Chapter 2

The Front End of Super fund: Site Discovery and Evaluation

CONTENTS

Boxes

Figure

Table

The Front End of Superfund: Site Discovery and Evaluation

INTRODUCTION

Before major site studies are done and long before cleanups start, uncontrolled sites spend 5 or more years in the front end of the Superfund pipeline. First someone has to discover a potential site, then others evaluate it to determine whether or not the Federal program will take action.

In this chapter OTA reviews both the site discovery and evaluation aspects of the Superfund program; how site discovery has remained the same and site evaluation changed during the first 8 years of Superfund, and the consequences for the future. OTA concludes that **Superfund's environmental mission could be measurably enhanced by the addition of a national site discovery program and the revamping of the evaluation process.** Both changes will involve spending more money at the front end of Superfund but can mean saving more money over the long run. The upfront costs of having better information sooner and making sounder environmental decisions, instead of expedient management choices, are small, compared to the cost of a few Superfund cleanups.

National Site Discovery Through Technology

Almost a decade after obvious problems like Love Canal drove the Federal Government to get actively involved in cleaning up environmental damage, we still do not really know how many sites need attention. As Russell Train, former administrator of EPA, said in 1987 in answer to the question, How serious is the problem? "Distressingly little is known, in fact, about the number of toxic waste sites and how serious a risk each site poses.

We do not know how many sites there are because there has been no comprehensive or systematic search for them. The sites that have been discovered are listed on various Federal and State inventories for which there are no listing criteria nor attempts to avoid duplications among lists.

Despite haphazard and passive site discovery, over 31,000 sites have been reported to EPA's CERCLA Information System (CERCLIS) inventory of sites that potentially require cleanup. Over the years of the program, 1,274 of these sites have been proposed for the National Priorities List (NPL); the list now includes 1,224 sites. All sites on the NPL qualify for cleanup funding under the Superfund program, except the 115 sites belonging to Federal agencies.

For some years EPA has maintained that a maximum of about 2,000 sites will ultimately be Superfund's responsibility. Those estimates have always been constrained by the choices that EPA has made about the Superfund program. OTA said in 1985 that ultimately 10,000 sites might be on the NPL. That estimate, still valid, assumed that EPA would have an active site discovery program and would apply environmental criteria to the site evaluation process. It did not take EPA's deferral of cleanups to other programs into account. Today's CERCLIS inventory and its growth rate--despite passive site discovery-imply that over 4,000 sites could be on the NPL by the year 2000.

Enough sites to keep the Superfund program challenged? Of course, but the real question ought to be: Have we found them all? OTA and many others don't think so, and data from the Superfund program shows that the worst sites are not necessarily accounted for. Both site discovery and the way sites are inventoried can be improved—and the costs of doing so contained-by the use of technology.

Active, comprehensive, and systematic site discovery could be built around a method such as historical aerial photography analysis. A national program could be supplemented by State efforts using traditional site discovery methods. With funding assistance from the Federal Government and in combination with improved preremedial site analysis, this could move the Nation quickly to finally knowing with more certainty the true size of the cleanup problem. But the history of the Superfund program so far tells us that a **comprehensive site discovery program will not occur unless Congress gives explicit direction to EPA, or some other authority, to proceed.**

[1]Russell Train, ' 'Big Questions Facing the Cleanup, " *EPA Journal*, January/February 1987, p. 8.

Site Evaluation: A Management or Environmental Decision?

Most observers express great concern with one point in Superfund's site evaluation process: the on/off NPL decision, whether or not a site is on the NPL. Few, however, seem to care about what happens before that point. OTA concludes that there are many reasons to suspect that the entire site evaluation process is biased against making sound environmental decisions.

Should the Superfund program bear the burden of **determining** whether or not discovered sites are potential environmental threats, even though the program may not have the authority or resources to respond to all those threats? Or, should the Superfund program's site evaluation process be restricted to only finding potential Superfund sites? If so, how early in the process is it reasonable to make that decision? When it is made, how can the public best be notified when a site is rejected for management reasons even though it poses a threat to public health and the environment?

There is evidence that past Superfund site evaluations have produced many false negative decisions. That means that sites have been rejected when they really require cleanup. **OTA estimates that from 240 to 2,000 false negatives may exist so far.** Recent changes in the process-meant to cope with demands to work faster-are likely to aggravate this problem. Another characteristic of site evaluation is regional inconsistency; across the Nation wide differences exist in the efficiency and apparent environmental effectiveness of site evaluation.

SITE DISCOVERY AND INVENTORIES

This section sums up what the Nation knows and does not know about potential sites; how many sites *might* need to be cleaned up? The ad hoc nature of EPA and State searches and reporting systems is a result of EPA (and Congress) not having paid much attention to (or spent much money on) site discovery since the Superfund program began in 1981. In the interim and without knowing the full extent of the potential universe, **EPA** (and Congress) have concentrated on evaluating sites that are known.

Further, what we do know about the potential size of the national problem is confused by a multiplicity of site inventories. OTA has identified other national inventories, besides the well-known CERCLIS and NPL lists associated with the Superfund program, In addition, each State has some kind of list or lists. Today, there is no way to know the extent of overlap of these lists. This situation will be exacerbated if EPA's proposed deferral policy (see ch. 4) for Superfund sites goes into effect. Increased deferral of sites out of the Superfund program and into other programs will create dynamic lists of sites as the authority for cleanup changes and sites are moved back and forth among programs. Accountability at the national level will become more difficult. Some sites deferred to other programs will essentially disappear. Not only are the other cleanup programs less visible to Congress and the public than is Superfund, but few of their site inventory systems are as available. Some programs do not have formal inventories.

Why Should We *Want to Know?*

The major, longstanding bureaucratic argument against site discovery has been that we know enough without it to keep us busy for years. But this argument ignores the environmental mission of the Superfund program. And it perpetuates the crisis atmosphere around Superfund. Without a strategy to decide which sites get attention first, newly discovered sites that engender a lot of publicity tend to push existing work aside.

The number of sites potentially requiring any kind of cleanup defines the magnitude of the national problem, its ultimate cost and length. The number of sites *actually* needing cleanup will always be less than any inventory of known *potential* sites, such as CERCLIS. But, all potential sites consume national resources because they all require some type of evaluation to separate out duplicate and obviously nonhazardous sites from hazardous sites requiring some kind of attention.

From a policy perspective, knowledge of the outside bounds of site cleanup determines the necessary scope and, therefore, acceptable pace and financial impact of a national program. Unrealistically low estimates of potential sites lead to underestimates of sites requiring cleanup and low estimates of resource needs. If a site is a problem today, it will probably be a worse problem tomorrow. **Ignoring potential sites only saves resources in the short term and decreases protection of human health and the environment.** Overestimates of the size of the national problem that overwhelm systems and resources result in an unnecessarily slow pace of cleanup.

A relatively small number of potentiai sites may mean that limited funds can be used quickly to clean up the universe of sites, but a small number devalues the importance of a priority system for determining which sites get attention sooner. The larger the number of potential sites becomes, more care must be taken in assigning priorities to sites and managing limited funds. Issues of environmental protection and cost-effectiveness become more difficult to balance; thoughtful development of a long-term strategy becomes critical.

What Have We Done So Far?

In 1982, the General Accounting Office (GAO) stated that ''a national hazardous waste site inventory does not exist. "[2] In 1985, GAO stated, "A complete inventory of hazardous waste sites does not exist. And, in late 1987, GAO said, "While still not fully understood, the extent of the nation's potential hazardous waste problem appears to be much larger than is indicated by EPA's inventory of sites.

These statements are not surprising since an EPA official stated in 1981 that a comprehensive search for sites needing cleanup was "against EPA policy. ''5 This attitude was still policy in 1985 when the then director of the Superfund program said:

> There is no national policy that says go out and aggressively look at sites. I'm not sure, if we had such a policy, that we would have more to deal with then we currently do, however. The national inventory (not the NPL) has grown about 3000 sites a year. It is growing faster than we have the resources to assess and inspect those that come to our attention, I'm not sure, frankly, what more we could do. Or how one would go about actively investigating for the presence of new sites ...6

Congressional Direction and Funding

Congress has been largely silent on site discovery. In CERCLA, little attention is paid to site discovery except to say that the NCP shall include ''methods for discovering and investigating facilities. "[7] GAO and **OTA analyses** at the time of Superfund reauthorization concluded that the 'methods' being used by EPA were not producing comprehensive information. But site discovery was not art issue during reauthorization and nothing was added in the Superfund Amendments and Reauthorization Act of 1986 (SARA) to prompt more action from EPA.

Sufficient site discovery funding levels have been *authorized* under the Resource Conservation and Recovery Act (RCRA) since 1980 but Congress has only *appropriated the* funds once,[8] When the authorized level was $20 million, Congress appropriated $10 million for fiscal year 1983 for State site inventory programs. Subsequently, Congress raised the authorized level to $25 million per year for fiscal years 1985 through 1988. None of that money was ever appropriated. If it had been appropriated *and* directed toward site discovery, the question of how large the potential universe is might have been answered by now.

The fiscal year 1983 funds were actually drawn from the Superfund trust fund. They were earmarked

---

2U.S. General Accounting Office, ''Environmental Protection Agency's Progress in Implementing the Superfund Program, ' GAO/CED-82-91.

3U.S. General Accounting Office, *EPA's Inventory of Potential Hazardous Wastes Sites Is Incomplete,* GAO/RCED/-85-75 (Gaithersburg, MD: U.S. General Accounting Office, Mar. 26, 1985).

4U.S. General Accounting Office, *Superfund: Extent of Nation's Potential Hazardous Waste Problem Still Unknown,* GAO/RCED/-88-44 (Gaithersburg, MD: U.S. General Accounting Office, December 1987).

5Morgan Kinghorn, then EPA Comptroller, as noted in telephone log of Vem Webb, then director, Environmental Photographic Interpretation Center, Aug. 25, 1981.

6William N. Hedeman, Jr., as quoted in ''Superfund Chief Outlines Strategy on Hazardous Waste cleanups,' *Chemical & Engineering News,* June 3, 1985, p. 17.

7CERCLA Section 105(1).

8RCRA Section 3012, Hazardous Waste Site Inventory for site discovery *and* evaluation. Congress referred to that appropriation as a "one time event. '

There was no followup accounting by EPA (or examination by Congress) to ascertain how useful the funding was for site discovery purposes. However, EPA has said that activities were funded ". . . in the following order of priority: preliminary assessments, site inspections, responsible party searches, **discovery,** and site inspection follow-up" [emphasis added].[10] Given such low priority, little of the $10 million was likely spent on site discovery.

EPA and Site Discovery

EPA has never requested funds from Congress for site discovery. EPA has no site discovery program, has no budget for site discovery, and does not allow States to spend Superfund monies for site discovery. Instead, EPA has relied on varied State-funded efforts and a few regional investigations to identify potential sites.

Traditionally, EPA officials give two reasons why the program devotes no effort to site discovery. First, there is no need for a site discovery program because the extent of the problem is known; the worst sites have been found already. Second, the Superfund program has enough to do just evaluating the known sites. Discovering more would simply choke the system.

EPA has implied that some kind of a site discovery program once existed. In 1984, EPA told Congress that it had shifted the emphasis in the Superfund program away from site discovery. 'These changes reflect EPA's belief that many of the sites posing more serious problems have been identified and EPA resources should increasingly focus on further assessment and inspection of these sites. ''[11] **Today, EPA's policy is the same even though it has abandoned the idea that the extent of the problem is known. GAO** reported in 1987:

> EPA officials now recognize that many more hazardous waste sites may exist, [but] they believe a higher priority is to meet the deadlines imposed by SARA for assessing and evaluating those sites already included in the CERCLIS inventory .12

In other words, **resources and SARA schedules are determining the size of the inventory rather than the inventory size establishing the funding level.** This management, rather than environmental, perspective of the CERCLIS inventory ignores the higher future costs (both resources and public health and environmental damage) of not fully understanding the extent of the problem today. Moreover, limiting the size of the inventory and the NPL diminishes the need to develop a strategy for distinguishing between near- and long-term remedial actions.

EPA reported on the extent of site discovery to Congress in 1984 with a list of site discovery methods in use: [13]

1. required reporting under CERCLA Section 103 of known sites (following enactment of CERCLA) or subsequent releases;
2. government investigation under CERCLA response authority;
3. reporting by permit holders under other statutes when required;
4. inventory efforts (i.e., RCRA Section 3012) or random observations and reports; and
5. ''other sources,'' including formal analysis of various industries.

EPA said the one-time reporting requirement of Section 103 (#1 above) had been a major source of initial site discoveries. Since then, EPA said it was

[9] House of Representatives, Conference Report [appropriations for HUD and Independent Agencies for fiscal year 1983], H.Rep. 97-891, Sept. 29, 1982.

[10] U.S. Environmental protection Agency, ''The Effectiveness of the Superfund Program, CERCLA Section 301(4(1)(A) Study,'' December 1984, p. 1-14. According to Frank Wolle, who was then with EPA's Environmental Photographic Interpretation Center, the funds were actually used to redo RCRA inventories conducted in 1976 and 1980 that were considered to be invalid. Thus, the monies were not spent on gathering any new information.

[11] Ibid., p. 1-13.

[12] U.S. General Accounting Office, *Superfund: Extent of the Nation's Potential Hazardous Waste Problem Still Unknown,* op. cit., footnote 4, p. 28. In a rare public statement, an EPA official from Region 2 lamented on national TV in July 1989 that just as we think we know where all the sites are another one becomes known.

[13] U.S. Environmental Protection Agency, ''The Effectiveness of the Superfund Program, CERCLA Section 301(a)(l)(A) Study,' op. cit., footnote 10. These "methods" are the same as those listed by EPA in the National Contingency Plan.

relying on inventory efforts (#4) and other sources (#5) for discovery. For the future, EPA said it was:

> .,. currently developing a method to systematically evaluate various industries to determine categories of waste generators which are more likely to involve hazardous release problems that require Superfund action.[14]

The new method may have been developed, but industries have not been systematically evaluated by EPA since 1984. OTA could locate only one industry study: "U.S. Production of Manufactured Gases: Assessment of Past Disposal Practices," completed in 1988.

EPA did contract with Booz-Allen & Hamilton to conduct a study on site discovery. The results are contained in a 1987 draft report, which begins:

> To date, discovery and identification of releases or threatened releases of hazardous waste have been reported to the EPA through a wide variety of uncoordinated channels . . . The magnitude of the hazardous waste problem on a national level remains unknown and EPA is unable to forecast the resources required to understand and mediate this problem. An active site discovery program is needed to better forecast future EPA resources and schedules in which Superfund's overall objectives may be met.[15]

The study remains in draft form; the need expressed in it for an ''active site discovery program' is not EPA policy and does not reflect official EPA thinking. Two followup studies mentioned in the report-to develop management options and guidance for a site discovery program-were not done. Followup work was confined to reviewing discovery techniques and drafting some guidance.

The Aborted 200 Cities Plan

There were some technical people in EPA in the early days of the Superfund program who saw value in site discovery. EPA's Environmental Photographic Interpretation Center (EPIC) had a comprehensive site discovery program-the 200 Cities Hazardous Waste Site Discovery Plan-underway in 1980. It lasted barely a year; in August 1981, EPA's comptroller told the laboratory to cancel the project because "we already have more sites than there is money for so we do not need more.[16] EPIC asked the Superfund program in March 1982 to support its $850,000 request for funding the project. According to the director of EPIC, the Superfund program director told him that there was not to be a 200 Cities project.

There was some initial support for the EPIC project within EPA's Office of Research and Development (ORD). EPIC had $161,500 in fiscal year 1981 for a pilot program and received guidelines from ORD for an expanded program in June 1981. The guideline document called for the 200 city search to be completed and 50 sites analyzed by 1985. An accelerated program was also outlined in which 100 of the most serious sites found would receive detailed analysis. Total funding through fiscal year 1986 was estimated at $6.4 million for discovery and site analysis. Of the four city inventories selected for the pilot project, two were finished.

The 200 Cities plan was based on EPIC's experiences since 1973 in developing the use of historical aerial photography for environmental purposes. By the late 1970s, it was an established technique for site characterization. One intent of the initial 200 Cities pilot project was to demonstrate the practicality of using the technique to conduct a *site discovery* program and to determine if the criteria for selecting the 200 cities were valid (see box 2-A). Figure 2-1 shows how a series of aerial photographs, the basis for HAP analysis, can uncover past disposal practices no longer visible.

State Site Discovery

State site discovery efforts are funded by States. States have become, by default, fully responsible for collecting national data. But EPA does not allow States to use Cooperative Agreement (CA) money for site discovery, because CERCLA Section 104b confines its planning authority to only *site-specific* actions. Thus, since site discovery actions-by their very nature--cannot be site specific, EPA reasons that States cannot use CA funds for site discovery

[14] Ibid., p. 1-8.

[15] U.S. Environmental Protection Agency, Office of Solid Waste and Emergency Response, ''CERCLA Site Discovery Program Evaluation," draft prepared by Booz, Allen & Hamilton, Inc., Mar. 5, 1987, p. 1. EPA told OTA that the report never went from draft to final form because of resource constraints in the preremedial program.

[16] Morgan Kinghorn, then EPA comptroller, as noted in telephone log of Vern Webb, then EPIC director, Aug. 25, 1981.

Box 2-A *Historical Aerial Photography* and Site Discovery

Site discovery using historical aerial photographs allows a trained analyst to peel back layers of construction and vegetation in a particular area that have occurred over time and hide past practices and to review those practices. This process of looking back in time uncovers sites where wastes have been dumped. The technique cannot confirm that hazardous constituents are present at a site. Instead, a probability of their presence is inferred from activities observed or if the site has remained scarred over a long period of time and does not support vegetation.

Historical Aerial Views of the United States

The United States has been extensively photographed from the air since at least the 1920s and especially after 1938. The photographs have been taken primarily for mapmaking and soil survey purposes. After World War II, aerial photography increased with the availability of surplus military reconnaissance equipment and pilots and crewman trained in the technique during the war. For maps, a series of overlapping photographs are taken of the terrain as an aircraft progresses along a straight flight path. The photographs are taken from a height of 12,000 feet and each one covers an area 11.5 square miles. Viewed through a stereoscopic microscope, a three-dimensional view of the terrain is obtained. Towns and cities undergoing the greatest development and expansion have been the most recoded areas, generally on a three-to-five-year cycle.

The collection of historical aerial photography (HAP) is now stored in five government archives: the National Archives in Washington, DC; the Earth Resources Observation System in Sioux Falls, South Dakota; the USDA's Agricultural Conservation and Stabilization Service in Salt Lake City, Utah; the National Atmospheric and Oceanic Administration in Rockville, Maryland; and the Tennessee Wiley Authority's Mapping Services Branch in Chattanooga, Tennessee. Additional photos and index services are sold through commercial firms. In the past, some commercial negatives have been destroyed to reclaim the silver value. For instance, during the speculation of silver by the Hunt brothers in the late 1970s, the value of the silver in film rose ten-fold and millions of feet of commercial photographic records were lost. An unknown amount of aerial photographic film was destroyed as well.

The use of historical aerial photography for purposes other than mapmaking is not new. In the past, it has been used to:

- analyze coastline erosion or change over time;
- note changes in watersheds following darn construction;
- collect evidence for litigation involving land change, such as wetlands; and
  locate historical landmarks, such as Civil War fortifications.

When EPA's Environmental Photographic Interpretation Center (EPIC) was established in 1973 part of its charter was to develop techniques and methods for the extraction of environmental information from the historical aerial photographic libraries of other Federal agencies. At the time, many people considered historical aerial photography to be of no value in addressing environmental problems. Instead, the prevalent belief was that aerial photographs could be used to record current environmenal events such as oil spills, fires, and agricultural practices rather than uncover practices of the past. Work at EPIC has proven the value of HAP analysis as a site discovery technique.

Using Historical Aerial Photography for Site Discovery

The process of using aerial photographs to locate unknown hazardous waste disposal sites involves: 1) selecting an area of high probability; 2) acquiring a historical series of aerial photographs of the area; 3) analyzing the time series of photographs for indicators of hazardous waste activities; and 4) for any area where indicators are four@ following up the analysis with ground investigations to ascertain whether hazardous substances are present.

Areas of High Probability-For the 200 Cities project at EPIC, a number of criteria were used to select areas with the highest potential of having abandoned hazardous wastes. Criteria included: 1) knowledge of where hazardous waste had been generated through manufacturing, 2) knowledge of traditional waste disposal areas, 3) consideration of transportation methods and corridors between the 1930s and 1960s, and 4) cancer incidence rates. Four cities (one each in EPA Regions 1,2,3, and 4) were selected for the pilot project: Worcester, Massachusetts; Buffalo, New York; Charleston, West Virginia; and Chattanooga, Tennessee. They were chosen because of their population density, concentration of chemical manufacturers or users, and health data.

Selection of Photographs-A time series of photographs, usually in 4-to-5-year increments, is selected and printed **from negative transparencies.** A **current aerial photograph may have to be taken.**

Analysis of Photo Series--A photo analyst, using a backlighted table and magnifying stereo optics, scans each **transparency for indicators of hazardous waste activities,** such as ground scars, indiscriminate dumping, waste ponds, landfills, quarries, ground stains, junk yards, etc. By viewing successive years of photographs, an analyst can see changes that have occured over time and are no longer visible on the surface.

Ground Investigations--Like all discovery methods, once a suspect site is located through photo analysis, the site must be investigated on the ground. For instance, all sites placed in the CERCLIS inventory must undergo a Preliminary Assessment to determine if there is any evidence of an environmental problem and to assure that the site reported is not a duplicate of another one in CERCLIS. In the case of sites discovered through HAP, the same process must occur. However, HAP sites also provide information not available through other discovery methods, such as the exact location and extent of the possible contamination.

purposes. But EPA has not invoked this interpretation to explain its own lack of attention to site discovery, and there is no reason to believe that, if EPA allowed States to use CA funds for site discovery, anyone—including Congress-would be concerned.

EPA also does not allow States to use Core Program Cooperative Agreement (CPCA) funds—money that is explicitly *not* site specific-for site discovery. In this case, EPA reasoned:

> Site discovery is not eligible for funding at this time. It is OERR's [Office of Emergency and Remedial Response] opinion that all sites presently listed in CERCLIS should be addressed prior to any funding for site discovery under the CPCA.[17]

This decision by EPA was made despite admitting, in the same memorandum, that CPCAs are a result of congressional "intent to increase the scope of Cooperative Agreements. ' '[18] The congressional SARA Conference Report mentions "site inventory and assessment efforts" as a class of activity that may be included in these expanded Cooperative Agreements.[19]

The consequence of delegating site discovery and funding to States is that there are 50 different site discovery programs. According to the 1987 Booz-Allen/ EPA study:

> [State] discovery efforts have not been consistent. They range from minimal efforts with little interest for change to an active state effort based on state "Superfund" legislation, which may result in site lists larger than CERCLIS.[20]

Eighteen out of fifty States surveyed by Booz-Allen claimed to have an active, as opposed to a passive, program. Booz-Allen's review of the site **discovery** methods used by States, however, shows that passive methods predominate. Citizen complaints topped the list of methods (47 States) with referrals from other programs second (38 States). Third on the list (22 States) was some kind of survey review (e.g., a records search), which, of the three most used methods, is the only active one. Other active methods (e.g., aerial photography, property transfer regulations, and special studies) were used by less than a quarter of the States.

Comprehensive, Active Site Discovery

The existing structure for site discovery and listing of sites that may need cleanup is largely a disconnected maze. Various authorities (Federal and State) seek out sites. The degree of effort and comprehensiveness each applies to the task varies widely; each approach lies along a spectrum from passive to somewhat aggressive. Once discovered, knowledge about potential sites is handled in many

[17] U.S. Environmental Protection Agency, "Final Guidance on State Core Program Funding Cooperative Agreements,'' Directive 9375.2-01, Dec. 18, 1987, transmittal memorandum p. 3.

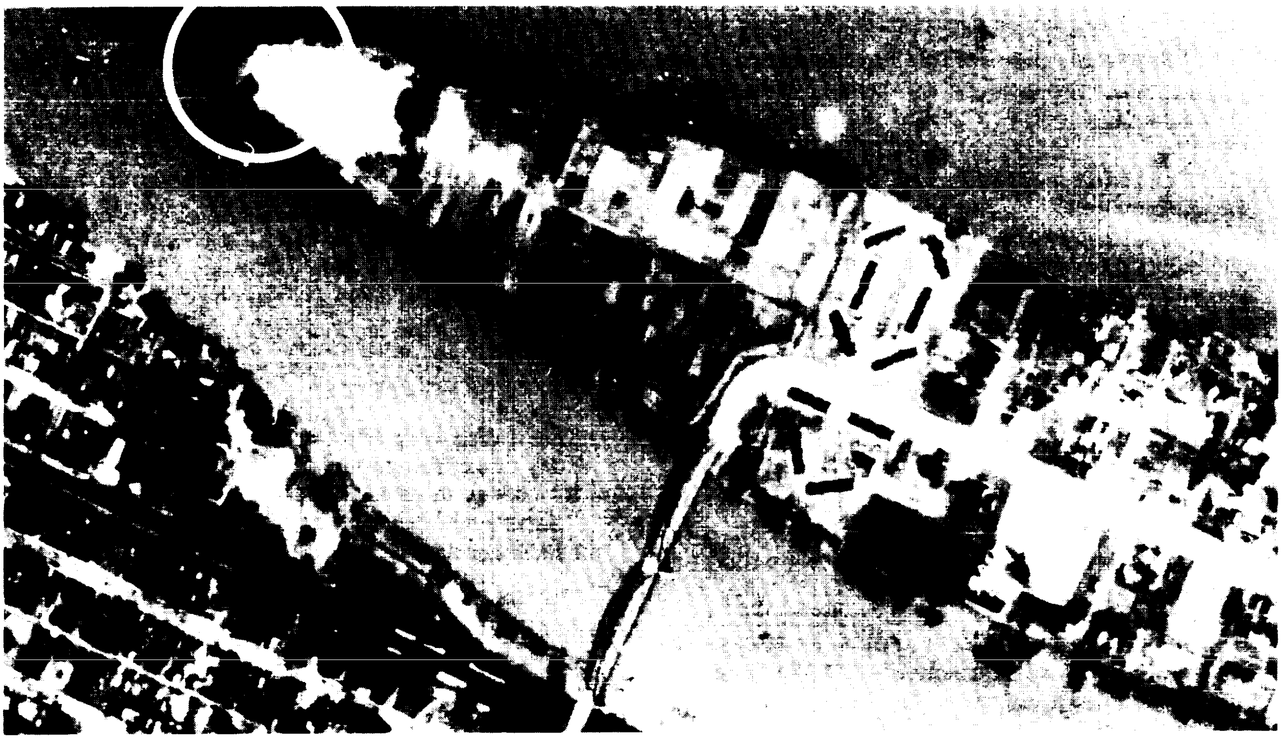
[18] Ibid., p. 1.

[19] House of Representative, Conference Report [on amendments to CERCLA], H.R. Rep. 99-962, p. 195.

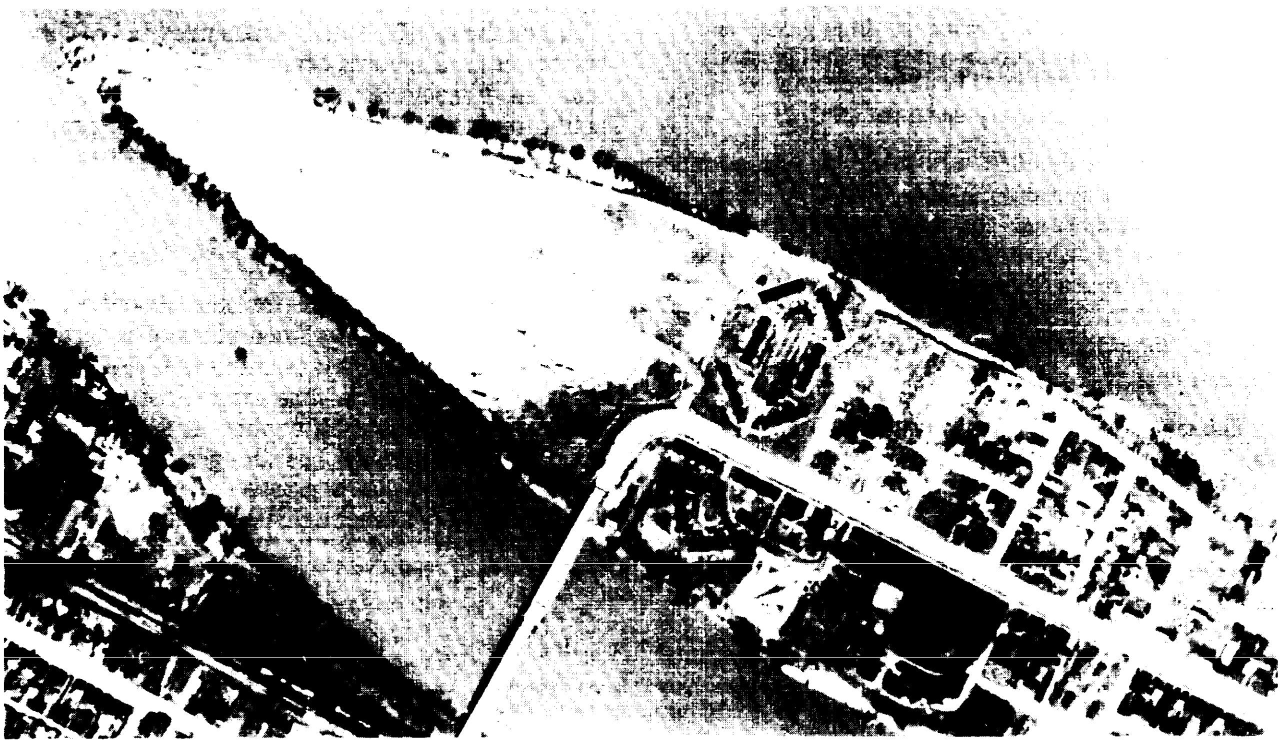
[20] U.S. Env'ironmental Protection Agency, ''CERCLA Site Discovery Program Evaluation, " op. cit., footnote 15, p. 26. OTA found in a review of many Stale programs' annual reports that site discovery efforts were rarely mentioned. The topic, thus, appears to be of little interest at the State level.

20-011 0 - 89 - 3 : QL 3

Figure 2-l--Land Disposal In the 19509 Hidden **by Land Use In the 1970s**



1948



1952

These three photographs are from a larger series of historical aerial photographs of one end of Neville Island in the Ohio River near Pittsburgh, PA. A number of chemical plants are located at the opposite end of the island. Across from t he island are industrial plants. A 1938 photograph (not included) shows that before the war the island was a large farm. Over time, land use on the island changed to a mix of residential, light and heavy industrial, and recreation.

Even viewing the photographs without the aid of a stereoscopic microscope, as is done in HAP analysis, it is possible to see activity taking place overtime t hat eventually created a need for a cleanup. In the first photograph-taken in April 1948--victory gradens from WWII still remain the dominate feature, although at the tip of the island (circled area) t he gardens have been destroyed and t he surface appears disrupted. It may indicate the beginning of land disposal. In the second photograph—April 1952—farming here has ceased and this portion of the island has become a disposal **site** for solid and liquid wastes.

By October 1973 (the third photograph), however, visual evidence of dumping has almost disappeared. Vegetation has returned to most of the area, although in the circled area **(where liquid dumping** has taken place) the soil does not support much vegetation. By 1979 the area had been donated to the local community for a park. Cabins, roads, trails, and a parking lot further obscured past waste disposal practices.

1973

ways, from entering the information into a database in a timely manner to letting the information sit in a file drawer for years. Communication among all these discovery and inventory systems ranges from nonexistent to formal handoffs. Thus, sites discovered in one system but judged not eligible may be officially turned over to another authority or may be simply ignored, even though some cleanup may appear necessary.

Often site discovery methods do not respect regulatory or legal boundaries. Historical aerial photography (HAP), for instance, discovers sites on evidence of land use over time. HAP does not necessarily distinguish between underground tanks that store a nonhazardous or hazardous material. HAP does not distinguish between a RCRA hazardous waste and a CERCLA hazardous substance. HAP also does not tell whether the private or public sector is responsible for creating a problem or whether or not the owner of the property has had a RCRA permit or not. It does, instead, provide some *environmental* evidence of a problem.

The undiscriminating nature of many site discovery methods suggests that it is better to think of site discovery not as the responsibility of Superfund or RCRA or some other program but as a national responsibility. A national site discovery program, whether solely a Federal function or in cooperation with States, could produce inventories of potential sites to feed into a variety of cleanup programs.

A comprehensive, active site discovery program will, of course, cost money. But, as discussed above, such a program can be cost-effective in the long run, and the costs seem small compared to what is routinely spent to study and clean up individual sites. From among a variety of methods used in the past to discover sites, some could be part of an active program, and existing technology can be adopted to minimize the immediate program costs. OTA has estimated, for instance, that resurrecting and retargeting the 200 Cities plan today would cost about $100 million over a 5-year period. The cost includes detailed analysis of an estimated 7,500 found sites—part of the work now done by EPA during the preliminary assessment (PA) phase of site evaluation.

Once sites are discovered, the information found needs to be aggregated to make the necessary followup efficient and to avoid duplication by EPA and others. One solution is to develop a nationally consistent listing system or compatible ones that can be used by all cleanup authorities (see later discussion). Once sites have been evaluated as needing some cleanup, OTA has suggested (see ch. 1, option 18) that a national cleanup list be maintained.

Technical Options for Comprehensive Searches

In heavily populated areas, it may be true that obvious uncontrolled sites have been found. People have smelled them or seen them leaking.[21] To discover those sites not yet emitting odors or liquids, techniques can be applied that sweep large geographic areas and collect information about what may be occurring under the surface. Such ways include using knowledge about contaminated ground and surface waters as indicators and peeling back layers of earth through historical aerial photo analysis.[22] California has conducted studies in two counties using the contaminated water technique. EPA and others have used historical aerial photo analysis. OTA has chosen to review the latter for its potential as a cornerstone for a national site discovery program.

HAPPI: A Historical Aerial Photography Program Initiative

Despite the halt of the original EPIC pilot project to initiate the proposed 200 Cities plan in 1981, EPA and others have used historical aerial photography (HAP) to discover, verify, and characterize hazardous waste sites. The technique has not been used comprehensively for site discovery, however, Instead, various government authorities (i.e., the U.S. Army, some EPA regions, and some State governments) have used the method on an occasional basis or for special projects.

[21] Or, have followed up some intuition. This happened in Arlington, VA, in 1989 when a piece of industrial property was being reviewed for sale to the U.S. Navy. Testing based on suspicion proved that PCBs had been disposed on the land, presumably by a metal recycling firm that had leased the property.

[22] Remote sensing by satellite is often thought of as another way, however, the technique has shortcomings such as poor resolution and the inability to view subsurface conditions.

OTA has evaluated some of the projects and found that most evidence suggests the technique, when properly applied, is a very useful tool. Because of past inattention and lack of follow up evaluation, a new pilot program is advisable, especially one that carries sites through at least a PA type of evaluation. Ultimately, the best use for HAP would probably be in examining areas with the highest probability of having abandoned hazardous wastes since cost prohibits and logic denies a blanket examination of the entire country. HAP is one of the few methods of site discovery that can systematically analyze a given area for unknown and abandoned sites. One analyst says, ''Historical aerial photographs may be the only source of reliable information for identifying active and inactive landfills. ' '[23] Another expert calls HAP the most ''impersonally recorded document available' for conducting inventories.[24]

Other methods tend to be biased and limited by the information used for discovery. For instance, surveys are limited by the accuracy and completeness of responses, historical document searches are limited by how well records were kept and have been maintained, and property transfer regulations are limited by the types of property (usually existing commercial or industrial) included in the regulations and only are triggered if property is sold.

Historical aerial photography analysis is limited by the extent of the availability of historical photographs and trained analysts. As long as the existing archives of aerial photographs are preserved and maintained, there does not appear to be any shortage of photographs. Trained analysts are another matter; in the initial stages of a major government program, demand could outstrip supply.

If the pilot phase had been completed before the 200 Cities plan had been canceled, a wealth of data would be available today to analyze its potential in full. Since the termination of the 200 Cities plan, EPIC has used aerial photography analysis almost exclusively (about 90 percent of its work) for site characterization for EPA's RCRA and Superfund programs. Still, site discovery work has not been totally abandoned, and OTA has based its analysis on available information. We have included in our analysis: 1) existing results from the 200 Cities project; 2) the U.S. Army's Toxics and Hazardous Materials Agency work under the Facilities Restoration Program; 3) a Monroe County, New York, project; 4) a Memphis, Tennessee, emergency response project; 5) an EPIC inventory of Love Canal, New York; and 6) several inventories conducted for EPA Regions 2, 3 and 4. Brief summaries of the projects are presented in box 2-B.

These projects show that the HAP method locates sites found through other discovery methods and adds sites to existing inventories generated through other methods. In other words, HAP **could replace many commonly used site discovery methods.** There is no information available to determine whether the HAP method generates any more or fewer false positive sites (sites that do not need to be cleaned up) than any other site discovery method. This void exists because followup on HAP has seldom been done, or, if done, records have not been kept, so that determinations of false positive rates cannot be made. Similarly, no records have been kept on the false positive rates of other methods.

Although information is available on the cost of various HAP projects, there is little cost data available on conducting site discovery through other methods, Cost comparisons should be made on the basis of the costs of discovering true hazardous waste sites. There is, however, no information available on the costs of discovering true hazardous waste sites through HAP or any other means.

Another aspect of site discovery is the time required to complete a project. A State of California project using traditional methods took 33 months to search for sites in two agricultural counties; 9 sites were located. The time and cost of the U.S. Army searches of military installations depend on the size of the installation; some have been ongoing since the early 1980s. In comparison, the duration of the 10 HAP Re-Look projects reviewed by OTA were all less than a year,

Pros and Cons of HAP—No one questions HAP's viability as a site discovery technique. Its lack of use is a consequence of site discovery being

[23] Thomas L. Erb et al., "Analysis of Landfills With Historic Airphotos, " *Photogrammetric Engineering and Remote Sensing, vol. 47, No. 9,* September 1981, p. 1364.

[24] Frank R. Wolle, formerly with EPA's Environmental Photographic Interpretation Center, personal conversation, January 1988.

Box 2-B-Historical Aerial Photography (HAP)-Project Summaries

EPA Region 3:84 sites Added to CERCLIS; Other projects Await Screening

Two of the four inventories planned by the 200 Cities pilot project were done. The HAP inventory of Buffalo, New York has not had a followup to verify the results. The inventory of Charleston, West Virginia was released to EPA Region 3 in 1983 but was not used until the Bhopal, India, chemical release incident in 1984 heightened concern about a similar plant near Charleston. Field teams then checked the 74 areas identified in the report, many of which had multiple sites. As a result of the followup, 84 sites were added to the 56 Buffalo sites already in the CERCLIS inventory. Acceding to Region 3, about 25 percent of the HAP sites required an Site Inspection (SI) and four sites may receive a listing SI. In other words, 5 percent of the 84 sites may end up on the NPL.

The value of the inventory convinced Region 3 officials to fund two more inventories (southwestern Pennsylvania and the Elizabeth River watershed). Because of resource constraints, the new inventories were not finished by EPIC until 1987. As of April 1989, Region 3 had not done the followup on the southwestern Pennsylvania inventory because it has a lower priority than getting normal preremedial work accomplished.[1]

The Elizabeth River inventory covering 1937 through 1985 turned up 650 potential problem areas. Over half were eliminated as potential Superfund sites; the scope of the inventory had been purposefully broad, including such areas as grain storage systems, gas and petroleum storage facilities, and other pollution sources that do not qualify for the program. Almost 300 potential Superfund sites have, as of April 1989, not been added to CERCLIS because the necessary field investigations have been delayed due to funding priorities.

U.S. Army Re-Look Project: HAP Equals and Betters Other Site Discovery Methods

The U.S. Army's site discovery program has evaluated hundreds of Army installations using HAP as well as methods such as exhaustive record searches and interviews with current and former employees. The Army considers its search and interview methods possible of yielding a 90 percent discovery rate because of its penchant for recordkeeping and a stable workforce.

In a program called Re-Look, the Army asked EPIC to verify the site discovery program's accuracy using HAP. In a random sample of 10 (one-third) of the Re-Look projects, HAP added 25 new sites to the Army's inventory. HAP was particularly useful in finding errors in the previous searches at large installations; at some small installations HAP found all the known sites (i.e., had an equal discovery rate). At three large installations, 23 previously unknown sites were discovered. Eleven sites were later confirmed as hazardous waste sites; others were placed in a lower priority for additional screening.

Monroe County: HAP Project Added 33 Sites to Existing Inventory

In Monroe County, New York, record searches, interviews of residents, and an advertising campaign requesting citizen reports had produced a list of 10 landfills for a RCRA inventory. In a 1981 project to test HAP as a site discovery method, nine of the 10 reported landfills were identified and 33 additional sites were found. The one landfill not found by HAP was an incorrect entry in the inventory.

Of the 42 sites, the HAP process classified 12 sites as dumps or landfills, 19 as possible dumps or landfills, and 11 as unspecified sites. After followup interviews and field inspections, 22 sites were confirmed as dumps or landfills, 11 were classified as possible dumps or landfills, and six remained unspecified. Three sites were eliminated from the inventory because they contained clean fill.

Region 4: Eight Percent of HAP Sites with Sampling Hazardous

A HAP project, initiated for emergency response planning, discovered 350 potential sites in and around Memphis, Tennessee, in 1980. Field investigation of the sites was limited to visual observations samples at 44 sites were subsequently taken, and 29 sites were found to pose some degree of hazard.

Despite some prescreening, all 350 sites were entered into CERCLIS, at least doubling the number of Tennessee sites in the inventory at that time. Since all PAs have been done for sites entered into CERCLIS prior to October 1986, PAs have been done on all the HAP sites. This set of data of sites with PAs could be used to compare the false negative/positive rate of the HAP process v. other discovery methods. Unfortunately, once in

CERCLIS the HAP sites cannot be distinguished from all other entries. And Region 4 has misplaced the EPIC map sheets with overlays so that it is not possible to separately **identify the HAP sites.**

Love Canal, New York: HAP Confirms Known Sites and *Adds 55* More Sites

As a part of the **EPA** investigation at **Love** Canal in 1980, HAP was used to verify 107 sites identified by a task force. EPIC located and confirmed 46 of the 107 sites as potential hazardous waste sites. In addition, *55* sites were discovered that the task **force has missed. Of the61 task force sites that HAP did not confirm, some were were inside buildings and thus invisible. Others HAP confirmed as negative findings by the task force.** For instance, **for one very large disposal area in the task force inventory, HAP could not find any sign from 1938 to the present that any dumping had occured.**

Regions 2 and 4: HAP Inventories Not Used

EPA Regions 2 and 4 have engaged EPIC to do nine inventories since 1981. However, no followup work has been done in the regions, so it is not possible to ascertain the accuracy of the HAP analysis. The inventories have **not been used as sources of potential sites.**

Three inventories done for Region 2 found 1,341 potential sites. **A removal action was** later taken on one site when buried drums-identified by HAP but ignored—were uncovered during subsequent housing construction. Instead of using the inventories as intended, the region considers them as a source of supplemental information on sites discovered by other means. (The New York State cleanup program, which has a site discovery project, was not aware of the availability of the inventories until OTA happened to **contact** a *State* official asking questions about its status.)

EPIC HAP analysis for Region *4* located 2,076 potential hazardous waste sites, of which 873 are sites containing liquids. OTA could not locate anyone in Region *4 with* any knowledge **of how the** inventories have been used or whether any followup has been conducted

[1]**Regional performance is assesed** on **whether or** not **targets** (i.e., numbers of **PAs** and **SIs) are met. Doing the followup on an** EPIC **inventory,** despite its **potential environmental significance, would not count. Targets are set based on** numbers of situ **already** in the **CERCLIS inventory.**

a low priority in the Superfund program and being of little concern to Congress or public interest groups.

Most of those who have used HAP to assist in site discovery work see its value; no one disputes its ability to comprehensively and efficiently survey an area and uncover valuable environmental information. The major argument against its use is its drain on resources; finding hundreds of *potential* sites in a given area means that hundreds of sites have to be checked out. But the same argument is used against active site discovery, in general. **Hundreds of potential sites will need hundreds of assessments today but assessments today can save millions of dollars tomorrow if finding truly hazardous sites sooner rather than later means that the cost of cleaning them up is minimized.** Identifying hazardous sites faster also means that protection of human health and the environment is enhanced.

The following statement is representative of several that OTA received from EPA and other officials who have used HAP. The same kind of statement could be made about any site discovery method:

> The major drawback is that historical analysis [is] only a screen tool. A completed, historical aerial photography study does not define potential CERCLA hazardous waste sites exclusively. As a result, a significant amount of follow-up work is necessary to define the true number of potential CERCLA hazardous waste sites. This additional work includes screening out obvious non-CERCLA sites, file search, mapping, cross-referencing other environmental data bases, and offsite reconnaissances. These subsequent activities are both time and resource intensive.[25]

Like all other discovery methods, the HAP method locates potential sites that may or may not be true hazardous waste sites. HAP and other methods

[25]Stephen Wassersug, director, Hazardous Waste Management Division, EPA Region 3, letter to OTA, February 1988.

will discover sites regardless of their appropriate cleanup authority. Because of the comprehensive sweep of an area that is possible using HAP and the resultant large numbers of discovered potential sites, its use may require that more resources be devoted in the field to verify sites than would be necessary with other methods. One analyst has estimated that competing the **200 Cities** project could add another 30,000 potential sites to CERCLIS, doubling the inventory. It took EPA, under current policies, 7 years to complete the PAs for 24,185 sites, a task that was accomplished only because Congress mandated **EPA to** do SO by January 1988.

It is important to note that thousands of HAP discovered sites would not be added to the CERCLIS inventory overnight. They would be added over 5 to 10 years. **Under the current passive site discovery program, the inventory grows at a rate of 2,000 sites per year. An active site discovery program, using HAP and other methods, could at least double the rate of growth.** Thus, EPA could require twice as many resources as are currently devoted to preremedial work (about $50 million per year) to keep up with the discovery rate. However, some of this need for additional resources could be offset by adopting a more efficient site evaluation process (see later discussion and app. 2A).

Other drawbacks of HAP mentioned to OTA include:

- "6 . . . it is only cost effective in highly industrial areas or areas where many sites are clustered together. '[26]
- "EPA has no dedicated resources to verify critical information such as street addresses for many of these sites. ' '[27]

The first problem is resolved by carefully selecting the areas in which to use HAP, such as was done under the 200 Cities proposal. The second problem of identification is one aspect of ground proofing unique to HAP. HAP discovered sites are identified by map coordinates; sites are listed in CERCLIS by street address (and assigned code number). Today off-the-shelf commercial computer programs are available that will match coordinates (latitude/longitude) with street addresses.[28]

Beyond its site discovery capabilities, HAP can provide additional valuable information, not available from other site discovery methods, that is pertinent in later stages of site cleanup analysis. Some of this information is available directly from the site discovery work, other information can be provided with additional photo analysis. For instance, U.S. Army documents say that, even in cases where HAP projects did not locate additional sites, "the study was very useful in confirming the existence and a real extent of various potential sites identified in the initial assessment report."[29]

One expert has classified the information about a landfill derivable from HAP as: 1) existence (i.e., the location, extent, and possible nature of a landfill), 2) general or detailed temporal land use and land cover information, and 3) physical environmental aspects (i.e., the geology, soils, and surface and subsurface drainage). The expert says,

> In general, [HAP] can provide the most efficient, complete source of information regarding the physical environment, particularly, in the absence of soil survey or surficial geology reports.[30]

The original EPIC inventory done for EPA Region 3 (as part of the aborted 200 Cities project) was later used for other projects in the Charleston, West Virginia, area. For the Kanawha Valley Integrated Environmental Management Project, the study provided spatial relationships of potential and actual hazardous waste sites and of production facilities to populated areas. In another study on concentrations of dioxin in fish, the inventory provided valuable information for developing a sampling plan of river sediments by pinpointing areas of high potential sources of contamination.

[26] Barbara Metzger, EPA Region 2, letter to OTA, January 1988.

[27] Narindar M. Kumar, EPA Region 4, letter to OTA, May 1988.

'Mapping Information Systems Corp. (MAPINFO) offers one that covers 330 metropolitan areas.

[29] U.S. Army Toxic and Hazardous Materials Agency, "Update of the Initial Installation Assessment of Green River Launch Complex, UT, " Nov. 12, 1987.

[30] Erb et al., op. cit., footnote 23.

Inventories, Lists, and Estimates

Once sites are discovered, the information needs to be put somewhere. The Superfund CERCLIS inventory is perceived as *the* national list, but it is not. The CERCLIS inventory is, more accurately, the Superfund *remedial* list. There is also the Emergency Response Notification System (ERNS), the Federal Facilities Hazardous Waste Compliance Docket, and various inventories associated with other Federal cleanup programs, such as RCRA corrective action. In addition, most States have some type of inventory; many have multiple lists. All of these inventories or lists have been created independently of one another and few are compatible with CERCLIS.

Sometimes a site is included in more than one inventory. For example, there are Leaking Underground Storage Tank (LUST), RCRA, and Federal facility sites in CERCLIS. But, not *all* RCRA, LUST, and Federal facility sites are in CERCLIS. While all known potential Federal sites are included in the Federal docket, EPA does not have, as yet, an operational RCRA correction action database and has no plans to create a national LUST inventory.

National Lists

CERCLIS was originally created by combining three separate databases (13,392 sites) in 1982 and a group of sites reported by States in 1983 that brought the total to over 15,000 sites.[31] This list now grows at a steady rate of about 2,000 sites per year (see figure 1-1 in ch. 1) despite the lack of an active national discovery program or any consistency on how and when reporting occurs.

To be evaluated by the Federal Superfund program, a site is supposed to be placed in CERCLIS.[32] Actual site entry into CERCLIS is the responsibility of EPA regional offices. But, how the information flows to that point is dependent on who discovers the information and how it is reported. **There is no national criteria or guidance on the timeliness of entering a site into CERCLIS or if any prescreening should take place. The** Booz-Allen/EPA report on site discovery identified a number of States (e.g., California, Florida, Wisconsin, Ohio, West Virginia, and New Jersey) that prescreen sites prior to turning them over to EPA for placement in CERCLIS.

GAO investigations of just five States in 1985 showed that 837 sites on State lists were not in CERCLIS. By 1988, a followup report confirmed that 494 of those 837 were still not in CERCLIS.[33] States do not report all potential sites for various reasons, Some States claim that not reporting sites to CERCLIS gives them an edge in negotiating to get potentially responsible parties to clean up sites. Sometimes States do not report sites they feel will never qualify for the NPL, Florida, according to the Booz-Allen/EPA report, ''only adds sites to CERCLIS if the State believes that the site will require [a Site Inspection]."[34] California, according to the same report, only submits those sites ''it wants to address through a CERCLA cooperative agreement.'[35]

Lack of listing guidance has backfired in a way probably not intended by Congress when it mandated (through SARA in 1986) schedules for the preremedial process. EPA subsequently added a policy that all sites must have a PA within a year of entry into CERCLIS. The nature of CERCLIS as a list of all reported potential sites is changing. Sites are now sometimes held outside of CERCLIS until regions have some confidence that resources are available for a PA within a year. An example is the EPA decision to move some 3,000 potential RCRA corrective action sites through the Superfund PA/SI process. Since this move requires the entry of thousands of sites into CERCLIS, the sites are being

---

31 CERCLIS was called ERRIS until the mid- 1980s. In the last few years, CERCLIS has become much more than simply an inventory of potential Superfund sites. It is now THE database of the Superfund program and contains the data for tracking NPL and non-NPL sites through the program.

32 Conventional wisdom is that a site *has* to & in CERCLIS to move into the Superfund program. This is not necessarily true. In 1984 EPA identified 19 California sites that had been submitted directly for HRS evaluation without having been placed in CERCLIS. [U.S. Environmental Protection Agency, "Extent of the Hazardous Release Problem and Future Funding Needs, CERCLA Section 301(a)(l)(C) Study, " December 1984, p 4-.?,) Also, in an analysis of the time it took to move a set of sites from discovery to proposal for the NPL in June 1988, OTA could not find about a dozen of the NPL sites in CERCLIS.

33 U. S. General Accounting Office, *EPA's Inventory of Potential Hazardous Waste Sites Is Incomplete*, op. cit., footnote 3; and *Superfund Extent of Nation's Potential Hazardous Waste Problem Still Unknown,* op. cit., footnote 12.

34 U.S. Environmental Protection Agency, ''CERCLA Site Discovery Program Evaluation, ' op. cit., footnote 15, p. 5.

35 Ibid., p. 10.

phased in to relieve the pressure caused by the one-year policy. Thus, RCRA sites are being prescreened prior to entry; high priority sites were scheduled to be entered in fiscal year 1989 and medium and low priority sites in fiscal years 1990 and 1991.[36] This cautious approach to CERCLIS entry appears to duplicate workload; one outcome of the PA done once a site is in CERCLIS is to give sites priority labels. In another example, an inventory of potential sites was discovered through historical aerial photography and given to EPA Region 3 in 1987. As of April 1989, the data was still being held outside of CERCLIS until it can be prescreened prior to entry. Again, the bureaucratic reason is to avoid creating a workload for which there are no resources within the required one-year timeline.

Holding sites outside of CERCLIS has several effects. From a management perspective, it helps meet SARA targets and conserves existing resources. If, however, the preremedial program has more work than it can handle, which is justified environmentally, another approach is for EPA to ask Congress for an increase in funding. Otherwise, the practice of circumventing congressional intent of speedy PAs serves to avoid rather than solve potential problems. The practice devalues CERCLIS as a timely inventory and source of knowledge of the national extent of the problem. **Delays in entry will artificially cause the growth of CERCLIS to decline, falsely implying that site discovery has peaked.**

Other Lists or Estimates

CERCLIS is the largest single Federal inventory, but it is not the only source of potential Superfund sites. Other Federal and State inventories, lists, or estimates include:

- ***Federal Facilities Hazardous Waste Compliance Docket. The*** docket lists Federal facilities or sites that may require cleanup. Reporting is required under CERCLA every 6 months. Reporting is on a facilities basis (except for sites belonging to the Department of the Interior (DOI)); each listed facility may have one, a few, or hundreds of sites. The Department of Defense's (DOD's) Rocky Mountain Arsenal, for instance, has 165 sites, As of November 1988, the last update, the docket contains 1,170 facilities; 115 sites are on the NPL.
- ***GAO Review of Civilian Agencies.*** This September 1986 report reviewed agency data and identified 1,882 potential sites (excluding some 7,000 sites believed to be DOD's responsibility). The study implies that over half of the civilian agency sites (about 1,000) will need some cleanup.[37]
- ***Individual Agency Estimates.*** Federal agencies are required to report to Congress annually (under CERCLA) on the status of their cleanup programs. Data from those reports is more recent than the GAO report data. In fiscal year 1988 reports DOD listed 8,139 sites; DOE, 1,700 sites; and DOI 254.
- ***RCRA Corrective Action.*** A formal national inventory does not yet exist. There may be anywhere from 2,000 to 5,000 RCRA sites requiring some kind of cleanup; differing estimates depend on the counting scheme. A 1989 EPA document accounts for 5,081 known RCRA facilities. Using a 1987 GAO estimate, at least one unit in 52 percent of the facilities, or 2,626 facilities, will need remediation. In November 1988, EPA estimated that 29 percent of 80,000 units in RCRA facilities, or 23,066 units, will need to be cleaned up. A unit (called a Solid Waste Management Unit in regulatory jargon) can range from a small tank to a large landfill. Another EPA document says that close to 5,000 *closing* RCRA facilities are potential sites. None of the above accounts for the thousands of municipal landfills that may require cleanup and for which there is no national listing.
- ***Mining Sites. The DOI'S*** Abandoned Mine Lands Remediation Program has not inventoried noncoal mines needing cleanup. GAO estimated in 1987 that there could be 22,339 hazardous waste sites at mines and processing facilities, over 90 percent of them located at

[36] U.S. Environmental Protection Agency, Office of Solid Waste and Emergency Response, ''Annual Report, Fiscal Year 1988. ' EPA/68-01-7X9, November 1989, p. 31. The sites scheduled for entry and PAs in fiscal year 1989 had not been entered by July 1989.

[37] U.S. General Accounting Office, *Superfund Civilian* Federal Agencies *Slow* to *Clean up Hazardous* Wrote, *GA O/RCED-87-153* (Gaithersburg, MD: U.S. General Accounting Office, July 1987).

closed or abandoned mines.[38] Of the 24 sites in the Uranium Mill Tailings remediation program, 22 have yet to be cleaned up.

- *Leaking* ***Underground Storage Tanks. The*** estimated number of potential leaking underground storage tanks ranges from 300,000 to 400,000. Discovery and development of inventories has been left to State programs. (When writing Underground Storage Tank regulations, EPA rejected a suggestion that active discovery be required.)
- ***Asbestos Abatement.*** EPA has estimated that over 44,000 public schools contain asbestos that may need attention. Between 300,000 and 700,000 public and commercial buildings have asbestos that may have to be removed. This hazardous material gets deposited primarily in municipal landfills (see ch. 4).
- State *Lists.* Many States have inventories; usually a State list (or lists) contains more sites than the number of that State's sites in CERCLIS. The results of an Association of State and Territorial Solid Waste Management Officials (ASTSWMO) survey taken in July 1986 reported a total of 31,910 confirmed and suspected sites in 45 States and Puerto Rico.[39] This number compares with 24,544 sites listed in CERCLIS, as of October 1986, for those same States and Puerto Rico. If the same State/CERCLIS ratio holds today, States have over 40,000 sites inventoried. State inventories vary widely in content (types of sites listed), in knowledge level on sites (confirmed or only suspected problems), and whether or not they include sites that are also in CERCLIS.

It does not have to be this way. New Jersey has recognized that because of multiple and incompatible lists, "the actual number of sites requiring remediation [by the State] is unknown."[40] As part of a new strategy to coordinate various State cleanup programs, a computer database-a Comprehensive Site List—has been established. Data input will be done by and be accessible to individual programs. The central database replaces individual program inventories that caused overlap or duplicate counting and collection of data.

SITE EVALUATION

Site evaluation in the Superfund program starts with a preliminary assessment (PA) and doesn't really end until a remedial investigation/feasibility study (RIFS) has been completed. (But, to some extent, site evaluation continues through remedial design and implementation, especially for complex sites,) This section reviews the conduct, status, and outcomes of *preremedial* Superfund site evaluations; the PA and site inspection (SI) stages and the use of the Hazard Ranking System (HRS) to score sites for the NPL on/off decision. Along the way from a PA to scoring, the majority of sites in CERCLIS get rejected; only 10 percent make the NPL. Those few sites, however, are not the only ones that pose threats to public health and the environment. They are also not necessarily the worst sites.[41] OTA has estimated that from 240 to 2,000 sites may have missed being placed on the NFL because of false decisions made during preremedial screening.

Environmental Fulcrum Shift

The goal of preremedial site evaluation has always been—and still is—to ultimately decide which sites belong in the Superfund program. Over time, early decisionmaking has shifted from an environmental bias (''Does the site need cleanup'?' to a management bias ('Will the site qualify for the NPL?"). A site no longer moves beyond the *first* screening step unless a case can be made that it may warrant Federal attention (i.e., has a probable HRS score of at least 28.50). A PA no longer simply determines whether or not a site is a threat or not.

The shift, part of the narrowing of Superfund discussed in chapter 4, has occurred quietly without any public discussion. There **has been little public notice of the shift because the public doesn't pay much attention to preremedial activities and because public statements by EPA imply that**

[38] Ibid., p. 16.

[39] Association of State and Territorial Solid Waste Management Officials, ' 'State Programs for Hazardous Waste Site Assessments and Remedial Actions, " June 1987, p. 1.

40 New Jersey Department of Environmental Protection, "Case Management Strategy Manual, " draft, May 1989.

[41] The term *worst sites* is rarely defined. It can mean a complicated site, a site that is expensive to clean up, or one that poses high risks to the surrounding community. It could also mean a site that poses current risks as opposed to a site that only poses potential future risks.

nothing has changed. The shift may have been occurring before SARA was passed in 1986 but was certainly enhanced by congressionally mandated activity levels for site evaluation.

The management bias means that the Superfund program has fewer sites to deal with not only during the preremedial process but in all following stages of Superfund. The shift saves money for the Superfund program and makes it easier for EPA to meet targets mandated by SARA. In 1988, EPA stated:

> A key management initiative in fiscal year 1987 was a strategy designed to expedite the pre-remedial process by focusing attention **on early decisions** to ensure that fewer low priority sites reach the resource-intensive stages of the pre-remedial process [emphasis added] .42

The environmental consequence is that, as the kinds and numbers of sites rejected early by EPA grows, increasing numbers of those sites probably truly need attention. The earlier they are rejected, the less anyone will know about them. If ignored, the future costs of cleaning up those that do need cleanup will probably be greater than they would be today and, in the interim, protection of public health and the environment has been reduced.

The Superfund site evaluation process is what the health care field calls screening. Screening generates two kinds of correct (true) and incorrect (false) outcomes. True positive and true negative decisions are the desired information. But, there is always a probability of making false positive and false negative decisions. For Superfund and in a strict environmental sense, false *positive* sites are those that do not pose a threat but are judged to be threats. False *negative* sites are sites that are a threat but are judged not to be.[43] EPA worries more about whether a site is judged a problem, when it is not, than whether a site is judged not a problem, when it is. False positives mean that money is spent unnecessarily in site analysis. False negatives may cause harm to human health and the environment until sites eventually resurface for attention, and then cleanup costs will be higher. False negatives also can downplay the extent of the cleanup problem to Congress and the public by underestimating the number of sites requiring cleanup.

Screening in Superfund is done in a series: PA, SI, HRS scoring, RIFS. Series processes tend to generate more false negatives than false positives (see app. 2A). Despite this inherent bias in the Superfund process, EPA assumes that false negative outcomes are minimal. EPA **has never assessed the 8-year universe of rejects from its process to determine how well it performs, environmentally.** On the other hand, EPA has spent time and effort to assure the lowest possible numbers of false positive outcomes.

A Better Environmental Priorities Initiative

EPA's Environmental Priorities Initiative (EPI) is a minor adjustment in the Superfund site evaluation process, It is a good example of a move by EPA to integrate two cleanup programs (see ch. 4). EPI does not appear to have the environmental significance that its name implies or EPA claims, however. But, a broader initiative with true environmental focus could have.

As it now stands, EPI partially integrates the RCRA and Superfund programs by evaluating under the existing Superfund preremedial system those sites covered by either program. At a point during the SI stage, a management decision is made about whether Superfund or RCRA has responsibility for a site.

In combination with a national site discovery program, this kind of integration could encompass all cleanup programs. To regain art environmental focus, the PA would need to be returned to its original threat/no threat role and sites would be kept in the system until a case could be made that a threat *does not* exist. At the decision point, sequential decisions would be made: Is it a threat or not? If not, a site would be tagged no further action (NFA). If yes, which program has the authority to act? Then, a formal notification would send the site to the proper authority *and the* appropriate indication would be entered into the CERCLIS database so that tracking would be feasible.

42U.S. Environmental Protection Agency, Office of Solid Waste and Emergency Response, ''Superfund Advisory,'' Winter 1988, p. 1.

43If the intent is to find NPL sites, then false negatives are sites that qualify for the NPL but have been rejected.

Just where the environmental/management decision point should be assigned is a key element of a better EPI. What may be required is to throw out the PA/SI concept, to view site evaluation as a continuum rather than two distinct steps. Instead of a system that says that at a specific point sufficient information is always available to make a binary decision, a better system could be built around the principle that *when enough information is available* a decision will be made. Unlike the current Superfund evaluation process, this system would require the development and use of experts as decision-makers.

State programs, if they had sufficient resources to do so, might be inclined to make better judgments about whether a site is a threat or not. States, unlike EPA, have no one to defer sites to. If a site is a problem and does not qualify for the Superfund program, the State itself will eventually clean it up.

The Changing PA

It is primarily at the PA stage and to some degree at the SI stage where the environment-to-management shift has occurred in the Superfund program. The change has crept in with changes in the definition of a PA and its outcome and with the use of the HRS to prescore a site. That a change has occurred is obvious by the change in the language used to tag a site that is rejected during the preremedial process (see box 2-C). No longer does a PA simply say whether or not a problem potentially exists, leaving it up to the more extensive information of the SI stage to make a judgment about NPL qualification. The PA now concludes whether or not a site may qualify for the NPL. **With little change in the information available, EPA has cast the PA in a new role for which it is inappropriate.**

According to the 1982 National Contingency Plan (NCP), the PA originally was a method for the *removal program* to assess whether: 1) no threat was present at a site, 2) a threat required immediate attention, or 2) a potential threat should be turned over to the remedial program. This concept is similar to that in the medical field where a process called *triage* separates patients into three categories: no attention, immediate attention, and later attention.

Box 2-C—From NFA to NFRAP

Concurrent with the changes in the PA have been changes in the nomenclature for sites rejected by the preremedial process. The original no further action **(NFA) was changed** around *1986* to no further remedial action planned (NFRAP).

An NFA had an environmental meaning: no threat was present at the site. The current NFRAP is more a policy statement and can mean, according to **a 1988 EPA documental**

1. **sites** that never received CERCLA hazardous **substances;**
2. sites where the CERCLA hazardous substances are clearly not releasing, and have no potential to release, into the environment and *no removal action is* required;
3. sites where EPA is not legally authorized to respond to the release; and
4. sites with no reasonable potential to score 28.5 or higher upon application of the current HRS at the end of an SI.

EPA further explained in another 1988 document: "Note that the NFRAP designation does not mean that there are no environmental hazards at the site. There may be hazards but these hazards may not be of sufficient magnitude for NPL listing purposes."[2] A 1989 EPA document is more succinct; it defines NFRAP as "those sites with no reasonable potential to score above the HRS cutoff. ''3

[1] U.S. Evironmental protection Agency, Office of Solid Waste and Emergency Response, ''Preliminary Assessment Guidance, Fiscal Year 1988," directive 9345.0-01, January 1988.

[2] U.S. Environmental protection Agency, Office of Solid Waste and Emergency Response, ''Pre-Remedial Strategy for Implemen ting SARA," directive 9345.2-01, Feb. 12, 1988, p. 5.

[3] U.S. Environmental Protection Agency, Office of Solid Waste and Emergency Response, ''Regional Pre-remedial Program Objectives for FY 89 and First Quarter of FY 90," directive 9345.242, Mar. 10, 1989, p. 3.

By 1985, the PA had become the province of the remedial program, but the triage concept remained. That year **the** NCP listed three purposes of a PA: 1) to eliminate nonthreatening sites from further consideration [no attention], 2) to determine any potential need for removal action [immediate attention], and 3) to establish priorities among sites requiring Site Inspections (SIs) [later attention].[44]

4450 F~r~ Re~ster *47972,* Nov. 20, 1985

As late as April 1987, the PA still retained the triage concept. EPA stated in a training manual then in use:

> A PA is not intended to give a full or complete picture of a site and its associated problems. A PA is by design, a relatively quick, low-cost review of relevant available data to determine whether the site potentially poses a problem and, if so, what type of follow-up work should be undertaken to further assess the site.[45]

And, in public documents, EPA said that an SI proceeded ''if a preliminary assessment turns up evidence that a site may pose a threat . . .' '[46]

Thus, the official use of the PA has been to determine whether a site needs emergency attention, whether it should move on to the second screening step (an SI), or whether it is dropped from further consideration because the site does not pose a problem. In a 1984 report to Congress, EPA defined sites rejected by the PA—no further action sites—as sites that ''pose no threat to public health or the environment, and thus warrant no further investigation or remediation. '[47] Examples of such sites are sites reported to the CERCLIS that do not actually exist, have already been identified under a different name, and demonstrably contain no hazardous substances. Dropping such narrowly defined types of sites early is environmentally appropriate and cost-effective.

Under December 1988 proposed rules for the NCP, the PA appears to be unchanged, except that, in the preamble and a section called ''Point of Clarification, ' EPA is explicit about the entire preremedial process being one to determine whether or not sites ''warrant remedial action, " The phrase ''warrant remedial action' clearly means whether or not a site qualifies for the NPL, whether it will attain an HRS score of 28.50 or greater and will not be deferred to another authority (e.g., is not a RCRA site). No longer is a *potential* threat sufficient cause to keep a site in the evaluation process. Only if a site shows evidence of a *significant* threat may it move to the SI stage of site evaluation, according to the proposed NCP.

Today's PA (and SI) appear to be the result of studies such as one done for EPA in 1987 by Ecology & Environment, Inc. The study, "Workload and Resource Requirements for Preliminary Assessments, Site Inspections, and Hazard Ranking System Evaluations Under SARA," was funded by EPA after SARA mandated EPA to meet specific target dates and objectives.[48] This document, like other similar EPA documents, does not discuss or mention the possible environmental effects of changing the preremedial process.

At the time that such internal EPA documents were changing the concept of the PA, public statements were projecting the original environmental image of a PA (and SI). EPA told Congress in 1988 that the information collected for a PA:

> . . . is then evaluated to determine whether the site has handled hazardous substances and if those substances have the **potential** to affect human health or the environment . . . If a PA indicates that there may be a release of hazardous substances that may threaten human health or the environment, EPA then recommends a site inspection (SI) to better understand the problem. On completion of the SI, if the site still poses a **potential** threat, it is scored . , . [emphasis added] .49

Compare the above "potential to affect human health or the environment' determination that keeps a site in the screening system with a March 1989 directive to EPA regional offices that says, at the end of a PA (and a screening SI), sites ' 'with no reasonable potential to score above the HRS cutoff will be rejected from the system.[50]

---

[45] U.S. Environmental Protection Agency, Hazardous Site Evaluation Division, ''sup-f~d PA/SI Training Course,' ' Section 1.1, undated but used *in* training sessions on Apr. 22-24, 1987.

[46] '' Steps in Cleaning Up a Superfund Site, " *EPA Journal*, January/February 1987, p. 17.

[47] U.S. Environmental Protection Agency, Office of Emergency and Remedial Response, ''The Effectiveness of the Superfund Program—CERCLA Section 301(a)(l)(A) Study, " December 1984, p. 1-9.

[48] See SARA Sections 105 and 116.

49 U.S. Environmental protection Agency, '' Superfund Advisory, ' op. cit., footnote 42, pp. 5-6.

[50] U.S. Environmental Protection Agency, Office of Solid Waste and Emergency Response, ''Regional Pm-remedial program Objectives for Fiscal Year 1989 and First Quarter of FY90,'' directive 9345.2-02, Mar. 10, 1989, p. 3,

HRS Prescoring

The Hazard Ranking System (HRS) was designed to score a site with data collected from the SI stage of the preremedial process; the score is used to make the on/off NPL site decision. Even this documented, quality assured set of information does not (and was never intended to) fully characterize a site and the risks it poses. That occurs later during a Remedial Investigation (RI).

Now, EPA has moved the HRS up to the PA stage. The HRS is being used to *prescore* (or, estimate the HRS score for) a site when most of the available information comes from existing records. There is no site sampling done for a PA. EPA has defined two types of prescores. A *preliminary* HRS score is a minimum value; missing data is assigned a zero. A *projected* score is a possible score with missing data estimated by the evaluator. The PA guidance document says that the prescore will "be used to assign a priority to the site for an SI or to eliminate the site from CERCLA remedial activity."[51] **Clearly, the PA has changed from being a threat/no threat decision to being a method to eliminate sites as early as possible from the Superfund program.**

To make the ruling of *significant threat* to move a site from a PA to an SI, agencies conducting the PA have been told, through the proposed NCP, that they ''may use a combination of a preliminary HRS score and best professional judgment. The latter tool—professional judgment—is, however, only to be a supplement ''to the preliminary score in making decisions about whether or not to proceed to the next phase of evaluation."[5z] EPA's *Preliminary Assessment Guidance Fiscal Year 1988* requires the use of preliminary and projected HRS scores as a basis for site decisions. **Thus, the NCP proposal is not only codifying the use of prescoring but appears to be making it more difficult than does the guidance to move a site forward.**

Adding a quantitative measure and, perhaps, professional judgment to the PA evaluation appears to be a step toward objectivity. And, EPA is designing a computer program to standardize and reduce the workload. However, the numerical outcome of the HRS--even when used after an SI—is based on subjective decisions, and because the information available at the PA stage is the poorest, the HRS prescore is most uncertain. Since the PA relies on existing information, evidence of contamination may not be uncovered without sampling, which does not occur until an SIq.[53] Confirming this, a memorandum from two EPA officials said: ''Often it is difficult at the PA stage to recommend no further action without field visits and sampling.'[54] PA data collections and evaluations are done by entry-level employees, adding uncertainty both in the application of the HRS and the use of professional judgment.

Another point is that getting HRS scores high enough to qualify for the NPL is sometimes an art. It is not an uncommon practice to recalculate HRS scores repeatedly until a sufficiently high score is obtained so that a decision can be made to proceed with formal scoring.[55] One regional official told OTA that it is "always possible" to get the crew to go back to the site and get more information to raise the score. This is an overstatement because scores from repeated attempts would eventually approach a ceiling. However, the practice points out that the

51 U.S. Environmental Protection Agency, Office of Emergency and Remedial Response, *Preliminary Assessment Guidance Fiscal Year 1988*, Directive *9345.0-01*, January 1988, p. 14.

5253 Federal *Register* 51394, Dec. 21, 1988, p. 51413.

53 Sampling is not always done for an SI either. An EPA Inspector General report cited several instances where sampling did not occur In one audit of State work that had not received adequate Region oversight, the [G sad, ''The most common deficiency was the failure m perform the necessary sampling during the S1 process. [U.S. Environmental Protection Agency, "Capping Report on EPA, Office of the Inspector General Audits of Superfund Cooperative Agreements for Fiscal Years 1985 through 1987," Mar. 29, 1988.] Also an EPA contractor study found that sampling was not done and no pre-existing analytical data was available for 22 percent of 212 sites that had received SIs. [U.S. Environmental Protection Agency, "preliminary Assessment and Site Inspection Program Quality Assurance Review," draft, Sept. 11, 1987.]

54 U.S. Environmental Protection Agency, ''Guidance on Preliminary Assessments and Site Inspections Under CERCLA,'. draft. no date, sent to the all EPA Regions on Sept. 8, 1987, by Gene Lucero (Office of Waste programs Enforcement) and Henry Longest (Office of Emergency and Remedial Response), p. 10. In the past, site visits were not an official part of a PA. Now, they *might* occur. Under the proposed NCP: "A PA shall . *include* an off-site reconnaissance as appropriate. A PA may include an on-site reconnaissance where appropriate. [53 Federal Register 51502, Dec. 21, 1988.]

55 An EPA Inspector General report dated March 1988 documents several examples. State program officials also gave OTA numerous examples of multiple sampling to finally obtain the necessary data. For instance, William De Vine, Louisiana Department of Environmental Quality, said that air was sampled numerous times at the Dutchtown site m Louisiana before a release could be documented.

aggressiveness with which information is collected and the comprehensiveness of the information affects the ultimate HRS score. Using an HRS prescore *before* most of the information is collected can prematurely bias the ultimate fate of a site.

This is not to say that prescoring does not have value during a PA. As a tool, it can help focus future data collection and sampling efforts. For example, if prescoring indicates that no information is known about contaminant migration paths, the SI can be designed to look specifically for them. EPA's contractor in Region 5 has been using prescoring expressly for this purpose since mid-1984. Prescoring can also be used as an indicator of environmental threat to help decide which sites with PAs should get SIs first,

What prescoring cannot do is make a definitive determination of the possible environmental threat of a site nor does it necessarily forecast a site's ability to make the NPL. A preliminary score, as a minimum score, can tell whether a site appears to be a significant threat, if the information is valid. **In an environmentally biased system, the default option would be to always move sites forward unless a case can be made that a site is not a threat.** That decision would require an HRS prescore that is *maximized* and still shows no threat.[56]

Focusing so narrowly and early in the process on only the data needed for HRS scoring also detracts from a positive step EPA may be trying to take and that OTA has suggested as a policy option. That is, to link the site evaluation process with the RI to avoid the current duplication of effort that occurs as sites move from the preremedial to the remedial phase of evaluation.

Two SIs and a Deferral Point

EPA now has two SIs: a screening SI (SSI) and a listing SI (LSI). The goal is to flush out Superfund false positive sites that have managed to get beyond the PA stage (i.e., have received a high or medium priority rating when in fact they will not qualify for a Superfund remedial action). From EPA's management perspective, there is no point in spending resources on collecting more data for these sites. If they need to be cleaned up, Superfund won't be doing it.

Introducing a new screening stage does provide an opportunity to find false positive sites that have gotten through the previous stage and to prevent them from moving to a more expensive stage. It also means that additional false *negative* decisions will be made because all screening stages make both types of false decisions. Under the old one-SI system, all sites that made it through the PA stage got a full SI, lessening the chances of making false negative decisions.

Now, under an SSI all sites get a ''refined" I-IRS prescore. In conducting an SSI, the ''rigorous ILSI] data quality objectives (DQOs)" do not have to be met.[57] Thus, like during the PA, the incompleteness of the data for the prescore may bias the outcome. Not having to adhere to DQOs, however, saves resources. Based primarily on the new prescore, sites with SSIs will either be rejected (get an NFRAP designation), be recommended for an LSI, or get deferred to another authority. EPA says that deferral at this point ''indicates that the site has the potential to score above the cutoff score for NPL listing but the release could be more effectively addressed by another statute or authority. "[58] For now, this is a CERCLA/RCRA decision point. However, it is designed to accommodate more extensive deferrals, if a comprehensive deferral policy is implemented (see ch. 4).

What Are the Outcomes?

Over the last few years, the budgets for site evaluation have increased some. For fiscal year 1990, EPA has requested almost $47 million to pay

[56] In one of a series of papers prepared for the Massachusetts Contingency Plan, the question of proper bias was discussed as: ''. whether the Department should assume, when little information is available, that a substantial hazard exists *or* whether it should designate a site as a priority disposal site only when available information indicates a substantial hazardous exists. [''Site Classification System, *Massachusetts Contingency Plan*, Discussion Papers, December 1987.] Note that not included is the Superfund program concept that when little information exists a site can be classified as not hazardous.

[57] U.S. Environmental Protection Agency, Office of Solid Waste and Emergency Response, ''Pre-Remedial Strategy for Implementing SARA," Directive 9345.2-01, Feb. 12, 1988, p. 8.

[58] U.S. Environmental Protection Agency, Office of Solid Waste and Emergency Response, ''Regional Pre-remedial Program Objectives for FY 89 and First Quarter of FY 90, " Directive 9345,2-02, Mar. 10, 1989, p. 5.

contractors for PAs, SIs, and HRS scoring work; in fiscal year 1987, $34 million was appropriated.[59] Still, the annual cost of site evaluation is comparable to cleaning up one large site, and taking sites from a PA to the NPL consumes only about 3 percent of the annual Superfund budget. And, the budget increases for preremedial work may not be keeping pace with an increasing effort required.[60]

At the regional offices, the internal EPA workforce in terms of numbers of full-time equivalents (FTEs) is very small. It has decreased from a high of 47 FTEs in fiscal year 1988 to 41 FTEs in fiscal years 1989 and 1990.[61] The near leveling of PA/SI funding and decrease in staffing may reflect EPA's goal to change the process to reduce the overall workload, i.e., evaluate fewer sites.

Site evaluation is not done by EPA staff. Staff supervise the work of Field Investigation Team (FIT) contractors and State government agencies, whose employees or contractors do site evaluations. EPA's FIT contracts have been held throughout the history of the program by two firms: Ecology & Environment, Inc., and NUS Corp. The current contracts run for 5 years from November 1986 to October 1991 and are valued at $154 million and $130 million, respectively.

Like the other portions of the Superfund program, the accomplishments of site evaluation are recorded as numbers of PAs, SIs, and NPL listings completed per fiscal year. These numbers are not only used to show the progress of the program to the outside world but also internally as regional performance measures. Table 2-1 shows the numbers from fiscal years 1980 through 1988.

Additionally, since SARA imposed mandatory activity levels in 1986, EPA's performance has been judged on whether or not those schedules are met. The first requirement, that PAs be completed for all sites in CERCLIS as of the date of enactment of SARA, was met on schedule by January 1988. Within that same year, however, EPA had stated that the other deadlines—all necessary SIs and HRS scores by January 1989 and October 1990-respectively, could not be met. Actually, EPA Regions 6 and 10 did complete the SIs on all their pre-SARA sites by the deadline. Those two regions had only 6 percent of the total backlog. Thus, 94 percent of the SIs were not done as required by SARA.

Congress also placed continuing pressure by saying EPA should complete all necessary HRS scores within 4 years for sites entered into CERCLIS after SARA was enacted. EPA told the House Appropriations Committee in early 1987, that it would be able to meet that goal.[62] Historically, EPA told the committee, to go from CERCLIS entry through HRS scoring may take as little as 2 years or as long as 5 years. The conclusion from an OTA analysis differs. OTA reviewed all 229 sites proposed for the NPL on June 24, 1988. The *average* time from discovery date to proposal was 5.5 years. For 54 percent of the sites it took 6 to 10 years. While for 20 percent of the sites it took 3 years to complete the process, an equal number took 8 years.

The new PA/SSI/LSI/HRS scoring screening will be more time-consuming than the old PA/SI/HRS scoring screening was. To shorten the time it takes to move a site through the process, EPA could reduce the time a site sits between stages.[63]

***The Persistent SI* Backlog—Meeting** the initial SI and HRS deadlines has been impossible due to a persistent backlog of sites awaiting SIs and an apparent unwillingness or inability on EPA's part to fund SIs at the level needed to resolve the backlog. Even if full funding were available, a question remains as to whether or not the technical expertise

59The fiscal year 1990 request is $1 million more than the estimated budget for fiscal year 1989.

60Minnesota claims that EPA's preremedial process has changed over the past several years and has become more time-consuming The effect for Minnesota is a reduction in the *number* of sites identified for placement on the NPL. [' 'Minnesota Pollution Control Agency's Report on the Use of the Environmental Response, Compensation and Compliance Fund During Fiscal Year 1988," November 1988, p. 4.]

61In comparison, at EPA headquarters the Site Evaluation Division of the Office of Emergency and Remedial Response (OERR) has over 30 professional positions in three branches.

62"A Report to the Committee on Appropriations, U.S. House of Representatives, on the Status of the Environmental Protection Agency 's Superfund program," March 1988, Appendix I, p. 31.

63According to EPA data, the actual amount of time to do the work to complete a PA, SI, and HRS Scoring has averaged about 4 months. Thus, a site that made it through the process in the average 5.5 years sits around m the pipeline for over 5 years.

Table 2-1—Pre-Remedlal Program Accomplishments (numbers of sites)

Fiscal year:	1980	1981	1982	1983	1984	1985	1986	1987	1988
Sites in CERCLIS	8,000	10,500	13,386	16,309	18,884	22,621	25,194	27,571	29,987
PA completions'									
Fiscal year total	2,204	1,072	1,209	1,809	4,447	5,181	4,262	4,001	2,953
Cummulative total	2,204	3,276	4,485	6,294	10,741	15,922	20,184	24,185	26,913
Fraction of CERCLIS with PAs	0.276	0.312	0.335	0.386	0.569	0.704	0.801	0.877	0.897
SI completions'									
Fiscal year total	613	428	566	642	1,308	1,618	1,267	1,343	1,258
Cumulative total	613	1,041	1,607	2,249	3,557	5,175	6,442	7,785	9,048
Fraction of SIs required	0.348	0.397	0.448	0.447	0.414	0.406	0.399	0.402	0.562
Number of SIs requiredb.	1,761	2,622	3,587	5,031	8,592	12,746	16,145	19,366	16,100
SI backlog[b]	1.148	1,581	1,980	2,782	5,035	7,571	9,703	11,581	7,052
Percent of sites with PAs that require SIsb c	80	80	80	80	80	80	80	80	60

aAll sites in CERCLIS must have PA; Sites with NFA after PA do not get an SI
bCalculated by OTA from EPA data in rows above
cOne explanation fort ha same, constant percent over 7 years and a reduced percent for FY88 is that the EPA numbers for SIs were generated using 80 percent FY80-87 and 60 percent in FY88.

SOURCE: Office of Technology Assessment, 1989; using EPA data.

would be available, as well.[64] The situation is compounded by the fact that EPA is already over one year behind the SARA date of April 1988 in finishing the development of the new HRS.[65] The scope of the HRS determines the data collection needs for the SI, and although sites scored prior to the availability of the new HRS do not have to be rescored using the new HRS, there will have to be a phase-in period.

Since the CERCLIS list grows at a rate of about 2,000 sites per year, EPA must complete about 2,000 PAs per year to comply with its own policy of completing PAs within one year of CERCLIS entry.[66] EPA's performance, budgets, and projections clearly show that it can complete those PAs. Historically, the program has rejected 20 percent of the sites for which PAs are completed, Thus, 80 percent of all PAs have needed SIs. To keep up, EPA should have been doing about 1,600 SIs (80 percent of 2,000) per year. Only in 1985 has this happened (see table 2-l).

So far, the program has cumulatively completed 56 percent of the SIs necessary. That rate jumped from 40 percent at the end of 1987 because of a PA reassessment conducted in 1988 during which some 3,000 sites were reclassified as not requiring SIs (see later discussion).[67] Before the reassessment, EPA had over 11,500 sites awaiting SIs; that is, almost 4,000 *more sites* awaiting SIs than had been completed in the previous **8 years** of the program!

Now, there may be over 7,000 sites in the SI backlog. And, it may start growing again because EPA's projections for SIs do not appear to take the backlog into account. For fiscal years 1989 and 1990, EPA expects to complete 1,325 SIs each year.[68] At that rate, if 1,600 SIs are required a year, EPA will be adding 275 sites to the backlog each year. If EPA is estimating that only 66 percent of sites with PAs will require SIs, the backlog will persist at the current level. Even if only 50 percent of the incoming sites per year require SIs, only 325 of the backlogged SIs could be done each year.[69] At

[64] The FIT contractor in Region 5 estimated in 1987 that an additional 200 technical staff (tripling their staff level at that time) would be needed to get the SIs required by SARA done in time. [Ecology & Environment, Inc., ''Analysis of and Solutions to Problems Related to the Completion of the SARA Mandated Site Inspection Goal, " Jan. 26, 1987.]

[65] Rules were proposed December 1988; final rules are expected in February 1990.

[66] Congress, through SARA, has required EPA to complete a PA petition (a citizen's request for a PA) within one year. EPA has made that one-yew rule a policy for *all* PAs.

[67] As table 2-1 shows, the reassessment also reduced the cumulative percentage of sites requiring SIs from 80 to 60 percent of PAs completed

[68] U.S. Environmental Protection Agency, ''Justification of Appropriations Estimates for Committee on Appropriations, Fiscal Year 1990, " Other, internal EPA documents show the fiscal year 1990 S1 target to be 1,250 sites.

[69] EPA's workload report calculated that 1,211 SIs would have to be done each year to meet SARA's requirement of 5 years. This assumes that only 60 percent of the sites with PAs—instead of the historical 80 percent-would need SIs.

that rate, it would take 21 years to eliminate the backlog.

The reality is that to eliminate the SI backlog in, say, 5 years while handling the normal flow of required SIs from the PA stage, EPA needs sufficient resources to complete about 3,000 SIs per year. Assuming that no cost efficiencies could be found in conducting SIs, doubling the number of SIs per year amounts to a doubling of the current SI funding level to over $70 million per year for 5 years.

If Not By Adding Resources, How?—Not having asked for or been given budget levels and staffing to rapidly reduce the SI backlog, EPA is trying to manage away the problem of a mismatch between the numbers of sites to evaluate and schedules imposed by Congress.[70] EPA says it has developed a strategy that "reduces the overall pre-remedial workload while increasing resources available for the highest priority sites. "[71] The PA/SI process outlined in the proposed NCP is the result of this strategy. Figure 2-2 is OTA's version of the new flow for site evaluations.

The necessary program changes are, according to EPA's preremedial strategy document, to:

1. More effectively screening out sites that do not require SI through improved PA procedures.
2. Adjusting the way we conduct SIs so that we are more efficient in applying resources appropriately.
3. Increasing the resources available to do PAs, SIs, and HRS scoring packages .72

The first change, ''improved PA procedures, ' has been made by changing the criteria for rejecting sites at the PA stage, and the second change, ''adjusting . . . SIs,' ' means splitting the SI into two stages. While the third change, ''increasing. . . resources," implies increased budgets (and they have been increased), according to the preremedial strategy document, EPA planned to reduce FIT contractors' overhead by 10 percent and to ''exercise future FIT [contractor levels of effort] options now' to increase available funds up to 25 percent, nationally .73 Funds were also to be made available by discouraging the use of FIT resources for non-preremedial work.

The EPA analysis assumed that the current levels of effort allowed for the PAs and SIs were appropriate and that if new work was added, the levels of effort would have to be increased.[74] However, added workload because of increased pace does not necessarily require increased resources. Efficiencies might have been found by evaluating the validity of the established levels of effort for PAs and SIs. One State official has disputed EPA's assumption of a need for increased levels of effort. A bureau chief in Iowa's Department of Natural Resources claims that his staff can complete PAs in about 40 (instead of 100) hours and SIs in 240 (instead of 400) hours. And, EPA could have searched for cost efficiencies by reconsidering the PA/SI process itself. But that option might have been foreclosed by SARA, which by establishing schedules effectively codified the existing process.

One tactic EPA has employed is a one-shot attempt at reducing the level of need for SIs. **To cut down the SI backlog, EPA had the regions retroactively apply the new PA-to-SI rules to sites that had already received PAs.** Many of those sites had been evaluated in the days when the PA decision was a threat/no threat decision. In 1987, EPA identified over 8,000 sites awaiting SIs and in a serious of documents, asked or required the regions

70According to the House Appropriations Committee report cited earlier, in one fiscal year EPA's request fOr increased funding was denied by the Office of Management and Budget.

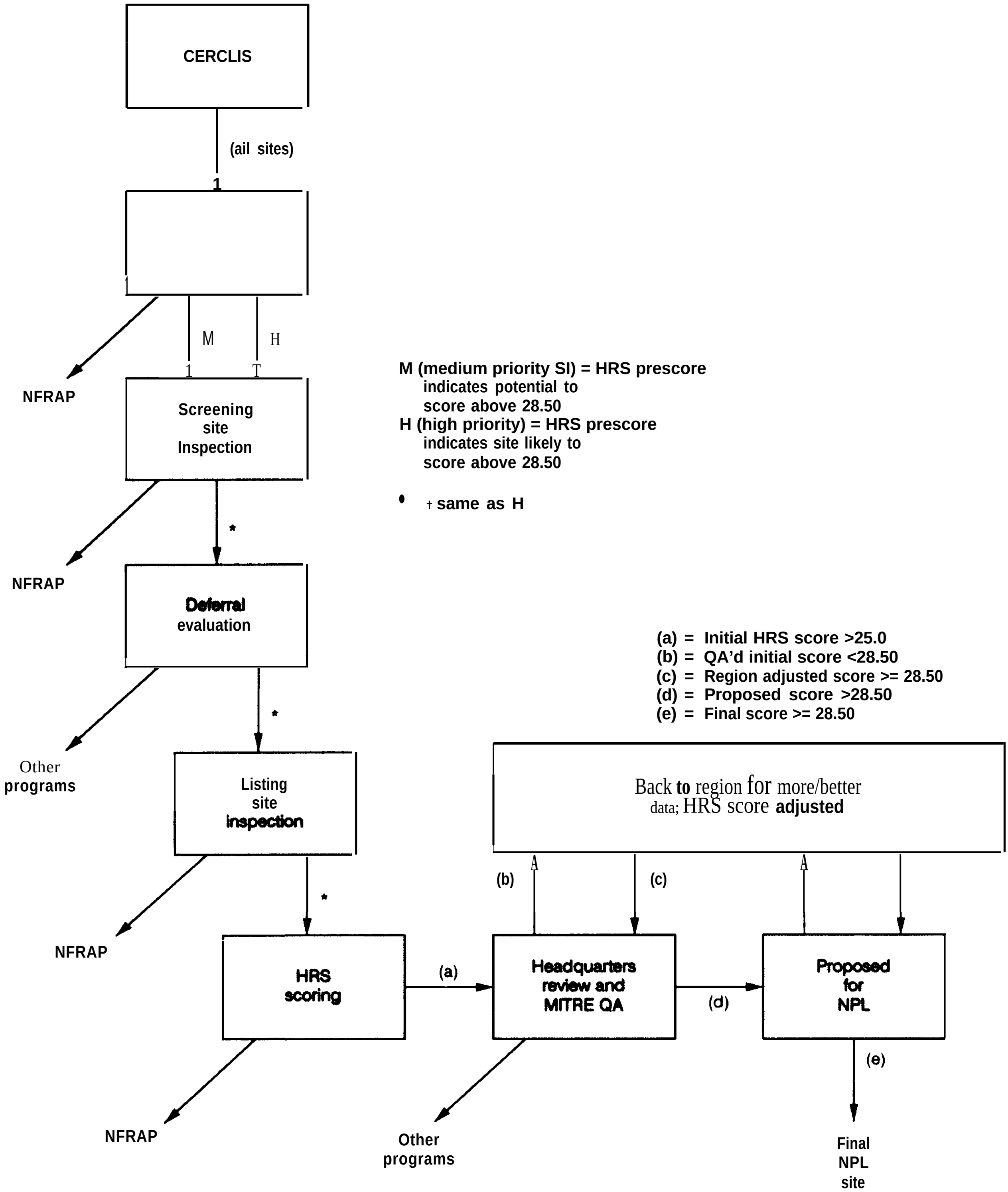
71U.S. Environmental Protection Agency, ''Pre-Remedial Strategy for Implementing SARA, " op. cit., footnote 57, p. 3.

72Ibid., p. 3.

73This Option was suggested by EPA's contract study on workload needs. Although it appears in the EPA's later preremedial strategy document, it was apparently initially rejected by EPA bemuse ''a shortage of experienced staff exists. [Ecology & Environment, inc., ''Workload and Resource Requirements for Preliminary Assessments, Site Inspections, and Hazard Ranking System Evaluations Under SARA, prepared for EPA, October 1987, p. 21.]

74It is clear from comparing the Ecology & Environment report for EPA with the preremedial strategy document that EPA relied heavily on the contractor's conclusions to make decisions about how to find the resources with the current program to increase the workload

Figure 2-2-Revised EPA Preremedial Process



NFRAP (no further remedial action planned) = sites with no reasonable potential to score above 28.50

SOURCE: Office of Technology Assesment, 1989.

to reevaluate these sites.[75] In August 1987, for instance, EPA said:

> A new criterion applies to this re-evaluation. Those sites which do not have a reasonable chance for scoring high enough to be listed should be rated as ‘ ‘no further action This differs from the past tendency to designate ‘‘no further action’ only if there was no hazard potential. The remaining sites should be designated as “medium’ or “high’ priority based on their preliminary **HRS** score.[76]

In a later document, EPA said that although the 8,000 sites already had been given low, medium, and high priority ratings, they must be reassessed “against the new criteria’ and given NFRAP, medium priority, or high priority .77 Regions were given the option of simply reclassifying low priority sites as NFRAP sites without any reassessment.

The reassessment was highlighted in the OSWER *Annual Report, Fiscal Year 1988* as having been necessary to ‘‘more accurately assess the future SI workload. Out of approximately 5,000 low priority PAs re-evaluated, it was determined, said EPA, ‘‘that no further action was necessary at approximately 3,000 sites, ’ saving ‘‘substantial resources for use at more serious sites. ’[78] **What this means is that the *Superfund program* is no longer responsible for 3,000 sites. While the Superfund program may have saved 1.4 million hours of contractor time by avoiding SIs, someone else (State programs, probably) will have to spend a portion of that time and the money to discover which sites are problems.**[79]

False Negatives and Regional Comparability

With SARA, EPA was forced to review the site evaluation process. As discussed above, the reviews have all focused on how to match increased workload needs with available resources. There has been no assessment of the environmental effects of the process or how a speeded up program might affect environmental outcomes, In particular, EPA has not asked the question: How many false negative decisions are we generating?

In any screening process, some false positive and false negative decisions are unavoidable but mechanisms can be built in to minimize these errors. (See OTA’s comparison of the Superfund screening process to a health care model in appendix 2A.) The EPA process in use today is biased toward finding false positives rather than false negatives for two reasons, First, false positives in any serial screening process stay in the universe being evaluated, providing further chances to find them. But false negatives are shunted out of the universe being evaluated and join true negative sites in a universe that receives little, if any, attention. Second, EPA tries to minimize false positives to save unnecessary costs to the trust fund.

No one knows the real frequency of false decisions because EPA has not kept records. But EPA apparently assumes that false *positives* exist at a level that needs to be reduced. One contractor report for EPA looked at the feasibility of spending more resources during SIs to reduce the number of false positives; the issue of false negatives was not raised.[80] In the 1987 workload report (see above), false negatives were mentioned once:

75 There is an obvious discrepancy that we have not been able to resolve between EPA’s 1987 backlog number of 8,000 sites and OTA’s earlier calculation, based on EPA data (see table *2-1)* of over 11300 sites, as of the end of fiscal year 1987. To add to the discrepancy, EPA workload report, **dated** *October* **1987, said that there were 13,719** sites in CERCLIS awaiting SIs. And a 1989 EPA document says there were 7,150 backlogged SIs as of **October 1988.** Since this date is after the PA reassessment was completed and 3,000 sites had been eliminated, it implies that 10,150 sites, rather than 8,000 sites awaited SIs.

76 U.S. Environmental Protection Agency, ‘‘Attached Draft Pre-Remedial Strategy for Implementing SARA, ’ memorandum from Stephen A. Lingle, director, Hazardous Site Evaluation Division, to Superfund Branch Chiefs, Regions I-X, Aug. 14, 1987.

77 U.S. Environmental protection Agency, ‘‘Pre-Remedial Strategy for Implementing SARA, ” op. cit., footnote 57, p. 5.

78 U.S. Environmental protection Agency, “Office of Solid Waste and Emergency Response, Annual Report, Fiscal Year 1988, ” EPA/68-01-7256, November 1988, p. 9.

79 EPA’s workload report in 1987 estimated that it would take 20 hours to reassess a site PA versus 500 hours to perform an S1 for a site. Thus, if these estimates are valid. Superfund spent 100,000 hours on 5,000 reassessments and qualified 2,000 sites for SIs (which will take l million hours to do) instead of spending 2.5 million hours on SIs for all 5,000 sites. [Ecology & Environment, Inc., “Workload and Resource Requirements *for Preliminary Assessments,* Site Inspections, and Hazard Ranking System Evaluations Under SARA, ’ op. cit., footnote 72, p. 19.1

80 Stuart Haus (MITRE Corp.), ‘‘~~ysls of the Impact on Superfund Program Costs of Increased Expenditures for Site inspections, draft, Oct. 28, 1986.

> It is **assumed** that the EPA Regional offices tend to err on the side of caution, and that the incidence of false-negatives is rare [emphasis added] .81

Later, in response to congressional questioning, EPA agreed that there was a statistical possibility of false negative decisions but, reiterated that they were minimized by a' 'conservative approach' to making NFA decisions. EPA further minimized the impact of errors by stating that sites are not precluded from "reentering our system if new information indicates a mistake was made" earlier.[82] This ignores the possible added costs of getting a site under control later rather than sooner.

What Is the Rate of False Negatives?

Many people working in EPA and State programs can provide anecdotal information about false negatives. OTA's comparison of a common screening model used in the health care field and Superfund's site evaluation process shows the inevitability of false decisions (see app. 2A). And, OTA's analysis of the changes in the preremedial process raises questions about the potential for making more false decisions.

While there have been no records kept on false negatives in Superfund, OTA has found some information about false negatives in EPA contractor studies. Using these studies and other information, OTA has estimated that between 240 and 2,000 false negative decisions may have been made so far (see box 2-D).

Evidence ***From*** Studies-In a paper assessing how Region 5 could meet the SARA SI goal by January 1989, EPA's FIT contractor tested the assumption that a PA conclusion is predictive of a site's ultimate HRS score.[83] The point of the exercise was not to calculate false decisions but to make a case for changing the way SI workloads were assigned. The PA conclusions (low, medium, high priorities) for 308 sites were compared with each site's eventual HRS score, The correlation was poor and indicated a problem with both false positives and false negatives. Only 30 percent of the 104 sites with a high priority rating after the PA ended up with an HRS score of at least 28.50. Viewed another way, 70 percent of high priority PAs missed the NPL; the 46 percent (48 sites) that got HRS scores of zero were clearly false positives at the PA and SI stages but were eventually caught by HRS scoring.

At the other end of the spectrum, while 57 percent of the 30 low priority PAs had HRS scores of zero, 10 percent of them received HRS scores higher than the NPL cutoff of 28.50. These sites managed to make the NPL because, for some reason, they stayed in the screening process long enough to be among the 308 sites that got scored, Under the new preremedial process, HRS prescores would eliminate them from further consideration at the PA stage. The study did not consider sites that got dropped out (NFAs) before being scored.

The same contractor stated, in another study: ''False-negatives are those sites that are erroneously classified as NFA **or** low-priority sites after the PA' [emphasis added].[84] This is simply a statement of general reality in the program at the time, that low priority sites often did not move beyond the PA stage. (EPA made this practice explicit policy by eliminating the low priority category in 1988 and, as discussed above, suggested to regions that low priority PAs when reassessed could be simply designated as NFRAP sites.)

Thus, **it is not unreasonable to assume that up to 10 percent of the sites judged as NFAs, in addition to low priority sites, might have made the NPL.** With 26,913 PAs completed through fiscal year 1988 and a historical NFA rate of 20 percent, at least 538 sites (2 percent of PAs completed) may be false negatives. This estimate would be higher if data on low priority PAs were available.

There is other evidence that some NFAs might be false negatives. A Booz-Allen contract study re-

[81] Ecology & Environment, he., ''Workload and Resource Requirements for Preliminary Assessments, Site Inspections, and Hazard Ranking System Evaluations Under SARA, " op. cit., footnote 72, p. 18.

[82] House of Representatives, ''preliminary Findings of OTA Report on Superfund,' hearing before the Subcommittee on investigations and Oversight of the Committee on Public Works and Transportation, Apr. 20, 1988, p. 270.

[83] Ecology & Environment, Inc., "Analysis of and Solution to Problems Related to the Completion of the SARA Mandated Site Inspection Goal," Jan. 26, 1987. The contractor used all HRS packages that it had completed over a 15-month period for which PA priorities were available.

[84] Ecology & Environment, Inc., ''Workload and Resource Requirements for Preliminary Assessments, Site Inspections, and Hazard Ranking System Evaluations Under SARA," op. cit., footnote 72, p. 18.

Box 2-D-How *Many False Negative Decisions?*

From 240 to 2,000 false negative site decisions may have been made so far in the Superfund program. OTA used several sources of information to arrive at these estimates. Each source covers a different set of the Superfund three-stage screening process. Each provides a different estimate. The first two estimates are based on existing studies which **used actual data from the program. The third is a model; the numbers** used are similar to the number of sites that have passed through the screening process. The fourth estimate is based on known errors caught during the *quality assurance* for assigning HRS scores to potential NPL sites.

1) A study by an EPA preremedial contractor in Region 5 covered the **PA stage only.** OTA'S extrapolation from that study gives an estimate of 538 *false negative decisions,* or **2 percent of the PAs completed** through fiscal year 1988.

2) A study done for EPA by Booz-Allen & Hamilton covered both PA and SI stages. OTA's use of that study data provides an *estimate of* 2,056 *false negative decisions* for PAs and SIs completed through fiscal year 1988.

3) An OTA comparison of the preremedial process with a health care field screening model covered all three stages (through the RIFS). The model using conservative assumptions estimates that, for 10,000 sites evaluated (of which 1,200 are problems and would qualify for the NPL), 240 *false negative decisions would be* made.

4) OTA review of two different data sets of sites with SIs moving through the HRS scoring phase of evaluation reveals an 18-20 percent error rate in calculating scores (see pp. 34-36). **EPA data implies that, through fiscal year 1988, just over 7,000 sites have received NFAs after an SI. A 5 to 10 percent error rate in underestimating preliminary HRS scores for these sites would produce 350** *to 700 false negatives. This assumes the* error rate is lower or the same for these sites. Actually, the error rate could be higher since the data can be weaker and less attention may be paid to fine tuning preliminary scorns.

viewed PA and SI files in eight EPA regions to assess how well regional work-reflected headquarters guidance.[85] In each region, a random set of PA and SI files was selected for evaluation. As part of the study, Booz-Allen checked to see if NFAs were justified by data in the files. In 406 PA files, 19 percent had NFA recommendations, and Booz-Allen concluded that 28 percent of those decisions (22 out of 79) were not supported by PA file contents. Thus, 5 **percent of the PAs completed may be false negatives.** Regionally, the percentage of possible false negatives, out of PAs completed, ranged from zero to 16 percent. Out of a total of 212 SIs reviewed, 31 percent of the 49 NFAs, or 8 **percent of the SIS completed were possible false negatives.** For SIs the range of possible false negatives across the regions was larger than that for PAs; from zero to 30 percent.

If the Booz-Allen data is extrapolated to the total numbers of Superfund PA and SI decisions, estimates of the national count of possible false negatives can be made. As of the end of fiscal year 1986, the period when the Booz-Allen study ended, EPA says it had completed 20,184 PAs and 6,442 SIs. Using the Booz-Allen rates, 1,009 PAs and515 SIs could be false negatives. Between then and now, another 556 false negative decisions may have been made for a total of 2,056 sites. This does not necessarily mean that over 2,000 sites might be added to the NPL but that 2,000 problem sites may be hidden among the universe of sites rejected by the Superfund program through fiscal year 1986. Some of them may qualify for the NPL,

Anecdotal Information-Every region, every State can provide examples of false conclusions. Anecdotes, however, are only possible when sites once judged no problem become evident as problems. In other words, until they resurface, false negative sites are unknowns. Known turnarounds include:

[85] Booz-Allen & Hamilton Inc., "U.S. Environmental Protection Agency Preliminary Assessment and Site Inspection Program Quality Assurance Review," draft, Sept. 11, 1987. The study did not include Regions 4 and 9.

- The Illinois Environmental Protection Agency provided OTA with a list of 12 sites that had been tagged after the PA with a low priority or NFA designation. Two sites (one low priority and one NFA) are now on the NPL; the balance are being cleaned up under the State program.
- In Region 8, the Martin Marietta (Denver Aerospace) site had a PA/SI completed early in the Superfund program that did not involve any sampling. Contamination was later discovered when monitoring wells were required under RCRA. After a second SI was completed, the site was proposed for the NPL in 1985 with an HRS score of 46.01.
- The California State program told OTA of a San Diego site that was evaluated using existing county information and determined to be no threat. Later, when the property was sold and construction began, an old incinerator was found and a cleanup is now required.
- In a survey prompted by SARA, EPA regional staff identified a dozen sites that they felt should be on the NPL but were not because they had not qualified.

Why Regional Differences?—The wide variance in NFA rates by EPA region, reported in various EPA documents, may indicate that regions get distinctly different kinds of sites to evaluate. The wide ranges of unsupported NFAs in the Booz-Allen study, however, indicate that it is more likely that the differences are a product of varying regional and contractor staffing problems and the fact that, until January 1988 an official PA guidance document did not exist. (The SI guidance document awaits the new HRS.) The first directive covering site evaluation was issued in February 1988, 7 years *after* the program began to evaluate sites. The consequence is that where you live may determine how many sites are being ignored.

Turnover and lack of skills, as reported in OTA's *Assessing Contractor Use in Superfund*,[86] will certainly decrease the accuracy and reproducibility of PA, SI, and HRS decisions. A MITRE official, for instance, told OTA that turnover in the EPA regions severely affects their ability to assure that properly trained employees do HRS scoring packages.[87]

For the first 8 years of Superfund, EPA regions had some direction from a State participation manual issued in 1985 and in some training manuals. FIT contracts also provide some written detail on PAs and SIs. EPA did issue PA and SI forms for use by the regions that could have provided some consistency, but not all regions used the forms and revised forms have been ignored in some cases. The Booz-Allen study found that only one of eight regions used a form on all PAs, Region 9, which was not included in the Booz-Allen study, told OTA that the FIT contractor does not use a form in reporting PAs.

Issuance of the recent documents may improve the consistency across regions or it may not. Comments made to OTA by an EPA headquarters official in the Hazardous Site Evaluation Division indicated that while headquarters would prefer all regions to conduct their work in a consistent manner, EPA is unwilling to require them to do so. Guidance documents do not assure consistency unless they are followed. Assuring national consistency may require periodic evaluations of regional performance.

Where Do False Negatives Go?—False negative sites from Superfund end up in the universe of sites rejected by the site evaluation process. The universe includes NFA (now NFRAP) sites as a result of PAs and SIs; sites rejected when their official HRS scores fall below 28.50, either before being proposed for the NPL or afterwards when MITRE Corp. does the quality assurance; and sites that EPA has decided—on a policy basis-do not belong in the Superfund program.

Since an estimated 90 percent of the sites in CERCLIS don't make the NPL, the universe of rejects is now approximately 17,000 sites. OTA has calculated (see above) that the false negatives within this universe may total over 2,000 sites. Currently, there is no easy way to track the fate of these sites. In the CERCLIS database an NFA entry indicates the site decision. In the early years, when sites were rejected primarily because no environmental threat

[86] U.S. Congress, Office of Technology Assessment, *Assessing Contractor Use in Superfund*, OTA-BP-ITE-51 (Washington, DC: U.S. Government Printing Office, January 1989).

[87] MITRE is the firm that developed the HRS and holds the contract from EPA to do the quality assurance on all HRS scoring packages completed by regional offices (see later discussion).

existed, there was no need to develop any kind of a notification system other than the CERCLIS entry. But, as EPA's own statements show, today's NFRAP is not an indication of lack of hazard. Although 1988 EPA documents tell the regions that they should notify States of sites that are rejected, there has never been, nor is there anticipated to be, a formal national notification process.[88]

In the last couple of years, States have increasingly taken on more of the task of site evaluation. When States do the PAs and SIs, they know when sites are rejected and presumably assume responsibility for those sites that are problems. But, once a site is judged NFA or NFRAP by EPA, funding by Superfund effectively dries up.[89] For the sites rejected in years prior to State involvement, any State has two choices: 1) to verify all EPA decisions (i.e., search for false negatives), or 2) to assume EPA decisions are correct and focus on the more obvious problem sites—those that have moved through the preremedial process and have at least acquired an estimated HRS score.

While State cleanup resources vary widely, they are often extremely limited. Taking the first path—reevaluation-is most unlikely. Louisiana, for instance, has 297 sites that EPA has tagged as NFAs. While the State cleanup program admits that an unknown number of those sites may require some attention, there are 209 sites yet to receive evaluation. Louisiana is a State program with few resources; a 1989 report poses the question: ''where will the necessary resources be found?"[90] Initiating 209 site evaluations for which Superfund funding is available has a higher priority than reevaluating 297 NFA sites for possible false negatives. Conversely, the New Jersey State program's strategy document states that no site should be listed as NFA ''without file documentation and sampling to justify no action."[91] This is a recognition that a State does not really have the option to defer sites away to someone else.

The HRS and the NPL

The HRS score calculated for a site determines whether or not it will go on the NPL. Being *on the NPL* (a score of 28.50 or more) means that trust fund money may be spent for remedial action. Being off *the NPL* (less than 28.50) means trust fund money cannot be spent for remedial action. The NPL also has informational significance; NPL sites receive considerably more public attention than non-NPL sites. The critical on/off decision is an EPA policy enunciated in the NCP. Congress, through SARA, expressed concern about the HRS and its use by requiring EPA to examine several questionable effects of the HRS and to revise it by October 1988 (See box 2-E).

Scoring

Once a site has received an SI, the EPA region prepares an HRS scoring package. For those sites scoring at least 25.00, the package may be submitted to EPA headquarters, which turns it over to the MITRE Corp. for quality assurance (QA).[92] If verified at 28.50 or greater the site is eligible for proposal for the NPL. Sites get returned to regions if the QA determines that the score is less than 28.50 or if the information does not support the score. Sites can also get returned to regions if EPA determines that the site does not qualify for the Superfund program (i.e., is exempt by the statute or policy).

Once formally proposed for the NPL, a site is subject to public comment, which may push the score, up or down. If adjusted below 28.50, the site is removed from the proposed list. If not, sites are eventually placed on the final list.

EPA says the HRS process currently costs an average of almost $60,000 per site, including the PA

88 A 1989 EPA document says that EPA *must* notify States when a site is given an NFRAP after a PA and *should* inform States when a site is given an NFRAP after an SS1.

89 Not officially, however, since the NCP rule is that Superfund monies cannot be used for *remedial action* at non-NPL sites. That means that, legally, States could apply for funding to continue site evacuation when EPA stops doing so.

90 Louisiana Department of Environmental Quality, *Progress and Problem: Cleaning Up Louisiana's Inactive and Abandoned Hazardous Waste Sites*, A Report to the Louisiana Legislature, April 1989, p. 1.

91 Department of Environmental Protection, "New Jersey's Case Management Strategy for Hazardous Waste Programs Remedial Actions, " June 1987.

92 Review of MITRE data indicates that only 12 sites submitted for QA since 1984 had scores less than 28.50.

Box 2-E--SARA ***and the Hazard Raking System***

In The Statute:

- **Section 105 (c)(l): ".. . assure, to the maximum extent feasible, that the hazard ranking system accurately assesses the relative degree of risk to human health and the environment posed by sites and facilities subject to m-View"**
- **Section 105 (c)(2) :"... ensure that the human health risks associated with the contaminati"on or potential contamination. . . of surface water [used for recreation or potable water consumption] are appropriately assessed. . ."**
- **Also, in Section 125,** EPA was ased to revise **the HRS to "assure appropriate consideration" of specific site characteristics of facilities that generate wastes such as fly ash, bottom ash, and slag and that emit flue gases.**

In the Conference Report, Congress asked EPA to determine:

- **the "effect of establishing a threshold value of 28.5 for facilities to be included on the [NPL]," and**
- **"whether a new threshold value should be established."** [1]

l s ~ ~~ pp. m-m.

and SI.[m] EPA pays up to $4 million per year for the MITRE **QA** services. Based on the average number of sites processed each year, QA alone may cost over $12,000 per site.

At Least 28.50!

The cutoff score of 28.50 has no technical basis. It is an arbitrary number; or, as EPA calls it, a management tool.[94] The number was selected in 1982 to come as close as possible to the 'at least 400 sites" required by CERCLA for the first list.[95] In essence, then, the hazard level of the first set of sites from which the initial 418 NPL sites came serves as the basis for inclusion on the NPL of all sites that have followed. If the hazard level of that collection had been lower, then some sites subsequently rejected for the NPL would be on the NPL.

The cutoff score is often reported as 28.5 instead of its real value: 28.50. The zero plays a major role and implies a certain numerical precision even though there is no possible technical rationale for taking the number to the hundredths, or even tenths.[96] Using two decimal places means that a site with a score between 28.45 and 28.49 does not get rounded up to 28.5 and get on the NPL.

EPA is aware of the HRS score's lack of precision but has not used that information to convert the cutoff *score* to a whole number or a range. In a study prepared after SARA, EPA commented: "because of the uncertainties associated with the HRS, it is possible that a site scoring 35, for example, is more hazardous in terms of absolute risk than a site scoring 36."[97] If so, then there maybe no difference between a site at 29 and one at 28. Furthermore, EPA groups the sites on the NPL instead of listing them by HRS score ''to emphasize that minor differences in scores do not necessarily indicate significantly different levels of risk. "[98]

Once calculated, the HRS score does not serve any official function other than to make the NPL decision. EPA repeatedly states that the HRS only measures relative risk and that the score is not used as a way of making priority cleanup decisions. However, OTA showed in its 1985 report, *Superfund Strategy,* that, even if the method works, the resultant score is **not** a measure of relative risk

93 53 Federal Register 51%2, Dec. 23, 1988.

94 In a Report to Congress, EPA stated that the cutoff score was not chosen because ''. . . it represents any threshold in the significance of the risks presented by sites. " [U.S. Environmental Protection Agency, Office of Solid Waste and Emergency Response, Progress *Toward Implementing Superfund, Fiscal Year 1987, Report to Congress,* EPA 540/8 -89/003, April 1989, p. 27.]

95 EPA had a list sites that had been scored. Using *28.50 as the* cutoff score generated a proposed list Of418 sites.

96 Any mathematician, scientist, or engineer knows that the solution to any calculation can only carried to as many decimal places as the least accurate number for any data used in the calculation. Many numbers used in the HRS calculation are whole numbers, i.e., 3 or 5.

97 U.S. Environmental Protection Agency, ''HRS Revisions Support: SARA Studies on HRS Scores and Remedial Actions, HRS Scores and Potential Dangers, and the Effect of the 28.5 Cutoff Score," November 1987, p. 9.

98 53 Federal Register 51962, Dec. 23, 1988, p. 5 IQ&$. Each group has 50 sites; each time new sites are added to the NPL, the sites are reordered into new groups of 50.

because not all sites receive complete scores.[99] HRS scores are a combination of three possible routes of exposure--groundwater, surface water, and air. Air subscores are frequently not calculated if the two water routes or even one of them provides enough information to push a site score over 28.50.

For this report, OTA reviewed the sites submitted to MITRE for QA since 1984. Ninety-four percent of sites were submitted without air subscores and 97 percent of the sites with final HRS scores did not have air subscores. Conversely, 98 percent of the final sites *do have* groundwater subscores. Thus, the resultant HRS score says, at the most, something about the relative risks of sites due to contaminated groundwater.

It is true, however, that the HRS score does not set priorities.[100] Higher ranking **sites do** not necessarily **move through the system first or faster.** On the whole, there is no discernible relationship between a site's HRS score and, say, the start of its RIFS. It is, in fact, possible to find cases where an inverse relationship exists, such as occurred with three sites Region 7 proposed for the NPL in April 1985.

Despite its lack of technical foundation and usefulness after the fact, the cutoff score has taken on serious meaning within the context of its use. Regions, States, public interest groups, industry, and others are at times keenly interested in making sure that a particular site gets on the list or stays off. States with few resources or without an enforcement program to clean up sites on their own may prefer to have as many sites as possible on the NPL so that they only have to pay the CERCLA 10 percent match rather than the full cost of cleanup. Some States, like California, make decisions in advance about whether or not they want a site on the NPL and in the Superfund program. If not, they intentionally keep sites off the CERCLIS, which eliminates them from consideration and scoring, Having a site on the NPL projects a negative image to the public, and a company may have a strong interest in not having its property listed.[101] Communities have been known to press to keep a site off the list because of the stigma and possible negative affect on their economic welfare. Congress can intervene; through SARA the Silver Creek Tailings site in Park City, Utah, was effectively removed from the proposed list.[102]

This jostling to be on or off the NPL position adds inefficiency to and detracts from Superfund's environmental role. As discussed previously, **calculating an HRS score is not a science.** When scores are close to the cutoff, it is not uncommon for regional offices or State agencies to seek out more information to move the score above the cutoff. A Kansas State official, who was formerly with the New Jersey State program, told OTA that he got 65 sites on the NPL by *sampling judicious).*[103] There is no way to calculate the national costs of pushing sites over 28.50, because the data to do so is spread among files in 50 States and 10 regional offices. It could be significant if each entity spends extra money and time on just a few sites each year.

Because the setting of an HRS score on a site has been made a regulatory procedure, the score must be proposed for public comment. This can cause reworking of the number and has led to an official change of at least 224 site scores. Twenty-three of those sites were removed from the proposed list because their revised score was below 28.50. For 66 of the 224 (30 percent), the difference between the proposed and revised scores was less than 1 point and ranged from 0.01 to 0.99, up or down. In none of these cases did the reworked score affect the site's NPL status. Obviously, the level of effort required to rework scores varies, but EPA was unable to give

[99] U.S. Congress, Office of Technology Assessment, *Superfund Strategy,* OTA-ITE-252 (Springfield, VA: National Technical Information Service, April 1985), p. 163.

[100] In California it does: "... cleanup priorities are now generally established based on the HRS migration score . . .' [California State Department of Health Services, *Expenditure Plan for the Hazardous Substance Cleanup Bond Act of 1984,* revised January 1988, p. 51.] In Louisiana, priorities are based on whether or not a PRP exists. In Minnesota, sites in the cleanup pipeline have precedence over sites that have not yet started the process.

[101] Not being listed does not prohibit EPA from taking CERCLA enforcement action, however.

[102] SARA, Section 118(p). See also, "Selection for Superfund List Puts Utah Resort in Dumps: As EPA Reconsiders, Property Values Plunge," *Washington Post, Mar. 2,* 1987, p. A1.

[103] Or, as a New Jersey document states: "Since the criteria for placement (HRS) was relatively undefined, NPL placement was easily accomplished. [Department of Environmental Protection, "New Jersey's Case Management Strategy for Hazardous Waste Programs Remedial Actions," June 1987.]

OTA a rough estimate of the cost and staff time (for either EPA or MITRE).[104]

What Are the Results?

So far, almost **2,000** sites have gone through the HRS scoring QA system. As of July 1989, 1,274 sites have been proposed for the NPL. Accounting for removals and deletions from the list, the NPL stands at 1,224 proposed and final sites. Figure 2-3 shows NPL actions by EPA from fiscal year 1983 through 11 months of fiscal year 1989.

CERCLA requires EPA to update the NPL at least once a year. Since 1983 there has been at least one update a year for the proposed or final list.[105] In addition, at irregular intervals sites have been removed (from the proposed list) or deleted (from the final list). NPL removals are part of the HRS scoring process, while deletions occur after remedial actions have been completed. Deletions can also happen because the Remedial Investigation/ Feasibility Study (RIFS) shows that no remedial action is necessary.

There are a number of aspects of HRS scoring that deserve some attention. The managerial significance but questionable relevance of the cutoff score has already been discussed. Another aspect is the reworking of scores. This is driven by the cutoff score and the need to calculate THE score for a site. MITRE checks the region's work before a site is proposed and then, as discussed above, scores can be altered between proposed and final listings. When a score is revised, it is because some kind of error has occurred. It could be caused by poor math, improper use of the HRS, or inaccurate or incomplete information. While the two revision points catch those errors, examination of some data raises questions about why regional work varies so much and what the errors and the variances mean regarding sites that do not make it to the first or second revision point.

The difference between the number of sites submitted for QA (1,970) and sites proposed (1,223) through early 1989 says that almost 40 percent of the sites submitted by regions have not been proposed. Some were rejected for policy reasons. Some have not yet been verified. These two categories may represent up to half of the nonproposed sites.[106] Then, almost 400 of the submitted sites may have been rejected because of errors in the regional offices. QA determined that their scores were below 28.50 or there was not proper documentation backup. **If 20 percent of the submitted sites have errors, for how many sites that do not get submitted (i.e., judged NFRAP after an SI or PA) are errors made?** Remember that there is less documentation for and attention paid to these sites for which HRS scores are only estimated.

The fact that the 224 scores discussed earlier could be changed between proposed and final listings is indicative of the lack of precision to an HRS score. More information and reevaluated information can change an HRS score, up or down, by as little as 0.01 and as much as 31.57 points. In terms of the on/off NPL decision, all of these errors are significant since 0.01 can make the difference between 28.49 (off the NPL) and 28.50 (on the NPL). From a risk perspective, however, because 79 percent of the errors are between 0.01 and 10.0, the errors may be insignificant. Should the Superfund program be spending money looking for and adjusting site scores, especially when they have no relevance afterwards?

While the 224 changes represent a national error rate of 18 percent in proposed scores, the rate ranges from 8 to 37 percent by region. Nationally, most of the errors (54 percent) are on the plus side (i.e., the proposed score is higher than the final score), but they range from 37 to 66 percent on a regional basis.[107] Four regions have higher minus than plus

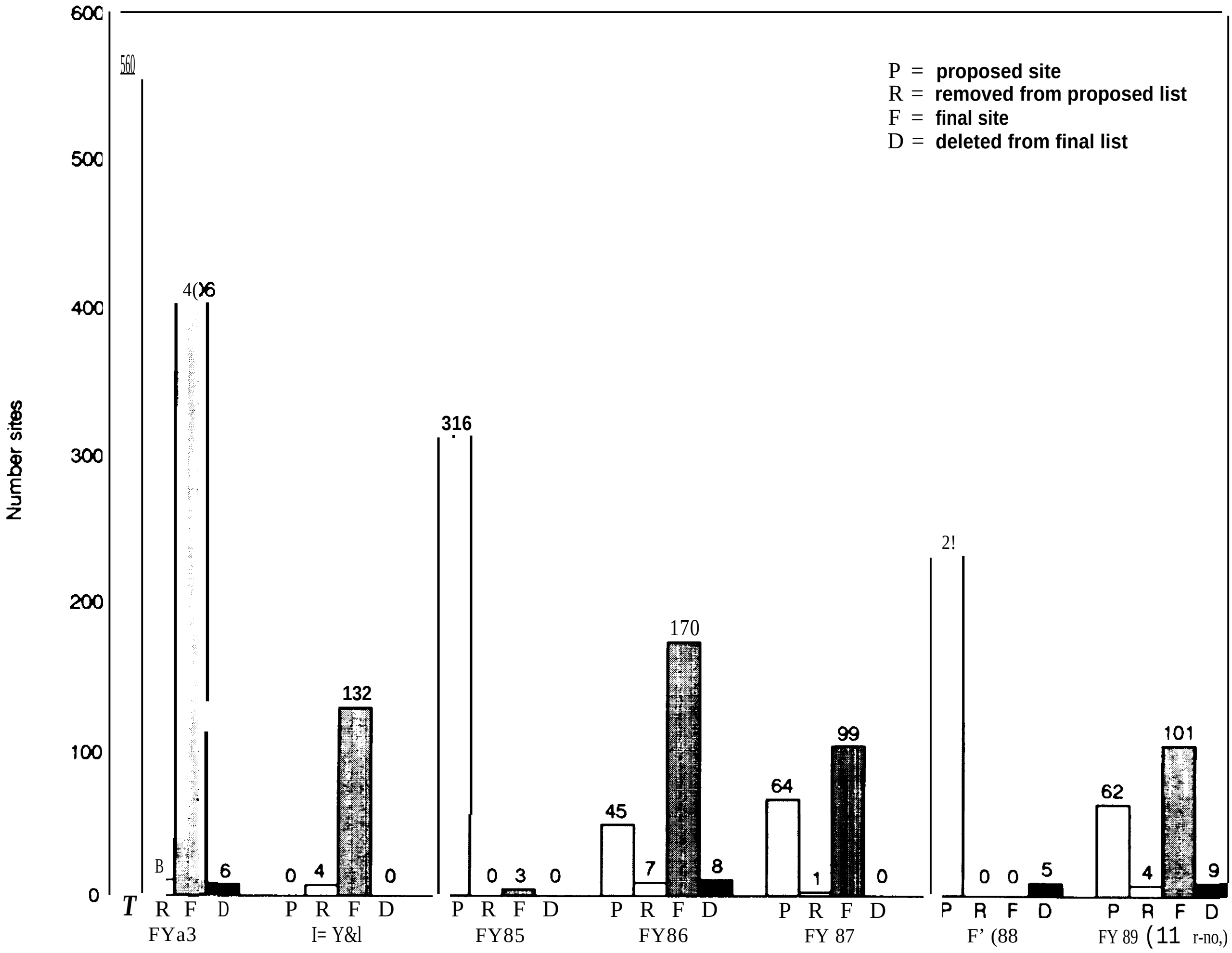
[104] EPA did provide OTA with detail on five score revisions. Obviously, the amount of time and resources necessary to respond to public comments is proportional to the complexity of the comments rather than to the eventual score change. Of *the five* scores, two were each changed 0.01 points (one went up, the other down) because of rounding errors discovered while considering comments. The other three were revised (maximum, 0.14 points) because of comments received regarding the surface water subscores.

[105] The proposed listing may be more important than a final listing. EPA does not necessarily wait until a site gains final status to start the RIFS or the search for PRPs.

[106] This is a *very* conservative estimate. For instance, less than 4 percent of the sites submitted for QA in 1988 and 1989 still wait to be proposed.

[107] This national trend for overestimating scores may be confirmed by MITRE data on all initial (ss submitted by regions for QA), proposed, and final site scores. On average, as sites move through the QA process, they systematically decline. It may be, however, that score depression is a product of the QA process, rather than regional tendencies to overestimate initial scores, especially since score depression occurs between proposed and final scores *aim.*

Figure 2-3-National Priorities List Actions, Fiscal Years 1983-89



NOTE: Shading indicates that more than one update occurred in a fiscal year.

SOURCE: Office of Technology Assessment, 1989; based on EPA data.

error rates which means they are more likely to underestimate HRS scores. Finally, **if 18 percent of proposed scores are in error, that strongly suggests a significant error rate on sites that don't make it that far.**

False Negatives and False Positives

The error rates discussed above suggest that false decisions (both positive and negative) can persist through the HRS scoring stage. EPA does not know how many false negatives scoring has created in 7 years of use but does know for sure that at least three false positives exist because three sites have been deleted from the NPL after an RIFS was completed.[108] While EPA is "concerned that lowering the [cutoff score] might substantially increase the number of [false positives] . . .," the agency has not expressed a corresponding concern about false negatives.

108 53 F-dRe@~r51962, Dec. 23, 1988, p. 519M, There may be more than three false positives. OTA reviewed all RODs issued in the program. At least nine sites have been deleted because no remedial action was recommended in the ROD.

EPA's Science Advisory Board (SAB) assumes both types of false decisions have been made and has suggested that EPA establish a review process:

> Screening models like the HRS must be simple. They do not have much resolving power and therefore, some false positives and false negatives are inevitable. Because of this limitation, HRS scorns should not be overemphasized. A process should be established either to review sites subject to scoring or to review HRS scores in an attempt to spot fake positives and negatives.[109]

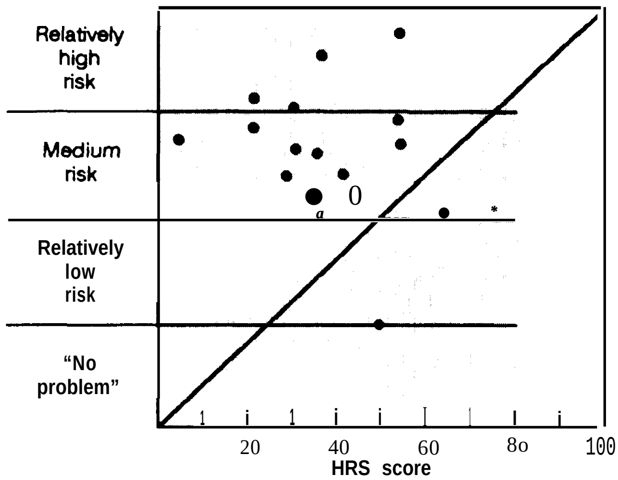
Some studies and certain aspects of the HRS point to the possibility of false decisions. In one study, EPA compared current HRS scores with potential dangers at 32 sites using a risk assessment methodology. The results showed poor correlation between the HRS score and the potential danger at a site. The study concluded:

> Based on the sites used in this study, potential false-negative results (sites that pose potentially significant risks, but receive HRS scores below the cutoff) are more common than false-positive results (sites that receive HRS scores above the cutoff, but do not pose potentially significant risks). *10

For another study, EPA's Office of Policy Planning and Evaluation (OPPE) assembled a panel of EPA experts from the various disciplines involved in site evaluation. The panel studied documentation on a set of Superfund sites, developed its own ranking methodology, and scored the sites by consensus along a spectrum from high risk to no problem.[111]

When the panel's ranking of sites is compared with the sites' HRS scores, the HRS is shown not only to be a poor predictor of risk-as judged by experts-but also to be source of many false negative decisions, As figure 2-4 shows, 6 of 18 sites (33 percent) with scores at or below the cutoff were judged by the panel as medium or high risk. Another conclusion is that the HRS underestimates risk as most of the panel's decisions placed the sites above the diagonal line that represents concurrence between the panel and the HRS.

Figure 2-4-EPA Panel's Ranking of Sites v. HRS score



SOURCE: Environmental Protection Agency.

On the NPL are three examples of the HRS as a poor predictor of risk. Under CERCLA, each State is allowed to place one site on the NPL regardless of how it scores. Three such sites, with scores of 5.49, 8.27, and 17.68, have moved through the system and received RIFSs, and RODS have been issued. In all three cases, EPA has decided that a remedial action is necessary. If the sites did not present a risk, EPA could have decided that no action was necessary and deleted them from the NPL. While the cost of cleaning up these sites ranges from $1 to $2 million and is below the average for Superfund sites, the sites are not dissimilar to many sites that qualify for the NPL on the basis of higher scores.

Two other problems with the current HRS, which may be corrected by the pending new HRS, are also creating false negatives. When there is not enough information to assign a number to a factor, a default

[109] U.S. Environmental Protection Agency, Office of the Administrator, Science Advisory Board, 'Review of t-be Superfund Hazard Ranking System,' ' SAB-EC-88-008, January 1988, p. 6.

[110] U.S. Environmental Protection Agency, ''HIM Revisions Support: SARA Studies on HRS Scores and Remedial Actions, HRS Scores and Potential Dangers, and the Effect of the 28.5 Cutoff Score," op. cit., footnote 97, p. 50.

[111] Applied Decision Analysis, Inc., "A Site-Ranking Panel Evaluation of the Relative Risk Posed by Twenty Superfund Sites, " draft, July 14, 1987, pp. 44-45.

value has been used.112 In 1982, EPA advised scorers to assign a default value of zero, [113] The result for toxicity is, according to an EPA scientist:

> A default value of zero would enhance the possibility of false negatives in the absence of toxicity data, while a default value of 5 would tend to enhance the possibility of false positives, A mid-range default value of 3 in the absence of appropriate toxicity information would reduce any directional bias toward either false positives or false negatives.[114]

In the proposed HRS, some default values have been adjusted. For example, a default value of 3 (a midpoint in the toxicity scale) has been proposed for a pathway's toxicity factor value when 'appropriate toxicity data for scoring does not exist for any hazardous substance relevant to that pathway .''[115]

Under the current HRS, when a site appears to have only a direct contact threat, the site will not make the NPL.[116] But, two EPA contractor studies have shown that a threat of direct contact is a major rationale for remedial action. EPA has in the past added two sites to the NPL, using provisions in CERCLA other than the HRS, when their scores were below the cutoff, because of their direct contact threat. [117] If direct contact is added as a new pathway in the proposed HRS, these kinds of sites may have a better chance of being on the NPL. However, among the universe of sites scored under the current HRS, they are false negatives. Some, but not necessarily all, may have been cleaned up by removal actions.

How Will the New HRS Change Superfund?

Many concerns had been raised about the efficacy and fairness of the HRS. Box 2-E lists the concerns expressed by Congress in SARA. A summary of major problems with the HRS from OTA's *Superfund Strategy* are included in box 2-F. Box 2-G contains a summary of recommendations made by EPA's Science Advisory Board (SBA).

Has EPA Resolved the Concerns ?-For this report, OTA could not do a comprehensive analysis of the new HRS because the new HRS does not yet exist. EPA proposed the new HRS in December 1988 but does not plan final action on the rules until February 1990 (almost 2 years beyond the promulgation date that SARA specified).[118] And, so far, EPA has not tackled a major issue—the algorithm—cited by the SAB:

> The Subcommittee places special emphasis on the algorithm issue because it is impossible to review the components of the HRS without considering how the components fit together.[119]

In the proposed rules, EPA recognized SAB's concern and said about suggested changes:

> EPA is planning to evaluate and possibly test such changes in the algorithm prior to promulgating a revised FIRS,[120]

The SAB was concerned about the algorithm because it is the basic logic of the model. The board implied that the way the current HRS was designed may have been a backwards approach. A better way, according to the SAB, is a risk assessment approach that begins with an understanding of how to list sites quantitatively if all needed information and resources were available. This risk assessment model is then transformed into a scoring system and simplified to operate at reasonable cost and with sparse information. [121]

Despite outside concern and advice, EPA has postponed consideration of the basic logic of the

112A factor in the HRS is a point at which a number is assigned based on the judgment of the person doing the scoring package. For instance, a factor could translate toxicity data mto a numerical value for calculation purposes.

11347 Federal Register 31222, July 16, 1982.

114Chris &Row, chief, EPA's Chemical Mixtures Assessment Branch, letter to Scott Parrish, acting chief, Hazard Ranking and Listing Branch, Sept. 22, 1987.

115U.S. Environmental Protection Agency, preamble to the HRS proposed rule, draft, ca. February 1988, p. 40.

116A direct contract threat means that a person could come into direct contact with toxic substances at the site.

1 17The sites and scores are Landsdowne Radiation site, PA (20.32) and Quail Run, MO (21.19).

118SARA Section 105(c)(1) required the new HRS to be promulgated by April 1988 and to be in effect by October 1988.

119U.S. Environmental Protection Agency, "Review of the Superfund Hazard Ranking System, " op. cit., footnote 109, p. 5.

12053 Federal Register 51%2, Dec. 23, 1988. p. 51970.

121See, U.S. Environmental Protection Agency, "Review of the Superfund Hazard Ranking System, " op. cit., footnote 109, Appendix A5.

Box 2-F—The Office of Technology Assessment's 1985 Comments on the Hazard Ranking System[1]

OTA summarized other criticisms of the Hits at that time. Problems identified were that the HRS:

- had a bias toward human exposure at the expense of the environment
- had a bias against low density populations;
- required documentation for air releases but none for water;
- scores were based on site contaminantt rather than known or potential released contaminants;
- averaged route scores creating a bias against a site with only one route score greater than zero; and
- considered only waste quantity rather than quantity and distribution.

[1] U.S. Congress, Office of Technology Assessment, *Superfund Strategy*, OTA-ITE-252 (Springfield, VA: National Technical Information Service, April 19ss).

Box 2-G-Summary of EPA's Science Advisory Board Hazard Ranking System Recommendations

When the Superfund program requested the Science *Advisory* Board (SAB) to review the HRS, only three specific issues were posed: types of toxicity to address and how, relevant distance from a site for air pollutants, and large volume wastes and waste concentrations. On these issues, SAB recommended that:

1. the toxicity rating scale in use be replaced by multiple measures of toxicity and exposure measures to be improved;
2. since a potential for air release seemed appropriate and possible, a scoring system weighing population exposure in concentric rings be employed; and
3. although applying the HRS to mining (large volume) sites has not treated *them with* systematic error, it does have the potential to do so and could be improved with the adoption of several factors.

But, the SAB subcommittee that studied the HRS chose to also address some fundamental issues. For instance, the SAB said: "Improving the algorithm could potentially do more to improve the HRS than fine-tuning individual Components."[1]

[1] U.S. Environmental Protection Agency, Science Advisory Board, "Review *of* the Superfund Hazard Ranking System," SAB-EC-88-008, January 1988, p. 20.

HRS. Meanwhile, the proposed HRS is replete with the "fine tuning" that the SAB thought to be less important than the algorithm. These changes apparently will result in "vast new data requirements' that also concerned SAB. **EPA says that the new HRS will cost almost $150,000 per site; that is, two and a half times ($90,000 more than) the current HRS costs.** EPA expects the new HRS to add $56 million in total costs to the program.[122]

EPA says it **"expects** that the changes will result in increased accuracy in assessing the relative degree of risks to public health and the environment **for certain sites** [emphasis added].''[123] EPA also says that "at this point, it is impossible to predict whether the revised HRS would result in more or fewer sites being included on the NPL. " Unanswered is Congress' concern that the HRS *accurately as* possible assess the relative degree of risk posed by sites.

EPA did not, according to the SAB, properly assess the current HRS prior to proceeding with changes. Also, EPA has apparently not tested the proposed HRS to assess its impact (other than cost of using it). EPA's request of the SAB was made after an advance notice of a proposed rule on the HRS had been published (April 1987) and public comments received. That is, the process of revising the HRS was well underway at the time of SAB's involvement. Thus, SAB made suggestions about how to better proceed with the next revision. On the top of its list was the need for an ''empirical retrospective evaluation of how successfully the FIRS predicts risk . . . based on an in-depth technical review."[124] What SAB had in mind was a review comparing sites' HRS scores with the knowledge gained as a result of their RIFSs. In other words,

[122] 53 Federal Register 51%2, Dec. 23, 1988, p. 52002.

[123] Ibid., p. 51966.

[124] U.S. Environmental Protection Agency, "Review of the Superfund Hazard Ranking System, " op. cit., footnote 109, p. A6-1.

EPA should ask the question "How does it err in practice?" before trying to fix it.

The copious proposed changes in the kinds of data used and how used to calculate an HRS score have been backed up with a plethora of contractor studies. But, the algorithm remains the same and there has been no analysis of the *combined affect* of the changes. The debate will continue as to whether or not the HRS can provide an appropriate or accurate measure of relative risk among sites. As the SAB said:

> Each step in this process affects the final score and, therefore, how well the HRS discriminates between sites of greater and lesser risk to human health and the environment. [{x]

As to the steps, EPA appears to have adopted some of the SAB's recommendations regarding toxicity, air releases, large volume sites, and waste concentrations. The proposed HRS has four instead of three pathways: surface water, groundwater, air, and onsite exposure. The new onsite (or direct contact) pathway may improve one shortcoming of the current HRS, as discussed earlier. Currently, direct contact is only used to determine whether or not a site needs a removal action.

In the HRS, current and proposed, each pathway is made up of three categories (release, waste characteristics, and targets) and each category has a number of factors. As described by the SAB:

> After a numerical value is assigned to each factor, it is multiplied by a weight to obtain a factor score. Factor scores within the same category are added. Scores for the categories are multiplied together. This procedure yields a score for the pathway. The pathways are then combined through a method called quadratic averaging. [126]

Most of the fine tuning of the HRS has involved changing factors or how to obtain the numerical value assigned. The structure of the groundwater and air migration pathways are unchanged although some of the factors of the three categories have been changed. A mobility factor has been added to both pathways to account for exposure. Potential releases are now calculated for the air pathway, and the distance to sensitive environments has been increased. The surface water pathway has been substantially expanded and now consists of four threats: drinking water, human food chain, recreational, and environmental. Under food chain, bioaccumulation and fishery use are considered for the first time.[127]

A New Cutoff?--In response to congressional concerns about the cutoff score, EPA says that a study did ''indicate that **some sites** with scores below the cutoff can also pose potential dangers to human health and the environment' [emphasis added].[128] This is in contrast to EPA's stronger statement in 1984 to Congress: **"Many** of the sites that score below the 28.5 HRS cutoff still pose some threat to human health" [emphasis added],[129]

Review of State programs yields information that "some sites' is a major understatement. In **reality, many sites with scores well below the cutoff do pose current or potential dangers.** For instance, Minnesota uses the HRS to score all sites, whether for the Superfund program or its own program. The scores for 118 sites in the State program run from 2 to 51.130 Illinois also uses the HRS to score all sites; sites with scores greater than or equal to 10 gain placement on the State Remedial Action Priority List. The list currently has sites with scores from 10 to 28.16. Further, the State acknowledges that sites with scores less than 10 may ''present immediate threats' and are handled through removal actions. [131]

EPA has proposed that the revised HRS have a cutoff that is "functionally equivalent" to 28.50 because ''EPA believes that the current cutoff score

125Ibid., p. 10.

126Ibid., p. 10.

127This new data may make it easier for marine sediment sites to gain NPL status. Until such sites are scored under the new HRS, however, it is not possible to know for sure.

12853 Federal Register 51%2, Dec. 23, 1988, p. 51966.

129U.S. Environmental Protection Agency, Office of Solid Waste and Emergency Response, ''Extent of the Hazardous Release Problem and Future Funding Needs—CERCLA Section 301(a)(1)(C) Study," December 1984, p. 2-2.

130A Minnesota official told OTA that some of the sites with scores over 28.50 in the State program were municipal landfills and had not been submitted to EPA because at the time the policy was to reject such sites regardless of their scores.

131Illinois Environmental Protection Agency, Division of Land Pollution Control, *Cleaning* ///1/10 is, April 1988, p. 15.

has been a useful management tool" [emphasis added].[132] But, EPA has not determined the current HRS' false positive/negative rates, dealt with the algorithm, or evaluated how sites will fare under the new HRS. Thus, an equivalent cutoff maybe a useful management tool, but its environmental implications are not at all clear. Two of EPA's three suggested ways of determining equivalency are designed to produce an NPL of the same size that would be produced under the current HRS. The third way would attempt to produce the same level of quantitative risks for sites evaluated with the old and new HRS.

EPA has not grasped the nettle of the improbability of ever finding *a single point* (especially one with two decimal places) above which sites can be judged to present substantially more risk than those below. It may, however, be possible to design a more equitable system with two points (see OTA's option 20 in ch. 1).

Effects of the Delay-Congress specifically exempted EPA from having to reevaluate sites that have been listed on the NPL prior to the effective date of the new HRS, which Congress set at October 17, 1988. It may have been that the SARA schedule for revision was unrealistic, but the fact remains that the delay in issuing the new rules is causing problems in scheduling SIs. Further, while the method chosen by EPA to switch to the new HRS appears designed to save money, it may generate false positives.

SI data collection is dependent on the data needs of the HRS. The new HRS will require the collection of different information (and, perhaps, more information) than the current HRS does. This means that at some point EPA has to define and start new SI data collection. To avoid having a period during which no SIs and scoring packages are done, EPA has devised a phase-in of the new HRS. It is biased against low scoring sites and toward finding false positives.

EPA is assuming that a non-NPL site under the current HRS will be a non-NPL site under the new one but that some NPL-bound sites under the current HRS will be rejected by the new HRS. Sites that score below 25.0 during the transition will not be reevaluated. An exception is allowed for sites that have an element (e.g., direct contact pathway) that might allow them to score high enough for the NPL under the revised HRS. These exceptions ''should be infrequent" according to EPA.[133] However, sites that do score at least 25.0 (i.e., could be submitted for QA) are to be reevaluated using the new HRS. Thus, positive sites under the old HRS have to pass another screening and be judged positive under the new HRS, as well, to make the NPL.

Instead of just 2 years after SARA of current HRS evaluations, EPA will have had almost 4 years, if the new HRS is effective in February 1990. Between SARA's enactment and mid-1989, 458 more sites have entered the MITRE QA system and 355 sites have been proposed for the NPL using the existing HRS. The NPL updates in May and July 1989 would have been under the new HRS if EPA had finished it on schedule.[134] Because of the uncertainty of the effect of the new HRS, it is unknown whether more or fewer than 62 sites would have been proposed if the new HRS was used.

The NPL Grows

The NPL's annual growth and its eventual size depend on how thorough EPA is in discovering sites, maintaining inventories, and evaluating potential sites. So far, site discovery has been ad hoc, inventories incomplete and incompatible, and evaluation has been driven toward limiting the growth of Superfund. Still, the NPL continues to grow, proportionally increasing the responsibilities of Superfund.

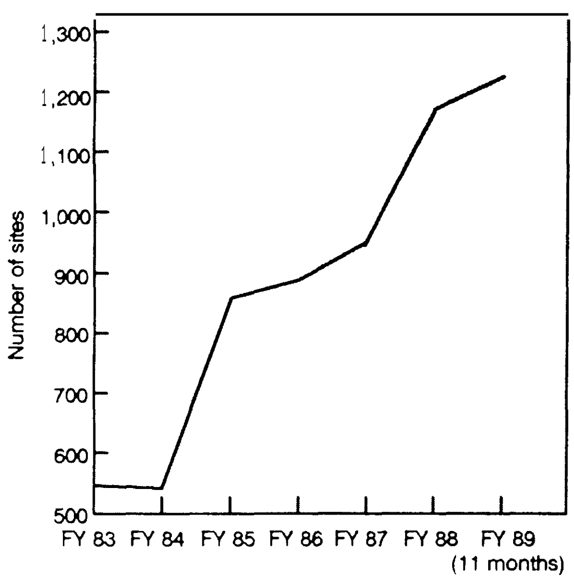
NPL growth is also affected by the rate at which EPA moves sites through the HRS scoring stage, primarily by the numbers of sites proposed (see figure 2-3). The total number of sites on the NPL at any one time is the sum of the proposed and final sites less those that have been removed or deleted, Figure 2-5 shows NPL growth from fiscal year 1983 (546 sites) through 11 months of fiscal year 1989

132 53 Federal Register 51%2, Dec. 23, 1988, p. 51966.

133 U.S. Environmental Protection Agency, ''Pre-Remedial Strategy for Implementing SARA," op. cit., footnote 57, p. 6.

134 Actually, everything is behind. The July 1989 update had been scheduled for early 1989, and a last up&uc was to happen in the summer of 1989 but, by mid-September, had not.

Figure 2-5 Growth of the National Priorities List



SOURCE: Office of Technology Assessment, 1969; baaed on EPA data.

(1,224 sites).[135] Although the growth has been uneven and sporadic, EPA has complied with the CERCLA requirement to update the list annually but did not add enough sites to meet the 1,600 to 2,000 level by January 1988 suggested by the SARA Conference report.

Another way to analyze the NPL is by the quality of the sites. EPA claims that HRS scores are a measure of relative risk among sites. EPA has also long maintained that the worst sites have been found. If so, one would expect the HRS scores to decline over time. OTA has averaged the scores of sites proposed in each fiscal year.[136] Figure 2-6 shows that the average score declined somewhat from 1983 to 1986 but from 1986 to 1989 has increased. The changes-up and down—are all within a 6-point spread that is probably insignificant for the HRS. **It is not possible to conclude-if HRS scores are a valid measure--that sites now coming through the system pose less (or more) threat than those in previous years.**

OTA also looked at the spread of HRS scores (see figure 2-6) and the distribution of scores in each fiscal year. These data suggest that the HRS scores may be approaching an equilibrium. The spread in scores (minimum to maximum score) shows a trend similar to the average scores. The spread compressed between 1983 and 1986 and then expanded between 1986 and 1988.[137] In 1989, the spread compressed again although the average is up, slightly. The distribution of scores is roughly similar for all years except the first year (1983). In 1983,44 percent of the scores were between 28.50 and 40.00. For all other years, from 60 to 73 percent of the scores were in that range. Conversely, 27 percent of the scores were greater than **50.00** in 1983; for all other years, 7 to 15 percent were greater than 50.00.

It is important to note that the changes in HRS scores averages, spread, and distribution over time do not necessarily reflect differences in the character of the sites. They can be the result of changes in policy (e.g., for a number of years EPA discouraged Regions from submitting landfill sites) or other factors. While the highest scoring sites (scores greater than 70.00) were all proposed in the first 2 years, all nine of these sites have high air subscores. As discussed earlier, 97 percent of the NPL sites processed by MITRE since 1984 have air subscores of zero. A zero air subscore may mean that air migration is not a problem or that an air score was not necessary to push a site score over 28.50.138

Estimates for the Future

Looking to the future, it appears that-if the environmental mission of the Superfund program is regained-the size of the NPL should increase substantially. OTA, GAO, and EPA data all point toward growth. **Today's CERCLIS inventory and its growth rate implies an NPL of over 4,000 sites by the year 2000. Ultimately, with a national site discovery program, minimal deferrals, and a need for secondary cleanups, over 10,000 sites could qualify for cleanup under Superfund.**

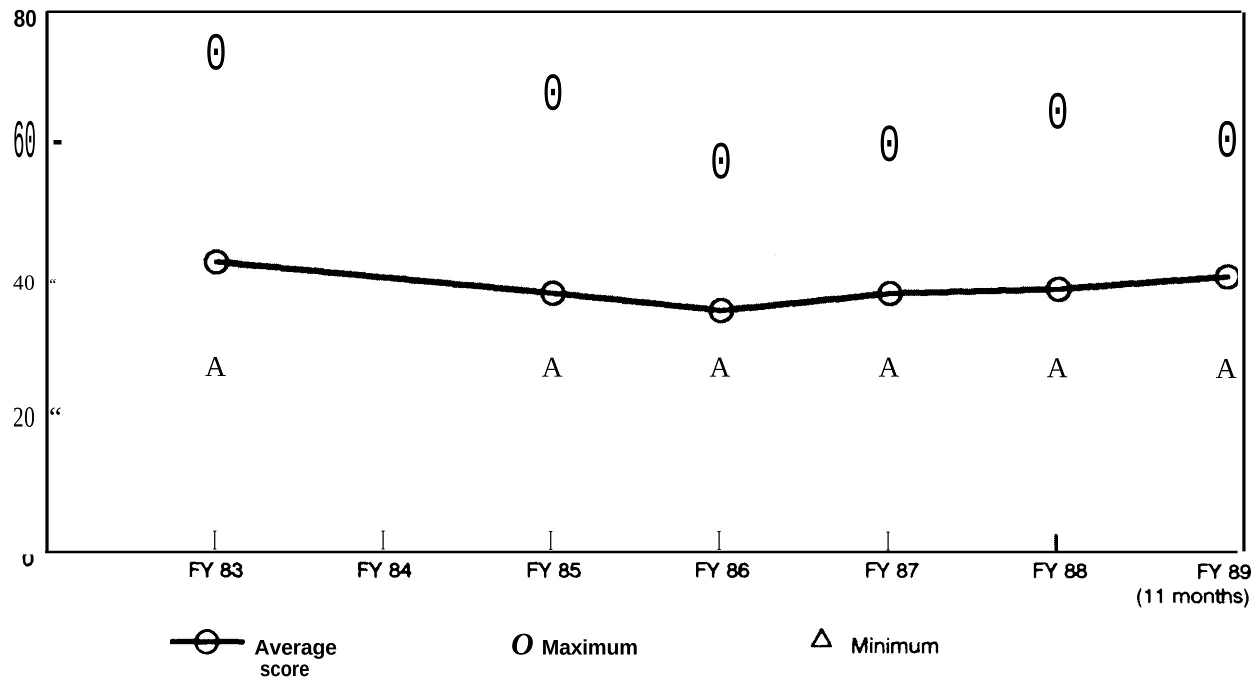
135Another count of the NPL is 1,251. That is, the number of sites currently on the NPL plus those that have been deleted. In the future, if the numbers of deletions per year increase, the growth rate of the NPL could appear to attenuate.

136The data are for 7 instead of 8 years because no sites were proposed in fiscal year 1984.

137The highest NPL score occurred in the first fiscal year of the NPL, 1983. Because of the cutoff score, the minimum always hovers around 28.50.

138Changing policies on the quality (observed or monitored) of air data needed for the HRS has affected the difficulty of obtaining air data.

Figure 2-6HRS Scores of NPL Sites (averages and minimum/maximum)



SOURCE: Office of Technology Assessment, 1989; baaed on EPA data

EPA's Projections-The Superfund program's projections do not agree with OTA's conclusions **about the future size** of the NPL. EPA makes its projections based on the choices it has made for the program. Those choices include no active site discovery, deciding that the preremedial program should make management rather than environmental decisions, and explicit or implicit deferral of clean-ups to other programs. For instance, EPA told Congress in 1988 that it was not possible to give an estimate for the future size of the NPL because "future changes in the preremedial program. . . will likely revise [the traditional] percentage" of CERCLIS sites that make the NPL.[139]

The Superfund program estimate of the NPL has always hovered around ***2,000 sites*** despite a growing CERCLIS inventory. An assessment in 1983 projected an eventual CERCLIS inventory of 22,000 sites and an NPL of 1,400. Uncertainty about types of sites to be included in the NPL in the future produced an upper bound estimate of 2,200 sites. Today, EPA says that, with an inventory of 31,000 sites, there will be 2,100 NPL sites by the year 2000.

EPA says that about 5 percent of the *evaluated sites* end up on the NPL. How EPA arrived at *historical* average of 5 percent of evaluated sites is unclear since at the same time EPA presented data showing that the rate stood at 11 percent by 1986 and 7 percent by 1989.[140]

CERCLIS and the NPL--Using the size of CERCLIS to estimate the NPL is problematic, Not only do changing site evaluation and listing policies change historical averages but there is the pipeline effect. It can take 5 years or more for a site to move from CERCLIS entry to the NPL. Additionally, CERCLIS is not *the* master list of potential sites for the Superfund program. But, most importantly, as the Superfund program's long adherence to an NPL of about 2,000 sites shows, the size of the NPL is a product of choices made about how large the NPL *should* become. Thus, a choice of whether or not to

[139] House of Representatives, "Preliminary Findings of OTA Report on Superfund," op. cit., footnote 81, p. 270.

[140] U.S. Environmental Protection Agency, "A Management Review of the Superfund Program," June 1989, p. 1-6,

conduct a site discovery program alters the size of the NPL.

OTA calculated a cumulative rate of *9* percent of CERCLIS sites making the NPL from 1983 through 1988 (using EPA data from table 2-1). OTA took the numbers of sites evaluated each year and compared it with the NPL of the following year to take into account some of the time lag between evaluation and final placement on the NPL. On a noncumulative basis, the initial rate was **22** percent in 1983 (the first NPL year) and for 1988 was 11 percent. Thus, historically-by either measure-a declining percentage of CERCLIS sites have become NPL sites, but the EPA average rate of 5 percent has not yet been encountered.[141]

OTA estimates that 10 percent of the sites on CERCLIS, or over 4,000 sites, could be on the NPL by the year 2000. The 10 percent rate assumes that the preremedial process is improved such that sites are evaluated on an environmental rather than management basis and that cleanup deferrals are minimal. Thus, if the CERCLIS could be frozen at 31,000 sites, the NPL could grow to at least 3,100 sites by the year 2000. But, a CERCLIS growing by 2,000 new sites each year will eventually contribute another 200 NPL sites per year. Taking evaluation time into account, 1,000 of those sites (or, 5 years worth) could be on the NPL by the year 2000. A national site discovery program could add several thousand more sites to the NPL.

CERCLIS may becoming an increasingly poor indicator of the potential size of the Superfund program. First, EPA plans-under the Environmental Priorities Initiative—to enter some 3,000 sites the agency clearly intends to defer to the RCRA corrective action program. That action, however, will broaden the concept of the CERCLIS inventory and move it toward being more of a national inventory. Second, as discussed earlier, the bureaucratic response to the policy to complete a site's PA within a year of its entry into CERCLIS has been to hold up site entry. The result may be a decline in the 2,000 sites per year growth rate.

Calculations based *only* on known CERCLIS/ NPL data can underestimate the future size of the Superfund program (and national cleanup needs), Despite the move to add RCRA sites, CERCLIS is not a master list. But, using combinations of lists can result in overestimates. As previous discussions show, lists of potential sites abound and no cross checks have been made for double counting either among these lists or between each list and CERCLIS.

Further, when and if cleanups fail in other cleanup programs, they may become a new source of NPL sites: secondary cleanups (see ch. 4). OTA estimates that failures from other cleanup programs could add over 1,000 sites to the NPL. This estimate assumes that, while 10 percent of CERCLIS sites become NPL sites, an additional 20 to 30 percent actually require cleanup.[142] A failure rate of only 5 to 10 percent of those cleanups, could add 410 to 1,230 sites to the Superfund program.

Ultimately, An NPL of 10,000 Sites or More?— OTA's 1985 estimate of 10,000 or more Superfund sites remains valid. It assumes that the Superfund program has an improved preremedial process, active site discovery, and minimal deferrals. In terms of the national cleanup problem, the 10,000 site projection is a major underestimate (see ch. 4).

The original OTA estimate was based on a conservative analysis of only three categories of potentiai sites. It produced a total of 8,000 sites: 1) solid waste facilities (5,000 sites), 2) groundwater problems created by RCRA Subtitle C hazardous waste facilities (1,000), and 3) an improved site analysis and selection process for the NPL (2,000). EPA's own estimate at the time of a maximum 2,000 sites, which did not include the OTA categories, was added to the 8,000 figure to arrive at a total estimate of 10,000 sites.

Since 1985, when the OTA analysis was done, the OTA categories still remain as potential problems for the Superfund program:

- For **solid waste facilities** (active and closed municipal and industrial landfills and surface

[141] It is possible to obtain a cumulative rate of 3 or 4 percent by comparing the number of NPL sites in one year with the CERCLIS sites in the same year but that ignores the evaluation and time lag effect.

[142] In 1985 OTA presented State data that an estimated 40 percent of potential sites would require a cleanup. See, OTA's *Superfund Strategy*, op. cit. Also, DOD data for fiscal year 1988 shows that 30 percent of that agency's potential sites will require a cleanup.

impoundments), there has been no comprehensive study, subsequent to OTA's, to identify problem sites. In fact, municipal landfills were actively kept off the NPL (and thus not evaluated) by EPA until external pressure caused a policy reversal in August 1987. Even so, few landfills have been added to the NPL since the policy was changed.[143] A more recent, proposed policy is to defer such sites to State Subtitle D correction action programs. Proposed Federal rules for those cleanups, however, only cover new and existing landfills; closed landfills and other types of solid waste facilities would still, presumably, qualify for the Superfund program.

- Estimates for **hazardous waste RCRA Subtitle C cleanups** range from 2,000 to 5,000 sites now, but those sites are being actively deferred by EPA to the RCRA corrective action program. Since 1983, EPA has designated about 80 sites proposed for the NPL as possible RCRA corrective action sites; about a dozen have made the final list. In June 1988 EPA proposed to officially designate 30 proposed NPL sites as RCRA Subtitle C corrective action sites and 15 as NPL sites.[144] For all future sites moving through Superfund site evaluation, EPA will decide whether or not they qualify for the RCRA program. If so, they will become RCRA rather than Superfund sites. GAO estimated in 1987 that 818 sites would fail to get cleaned up under RCRA corrective action and end up in the Superfund program.
- The **selection process for the NPL** remains the same. Site analysis has actually been adjusted such that fewer, instead of more, sites should be expected to move far enough through the process to receive NPL evaluation. OTA concluded in 1985 that if EPA paid more attention to environmental factors, more sites would end up on the NPL. Congress has required EPA to revise the HRS, but those new procedures will not be effective until February 1990, or later. As the discussion on the proposed HRS has shown, whether or not the new HRS wili improve environmental decisionmaking is unknown.

OTA has identified two new categories of sites that could add work to the Superfund program: 1) newly created sites, and 2) secondary cleanups. Illegal dumping still occurs.[145] California recently cited its Transportation Department for dumping toxic and other waste materials into a pit at a maintenance yard for 10 years. The practice only ceased in May 1989. Firms that legally operate outside of the regulatory system are also creating new sites. One example is bankrupt firms that have used hazardous substances but were not required to have a RCRA permit because they did not store, treat, or dispose of hazardous wastes and, thus, received no enforcement actions while in operation. The need for secondary cleanups will occur when impermanent cleanups done by programs other than Superfund fail.

Uncertain Future--The maximum number of *potential* sites, from which eventually come CERCLIS and then NPL sites, is approaching 500,000. Ten years ago, the maximum base number was 50,000. It was estimated by an EPA contractor who concluded that from 30,000 to 50,000 hazardous substance sites existed.[146] The report was roundly criticized at the time as an inflated estimate. Then, in 1984, EPA said in its Report to Congress that systematic investigation efforts could expand the universe of problem sites and thus increase the response needs of the Superfund program. EPA estimated those "problem sites" to be between 131,000 and 379,000 from a larger universe of known sites in five categories: RCRA Subtitle C TSD facilities, municipal landfills, industrial landfills, mining waste sites, and leaking under-

143As of May]986, 184 municipal solid waste landfills were on NPL according to the Subtitle D regulations proposed Aug. 30,1988. CERCLIS data, as of July 1988, classifies 220 NPL sites as landfills.

144Even though the 15 sites t. remain on the NPL had been on the NPL for a number of years, EPA chose to repropose them. They now wait again to become final sites.

145Data from the removal program shows an increase in 1987 over 1986 in removal actions at illegal dump sites. [U.S. Environmental Protection Agency, *Progress Toward Implementing Superfund Fiscal Year 1987, Report to Congress,* op. cit., footnote 94.] Also, New York City Environmental Police unit has a 22-member force working fulltime to seek out illegal dumps. ["Toxic Avengers, " *Discover*, August 1989.]

146Fred C. Hart, Inc., "Preliminary Assessment of Cleanup Costs for National Hazardous Waste Problem, " Feb. 19, 1979, as cited in EPA's "Extent of the Hazardous Release Problem and Future Funding Needs, CERCLA (301)(a)(l)(C) Study, ' op. cit., footnote 32, p. 1-2.

ground storage tanks.[147] EPA did not estimate how many of these sites would eventually require cleanup; only that ''some subset would require more intensive investigation, and a subset of those could require removal or remedial response by Superfund.[148]

In 1987, using EPA data, GAO recalculated the number of potential sites and arrived at a new range of 130,000 to 425,000. This group includes: RCRA Subtitle C and D facilities, mining waste sites, underground leaking storage tanks (non-petroleum), pesticide-contaminated sites, Federal facilities, radioactive releases, underground injection wells, town gas facilities, and wood preserving plants.[149] Again, no estimate was attempted of how many of these sites would actually require any cleanup. They are, however, classes of sites which currently are handled by the Superfund program. The estimate does not include classes of sites that are the exclusive purview of other cleanup programs, such as LUSTS with petroleum. Federal agency sites are included on the list. Although the Superfund trust fund is not used to pay for those cleanups, EPA incurs related costs due to its responsibility for oversight of Federal agency cleanups.[150]

OTA has updated two categories of the 1987 GAO estimates for a new upper bound of at least 439,000 potential sites. GAO's estimate for LUSTS containing hazardous substances was 10,820 in 1987; using 1988 data from the Office of USTS that estimate should be about 20,000 tanks. Federal facilities now contain over 10,000 known sites, instead of the 5.800 estimated by GAO.

If 10 percent of these *potential* sites do require cleanup, the Superfund program could be facing a total NPL of from 13,000 to 43,900 sites. If only 5 percent, then from 6,500 to 21,950 sites. These are not necessarily the worst case national scenarios because they do not account for any sites currently resigned to other cleanup programs-such as LUST—some of which could eventually become Superfund sites (see ch. 4).

Comments on the RCRA corrective action program by the General Accounting Office in a 1989 discussion paper make clear why sites in other programs may eventually have to be redone by the Superfund program:

> Preliminary indications are that over half of the 5,000 operating hazardous waste facilities are leaking and causing contamination . . the pace of cleanups has been slow. in part because there is no overall strategy to deal with the problem . . . the agency has not been able to devote sufficient resources towards its corrective action program . . . remedy selection has often been conflicting and inconsistent, with no clear criteria for selecting a remedy that is most protective of human health and/or the environment . . . The longer these problems persist and remain unresolved, the greater the likelihood that operators will be unable to take corrective action and that the facilities will become Superfund sites.[151]

147 U.S. Environmental Protection Agency, "Extent of the Hazardous Release Problem and Future Funding Needs: CERCLA Section 301(a)(1)(C) Study," op. cit., footnote 32, p. 5-3.

148 Ibid., p. 5-2.

149 GAO only included those RCRA Subtitle C facilities they felt would end up in the Superfund, rather than RCRA corrective action, program.

150 In a response t. questions posed by the House Committee on Public Works and Transportation, EPA stated in 1988 that the maximum number of Potential sites m the 1987 GAO study should be 84,000. EPA argued that the count for three categories should be decreased not because the sites might not be contaminated but bezause-bureaucratically -+hey should not be listed m CERCLIS. EPA subtracted RCRA Subtitle D facilities on the basis of a Superfund policy that had been rejected by EPA in 1987. EPA objected to GAO's count of Federal *sites* rather than *facilities* because the NPL lists Federal facilities. However, Federal agencies inventory numbers of potential sites not facilities, EPA eliminated the bulk of injection wells counted by GAO because they are "non-hazardous by definition.'' GAO justified the inclusion of these wells because of evidence that Class 5 wells have a low to high probability of being contaminated. [''Prcllmlnag Findings of OTA Report on Superfund,' op. cit., footnote 81, p. 269.]

151 U.S. Congress, General Accounting Office, ''Major Enviromental Issues: 1991-1994," discussion papers, September 1989.

APPENDIX 2A
A HEALTH CARE MODEL FOR SUPERFUND SCREENING

Introduction

Because the Superfund program does not collect the proper data, OTA cannot make a definitive analysis of the environmental effectiveness of its screening process. Analogies are possible, however, with the health care field where screening tests are routinely used to detect the presence of illness (e.g., mammography for breast cancer) or risk factors (e.g., high cholesterol levels) that may require treatment. The efficacy of medical screening tests receives a lot of attention and making improvements is often high priority research. In addition, cutoffs are set with explicit consideration of the costs of missing cases and of incorrectly labelling a healthy person as diseased. The Superfund program, which justifies cleanup decisions on protection of human health, has not sought the same high standards in its approach to screening sites and setting a cutoff.

A potential Superfund site must pass three levels of screening prior to site cleanup: 1) the preliminary assessment (PA), 2) the site inspection (SI) and HRS scoring, and 3) the remedial investigation and feasibility study (RIFS) and Record of Decision (ROD). Starting with the PA, screening is simple (only existing information is used), but at each of the two higher levels, increasingly more and better information is used. A site must be judged positive **at** each screening stage in order to enter the next screening stage and, finally, to receive a cleanup.

At each stage, some sites are eliminated and labeled NFA—No Further Action. * Elimination does not necessarily mean that a site is free from public health or environmental problems. First, because no screening process is perfect, some sites are judged negative-not requiring cleanup-when in fact they are positive. Second, sites are eliminated because it is estimated or shown that they will not score at least 28.50 using the HRS. Third, sites are eliminated because, for statutory or policy reasons, they are not covered by the Superfund program. All of these classes of rejected sites have been lumped together in the universe of sites that don't need cleanup.

Based on a health care model, two characteristics and two outcomes can be used to assess how well the three Superfund screening stages perform, both independently and collectively:[2]

- Sensitivity—What is the probability that screening will identify the *true positive sites,* i.e., those sites requiring cleanup? This is the valid hit rate of the screening method.
- **Specificity-What** is the probability that the screening method will identify the *true negative sites,* i.e., nonproblem sites that do not require cleanup? This is the valid reject rate.
- Accuracy—What is the probability that a decision made by the screening is correct? Accuracy is a dependent variable. It is affected by sensitivity, specificity, and the fraction of sites needing cleanup,
- **Precision-Is the** screening decision reproduce - ble if different people or offices examine the same site?

Embodied in sensitivity and specificity are two fundamental pitfalls: making false positive and false negative decisions. It is critical to understand that a screening test is not necessarily equal in its abilities to detect problems and nonproblems; sensitivities and specificities may differ widely.

For example, using x-ray examination (mammogram) to screen for breast cancer may have a relatively high *sensitivity* (find a high fraction of cancers) but a lower *specificity* (identify many noncancers as cancers). Although such false positives present many problems, in this case it may be better to have false positives than to have false negatives. Any positive finding can be followed up with more sophisticated testing. But a negative finding leaves the system without a second chance, and its true nature only becomes apparent if symptoms appear later when cure (remediation) is more difficult. Another form of medical examination may have greater *accuracy* because, for example, it can detect smaller size cancers. High *precision* would

[1] The term now used is NFRAP, no further remedial action planned.

[2] See, for instance, R.M. Thorner and Q.R. Remein, *Principles and Procedures in the Evaluation of Screening for Disease,* Public Health Mono@@ No. 67, May 1967,

mean the test would detect the same cancer if performed by different people at different times and under different conditions.

False Positives and False Negatives in Superfund

In Superfund, a *false positive is* a site that is selected for cleanup but really does not need one. Money is wasted and an opportunity cost may be paid because other sites do not get the cleanup attention they require in a timely way. A *false negative is* a site that is eliminated from the system even though it really needs cleanup. In this case, near-term costs are avoided but long-term costs, including environmental damages, are very likely to grow.

Improving the environmental performance of Superfund screening process means reducing false negatives, It is necessary to find out at what level, where, and why the process produces false negatives. When the problems are found, ways to solve them must also be found and then resources must be devoted to do the job. Management must *want* to evaluate and improve system performance, but current pressures are to meet numerical quotas with fairly constant budgets. There are no allowances for reassessing what has been done,

The result for Superfund is that there has been almost no critical examination of the efficiency and accuracy of screening procedures nor of alternative screening strategies. Some effort has been made at determining false positives but not false negatives.

By nature, the system responds to positive test results and not negative ones; the system has several chances to detect a false positive (because it stays in the system), but not a false negative (because it departs). But to detect a false negative (and to measure sensitivity and specificity) requires that money be spent on evaluating the successive levels of screening and, perhaps, the ultimate remediation to assess whether sites labeled as negatives and positives are really so. Clearly, no system would expend such effort on *all* findings because that would eliminate the reason for conducting screening tests, whose costs are supposed to be small relative to the final cleanup, In the health care field, research is conducted on smaller numbers of subjects in order to establish the sensitivity, specificity, accuracy, and precision of screening tests. The results are compared with some "gold standard, " usually a definitive diagnostic test or the closest thing to it. The costs and benefits of improving the efficiency and accuracy of a test are dealt with explicitly.

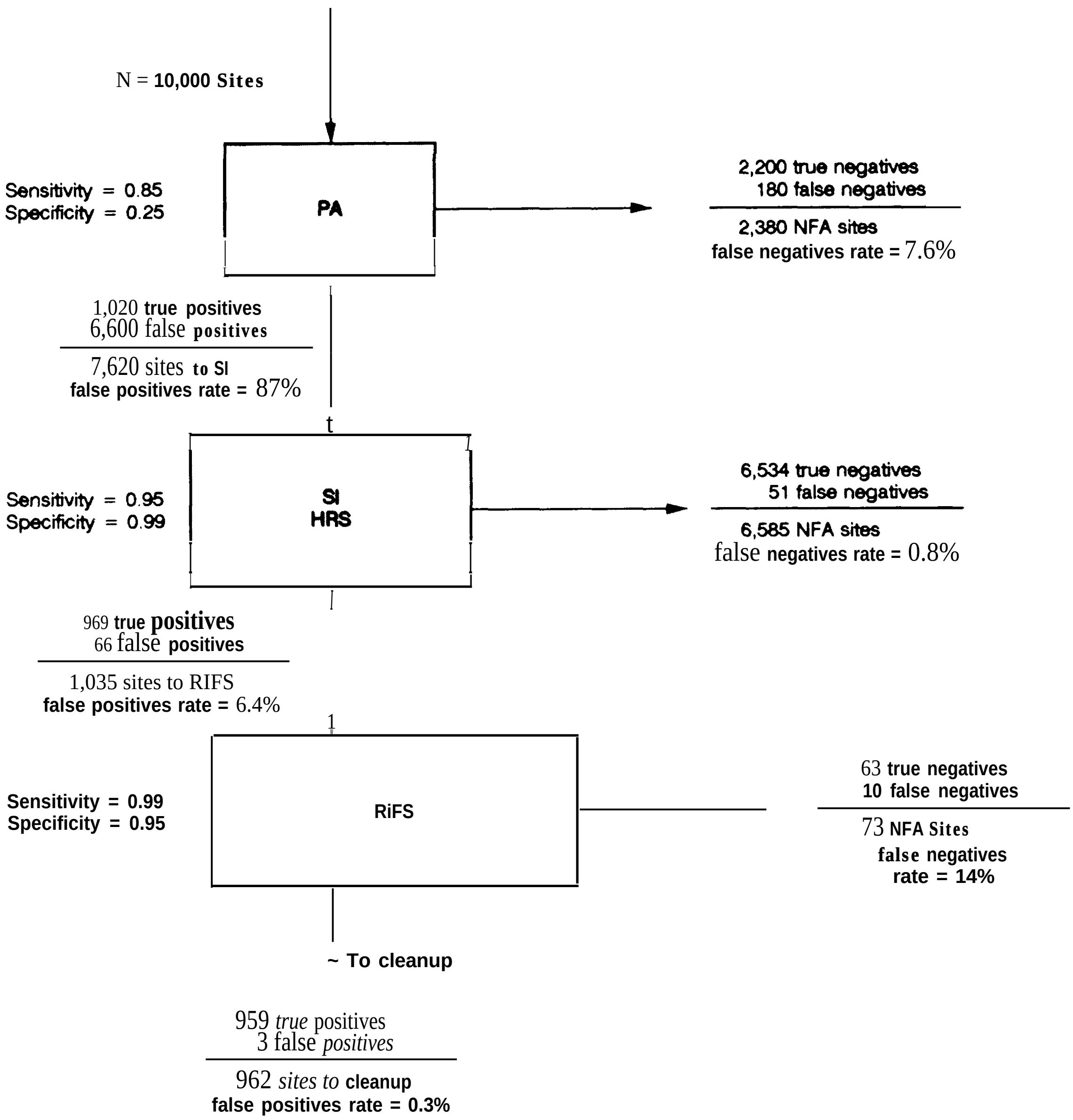
A Model of Superfund Screening

The percentage of false positives and false negatives for each screening stage can be calculated when three things are known: the true incidence rate (percentage of sites in the inventory of possible cleanup sites that actually require cleanup) and the sensitivity and specificity of each stage. Either special tests must be done or some reliable historical information must be used to obtain these figures; it is not possible to get them from current Superfund records. OTA has assumed values to gain insight into the nature of the current system, to illustrate possible problems, and to suggest strategies to solve the problems.

With the model shown in figure 2A-1, the three serial screening stages in Superfund allow for calculations-analogous to a mass balance-to track the disposition of sites as they move through the system. Positives from one stage pass on to the next, while negatives leave the system. As with any model, some details and richness of the real case are simplified or ignored. For example, OTA has combined the SI and HRS scoring into one stage, because for the most part the same information has been used for both, although in reality some sites are eliminated after the SI. Now, with two SIs (a screening and a listing SI) the level of the information is different and EPA has added another rejection point.

It is important to emphasize that there is no one correct result from the model. Numbers are assumed for key variables. For figure 2A-1, OTA assumed a CERCLIS inventory of 10,000 evaluated sites and a true incidence rate of 12 percent, or that 1,200 sites really require cleanup. We assumed that at the PA stage the *sensitivity is* good, but not exceptionally high, and resigned it a value of 0.85 (i.e., a 15 percent miss rate on the true problem sites) and the *specificity* is rather low, a value of 0.25 (i.e., 75

Figure 2A-1 -Estlmation of True/False Positives and Negatives



SOURCE: Office of Technology Assessment, 19S9.

percent of nonproblem sites pass through to the SI stage).

As figure 2A-1 shows, 7,620 sites pass the PA stage; 7.6 percent of the NFA sites are false negatives and 87 percent of the sites approved for an SI are false positives. That is, 180 NFA sites really *do* require cleanup and 6,600 sites that go on to the next stage *do not* really require cleanup.

At the SI/HRS stage the level of information is improved and the sensitivity increases to 0.95. The specificity increases to 0.99. Figure 2A-1 shows that 1,035 sites pass this second screening stage. The false negatives are 0.8 percent, and the false positives are 6.4 percent. The number of true problem sites missed (false negatives) is 51.

At the RIFS/ROD stage, the quantity and quality of information are greatly improved. OTA assumed that the sensitivity increases to 0.99 and the specificity is 0.95. The specificity was decreased somewhat to reflect the likelihood that investigators would have some reluctance to reject a site at this last level of screening after so much investment has been made in the site. Figure 2A-1 shows that 962 sites pass through the RIFS/ROD stage to actual cleanup; 14 percent of the negatives are false and 0.3 percent of the positives are false. Ten more true problem sites (false negatives) are missed.

Overall, out of the 1,200 true problem sites, 959 are detected and 241 sites are missed. The number of unnecessary cleanups is three. A total of 9,038 sites are eliminated. **It is only the number of false positives at the last screening stage that results in unnecessary cleanups. But false negatives drop out at each screening stage and accumulate.** Thus, for the three-stage process, 0.3 percent of the positives are false while 2.7 percent of the negatives are false. The overall sensitivity for detecting true problem sites is 0.80, less than for any of the individual stages. Thus, out of 1,200 true problem sites, 20 percent (241 sites) are missed because of the cumulative effect, The specificity is very high at 99.97 percent (i.e., nearly all the nonproblem sites are rejected); only 3 false positives get through the last stage.

Applying Results to Superfund

The model and the numbers assumed and calculated simulate current Superfund results. About the same size NPL is created (roughly 1,000 sites) from about the same universe of inventory sites and sites examined through the three screening stages. Available data show that the historical NFA rate at the first screening stage (the PA) has been about 20 percent (24 percent in the model) and that the NFA rate at the last screening stage is about 8 percent (7 percent in the model). Other numbers might lead to the same overall performance; therefore, the model shows several important things about the *possible* behavior of the current Superfund screening system:

- It probably does a good job of minimizing false positives; that is, very few totally unnecessary cleanups result, although responsible parties asked to pay for cleanups sometimes believe otherwise.
- It may do a poor job of minimizing false negatives; that is, a rather large number of sites that require cleanup can be missed with no indication that they exist; they are buried within a large number of true nonproblem sites. While some State programs may do a good job at detecting which are problems, not all can (see chs. **2** and 4).
- It is impossible for the second two screening stages, with their higher levels of sensitivity, to overcome or offset the inefficiency of the first screening stage where 180 out of the total of 241 false negatives are created; only 10 false negatives stem from the last screening stage.
- Most false positives come from the first stage and very few from the second two screening stages.

Ways to Minimize False Negatives

The results of the model suggest two fundamental strategies to cut down the number of false negatives and their attendant problems of high future cleanup costs and damage to human health and the environment. A High Risk Site Strategy creates a parallel track of sites and a Better Information Strategy results in a two-stage process,

High Risk Site *Strategy-A* case can be made to circumvent the three-stage screening system and its inherently lower overall sensitivity and very long evaluation time by going directly to the third stage. In the health care field a subpopulation with a higher incidence of a certain disease is identified, and this high risk groups is sent directly to a more advanced stage of screening. The key is to use preexisting information to define the subpopulation.

This strategy offers Superfund a parallel route to cleanup for some sites with a higher incidence of risk. Example subpopulations include: 1) sites that have been identified through historical aerial photographs, analysis of which clearly shows past hazardous waste management practices that lead to contamination; and 2) sites that have received emergency or other removal actions and that professionals who have worked onsite believe need a cleanup.

A high percentage of such sites are apt to require cleanup, perhaps 70 to over 90 percent. They could be **skipped directly** to the third screening stage **where the very** high sensitivity would confirm almost all of them. For example, for 1,600 such sites with an incidence rate of 75 percent (i.e., 1,200 true problem sites, the same as in figure 2A-l), 1,188 sites would be correctly detected, and 12 sites would be missed. This strategy results in a 1 percent miss rate rather than the 20 percent miss rate for the current system.

Better Information Strategy--lt is conceivable that the first screening stage, where most false negatives are created, can be improved to raise its sensitivity. This approach runs counter to the basis of the current three-stage screening process which is simple and low-cost (typically days or weeks and a few thousand dollars) at the first stage and gets much more costly and longer (typically a few years and several hundred thousand dollars) at the third stage. Improving the first screening stage means spending more money on a very large number of sites—an activity that could make subsequent stages redundant. This redundancy seems to be happening now; for example, pushing the use of the HRS into the PA stage. But, the current use of the HRS at the **PA stage does** not include using better information.

Making a significant improvement in PA screening so that its sensitivity increases is worth considering. For example, if its sensitivity is increased from 0.85 to 0.95, the PA and SI/HRS stages can be combined into a new, more efficient first stage with the benefits of the current second stages That is, this strategy uses the second and third stages of the original model. The result is that, after the third stage, 71 true problem sites are missed and there are 4 false positives in the total of 1,133 going to cleanup.

Instead of missing 241 sites in the current system and 12 sites in the High Risk Site strategy, 71 sites are missed in the Better Information strategy. The sensitivity for the Better Information strategy is 94 percent (as compared to 99 percent in the High Risk Site strategy and 80 percent in the current system). However, the basis for comparison is not quite the same for all three cases, even though there are 1,200 true problem sites in each. Although the current system and the Better Information strategy could be used on the same group of randomly selected sites within the entire Superfund inventory, the High Risk Site strategy is used on a selected group of sites for which it is known that the percentage of sites needing cleanup is high. The advantage of the High Risk Site strategy, therefore, depends on having information which reliably predicts that a site is a true problem requiring cleanup.

Comparison With EPA Attempts To Improve Screening

One of the paradoxes of Superfund is that the PA, the first screening stage traditionally gets the least attention, uses the least and probably worst information, and probably is implemented by the most junior, inexperienced people. This condition means that the sensitivity and specificity are probably low, as reflected in OTA's modeling of the current system above. EPA's interest (as detailed in ch. 2) is to reduce the workload of the SI/HRS, the second, more expensive screening stage. Thus the drive is to improve the specificity (i.e., finding nonproblem sites) of the PA rather than the sensitivity (i.e., finding problem sites). But, it is the sensitivity that determines how many false negatives are created.

The model helps to assess the effects of EPA's expansion of PA screening to reduce the workload at the SI/HRS stage; that is, to reduce as early as possible false positives. The subtlety is that it is possible to increase the specificity without increasing the sensitivity; that is, by increasing the specificity the number of false positives is decreased, but not the number of false negatives. If something is done to better detect true nonproblem sites, such as using crude field sampling and analysis to show no contamination, a specificity of 0.75 instead of 0.25 (as in the current system model) for the first screening step could be assumed. This reduces the number of false positives from 6,600 to 2,200. The number of sites going on to the SI/HRS stage-the workload-decreases from 7,620 to 3,220 sites, a 58 percent reduction. The false positive rate drops from 87 to 68 percent. The false negative rate decreases from 7.6 to 2.7 percent even though the actual

[3] This is similar to the suggestion in ch. 2 that screening be a continuum rather than one in which at a specific point, say the PA stage, it is assumed that enough information exists to make a site decision.

number of false negative sites remains constant (because a larger NFA base is created).

However, the real danger is that whatever action is used to increase the specificity also affects the screening's sensitivity. In fact, the likelihood is very great that the sensitivity would decrease, thereby increasing the rate of false negatives. For example. in a modified PA, using crude sampling and analysis with high detection limits and very few samples per acre could miss major contamination. Similarly, as is currently the situation, application of the HRS scoring at a time when very little reliable information exists could reduce the sensitivity as much as the specificity is increased.

Reducing the sensitivity, such that more false negatives are created, has an apparent effect of also increasing the number of sites eliminated and reducing the workload for the next screening level. If the sensitivity decreases from 0.85 to 0.80 at the PA stage, then an additional 60 false negatives are created (240 as compared to 180 in the current system).

Accuracy in Superfund Screening

The degree to which screening decisions are accurate is yet another problem. In Superfund, the increasingly critical factor is the use of the HRS. Just as in some medical tests, a cutoff score is used to decide whether a site is "bad" enough to merit cleanup. As discussed in chapter 2, the HRS has been criticized for many years by many people. No available evidence has established a valid relationship between the score—and certainly not the arbitrary cutoff value of 28.50--and the actual threat to human health and environment.

It is not clear that there is *any one* cutoff score that would accurately indicate that a significant environmental threat exists or does not exist. If there is one point of uncertainty below which no cleanup problem probably exists and another point above which there is almost certainly a need for cleanup, these points among HRS scores has not been determined. If such points were determined, then the policy decision would be whether or nor to consider sites between the points as positives or negatives.

Precision in Superfund Screening

The higher the precision of screening tests, the greater their reproducibility. No matter who applied the screen, or when, the result would be the same. In Superfund, good precision would mean that it would not make any difference, with regard to whether a site is judged to require cleanup, what EPA region a site is in, nor what contractor or State office performed the work, nor which people did the work, nor when the work was done. Unfortunately, no specific attempts (such as having some sites evaluated by different offices) have been made to evaluate the precision of the three-stage screening process and, especially, the HRS. What data exist, however, suggest a substantial level of inconsistent results in every aspect of Superfund implementation (see chs. 2 and 3). Thus, the probability is high that a site which is judged to be positive or negative could receive just the opposite label if it was examined at a different time by different people in a different office.

Conclusion

The examination of sensitivity and specificity in combination with an overview of problems with accuracy and precision in Superfund's screening steps leads to an unsettling conclusion. No detailed data have ever been intentionally gathered that could rate the worth of Superfund's screening steps. However, whatever analysis can be done with the meager information available leads one to suspect that current screening efforts may miss substantial numbers of sites that really require cleanup. But the system is much less likely to result in cleanups that are really unnecessary.

In fact, for every one unnecessary cleanup, the model used here suggests 80 sites are not cleaned up that should be. The costs of better or extra screening designed to minimize false negative sites seem small compared to the higher costs of delayed cleanup. Doubling preremedial costs, for instance, to 6 percent of the Superfund budget would add about $40 million to screening. That is comparable to today's average cost of cleaning up a site, *$30* million. If improved screening found most of the 80 sites, it is possible that hundreds of millions in future clean up costs could be saved.

When the problems of poor accuracy and precision are also taken into account, then it is clear that the margin of error in any estimate of false negatives and false positives is probably very large, even as much as plus or minus 100 percent. This means that if all errors were random there may be no significant problem of false negatives or the rate might be twice as large as the estimates made with the model used here. On the other hand, errors may be systematic, not random. Indeed, some actions have been discussed that bias results, and the pressures on Superfund point to a predilection to ignore false negatives while attempting to minimize false positives. Short-term costs are being minimized at the expense of higher long-term costs.