Refinements of Adaptive Sampling in Site Decontamination

Myron J. Katzoff, Abera Wouhib and Joe Fred Gonzalez, Jr.
National Center for Health Statistics, 3311 Toledo Road,Hyattsville, Maryland 20782

Contact: Myron J. Katzoff (mjk5@cdc.gov)

Key Words: adaptive sampling, simulation, microparticle contamination

1. Introduction

In this paper, we report on our investigation of the properties of a few basic adaptive sampling procedures for site decontamination with the aid of a computer simulation model. The idea of using adaptive sampling procedures in the manner discussed here was suggested in [1]. We are particularly concerned with the situations where almost nothing may be known about building features such as air-flow paths or foot-traffic patterns. The sampling procedures are adaptive in the respect that they take advantage of statistically relevant information for adding sample units that becomes available before, during or after the selection of an initial sample. Our greatest interest is in maximizing the proportion of population units that contain lethal concentrations of a pathogen that are contained in the final sample. Adaptive procedures ensure that a higher proportion of these "important" population units are contained in the final sample than would be the case when using a corresponding non-adaptive procedure. This sample (unit) coverage feature of adaptive procedures has not been fully exploited in the applications that are of interest to us. Our investigations indicate that sample coverage depends significantly on several features of the initial sample; in particular, the initial sample must be well-dispersed throughout the collection of elementary sample units. This paper will discuss the strengths and deficiencies of the basic procedures we have investigated and how their effectiveness depends upon design-features of the initial sample.

2. Review of Simulation Model Structure

In [2], we described our computer simulation model as consisting of the following modules:

(1) define and label sampling units;

(2) create spore-clouds or clusters;

(3) select an initial sample;

(4) define and apply criteria for linking and adding sampling units to the initial sample; and

(5) formulate and quantify estimators.

In module (1), a building was represented as a large parallelipiped which was further subdivided into the cells of a uniform grid to represent offices or rooms inside the building. Our hypothetical building has ten floors on each of which there are ten corridors with ten rooms; thus, our hypothetical building has 1000 equal size rooms. In module (2), two clouds or clusters of spores were created as elliptical clouds of sample points from two trivariate normal distributions with randomly chosen location parameters corresponding to points within the boundaries of the building but with the same covariance matrix. The "spores" of one cloud result from 200,000 random draws from one of these distributions; similarly, the second cloud was created by making 300,000 draws from the second distribution. Points that fell outside of the building boundaries were simply discarded leaving a total of 479,000 sample points over the interior of the building. Module (3) defined the procedure for selecting an initial sample of units (cells, offices or rooms) inside the building; and it is this module to which we give different realizations in this paper. Module (4) articulates the rule for adaptively adding sample units throughout all the simulations:

> Examine each cell in the initial sample to determine if its spore-count exceeds the Critical Minimum Number (CMN) spore-count; and when it does, add cells which share a common boundary with it to the sample; check the spore-count of the cells just added and add the cells adjacent to them if they too contain more than the CMN spore-count; the process of adding cells continues until no further cells can be added thus producing the final sample of cells.

The views expressed in this paper are those of the authors and do not necessarily reflect the policies of the National Center for Health Statistics.

Our simulations implement the simplifying assumption that there exists a CMN spore-count which is sufficient to cause anthrax in humans. (In fact, the lethal exposure can vary from individual to individual and may, indeed, depend upon the values of variables which measure the health status of a person.) For our simulations, we used a CMN of 500 so that a cell with less than 500 spores was treated as not having a harmful concentration of spores. In module (5), for each run of the simulation, we tabulate the numbers of rooms associated with each cloud that contain more than the CMN spore-count and that were included in the initial sample, the total number of rooms associated with each cloud after adaptively adding units that were included in the final sample and the corresponding spore-counts for all the rooms from each cloud in the final sample.

Additional details on the computer model structure can be found in [2]. Further details on the outputs generated by module (5) will be provided in our discussion of simulation results.

3. Simulation Results

Four design variations for drawing an initial sample were studied with the computer simulation model: (1) three-stage simple random sampling (SRS); (2) stratified sampling; (3) systematic sampling; and (4) zone sampling. In each case, initial samples of size 100 were drawn without replacement. The most extensively studied procedure was the first: three-stage SRS without replacement. Our examination of this procedure allowed for two specifications of the Gaussian cloud covariance matrix: one for which the determinant was 0.49 and a second, which created a pair of more dispersed clouds, for which the determinant was 7.29. The performance of the adaptive procedure did not degrade with increased dispersion (as measured by the determinant of the covariance matrix); in fact, the increased dispersion yielded slightly better coverage of the final sample because of greater numbers of "important" units in the population itself. The sample selection options also provided for examination of the consequences of including the units corresponding to the mean vectors with certainty to emulate prior knowledge of the source-units where contamination first occurred. Taking mean-vector units with certainty guaranteed the location of all the critical units regardless of which initial sample was drawn. Since such knowledge may not always be available when authorities demand commencement of the decontamination process, it was natural to study sampling procedure outcomes absent that information.

In the tables presented in this paper, we show results for the first 50 simulations in which a total of 145 rooms were known to have spore counts in excess of the CMN. These are indicative of our general conclusions given in the following section. In Table 1, there is a summary for each simulation run for the three-stage SRS case. Even though we observed that this design provides samples with better coverage than samples drawn in accordance with ordinary SRS (not shown), the full collection of (100) runs for Table 1 indicates that adaptively adding units would uncover all 145 rooms only about 67 % of the time. It was ascertained from inspection of the tables that the first design does not yield coverage as effective as design variations (2) and (4). Moreover, tinkering with the sampling parameters suggested that the three-stage SRS design would, in general, require larger sample sizes than alternatives (2) and (4) to be fully effective. Tables 2 and 3 indicate that stratification is beneficial even if only half as many strata are created in one instance as in another Table 4 shows that if the spreading of microparticles from source points is credible, great care should be exercised in the use of systematic selection. The example shown of an initial sample of 100 units with a skip interval of 10 illustrates that ignoring the geometry of the structure can yield extremely unsatisfactory outcomes. (The reader should note that systematically sampling every tenth unit results in stacks or slabs of units in which the position of the first unit determines the location of the entire slab. From this, we intuit that as cloud dispersion decreases, the likelihood that the use of those parameters in systematic sampling will yield poor results or be unproductive altogether increases.) Sampling in zones (or groupings of consecutive units) makes extensive use of the lesson about stratification. The outputs in Table 5 reflect the results obtained by creating 100 zones of 10 consecutive units each and choosing one unit at random in each zone. (Note that this is a departure from zone sampling as described in [3], pp.127-130. For the purpose of assessing the variance of certain key estimates, it might be prudent to create only half as many zones and to select two units without replacement per zone; thus, for our example, we would create 50 zones and choose two units per zone. For that option, observe that it is easy to assign numerical labels to the units in the building and that the units of the initial sample can be designated by a well-known sequential procedure in advance of testing for the presence of spores.)

4. Conclusions

Controlling the spread of the initial sample(*i.e*, maximizing the spread) and random selection are essential for a high frequency of success with the sampling procedures we have studied. If spores disperse in a manner similar to our Gaussian assumption, adaptive spatial sampling can be an effective method for locating places where there are lethal concentrations of spores. Since it may be possible to trace the movements of persons who have had exposures that made them sick, precise information about where they were exposed can be worked into an adaptive procedure to enhance its effectiveness. Although we have not shown results for the case where "clumping" occurs because of environmental conditions, adaptive sampling procedures investigated with our model do not seem to lose their effectiveness in the case of clumping. As a final point we note that better knowledge of the physics of interzone airflows for a structure would increase the range of adaptive procedures that are applicable in microparticle removal; this would be especially helpful in developing link-tracing designs for handling a truly large problem.

Table 1: Iterations for Three-Stage SRS

It.	cloud1	rooms1	cloud2	rooms2	rooms	spores
1	5	75	9	70	145	452686
2	3	75	24	70	145	452686
3	11	75	7	70	145	452686
4	7	75	15	70	145	452686
5	16	75	0	0	75	271710
6	7	75	10	70	145	452686
7	19	75	4	70	145	452686
8	9	75	7	70	145	452686
9	0	0	6	70	70	180976
10	7	75	11	70	145	452686
11	14	75	2	70	145	452686
12	4	75	18	70	145	452686
13	8	75	5	70	145	452686
14	17	75	6	70	145	452686
15	8	75	3	70	145	452686
16	18	75	6	70	145	452686
17	5	75	7	70	145	452686
18	8	75	3	70	145	452686
19	3	75	1	70	145	452686
20	3	75	0	0	75	271710
21	14	75	0	0	75	271710
22	0	0	18	70	70	180976
23	7	75	9	70	145	452686
24	1	75	12	70	145	452686
25	11	75	0	0	75	271710
26	9	75	3	70	145	452686
27	16	75	10	70	145	452686
28	3	75	9	70	145	452686
29	19	75	3	70	145	452686
30	9	75	3	70	145	452686
31	5	75	11	70	145	452686
32	7	75	3	70	145	452686
33	12	75	10	70	145	452686
34	20	75	7	70	145	452686
35	2	75	6	70	145	452686
36	24	75	5	70	145	452686
37	0	0	13	70	70	180976
38	8	75	0	0	75	271710
39	8	75	5	70	145	452686
40	0	0	19	70	70	180976
41	7	75	0	0	75	271710
42	10	75	3	70	145	452686
43	9	75	0	0	75	271710
44	0	0	7	70	70	180976
45	6	75	0	0	75	271710
46	1	75	3	70	145	452686
47	20	75	0	0	75	271710
48	5	75	27	70	145	452686
49	10	75	6	70	145	452686
50	0	0	14	70	70	180976

References

[1] Katzoff, Myron J.; Sirken, Monroe G. and Thompson, Steven K.(2002). Proposals for Adaptive and Link-Tracing designs in Health Surveys. *Proceedings of the Section on Survey Methods Research, American Statistical Association*, p.1772.

[2] Katzoff, Myron J.; Wouhib, Abera; and Gonzalez, Joe Fred, Jr.(2003). Using Design-Based Adaptive Sampling Procedures in Site Decontamination. *35th Symposium on The Interface: Computing Science and Statistics, security and infrastructure protection.*

[3] Kish, L.(1965). *Survey Sampling.* John Wiley & Sons, New York.

[4] Thompson,S.K.(1992). *Sampling.* New York: John Wiley & Sons.

[5] Thompson, S. and Frank, O.(2000). Model-based estimation with link-tracing sampling designs. *Survey Methodology*, **26**, 87-98.

Table 2: Each Floor As A Stratum

It.	cloud1	rooms1	cloud2	rooms2	rooms	spores
1	10	75	8	70	145	452686
2	5	75	1	70	145	452686
3	10	75	6	70	145	452686
4	11	75	5	70	145	452686
5	8	75	9	70	145	452686
6	9	75	5	70	145	452686
7	13	75	7	70	145	452686
8	10	75	4	70	145	452686
9	8	75	6	70	145	452686
10	9	75	6	70	145	452686
11	5	75	12	70	145	452686
12	12	75	9	70	145	452686
13	7	75	4	70	145	452686
14	8	75	6	70	145	452686
15	11	75	4	70	145	452686
16	8	75	13	70	145	452686
17	5	75	4	70	145	452686
18	12	75	10	70	145	452686
19	8	75	5	70	145	452686
20	7	75	8	70	145	452686
21	4	75	7	70	145	452686
22	12	75	4	70	145	452686
23	5	75	6	70	145	452686
24	5	75	8	70	145	452686
25	11	75	4	70	145	452686
26	6	75	5	70	145	452686
27	10	75	6	70	145	452686
28	4	75	7	70	145	452686
29	3	75	9	70	145	452686
30	12	75	4	70	145	452686
31	2	75	7	70	145	452686
32	7	75	7	70	145	452686
33	10	75	5	70	145	452686
34	4	75	6	70	145	452686
35	9	75	9	70	145	452686
36	11	75	4	70	145	452686
37	4	75	4	70	145	452686
38	11	75	3	70	145	452686
39	5	75	4	70	145	452686
40	14	75	14	70	145	452686
41	5	75	8	70	145	452686
42	6	75	9	70	145	452686
43	6	75	4	70	145	452686
44	8	75	14	70	145	452686
45	7	75	8	70	145	452686
46	5	75	11	70	145	452686
47	5	75	10	70	145	452686
48	8	75	10	70	145	452686
49	7	75	11	70	145	452686
50	15	75	8	70	145	452686

Table 3: Consecutive Pairs of Floors As Strata

It.	cloud1	rooms1	cloud2	rooms2	rooms	spores
1	3	75	7	70	145	452686
2	11	75	7	70	145	452686
3	8	75	13	70	145	452686
4	4	75	7	70	145	452686
5	6	75	6	70	145	452686
6	13	75	5	70	145	452686
7	7	75	10	70	145	452686
8	9	75	7	70	145	452686
9	11	75	4	70	145	452686
10	15	75	12	70	145	452686
11	5	75	5	70	145	452686
12	7	75	11	70	145	452686
13	12	75	9	70	145	452686
14	10	75	12	70	145	452686
15	8	75	5	70	145	452686
16	5	75	1	70	145	452686
17	7	75	3	70	145	452686
18	9	75	6	70	145	452686
19	7	75	4	70	145	452686
20	14	75	4	70	145	452686
21	14	75	3	70	145	452686
22	14	75	6	70	145	452686
23	2	75	4	70	145	452686
24	5	75	8	70	145	452686
25	6	75	5	70	145	452686
26	8	75	13	70	145	452686
27	3	75	11	70	145	452686
28	4	75	8	70	145	452686
29	6	75	6	70	145	452686
30	5	75	8	70	145	452686
31	8	75	9	70	145	452686
32	8	75	7	70	145	452686
33	7	75	7	70	145	452686
34	7	75	8	70	145	452686
35	5	75	7	70	145	452686
36	4	75	13	70	145	452686
37	7	75	5	70	145	452686
38	12	75	12	70	145	452686
39	10	75	4	70	145	452686
40	5	75	9	70	145	452686
41	6	75	6	70	145	452686
42	3	75	6	70	145	452686
43	14	75	9	70	145	452686
44	10	75	6	70	145	452686
45	9	75	10	70	145	452686
46	10	75	8	70	145	452686
47	9	75	9	70	145	452686
48	11	75	1	70	145	452686
49	4	75	9	70	145	452686
50	10	75	9	70	145	452686

Table 4: An Example with Systematic Sampling

It.	cloud1	rooms1	cloud2	rooms2	rooms	spores
1	0	0	27	70	70	180976
2	29	75	0	0	75	271710
3	1	75	21	70	145	452686
4	0	0	22	70	70	180976
5	0	0	0	0	0	0
6	0	0	27	70	70	180976
7	0	0	0	0	0	0
8	25	75	0	0	75	271710
9	0	0	0	0	0	0
10	1	75	21	70	145	452686
11	1	75	21	70	145	452686
12	0	0	0	0	0	0
13	0	0	27	70	70	180976
14	0	0	22	70	70	180976
15	0	0	22	70	70	180976
16	0	0	0	0	0	0
17	1	75	21	70	145	452686
18	0	0	27	70	70	180976
19	0	0	0	0	0	0
20	0	0	0	0	0	0
21	0	0	0	0	0	0
22	20	75	0	0	75	271710
23	0	0	27	70	70	180976
24	0	0	0	0	0	0
25	0	0	0	0	0	0
26	0	0	0	0	0	0
27	20	75	0	0	75	271710
28	0	0	0	0	0	0
29	0	0	0	0	0	0
30	0	0	0	0	0	0
31	0	0	0	0	0	0
32	25	75	0	0	75	271710
33	25	75	0	0	75	271710
34	25	75	0	0	75	271710
35	0	0	22	70	70	180976
36	20	75	0	0	75	271710
37	20	75	0	0	75	271710
38	1	75	21	70	145	452686
39	0	0	0	0	0	0
40	0	0	0	0	0	0
41	25	75	0	0	75	271710
42	0	0	22	70	70	180976
43	1	75	21	70	145	452686
44	0	0	0	0	0	0
45	20	75	0	0	75	271710
46	0	0	0	0	0	0
47	1	75	21	70	145	452686
48	0	0	22	70	70	180976
49	0	0	0	0	0	0
50	29	75	0	0	75	271710

Table 5: Zone Sampling: One Unit Drawn Randomly Per Zone

It.	cloud1	rooms1	cloud2	rooms2	rooms	spores
1	12	75	7	70	145	452686
2	7	75	6	70	145	452686
3	6	75	4	70	145	452686
4	9	75	6	70	145	452686
5	11	75	7	70	145	452686
6	7	75	8	70	145	452686
7	5	75	7	70	145	452686
8	7	75	5	70	145	452686
9	3	75	6	70	145	452686
10	10	75	6	70	145	452686
11	9	75	4	70	145	452686
12	9	75	9	70	145	452686
13	8	75	7	70	145	452686
14	10	75	9	70	145	452686
15	9	75	5	70	145	452686
16	9	75	8	70	145	452686
17	4	75	6	70	145	452686
18	10	75	5	70	145	452686
19	5	75	8	70	145	452686
20	1	75	4	70	145	452686
21	8	75	5	70	145	452686
22	10	75	10	70	145	452686
23	6	75	7	70	145	452686
24	4	75	5	70	145	452686
25	7	75	5	70	145	452686
26	9	75	5	70	145	452686
27	5	75	4	70	145	452686
28	8	75	6	70	145	452686
29	4	75	7	70	145	452686
30	10	75	6	70	145	452686
31	10	75	10	70	145	452686
32	6	75	10	70	145	452686
33	6	75	7	70	145	452686
34	9	75	7	70	145	452686
35	10	75	7	70	145	452686
36	9	75	9	70	145	452686
37	6	75	5	70	145	452686
38	9	75	10	70	145	452686
39	5	75	8	70	145	452686
40	5	75	7	70	145	452686
41	7	75	7	70	145	452686
42	7	75	4	70	145	452686
43	7	75	4	70	145	452686
44	6	75	5	70	145	452686
45	7	75	6	70	145	452686
46	12	75	5	70	145	452686
47	10	75	4	70	145	452686
48	9	75	4	70	145	452686
49	10	75	5	70	145	452686
50	6	75	7	70	145	452686