An Agricultural Law Research Article

Numbers, Numbers Everywhere—And Not a Drop of Meaning

by

Allan S. Felsot

www.NationalAgLawCenter.org

ALLAN S. FELSOT*

Numbers, Numbers Everywhere—And Not a Drop of Meaning

J.A. Paulos, Temple University mathematics professor and presidential scholar, observed that many people seem to suffer from innumeracy, or "the inability to deal comfortably with the fundamental notions of number and chance."[1] Paulos' basic premise was that misperception of the magnitude of numbers and ignorance of the mathematics of probability leads to poor public policy and a tendency to believe in pseudosciences like astrology and parapsychology.[2] The inability to grasp the magnitude of numbers is very relevant to the prevailing attitude regarding chemicals in the environment and their potential health effects. Just as we talk about multi-billions of dollars for a seemingly worthy government program without blinking an eye, so we alarmingly focus on parts per billion (ppb) of a synthetic chemical substance without a conceptualization of just how small that amount is. The consequence for our society is that we are too quick to spend billions of dollars and equally as quick to condemn a part per billion of a synthetic chemical without rationally understanding the costs or benefits associated with either number. The latter misunderstanding has led to exceedingly lower and lower standards for chemicals, especially pesticides, in water and food.

A basic tenet of toxicology is that adverse effects of a chemical are directly related to its dose.[3] Thus, to properly assess a chemi-

---

* Professor, Food and Environmental Quality Laboratory, Washington State University, Richland, WA.

[1] JOHN ALLEN PAULOS, INNUMERACY: MATHEMATICAL ILLITERACY AND ITS CONSEQUENCES 3 (1988).

[2] Id. at 4.

[3] See TOXICOLOGY, THE BASIC SCIENCE OF POISONS 3-10 (Louis Casarett & John Doull eds., Macmillan 1975). See generally WAYLAND J. HAYES, JR., INTRODUCTION TO HANDBOOK OF PESTICIDE TOXICOLOGY 1-37 (Wayland J. Hayes, Jr. & Edward R. Laws, Jr., eds., 1991).

cal's hazard, we need to know its concentration or amount in the environment. Forty-five years ago we had the capability of detecting pesticides in soil and plants at levels of parts per million (ppm); few laboratories could detect anything at levels a thousand-fold lower, or parts per billion.[4] Since then, analytical instrumentation has undergone an evolution in capability that now allows routine detection of not only ppb, but increasingly allows detection of parts per trillion (ppt), and with some sophisticated work, parts per quadrillion (ppq).

As a result of engineering ingenuity permitting detection of nearly inconceivably small amounts, analytical chemists give policy makers reams of numbers. Most policy makers, having never experienced firsthand the art of analysis, treat the numbers as if they are concrete realities, when in fact the numbers represent probabilities. Coupled with a biological understanding that lags behind our ability to measure minuscule concentrations of many chemicals, regulatory standards have begun to overreach the true significance of the numbers, rendering them worthless pawns in an attempt to set an arbitrary social, economic, and political objective.[5] Thus, policy becomes uncoupled from science.

---

[4] This statement is based on a critical analysis of the origins of environmental toxicology that was undertaken by reviewing literature dating back to 1908; the analysis was published as Allan S. Felsot, *Early Contributions of Insect Toxicology to the Evolution of Environmental Toxicology*, 33 ILL. NAT. HIST. SURV. BULL. 199-218 (1985).

[5] The regulatory standards for a class of chlorinated compounds known as dioxins serve as an example of standards that go beyond the ability to understand hazard to humans by virtue of the incredibly low level at which they are set. Dioxins actually comprise seventy-two different compounds, called congeners, each having its own level of toxicity. The regulatory standard for residential soil in Washington State under the Model Toxics Control Act is based on the sum of all the dioxins, but each dioxin's concentration is normalized to the most toxic of all dioxins, TCDD (tetrachloro-para-dibenzodioxin). TCDD, also generically called "dioxin," gained notoriety during the Vietnam War because it was an unintentional contaminant of the herbicide formulation Agent Orange that the U.S. Air Force indiscriminately sprayed over millions of acres. The expression used for the sum of all the dioxins is termed the toxic equivalency (TEQ). *See generally* K.C. Jones & A.P. Stewart, *Dioxins and Furans in Sewage Sludges: A Review of Their Occurrence and Sources in Sludge and of their Environmental Fate, Behavior, and Significance in Sludge-Amended Agricultural Systems*, 27 CRITICAL REVIEWS IN ENVTL. SCI. & TECH. 1, 6 (1997). The problem with standards based on the TEQ is that numerous dioxins are now known to occur naturally as a result of biomass burning (for example, from burning wood). Thus, it is likely that soils in areas of historical forest fires may already be contaminated with dioxins that add to the TEQ. The fact that dioxins are now known to occur naturally raises questions about their biological significance at very low levels. However, there is no doubt that the TCDD isomer at sufficiently high levels can cause a human skin disease known as chloracne. Open to debate is

Science is actually a process of discovering the physical, chemical, and biological aspects of our universe by measurement. The scientific process recognizes the uncertainty or limitation of our measurements to a complete understanding of the phenomena we wish to know. Yet, a democratic society has the right to set certain performance or behavior standards, and so depends on the numbers, i.e., scientific measurements, as guideposts. But if the guideposts are false, like a road sign pointing in the wrong direction, then policy is doomed to fail its objectives.

A bridge of numerate understanding (i.e., numeracy) is needed between the scientists who create the numbers through their measurements and the policy makers who depend on these numbers for social and political objectives. The foundation of such understanding rests on two premises. First, as we push regulatory standards to lower and lower levels, measurement error increases to the point of making inconclusive any decisions about whether the standards have been met. Second, mere capability to detect a substance is not equivalent to a knowledge of the biological hazard, if any, of the substance. Regarding this latter premise, the fact that a relationship between dose and biological response has served as an initial basis for creating a regulatory standard itself indicates that there are amounts of substances below which hazard is likely nonexistent. Thus, the second premise is implicit in any regulatory standard. Less obvious is the problem of error in analytical measurement and how that might affect decision making.

My objective is to explore the consequences of increasingly sensitive analytical capabilities to our perception of chemical contamination and our desire to regulate it. First, I will place the numbers associated with our measurements into a perspective of magnitude that will hopefully stimulate thinking about the difficulties facing the analytical chemist or scientist who wishes to decipher the biological significance of chemical concentrations in the environment. Second, I will illustrate consequences of increased analytical capability on our perceptions of contamination, and argue that our perceptions may be forcing impractical expectations of management. Third, to understand the problems of the analytical process itself, I will provide an overview to explode the myth of "the black box"—the simplistic perception of

how much the notoriety of dioxins has influenced the desire to regulate it at very low levels.

environmental sample in and useable information out. Finally, I will show that residue numbers, i.e., concentrations of chemicals in environmental media, are not absolute entities, but estimates of reality based on repeated measurements. For that reason, I choose to consider the numbers representing our measurements as virtual realities.

I

EXPRESSING RESIDUE NUMBERS

The contaminants detected in the physical environment and biological tissues are called residues.[6] The term residue is most commonly associated with pesticides. The pesticide itself is the active ingredient of a formulation that may be poured into a tank and then sprayed onto a field.[7] Once the pesticide leaves the sprayer, it becomes a residue.[8] Likewise, industrial chemicals residing in locations remote from places of their intended use can also be considered residues.[9] Thus, polychlorinated biphenyls (PCBs), common insulating fluids in electrical transformers, become PCB residues once the transformers leak and fluid moves into soil or water.

The amount of a contaminant in any environment is expressed as a residue concentration. Concentrations are represented as the mass (i.e., weight) of contaminant in a specific unit volume or mass of soil, water, air, or biological tissue (i.e., units of media). For contaminants dispersed in the environment, as opposed to being concentrated as waste in a landfill, the masses are so small that they are most easily expressed in metric units as fractions of a gram. A gram itself is approximately 0.04 (four hundredths) of an ounce. For example, mass units commonly used for contaminants are milligrams (one thousandth of a gram), micrograms (one millionth of a gram), or nanograms (one billionth of a gram).

---

[6] The term residue has historically been used to describe the intact chemical or its associated transformation (i.e., degradation) products in the environment, including all organisms. As such, residues are considered contaminants. *See Glossary* to 68 PURE & APPLIED CHEMISTRY 1167, 1173 (1996); John M. Geary, *Introduction* to 1 PESTICIDES MONITORING J. 1 (1967); *Preface* to 65 RESIDUE REVIEWS at vii (Francis A. Gunther ed., 1976).

[7] *See generally Glossary, supra* note 6, at 1184 (1996).

[8] *See generally id.* at 1186.

[9] *See id.* at 1173 (defining contamination).

The volumes of media in which contaminants reside can be expressed as liters (one liter is approximately one quart), and the masses can be expressed as either grams or kilograms (one thousand grams, or approximately 2.2 pounds). The concentrations of pesticides routinely found in water, for example, are in the range of tens of nanograms per liter (e.g., 10 ng/L) and occasionally micrograms per liter (μg/L).[10] In soil, immediately after application of a pesticide, concentrations of 1 mg per kilogram (1 mg/kg) or higher are common.[11] For air, concentrations are expressed as nanograms or micrograms per cubic meter, the cube of linear measurements comprising a volumetric measurement. The concentrations of chemicals in an organism are expressed as the mass per mass of body weight or tissue weight. Thus, studies commonly report concentrations of DDT in humans as milligrams per kilogram (mg/kg) of fat.[12] Note that 1 μg/g (one microgram per gram) represents the same concentration as 1 mg/kg.

For convenience, any concentration expressed as a mass per unit volume or weight can also be expressed as a proportion. The most familiar expression of proportion is percent, or parts per hundred (1% is one part per 100 parts, commonly expressed mathematically as the base). Common expressions for concentrations as proportions are parts per million (ppm), parts per billion (ppb), and parts per trillion (ppt). Any one proportion can be translated to another by multiplying or dividing by one thousand. For example, 1 ppb is 1000 ppt.

When using the proportional expressions of concentration, one has to specify whether the base units are volumes (milliliters or liters) or weights (i.e., grams or kilograms). Concentrations in water corresponding to ppt, ppb, and ppm are 1 ng/L, 1 μg/L, and 1 mg/L, respectively. Because one kilogram of water occupies a volume of one liter (i.e., the density of water is 1 kg/L), soil residue concentrations expressed as ppt, ppb, and ppm are equivalent to 1 ng/kg, 1 μg/kg, and 1 mg/kg, respectively.

---

[10] JACK E. BARBASH & ELIZABETH A. RESEK, PESTICIDES IN GROUND WATER: DISTRIBUTION, TRENDS AND GOVERNING FACTORS 257, 258, 360-65 (1996). *See generally* S.J. LARSON, ET AL., PESTICIDES IN SURFACE WATERS: DISTRIBUTION, TRENDS AND GOVERNING FACTORS 27-112 (1997).

[11] Knowing the rate of pesticide application and the depth of soil being sampled, the actual concentration can be calculated. For example, applying 3 pounds of pesticide uniformly over an acre will result in a concentration of 1 mg/kg in the top three inches of most soils (depending on the soil density).

[12] *See, e.g.*, A. Bevenue, *The "Bioconcentration" Aspects of DDT in the Environment*, 61 RESIDUE REVIEWS 37-111 (1976).

II

THE MAGNITUDE OF SMALLNESS

The concentrations of medicines that are therapeutic and the concentrations of toxicants that are harmful are most frequently large enough to be expressed as tens, hundreds, or even thousands of milligrams per kilogram of body weight. For example, a body dose of 192 ppm (i.e., 192 mg per kg of body weight) of caffeine has proven toxic to rats.[13] A number of herbicides and fungicides are actually less toxic than caffeine,[14] but some chemicals like parathion, now banned in the United States, can be toxic at concentrations of 10 ppm (i.e., a dose of less than 10 mg per kg body weight).[15]

Contaminant residues in air, soil, and water, however, are present in concentrations at thousands, millions, and even billions of times less than the known toxic concentrations expressed on a body weight basis. For example, DDT related residues are found in some water bodies at concentrations approaching 10 ppt (i.e., 10 nanograms per liter of water) or less.[16] To put this concentration in perspective to its hazard, consider how much water an individual would have to consume to cause a dose or body concentration that is likely harmful. The Environmental Protection Agency (EPA) has defined the amount of DDT residues that can be consumed daily for a seventy year lifespan without any appreciable risk (from any toxicological endpoint, including cancer) as 0.5 µg DDT per kg body weight.[17] Thus, a 10 kg child consuming water contaminated with 10 ppt of DDT would have to consume 500 L of water (132 gallons) daily just to receive a dose that leads to increased risk for adverse health effects. Even

---

[13] LYNN E. BUTLER ET AL., 1 HANDBOOK OF PESTICIDE TOXICOLOGY 49 tbl.2.2 (Wayland J. Hayes, Jr. & Edward R. Laws, Jr. eds., 1991).

[14] For example, the herbicide glyphosate, which is the active ingredient in the popular formulation marketed as "Round-up", is toxic to rats at doses of 5600 mg/kg body weight. LYNN E. BUTLER ET AL., 3 HANDBOOK OF PESTICIDE TOXICOLOGY 1340 (Wayland J. Hayes, Jr. & Edward R. Laws, Jr. eds., 1991).

[15] LYNN E. BUTLER ET AL., 2 HANDBOOK OF PESTICIDE TOXICOLOGY 1042 (Wayland J. Hayes, Jr. & Edward R. Laws, Jr. eds., 1991).

[16] JOSEPH F. RINELLA ET AL., PERSISTENCE OF THE DDT PESTICIDE IN THE YAKIMA RIVER BASIN WASHINGTON 10-11 (United States Geological Survey Circular 1090, 1993); D. DAVIS, WASHINGTON STATE PESTICIDE MONITORING PROGRAM 52 (United States Geological Survey 1998); WASHINGTON DEPARTMENT OF ECOLOGY, 1996 SURFACE WATER SAMPLING REPORT 16 & apps. (PUB. NO. 98-305).

[17] U.S. EPA Integrated Risk Information System (IRIS) Database (visited Nov. 9, 1998) <http://www.epa.gov/ngispgm3/iris/subst/0147.htm#I.A>.

at concentrations of DDT in water of 1 ppb, an extraordinarily high concentration rarely found today in water, a 10 kg child would still have to consume 1.3 gallons a day to cross the threshold of appreciable risk.

To emphasize further the smallness of contaminant residue concentrations, we can translate the contamination level of water into a percentage purity. If a glass of water contained only 10 ppt of DDT and nothing else, then the purity of the water would be 99.999999999%. Such purity levels cannot be achieved without great expense. Indeed, older studies of the ability of municipal water treatment to clean DDT contaminated river waters before distribution into a public supply showed that such levels of cleanup were impossible to achieve, even after carbon filtration.[18]

In a democratic society, a choice can be made to manage 1 ppt of DDT in water as if it is still too hazardous. Indeed, risk management, being distinct from the scientific process called risk assessment, is not necessarily based on scientific input, and is more likely to be propelled by social, economic, and political concerns. So, let us suppose that regulators now decide that a thousand-fold lower concentration of DDT, or 1 part per quadrillion (1 ppq or one trillionth of a gram, or picogram, per liter) assures us of absolute safety. While 1 ppq may seem reassuringly minuscule, perhaps knowing the number of molecules that this concentration represents would change this perception. For any substance, the number of molecules in a liter volume containing the equivalent of its molecular weight in grams is equal to 6.023×10^{23} (that is equivalent to the number 6.023 with twenty zeroes added after the three). By extrapolation, 1 ppq would still contain over one billion molecules of DDT (1×10^{9}). Thus, from the perspective of numbers of molecules, even lowering the regulatory standard a thousand-fold still permits an incredibly large number of molecules. Current analytical equipment would routinely fail to detect one billion molecules of DDT per liter. None of the preceding argument should be misconstrued as approving of the use of DDT, nor providing an excuse for not instituting the best available technology for protecting water. However, the magnitude of smallness of the concentrations of common contaminants today and the virtual lack of biological relevance must

---

[18] *See* John J. Richards et al., *Residues in Water: Analysis of Various Iowa Waters for Selected Pesticides: Atrazine, DDE, and Dieldrin—1974*, 9 PESTICIDES MONITORING J. 117 (1975).

be appreciated prior to attempting risk management. Failure to consider this perspective, coupled with lack of appreciation for our incredible powers to detect substances, has skewed our risk perception and substantially affected risk management.

III

THE CONSEQUENCES OF ANALYTICAL CAPABILITY

First, we now know that pesticides and other contaminants occur in places where we had never seen them before. For example, many scientists were somewhat surprised when the pesticide aldicarb was found extensively in groundwater in several locations of the U.S. during the late 1970's. The EPA had not emphasized ground water monitoring in agricultural regions prior to 1979.[19] Furthermore, aldicarb was commonly thought to break down so quickly that its residues would be sufficiently low to be undetectable. An earlier published report of atrazine in shallow ground water wells along the Des Moines alluvial plain was overlooked.[20] These reports from Iowa were nearly unnoticed harbingers of the plethora of pollution studies to pour from university and government laboratories during the next two decades.

Based on current concepts of environmental chemistry, soil physics, and hydrology (water flow through soil), certain pesticides had probably always been transported to groundwater within short times after their first use. But now, the residues of the pesticides are easily detectable because our instrumentation has advanced. The frequency of atrazine detections in ground water has been shown to be inversely correlated to the analytical reporting limit.[21] The reporting limit is the lowest concentration of chemical that the analytical chemist is willing to report with qualitative and quantitative confidence of its accuracy. When the reporting limits were 1 ppb, only 1-2% of sampled midwestern wells were reported to have atrazine.[22] When the reporting limit was lowered to 0.003 ppb (equivalent to 3 ppt), frequency of atrazine detections increased to 46%.[23]

---

[19] *See* S.Z. Cohen et al., *Potential for Pesticide Contamination of Ground Water from Agricultural Uses*, in 259 TREATMENT AND DISPOSAL OF PESTICIDE WASTE, 297, 297-325 (Raymond F. Krueger & James N. Seiber eds., 1984).

[20] Richards, *supra* note 18, at 117-23.

[21] D.W. Kolpin et al., *Pesticides in Near-Surface Aquifers: An Assessment Using Highly Sensitive Analytical Methods and Tritium*, 24 ENVTL. QUALITY 1125 (1995).

[22] *Id.*

[23] *Id.*

The ability to detect contaminant residues in places we did not suspect leads to a second consequence of our analytical ingenuity. We have developed the notion that synthetic chemicals are now everywhere in our environment. Compared to the abundance and distribution of all chemicals, natural and synthetic, however, the amounts of synthetic chemicals are minuscule. Indeed, of the numerous chemicals that can be detected in natural waters, only small percentages are contaminants we are familiar with and regulate. The vast majority, estimated as at least 85% of organic chemicals present,[24] remain unidentified and probably of natural origin. But simply reporting more contaminant detections does not change any hazards. For example, the biological significance of the atrazine detections described previously is obscure.[25] The quantities of all the residues detected were below currently established regulatory standards for protection of drinking water.[26] As argued before, the pesticides were likely always there, apparently at levels assessed by regulators as tolerable, but the perception of the hazard changed. While the perception of a world highly contaminated with synthetic chemicals is understandable, toxicological principles argue against equating the mere presence (or detection) of a chemical with having a biological effect.

A third consequence of our ability to detect ever smaller concentrations is a tendency to lower regulatory limits to increasingly unrealistic levels, especially in consideration of the uncertainty of measurements of very small concentrations. In some cases, regulatory criteria are set at the very limit of detection, which is defined as the lowest concentration in a sample matrix that can be determined to be statistically different from the sample matrix without the chemical.[27] Analytical chemists are horrified by this trend because they know that the error of false detections (both positive and negative) is extremely high at such levels. A similar problem arose when regulations like the

[24] H.F. Kraybill, *Assessment of Human Exposure and Health Risk to Environmental Contaminants in the Atmosphere and Water with Special Reference to Cancer*, J. ENVTL. SCIENCE & HEALTH, pt. C, 175, 175-232 (1983).

[25] Richards, *supra* note 18, at 117-23.

[26] The regulatory standard for drinking water is called the Maximum Contaminant Level (MCL), and is set in the Safe Drinking Water Act. For atrazine, the MCL is 3 ppb. *See* UNITED STATES EPA, DRINKING WATER HEALTH ADVISORIES 43-67 (1989).

[27] Lloyd A. Currie, *Detection: Overview of Historical, Societal, and Technical Issues*, *in* DETECTION IN ANALYTICAL CHEMISTRY 1, 17 (Lloyd A. Currie ed., 1988).

former Delaney Clause of the Food Additives Amendment to the Federal Food, Drug, and Cosmetic Act mandated that no residues could be present in a processed food commodity.[28] From the viewpoint of analytical chemistry, such a demand created a condition that was technically impossible to meet.

Yet, regulators forge ahead with making policy, seemingly incognizant of the increased probability for mistakes at concentrations near the limits of detectability. With renewal of the Clean Water Act and its system of issuing National Pollution Discharge Elimination System Permits, policy makers lowered water quality based effluent limitations (WQBELs) to levels that analytical measurements would have trouble detecting with the requisite statistical confidence.[29] EPA had to issue a guidance policy for permitting, monitoring, and enforcing WQBELs when set below the analytical detection and quantitation levels.[30] The bravura of the policy and EPA's attitude spoke volumes: "The problem is that science has not kept pace with policy in this area."[31] Ever optimistic, the EPA guidance document also stated that "[i]n the future, as analytical chemistry improves, the . . . MDLs [Method Detection Limits] will become more sensitive and approach the lower WQBEL."[32] Whether the WQBELs themselves were based on science is open to debate, but to set legally enforceable standards below competent measurable levels hardly seems scientific or logical. How can one ascribe damage to a contaminant when one cannot measure the contaminant?

IV

DETECTION AND QUANTITATION ARE EASIER SAID THAN DONE

There is something macho about being able to detect a ppt of a pesticide in water. Regardless of the thrill, analytical chemists today are mandated to achieve ultra low detection levels by virtue of prevailing regulatory standards. For example, the aquatic

---

[28] The Food Additives Amendment of 1958, 21 U.S.C. § 348(c)(3)(A) (1998).

[29] J.F. Ryan, *Method Limbo—How Low Can We Go?*, 6 TODAY'S CHEMIST AT WORK, Issue 3, at 38-48 (1997).

[30] U.S. ENVIRONMENTAL PROTECTION AGENCY, NATIONAL GUIDANCE FOR THE PERMITTING, MONITORING, AND ENFORCEMENT OF WATER QUALITY-BASED EFFLUENT LIMITATIONS SET BELOW ANALYTICAL DETECTION/QUANTITATION LEVELS (unpublished draft) (Mar. 22, 1994).

[31] *Id.* at 1.

[32] *Id.* at 6.

ecological criterion or guideline for DDT levels in water has been set at 1 ppt.[33] Yet, the analytical capability to reach this level is barely sufficient. Laboratories strain the credulity of an analysis to keep up with ever lower regulatory standards. For example, the ability to detect pesticides at concentrations suggested by the EPA Ambient Water Quality Criteria for Protection of Aquatic Organisms differs widely among government laboratories. For the pesticides azinphos-methyl, chlorpyrifos, diazinon, and DDT, the EPA criteria are 10, 41, 9, and 1 ppt, respectively.[34] The reporting limits for the USGS laboratory of the NAWQAP (National Ambient Water Quality Assessment Program) were 1, 4, 2, and 1 ppt for the four pesticides.[35] A USGS research lab in Denver, Colorado published a paper listing the reporting limits for conventional analysis of chlorpyrifos, diazinon, and DDT as 47, 45, and 35 ppt.[36] The Washington Department of Ecology, in its publication of surface water quality monitoring data for 1994, listed reporting limits based on quantitation as 160, 60, 70 and 50 ppt for azinphos-methyl, chlorpyrifos, diazinon, and DDT, respectively.[37]

Because regulatory standards and guidelines are set near the analytical limits of detection, we must understand the pitfalls of pushing risk management beyond the limits of what our analytical technology can provide. These problems become clearer in a discussion of what happens during analysis.

A. Sampling Problems

When we try to measure the amount of something in air, water, soil, or organisms, we can only sample or collect a small piece of it. Thus, what we collect and measure is only a represen-

---

[33] *See generally* L.H. Nowell & Elizabeth A. Resek, *National Standards and Guidelines for Pesticides in Water, Sediment, and Aquatic Organisms: Application to Water-Quality Assessments*, 140 REV. OF ENVTL. CONTAMINATION & TOXICOLOGY 1 (1994).

[34] *Id.* at 77-93.

[35] *See* R.J. WAGNER ET AL., AGRICULTURAL PESTICIDE APPLICATIONS AND OBSERVED CONCENTRATIONS IN SURFACE WATERS FROM FOUR DRAINAGE BASINS IN THE CENTRAL COLUMBIA PLATEAU, WASHINGTON AND IDAHO, 1993-1994, UNITED STATES GEOLOGICAL SURVEY 52 (Water-Resources Investigation Report 95-4285 1996).

[36] *See* Gregory D. Foster et al., *Determination of Dissolved-Phase Pesticides in Surface Water from the Yakima River Basin, Washington, Using the Goulden Large-Sample Extractor and Gas Chromatography/Mass Spectrometry*, 27 ENVTL. SCI. & TECH. 1911-17 (1993).

[37] *See* DAVIS, *supra* note 16, at apps.

tation of what the true, absolute concentration is. Because we cannot bring the whole stream into the lab, our measurements will have error. Error is a statistical way of saying that we have not sampled the true concentration, but only a representation of it. Every time we sample a body of water and analyze it for a pesticide, we obtain a different result. A collection of repeated concentration measurements describes the variation among individual samples. By taking repeated samples, however, we can estimate the true pesticide concentration, and we can also describe the likelihood that our collection of measurements is a good estimate of that concentration.

Thus, the first problem in analysis is actually the error (or sampling variation) in collecting the sample. If a pesticide is concentrated in a certain place in the stream (for example, residing in a slow moving pool rather than the faster moving middle), and we do not sample that place, then our estimate of the true concentration would be in serious error. On the other hand, if we sampled only the pool containing the high pesticide concentrations, then our impression of the overall pesticide concentration in the stream would be biased to the high side. So before the water sample comes through the door of the analytical laboratory, errors, although unintentional, have already been made.

The limitations to accurate analysis imposed by the variations of environmental sampling may partially be overcome by increasing the numbers of samples, and sampling throughout the whole area we are interested in understanding. Increased sampling numbers improve the confidence in the estimate of the average concentration and better define the distribution or range of possible concentrations. Each additional sample analyzed, however, comes with a cost that must be balanced with the benefit derived from having better estimates of precision (repeatability of sampling) and accuracy (how close the average estimate is to the true concentrations).

B. Isolating the Contaminant from Its Matrix

Let us assume that a field inspector has done a very good job of representatively sampling a stream, has composited all the individual samples into one bottle, and then returned them safely to the lab. The analyst then has two objectives: accurately determine the identity of the contaminants (qualification), and determine how much is present (quantification).

Analysis can be boiled down into five mechanical steps: sampling, extraction, cleanup, concentration, and instrumental analysis. In the laboratory, only part of a sample may actually be analyzed. Thus, not only does sampling error occur in the field, error may occur in the lab if the sample is not homogenized in some way. Laboratory sampling error can be a problem for soils and biological materials, but is usually not for water.

The analyst extracts the pesticide by transferring it from its matrix (i.e., water, soil, etc.) into an organic solvent. Extraction is usually accomplished through liquid-liquid partitioning (water), absorption onto a special material (solid phase extraction of water), mixing the matrix with the solvent (soil), or high speed homogenization (biological material). The extracting solvent is separated from the matrix by filtering or centrifugation, and then its volume is reduced (i.e., concentrated) to increase the ability to detect the pesticide. Unfortunately, this latter step also increases the concentration of other compounds (co-extractives) that naturally occur in the sample.

In addition to the sample matrix containing extraneous compounds that are extracted along with the contaminant of interest, the solvents used are much less pure than the water that we wish to analyze. For example, water is usually extracted with an immiscible organic solvent (i.e., water and the solvent do not mix) known as methylene chloride. The methylene chloride commercially available for pesticide and other contaminant analysis has a certified purity of 99.9%. In other words, for every volume of methylene chloride, there are 1000 ppm of impurities. If a water sample contains 1 ppb of a contaminant, then we are trying to analyze a matrix of 99.9999999% equivalent purity with a solvent only 99.9% pure.

Organic solvents left from an analysis must be disposed of as hazardous waste. Ironically, the analyst creates several million times more hazardous waste by use of organic solvents than the equivalent amount of hazardous material he or she is trying to extract. Such a seemingly absurd predicament has been termed analytical "damage."[38]

If the co-extractives in the sample matrix or in the organic solvents used during extraction were going to interfere with the in-

---

[38] *See* H. Steinwandter, *Development of Microextraction Methods in Residue Analysis, in* EMERGING STRATEGIES FOR PESTICIDE ANALYSIS 3, 3-38 (T. Cairns & J. Sherma eds., 1991).

strumental analysis, then they would have to be removed by a cleanup method. Some instrumental methods are not very sensitive to the extraneous compounds, while others are essentially worthless unless the junk is removed. Fortunately, water samples have fewer co-extractives than soil and biological material, and so water does not have to be processed further, which saves time and reduces analytical costs.

During the course of each analytical step, opportunity abounds for small losses of the contaminant residue. Each time a solution is transferred from one container to another, nearly imperceptible spills or failure to transfer 100% may occur. Thus, the mechanical processes of analysis create error that is added to the error accumulated from the field sampling. The more complex the analysis, such as the necessity to clean up a sample or other increased handling requirements, the more opportunity exists for error in measurement.

C. Instrumental Analysis

Once the sample is cleaned up and reduced in volume, it is then passed through an instrument that tentatively identifies it. Based on this tentative identification, the quantity can be calculated. Most of the instrumentation relies on some form of chromatography, which is the separation and detection of like molecules from a population of many different molecules.

For the analysis of pesticides and other contaminants in water, the gas chromatograph (GC) is the most widely used instrument. The GC consists of a long glass-like column on which different molecules are separated by being pushed along in a gas stream. The inside of the column is coated with a viscous liquid of very high boiling temperature, known as the stationary phase. The molecules interact with the stationary phase, moving back and forth from the gas phase. Because different chemicals have different affinities for the stationary phase, they become separated from one another as they travel along the length of the column. Like molecules will travel together but arrive at the end of the column at times different from other molecules. Arrival at the column end is known as elution, and the eluted gas and pesticide molecules pass through a detector that is designed to recognize certain kinds of atoms like chlorine, phosphorous, nitrogen, or sulfur. Every time molecules pass through the detector, an electrical signal is produced. The electrical signal is then sent to a

computer for visualization and further analysis. Thus, what the analyst actually sees is a change in electrical signal upon elution of a chemical through the detector.

Although the GC can detect different molecules, it cannot tell the analyst an absolute identity, which is the specific chemical structure. To identify a compound by GC, the analyst must have already guessed what might be in the sample. For pesticides, the task is somewhat simplified because only so many pesticides are likely to have been used in a particular place. However, the more possible compounds that might be involved, the more difficult to recover and separate them with one analytical method.

Once a list of likely candidate contaminants is developed, the analyst would prepare solutions of the possible array, inject them into the GC, and determine their characteristic patterns of elution. This pattern is described by the time it takes for the molecules to travel along the length of the column and pass through the detector (known as the retention time). Although different kinds of chemicals have specific retention times, some chemicals behave identically with other chemicals, making GC less than adequate as a qualitative instrument. Gas chromatography is actually a quantitative tool, and depends on calibrating the instrument with known amounts of the contaminant that the analyst suspects might be in the sample.

For environmental analysis, absolute identification of a chemical is made using a mass spectrometer. The pesticide is usually introduced into the mass spectrometer through a gas chromatograph. In the spectrometer, molecules are bombarded with electrons into fragment pieces. The pieces are collected according to their mass. The analyst is trained to put the pieces together, as if working a puzzle, to deduce the chemical structure.

D. Making Detection and Quantitation Decisions

Although the steps of extraction and detection seem a lot of work, analysts are only at the beginning of meeting their objectives. First, they must validate the extraction method to ensure that it is reliable and efficient. The analysts must then determine what the detection limit is (i.e., how little can be seen), and what concentration they can accurately calculate. The problem, however, stems from the co-extractives mentioned earlier, as well as the electrical background signals of the detectors. The analyst must determine whether a signal from an analytical instrument

during the measurement of a chemical is above the signal given by the measurement of a sample known not to contain the chemical (i.e., above the "blank" or background noise). When concentrations are substantially above the detection limits of the machine, the background noise is not a problem. However, when the concentration is very low, i.e., close to the detection limit, distinguishing the background signal from the contaminant signal becomes much more difficult. It is at this juncture that mistakes in detection (i.e., false positives or false negatives) can be made.

The chances of making false positives (i.e., saying a contaminant residue is present when it really is not) and false negatives (i.e., saying a contaminant residue is absent when it is really present) are overcome by repeatedly conducting an analysis at the lowest possible levels of concentration. In other words, analysts will take a background sample and deliberately add very low amounts of a chemical to it. Then they will extract and analyze the chemical repeatedly. Repeated analysis will lead to an average concentration and a measurement of the analytical variation or error that is expressed as the standard deviation. For example, if an analyst adds enough DDT to water to produce a concentration of 1 ppt, the lowest limit of machine detection, and then repeatedly analyzes that water, residue concentrations like 2, 2, 4, 5, 3, 2, and 1 ppt may conceivably result. Thus, although the analyst knows the water has an equivalent of 1 ppt DDT, the natural error factors associated with analysis allows only an estimate of what is really there. The closer the true concentration of a chemical is to the lower limits of the detection method, the greater the probability of making a mistake in detecting the chemical.

Even under controlled conditions where an analyst adds a known amount of a chemical to a matrix, a range of possible outcomes results from the analysis. By repeatedly analyzing multiple samples, the analyst learns about the variation of these outcomes. An average outcome value can be calculated, but if all the values are plotted, the resulting graph approaches the shape of a normal distribution or bell shaped curve, which is also called a probability density function.[39] The graph also represents the frequency of different concentration values that can be expressed as the number of standard deviations away (plus or mi-

---

[39] LAWRENCE H. KEITH, ENVIRONMENTAL SAMPLING AND ANALYSIS: A PRACTICAL GUIDE 100 (1991).

nus) from the average (mean) concentration. Three standard deviations away from the average concentration would represent approximately 99.7% of all possible outcomes of repeated measurement.[40]

If contaminant concentrations are only estimates based on probability density functions, then a chance exists of making a wrong decision about whether a contaminant is truly detected and its true concentration. Picture two water samples, one having no contaminant, and the other having a contaminant concentration close to the lowest amount a method can detect. Repeated analysis of each water sample produces two probability density functions that overlap. The area of overlap represents the error associated with making a wrong decision about whether the contaminant is present or not. To reduce this area of overlap, and thereby reduce the risk of making a wrong decision, either positive or negative, analysts will set their method detection limit (MDL) to the equivalent of three times the standard deviation of repeated measurements of a matrix sample not containing the contaminant.[41] Statistically speaking, this will allow conclusions about detection to have only a 0.1% risk of deciding a contaminant is present when it really is not.[42] However, according to probability theory, an error risk of 50% still exists for concluding that nothing is present when in fact it might be. Therefore, proficient analysts will set the reporting limits to six times the standard deviation of repeated measurements near the MDL. This new limit is called the reliable detection limit (RDL). Using this definition, the RDL guarantees only a 0.1% risk of a false negative and false positive.[43]

Once the analyst has concluded that a contaminant is present, then the calculation of its concentration proceeds. To ensure that the concentration has been quantitated with less than 1% error, the limit of quantitation (LOQ) has been set at ten times the standard deviation. At this level of confidence, the calculated concentration has an uncertainty of plus or minus 30%.[44]

An example of how the RDL and LOQ are set follows. For compounds like DDT that are ubiquitous, analysts will add DDT

[40] *Id.* at 101.
[41] *Id.* at 111.
[42] *Id.* at 103.
[43] *Id.* at 102, 103.
[44] *Id.* at 109.

to a water sample in amounts very close to the ability of the analytical instrument to detect it. The source of the water is important. When first developing a method, analysts will likely use distilled water, which is free of DDT residues. Eventually, analysts would try the method by adding DDT to natural water containing a lot of unknown organic compounds that could interfere with the DDT analysis. The analysts will conduct the experiment on at least seven water samples, and then repeatedly measure the resulting DDT concentration. They will then calculate the average concentration obtained and the standard deviation. The standard deviation will be multiplied by three to obtain the MDL, and by six to obtain the RDL. A realistic array of DDT concentrations resulting from analysis of distilled water might be 1.0, 1.1, 1.2, 1.3, 2.0, 1.6, and 1.5 ppt. The average concentration is 1.4 ppt and the standard deviation is 0.34 ppt. Thus, the RDL is 2.1 ppt, higher than what the analysts added. The LOQ would then be calculated as 3.4 ppt.

If natural water was used to conduct the experiment with DDT, the resulting concentrations could be higher because of co-extracted unknown organic compounds that change the background signal of the analytical instrumentation and generally cause more error in measurement. Thus, resulting DDT concentrations could conceivably be 1.6, 5.0, 1.0, 1.0, 1.5, 2.9, and 3.2 ppt. In this case, the average would be 2.3 ppt, but the RDL would be 8.8 ppt. Setting the RDL much higher would minimize the risk of false positive and false negative decisions to 0.1%. The LOQ would be 7.9 ppt, indicating that the true quantity of DDT lies between the interval 5.5 and 10.2 ppt.

Although the aquatic ecological criterion of 1 ppt for DDT is not legally enforceable,[45] and is just a guideline, it is clear that the true concentration, even for a lab proclaiming a detection level of 1 ppt, cannot accurately be known at such low levels. For higher concentrations, say 10 or 100 ppt of DDT, the error of measurement decreases.[46] Thus, if a true concentration is substantially above the RDL, and ideally the LOQ, the user of the generated numbers can have great confidence that they are accurate qualitative and quantitative estimates as practically possible.

---

[45] See Nowell & Resek, supra note 33, at 6-7.

[46] KEITH, supra note 39, at 104.

V

How Well Do Different Analysts Perform?

Valid scientific experiments routinely employ quality control procedures. During the sampling process, the researcher will add known on-site amounts of contaminants of interest to selected samples (known as field spike samples).[47] This procedure accounts for breakdown of contaminants during the trip back to the lab. In the lab, after the analytical procedures have been developed and the MDL has been determined, the analyst will routinely take some of the matrix and add known amounts of the suspected contaminants. These samples will be extracted along with the experimental samples. If any samples must be stored for prolonged periods prior to analysis, the analyst will add the contaminant to extra matrix samples and store these with the samples. Later, these storage control samples are analyzed to determine if contaminant breakdown and loss had occurred during storage, which occasionally happens, even in samples stored at subfreezing temperatures.

Quality assurance procedures are analogous to bookkeeping. The flow of samples, e.g., who handled them and when, is carefully documented. The analytical steps and notations of any deviations from procedures are also documented. Records of all instrumental output are cumulated and calculations verified. All the raw data along with the final report are archived for future reference.

Today, labs that generate numbers to be used in regulatory decisions and enforcement, whether to register a new pesticide or determine compliance with the Safe Drinking Water Act, are mandated to operate under Good Laboratory Practices (GLPs, as defined by the Federal Insecticide, Fungicide, and Rodenticide Control Act)[48] or Contract Laboratory Practices (CLPs, as defined by the Resource Conservation and Recovery Act).[49] Under this set of mandates, standard operating procedures are developed, documented, and implemented for every aspect of lab operation and sample analysis. The requirement for this extensive documentation grew out of fraudulent data collection cases

---

[47] KEITH, *supra* note 39, at 30.

[48] Good Laboratory Practice Standards, 40 C.F.R. § 160 (1997).

[49] Identification and Listing of Hazardous Waste, 40 C.F.R. § 261 (1997); Guidelines Establishing Test Procedures for the Analysis of Pollutants, 40 C.F.R. § 136 (1997).

during the late 1970s at a contract toxicology testing laboratory in Chicago, Illinois.[50] The purpose of GLPs and CLPs was to create an audit trail so that investigators could track how data were generated. While data derived under GLPs and CLPs may be tracked by a third independent party (e.g., an auditor), they do not guarantee that the best available technology is used, nor that the design of experiments is optimal. Thus, the data may be auditable, but its quality may still be open to question.[51]

Despite many contract and government labs today having implemented standard operating procedures consistent with GLPs and CLPs, the actual methods chosen for an analysis will vary, and thus the results among different labs will also vary. Different methods lead to different RDL and LOQs. Errors in analysis occur differently among different operators. Even labs employing similar analytical procedures for regulated contaminants achieve different detection levels and different efficiencies in analysis for the array of possible matrices.[52] Thus, it is instructive to examine expectations for agreement in analysis among different labs.

Interlaboratory studies are routinely conducted by the Association of Official Analytical Chemists (AOAC), EPA, and occasionally, university researchers. A study reported by Norris in 1986 attempted to determine the accuracy and precision of analyses for the pesticide 2,4-D and picloram in streamwater by ten contract laboratories.[53] The levels of these pesticides are used by forest resource managers for assessing the potential impact on water of herbicide drift in replant areas. Norris's interlaboratory comparisons involved preparing quadruplicate samples of water

---

[50] *See* Sharon Begley et al., *Scandal in the Testing Lab*, NEWSWEEK, May 30, 1983, at 83; David L. Dull & Francisca E. Liem, *Is the Federal Insecticide, Fungicide and Rodenticide Act Good Laboratory Practices Program at a Crossroads?*, in GOOD LABORATORY PRACTICE STANDARDS 375, 376 (W.Y. Garner et al. eds., 1992).

[51] Maureen S. Barge, *Good Laboratory Practices and the Myth of Quality*, in GOOD LABORATORY PRACTICES: AN AGROCHEMICAL PERSPECTIVE 41, 41-46 (Maureen S. Barge & Willa Y. Gardner eds., 1988).

[52] *See generally* Keijo I. Aspila et al., *Interlaboratory Quality Control Study of the Analysis of Water for Pesticides*, 60 J. ASS'N OFFICIAL ANALYTICAL CHEMISTS 1097, 1097-1104 (1977); Virlyn W. Burse et al., *Assessment of Methods to Determine PCB Levels in Blood Serum: Interlaboratory Study*, 66 J. ASS'N OFFICIAL ANALYTICAL CHEMISTS 40, 40-45 (1983); Raymond R. Edwards et al., *A Performance Evaluation of Certified Water Analysis Laboratories*, 49 J. WATER POLLUTION CONTROL FED'N 1704, 1704-1712 (1977).

[53] L. A. Norris, *Accuracy and Precision of Analyses for 2,4-D and Picloram in Streamwater by Ten Contract Laboratories*, 24 WEED SCIENCE 485, 485-489 (1986).

with known amounts of 2,4-D and picloram. The samples were sent to individual labs that used their own methodologies to conduct the analyses. The overall bias in reporting average concentrations was negative, i.e, the labs on average reported anywhere from 2 to 92% less pesticide than was placed into the test samples. Two laboratories could not detect any pesticide. One lab reported nearly 1000% more pesticide than added! Within any lab, concentrations obtained were also variable. As an example, one lab reported four replicates of a sample containing 50 ppb picloram to have 13, 16, 44, and 37 ppb, clearly a negative bias that would lead to the conclusion of less contamination than was present[54]. On the other hand, another laboratory assessed the same concentration of picloram as 85, 170, 100, and 110 ppb, leading to an opposite conclusion about contamination.[55] All of the labs were assured to have the capability to detect at least 1 ppb of either pesticide. Thus, even when the true concentration was at least tenfold higher than the detection limits, considerable error still occurred.

A recent study by G.C. Su published in 1998 examined the effects of uncertainty of measurement in laboratories using empirically derived detection limits as opposed to statistically derived limits.[56] The empirically derived limits involved proving that the MDL, as defined by EPA rules, can be achieved and then repeatedly adding the equivalent amount of contaminant to an actual matrix to determine the validity of the MDL. The amount recovered became the RDL, which Su called the reporting detection limit, an empirically derived quantitatable limit. The statistically derived quantitatable limit is equivalent to determination of the LOQ as described above (i.e., ten times the standard deviation associated with repeated analysis of the lowest concentration detectable). The EPA terms this level the ML (minimum level).[57] An alternative statistically derived limit was derived from using a series of concentrations while weighing the lowest concentration heavier in the calculation. This method gave an alternative minimum quantitatable level. Su found that for 43 different contami-

[54] *See id.* at 486 tbl.2 (discussing contract laboratory 2).

[55] *Id.*

[56] *See* G.C.C. Su, *A Comparison of Statistical and Empirical Detection Limits*, 81 J. OF THE ASS'N OF OFFICIAL ANALYTICAL CHEMISTS INT'L 105 (1998).

[57] UNITED STATES ENVIRONMENTAL PROTECTION AGENCY, 821-B-93-001, GUIDANCE OF EVALUATION, RESOLUTION AND DOCUMENTATION OF ANALYTICAL PROBLEMS ASSOCIATED WITH COMPLIANCE MONITORING (1993).

nants the empirically derived quantitatable limit and statistically derived limits agreed within a factor of five about 70% of the time.[58] He concluded that statistically derived quantitatable limits were valid, but that the analyst needed to realize there would be an upper uncertainty factor of five at the lowest concentration at which an instrument is capable of quantitative analysis.[59] The results indicated that when regulatory standards are close to the detectable limits, allowance must be made for the uncertainties of measurement. Thus, Su suggested that when the detected contaminant is above the LOQ (which is the ML), but less than five times the LOQ, further investigation and corrective measures should be taken; accordingly, enforcement should not be applied until the contaminant levels are more than five times the LOQ.

CONCLUSION

A lot of hard work goes into estimating concentrations of contaminants in the environment. High concentrations can be estimated with comparatively little error. At very low concentrations, however, inaccurate conclusions about whether or not a residue is present and its actual amount are easy to make. We can only estimate how close we are to the true concentrations by application of probability statistics to the analytical process. The frequency of different concentrations calculated from repeated measurement can also be thought of as the probability of their occurrence. Thus, the reported residue number does not really represent one "hard" or tangible thing. Rather, it represents a statistically defined probability of whether or not the analyst has captured a fair representation of the true concentration. When a new set of matrix samples is analyzed, a new probability density function results along with a new estimate of the average concentration. For this reason, residue numbers are like virtual realities, i.e., the numbers representing the actual contaminant exist only as long as the analyst is conducting repeated measurements. The estimate is not fixed and will differ from estimates measured in subsequent repeated samplings, even when the concentration has been intentionally added by the analyst.

Our analytical technology is truly remarkable, and has advanced very quickly. Unfortunately, as our knowledge of analyt-

---

[58] Su, *supra* note 56, at 109.
[59] Su, *supra* note 56, at 110.

ical detection advances, our biological understanding lags behind, especially for the very low concentrations of contaminants frequently reported in different matrices. What results are reams of numbers with little indication of actual hazard, and a mere appreciation of the quantitatable uncertainty they reflect. Unless we overcome our innumeracy and seek greater understanding of the relationship of contaminant residues to biological effects, we will be doomed in our desperation to seek nearly nothing in everything.