Applied Occupational and Environmental Hygiene

ISSN: 1047-322X (Print) 1521-0898 (Online) Journal homepage: https://www.tandfonline.com/loi/uaoh20

Analysis of Strategies for Reconstructing Exposures

Nurtan A. Esmen

Analysis of Strategies for Reconstructing Exposures

Nurtan A. Esmen
University of Pittsburgh, Pittsburgh, Pennsylvania 15261; Present address: Esmen Research & Engineering, 2531 Wickline Road, Gibsonia, Pennsylvania 15044

Many retrospective epidemiologic studies require the reconstruction of exposures over long periods of time on the basis of incomplete or absent data. The modern requirements of acceptably accurate, quantitative exposure estimations for the construction of exposure–response relationships and extrapolation of estimated risks to untested regions suggest the existence of a conflict between the study requirements and the more or less qualitative reconstruction methodologies.

While the importance of exposure reconstruction is accepted, there is no structured analysis of the logical basis of the reconstruction process. In an attempt to analyze exposure reconstruction strategies (ERS) mathematically, a formal set of properties for ERSs are developed. These properties are presented in the form of three necessary and six desirable/undesirable general attributes of any exposure reconstruction process. Even a cursory analysis of these formal attributes suggests that a careful analysis of the exposure reconstruction method utilized in an epidemiologic study is necessary. Without such an analysis, the reconstructed exposures may lead to exposure–response relationships which are more hypothetical than real. Since no analyses of reconstruction strategies have been reported, the frequency of erroneous procedures for the exposure reconstructions reported in the literature is unknown. It is also shown that a significant amount of fundamental methodological research is needed to develop tests for the reliability, accuracy, and sensitivity of exposure reconstruction methods. **Esmen, N.A.: Analysis of Strategies for Reconstructing Exposures. Appl. Occup. Environ. Hyg. 6:488–494; 1991.**

Introduction

Today and in the future, researchers in occupational health have to deal with more subtle manifestations of exposure to potentially hazardous substances. As the subtle effects of exposure become more fully integrated into the milieu of chronic diseases that can result from many different causes, the relationships between the causative factor(s) through exposure and the outcome become highly complicated entities possibly spread over a long period of time. Epidemiologic studies are probably one of the most useful investigative methodologies in determining the relationship between exposure to a potentially hazardous substance and its subtle manifestations. The paradigm of occupational epidemiologic studies includes the estimation of exposures in a manner that facilitates the determination of an exposure (dose)–response relationship. In a well-designed, prospective experiment, the measurement of exposures and other related parameters may be assumed to be possible. In contrast, in retrospective epidemiologic investigations of diseases that have suspected causative or aggravating occupational factors, the exposures are normally estimated from the outcome of ***undesigned*** experiments vis-à-vis exposure measurement. In fact, many retrospective epidemiologic studies require the reconstruction of exposures over long periods of time on the basis of incomplete or absent data. More importantly, if a quantitative risk estimation is desired and if the dose–response relationship generated is to have a sufficient predictive accuracy, then the quantitative reconstruction of exposures ought to be acceptably accurate. Although the importance of exposure reconstruction is accepted by modern occupational health epidemiology, there is no structured analysis of the logical basis of exposure reconstruction.

The availability of nearly complete exposure measurement data over a given study period is such a rare occurrence that these fortunate occasions need not be considered here. The main concern, the reconstruction of exposures over long time periods in an everchanging environment, requires the use of estimation and modeling techniques that are more or less qualitative and, in most cases, unverifiable over the entire time span of the study. This suggests the existence of a conflict between the data requirements for the study and the available data-gathering or data-generating techniques.

In a study, a researcher may choose to assign exposures on the basis of a combination of estimation techniques such as employee interviews, mathematical modeling, extrapolation from a few existing exposure values to all exposures, or other similar techniques. By these choices, the researcher makes a strategy decision. At this point, it is essential to raise several important questions. Was the strat-

 1047-322X/91/0606-488$2.00/7

egy appropriate? Did the strategy produce results essentially better than a random assignment of exposures? Is one set of methodological exposure assessment strategies better than another feasible set? In order to answer these questions, it is necessary to develop a framework by which exposure reconstruction strategies can be compared. An acceptable strategy for reconstructing exposures, i.e., an acceptable plan for the selection of data-gathering and/or data-generating methods, inherently assumes that a number of methods are available for the reconstruction of exposures. Without a loss of generality, it may be further assumed that none of these methods produces impossible answers. In a very general sense, it is desirable to have a strategy that is accurate, sensitive, and efficient. Unfortunately, such generalized statements are not sufficient to characterize or compare strategies. Comparative analyses of strategies require a set of attributes that can be assigned to any strategy. Perhaps such attributes can be developed through an organized investigation of fundamental issues such as dose or exposure or misclassification of exposures. The following analysis is an attempt to analyze exposure reconstruction strategies mathematically. Although the analysis raises many questions and answers but a few, it is presented to stimulate discussion in order to provide a firm, logical basis for exposure reconstruction.

Mathematical Model of the Exposure Reconstruction Process

In the most general sense, exposure to an agent can be defined as the quantifiable presence of that particular agent at a specific moment in a manner or form such that the agent can come into contact with one or more target organs in the body of a specific receptor through one or more routes of contact, each with its own distinct set of characteristics of exposure to dose transformation. This generalized statement has important implications with respect to the definition of the functions that relate response to dose to exposure; however, without some restrictions, the estimation or measurement of exposures may become hopelessly complicated.(1) In an attempt to simplify the mathematical and logical formulations of exposure estimation or reconstruction, exposure will be more naively defined as: the quantifiable presence of an agent in an appropriate form with respect to a single route of entry for a specific person, at a specific time and activity class (e.g., work, leisure, etc.). The application of this concept to a work-related exposure suggests that, at any given moment, the exposure of a worker is a result of a large number of variables, e.g., the task the worker is performing, the generation of the contaminant, the air exchange where the worker is, and so on. These variables, each dependent on the worker and on time, define the deterministic expression of exposure for that worker as a function of time. The exposure values which result from this function evaluated over specific intervals of time constitute a characteristic set of exposures for the worker.

Consider an arbitrary number of contemporaneous workers judged to be members of a group due to a shared characteristic. From the description of exposure above, each worker contributes his/her characteristic exposure set to the set of exposures which is the characteristic of this group. The set of exposures which constitute the combination of all characteristic worker exposures is said to be the true group exposure set, {S}.

Suppose that the true group exposure set for a group of workers is not known. If all the variables that generate the exposure values for each worker are known precisely for all workers, then, at least theoretically, the true group exposure set can be reconstructed precisely. Suppose that some of the variables either are not known or can only be estimated with an arbitrary precision. Under this supposition, the reconstructed group exposure set, {B}, will be such that the probability of some of the members of set {B} not being included in set {S} is not necessarily zero. Consequently, assigning an exposure value b from set {B} to a worker from the group carries a probability that b is not a member of the true exposure set {S}. Therefore, the analysis of an exposure reconstruction process may be viewed as a systematic consideration of the properties of the exposure set {B} and the probability of the members of {B} also belonging to set {S}. This probability can only be estimated if the distribution of exposures within set {S} is known. As discussed below, estimation of the probabilistic relationship between the reconstructed group exposure set and the true group exposure set is one of the difficult problems of the analysis of exposure reconstruction strategies. The generalization of this model for multiple routes of entry or many groups or many agents is straightforward.

Comparative Attributes of Exposure Reconstruction Strategies

In order to make normative statements on a strategy, it is necessary to develop a formal set of properties for strategies. In addition, there must be a way to attach a value or utility to each property. These two conditions may be fulfilled by defining three classes of attributes for exposure reconstruction strategies (ERS). If an attribute of an ERS is such that, in its absence, the reconstructed data are rendered unusable, then the attribute is said to be a necessary attribute (NA).

The formulation of the first necessary attribute (NA1) for an ERS is rather straightforward:

NA1: An ERS must produce at least two exposure classes.

The necessity of this attribute is obvious. It specifies that, at the end of the exposure estimation, the results can be used in an epidemiologic study. However, in and of itself, this attribute does not say anything about the desirability or undesirability of the strategy utilized or the usefulness of the results obtained.

In order to examine questions involving the usefulness of generated exposures, consider a study where the specific assumptions with respect to the responsibility of the agent considered in exposure reconstruction vis-à-vis the observed disease outcomes cannot be made on biological, physical, or chemical grounds (indistinct study). In such a study, it would be necessary to find the proper causative agent by the end of the study so that the results of the study can be generalized. The implications of this statement may be illustrated by the following hypothetical example. Suppose in an epidemiologic study, particles of a chemical, which coexist with a number of other nonparticulate matter, are selected as the possible causative agent for respiratory irritation. Further suppose that the degree of respiratory irritation in each subject is correctly classified, while the estimated exposure levels of the particles have a certain unknown rate of misclassification. If the study shows a positive relationship between exposure to the particles and respiratory irritation, then it is usually claimed that the chemical is a causative agent of respiratory irritation with a positive slope of the dose–response curve. Any interpretation or extrapolation of this claim requires several important hierarchical assumptions. The first assumption is that the chemical is indeed a causative agent and not a surrogate for a coexisting agent, e.g., an irritant gas. It is important to note that the surrogate is neither a modifier nor a confounder; i.e., the surrogate itself does not produce the observed response. If the first assumption is incorrect, then in other situations involving the chemical but not the irritant gas, the assigned risks would be fictitious; conversely, in situations involving the irritant gas but not the chemical, risks would be neglected. Furthermore, if it can be assumed that the chemical is very likely to coexist with the irritant gas, then the direct extension of the results of the specific epidemiologic finding to general conditions requires the assumption that the relationship between the chemical and the irritant gas is linear with a positive slope, and the absence of the chemical implies the absence of the irritant gas. Under this assumption of linearity, the estimated risk may differ from one condition to another, but the dose–response trend will be preserved. If the linearity assumption is not fulfilled, then without the specific knowledge of the functional relationship between the chemical and irritant gas, it would not be possible to make a valid claim with respect to risk or dose–response beyond the observations of the first epidemiologic study. If the relationship between the causative agent and the surrogate used in the exposure reconstruction has a negative slope, then it is possible to obtain exposure values where the low exposures would be a better indicator of response than the high exposures. Clearly, such an exposure reconstruction would not be useful. In fact, a true exposure–response relationship may be obscured and a real causative agent may be concealed. Thus, in an indistinct study, the necessary attribute which prevents the interpretive errors may be stated as:

NA2: In an indistinct study, an ERS should enable the researcher to select the proper agent(s) for exposure reconstruction independent of the *a priori* choice of the agent(s).

In comparison to a random assignment of exposures, if reconstructed exposures by a given procedure are equally or less well correlated with the true exposures, then it might be claimed that the exposure reconstruction process used was, at best, a waste of time. The formal statement of an attribute that considers this aspect of exposure reconstruction strategies may be developed as follows. Ideally, for a selected agent and a selected group of workers in a specific workplace, all existing information can be used to form the basis of a complete set of reconstructed exposure data by the use of all theoretically possible models and methods. The data on the exposure-generating parameters, theoretical models, and subjective and objective methods of estimation constitute an all-inclusive set of all-knowable truths (available data set) and all-possible conjectures (conjectured data set). In order to remove some theoretical difficulties, it will be assumed that the conjectured data set is finite. Let the proportion, p, be defined as the proportion of the inclusive data set which is in the available data set (see the Appendix). The reconstructed exposure set as a function of the proportion p, $\{B(p)\}$ can contain values which are not members of the true exposure set, $\{S\}$. Thus, a probabilistic measure of correctness, (C), may be defined by the probability of the expected value of $\{B(p)\}$ to be within the true exposure set $\{S\}$:

$$C = \Pr\left(E[\{B(p)\}] \varepsilon \{S\}\right) \quad (1)$$

It may be noted that as C approaches zero, most or all of the set $\{B(p)\}$ is outside of set $\{S\}$; conversely, as C approaches 1.0, most or all of $\{B(p)\}$ is contained in $\{S\}$. Thus, the reliability (R) of an ERS may be defined as:

$$R = \frac{C}{1 - C} \quad (2)$$

It should be noted that if $R = 1$, then the ERS predicts as a random assignment of exposures from a plausible interval of values, and if $R < 1$, then the ERS performs worse than a random assignment. These general considerations lead to specific definitions of a necessary attribute.

NA3: At least for some $p < 1$, an ERS must produce $R > 1$ almost everywhere (a.e.).

This necessary attribute specifies that the reconstructed exposures should behave better than a random selection of exposures from a plausible interval ($R > 1$) and requires that a solution can be found without a complete knowledge of the parameters ($p < 1$). The phrase, "at least for some p," allows for the existence of a range (however small) for the proportion of parameters needed to be known in order to satisfy this attribute. The mathematical notation "almost everywhere" means that when NA3 is satisfied, the set of those reconstructed values for which $R > 1$ or $p < 1$ is false is a set of measure zero (see Appendix).

Any strategy with all the necessary attributes will ascribe an inherent utility for the reconstructed data. Consequently, if the presence of an attribute enhances the utility of the reconstructed data, then it is said to be a desirable attribute (DA) and if the presence of an attribute detracts from the utility of the reconstructed data, then it is an undesirable attribute (UA). It must be noted that the negation of a DA is not necessarily an UA. The DAs and UAs relate to more subjective attributes such as efficiency, accuracy, sensitivity, etc. In order to list these attributes, several definitions given below are necessary. In the reconstruction of exposure data, it may be intuitively suggested that only the exposure-generating parameters with non-negligible probabilities of occurrence would be needed to obtain almost all of the information necessary for successful reconstruction of exposures. Consequently, the evaluations which use the conjectured data set may be dominated by a relatively small number of subsets (consequential subsets).

DA1: An ERS that can be reduced to the estimation procedures and/or parameters of the consequential subset is said to be efficient.

This attribute suggests that if an ERS can identify whether or not a procedure or a parameter would contribute a negligible amount to the final estimation, then the researcher would be able to carry on the reconstruction using the minimum number of operations without that procedure or parameter. An example of this can be given. In general, the dependency of an exposure to a chemical agent is not significantly influenced by barometric pressure. Although barometric pressure is a parameter, in most cases, it may be shown to be "not a member" of the consequential subset.

DA2: In an efficient ERS, if each step in the reconstruction process after the first n steps contributes an equal or less amount to the magnitude of the reconstructed variable than the previous selection, then the ERS can be optimized.

If an ERS has this attribute, the order of operations may be selected to achieve reconstruction in the most convenient way. If an ERS can be shown to be efficient, a simple simulation using arbitrary ranges for each conjectural step may be used to quantitate the contribution of each step to the magnitude of the estimated exposure value.

DA3: If for an ERS there is a $0 < a < 1$ and $a < p < 1$ implies that $R >> 1$ a.e., then the ERS is accurate.

This desirable attribute suggests that in the ERS used, if the available parametric data (p) are in excess, a proportion "a" results in reconstructed exposures that are much better than a random assignment; then it can be expected that for any $p > a$ used in a reconstruction strategy, the correlation between reconstructed exposures and true exposures will be very high. As discussed later, the determination of the critical proportion, a, and testing for this attribute is problematic. Even so, it may be stated, at least theoretically, that an ERS with this attribute would be able to predict the exposures accurately with some available data. Of course, the smaller the value of a, the less available data would be required.

It is realistic to assume that the reconstructed exposure level data set for any p would be bounded. Obviously, exposure levels cannot be less than zero; thus, there exists a lower bound, and obviously, physical laws would require a less than infinite exposure levels. Similarly, exposures are also bounded by zero, and they are also less than infinite. Consequently, the span of reconstructed exposures for a given proportion p, I(p), may be defined as the difference between the maximum and the minimum of the reconstructed exposure data. This will lead to another desirable attribute.

DA4: If the magnitude of I(p) determined by an ERS is small for some p, then it is sensitive.

This desirable attribute suggests that the reconstructed exposures for the group constitute a narrow band and the differentiation between groups of workers with different exposures would be easy because the likelihood of overlap would be small.

UA1: If an ERS is accurate only when it is not sensitive, or sensitive when it is not accurate, then it is said to have low utility.

This attribute is undesirable because it suggests that when one can differentiate between groups of workers easily, then the reconstructed exposures are not accurate as defined above. Conversely, when the reconstructed exposures are accurate as defined above, the possibility of overlap between groups of workers may be significant.

UA2: An ERS with the negation of DA3 is equivocal.

This attribute is undesirable because it suggests that, without complete available data, it may be impossible to obtain a reconstructed exposure data set much better than a random assignment.

The attributes given above were developed for the initial theoretical analysis presented here. Therefore, these attributes should be considered as a preliminary set that can be used as a starting point in future theoretical discussion of the analysis of exposure reconstruction strategies. It is also important to point out that an analysis of a given strategy using these attributes may not be readily apparent because many of the probabilistic statements included in the analysis may have to be estimated. These caveats do not detract from their utility in the analysis of ERSs because, to this author's knowledge, few exposure reconstruction processes hitherto reported have been analyzed with any rigor to provide the researchers with a qualitative or quantitative sense of confidence in the results.

One reported, in-depth analysis[2] of a manufacturing process-based estimation strategy as compared to an expert panel-based estimation strategy showed a significant difference between the two strategies. The process-based strategy for that particular situation was remarkably effective. With the lack of confirmation of the exposure reconstruction, unfortunately all too often, the exposure reconstruction results have been used in epidemiologic studies with assumed accuracy and reliability that may not be warranted.

Analysis of Exposure Reconstruction Strategies

In the reconstruction of job-specific exposures resulting from industrial processes, the researcher is faced with a large list of exposure parameters, e.g., process characteristics, process rate, ventilation, tasks which define the jobs, distribution of tasks, neighboring jobs, and the use of personal protection among other possible parameters. These parameters can be listed in great detail and precision even including plausible, but less likely, parameters, e.g., intentional exposure. Each of these exposure parameters may contribute to the reconstructed exposure. Clearly, this assertion requires the restriction that the functional relationship between exposure and exposure parameters is known. For many parameters, such relationships either may be known for general cases or may be reasoned from basic principles. If the functional relationship is not known and cannot be developed from basic principles, the researcher either has to make an *a priori* assumption that the parameter's contribution is not important or has to develop such a relation by one means or another (empirically, intuitively, at random). It must be noted that if the parameter cannot be ignored, and if there is no reasonable way to find the functional relationship between the exposure and the parameter, then NA3 or DA3 may be violated with no possible method of verification of the reconstructed exposure.

The tools available for estimating the values of each parameter include physical modeling, mathematical modeling, the use of general correction functions, existing data on some parameters, existing data on some surrogate parameter(s), employee interviews, similarity with other processes (professional judgement), and similarity with historically documented practices. The relative strengths and weaknesses of these procedures, their applicability, and the specific methodologies to be used in a study are governed by unique factors associated with each study. Although the properties of each method will make an important contribution to the de facto analysis of an ERS, in the more general theoretical sense, they do not affect the process of ERS analysis. The general analysis centers around comparing the properties of an ERS that uses these processes with the comparative attributes of other ERSs before the time-consuming and costly reconstruction process is undertaken. In doing so, one seeks answers to fundamental questions such as: Can the study be done? What quality results do we expect? Can we improve the strategy? Obviously, the first question to ask is whether the study can be carried out, i.e., does the ERS have all three necessary attributes? The theorem which enables the construction of a test for NA1 is given below without proof. It may be noted that the theorem is not restricted to estimated exposure levels; it may be applied to differences in exposure agents. In the latter case, the maximum and minimum used in the theorem would be treated in the set theoretic sense.

Theorem I

Let $\{Q\}$ be a subset of exposure parameters that define the job of a set of workers (e.g., task performed, etc.). The members of this subset, q_i, take on the values 0 or 1 depending upon the absence or presence of the parameter. Let $b(q_i)$ be the reconstructed exposure with b_{max} and b_{min} the maximum and minimum values of the reconstructed exposure, respectively. Then, the necessary condition NA1 is satisfied if, and only if, there is at least one $q_k \in \{Q\}$ such that

1. $q_k = 0$ for at least one job and $q_k = 1$ for at least one job.
2. $b_{max}(1, q_{i \neq k}) < b_{min}(0, q_{i \neq k})$ or $b_{min}(1, q_{i \neq k}) > b_{max}(0, q_{i \neq k})$.

The first condition specifies that at least one parameter is not shared between jobs. The second condition requires that the presence or absence of the nonshared parameter determines different levels (types) of exposure. A convenient test procedure utilizing this theorem is to assign guess-estimated ranges for all parameters that belong to set Q and test for the conditions provided by the theorem. While the test for NA1 is simple, testing for NA2 may be problematic. A distinct study satisfies NA2 directly, but an indistinct study requires extensive inquiry into such questions as: What agents (parameters) can be isolated? What are the correlations between the agents (parameters)? Can exposures for each identified agent or parameter be estimated? Quantitative tests for these types of questions are not known. The issues which can result from the careful examination of the questions raised may lead to some resolution, if their resolution is sought *both* by the exposure reconstruction team and by the epidemiologic analysis team.

Analytically, NA3 is also problematic. While the condition specifies that the reconstructed results must be better than a random generation of the results within a plausible interval, the basic premises require the knowledge of the probabilities for the conjectures made in the reconstruction of results. Even if the condition is weakened by replacing the "almost everywhere" requirement by a relaxed criterion (which includes hard to define terms such as "substantially" or "by and large"), the primary question becomes one of testing conjectures. At any point in the reconstruction process, the results can be partially tested by a number of methods including existing exposure records, re-creation of the processes in the laboratory, or prediction of exposure followed by an exposure assessment. These tests, in and of themselves, can be problematic. This may be illustrated by an example. Suppose that the only available method for the evaluation of an exposure parameter is an employee interview and the only verification test that can be performed is comparing the interview results to a recent or current actual evaluation of that parameter. Such a case would be wrought with questions. If the attribute is trivial with respect to the person's life, such as fraction of time spent in doing a task, how would one project the results of a short-term recall of a trivial event to a point far in the past? Can the degree of long-term recall of personally important events be used as a criterion to project a short-term recall of a trivial event to a point far in the past? Does a change in the attitude toward the job over time influence the results? Without answers to these and similar questions, any decision to accept or reject the procedure is arbitrary, because the basis of the NA3 comparison is not how much confidence one can attribute to the results but are the results almost everywhere ("substantially," "by and large," etc.) better than assigning a randomly determined value to each reconstructed point. For almost any estimation method, questions along the same lines can be raised. Many of these questions, unfortunately, do not have answers. Future research efforts which focus on the careful analysis of each exposure reconstruction method and performance of specific, controlled experiments designed to answer the questions raised would contribute significantly to the solution of the difficulties associated with NA3.

If the exposure parameter set is large, then the sheer number of calculations may result in an unrealistically lengthy process. If the ERS selected is an efficient one, then theoretically, the number of parameters considered may be reduced. The parameters which pertain to the description of jobs such as the tasks, occurrence of tasks, etc., can be reduced readily. The remaining set may be analyzed by somewhat weakening the DA1. If a researcher is willing to make certain assumptions, then the efficiency may be defined as the reduction of the parameter set to a consequential set under the assumptions made. The generation of the efficient set may be achieved by a sequential procedure. In this procedure, the assumptions are related to the guess-estimated ranges for all of the parameters. By using these ranges, exposures may be reconstructed for several points in time and for several jobs. The second reconstruction of exposures in the absence of each parameter singly may be compared to the initial set of reconstructed exposures, and the decision to keep or reject the parameter can be made by comparison to an *a priori* criterion. This procedure will also provide means of choosing the sequence of reconstruction operations which can be used to test the conditions of DA2.

If the preliminary analysis shows no or insignificant reduction in the number of parameters to be considered in the final analysis, the researcher has to make one of the three choices:

1. Accept the number of parameter as reasonable.
2. Declare the project to be too complicated to analyze meaningfully.
3. Reconsider the guess estimates.

The second choice suggests that under the assumptions (guess estimates) the researcher is willing to make, the reconstruction of exposures is not tractable and another facility or situation must be chosen or the project must be abandoned. Although such a decision may not be convenient, it would be scientifically correct. The third choice suggests that the researcher is willing to change the initial assumptions and reiterate the analysis. For obvious reasons, if the third choice is accepted, then a new set of guess estimates must be made for all of the parameters without consulting the first set of estimates or calculations. This process can be repeated as many times as necessary until the parameters are accepted or the project is declared intractable.

The problems associated with the analysis of an ERS with respect to the remaining four desirable and undesirable attributes involve the difficulties associated with the analysis of NA3. The data necessary in constructing possible test procedures require a detailed analysis of the reconstruction methods with respect to their predictive capabilities. Unfortunately, such studies are lacking. A detailed analysis of the data requirements for the proper description of each method would be beyond the scope of this article. It may be suggested that 1) carefully controlled experiments where the exposures are predicted before measurement, 2) compilation of current industrial exposures in conjunction with as many influences as can be recorded, and 3) testing of reconstruction processes with the few large data sets available may provide sufficient information for the construction of the needed analytical tools.

Discussion and Conclusions

The logical analysis of exposure reconstruction suggests that, in an epidemiologic study without a careful analysis of the exposure reconstruction strategies, the exposure (dose)–response relationships generated may be much more hypothetical than real. More importantly, without such an analysis, the probability of the truth of the conjectures included may not be clear. This is a troubling observation because many of the answers sought are not hitherto available, as was seen in the theory presented above. Fortunately, this statement does not imply that epidemiologic studies are generally unreliable. Instead, its clear implication is that, without extensive verification of the exposure estimation, the dose–response found in an epidemiologic study may be in significant error and extrapolation of the results may be unreliable. The converse of this statement can also be true. Consequently, the possible lack of an observed exposure–response relationship may be due to the errors in the exposure estimation and, thus, may obscure a relationship which is, in fact, present. In this context, it is proper to challenge the operational paradigm of industrial epidemiologic studies. Usually, an epidemiologic study is designed on the basis that exposures can be characterized as needed, and frequently, the exposures have been reconstructed as needed on utilitarian, but not necessarily rigorous, grounds. Without extensive reanalysis of these reconstruction strategies, it is not possible to estimate the frequency of occurrence of erroneous exposure reconstruction. The resolution of this problem may be sought in reversing the process of experimental design so that the epidemiologic design is selected on the basis of what exposure parameters can be satisfactorily defined and determined.

Some, perhaps most, of the questions raised in this paper do not yet have practical or theoretical answers. Considerable theoretical and applied research in defining the properties and limitations of exposure reconstruction methods will be needed to answer some of the important concerns of methodology. Extensive method verification of estimation techniques such as employee interview, adjustment of exposures on the basis of process modeling, and limitations of process modeling are lacking, and it is doubtful that significant advances in exposure reconstruction can be made without considerable effort in this area. The theoretical framework proposed here can serve as a basis for discussions on the subjects relating to the reliability, accuracy, and sensitivity of the exposure side of the equation and perhaps predicate the future developments in epidemiologic studies where the epidemiologic methodologies are developed in keeping with the limitations of the exposure reconstruction.

References

1. Esmen, N.A.: Limitations on Dose Estimation. Environ. Health Perspect. 42:3–7 (1981).
2. Greife, A.L.; Hornung, H.W.; Stayner, L.G.; Steenland, K.N.: Development of a Model for Use in Estimating Exposure to Ethylene Oxide in a Retrospective Cohort Mortality Study. Scand. J. Work Environ. Health 14(Suppl. 1):29–30 (1988).
3. Taylor, A.E.: General Theory of Functions and Integration, p. 232 ff. Blaisdell Co., New York (1965).
4. Kestelman, H.: Modern Theories of Integration. Oxford University Press, Oxford (1937).
5. Loeve, M.: Probability Theory, 3rd ed. Van Nostrand, Princeton, NJ (1963).

APPENDIX

Mathematical Definitions

Almost Everywhere (a.e.):[3] If a mathematical property holds for every point of a region except for a countable set of points, it is said to hold almost everywhere. This is because every countable set is measurable and has a Lebesque measure of zero. The inclusion of a.e. in the definition of the attributes was for the benefit of the theoreticians who may want to consider expanding the ideas proposed in this paper. The general reader and the practitioners can ignore this mathematical nicety without a loss of generality. The readers who are interested in more information on the mathematical concepts used here are referred to the references 3–5.

Proportion, p: Let $\{A\}$ = available data set
$\{G\}$ = conjectured data set
$N(\{X\})$ = cardinal number of a set $\{X\}$

Then $p = N(\{A\})/N(\{A\} \cup \{G\})$.