segments in the test program for the MX missile has specific objectives, but all are intended to establish and monitor the capability and reliability of the system. The test program involves monitoring the operational ground equipment, airborne equipment, and warhead. Included in these systems are numerous subsystems, such as the booster, post-boost vehicle, arming and fuzing mechanisms, guidance subsystem, and flight-control subsystem.

RESEARCH AND DEVELOPMENT TESTING

Testing starts while a weapons system is still at the R&D stage. The goal is to develop a missile that meets specified performance standards for such features as range and accuracy. Now under way, the MX missile's R&D test program includes 20 flight tests. Objectives in this phase evolve from limited early developmental testing through fully integrated operational testing. To date, ten MX flight tests have been conducted, with the latest two carried out from a modified Minuteman silo at the Vandenberg, California, test range. The current plan calls for completion of 16 flights before Initial Operational Capability (IOC) of the system in December 1986, with the remaining four flight tests leading into the Operational test phase of the program. 1/ According to a recent report by the General Accounting Office, the MX's flight test program to date has shown good performance. 2/ Problems that have occurred have centered on the third-stage extendable nozzle exit cone (ENEC), which gives the missile an important increment of range. This mechanism failed the first, third, and seventh flight tests. Later modifications appear to have solved the problems.

---

1. Initial Operational Capability for the MX is defined as deployment of the first ten missiles.

2. General Accounting Office, "Status of the Intercontinental Ballistic Missile Modernization Program (GAO/NSIAD-85-78)," July 8, 1985, p. 17.

OPERATIONAL TESTING AND EVALUATION

After successful R&D testing, Operational Testing and Evaluation proceeds in two phases over the life of the system. These tests are designed to simulate as closely as possible what the Defense Department terms "the operational environment" in order to ascertain the missile's ability to deliver its payload reliably and accurately. The United States does not test ICBMs from actual, operational silos, however, but rather, from launch facilities at test ranges--such as Vandenberg--designed to represent the operational silos.

Phase I

The first phase of OT&E is primarily concerned with establishing a statistical estimate, with a high degree of confidence, of the system's accuracy and reliability. 3/ This estimate serves for both operational and target planning, and it establishes an accurate baseline against which to measure any future degradation. When problems are uncovered in a missile test, they are corrected, if possible, before the next missile flight test. Although this does not make for an ideal statistical sample, it is an operational reality.

Owing to reductions by Congress in planned procurement rates, and a decision to maintain the initial deployment in December 1986, Phase I Operational testing for MX will not begin until fall 1988--more than a year after completion of R&D. The planned Phase I OT&E program for the MX will test 24 missiles over a three-year period (1988-1990).

Phase II

On successful completion of Phase I, the second phase of the Operational Test program begins with annual testing over the life of the system to monitor degradation in performance and reliability. 4/ A sample of missiles

---

3. Weapon system reliability refers to the proportion of successful launches--with success defined as the launch, completion of flight, and delivery of the warhead. Accuracy of the weapon is measured by its Circular Error Probable (CEP). CEP is defined as the size of a circle (its radius) centered on the target within which the warhead has a 50 percent probability of impacting.

4. If there are unexpected difficulties in Phase I, it may be extended and some planned Phase II missiles used for additional Phase I testing.

is randomly selected--and replaced--from the deployed force each year, taken to the test range at Vandenberg--along with a task force of technicians and crews from each missile's home base--and flight-tested there. A total of 83 missiles, about seven a year, are planned for testing during Phase II of the MX program.

Aging and Surveillance

The Aging and Surveillance component of the test program proceeds at the same time as OT&E testing, and it is designed to detect any aging and deterioration of components before the system shows signs of actual degradation. This is accomplished by extensive ground testing. The MX test program calls for one Aging and Surveillance missile a year.

In the past, Aging and Surveillance testing has been instrumental in the early detection of age-related defects. For example, the deterioration of the bonding material between the motors and casing of Minuteman II, Stage 2, and Minuteman III, Stages 2 and 3, currently being corrected, was discovered this way. The program has also been instrumental in confirming that Stage 1 of both missiles is sound.

Most of Phase II OT&E and Aging and Surveillance testing is conducted after production of the missile has been completed. Thus, significant deterioration problems might require the manufacture of new components or sub-components.

Even though the MX is similar in many ways to current U.S. ICBMs, it will still need thorough testing. Many of the missile components that have deteriorated and caused reliability trouble with the existing Minuteman missiles have completely new counterparts with the MX, and they might or might not be the source of new problems. For instance, the missile guidance system, the component causing the greatest reliability flaw with the aging Minuteman missiles, is a new advanced system for the MX using inertial reference. The motor casing for the second and third stages is a new graphite composition that might wear exceptionally well with time, or might suffer from unexpected brittleness. These types of problems are difficult to predict. Eight components on the MX were not used on previous ICBMs.

- Stages I-III Kevlar Motor Case
- Stage II and III extendable nozzle exit cones
- Stage IV surface tension propellant tank
- AIRS guidance system

- Unique Signal Device
- MK21 Reentry Vehicle
- Solid state radar fuze
- Aluminum/composite "aeroshell" reentry vehicle structure

THE SIZE OF THE MX TEST PROGRAM COMPARED TO OTHERS

Compared to the size of test programs for five other U.S. ballistic missiles, the MX program is modest in scale (see Table 1). The number of missiles designated for the MX test program is at or very near the lowest of any of the test programs. 5/ Flight testing during Phase II of Operational testing for the MX is set at about seven missiles a year, as compared, for example, with 12-16 for the Trident II.

The MX test program might be considered especially modest in size if, as has been the case with past ICBMs, their actual service life has exceeded the Air Force's initial estimate for the system (see Table 1). The Minuteman II only has 26 test missiles remaining and might be deployed through the end of the century. The last operational test of a Titan missile was carried out in 1969, yet the last of these missiles is not to be retired until 1987.

THE JOINT CHIEFS' STANDARDS FOR OPERATIONAL TESTING

Though comparisons can help judge the magnitude of the MX test program, a key test of adequacy is the ability to meet guidelines issued by the Joint

---

5. The Navy's test programs for the Trident I (C-4) and planned Trident II (D-5) are much larger than the other test programs. Navy test programs have additional considerations that might make direct comparisons misleading, such as the assumption of a 30-year service life, the use of multiple--"ripple"--launches to test crew and submarine adequately, and the necessity to examine missile performance over a broad spectrum of ranges and azimuths. These issues are examined in more detail in a CBO companion analysis of the Trident II test program. The Army test program for the Pershing missile would provide a better comparison, but the Army has classified its numbers.

TABLE 1. COMPARISON OF TEST PROGRAMS FOR SIX BALLISTIC MISSILES

	Numbers of Missiles					
	MX	MMI	MMII	MMIII	Trident I	Trident II
Total Test Missiles	143	290	205	284	263	418
Research and Development Tests	20	56	20	25	25	30
Operational Tests	108	210	171	242	208	356
Aging and Surveillance Tests	15	24	14	17	30 a/	30 a/
	Year Ending					
Original Projected Service Life	2003 b/	1969	1980	1985	1998	2029

a. In the case of the Navy, these are "pipeline" missiles; missiles that are in the process of being inspected and redeployed as submarines are overhauled. They include so-called Service Life Evaluation missiles that are similar to Aging and Surveillance missiles.

b. Service life is technically measured from Initial Operational Capability. However, in the case of the MX, the 15-year Operational Test program was delayed two years to 1988. Thus, Operational Testing is planned to extend over the period 1988-2003 versus 1986-2001. According to the Air Force, then, the MX system will likely be deployed beyond its technical service life of 15 years.

Chiefs of Staff. This section examines Operational Testing--the largest segment of the test program--to which JCS guidelines pertain. (This section does not examine other phases of testing. Research and Development testing is well under way, and the missiles for those tests have already been purchased. No statistical requirement exists for Aging and Surveillance test missiles; from a practical perspective these missiles have been instrumental in detecting age-related problems, and so devoting one missile a year to this type of testing might be reasonable.)

The CBO has analyzed the test program for the MX that would meet JCS guidelines--which focus specifically on the missile's reliability--in the two phases of Operational Testing. Testing will also establish the accuracy of the MX, but the JCS guidelines do not explicitly deal with accuracy; hence, testing for accuracy does not explicitly influence the size of the MX test programs. 6/

Phase I

The JCS guidelines require a high degree of confidence in the baseline estimate of overall system reliability developed during Phase I. Specifically, the JCS requires 90 percent confidence (in statistical terms) that overall system reliability be within 10 percentage points of the reliability observed in the sample tested. Reliability refers to the proportion of missiles that would be successfully fired; that is, missiles that would launch, complete all phases of flight, and deliver their payload. For example, if the Air Force tests 30 missiles with 25 successful results (reliability of 83 percent in the sample), it would have to demonstrate by statistical analysis that, 90 percent of the time, it would be correct in stating that the system's true reliability was at least as good as 73 percent. 7/

---

6. With Navy missiles, test requirements imposed by the Navy for accuracy do affect the size of the program.

7. Note that this guideline makes sense only for reasonably high degrees of reliability. At 10 percent reliability, for example, the criterion is met but is not meaningful. But ballistic missile reliability is assumed to be well above 50 percent.

For a given standard of reliability, the larger the sample size observed to meet that standard, the higher the confidence one can have that the observation accurately reflects overall system reliability. An example might clarify the principle. In five tosses of a fair coin, the probability of getting "heads" four times is about 16 percent--that is, about one-sixth of the time, one would be mistakenly led to predict that one has an 80 percent probability (four out of five) of getting "heads." But in ten tosses, the probability of getting "heads" a comparable eight times is only 4 percent--making it much less likely that one would be led to the same erroneous conclusion.

Thus, for a given observed reliability, only certain sample sizes can provide the level of confidence required by JCS guidelines in making a prediction about overall--or "true"--system reliability. The required sample size depends on the expected results. The Air Force most likely assumes, for planning purposes, that the reliability of the sample tested will be around 90 percent. (This is derived from JCS statistical guidelines and the original Air Force plan for 36 missiles in Phase I using Classical statistical techniques. It is also consistent with public estimates of missile reliability.)

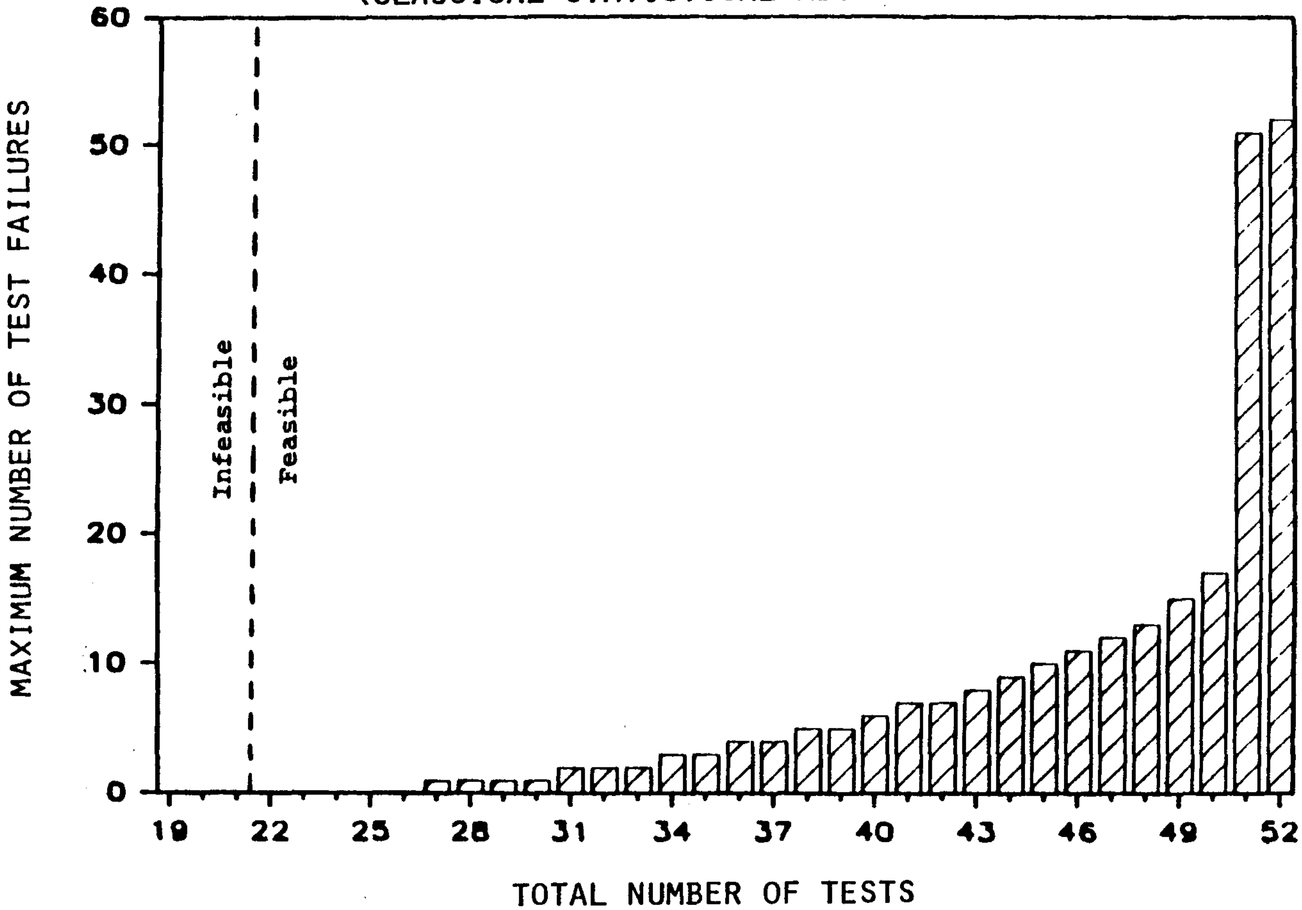
The statistical method that the Air Force chooses to use can also affect the sample size required to meet JCS guidelines. The remainder of this section looks at Phase I test program requirements, first using the "Classical" statistical method, and then using the "Bayesian" statistical method.

<u>Classical Method</u>. For various test sample sizes, Figure 1 shows the maximum number of test failures allowable for a Phase I test program that meets JCS guidelines under the Classical method. The figure illustrates that sample sizes must be larger than 21 missiles to have any prospect of satisfying JCS guidelines. (Again, the guidelines pertain to missiles with fairly high expected reliabilities; more than 50 percent. At very low reliabilities, the guidelines can always be met, but are not meaningful.) For sample sizes of 22 to 26 missiles, it is possible to satisfy JCS guidelines only if every trial is a success, indicating 100 percent reliability. On the other hand, with a large enough sample size--in this case, 51 missiles--one can always be confident that the success rate in the sample accurately represents the system's true reliability. In this case, one always meets JCS guidelines regardless of what the success rate is.

With the assumption that reliability of the sample will be around 90 percent, the smallest sample size that offers a chance of meeting JCS

FIGURE 1. MEETING JCS GUIDELINES FOR MISSILE TESTING

(CLASSICAL STATISTICAL METHOD)

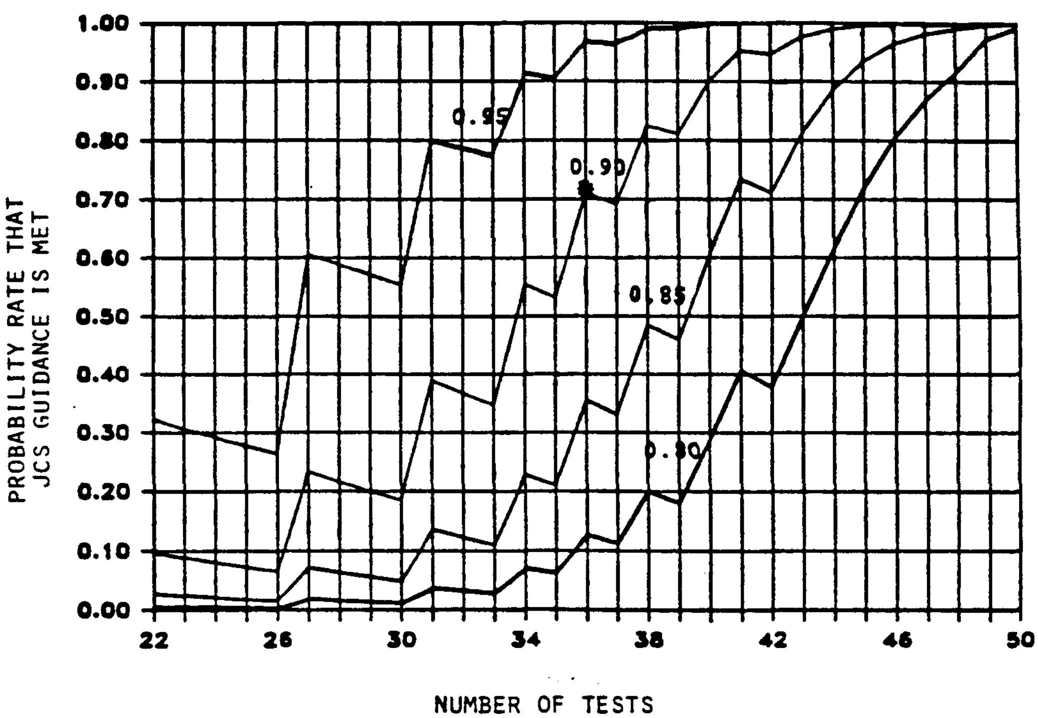


This graph displays the maximum number of failures that can occur in a Phase I missile flight test program that meets JCS guidance:

-- below 22 missiles, it is not possible to meet JCS guidance
-- at 40 missiles, 6 or fewer failures satisfies JCS guidance
-- above 50 missiles, any number of failures satisfies JCS guidance

guidance for Phase I is 36 missile tests with 32 successes. This minimum of 36 missiles exceeds the Air Force's plan for 24 missiles in Phase I. Moreover, a sample of 36 missiles does not guarantee meeting JCS guidelines. That only happens if, when the 36 missiles are tested, at least 32 are successful. But even if the overall system (true) reliability were around 90 percent, the particular sample tested would have at least 32 successes only about 70 percent of the time. Thus, even with a sample size of 36 test missiles, the Air Force would implicitly be accepting about a 30 percent risk of not meeting JCS guidelines. Figure 2 illustrates the probabilities, with different sample sizes, of meeting JCS guidelines given overall system reliabilities of between 80 percent and 95 percent.

FIGURE 2. PROBABILITY OF MEETING JCS GUIDELINES
(GIVEN TRUE RELIABILITY = 0.8...0.9)



* Even with a sample size of 36 missiles, the Air Force is implicitly accepting a 30 percent risk of not meeting JCS guidelines.

Bayesian Method. To justify a choice of 24 test missiles for Phase I in light of these results, the Air Force has applied a Bayesian statistical analysis. Bayesian analysis provides a formal framework for incorporating information other than sample information into a model for making inferences about the characteristics of a population--in this case, the overall MX system. By Classical analysis, only the sample results would be formally considered, although experience might still temper statistical prescriptions.

To illustrate the use of Bayesian statistics, consider an example in which a technician tests a sample of ten light bulbs from a large shipment, and finds five to be defective; this implies a 50 percent rate of defective bulbs. How would his assumption change if, in addition to the sample information, he knew that all past shipments of bulbs had been of extremely high quality, with an overall defective rate of only 5 percent? With this information, a technician using Bayesian statistics would develop a second--or "prior"--probability distribution based on the information about prior shipments, and essentially multiply it with the probability distribution from the sample data. The resulting distribution is a weighted compromise of the two and, according to a Bayesian, represents the best information at hand. The less sample information one has, the greater the relative weight or influence of prior information, and vice versa. In a case of a Bayesian analyst's not having prior information--hence, no empirical data to apply--results as would be expected are similar to the Classical method.

With constraints on test resources, the Air Force's apparent decision to incorporate prior information into a Bayesian framework, and thus reduce the sample size necessary in Phase I, seems reasonable. Prior information in the case of the MX could include performance data from R&D tests, engineering estimates, and data from component tests.

The Air Force did not, however, explicitly incorporate prior data into its Bayesian analysis. Instead, it used a mathematically-based prior distribution that heavily weights the probability of obtaining outcomes near 100 percent reliability in the sample and thus reduces the requisite sample size to meet JCS guidelines to 24 missiles. 8/ The prior distribution used by the Air Force does not appear to reflect actual, empirical information about MX or about earlier missiles.

---

8. The application of a mathematically-based prior is a subject of controversy even in the theoretical literature, and the CBO has not been able to find any practical justification that would be consistent with this case.

Although statistical justification for the Air Force's use of 24 missiles in Phase I is thus lacking, the decision reflects an operational judgment that might be reasonable. Based on its substantial experience with ICBMs, the Air Force believes that JCS guidelines for Phase II are not sufficiently demanding, and do not provide any additional flexibility for contingencies such as a longer service life. (This is discussed in more detail in the following section.) Given a limit on the total number of test missiles, the Air Force believes that added risk in Phase I is justified in order to shift additional missiles to Phase II.

Phase II

Air Force testimony indicates that for Phase II, JCS guidelines are less stringent than for Phase I and would allow a significant degradation in capability to go undetected. Moreover, the JCS guidelines do not provide sufficient detail to unambiguously interpret Phase II test requirements. Unlike Phase I, Phase II testing is concerned with detecting a change in system reliability. The Strategic Air Command has interpreted JCS guidelines for Phase II as: Test enough missiles to have a specified probability of observing at least one flight test failure caused by a specified degradation in reliability. 9/

According to Air Force testimony, six missile tests a year would be sufficient to meet minimal JCS requirements for Phase II. However, because the Air Force has never experienced as large a reliability degrade as covered by these guidelines, it has imposed a more stringent requirement for testing. The requirement is to test enough missiles to have a 90 percent probability of observing a failure associated with a force degrade of 10 percent or more over a three-year period. Based on this requirement, the Air Force plans seven missile tests a year and will pool the sample data over three years. Testing six missiles a year would provide a 90 percent chance of observing a degrade of 12 percent within three years.

But for several reasons the Air Force feels that the extra missiles are required for an acceptable Phase II program. First, as noted, the JCS guidance might not be sufficiently demanding, and the Air Force is concerned about detecting much smaller declines in reliability. Moreover, the Air Force wants some flexibility in case the MX remains in service longer than is currently planned, or in case of some necessary modifications later on. As stated earlier, the Air Force now plans to keep the MX about

---

9. The model is expressed as $P=1-(1-D)^N$ where P is the probability of detecting a failure, D is the amount of reliability degrade, and N is the number of missiles tested.

15 years, but Minuteman II has been in service about 20 years and might remain through the end of the century. If the MX should remain in service five more years than planned, then the currently planned buy of missiles for Phase II would provide about five missiles each year, or just under the minimum required to meet JCS guidelines.

As the above discussion suggests, the current MX test program is modest in scale by many standards, including JCS guidelines and in comparison with other test programs for ballistic missiles. Even in Phase II, for which the current plans do exceed JCS guidelines, there are reasonable arguments for the larger number.