APPENDIXES

Appendix A
MONITORING NONZERO LIMITS

In chapter 6, our discussion addressed the monitoring of limits on banned systems, e.g., the SS-20.[1] In verifying compliance with a ban, one can follow the line of reasoning made familiar in discussions of the Intermediate Range Nuclear Forces (INF) Treaty, that ‘‘the detection of a single SS-20, and not even the missile, just the launcher or a single-bay garage that was supposed to have been eliminated by the treaty, would be a violation of the treaty. ”[2] Other treaty limitations allow certain systems, but restrict their numbers. Such nonzero limits are much harder to monitor through aerial surveillance than are the total bans addressed so far in this report. Not only is the ‘‘If we see even one, then it’s a violation” dictum inapplicable, but even the more sophisticated notion of extrapolation from a sample is likely to fail.

Reporting of the results of public opinion polls has accustomed Americans to the power of extrapolation from polling of a manageably small sample to deductions about large populations. Some pollsters state their sample size and provide an estimate of the accuracy of the result: a common protocol polls about 1,600 Americans and returns what is describedasa‘‘3-percent margin of error’ and a‘ ‘95-percent confidence level. The Gallup organization, for example, describes these parameters as reflecting a 19-to-l chance that the response of the entire American population, if taken, would not have differed from the response of the 1,600 Americans by more than three parts in one hundred.[3] One may reasonably wonder whether aerial monitoring could hold out the hope of providing similarly accurate data on treaty-limited (but not banned) items on the basis of the modest sample size available from a single flight. In general, the answer is that it cannot.

Such an extrapolation scheme would use an aerial monitoring flight to examine part of the region in which Treaty-Limited Items (TLIs) were deployed! TLIs in this region would be counted during the flight and then a “population estimate” for the whole region would be made on the basis of proportionality: if 10 percent of the deployment region had been inspected, the enumerated TLIs would be construed as 10 percent of the total, leading to an estimate for the total. The flaw of this scheme is that it assumes an even distribution of TLIs,[5] whereas there is no reason to think that such an assumption would be true, and several reasons to think that it would be false. Communications and logistical arrangements, for example, might well be eased by the operation of mobile missiles in groups. More subtly, the missiles’ effectiveness as a deterrent would be enhanced by bunching them up and thus linking their fates: the plainer of a barrage or reconnaissance-strike attack would then have to contend with the possibility that all of the targets would survive.[7] Thus bunching has to be considered likely, weakening the aerial observer’s ability to estimate the total number of deployed TLIs based upon the number observed in some part of the deployment region. Finally, the other side might take deliberate steps to make sure that the sample population was-in one or more respects--simply not representative of the whole. A clever treaty violator could slant the results of inspections, allowing enough TLIs to be seen that the inspecting side would conclude that a plausibly large, but treaty-compliant, force had been fielded when in fact the true force was far larger than allowed by the treaty. In public-opinion polling, there is no “other side” to take such steps.[8]

In a similar vein, some have suggested that treaty veri.tiers borrow the statistical methods used by industrial “quality assurance” specialists. These methods also offer little hope in the case of monitoring treaty compliance, because they address the question of how to feed back information gained from product inspections into the manufacturing process, so as to reduce the number of defects produced. These methods are inapplicable in arms control treaty monitoring, because one side does the

[1] The SS-20 is banned under INF.

[2] The point was made in this instance by Major General William F. Burns, *The INF Treaty*, hearings before the Senate Committee on Foreign Relations, U.S. Senate, 100th Cong., 2d sess., Part 5, p. 88.

[3] Gallup on-line service. This description would strike a classical statistician as lacking in mathematical rigor because it comes close to assigning a probability to something which is in fact either true or false: the proposition that the response of the population as a whole is within 3 percent of that of the sample.

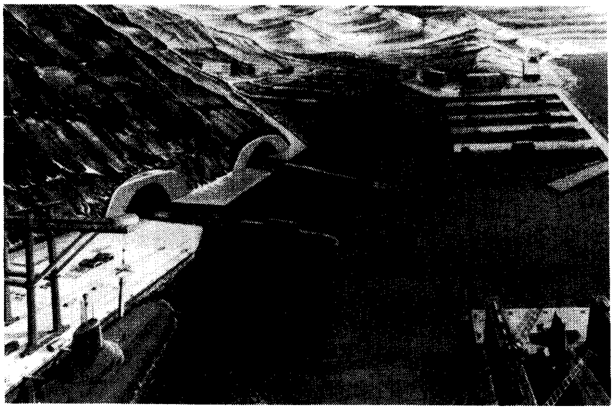
[4] Some treaty arrangements would facilitate Such monitoring by creating prescribed deployment regions fOr, say, land-mobile intercontinental ballistic missiles (ICBMs).

[5] As the pollsters caution US, sampling error is only one source of error in the final estimate.

[6] A logical mission for a mobile ICBM, with its heavy investment in survivability.

[7] This possibility—resulting from the time in which he happens to overlook the one subregion that contained the entire force—would loom larger in the attacker’s mind than the contrary case, in which all of the targets are found and destroyed.

[8] Though some pollsters have explained the discrepancy between their results and those on Election Day by an alleged reluctance on the part of voters to admit that they intend to vote for especially controversial candidates.



credi concep So ubm
Depa tmen De

Ba m au ch g bma d
b bidd b a m a C g m
p a pica as mo g
m e m gai a a ce a mo
allowed number exist even though not all
can be seen at once.

monitoring and the other side does the compliance (or violation).[9]

Tags

The **use of *tags*** (nontransferrable, nonduplicable identifying markers--analogous to automobile parking stickers--issued by the inspecting side to the owning side) has been proposed to facilitate monitoring. Tags might be made readable from aerial monitoring aircraft.[10] A tagging scheme would avoid difficulties posed by potential bunching of TLIs because it would not rely on statistical sampling: enough tags are handed out to account for the allowed number of TLIs, and any TLI found without a tag is presumed to be a violation of the treaty.

Even in the absence of a detected violation, however, statistical analysis of the tags found on the legitimate TLIs could enhance the tags' utility as a verification device. One such procedure,[11] conceptually similar to the capture-release-capture protocol used to estimate bird populations, would consider two successive "takes" of tag numbers as samples of the larger population. Those TLIs seen in the first "take" are considered to be "banded." Assuming that the chance of a TLI being seen in the second "take" is not increased or decreased by the inclusion of the TLI in the first "take," the proportion of "banded" TLIs in the second "take" ought to be equal to their proportion of the population as a whole. For example, if the first 'take' identified 30 TLIs, 2 of which appeared in a second "take" of 20 TLIs, one would conclude that the 30 TLIs represented 10 percent of the total population, yielding an estimate of 300 for the total. If this total is less than the allowed total, then (assuming that the other side has not voluntarily sacrificed some TLIs) the estimated "total population" is not the true total population but only the total subpopulation of which these observations have been made. In this case, the aerial monitoring needs to be expanded because its scope does not even cover all the legitimate TLIs. A total *greater* **than the allowed** total would indicate that, for some reason, TLIs seen in the first take were less likely than others to be seen in the second "take"; this finding would suggest that TLIs are being rotated through the region subjected to aerial monitoring, an ominous prospect.

With the possible exception of flights examining TLIs restricted to designated deployment areas, a single aerial monitoring flight is unlikely to see enough TLIs for the "banding" approach to be used. A generalization of the above method, however, could deal with the very small takes-perhaps of only one TLI even on a "good day' '-expected under some aerial monitoring regimes. Under this generalization, each sighting of a TLI would be logged and periodically-perhaps annually-the sightings would be totaled so as to create a listing of those TLIs seen once, those seen twice, those seen three times, and so on.

One would hardly expect sightings to be absolutely evenly distributed among TLIs. Through chance alone, some will be seen more than others. The *Poisson distribution* tells us how much "clumsiness" to expect in the repeat sightings. Fitting the observations to a Poisson distribution would reveal any departure from the expectation that the tendency of a TLI to be sighted is unrelated to its previous history of sightings. If the data failed to conform to a Poisson distribution because of an unduly small proportion of repeat sightings, one would have reason to suspect "hot-bunking,' '[12] or some other form of rotating TLIs through the region subjected to aerial monitoring. If, on the other hand, the data departed from a Poisson distribution by virtue of an overly large propensity of TLIs to be sighted repeatedly, then one

9As a saying now favored by quality assurance specialists holds, "Quality cannot be 'inspected in.' " The import of this saying is that inspection can only hope to fiiter out rejects, and cannot be used as a means of actually adding quality. Quality can only be added by manufacturers, not inspectors. Similarly, treaty compliance cannot be 'inspected in.' Moreover, the producers and the inspectors are, in the case of treaty compliance, on different sides. Such a contrast between acceptance testing and treaty monitoring was drawn by Patricia M. Lewis in "Verification of Conventional Forces in Europe," *IEEE Technology and Society Magazine,* December 1990/January 1991, p. 11.

10Tags are briefly discussed in OTA's unclassified *Verification Technologies: Measures for Monitoring Compliance with the START Treaty—Summary,* OTA-ISC-479 (Washington, DC: U.S. Government Printing Office, December 1990).

11Proposed b, Stephen Davis in "Verification and Compliance for Arms Control," *Comparative Strategy,* vol. 9, No. 4, 1990; pp. 44)3-413.

12In this method of cheating, an operating base supports excess TLIs by keeping some in the field at all times.

would suspect that the inspection process was somehow being manipulated so that the same TLIs, once seen, were presented over and over.

If the numbers of TLIs seen five, four, three, two, and one times each seem consistent with a Poisson distribution, one would be justified in extrapolating to a number seen zero times, and thus to an estimate of the total population size. The Poisson-based approach is particularly attractive because of this self-checking feature, by which the applicability of the imputed Poisson distribution can be checked in the cases of repeated sightings before it is used to estimate the number of TLIs never sighted.

Additionally, the Poisson system does not rely on any knowledge of how many tags have been given out.[13] For this reason, as well as the self-checking feature, it is promising as a means of interpreting sightings of "buddy tags.' ' Buddy tags, analogous to automobile license plates, uniquely identify the TLIs with which they are associated, yet (in some schemes) forego the elaborate precautions against duplication and transfer that complicate many conventional tagging schemes. Unlike these more technologically ambitious tags, buddy tags could easily be made large enough that they could be read from an airplane.

[13] For this reason, *it* would be well suited *to use in* conjunction with *intrinsic tags* (innate identifying characteristics Of individual TLIs—analogous to fingerprints—catalogued by the inspecting side). These hold out the promise of great security from illicit transfer or duplication. However, most proposed intrinsic tags, e.g. the marks left on the TLI by the tools used to make it, require such close examination of the TLI as to preclude the aerial reading of the tags.