"REVIEWED"

DESIGN FOR EVALUATION*

Arnold Court
California State University
Northridge CA 91330

Abstract. Non-randomized operational weather modification projects can be evaluated meaningfully only if they have been thoroughly planned in advance, stating precise intentions or goals in form amenable to later verification by procedures which also are specified in detail and use observations likewise fully specified, up to alternative actions in case objectives change or desired data are not available.

Designing an operational weather modification project so that its results can be determined is considered impossible by many at either end of the problem: the operating meteorologist who runs the program day-to-day and even hour-to-hour, and the practicing statistician who is asked to evaluate it when it's over. The meteorologist may say that he can't design the project because he doesn't know how it will develop, that goals and methods for attaining them will change as the weather unfolds, and so on. The statistician may insist that no valid conclusions can be drawn, with any sort of probabilistic confidence statement, unless the project is strictly randomized. But randomization, which requires that a substantial fraction of the apparently favorable opportunities be left untreated, usually is opposed by the meteorologist, and especially his client, who wants as much snow or water -- or as little hail -- as he can have.

While each of these attitudes is not without merit, the problem isn't hopeless. To obtain the best possible assessment of his degree of success, the meteorologist must plan carefully, in collaboration with his client and his statistician. First, he must think like a lawyer, who is (among many other things) a specialist in enumerating all possible eventualities and arranging accordingly. He must agree with his client as to the precise purpose of the project, without, of course, promising to accomplish it. For snowpack augmentation, 50 inches of water equivalent on 1 April, without blocking roads, might be specified. For a spring and summer rain increasing effort, the goal might be 10 inches in each of three months, but no more than 3 inches in any one day or 6 inches in three consecutive days.

These goals are then translated twice. The first is to the actual intent of the operation, often expressed not in actual amounts but as percentage increases of what would otherwise occur. The second translation is into a procedure for estimating these increases, positive or negative. This is where the statistician becomes involved, before the project starts.

*Presented at annual meeting, Weather Modification Association, Santa Barbara CA 24 April 1980.

Statistics provides perfectly precise answers to questions that weren't asked, and may be of no interest. This can happen unless advance planning properly expresses in statistical terms the physical questions of interest. Beginning students often are mystified by the basic concept of statistical testing, the null hypothesis: Why assume no effect, when the real question is how big it is. They soon learn that the null hypothesis can be rejected with greater precision than the increase can be estimated, and that a null hypothesis needn't specify equality: it can be that A does not exceed B by more than two inches, or 20 percent, or some such criterion. So the operator's intentions must be phrased in terms amenable to statistical test.

Measurement of the results also requires careful planning, the most detailed of all. If the goal is areal precipitation, its method of computation must be specified, down to alternate procedures if a precipitation gage or snow course is moved, lost, or abandoned. Changes in purpose, such as curtailing or extending the target area, must also be considered, and proper plans made.

Without randomization, the most crucial part of any weather modification evaluation is the method for estimating what would have happened if treatment had not been applied. Usually this is by regression, using predictors from outside the target, far enough away that the treatment doesn't affect them.

"Upwind control areas" aren't necessarily unaffected: winds at the surface, or near the tropopause, may be opposite to those at cloud level. Pre-treatment conditions, on the ground or in the atmosphere, are less suspect, but may be inappropriate for an operation lasting several days. Furthermore, they may have figured in the decision to initiate treatment.

Any regression procedure for estimating untreated conditions in the target should be based only on prior situations which would have been suitable for treatment: operational success can't be judged in comparison to all previous circumstances, many of which weren't treatable.

However, an entire month or season, part or all of which was treated, can be compared with corresponding untreated months, if general similarity is indicated by correspondence in storm types, air masses, circulation indices, etc. Again, such criteria for equivalence should have been specified before the project started, and certainly before its results have been examined in detail. The best and simplest way to avoid any bias, or even appearance of bias, is to specify in full detail the precise procedures to be used in acquiring and averaging data, inside and outside the target, and the statistical tests and criteria for the evaluation.

These complications are among the many reasons that randomization is the preferred method of evaluation: errors in the procedure for estimating area precipitation, or other criteria, tend to be the same for treated as control situations. When randomization is precluded, much more care must be taken to minimize these possible sources of error. Whether they can all be eliminated, or rendered insignificant, depends on the particular operation. Any intentional rain increase on an isolated island, without comparable control areas nearby, and without frequent radiosonde observations, may be completely undetectable. The same program over a single Iowa county, without other activities nearby, may, with proper advance planning, be amenable to rather stringent evaluation.

Unfortunately, most operational projects begin without adequate planning, and usually have barely enough funds even for proper operation, none for advance planning and careful evaluation. In such circumstances, all the operator can say is that he did essentially what he was paid to do--the best he could. If the customer wants a thorough evaluation of the results of the non-randomized project, he must pay for it substantially, in money for the detailed plan of analysis that must be prepared in advance, and in lead time for such proper experimental design.