Considering the Scenario of Incorrect Data Supplied During an Inspection Under the New START Treaty

Peter Sprunger, Ph.D.

Executive Summary

During the Senate hearings on the New START treaty, a question submitted for the record by Senator Lugar to Secretary Gates, Secretary Clinton, and Admiral Mullen requested the Administration to "Please provide for the record the analysis, including any statistical examination, done regarding the number of inspections required to have high, medium and low confidence of monitoring limits under New START. This material may be submitted in classified form if necessary." (1) The material was submitted in a classified form leaving those without appropriate clearances to guess at the answer. This paper provides an unclassified analysis to this question.

It is assumed that, for whatever reason, the data supplied to us by Russia regarding the number of deployed warheads under the New START Treaty is not 100% correct. What is the probability that the United States will catch this incorrect data and what are the implications? Over one hundred thousand Type-One inspection scenarios were simulated using open source estimates of Russia's nuclear force structure and by making assumptions regarding the capabilities of the United States' National Technical Means (NTM). Three possible scenarios of the future structure of Russia's nuclear force resulting from a decrease in deployed warheads are presented in detail here. They are 1) all warhead reductions come solely from ICBMs, 2) all warhead reductions come solely from SLBMs, and 3) reductions come from an equal reduction in ICBMs and SLBMs.

There are three conclusions that can be drawn from this work.

1) Because the probability of catching incorrect data increases, and therefore the resulting number of excess warheads decreases, as the information gathered by NTM increases, the first 70 Type-One inspections are of upmost importance to maximize our confidence in our NTM. Therefore, the primary focus of the first 70 Type-One inspections should be to maximize the capabilities and confidence levels of our NTM rather than to gain specific knowledge of Russia's nuclear force.
2) In all scenarios investigated, when the United States' NTM is such that greater than 75% of the supplied data is known prior to inspection, there is approximately less than a 0.2% probability of passing all inspections if the number of undeclared warhead exceeds 40.
3) The statement that Russia has a low probability of any military significant cheating contains two unknowns. What is considered a low probability, and what is militarily significant? Using the data presented in this paper, if one of those unknowns is determined, the other can be estimated. It is the author's estimation that either the United States' National Technical Means allows us information of over 75% of the current state of Russia's arsenal, or that 40 undeclared warheads is not considered militarily significant.

Introduction

During the Senate hearings on the New START treaty, the issue was raised about the United States' ability to adequately monitor the treaty to ensure that no significant military cheating occurs. The Administration stated for the record that any amount of cheating would be considered serious, but the military significance of said cheating would have to be put in context of the current situation. A question submitted for the record by Senator Lugar to Secretary Gates, Secretary Clinton, and Admiral Mullen requested the Administration to "Please provide for the record the analysis, including any statistical examination, done regarding the number of inspections required to have high, medium and low confidence of monitoring limits under New START. This material may be submitted in classified form if necessary." (1) The material was submitted in a classified form leaving those without appropriate clearances to guess at the answer. This paper attempts to do just that by providing an unclassified analysis to this question.

This paper supposes a scenario in which the information provided to the United States by Russia regarding its nuclear force structure is not 100% correct. More specifically, the scenario where the number of warheads on a given delivery vehicle is stated incorrectly in the supplied data. This could either be due to deliberate cheating, an unintended clerical error, or any other number of possibilities. The treaty allows for ten Type One (Type I) inspections per year "to confirm the accuracy of declared data on the numbers and types of deployed and non-deployed strategic offensive arms subject to this Treaty; the number of warheads located on deployed ICBMs and deployed SLBMs; and the number of nuclear armaments located on deployed heavy bombers." (2) This paper will only focus on the capabilities of the Type I inspection.

Furthermore, this paper will only focus on the Type I inspections carried out after the limits of the treaty are to have been met. The treaty requires compliance within seven years after entry into force. Therefore there is a maximum of 30 Type I inspections over the next three years to verify compliance with the treaty. While a discrepancy in the supplied data is a serious issue at any point in the treaty, if the discrepancy occurs within the first seven years it will not in any way imply non-compliance. Therefore, if non-compliance is the goal, there is no incentive to intentionally supply incorrect data within the first seven years. The first 70 Type I inspections not only help to maintain trust and transparency, but can also be used to increase the confidence in our national technical means (NTM).

The most difficult aspect of this calculation is determining the state of Russia's nuclear force structure over the span of the treaty. Considering that the Bulava is still being tested, it is probably safe to say that even Russia is fairly uncertain about what their force structure will look like in ten years. If the missile is a success, Russia may move towards a more marine-based nuclear deterrent and make the warhead reductions in their ICBM fleet (both silo-based and mobile). If the Bulava ends up being a complete failure, they may move towards ICBMs and make all the warhead reductions with the SLBMs. These are two extreme scenarios, neither of which are likely, but serve to represent the plethora of possibilities that could be considered.

The first assumption that is made to make this calculation manageable is that since Russia is already below the number of allowed delivery vehicles, there will not be any significant reductions in the number of delivery vehicles. Only a reduction in the number of warheads was considered. Relating to the assumption that there will be no reductions in the number of delivery vehicles, it was also assumed that each delivery vehicle would have at least one warhead on it.

Second, the bomber fleet was left out of this calculation. The counting rules of the New START treaty state that each bomber counts as one delivery vehicle and one warhead. These counting rules do not reflect reality as it is currently estimated that the bomber fleet consists of 76 airplanes that can deliver over 800 warheads. Even if the entire bomber fleet was eliminated, reductions in the number of warheads deployed on ICBMs and/or SLBMs would still need to occur to bring Russia within compliance.

From these two assumptions, it was assumed that the nuclear force structure, with the exception of the number of warheads on each delivery vehicle, will remain constant throughout the treaty. This is an assumption which will undoubtedly be proven incorrect but in light of all the uncertainties is probably the best guess of all possibilities. The estimates stated by Pavil Podvig on the Russian Strategic Nuclear Forces website were used as input to the calculation (3). This data is summarized in Table 1.

A third assumption that was made to simplify the calculation is that if the number of warheads on a delivery vehicle is stated incorrectly that the delivery vehicle was maximally MIRVed. This is a classic assumption that results in the data presented representing an upper limit on the number of excess warheads.

The final aspect that needs to be taken into account is our National Technical Means (NTM). NTM was taken into account by assuming that a certain percentage of the data was known prior to the inspection. Since this paper assumes no knowledge of the specific capabilities of our NTM, how this data is known is beyond the scope of this paper. Only the fact that it is known is important. The procedure to calculate the amount of data known via NTM was to first assume an average percentage that is known. For each base the number of delivery vehicles for which we already have complete information is the assumed percentage of known data times the number of delivery vehicles at the base. Fractions were rounded to the nearest integer.

Russia’s Current Estimated Nuclear Force Structure				
Base	Total delivery vehicles	MIRV	Max Warheads	System
Tatishchevo	41	6	246	SS-19
Tatishchevo	49	1	49	SS-27
Kozelsk	29	6	174	SS-19
Vypolzovo	18	1	18	SS-25
Teykovo	3	3	9	RS-24
Teykovo	18	1	18	SS-27
Yoshkar-Ola	27	1	27	SS-25
Dombarovsky	30	10	300	SS-18
Nizhniy Tagil	27	1	27	SS-25
Uzhur	28	10	280	SS-18
Novosibirsk	26	1	36	SS-25
Irkutsk	27	1	27	SS-25
Barnaul	36	1	36	SS-25
Gadzhiyevo	96	4	384	96 Delta IV, 6 subs
Vilyuchinsk	64	3	192	64 Delta III, 4 subs
Total ICBM	369		1247	
Total SLBMs	160		576	
Total Both	529		1823	
Total Bombers	76		76	

Total ICBM that can sustain a cut	131		1009	
Total Reducable Forces	291		1585	

Table I: Current estimates of Russia's Submarine and Rocket forces used in this paper. The data is taken from Russian Strategic Forces (3).

Calculation

There are three scenarios considered in this paper, two extreme cases and a middle ground. The two extreme cases are that all warhead reductions will come from ICBMs with no change to the SLBMs, and the second is that all warhead reductions will come from SLBMs with no change to the ICBMs. The middle ground scenario is that cuts will come evenly from both the SLBMs and ICBMs. Assuming no tubes are dismantled the cuts will come solely from a reduction of the MIRVing. As seen in Table 1, the total number of warheads on SLBMs, ICBMs, and the number of bombers (with a bomber only counting as one) is 1899. Therefore at least 349 warheads will need to be removed to bring Russia within compliance of the treaty. This represents a reduction of 18.4%, or retention of 81.6%.

It is not that simple however, because not all delivery vehicles have more than one warhead on them. There are currently 369 rockets and 1247 warheads on Russian ICBMs. However, 238 of these rockets are loaded with only one warhead. Therefore, the cuts to the ICBMs must come from the 131 MIRVed rockets with their 1009 warheads. If the cut of 349 warheads comes solely from ICBMs this represents a 34.6% reduction in the number of warheads on the MIRVed rockets. Similarly, if the warheads are removed from only SLBMS, that would leave 227 warheads out of the current total of 576 on the 160 MIRVed SLBMs. This represents a 60.6% reduction; or a 39.4% retention. Finally, if the cuts come from both SLBMs and ICBMs, the 349 warheads will come from the combination of MIRVed ICBMs and SLBMs. The two have a total of 1585 warheads and the 349 cut represents a reduction of 22%.

Numerically, the probability of catching incorrectly labeled delivery vehicles in any one given inspection is the number of incorrectly labeled delivery vehicles divided by the total number of delivery vehicles on the base. For example, consider the inspection of one submarine with 16 tubes, and there are two ballistic missiles that do not have the stated number of warheads on them, the probability of catching the error is 2/16 or 12.5% and the probability of the Russians passing the inspection is 87.5%. However, if the same incorrect data is supplied for a second inspection, the overall probability of passing both inspections drops to (2/16)*(2/16) or about 76.6%. If this is continued over a span of 30 inspections, the probability for passing all inspections is 1.8%. So while the probability of passing any one given inspection is rather high, the probability of passing multiple inspections is rather low.

However, we are not only interested in whether cheating is occurring, but to what extent. In the example above, there were four missiles that were incorrectly labeled in some way over two inspections. Let us suppose that they are also Delta IV's and the supplied data claimed that there were two warheads on each. These four missiles could have a maximum of 16 warheads on them collectively resulting in eight undeclared warheads. So, there is a 76.6% chance of hiding these eight warheads in this scenario. However, say the four incorrectly labeled SLBM's were not split up two and two, but rather all four were on one submarine. Then the probability of passing both inspections is (16/16)*(12/16) or 75%. It is important to notice that different configurations of the same numerical amount of incorrectly labeled data can result in different probabilities of passing all inspections.

Location of the incorrect data also changes the probability of detection. Consider two hypothetical bases, one with 20 delivery vehicles and the other with 30 delivery vehicles and between the two there is 1 incorrectly labeled launcher. If the incorrect launcher is on the base with 20 delivery vehicles there is a 95% chance of passing the inspection, if it is on the base with 30 delivery vehicles there is a 96.7% chance of passing. Therefore if a small number of delivery vehicles are mislabeled, it will be harder to find them on the bases with more delivery vehicles.

The calculation is easy to do post-mortem. It will be easy in 10 years to look back over the data collected and knowing which bases were inspected, calculate the probability that the United States missed detecting any incorrect supplied data. Furthermore, it will be easy to suggest improvements for the next treaty involving inspections. This paper takes the forward approach and attempts to suggest ways to strengthen the inspections in order to maximize the probability of catching any incorrect labeling and minimizing the number of undeclared excess warheads. This is done by simulating millions of different inspection combinations using Monte Carlo methods.

The procedure of simulating inspection scenarios went as follows:

1) Randomly generate a force structure consisting of the 13 ICBM and SLBM bases shown in Table I. The difficulty of this step was to simulate the number of warheads on each individual launcher assuming it was correctly labeled. In reality, it is very likely that any inconsistencies in the Russian force structure will be decoupled from the most likely way to pass an inspection. For this reason the number of warheads on each launcher was generated using a Poisson distribution. The use of the Poisson distribution injects randomness into the system caused by military realities not related to trying to pass an inspection and at the same time keeps the overall force structure rather uniform. Because it is assumed that each launcher has at least one warhead, the zero-point of the Poisson was set at one warhead and the average calculated using (MIRV capability) x (Retention Rate) - 1. In otherwords the Poisson distribution concerns only the amount of warheads on the launcher above one. If the force structure contained more than 1550 declared warheads, it was thrown out and a new one was generated.

2) Once a satisfactory force structure was determined, the number of incorrect delivery vehicles on each base was determined. This was also generated using a Poisson distribution with the average set using (Percentage of Incorrect Data) x (Delivery vehicles per base).

3) It was assumed that if the supplied data was incorrect for a given launcher, then the launcher would be maximally MIRVed. The number of excess undeclared warheads for each incorrectly label launcher was then the maximum number of warheads possible for that particular missile minus the number of warheads assumed on that missile as generated in step 1.

4) Then the process of inspections occurs. The probability of passing an inspection is calculated using the formula

$$P = \frac{M - I - K}{M - K}$$

where P is the probability of passing the inspection, M is the number of missiles at the given base, I is the number of incorrectly labeled missiles, and K is the number of missiles that we know all about prior to inspection with our NTM. This process was repeated 30 times for the 30 inspections. It was assumed that the actual process of choosing which base to inspect was decoupled from the process of maximizing our ability to catch incorrect data. For this reason, it was assumed that each rocket base and submarine was inspected one time and the remaining inspections were then randomly chosen.

5) The excess number of warheads for the given force structure and the probability of passing all inspections was recorded.

6) The five steps above were then repeated 10,000+ times in order to get a wide distribution of possible force structures and inspection scenarios.

Analysis

Figure 1 shows the probability for passing all inspections (vertical axis) as a function of the percentage of correct data supplied during the inspection (horizontal axis). The solid lines represent the median of the simulated data and the shaded areas represent the region containing 95% of the data. Three different sets of data are displayed representing three scenarios where different amounts of data are known via NTM. In red is the scenario with no NTM information, blue assumes 50% of the supplied data is known prior to inspection, and the green 75%. The data shown in Figure 1 assumes that the reduction in the number of warheads comes from both a reduction in ICBMs and SLBMs and that the bases to be inspected are chosen completely at random. It was also assumed that only one submarine was in dock at any given time, more will be said on this assumption later.

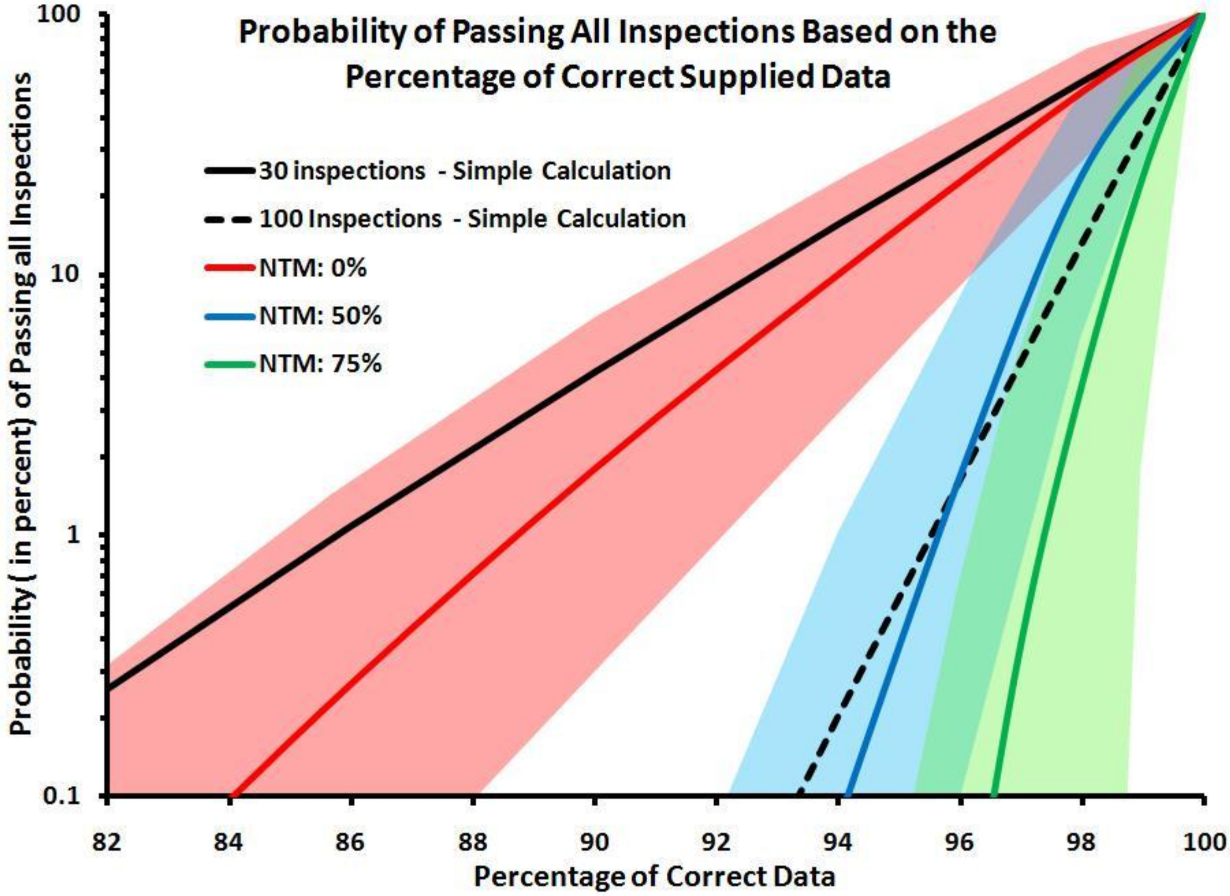

Figure 1: The probability for passing all inspections (vertical axis) as a function of the percentage of correct data (horizontal axis). The solid lines represent the median of the simulated data and the shaded areas represent the region containing 95% of the data. Three different sets of data are displayed representing three scenarios where different amounts of data are known via NTM. In red is the scenario with no NTM information, blue assumes 50%

of the supplied data is known prior to inspection, and the green 75%. The solid black line is the percentage of correct data to the 30th power, i.e. 30 inspections, and the black dashed line is the percentage of correct data to the 100th power. The data is taken assuming the warhead reductions come from both ICBMs and SLBMs and the bases to be inspected are chosen completely at random.

Obviously, as the amount of correct data supplied decreases, the probability of passing all inspections decreases. The data presented in Figure 1 show that with no NTM, the median of the data says that there less than a 2% chance of passing all inspections if the amount of supplied data is less than 90% correct. If the supplied data is less than 85% correct then there is less than a 1% chance of passing all inspections in 95% of the possible scenarios.

Figure 1 also demonstrates the value of NTM. As the amount of data that is known via NTM increases, the chance of passing all inspections decreases. If the supplied data is 96% correct, than with no NTM, there is a large probability, about 25%, of passing all inspections. However, if 50% of the data is known in advance, the probability decreases to 1.7%; and if 75% of the data is known, the chance of passing all inspections is less than 1 out of 1,000.

As the number of delivery vehicles at a given base increases and becomes very large, the discreteness of the data decreases, and eventually becomes negligible. At this point the probability to pass all inspections becomes (Percentage of correct data)^(Number of Inspections). The black lines represent this situation. The solid black line is the percentage of correct data to the 30th power, i.e. 30 inspections, and the black dashed line is the percentage of correct data to the 100th power. It is apparent that this simple formula can be used as a simple way to do an order of magnitude calculation to get an idea of the upper limit of the probability of passing all inspections.

The data shown in Figure 1 assumes that all the bases that were inspected and the order in which they were inspected were chosen completely at random. Not shown, for brevity, is the scenario that required that each submarine and every type of ICBM were inspected at least once. This restriction reduces the number of inspection configurations available and shrinks the area where 95% of the data lie. No significant change occurs to the median of the data. The data from the scenarios where reductions come from only SLBMs or ICBMs are nearly identical save for the primary difference of the size of the region containing 95% of the data. It is much larger in the case of the ICBMs reductions than the case of the SLBM warhead reductions. This is due to the larger number of configurations available when the reductions are taken solely from the ICBMs.

However, what is of real interest is the number of excess warheads that could result from not being able to detect errors in the supplied data. Figures 2, 3, and 4 show the probability of passing all inspections (vertical axis) as a function of the number of undeclared warheads (horizontal axis). Figure 2 shows the middle ground case when reductions occur evenly in both SLBMs and ICBMs while Figure 3 shows the scenario where reductions come solely from the SLBM fleet and Figure 4 shows data from the scenario where reductions come solely from the ICBM fleet. The solid lines represent the median of the simulated data and the shaded regions represent the area covered by 95% of the data. The red line and region represents the situation where no information from NTM is available, the blue represent the situation where 50% of the supplied data is known from NTM, and the green 75%. Note that the vertical axis of Figure 1 and the vertical axis of Figures 2, 3, 4 are identical.

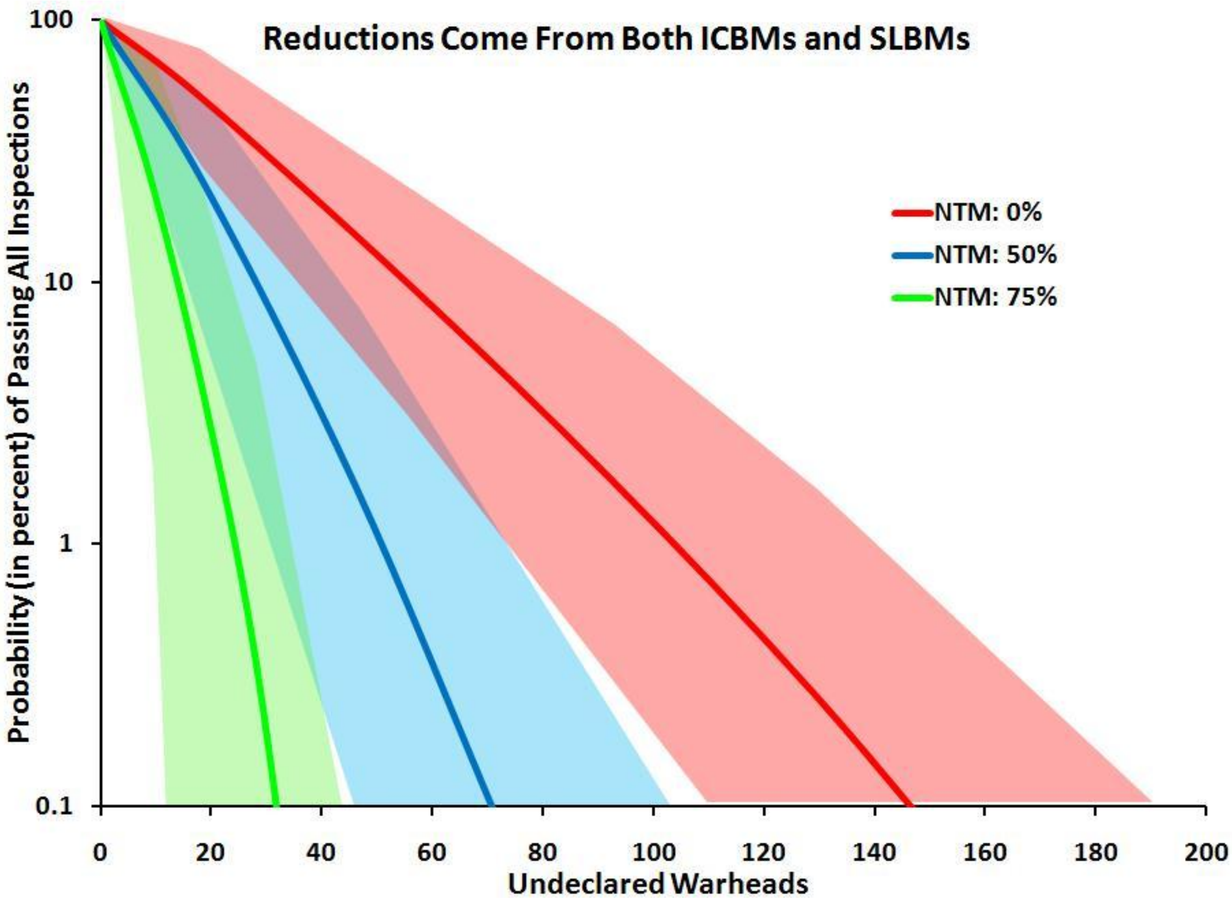

Figure 2: The probability of passing all inspections (vertical axis) as a function of the number of undeclared warheads (horizontal axis) if the reduction in the number of warheads comes from both SLBMs and ICBMs. It was assumed that the bases to be inspected and the order in which they were inspected were chosen completely at random. The solid lines represent the median of the data and the shaded areas contain 95% of the data. The red line and shaded region represents the scenario where no knowledge of the base is known prior to inspection, the blue 50%, and the green 75%.

Figure 2 shows for example, that if we have no prior knowledge from NTM of Russia's nuclear force structure and they have an excess of 140 undeclared warheads then in 95% of the structure scenarios there is less than a 1% chance that they will pass all inspections. If we know 50% of the data prior to inspection through our NTM, then there is the same probability of passing all inspections with only 80 undeclared warheads. If we know 75% ahead of time, then in 95% of the scenarios there is only a 1% chance that Russia will pass all inspections if they only have 40 excess undeclared warheads.

The data shown in Figure 1 represent the case where the bases were completely chosen at random. There is no requirement that each base be inspected, or each submarine be inspected, or one of every type of delivery vehicle be inspected. If it is required that each base and submarine be inspected at least once (and the rest of the inspections are performed randomly) these restrictions narrow the distribution of data and the region containing 95% of the data falls within the shaded region in Figure 1. The medians contain roughly 5% more warheads with this criteria.

Figures 3 and 4 show the data from the extreme cases where the reductions come solely from ICBMs (Figure 3) and solely from SLBMs (Figure 4). The lines and shaded regions correspond to their equivalents in Figure 2. Figures 3 and 4 required that each base and submarine be inspected at least once. They should also be read in the same fashion. The immediate difference to note is the relatively large distribution of the data the ICBM case and the relatively small distribution of data in the SLBM case. This is due to the assumption that if the number of warheads on a delivery vehicle is improperly stated, that the delivery vehicle is actually MIRVed. Because the SS-18s can carry up to ten warheads and the SS-19s can carry up to six warheads, there is a very large number of configurations possible. In the case of ICBM reductions, there was an average reduction of 3.5 warheads per SS-18 and 2.1 warheads per SS-19, and therefore there was an average excess of 3.5 undeclared warheads per SS-18 and 2.1 per SS-19. However, these are just averages and an SS-18 could have anywhere from one to nine excess warheads and a SS-19 could have anywhere between one and five excess warheads.

In contrast, the Delta III and Delta IV SLBMs can only carry up to three and four warheads respectively and required an average reduction of 1.8 warheads per Delta III and 2.4 from a Delta IV. But the range of undeclared excess warheads is only one or two for a Delta III and one to three for a Delta IV. This is a much more narrow range than in the case of the ICBMs and results in the relatively much narrower distribution of data.

This difference in MIRVing also explains the larger numbers of undeclared warheads in Figures 3 and 4 compared to Figure 2. When reductions came from both ICBMs and SLBMs the undeclared warheads were spread out over a larger number of delivery vehicles. Since they were spread out more, there was a greater chance to observe any incorrectly supplied data and therefore a smaller probability of passing all inspections.

An interesting feature that begins to show in Figures 2-4 is that as the NTM capabilities increase and the number of excess warheads increases there is a growing chance that there is no chance of passing a given inspection. This is shown very well in Figure 3 region where no NTM occurs. There is a sharp drop off once the number of undeclared warheads just above 90 undeclared warheads. The data at this point is somewhat uncertain because the exact point where the first configuration with 0% probability of passing all inspections was not determined, therefore the drop-off represents a likely scenario and should be thought of as a line to guide the eye.

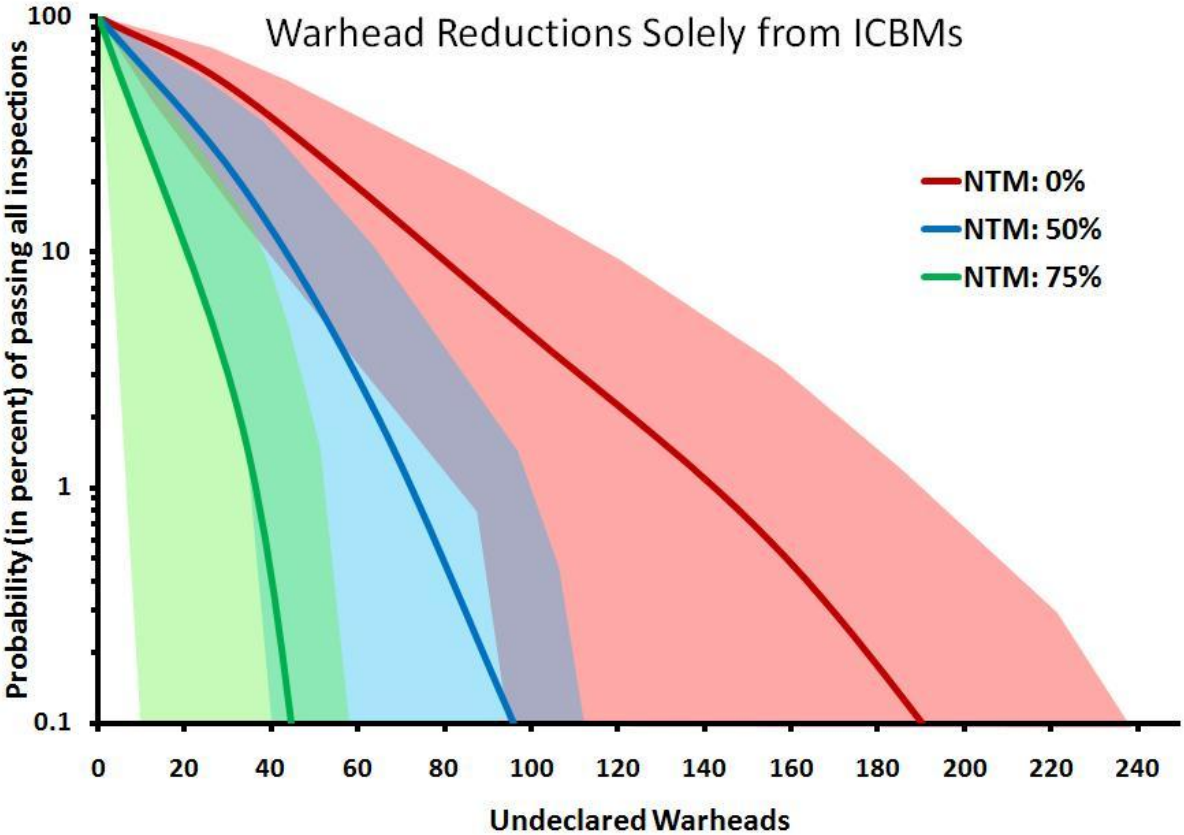

Figure 3: The probability of passing all inspections (vertical axis) as a function of the number of undeclared warheads (horizontal axis) if the reduction in the number of warheads comes solely from ICBMs. The solid lines represent the median of the data and the shaded areas contain 95% of the data. The red line and shaded region represents the scenario where no knowledge of the base is known prior to inspection, the blue 50%, and the green 75%. It was required that at least one inspection of each ICBM base occur.

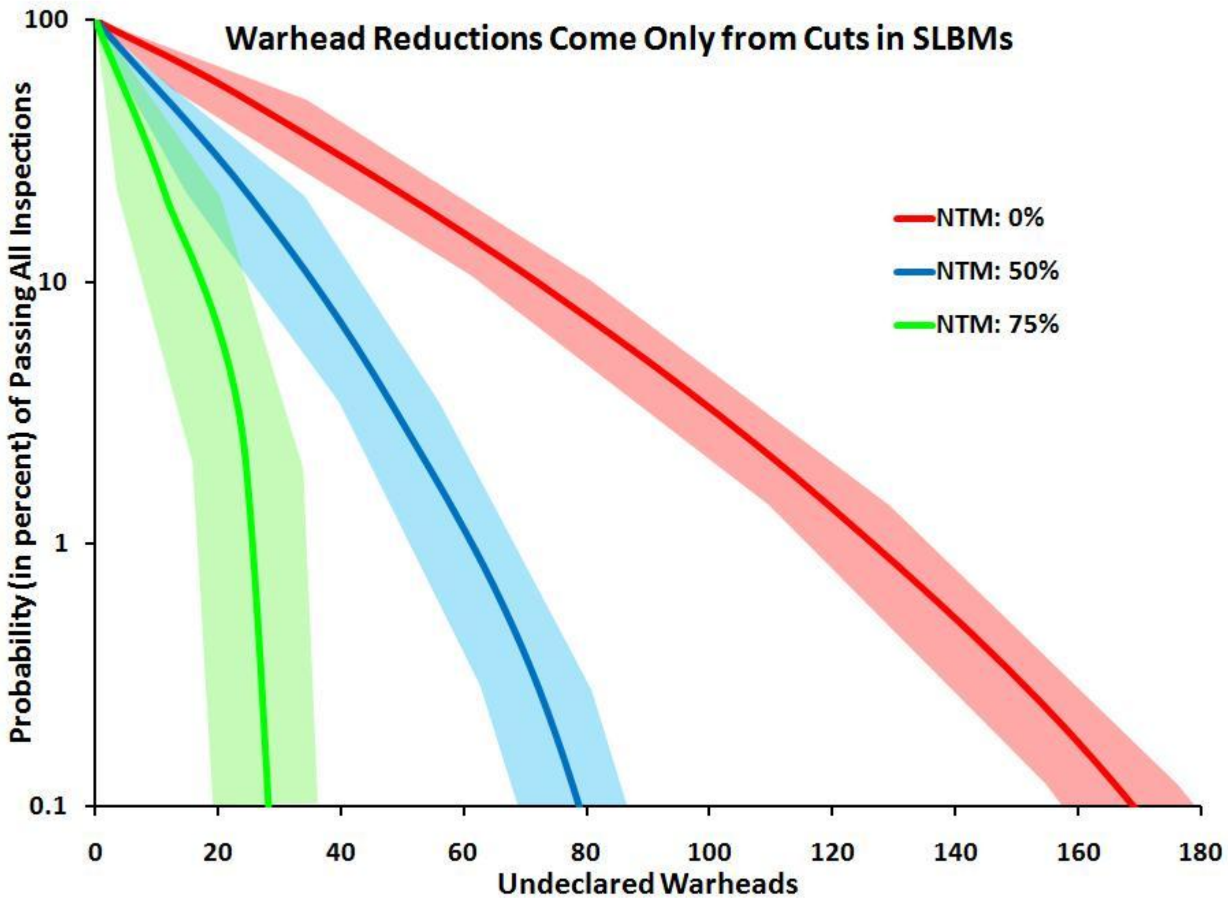

Figure 4: The probability of passing all inspections (vertical axis) as a function of the number of undeclared warheads (horizontal axis) if the reduction in the number of warheads comes solely from SLBMs. The solid lines represent the median of the data and the shaded areas contain 95% of the data. The red line and shaded region represents the scenario where no knowledge of the base is known prior to inspection, the blue 50%, and the green 75%.

Reduction Source	Median Number of Undeclared Warheads (0% NTM, 50% NTM, 75% NTM)		
	Detection Probability = 90%	Detection Probability = 99%	Detection Probability = 99.9%
Both ICBMs and SLBMs	58, 28, 15	104, 53, 24	146, 71, 33
ICBMs only	78, 45, 21	140, 73, 38	190, 96, 43
SLBMs only	75, 35, 18	127, 61, 25	168, 79, 28

Table 2: The median number of undeclared warheads broken down by probability of detection and amount of data known prior to inspection via NTM. The data shown are taken from Figures 2, 3, and 4. Note that the tabulated data are extrapolations between calculated data points.

Conclusions

There are a few conclusions that can be drawn from this work.

1) Because the probability of catching incorrect data increases, and the resulting number of excess warheads decreases, as the information gather by NTM increases, the first 70 Type I inspections are of upmost importance to maximize our confidence in our NTM. The effects of a a highly capable NTM can be seen in the data. In the situations investigated, when the United

States' NTM is greater than 75%, there is a very sharp drop off in the probability of passing all inspections as the number of undeclared warhead exceeds 40. Therefore, every effort should be made to maximize the capabilities in our NTM by using the first 70 Type I inspections to calibrate and test our NTM.

2) The data shown in Figure 2 show the case where absolutely no thought is given to which bases to inspect (assuming only ICBM and submarine bases are inspected). Should requirements be made that we inspect each base and/or submarine, the range of possibilities will shrink approximately 30% and the median of the possibilities will only increase slightly. Any thought that will be put into which bases to inspect will only improve our probability of detecting incorrectly supplied data.

3) The statement that Russia has a low probability of any military significant cheating contains two unknowns. What is considered a low probability and what is militarily significant? Using the data presented in this paper, if one of those unknowns is determined, the other variable can be estimated. It is the author's estimation that either the United States' National Technical Means allows us information of over 75% of the current state of Russia's arsenal, or that a large number of warheads is considered not militarily significant.

Works Cited

1. The New START Treaty; Senate Hearing 111-738. *GPO Access.* [Online] 2010. http://www.gpoaccess.gov/congress/index.html.
2. New START: Treaty Text. *Department of State.* [Online] http://www.state.gov/t/avc/newstart/c39903.htm.
3. *Russian Strategic Forces.* [Online] www.russianforces.org.