

How Much Warhead Reliability Is Enough for a Comprehensive Nuclear Test Ban Treaty?

David Hafemeister

I. Introduction

The National Nuclear Security Administration (NNSA) selected the winning design from the two nuclear weapons laboratories for the reliable replacement warhead (RRW) on March 2, 2007. The winning design by the Lawrence–Livermore National Laboratory was the more cautious design, and had been tested previously. The Los Alamos design was more creative, but had not been nuclear tested. With the Cold War over, NNSA is planning to make warheads that are less constrained in weight and, in principle, more reliable. The Congress and the Executive Branch have agreed that RRW will not be tested before it enters the stockpile. Of course, this does not guarantee that the decision not to test could not be reversed in the future. The JASON group will comment on the RRW designs during the next year and the American Association for the Advancement of Science will release its report on the RRW in March 2007. Our discussion is intended as background material to help understand the RRW decisions and reports. On March 6, a session at the Denver APS meeting considered the RRW and nuclear missions. Talks were given by John Harvey (Director of NNSA Policy and Planning), Lt. General C. Robert Kehler (Deputy Commander of STRATCOM), Bruce Tarter (Chair of the AAAS–RRW Study), Sidney Drell (Stanford), Ivan Oelrich (Federation of American Scientists). The need for the RRW has been called into doubt by the 2006 JASON report that concluded the following: [\[1\]](#)

“Most primary types have credible minimum lifetimes in excess of 100 years as regards aging of plutonium; those with assessed minimum lifetime of 100 years or less have clear mitigation paths that are proposed and/or being implemented...There is no evidence for void swelling in naturally –Pu samples over the actual and acceleratedδaged or artificially aged times scales examined to date, and good reason to believe it will not occur on times scales of interest, if at all. Systems with large margins will remain so far greater than 100 years with respect to Pu aging. Thus, the issue of Pu aging is secondary to the issue of managing margins.”

There is a strong consensus in the US that the *primary* mission of nuclear weapons is to deter nuclear attacks by other nations. However, there is also a strong consensus that nuclear weapons do not deter terrorism by non-state actors. These views were summarized by former Secretaries of State George Shultz and Henry Kissinger, former Secretary of Defense William Perry, former Chair of the Senate Armed Services Committee Sam Nunn and others, who commented in the Wall Street Journal of 4 January 2007 that “reliance on nuclear weapons for this purpose [deterrence] is becoming increasing hazardous and decreasingly effective.”[\[2\]](#) They also recommended ratification of the Comprehensive Nuclear Test Ban Treaty (CTBT) by “Initiating a bipartisan process with the Senate, including understandings to increase confidence and provide periodic review, to achieve ratification of the Comprehensive Test Ban Treaty, taking advantage of recent technical advances, and working to secure ratification by other key states.”

The main technical issue that blocked CTBT ratification in 1999 was the following: “Will nuclear weapons be sufficiently reliable if they are not tested for centuries?” This question is somewhat misleading since a nation can always withdraw from the CTBT under Article IX when its “supreme interests” are jeopardized. The other main CTBT issues have been or are being solved sufficiently for ratification by the Senate:

(1) CTBT Effective Verification. The CTBT will be “effectively verifiable” when the International Monitoring System is complete and because regional seismic monitoring has greatly improved, along with improvements with seismic arrays and analysis, interferometric synthetic aperture radar and cooperative monitoring. The level at which cheating could take place would not significantly threaten US national security, according to the Nitze–Baker criteria used for the INF and START I-II treaties.^[3] The National Academy of Sciences (NAS) 2002 study on the CTBT concluded the following on monitoring with a fully deployed primary seismic network:^[4]

Underground explosions can be detected and can be identified as explosions, using IMS data, down to a yield of 0.1 kt [tamped] in hard rock if conducted anywhere in Europe Asia, North America and North Africa. In some locations of interest, such as Novaya Zemlya, this capability extends down to 0.01 kt or less.

(2) CTBT (with–compliance) vs. no–CTBT vs. CTBT (with–evasion). The NAS panel examined these three situations for seven nations (Russia, China, India, Pakistan, North Korea, Iraq and Iran), concluding the following:^[5]

States with extensive prior test experience [Russia and China] are the ones most likely to be able to get away with any substantial degree of clandestine testing, and they are also the ones most able to benefit technically from clandestine testing under the severe constraints that the monitoring system will impose....Countries with lesser prior test experience and/or design sophistication would also lack the sophisticated test–related expertise to extract much value from such very–low–yield tests as they might be able to conceal....The worse–case scenario under a no–CTBT regime poses far bigger threats to US security interests—sophisticated nuclear weapons in the hands of many more adversaries—than the worst–case scenario of clandestine testing in a CTBT regime, within the constraints posed by the monitoring system.

(3) Nuclear Safety. Only one US nuclear weapon accident has taken place since 1968, which was the 1980 accident of a liquid–fueled missile. This accident did not spread radioactivity and is now irrelevant since all liquid–fueled nuclear missiles have been decommissioned. Only two of the 32 accidents spread considerable radioactivity, which were both aircraft accidents. Practically all (29 of 32) nuclear weapon accidents were with aircraft, which no longer carry nuclear weapons unless placed on alert. The least safe nuclear weapons (nuclear artillery and SRAMs) have been decommissioned and safety procedures have been modified for submarine weapons. A 1992 law required that the Defense Department to do a cost–benefit analysis on safety issues to determine whether new warheads that needed nuclear testing were cost effective. Both Republican and Democratic administrations have since testified that new weapons are not needed to enhance safety. There are no significant safety problems that require nuclear testing to resolve them.

II. NAS Panel Conclusions on Reliability.

- The NAS panel determined that, under these conditions, US warheads could remain safe and reliable without testing:
- Maintain a high-quality workforce.
- Stockpile stewardship and enhanced surveillance must examine components of weapons. Based on past experience, the majority of aging problems will be found in the non-nuclear components, which can be fully tested under a CTBT.
- The most likely potential source of nuclear-related degradation is the possibility that the primary yield falls below a minimum level needed to drive a secondary. NNSA has

concluded that plutonium pits have a minimum lifetime of 45-60 years [now 100 years] with "no life-limiting factors as yet recognized."

- In the past there were few underground nuclear explosions that explicitly served to check the reliability of weapons in the stockpile. Most nuclear tests were used to study and certify new designs and to examine weapons effects.
- Remanufacture to original specifications is the preferred approach for age-related defects, with a highly disciplined process to install few changes without changing the basic nuclear design.

The NAS panel continually asked weapon designers during classified briefings on the enduring stockpile whether testing was needed to resolve the issue under discussion. NNSA weapon scientists always responded that testing was not needed to solve the issue under discussion. The NAS panel concluded the following, based on their experience and the briefings:

Although a properly focused stockpile stewardship program is capable, in our judgment, of maintaining the required confidence in the enduring stockpile under a CTBT, we do not believe that it will lead to a capability to certify new nuclear subsystem design for entry in the stockpile without nuclear testing -- unless by accepting a substantial reduction in the confidence in weapon performance associated with the certification up until now, or a return to earlier, simpler, single stage design concepts such as gun-type weapons.

It seems to us that the argument to the contrary – that is, the argument that improvements in the capabilities that underpin confidence in the absence of nuclear testing will inevitably lose the race with the growing needs from an aging stockpile – underestimates the current capability for stockpile stewardship, underestimates the effects of current and likely future rates of progress in improving these capabilities, and overestimates the role that nuclear testing ever played (or would be ever likely to play) in ensuring stockpile reliability.

These conclusions are consistent with the fact that the United States has not needed to test in the 15 years since the testing moratorium began in 1992. Each year the US government has stated that it is “confident that the stockpile is safe and reliable, and there is no requirement at this time for nuclear tests.”^[6] The annual certification on stockpile readiness requires the Secretary of Defense (after advice from Strategic Command and the military services) and the Secretary of Energy (after advice from the three weapon laboratory directors and the NNSA Administrator) to determine whether all safety and reliability requirements are being met without the need for nuclear testing. These reports have always certified that the stockpile does not need testing for reasons of safety or reliability. The NAS panel concluded, with these caveats, that testing is not needed in future years: (1) A robust stockpile stewardship program, (2) no new weapon designs, and (3) the right of the United States to withdraw from the CTBT if the United States decides it must test to defend its national security.

About \$7 billion is spent annually to maintain the enduring stockpile (Table 1) and infrastructure under the Stockpile Stewardship Program (SSP) and the Lifetime Extension Program (LEP). Sidney Drell and Robert Peurifoy discussed the technical issues involved with a nuclear test ban.^[7] The main threat to warhead reliability is caused by non-nuclear components, which is usually observable without testing on these issues: insufficient tritium, faulty tritium bottles, corrosion of fissile material, degradation of high explosive, low-temperature performance, vulnerability to fratricide neutrons, radar, batteries, fuse switch, neutron generator, faulty cables, trajectory sensors, control systems, rocket motor, gas transfer valve, firing set, and pilot parachute. The warheads in the enduring stockpile have been tested 150–200 times.

Table 1. US Nuclear Warheads in the Enduring Stockpile (2006). Warhead types that are to be partially dismantled are marked with an *. This table does not include the B62 (580 warheads) and W84 (383 warheads), which are scheduled for full dismantlement. [R.S. Norris and H.M. Kristensen^[8]]

Type	Yield	Platform	Active	Inactive	Total
B61/3*	10-350 kt	airplane	200	186	386
B61/4*	10-350 kt	airplane	200	204	404
B61/7	10-350 kt	airplane	215	224	439
B61/11	10-350 kt	airplane	20	21	41
B83	1.2 Mt.	airplane	320	306	626
W76*	100 kt	SLBM	1712	1318	3030
W78*	335 kt	ICBM	785	20	805
W80/1*	150 kt	ALCM	1450	361	1811
W87	300 kt	ICBM	0	553	553
W88	475 kt	SLBM	404	0	404
TOTAL			5306	3193	8499

Eleven warheads of each type are annually taken to the Pantex facility, disassembled and examined for deterioration. The JASON group recommended a variety of measures to increase performance margins of warheads, beyond increasing tritium content in the warhead.^[9] Warheads will have to be rebuilt; the question is how often with 100-year pit lifetimes. The basic science of warheads and their viability are examined with the technologies listed below:

- visual observation for corrosion, deterioration, cracks and other issues
- chemical, electrical, ultrasonic, diamond-anvil, and other tests
- functional testing of components
- X-ray scattering to search for changes
- deep penetration digital radiography to detect flaws and cracks (core punch)
- laser scattering to study surface imperfections
- synchrotron-based spectroscopy and diffraction
- reassembled device without SNM tested to destruction (Joint Test Assembly)
- subcritical and hydrodynamic tests (Rebound, Holog, Joint Actinide Shock Physics Experimental Research, Atlas pulse power machine, critical assemblies at Device Assembly Facility).
- Dual Axis Radiographic Hydro Test (DAHRT)
- Advanced Simulation and Computing (ASC) program
- accelerated aging of pits with shorter-lived plutonium-238
- National Ignition Facility (not yet functioning)

III. NNSA Definition of Reliability

The United States has not tested each warhead type enough times to determine reliability with high confidence statistics, and certainly not for the effects of aging. Assume ten reliability tests were performed and all were successful. The reliability is not 100 percent with 100 percent confidence, but rather there is a 30 percent chance that reliability is less than 90 percent and a 10 percent chance that reliability is less than 80 percent.^[10] Thus, the United States has never known warhead reliability with precision when the warhead entered the stockpile, nor has the United States searched sufficiently for aging effects with confidence tests.

NNSA Definition of Reliability. “The reliability of obtaining the predicted yield of a nuclear weapon has never been assessed because there have never been enough performance [nuclear] tests to establish a statistical reliability. Thus, when a defect type impacting the nuclear explosive package is discovered, the yield performance is evaluated, but no reliability degradation estimate can be made. Therefore, no data is available regarding analysis relating to reliability degradation to predicted yields....In general terms, reliability is defined as the ability of an item to perform a required function. Implicit in the above definition of ‘required function’ for one-shot devices, such as nuclear weapons, are the required conditions and duration of storage, transportation, and function.”^[11] In other words, when a few successful tests give the design yield, the reliability of a warhead type is defined as 1.0, but *without* a confidence level. When actionable defects are detected, NNSA analysis reduces reliability of 1.0 by an amount R to give a reduced-reliability for each warhead type. NNSA set numerical bounds on reliability reductions $R\Delta$ for 164 actionable defects in 46 warhead types, mostly in the 39 retired warhead types:^[12] $\Delta R = 0-1\%$ (112 defect types), $R\Delta = 1-5\%$ (37), $R\Delta = 5-10\%$ (6), $R\Delta > 10\%$ (9).

The effect on secondary yield of radiant energy transfer from the primary stage is very nonlinear. A drop in primary yield by a factor of two, for example, could greatly reduce the secondary yield because critical pressures and temperatures may not be obtained. However weapon yield is not a “step function” that varies between two values, zero and certified yield. NNSA is concerned about catastrophic failure of an entire type. This is partially driven by the fact that yield on target is usually much larger than what is needed for particular missions, so the only issue is “does it work.” NNSA does not consider the criteria for nuclear missions in any depth since targeting is left to the Strategic Command. Since there are 7 warhead types in the enduring stockpile, a catastrophic failure of one type would shift responsibility to the other six types, with time to repair the catastrophic failure.

IV. Requirements for Reliability and Yield.

NNSA does not consider nuclear targeting for its annual certification report. Since the accuracy of missiles is a statistical phenomenon, statistical analysis is necessary to quantify destruction of targets to determine if warhead degradation is relevant or not. The ability to destroy a target depends on (1) the hardness H of the target (minimum destruction pressure), (2) the yield Y of the weapon, (3) the accuracy of the weapon (CEP , circular error probable), (4) the reliability R of the weapon system (0 to 1), and (5) the number n of warheads attacking a target (taking into account fratricide).^[13] The single-shot-kill-probability $SSKP$ is the kill probability of a single warhead on a known target with perfect reliability of $R = 1$. We initially assume lethal warheads with $SSKP = 1$, giving a kill probability for one warhead of $P_1 = R$. If n independent warheads from n missiles are used on a target without fratricide, the kill probability is $P_n = 1 - (1 - R)^n$. Reliability of $R = 0.5$ gives $P_2 = 0.75$ and $P_3 = 0.88$, and $R = 0.25$ gives $P_2 = 0.44$ and $P_3 = 0.58$. Except for the case of a pre-emptive attack against a large force, additional warheads on a target can be used for case of reduced reliability.

The kill capability of one W88 warhead of 475 kilotons with 100-meter CEP accuracy attacking a 2000-psi hard target silo with 0.9 reliability is $P_1 = 0.898$. If the W88 yield is reduced by 50%, $P_2 = 0.99$ and $P_3 = 0.998$, and if yield is reduced by 90%, $P_2 = 0.88$ and $P_3 = 0.96$. These results show that large yield reductions do not significantly change P_2 and P_3 .

Testing data obtained from DOE with a Freedom of Information Act (FOIA) request is discussed below.^[14] Some warhead types had problems during the early transition to miniaturized warheads with reduced mass and volume.^[15] The 1958–61 testing moratorium prevented tests at the time these new warheads entered the force.

Actionable defects, identified by stockpile stewardship and not by nuclear tests, are listed in Table 2; those marked with an * needed a retrofit or a major redesign. The last column gives the year of discovery of the defect after the first production unit (FPU). Three warheads were retrofitted: (1) B61, 3 years after FPU, (2) W80, 1 year after FPU, and W88 at 1, 1, 2, and 3 years after FPU. This and other data suggest that primary/secondary stages do not show significant aging problems once they have been in the field for a few years. The average age of discovery for the 6 retrofits in Table 2 was 1.8 years after FPU. Five retrofit types were for primaries and one was for a secondary. All six retrofits were from design flaws, causing yield reduction (in 4 cases), reduced safety (1 case) and non-applicable (1 case). The average discovery time for the retired warhead types was 1.9 years after FPU for the 33 primary and 1 secondary stages. Retrofits were caused by design flaws (33), aging (5) and production (1), which effected safety (19), reliability (6), yield reduction (5) and not applicable (4).

Table 2. Actionable defects for warheads in the enduring stockpile. Those that required a retrofit or major design change are marked with an *. This table does not include the 39 retired warhead types. Nuclear components are primary (p) and secondary (s) with number of generic events in parentheses. Causes are aging (A), design (D), and production (P), without needing the causes of field induced (F), unknown (U), and combination of design/production (C). Effects from the causes are safety (S) (nuclear detonation safety, nuclear material scatter, or personnel safety), operational yield reduction (O), and not applicable (na), and reliability reduction (R, which was not applicable here). [Table 6, FOIA–NNSA]

	primary(p) secondary(s) (number)	cause A,C,D,P,U	effect R,S,O,na	FPU: yr after
B61	p(2) s(2)	D*, P P, P	O*, O S, S	1980-86: 3*, 3 6,7
B83	p(0) s(0)			1983:
W76	p(2) s(1)	P, P P	na,O O	1979: 1,4 6
W78	p(3) s(0)	P, A, D	O, na, S	1980: 3,6,11
W80	p(2) s(2)	D*, P	O*,O na, na	1981-4: 1*, 1 1,5
W87	p(0) s(0)			1986:
W88	p(3) s(2)	D*, D*, D* D*, D	na*, S*, O* O*, O	1988: 1*, 2*, 3* 1*, 3

Table 2 suggests that primaries are much more vulnerable than secondaries. The two sets of data (retired and enduring warheads) show that the average age of discovery is less than two years after the first production unit. The full data set gives the main cause of diminished reliability, which results from failures of non-nuclear components, not failures of nuclear stages. Drell and Peurifoy quantified warhead reliability as follows: “Since the start of the current stockpile evaluation and reliability assessment program in 1958, about 13,000 weapon evaluations have been conducted. During this period, the failure rate of the nondevice hardware suggests an expected weapon failure rate of 1–2% for the stockpile.”[\[16\]](#)

Missile failure rates are larger, as pointed out by Richard Feynman, whose estimates were 2% for mature solid-fueled missiles and 4% for all solid-fueled missiles.[\[17\]](#)

These actionable defects for the enduring stockpile were all discovered by stockpile stewardship, except for the W80 cruise missile warhead, which revealed a cold temperature detonation problem. DOE was asked about the “four Product Change Proposals that required underground tests since 1970.” The FOIA response below stated that only 4 tests were used; 2 for enduring stockpile weapons and 2 for now retired warheads:[\[18\]](#)

1. B61/Mod-1 conversion to B61/Mod-7 (Underground testing was used to compare nuclear performance of the insensitive, IHE–primary relative to the former HE–primary being replaced) – 13 years post B61/Mod-1 FPU.
2. W68 (Underground testing verified a corrective change replacing the primary HE) – 7 years post FPU;
3. W79 (Underground testing confirmed a safety problem) – 7 years post FPU; and
4. W80 (Underground testing revealed a cold temperature detonation problem) – within 1 year FPU.

The FOIA response described Major Product Change Proposals for warheads in the enduring stockpile. Six of the 36 proposals affected the primary or secondary and 30 were for non-nuclear components. A new pit was incorporated for the B61 in the first year after FPU and high explosive specifications were changed 3 years after FPU. Thirteen years after FPU the B61/Mod-1 pit was modified for insensitive high explosive for Mod-7, the earth penetrator, which was nuclear tested. The W88 primary and secondary was modified during 1–3 years after FPU. The W80 primary was modified one year after FPU for cold-temperature performance.

Conclusion.

The data presented in this paper suggest that US nuclear warheads continue to be reliable, consistent with the annual certification by the Secretaries of Defense and Energy. Plutonium aging is no longer a significant issue, as shown by natural- and accelerated-aged plutonium samples. It is imperative that the missions for nuclear weapons be considered when modernizing and sizing the US nuclear weapons stockpile. NNSA is given very high requirements for yield and reliability by the Department of Defense. But these very high requirements are only relevant for a pre-emptive attack on Russia (perhaps China in the future). The Defense Department maintains these extremely high standards for this type of attack, but this policy leads the United States to reject the Comprehensive Nuclear Test Ban Treaty, an act which is counter-productive to the US goal of reducing the threat of nuclear proliferation. Secrecy and vagueness have prevented a relevant discussion on weapon reliability and its impact on the CTBT. The 6 December 2006 vote on the resolution favoring the CTBT in the UN General Assembly shows that practically all nations strongly prefer a completed CTBT, as they fail to understand US views on the reliability of nuclear weapons. The vote was 172 in favor to 2 against (Democratic People’s Republic of Korea and the United States), with 4 abstentions (Colombia, India, Mauritius, Syria). The data presented in this paper suggest that the 172 votes in favor of the CTBT are well justified by the facts.

David Hafemeister
Center for International Security and Cooperation
Stanford University
dhafemei@calpoly.edu

- [1] JASON, *Pit Lifetime* (McLean, VA: MITRE Corporation, November 20, 2006): 1, 16.
- [2] G. Shultz, W. Perry, H. Kissinger, and S. Nunn, “A World Free of Nuclear Weapons,” *Wall Street Journal* (Jan. 4, 2007): A15
- [3] D. Hafemeister, “Progress in CTBT Monitoring since its 1999 Senate Defeat,” and “Effective Verifiability of the CTBT,” to be published.
- [4] National Academy of Sciences, *Technical Issues Related to the Comprehensive Nuclear Test Ban Treaty* (Washington: National Academy Press, 2002).
- [5] NAS–CTBT study: 70–78.
- [6] L.F. Brooks, NNSA Administrator (Washington: Arms Control Association, January 25, 2006).
- [7] S.D. Drell and R. Peurifoy, “Technical Issues of a Nuclear Test Ban,” *Annual Review of Nuclear Particle Science* 44 (1994): 285–327.
- [8] R.S. Norris and H.M. Kristensen, “US Nuclear Forces 2007,” *Bulletin of Atomic Scientists* 63 (January/February 2007): 79–82.
- [9] JASON, *Primary Performance Margins* (McLean, VA: Mitre Corporation, 1999).
- [10] S. Fetter, *Toward a Comprehensive Test Ban* (Cambridge, MA: Ballinger, 1988): 89–105.
- [11] FOIA–DOE, E.A. Barfield, DOE Freedom of Information Act Office of Public Affairs, Number 95–207–C, to Hisham Zeriffi (Takoma Park, MD: Institute for Energy and Environmental Research, January 5, 1996). H. Zeriffi and A. Makhijani, “The Nuclear Safety Smokescreen,” *Physics and Society* 29 (April 2000): 3–6.
- [12] FOIA–DOE: 4–5.
- [13] D.W. Hafemeister, *Physics of Societal Issues* (New York: Springer, 2007): chapter 2.
- [14] FOIA–DOE: 3–6.
- [15] R.E. Kidder, *Assessment of the Safety of US Nuclear Weapons and Related Nuclear Test Requirements: A Post–Bush Initiative Update* (LLNL, December 1991). Also see, S. Fetter: pp. 34–42.
- [16] Drell and Peurifoy: 320.
- [17] R. Feynman and R. Leighton, *Classic Feynman* (New York: W.W. Norton, 2005): p. 466.
- [18] FOIA–DOE: 4.