United States General Accounting Office 133888

GAO

Testimony

For Release on Delivery
Expected at 10:00 a.m. Thursday, September 10, 1987

LIVE FIRE TESTING: EVALUATING DOD'S PROGRAMS

Statement of
Eleanor Chelimsky, Director
Program Evaluation and Methodology Division

Before the
Acquisition Policy Panel
Committee on Armed Services
House of Representatives

133888

GAO/T-PEMD-87-7 039928 - 133888

Mr. Chairman and Members of the Panel. It is a pleasure to be here today to present GAO's assessment of the Joint Live Fire Test program (JLF) and related DOD live fire testing programs. The results come from GAO's recently released study of live fire testing.[1]

JLF is a program in which foreign munitions are fired at combat-loaded U.S. weapon systems to determine their vulnerability, and U.S. munitions are fired at combat-loaded foreign weapon systems to determine those munitions' lethality. Testing began in FY 1985, and is scheduled to run through FY 1990. At least four groups have been identified as potential users of JLF test results: designers, tacticians, force level planners, and procurement authorities.

The program has two components: an armor/anti-armor component (JLF/Armor) and an aircraft component (JLF/Aircraft). Ships were excluded. The focus is on fielded systems. Last October, Congress mandated that developmental systems also undergo realistic live fire testing before proceeding beyond low rate initial production.

At the request of Chairman Bennett of the House Subcomittee on Seapower and Strategic and Critical Materials, we have addressed four

---

[1] GAO, Live Fire Testing: Evaluating DOD's Programs, GAO/C-PEMD-87-1 (Washington, D. C.: July, 1987) and GAO, Live Fire Testing: Evaluating DOD's Programs, GAO/PEMD-87-17 (Washington, D. C.: August, 1987)

issues: 1) the current status of each system and munition originally scheduled for live fire testing; 2) the methodological quality of the test and evaluation process; 3) the advantages and limitations of full-up live fire testing (that is, testing with combustibles on board), and the potential complementarity of other methods with full-up testing; and finally 4) improvements needed in live fire testing.

Our study covered JLF and other live fire testing not currently part of JLF, such as the Army's tests of armored vehicles removed from JLF. It did not cover testing of developmental systems under the new legislation, which was not yet implemented. Data were gathered between June and December, 1986, and consisted of observations of tests, interviews of DOD officials and outside experts in vulnerability and lethality testing and analysis, reviews of JLF and related live fire testing documents, reviews of literature on vulnerability and lethality model validation, and other literature as applicable. We reviewed all of the test plans and draft reports available during the period of our review.

ISSUE 1: What is the status of each system and munition originally scheduled for live fire testing?

There have been major slippages in the JLF/Armor test schedule, largely due to prolonged controversy between the Office of the Secretary of Defense (OSD) and the armor testers over the purposes of JLF and appropriate methods for conducting and analyzing tests. In

January, 1985 twelve different armored vehicles were scheduled to be tested within JLF. The first of these originally scheduled JLF/Armor tests began in January, 1987, almost two years behind schedule. The types of tests proposed have changed repeatedly, moving from inert tests (tests with combustibles removed) including surrogates for foreign vehicles, to exclusively full-up tests, and back again, as conflicts continued between JLF/Armor program officials and OSD over testing philosophy. This inability to agree on the basic design of the tests caused delays and uneasy compromises, as direction from OSD changed along with changes in personnel. Another major cause of delay has been the lack or shortage of targets.

In contrast to JLF/Armor, JLF/Aircraft personnel have planned and implemented their program without major conflict or interruption. The twelve tests scheduled for the first two years focused on components and subsystems. Their schedule has been delayed (principally due to lack of targets), but less severely than JLF/Armor. FY85 and FY86 testing was initially scheduled to result in nine reports by the end of FY86; as of March, 1987, however, only one was completed.

ISSUE 2: What has been the methodological quality of the test and evaluation process?

At the time of our review, there was little completed testing on which to judge the quality of the program. However, it was apparent that the technical capability to do full-up testing is not well developed. This seems to be a consequence of the historically low emphasis on full-up live fire testing in the U.S. Based on earlier live fire programs we examined, we concluded that the state-of-the-art of live fire testing has improved, but that some potentially solvable problems raised earlier have not been solved. For example, little progress has been made in the validation of vulnerability and lethality estimates derived from live fire testing.

With regard to the armor component in particular, we found that the inability of OSD and JLF/Armor officials to agree on a testing approach not only contributed to implementation delays, but also led to the waste of resources in repeated plan revisions. In effect, the testing approach taken by JLF/Armor in a 1986 draft master plan was quite similar to the first master plan, which had been rejected by OSD in 1984 because of inconsistency with the objectives of JLF.

The test planning process undertaken by JLF/Aircraft was generally well organized, thorough, and consistent with program objectives. However, individual JLF/Aircraft test plans omit key information, contain inconsistencies, and specify target requirements which exceed the availability of those targets.

Based on our review, we found that six issues have affected and will likely continue to affect the methodological quality of JLF and related live fire testing. These are: conflict over objectives, availability of targets, statistical validity, shot selection methodology, characterization of human effects, and incentive structure.

Conflict Over objectives. Guidance from OSD has been consistent in that neither improving nor validating models was ever intended to be the principal objective of JLF. The live fire testing legislation is equally clear, specifying that the primary emphasis be on testing vulnerability with respect to potential user casualties. However, we found that the JLF charter did not define live fire testing well enough to give test designers a clear direction. There have been several conflicting versions of the objectives of JLF and live fire testing in general.

The conflict over objectives reflects underlying differences between the interests of proponents of full-up testing who wanted to identify vulnerabilities of targets directly, and those who stressed the use of models to make estimates. These conflicts lead to largely incompatible approaches to testing. The failure of the two "philosophies" to co-exist is largely a function of resources. The first OSD program manager openly expressed his distrust of both models and modelers, and viewed them as impeding what to him was the primary objective--finding ways to reduce casualties. He viewed spending funds on model-oriented shots as a waste of the program's budget. The modelers, on the other hand, claimed they had waited years for an opportunity like JLF to supply their data needs. They wanted data to improve their models, which they feel are necessary for valid vulnerability and lethality assessment. They viewed spending the JLF funds on full-up shots as a waste of a unique opportunity.

Availability of targets. Both U.S. and foreign targets are in seriously short supply, and represent the principal constraint faced by all JLF test officials. For example, an early JLF test plan called for nine foreign vehicles of a type thought to be among the easiest to obtain. Only three were eventually obtained, after considerable delay. The problem applies to components as well as complete systems: one test plan for the Army's Blackhawk helicopter required four sets of flight control components, yet less than one complete set was obtained.

Such problems with target availability are in part the result of inadequate planning; there is no formal responsibility anywhere to provide targets and related support to JLF, and no funding was provided to purchase them. Consequently, test officials have had to spend a substantial portion of their time "selling" the program to skeptical service offices in order to obtain targets. It is important to remember that the bulk of this effort was directed not at obtaining new or operational hardware, but obsolete hardware that might otherwise have been discarded.

JLF has been further hindered by having to compete with others to obtain targets. For example, JLF/Aircraft was forced to vie with

NASA for F/A-18 prototypes, even though the Navy had granted JLF a formal acquisition priority and knew of its requirements for F/A-18s. Of sixteen aircraft, JLF received only two -- those in the worst condition. Negative attitudes among high DOD officials toward destructive testing were also reported to be a problem. Finally, though the systems and components that JLF does receive are frequently in poor condition, JLF provides no funds for restoration.

The Army removed several vehicles from JLF to conduct the tests themselves. In so doing, they took responsibility for obtaining targets, and supplied fully operational Bradley vehicles and M1A1 tanks off the production line (five M1A1 tanks have been made available for live fire testing). This clearly shows that while obtaining targets for live fire testing is currently a problem, it is indeed a resolvable problem (at least for U.S. ground vehicles). Given the right incentives, the services will supply targets.

Statistical validity. We found that in general, the sample sizes of JLF and related full-up live fire testing have not been large enough to produce statistically reliable results. Without statistical validity, there is no way to assess whether the test results are a fluke or would recur in repeated testing. Statistical validity would remain a major problem even if the number of targets listed in the test plans could be obtained. JLF planning presented minimal consideration of statistical issues, and the few applications of statistical analysis to live fire test data thus far are questionable. Several efforts are underway to make live fire tests more statistically interpretable.

As a substitute for statistical analysis, engineering judgment is heavily relied upon throughout the vulnerability and lethality assessment process. Unfortunately such judgment is subject to both individual and collective biases. Yet the most common form of indicator used to summarize live fire test results--probability of a kill given a hit ($P_{K/H}$)--is often based mainly on engineering judgment. Despite the fact that these judgments of $P_{K/H}$ have not been demonstrated to be reliable or valid measures, the users of output from vulnerability and lethality analysis are often unaware of their subjective nature.

Shot selection methodology. There has been controversy over how to select shots for live fire tests. We found that this essentially boils down to a conflict over two approaches: a judgmental approach which maximizes efficiency, and a random approach which minimizes bias. Each approach has its pros and cons. The test engineers tend to prefer the judgmental approach, but random selection is the only way to ensure that biases do not enter into the shot selection process, even inadvertently.

There have been several attempts to reconcile the two approaches. These attempts have sought to use technical principles of statistics and experimental design to resolve some of the methodological issues, or to provide a shot selection method that meets the concerns of both positions. However, we believe the shot selection problem will not be resolved by technical solutions alone. An interim solution might be to designate that some proportion of shots be selected judgmentally and others randomly, but ultimately, it appears impossible to agree on how to select live fire shots without first agreeing on the objectives of the tests.

Characterization of human effects. Despite claims by some officials that personnel vulnerability is well known, we found that the scientific capability to estimate human effects with confidence has not yet been achieved. This, and the fact that JLF plans have paid little attention to human effects, make it unlikely that JLF will produce precise estimates of casualties. Yet the current OSD program manager has cited the need to learn about crew survivability as his principal concern, and has asked the JLF program managers to emphasize it more in their test programs.

Incentive Structure. We believe that DOD's incentive structure is not especially conducive to realistic live fire testing. For example, there were no requirements for vulnerability testing of aircraft in the acquisition process prior to the passage of the live fire legislation, and DOD did not establish any linkage from JLF and related live fire testing to the procurement cycle. Moreover, the Bradley situation has shown that full-up live fire testing can threaten the "business as usual" of defense procurement in the U.S.

In this regard, an examination of live fire testing in other countries is informative. (Most of the section of our report dealing with live fire testing in other countries has, however, been classified.)

ISSUE 3: What are the inherent advantages and limitations of full-up live fire testing, and how do other methods complement full-up testing?

Advantages and limitations. We found that full-up live fire testing offers a unique advantage over all other methods of vulnerability and lethality assessment. It is the only method providing direct visual observation of the damage process under realistic combat conditions. The descriptions of directly observable damage that full-up testing provides are regarded as highly beneficial by users.

Full-up testing has already demonstrated some value by producing several "surprises", i.e., results that were not predicted, and might not have been detected by other means of testing or analysis. For example, the Air Force introduced a new hydraulic fluid, which laboratory tests had demonstrated to be less flammable than their standard hydraulic fluid. However, the JLF hydraulic fluid live fire tests with airflow (simulating the condition of an aircraft in flight) suggested the opposite: 30 percent of shots on the new fluid resulted in fires, compared to 15 percent of shots on the standard fluid.

The primary limitation of full-up, full scale live fire testing is cost. On a per shot basis, it is considerably more expensive than inert or subscale testing, primarily due to the high cost and limited availability of targets. Testing and restoration costs are also higher. Nonetheless, live fire testing costs are a very small percentage of total program costs. For example, even if the cost of live fire tests on the Bradley vehicle were to reach $50 million, that would still represent less than one percent of total program costs.

Another potential limitation of full-up testing is that it can yield less information about damage mechanisms per shot than inert or subscale testing, primarily because catastrophic kills may destroy the target and its components, along with much of the instrumentation used to record the damage. However, not all full-up shots result in catastrophic kills and not all catastrophic kills destroy the target. Further, such shots potentially yield more interpretable information than equivalent inert shots, which cannot provide unambiguous information about the occurrence of catastrophic events.

With regard to the expected benefits from live fire testing, JLF test officials note that such testing can have only a limited impact on developed systems because of "frozen" designs which are prohibitively expensive to change. For this reason, the main benefit of JLF and related programs may be in reducing vulnerability of future systems through lessons learned.

Other Methods. We concluded that subscale and inert testing have some distinct advantages over full-up testing, but provide only indirect evidence of effects on realistic targets, which must be inferred through an unproven analytical process (modeling). Analysis of available combat data has other advantages, but less scientific control, and is limited to systems that have been employed in combat. In sum, subscale testing, inert testing, and combat data can usefully supplement full-up, full scale testing but not substitute for it.

Models are widely used in vulnerability and lethality assessment. They are potentially useful in extrapolating beyond test results, and have a unique advantage over live fire testing in their applicability to systems still in the design phase. Live fire testing, by contrast, requires that at least a prototype has been built. However, we found that current models are inadequately validated and share numerous limitations. Prominent among these are a limited ability to model fire, explosion, multiple hits, ricochets, synergistic effects (interactions), and human effects. Some key mechanisms for producing casualties are poorly modeled, if at all, limiting the models' usefulness in predicting casualties or providing insights into casualty reduction.

We found that there are no clear mechanisms for using live fire test data to validate or revise vulnerability models. Models are frequently revised on the basis of test data, but the process is more informal and judgmental than the term validation would suggest. In addition, changes introduced into models after the Bradley Phase I tests provide an unknown level of protection from invalidation by test data, by limiting model predictions to a range of outcomes rather than a single result. It is doubtful that JLF or any future live fire testing will produce the kind or quantity of data needed to validate the sophisticated models currently in use. However, the accumulated data should enable checking whether model revisions improve prediction accuracy.

We reviewed studies that compared vulnerability model predictions with test or combat results. We found that claims that "on the average" models predict well can be misleading. On the other hand, claims that vulnerability models predict poorly have been somewhat overstated. Additionally, little attention has been paid to the different levels of accuracy required for different users' purposes. Because there are no clear criteria for success and failure in model prediction, proponents and opponents of modeling have both claimed support from the same data.

ISSUE 4: How can live fire testing be improved?

We believe that some important concerns arising from the uncertainties of JLF will be resolved for future systems by the FY87 live fire testing legislation. Specifically, the act establishes service responsibility for supplying targets, linkage to the procurement process, and a requirement for full-up and full scale testing. Future live fire tests should improve as a result, but other areas for improvement remain. With JLF still at an early stage of implementation, and the FY87 live fire legislation just taking effect, there is an opportunity now for any general lessons learned to be fed back into the live fire test design and implementation process.

We identified opportunities for technical improvements in the design, conduct, and interpretation of live fire tests. For example, DOD could test whether certain departures from combat realism that reduce the cost or difficulty of conducting full-up live fire tests do nonetheless preserve the generalizability of test results to realistic conditions. We also identified opportunities for general improvements to facilitate realistic live fire testing and the usefulness of its results. For example, rather than considering target costs in isolation, DOD could consider them in the context of total program costs, including the concept of a percentage set-aside for live fire testing. Additional suggestions for technical and general improvements are contained in Chapter 5 of our report.

RECOMMENDATIONS

In addition to the improvements I have mentioned, we believe there is a need to resolve the current conflicts about the purpose of live fire tests and to make clear that the objective of reducing the vulnerability and increasing the lethality of U.S. systems is the primary emphasis of testing. We believe that live fire testing has demonstrated its ability in this regard. Accordingly, we have recommended that the Secretary of Defense: conduct full-up tests of developing systems (in addition to existing systems), first at the subscale level as subscale systems are developed, and later at the full-scale level mandated in the legislation; establish guidelines on the role live fire testing will play in procurement; establish guidelines on the objectives and conduct of live fire testing of new systems; and ensure that the primary users' priorities drive the objectives of live fire tests. Modelers are secondary users.

The live fire legislation requires the services to provide targets for testing new systems, but there is no similar requirement for the fielded systems in JLF. Accordingly, we have recommended that the Secretary of Defense provide more support to JLF for obtaining targets.

Mr. Chairman, this concludes my prepared remarks. I would be happy to answer any questions that you or the other Members of the Panel might have.