Human Interindividual Variability—A Major Source of Uncertainty in Assessing Risks for Noncancer Health Effects

Dale Hattis[1] and Ken Silver[2]

For noncancer effects, the degree of human interindividual variability plays a central role in determining the risk that can be expected at low exposures. This discussion reviews available data on observations of interindividual variability in (a) breathing rates, based on observations in British coal miners; (b) systemic pharmacokinetic parameters, based on studies of a number of drugs; (c) susceptibility to neurological effects from fetal exposure to methyl mercury, based on observations of the incidence of effects in relation to hair mercury levels; and (d) chronic lung function changes in relation to long-term exposure to cigarette smoke. The quantitative ranges of predictions that follow from uncertainties in estimates of interindividual variability in susceptibility are illustrated.

1. INTRODUCTION

By this time the NOEL/safety factor approach, which incorporates a 10-fold factor to account crudely for possible interindividual differences within the human population, has been hallowed by long use and has an almost unassailable position in the habits of many regulatory toxicologists. Nevertheless, there are occasions where it is both feasible and desirable to develop procedures for estimating noncancer risks more quantitatively. Quantification of the incidence and intensity of noncancer effects could contribute to decision-making, in particular where actual or anticipated exposures are high enough to produce effects in a directly observable fraction of exposed people. Examples of this include the following.

- The design of improved protocols for the use of pharmaceuticals (where, ideally, the incidence of both beneficial and adverse biological effects for different subpopulations of patients can be modeled based in part on direct observations).[1]
- Assessments of the effects of ubiquitous ambient air pollutants such as ozone, carbon monoxide, lead, and acid particulates.[2,3]
- Assessments of the effects on lung function of chronic lung-damaging agents such as smoking, occupational exposure to coal dust, silica, etc.[4-6]
- Assessments of reproductive and developmental effects, some of which appear likely to produce some excess risk extending to very low dosages.[7]
- Assessments of the incidence of high-dose morbidity and mortality that could be expected to result from large releases of acutely toxic agents, such as chlorine or hydrogen sulfide.[8]

There is sometimes good reason to doubt the universal expectation of population thresholds that is built into the NOEL/Uncertainty Factor schema. In particular cases there may be some finite fraction of individuals who, because of disease or other causes, are marginal for biological functions affected by the chemical and who may be pushed beyond a functional threshold for

[1] Center for Technology, Environment, and Development, Clark University, 950 Main Street, Worcester, Massachusetts, 01610.
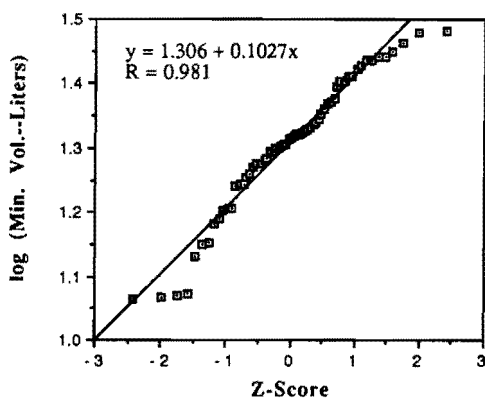[2] Boston University, School of Public Health, Boston, Massachusetts.

Fig. 1. Fit of a lognormal distribution to the minute volumes of 62 British coal miners—data of Jones *et al.*[9]

- Uptake: Individual differences in the environmental concentration needed to produce a given intake of toxicant into the body, e.g., due to differences in breathing rates, dietary habits, etc.
- Pharmacokinetic: Individual differences in the amount of uptake needed to produce a particular concentration–time product of active agent in the blood or at the site of action, e.g., due to differences in metabolic activation or clearance.
- Response: Individual differences in the dose at the active site that produces a similar risk of response.

The discussion below is organized around these major categories of heterogeneity.

an adverse effect by a small finite dose of the chemical. For example, for healthy workers there may indeed be a functional reserve capacity for oxygen delivery to the myocardium and, hence, a finite tolerance for a small impairment of oxygen delivering capacity for the blood due to carbon monoxide or agents that cause the conversion of hemoglogin to methemoglogin. However, for a worker who has just begun to experience a myocardial infarction, oxygen delivery to portions of the mycardium is known to be seriously compromised, and it is possible that a small difference in oxygen delivery capacity due to a modest blood carboxyhemoglobin or methemoglobin concentration could prove the difference between life and death for portions of the heart muscle that are suddenly forced to rely on collateral arterial vessels for oxygen supply.

In the discussion below, we illustrate (i) how we have attempted to use available data to give us some insight into the magnitude of likely interindividual variability in susceptibility (heterogeneity) in the cases of a few specific noncancer risk assessments, (ii) how we have attempted to assess quantitatively the several uncertainties in our estimates of human heterogeneity, and (iii) the magnitude of the uncertainties in risk that flow from our uncertainties on the issue of human heterogeneity.

2. HETEROGENEITY AT VARIOUS STEPS IN THE PATHWAY TO ADVERSE EFFECTS

A few classes of heterogeneity/interindividual variability in susceptibility can be defined as components along the pathway from environmental exposure through the production of adverse effects.

3. ILLUSTRATION OF EXPOSURE/UPTAKE HETEROGENEITY—INTERINDIVIDUAL DIFFERENCES IN BREATHING RATES IN BRITISH COAL MINERS

Breathing rate is a good example of a parameter that directly affects a person's primary uptake of an airborne toxicant. Other things being equal, a worker who breathes more air per unit time will take in more dust, etc., per unit time, for a given concentration of dust in air. One of the best data sets on workers' breathing rates that we have seen is in a report by Jones *et al.*[9] covering 62 underground British coal miners. Breathing rates were measured over periods that were typically about 90 min, and in most cases three replicate determinations were made. Data of this type allow us to illustrate three different problems in analyzing individual variability in this type of directly measurable parameter:

- How do we decide what statistical form to use for the population distribution of the parameter?
- How do we decide how much of the observed variation in a set of measurements is due to "true" interindividual variability and how much is due to measurement error?
- How do we determine the *uncertainty* in our estimate of the population variation of the parameter? That is, if we calculate a standard deviation, or a geometric standard deviation from individual measurements, how often could we expect to be wrong by various amounts in relation to the standard deviation we would calculate if we had measurements on an infinite number of people?
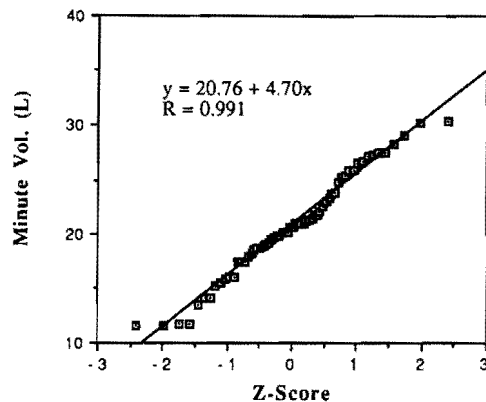
Fig. 2. Fit of a normal distribution to the minute volumes of 62 British coal miners—data of Jones et al.[9]

3.1. Analyses Using Alternative Statistical Forms

Figure 1 shows a lognormal probability plot of distributional data of this type.[3] A true lognormal distribution would be expected to result from a situation in which many factors each contributed in small ways to variation in the measured parameter, and the factors all acted multiplicatively. In practice, many parameters tend to be reasonably well described by this type of plot. However, in this case, as can be seen in the comparison with Fig. 2, a normal distribution provides a slightly better description of the observed variability. Elsewhere[10] we have made a tongue-in-cheek proposal for "laws" of uncertainty/variability analysis. The first of these "laws" is, "Nearly all parameter distributions

---

[3] To create this type of plot, the measurements are first arranged in order and given ranks i (1 through N). Then one calculates a "percentage score" for each ordered value as $100*(i-0.5)/N$. (This differs from the usual definition of a "percentile" in which the highest observation is assigned a score of 100.) Finally, from tables of probits or areas under a cumulative normal distribution, one calculates the number of standard deviations above or below the median of a normal distribution that would be expected to be associated with each "percentage score," if the distribution of values were in fact normal (Gaussian). In the regression line calculated from this type of plot, the intercept ($Z=0$) is an estimate of the median, and the slope is an estimate of the standard deviation. This procedure has the advantage of being able to accommodate truncated data (e.g., the presence of "nondetect" values)[26] or, as in the figure above, data that were originally provided only in the form of a histogram of the numbers of children above specific cutoff values. Where there are nondetect values, these contribute to the N used for calculation of percentage scores but such points are not plotted and therefore do not contribute directly to the fitted regression line. For further information see Ref. 27.

look lognormal, as long as you don't look too closely." Although phrased facetiously and deliberately overstated, these laws do represent regularities we have commonly observed in our practical experience. In this case, we mean to imply that where one does not have a strong mechanistic reason to prefer one type of distribution, it can help illuminate both the facts and the associated uncertainties to compare observations with expectations under a range of distributional forms.

3.2. Removing Estimated Measurement Error from Estimates of True Population Heterogeneity

The second issue—measurement (and, implicitly, other short term) variability—arises because as risk assessors we are not simply interested in describing the distribution of a set of observations. We want to use the data to help make inferences about how much different the true delivered doses and risks might be facing different people in the same environment. For this purpose, because an agent such as coal dust is expected to pose lung damaging hazards that depend primarily on doses delivered over an extended period (at least months, and more often years or decades), we need to estimate the distribution of long-term doses that are expected to result from differences in relatively stable characteristics of individuals and their job requirements. The distribution of our observations (Fig. 1) represents the combined effect of variation that results both from the true long-term variability we are interested in and various short-term perturbations, including measurement errors. In the fortunate special case where (i) both the true long-term variability and the measurement errors are normally distributed and (ii) the measurement error/short-term variability in our worker breathing rates can be calculated from the spread of the three replicates for the individual workers, then we can simply subtract the measurement variance from the total observed variance to obtain an estimate of true long-term variance. For the Jones et al.[9] data set our calculations indicate that about a third of the total variance represents likely measurement error and/or true short-term individual variability, leaving an estimate of the "true" long-term standard deviation of the individual breathing rates of about 3.84 L/min, rather than the observed overall standard deviation of 4.72 (this corresponds approximately, but not exactly to the standard deviation inferred from the plot in Fig. 2: 4.698). Thus, from the standpoint of a regulator who wished to consider the 2.5–97.5 percentile range on the expected long-term dose of dust contained in the air breathed by individual coal miners, the expected range

would be about 13.1–28.4 L/min—still over a twofold range, but less than the 11.3–30.2 L/min range that would be calculated without the correction for likely measurement and short-term variability.

3.3 Estimating Uncertainty in Estimates of Variability

The final issue we would like to illustrate with these data is how confident we should be in the original calculation of the overall standard deviation of 4.72 L/min itself. If we were to repeat the measurements on a large number of groups of 62 miners, and calculate a standard deviation each time, what would be the standard deviation of those standard deviations? Using available Monte Carlo simulation software, it is relatively straightforward to perform such an experiment on a computer. Based on 5093 trials in which the computer drew random normally distributed values for groups of 62 simulated miners, each of which had a true mean and standard deviation of 20.77 and 4.723 L/min, we found that the standard deviation of the simulated standard deviations was 0.42 L/min—about 9% of the best-estimate standard deviation, and leading to a 95% confidence range on the standard deviation of 3.88–5.56 L/min. In contrast, we know the mean breathing rate (20.77 L/min) to a much greater degree of precision. The standard error of this mean is about 0.60 L/min—only about 3% of the mean itself. This leads to our second tongue-in-cheek "law" of uncertainty analysis: "Any estimate of the variability of a parameter value will always itself be more uncertain than the estimate of the parameter value." Clearly, if a regulator wished to be very confident that he or she was protecting the miners with 98th or 99th percentile breathing rates, he or she would need to consider a much broader range of values than those we arrived at in the previous paragraph, based only on the best estimate of the standard deviation of long-term average breathing rates.

4. INTERINDIVIDUAL VARIABILITY IN SYSTEMIC PHARMACOKINETIC PARAMETERS

In our initial work for systemic pharamacokinetic parameters,[11] we sought to assemble data from studies of normal healthy adults on the variability in three particular types of pharmacokinetic parameters in humans, based primarily on studies of drugs.

- *Elimination half-lives* $(T_{1/2})$ are defined as the time required for a twofold reduction in the concentration of the substance in some compartment (usually blood) after absorption is complete. The longer the elimination half-life, the higher the concentrations that will be attained in the body if exposure is continuous or repeated at a frequency that is short relative to the half-life. Therefore, elimination half-lives are likely to be key determinants of susceptibility to toxicity from chronic continuous exposures.

- *Area under the curve* (AUC) is a plot of the plasma concentration versus time after exposure. The AUC integrates variability in the efficiency of absorption as well as metabolism and elimination. It is likely to be a good predictor of pharmacokinetic variability in susceptibility for effects that are linearly related to the amount of slowly accumulating products of reaction between the toxicant and resident macromolecules.

- *Peak concentration* (C_{max}) in blood represents an effective dose in cases of acute toxicity from an isolated single exposure to a substance. Like AUC, it also integrates information about absorption; in this case, both the efficiency and the dynamics of absorption.

4.1. Observed Variability in Groups of Normal Healthy Adults

Table I summarizes our results. It can be seen that in this case we have used geometric standard deviations to characterize the interindividual variability results. This is because we found that much more often than could be expected by chance, the data were positively skewed—showing a long tail of relatively high values. Additionally, a difficulty in applying a normal distribution to these data (or to the breathing rate data analyzed previously) is that inherently normal distributions must at some point take on negative values, which have no meaning for these parameters. Still, we have not definitively tested the performance of the lognormal distribution against other positively skewed distributions using these data.

It can be seen from the data in Table I that there are important differences among different chemicals/drugs in the degree of interindividual variability in these systemic pharmacokinetic parameters. In the usual case, where a regulator/risk analyst is considering a chemical

Table I. Variability in Human Systemic Pharmacokinetic Parameters Seen for Different Chemicals (Mostly Drugs) in Groups of at Least Five Healthy Volunteers[a] (Adapted from Ref. 11)

Range of geometric SD[a]	Corresponding probit slopes[b]	Human interindividual variability in pharmacokinetic parameters					
		Elim. half-life		AUC		C_{max}	
		No. obs.	Cum. %	No. obs.	Cum. %	No. obs.	Cum. %
1.778–3.16	2–4	2[a]	5	3	20	1	8
1.468–1.778	4–6	5	16	3	40	2	25
1.334–1.468	6–8	13	45	2	53		
1.212–1.334	8–12	10	68	5	87	5	67
1.155–1.212	12–16	8	86	1	93	2	83
1.000–1.155	16+	6	100	1	100	2	100
Total		44		15		12	
Median geometric SD		1.291		1.386		1.285	
95% range of observations[c]		1.109–1.858		1.117–2.748		1.072–2.506	

[a] Each of these observations represents a single chemical. Where there were multiple observations of interindividual variability for a partticular chemical, a combined geometric standard deviation was calculated by pooling the variances within the different experiments.

[b] This is the reciprocal of the log of the geometric standard deviation. Thus a log GSD of 0.5 corresponds to a log probit slope of 2.

[c] Assuming that the log(geometric standard deviation) values for the different chemicals are themselves distributed lognormally.
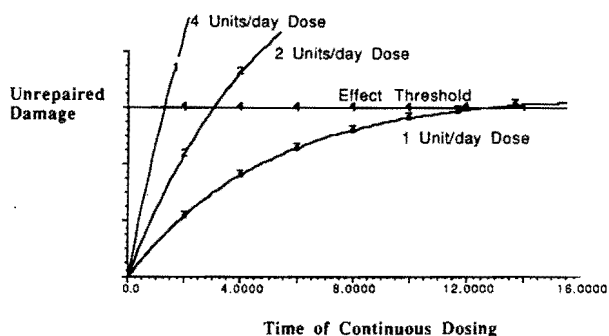
Fig. 3. Theoretical framework of slowly accumulating reversible damage.

for which there are no analyzable human data on pharmacokinetic variability, one essentially faces at least[*] the range of uncertainty represented by the 95% confidence limits given in the bottom line of the table. In the next section, we show how we used the upper end of the 95% confidence range for AUC to provide an indication of the likely uncertainty in projections we made for the chronic toxic risks of acrylamide.[12,13]

* We say "at least" because it should be stressed that the data in Table I were derived from studies in normal healthy adults and are, therefore, likely to understate the actual degree of variability that would be seen in a diverse general population, including elderly people, infants, people with various diseases, etc. On the other hand, the results in Table I reflect total observed variability; we did not attempt to make the kinds of corrections for measurement error that we illustrated in the previous section.

4.2. Implications of the Observed Distribution of Apparent Interindividual Variation in Pharmacokinetic Parameters—Use in a Risk Assessment for Acrylamide Neurotoxicity

For our acrylamide modeling work we analyzed some of the most classical studies of acrylamide neurotoxicity in animal systems[14–16] in which effects were seen to be produced at a lower cumulative dose if the acrylamide was administered at a higher dose rate. This implied some rate of repair of the incipient damage, whose buildup is illustrated in Fig. 3. From these data we calculated the apparent rate of repair of the incipient damage and the daily dose rate that would just barely be able to produce the various effects if the dosing were continued for the lifetime of the animals (baboons or rats; other data, collected in monkeys, were in somewhat different form but were analyzed similarly).

Because the results indicated no discernible pattern of change in calculated repair rates or long-term doses required to produce effects in different species, the doses of acrylamide that would be expected to produce different effects in *average* humans were projected directly from available primate data on a milligram per kilogram basis. The "best" (least unlikely) estimates of the doses expected to produce effects in smaller proportions of exposed people were calculated by assuming that humans will show a lognormal distribution of susceptibilities for specific effects with the same spread as was observed in the corresponding baboon and monkey experiments (Table II). (The long-term dose rates associated with risk

Table II. "Best" (Least Unlikely) Estimates of Interindividual Variability for Neurotoxic Effects of Acrylamide Derived from Limited Studies[a] in Primates

Species and parameter	Interindividual variability in the long-term dose rate required to cause a specific effect (Geometric SD)
Baboon	
Hindlimb weakness	1.230
Forelimb weakness	1.401
Monkey	
40-Hz vibration, 10% threshold increase	3.715[b]
150-Hz vibration, 10% threshold increase	1.349
Pickup test, 10% performance loss	1.995

[a] The experimental results reflected data from only six baboons and four monkeys.

[b] The large interindividual variability calculated here represents primarily the influence of one of the four available values.

levels of 10^{-4} and 10^{-6} were calculated as 3.72 and 4.75 geometric standard deviations below the estimated threshold dose for the median person.) To represent the uncertainties on this point, similar projections were also made assuming a higher degree of human interindividual variability—a geometric standard deviation of 2.748—based on the upper 95% confidence limit of measurements of interindividual variability for (AUC) from Table I.

The results are given in Table III. (It should be stressed that in all cases, the projections depended on an assumption that susceptibilities are lognormally distributed—for which there is no specific evidence.) It can be seen that the degree of human interindividual variability assumed has a dramatic effect in changing the doses that are expected to be associated with effect incidences of 10^{-4} and 10^{-6}. The "best-estimate" column in Table III shows that whereas the doses associated with a median risk for the different effects show a range of only about 4-fold, the 10^{-4} and 10^{-6} risk levels are associated with dose rate spreads of 200- and 700-fold, respectively, among the effects. The "plausible lower limit" projections (final column in Table III) serve to emphasize that if the interindividual variability in human susceptibility for acrylamide's more serious reversible effects is much larger than the best-estimate projections indicate (but within the range that has been observed for the variability of some pharmacokinetic parameters in humans, for a few percent of the chemicals studied), the risks of even the more serious forms of acrylamide peripheral

Table III. "Best" (Least Unlikely) and "Plausible Lower Limit" Projections of Exposures Potentially Associated with Acrylamide Toxicity in Different Proportions of Workers (Adapted from Ref. 13)

Data source, effect, and proportion of population expected to be affected	Long-term air concentrations (mg/m^3) needed for toxicity[a]	
	Best estimate[b]	Plausible lower limit[c]
Baboon[14]		
Hindlimb		
Proportion affected		
0.5	81	41
1×10^{-4}	38	1.18
1×10^{-6}	30	0.41
Forelimb		
Proportion affected		
0.5	66	25
1×10^{-4}	8.9	0.90
1×10^{-6}	5.1	0.31
Monkey[17] 10% increase in baseline threshold		
40-Hz vibration		
Proportion affected		
0.5	20	4.8
1×10^{-4}	0.16	0.041
1×10^{-6}	0.044	0.011
150-Hz vibration		
Proportion affected		
0.5	33	8.4
1×10^{-4}	11	0.41
1×10^{-6}	8.0	0.14
Pickup		
Proportion affected		
0.5	43	11
1×10^{-4}	3.3	0.45
1×10^{-6}	1.6	0.16

[a] On an occupational exposure schedule (5 days/week, 8 h/day).

[b] Based on the degree of interindividual variability observed in the corresponding primate data, an assumed breathing rate of 7.3 m^3 per 8-h work day, and a lognormal distribution of individual thresholds for effect.

[c] Based on the combined effects of uncertainties in the appropriate estimate of the degree of human interindividual variability and the repair rates observed in animals, a breathing rate of 10 m^3 per 8-h work day, and a lognormal distribution of individual thresholds for effect.

neurotoxicity (limb weakness) might approach 1 in 1 million at the then-current occupational standard (0.3 mg/m^3). The most modest effects on perceptual threshold and related motor functions indicated by the "pickup" test would be expected to be less rare. Clearly, it would lessen our uncertainty in population dose reponse considerably if we had some actual measurements of interindividual variability in humans.
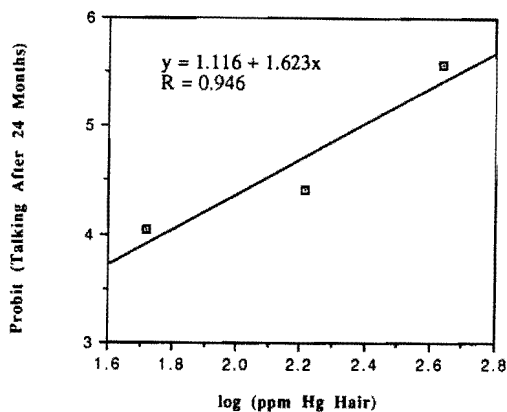
Fig. 4. Log probit dose–response relationship for talking after 24 months in relation to maternal hair mercury—data of Marsh et al.[19] Source: Institute of Medicine, *Seafood Safety*, Reproduced with permission.

5. VARIABILITY IN EFFECT PARAMETERS

Effect parameters come in two broad types—quantal (yes/no; present/absent) and continuous (e.g., measurements of a functional parameter that can take on any value within some continuous range). These are explored in turn in the two subsections below.

5.1. Observations of Apparent Interindividual Variability in a Quantal Effect Parameter—Incidence of Fetal Effects After Dietary Methyl Mercury Exposure

The strategy in using quantal data is generally to assume that an observable response is produced when some underlying continuous parameter (e.g., internal dose or damage) exceeds some critical threshold, as was done in the previous section for acrylamide. Changes in the frequency of quantal responses as a function of dose are therefore interpreted as changes in the fraction of the population whose individual thresholds have been exceeded. An unusual analysis of human data of this type was included in a report on *Seafood Safety* by the Institute of Medicine.[18] This addresses a key issue of whether (and if so by how much) interindividual variability in susceptibility for fetal/developmental effects differs from interindividual variability in susceptibility for effects in adults.

One of the best available (although still quite limited) human data sets for reproductive effects has been published by Marsh et al.[19] For 81 mother–infant pairs, these authors provide detailed information on the inci-

dence of a variety of fetal methyl mercury effects in relation to the maximal levels of mercury found in the hair of the mothers during gestation. The observations come from an Iraqi mass poisoning incident which resulted from the distribution of methyl mercury-treated Green Revolution seed grain. Observations of the children were made some years after birth. Maximum mercury concentrations were assessed by a series of sequential measurements along the hair shafts during fetal development. Log probit dose–response fits to these data [e.g., by simple regression analysis in Fig. 4 and by the more elaborate maximum likelihood procedure of Finney[20] in Table IVA] indicate very large amounts of interindividual variability in response—probit slopes of about 1. Analogous, but unfortunately not completely comparable analyses of data for a variety of adult effects, based on measurements of methyl mercury in blood, suggest probit slopes in the range of 2–8 (Table IVB). A probit slope of 1 would imply that 95% of the population would have thresholds for effect spread out over a span of about 10,000-fold in dosage—from 100-fold lower to 100-fold higher than the dose that would cause the effect in people of median susceptibility in an exposed population. Such a large amount of interindividual variation would imply appreciable risks (of the order of 10^{-5} to 10^{-2}) even at the much lower dosages that are present in the diets of people who consume relatively large amounts of fish with relatively large methyl mercury concentrations. A probit slope of 2 would suggest less, but still appreciable variability—with the thresholds of 95% of the population spread over a 100-fold range in dosage from 10-fold lower to 10-fold higher than the threshold for the median person.

Unfortunately, in addition to statistical uncertainties in the determination of these slopes from limited data, there are important questions of biological interpretation. A conclusion that the relationships represented in Fig. 4 and Table IVA represent true interindividual variability depends on an assumption that the biomarker of exposure used in this case—the maximum hair mercury found at any time during gestation—is the most appropriate direct causal predictor of response that can be developed. (Other possibilities might well include the concentration of mercury in maternal blood at a specific sensitive time during gestation or a weighted sum of concentrations × duration over a specific set of sensitive periods.) Any inaccuracy in the assessment of the relevant dose used for the least-squares regression analyses in Fig. 4 and Table IV would tend to cause a bias in the estimation of the probit slopes toward lower values (and hence higher estimates of interindividual variability and low-dose risks.)

Table IV. [a] Maximum-Likelihood Fits[b] of (A) of the Marsh et al.[19] Fetal and (B) the Iraqi Adult Methyl Mercury Effects Data Using the Method of Finney[20]

Effect	Background % response[c]	Probit slope	Slope SE	Intercept	ED_{50} (ppb blood)	ED_{50} geom. SE	χ^2	df[d]	p[e]
(A) FETAL EFFECTS									
Late walking	0	1.21	0.30	2.19	205	1.49	6.093	3	0.11
Late talking	7.3	1.76	0.71	0.81	244	1.38	0.689	1	0.41
Mental symptoms	2.4	0.99	0.76	1.88	1429	4.75	0.351	1	0.55
Seizures	0	1.10	0.53	1.54	1399	2.95	0.356	3	0.95
Neurol. score >4	0	0.85	0.27	2.42	1047	2.54	0.874	3	0.83
Average		1.18				Sum	8.363	11	0.68
(B) ADULT EFFECTS									
Paresthesias	7.5	2.17	0.63	−1.64	1145	1.24	1.155	2	0.76
Ataxia's	2.5	3.92	0.76	−7.67	1687	1.11	4.95	3	0.18
Visual changes	0	2.19	0.43	−2.24	2006	1.16	3.695	3	0.3
Dysarthria	5	4.67	1.36	−11.2	2952	1.10	5.608	3	0.13
Hearing defects	1.3	6.42	2.17	−18.05	3877	1.10	0.209	1	0.65
Deaths	0	7.58	3.19	−23.05	5007	1.18	0.83	1	0.36
Sum for adult effects							16.447	13	0.22
Sum for fetal and adult effects							24.81	24	0.42

[a] Source: Ref. 18 (reproduced with permission).

[b] The equation fit is Probit of excess risk over background = Intercept + (slope) * log(blood Hg as ppb). A "probit" is 5 + the number of standard deviations above the median of the cumulative lognormal distribution of the underlying thresholds; 5 was originally added so that toxicologists would not have to deal with negative numbers.

[c] Estimated from data in the lowest one to three dose groups.

[d] This is the number of dose groups available for analysis, less 2 for the number of parameters estimated from the data (the intercept and probit slope).

[e] This is the probability that a deviation as large as that observed between the log probit model and the data would have been expected by chance, even if the log probit model were a perfect description of the underlying dose–response function.
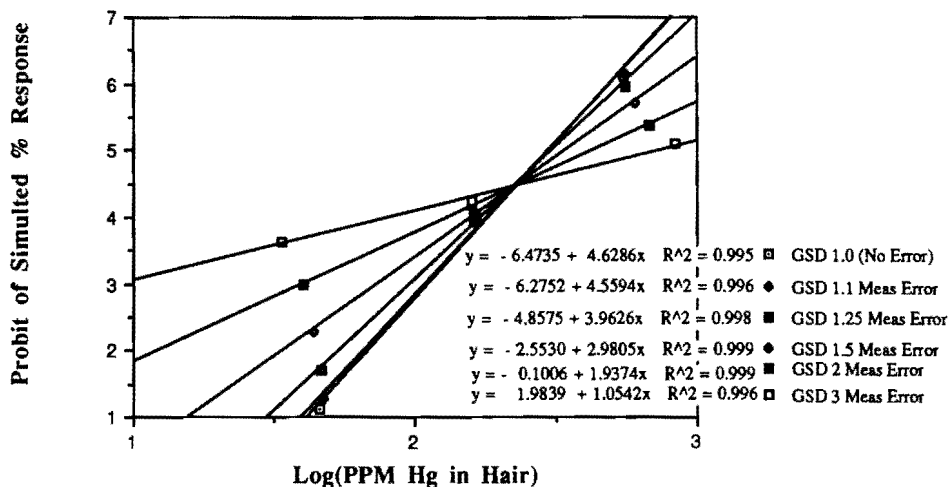

Fig. 5. Theoretical effects of grouping and lognormally distributed measurement/estimation errors on observed probit slopes for an effect with a "true" probit slope of 6.

Figure 5 illustrates the effects of various degrees of lognormally distributed uncertainty (expressed as geometric standard deviations) in estimating the probit slope for the "late-talking" dose–response relationship. For this figure, we have assumed a "true" probit slope of 6 but then introduced the indicated amounts of measure-
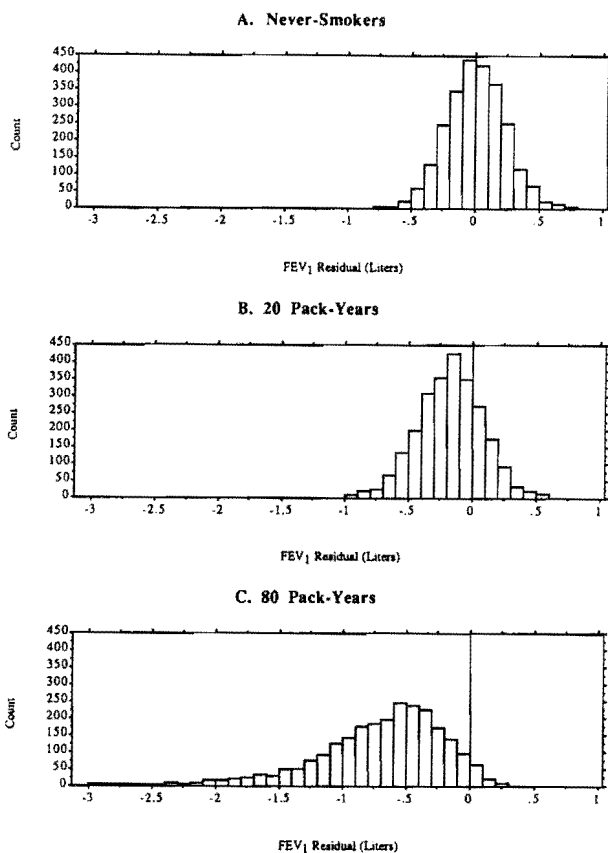
A. Never-Smokers

FEV₁ Residual (Liters)

B. 20 Pack-Years

FEV₁ Residual (Liters)

C. 80 Pack-Years

FEV₁ Residual (Liters)

Fig. 6. Simulated distributions of FEV$_1$ residuals in groups of 2500 people with normally distributed background variability and lognormally distributed (gsd = 1.9) differences in susceptibility to FEV$_1$ loss from smoking.

ment/estimating uncertainty in Monte Carlo simulations with 2000 trials for each line. The simulated dose and response data were then aggregated into the three ranges of estimated dose (20–80, 80–320, and 320+ ppm in hair) used in Fig. 4 for plotting. It can be seen in this case that (i) measurement/estimation inaccuracies as large as a geometric standard deviation of 3 would be needed to make a "true" probit slope 6 relationship appear to have a probit slope as low as 1, and (ii) some appreciable reduction in the apparent probit slope (from the input of 6 to the 4.6 seen in the GSD 1 "no error" line) is produced just by grouping the data into three ranges, even before introducing the estimation error. Analyses in which the data are disaggregated into more, smaller ranges create less distortion in the estimation of dose and response within the dosage categories and, thereby, give probit slopes that are closer to the "true" input slope. A further implication of this analysis is that if we are able to measure or estimate our uncertainties

in the individual "doses" estimated in epidemiological studies, we may be able to at least make an approximate reconstruction of the slopes of the underlying dose–response relationships.

5.2. Overall Interindividual Variability in a Continuous Outcome Parameter—Variation in Smoking-Related FEV$_1$ Loss

Continuous data, in contrast, present the analyst with a different kind of opportunity. The analyst can either (i) convert the data back into quantal form, by imposing one or more numerical cutoffs to define a "response" or responses of graded severity (this nearly always sacrifices some information), or (ii) utilize the continuous data directly. In the latter case, if the parameter measures some function, the baseline spread of the population in the continuous parameter shows directly how many individuals are how far from various degrees of impairment in that function. In addition, however, to the extent that individuals differ in *the amount of parameter change per unit exposure* to the agent, the analyst should see an increasing spread of the parameter values as the dosage increases to higher levels. This is illustrated with simulated data for the case of cigarette smoking and chronic changes in "FEV$_1$ residuals"[5] as a function of smoking dose in Fig. 6. Real data,[21,22] as analyzed by Silver and Hattis,[4,23] are similar in appearance.

In that analysis we made use of cross-sectional data from the Harvard University Six Cities Study[21] and a smaller earlier study in Tucson[22] to draw inferences regarding human interindividual variability in response to cigarette smoke. In the former study, lung function tests were performed on 8191 men and women 25 to 74 years of age who were randomly selected from the six participating communities. Information on current and lifetime cigarette smoking habits was gathererd by questionnaire. Using these data we asked,

- Is the degree of spreading of FEV$_1$ residuals with increasing cigarette dose greater than would be expected on the basis of (a) the "baseline" variability of FEV$_1$ residuals seen in never-smokers and (b) the likely variability in cigarette dose *within* dose categories?
- What degree of interindividual variability is most compatible with the data (expressed as a geometric standard deviation for an assumed log-

---

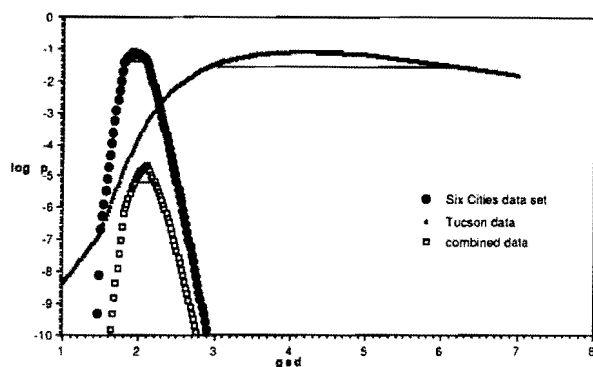[5] An individual's FEV$_1$ is the amount of air that can be breathed out in 1 s—a common measure of lung function.

Fig. 7. Log likelihood distributions for Six Cities[21] and Tucson[22] data sets, separately and combined. "Maximum-likelihood" estimates and 95% confidence limits (horizontal lines) for each curve: Six Cities data set, 1.9 (1.78, 2.12); Tucson data set, 4.2 (3.00, 6.42); combined data, 2.1 (1.90, 2.24).

normal distribution of individual rates of decline of FEV_1 per pack-year of cigarette smoking)? and

● What are the confidence limits around our estimates of interindividual variability in susceptibility to FEV_1 decline, considering each data set separately and combined?

Figure 7 presents in capsule form the results of our analyses of the two data sets. The "*p*" values shown were calculated from χ^2 comparisons of observed vs. expected distributions of FEV_1 residuals, where the "expected" distributions were calculated using a variety of different values for the geometric standard deviation of FEV_1 change per pack-year of smoking dose. The horizontal lines represent calculated 95% confidence range for each study separately and for a combined analysis of both studies.

The results from analyzing each study indicated that a geometric standard deviation of 1 (no interindividual variability) is incompatible with the data, but the estimates of gsd from are derived from our analyses of the two studies were unfortunately statistically incompatible with each other.

There are many challenges in analyses such as this. One must first do a very good job in estimating the relationships between the continuous variable under study (in this case FEV_1) and various confounding factors. Moreover, in the unexposed group, it is crucial to be able to describe accurately the distribution of departures from the basic prediction equation for different individuals. (This is the "baseline variability" with which the spread of values in the exposed groups are later compared.) In the case of the FEV_1 residuals for lung func-

tion, we found it necessary to use a mixture of two normal distributions, rather than a single normal distribution, to describe adequately our baseline variabilities for each of the data sets. Even then, the analysis depends critically on the precise departures of the number of observations at the tails of the fitted distributions from those expected.

In the light of these uncertainties, we cautioned that our numerical results must be regarded as crude initial estimates. Our analysis is significant more for the question we raised, and the potential we show for using a commonly collected type of cross-sectional data for addressing a central issue in the assessment of risk for noncancer effects, than for our final numbers.

DISCUSSION

We have illustrated both some of the problems and the unexploited opportunities for shedding light on different human risks by assessing likely human interindividual variability in parameters that contribute to susceptibility. Assessments of interindividual variability at various steps in the causal process from exposure to adverse effects are a vital component of risk assessments for noncancer effects. There is a considerable need to develop both more/better data and more sophisticated procedures for analysis in this area.

One obstacle to such progress arises from a basic attitude that is common among researchers. To many experimentalists, interindividual variability in the susceptibility of study subjects is an annoyance to be rid of to the greatest extent possible. Interindividual variability tends to increase the size of the sample of individuals that need to be studied to establish a causal connection between a particular exposure and a particular effect. Thus much toxicology is done on inbred strains of animals, reared under carefully controlled conditions and subjected to exposures beginning at a defined age.

Similarly, it is standard for at least the initial characterization of human pharmacokinetic parameters for drugs to be done in small groups of young adults, often of a single gender.[24] And it is fair to say that environmental epidemiologists as well, partly because of the difficulty of utilizing human data for establishing relationships in the first place between exposures and adverse health effects, have rarely focused on the question of providing quantitative insights into the degree of interindividual variability in susceptibility to various hazards.

Elsewhere we have pointed out that improved quantitative understanding of interindividual variability in

both beneficial and adverse effects of pharmaceuticals could make a substantial contribution to the better design of protocols for the practical use of drugs.[25] We have illustrated in this paper that uncertainties about the degree of interindividual variability make up an important part of our uncertainties in assessing risks for a wide variety of noncancer effects of environmental agents, and that, although the analysis of interindividual variability poses substantial challenges both for the collector of data and for the analyst, it is feasible to extract some information on variability from available or obtainable human data. However, progress in these areas depends on turning interindividual variability from a complication to be "controlled out" to the maximum extent feasible, to a primary object of study in its own right, as an important determinant of effects of possible social significance.

ACKNOWLEDGMENTS

REFERENCES

1. D. Hattis. "The Importance of Exposure Measurements in Risk Assessment of Drugs,"*Arch. Toxicol.* Suppl. **16**, 201–210 (1994).
2. J. V. Hall, A. M. Winer, M. T. Kleinman, F. W. Lurman, V. Brajer, and S. D. Colome. "Valuing the Health Benefits of Clean Air," *Science* **255** 810–817 (1992).
3. D. Hattis, S. Abdollahzadeh, and C. A. Franklin, "Strategies for Testing the 'Irritation-Signaling' Model for Chronic Lung Effects of Fine Acid Particles," *JAPCA* **40**, 322–330 (1990).
4. K. Silver and D. Hattis. "Methodology for Quantitative Assessment of Risks from Chronic Respiratory Damage: Lung Function Decline and Associated Mortality from Coal Dust," CTPID 90-9, (MIT Center for Technology, Policy, and Industrial Development, Cambridge, MA, May 1991).
5. I. A. Greaves and M. Shenker. "Tobacco Smoking,"in J. D. Brain, et al. (eds.), *Variations in Susceptibility to Inhaled Pollutants* (Johns Hopkins University Press, Baltimore, 1988, pp. 182–203).
6. J. Koenig. "Pulmonary Reaction to Environmental Pollutants," *J. Allergy Clin. Immunol.* **79**(6), 833–843 (1987).
7. M. Ballew and D. Hattis, "Reproductive Effects of Glycol Ethers in Females–A Quantitative Analysis,"CTPID 89-7 (MIT Center for Technology, Policy, and Industrial Development, Cambridge, MA. July 1989).
8. N. A. Eisenberg, C. J. Lynch, and R. J. Breeding. *Vulnerability Model: A Simulation System for Assessing Damage Resulting from Marine Spills,* Report by Enviro Control, Inc. to the U.S. Coast Guard, DOT-CG-33377-A, National Technical Information Service AD-A015 245 (1975).
9. C. O. Jones, S. Gauld, J. F. Hurley, and A. M. Rickman. Personal differences in the breathing patterns and volumes and dust intakes of working minors. Report to the Commission of the European Communities, TM/81/11 (Environmental Branch, Institute of Occupational Medicine, Roxburgh Place, Edinburgh, Scotland, 1981).
10. D Hattis. "Three Candidates Laws of Uncertainty Analysis" (Letter to the Editor), *Risk Anal.* **10**, 11 (1990).
11. D. Hattis, L. Erdreich, and M. Ballew. "Human Variability in Susceptibility to Toxic Chemicals—A Preliminary Analysis of Pharmacokinetic Data from Normal Volunteers," *Risk Anal.* **7**, 415–426 (1987).
12. D. Hattis, and K. Shapiro. "Analysis of Dose/Time/Response Relationships for Chronic Toxic Effects—The Case of Acrylamide," National Technical Information Service No. NTIS/PB89-109581, (MIT Center for Technology, Policy, and Industrial Development, (Cambridge, MA, July 1988).
13. D. Hattis and K. Shapiro. "Analysis of Dose/Time/Response Relationships for Chronic Toxic Effects—The Case of Acrylamide," *NeuroToxicology* **11**, 219—236 (1990).
14. A. Hopkins. "The Effect of Acrylamide on the Peripheral Nervous System of the Baboon," *J. Neurol. Neurosurg. Psychiat.* **33**, 805–816 (1970).
15. P. M. Fullerton and J. M. Barnes. "Peripheral Neuropathy in Rats Produced by Acrylamide," *Br. J. Industr. Med.* **23**, 210–221 (1966).
16. M. L. Kaplan and S. D. Murphy. "Effect of Acrylamide on Rotarod Performance and Sciatic Nerve β-Glucuronidase Activity of Rats," *Toxicol. Appl. Pharmacol.* **22**, 259–268 (1972).
17. J. P. J. Maurissen, B. Weiss, and H. T. Davis. "Somatosensory Thresholds in Monkeys Exposed to Acrylamide," *Toxicol. Appl. Pharmacol.* **71**, 266–279 (1983).
18. Institute of Medicine. *Seafood Safety* (Committee on the Evaluation of the Safety of Fishery Products, Food and Nutrition Board, National Academy Press, Washington, DC, 1991).
19. D. O. Marsh, T. W. Clarkson, C. Cox, G. J. Myers, L. Amin-Zaki, and S. Al-Tikriti. "Fetal Methylmercury Poisoning. Relationship Between Concentration in Single Strands of Maternal Hair and Child Effects," *Arch. Neurol.* **44**, 1017–1022 (1987).
20. D. J. Finney, *Probit Analysis* (Cambridge University Press, Cambridge, 1971).
21. D. W. Dockery, F. E. Speizer, B. G. Ferris, J. H. Ware, T. A. Louis, and A. Spiro. "Cumulative and Reversible Effects of Lifetime Smoking and Simple Tests of Lung Function in Adults," *Am. Rev. Resp. Dis.* **137**, 286–292 (1988).
22. B. Burrows, R. J. Knudson, M. G. Cline, and M. D. Lebowitz. "Quantitative Relationships Between Cigarette Smoking and Ventilatory Function," *Am. Rev. Resp. Dis.* **115**, 195–205 (1977).
23. D. Hattis and K. Silver. "Use of Biological Markers in Risk Assessment," in P. Schulte and R. Perera, (eds.), *Molecular Epidemiology: Principles and Practices* (Academic Press, New York, 1993), pp. 251–273.
24. A. A. Lai, R. J. Fleck, and N. B. Caplan."Clinical Pharmacokinetics and the Pharmacist's Role in Drug Development and Evaluation, in A. E. Cato (ed.) *Clinical Drug Trials and Tribulations* (Marcel Dekker, New York and Basel, 1988), pp. 79–98.
25. D. Hattis, "The Importance of Exposure Measurements in Risk Assessment of Drugs," *Arch. Toxicol.* (in press).
26. C. C. Travis and M. I. Land. "Estimating the Mean of Data Sets with Nondetectable Values," *Environ. Sci. Technol.* **24**, 961–962 (1990).
27. R. O. Gilbert. "Characterizing Lognormal Populations," in *Statistical Methods for Environmental Pollution Monitoring* (Van Nostrand Reinhold, New York, 1987), pp. 164–176.