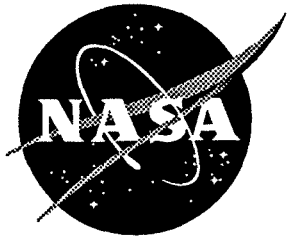
NASA Contractor Report 194950



A Review of an Updated Synthesis of Noise/Annoyance Relationships

James M. Fields
Georgia Institute of Technology, Atlanta, Georgia

Contract NAS1-19061
July 1994

National Aeronautics and
Space Administration
Langley Research Center
Hampton, Virginia 23681-0001

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

This report evaluates the evidence supporting the "dosage-effect relationship for the prevalence of annoyance due to general transportation noise" that was originally presented in 1978 (Schultz, 1978) and more recently updated in 1991 (Fidell, Barber and Schultz, 1991). After examining the synthesis methods and many of the constituent studies' reports the following conclusions have been reached:

1 The best objective description of the synthesis curve is that the synthesis curve represents one research group's judgements about the evidence from 29 data sets (26 surveys) on the relationship between a high degree of annoyance and Day-Night Average Sound Level (DNL).

2 The synthesis curve does not have a clear, definite meaning because it is not based on objective, reproducible study techniques. The curve is NOT a measurement of the relationship between DNL and the percentage of the population that would describe themselves as "highly annoyed".

3 The inaccuracies and ambiguities in the two synthesis publications are sufficiently important that the synthesis articles should not be relied upon for information about the individual surveys or for a description of the synthesis methodology.

4 The number and importance of the unexamined differences between the surveys' annoyance scales, noise estimation procedures and study conditions are too great to provide an accurate comparison of the dose/response relationships in different surveys.

5 Departures from standard methodological principles for selecting study data mean that the results do not provide information about the degree of similarity in dose/response relationships.

6 The effect of the methodological weaknesses is not known. It is not known whether a more rigorous methodology would give a different estimate of the relationship between noise level and empirically equivalent measures of high annoyance.

7 The single, most difficult obstacle to an objective synthesis is the absence of transfer functions for linking divergent annoyance scales.

1. INTRODUCTION

Over 300 surveys of residents' annoyance with environmental noise have been conducted (Fields, 1991). The interest in the relationship between noise level and residents' annoyance responses has lead to a number of studies that have attempted to compare and combine the results from several community surveys. Most of these have focused on only a small number of surveys. For example, Alexandre examined five surveys (Alexandre, 1973) and a U.S. Environmental Protection Agency document examined four surveys (Information..., 1974). However none of these studies have provided a widely accepted basis for summarizing a general relationship between noise and annoyance.

Two sequential attempts have been made to examine a larger number of surveys and to derive a universal dose/response relationship from those surveys. The authors described these studies as providing a "synthesis" of the results from the individual surveys. The first (1978) synthesis article reported the results from 12 data sets and produced the widely-know "Schultz curve" (Schultz, 1978). The updated synthesis (1991) added 17 data sets to the analysis to provide a total of 29 data sets (Fidell, Schultz, and Barber, 1991). Both of the synthesis articles present a relationship between Day-Night Average Sound Level (DNL) and the percentage of the population that is classified as "highly annoyed". The current report provides an assessment of the 1991 updated synthesis and thus implicitly of the original 1978 synthesis.

The data that these synthesis studies analyzed are plotted in Figure 1 together with three curves that have been used to summarize the data points. Two of the curves come from the two previously referenced synthesis efforts while one comes from a more recent analysis of a subset of the 29 data sets (Federal..., 1992: 3-6). This later analysis is not evaluated in this report. Appendix D provides more information about these curves.

1.1. Objective of this review

The present assessment is undertaken for three purposes: (1) to determine what can be concluded from the 1991 synthesis, (2) to provide assistance in understanding and using the information provided in the synthesis articles, and (3) to evaluate the methodology that was used for the synthesis. The present assessment does not attempt to comprehensively assess the accuracy of all of the information in the articles, nor does it attempt to derive a dose/response relationship.

1.2. Conventions followed in this review

It is often difficult to understand the structure of the data that has been reviewed in the synthesis articles. The following terms should clarify the discussion:

In this review, as in social science generally, survey refers to a single, coordinated data collection exercise that may have been conducted at a single time (cross-sectional design) or at several different times in the same locations (panel design). For this particular article the term data set refers to a set of data points that have been identified as a separate entity in the synthesis publications. Usually a single survey yields a single data set. However, this is not always the case. The Canadian Aircraft/Traffic comparison survey (CAN-168) for example is counted twice in the 1991 publication, once for one aircraft noise annoyance and once for road traffic noise annoyance. The synthesis articles are somewhat confusing because they often refer to multiple data sets from the same survey as multiple surveys.

Although most surveys can be identified by their titles, each survey is also identified in this article by a unique two-part catalog identification number (Fields, 1991). The first three letters abbreviate the country name and the last three numbers are a simple sequential identification number. The catalog also contains longer, unique titles for each survey. In this report, however, the shorter titles used in the synthesis publications are used even though these titles do not distinguish between all the surveys in the catalog.

An individual annoyance response is one respondents' evaluation of the annoyance due to one noise source. Although a single respondent (i.e. person) usually contributes only a single annoyance response to the synthesis, there are surveys in which a single respondent provides multiple individual annoyance responses usually by either rating several noise sources separately or by participating in more than one phase of the same longitudinal study. In the synthesis publication, the term "interview" is sometimes used to indicate that there are multiple individual annoyance responses.

Individual annoyance responses are aggregated into groups that are described as data points in the synthesis analyses. The data points vary in numbers of interviews and in their structure. These groups of individual annoyance responses are actually formed in any of four different ways: (1) study area -- responses are aggregated within a single neighborhood or other study area; (2) noise category -- responses come from within the same noise exposure interval but are drawn from several different study areas; (3) study area by phase -- responses from a particular study area during each interviewing phase are aggregated separately (e.g. the Burbank study (USA-203) generated 20 data points because interviews were conducted in four areas at five different times); (4) noise category by type of noise descriptor -- the same responses are presented twice, once for each of two definitions of the noise exposure (Second Heathrow Survey, UKD-042).

2. DESCRIPTION OF THE DATA

The two synthesis articles analyze the relationship between estimated values of DNL and the percentage of highly annoyed individual responses in each data point. This section will present evidence about each of the three elements of that relationship for each survey: (1) the definition of the data point, (2) the measurement of annoyance and the division point for high annoyance, and (3) the estimates of the noise exposure, DNL. The section concludes with an enumeration of some of the types of errors and misunderstandings present in the synthesis articles.

Table 1 provides summary information for the 29 data sets that were labeled "clustering" data sets in the synthesis articles. The table does not list the nine data sets that were labeled "non-clustering" data sets in the 1978 synthesis article. The 29 data sets in Table 1 are drawn from the results of 26 surveys. Three surveys provided two data sets each because respondents rated two noise sources in each survey (Catalog Identification Numbers CAN-168, SWE-165, SWI-053). One survey (UKD-008) appears once in this list but was also described as a "non-clustering" survey for a different division of the annoyance scale.

The data sets in Table 1 are grouped by the type of annoyance scale. The first groups of annoyance scales come closest to directly asking about "high" annoyance. The last groups use quite different concepts.

The first columns of Table 1 provide three methods for identifying each data set. The first column gives the name for the survey that was used in the synthesis articles. The primary bibliographic reference for each study is provided in parentheses in this column. The second column references the synthesis publication by year and page number ("1978" for Schultz, 1978 or "1991" for Fidell, Barber and Schultz, 1991)). The third column provides a cross-reference to a unique identification number to differentiate the survey from all other surveys in a comprehensive list of surveys (Fields, 1991).

The fourth column gives the total number of individual responses that are included in the synthesis analysis. If the number reported in the synthesis publications was incorrect, a (C) is entered to indicate a "Correction". Such corrections are often minor. A note is added if the number of responses is greater than the number of respondents. The number of respondents is not reported for these studies. Multiple responses from a single respondent occur either because residents were interviewed at intervals to determine whether their views had changed (two data sets) or because attitudes towards two different noise sources are reported separately (six data sets drawn from three surveys).

2.1. Definition of data points

Three of the columns in Table 1 describe the data points. The synthesis articles imply that each data point represents a study area. In fact a data point can represent any one of the four types of groupings of respondents that were defined earlier in Section 1.2. Each type is noted in the "Type of data point" column in Table 1. The most frequent type is a noise category group in which the responses come from within the same noise exposure interval but may be drawn from many different study areas. Other data points represent study areas (responses are aggregated within a single neighborhood or other study area), area-by-phase groups (responses from a particular study area during each interviewing phase are aggregated separately), or category-by-descriptor (the same responses are presented twice: once for each of two definitions of DNL).

The number of individual responses that are represented by each data point varies from 1 to at least 1,186. The range for each survey is given in the "Data point.. sizes" column in Table 1.

The number of data points for each survey is provided in a separate column. Since the synthesis article gives each data point an equal weight in the analysis (regardless of the number of interviews), the number of data points, not the number of interviews, determines the impact that each particular survey has on the synthesis.

2.2. Measurement using annoyance scales

The synthesis article includes surveys that have used four fundamentally different methods to measure attitudes: verbal rating scales, numerical rating scales, multiple-item indices and ranking scales. The four types are indicated in the first of the three columns under the "Annoyance measurement information" heading in Table 1. The data sets in Table 1 are grouped by these four major types and, within major type, by subtypes. The same groupings are repeated in Table 2 when the exact wording for each survey annoyance question is presented.

Verbal rating scales The verbal scales are all based on questionnaire items that present respondents with a choice of a limited number of words (from three to eleven in these surveys) to typify their degree of annoyance. These verbal rating scales differ in the question that is asked, the number of words (points) from which the respondent can choose, the valance of the words (i.e. whether both positive and negative descriptors are offered), the actual words that could be chosen, and the use of an annoyance screening question. Some of these distinctions are highlighted by the first seven subgroupings of data sets in Table 1 and Table 2.

The single survey (AUL-210) under the sub-heading "Highly annoyed Verbal Scale" is the only survey in which the word "highly" was offered as a response alternative to respondents. The exact wording of the survey question is presented in Table 2. When the synthesis

authors refer to the percentage "highly annoyed", they are not, except for this single survey, quoting the survey questionnaire, but instead providing the synthesis authors' own term to typify such words as "considerably", "very", and "extremely".

The single survey under the heading "Single Verbal 3-Point Scale" (DEN-075) asks a single question in which the respondents are presented with three alternatives (including "Not at all annoyed") to describe their degree of annoyance. Unfortunately, as is indicated in the "Comments" column, the synthesis authors misunderstood this question. The question does not ask about the traffic noise (as the authors thought) but rather about noise generally.

The two surveys under the heading "Single Verbal 4-Point Scale" (FRA-016, UKD-116) ask a single question in which the respondents are presented with four words (including "not at all annoyed") to describe their degree of annoyance. Although this scale is represented by only two surveys in the synthesis articles, it has been used in many European surveys making it the most widely used question in the world.

The seven data sets (six surveys) under the heading "Joined Verbal 1+3-Point Scale" each draw on the combination of answers to two questions. First respondents are asked whether they are annoyed and then, if they are annoyed, which of three verbal descriptors best describes their degree of annoyance. While both this and the preceding Single 4-Point Scale yield four points, the answers might differ if the respondents are influenced by the number of alternatives from which they may choose at a single time. Choosing "not annoyed" when there are only two alternatives may have a different meaning than choosing "not annoyed" when there are four alternatives.

It is important to remember that the initial screening question in the 1+3-Point Verbal Scale concerns annoyance not audibility. Many surveys have initial screening questions about whether a respondent can hear a noise source. Whether or not a "hearing" screening question was used is not noted in the tables. The use of a "hearing" question is expected to be unimportant because only a small percentage of the respondents report that they can not hear the rated noise sources in these surveys. These annoyance scales do not distinguish between respondents who can not hear a noise and other respondents who hear the noise but are not annoyed.

The next two types of scales in Table 1 parallel the previous two types except that now there is an additional scale point for either the single question (Single Verbal 5-Point Scale) or for the second of the two joined questions (Joined Verbal 1+4-Point Scale). Some of these scales used the same words (for example "very") that were used in the previous 4-point scales. If respondents respond to both the number of scale points as well as the word offered, then the same word "very" could be capturing a different degree of annoyance for the two types of scales.

The "Bipolar Verbal 5/9-Point scale" appears in three data sets from two surveys (CAN-121, CAN-168). All of the previous scales (unipolar scales) asked about only negative reactions to

noise. These three data sets also record positive reactions to noise. The "5/9" designation indicates there is some ambiguity in the number of scale points that are most relevant for characterizing the scale. The respondents choose from nine points but only five of these represent the types of negative reactions that are similar to those offered in the previously described unipolar scales.

Numeric scales The numeric scales provide verbal labels for only the end points of a scale and then allow the respondent to indicate the distance between those end points. The psychological distances between intermediate points on the scales are no longer confounded by the verbal labels for intermediate points. However, the verbal labels assigned to the end points could still affect the meaning of the ratings in ways that could only be estimated from empirical studies.

The first five "Numeric scales" all provide respondents with numbers to choose their degree of annoyance. The last two numeric scales provide respondents with an unmarked line and ask respondents to mark their position between the labeled endpoints of the line.

Multi-item indices The first three multi-item indices represent composite scores that are computed by assigning numerical values to the answers on each of several questions and then combining the scores from these constituent questions to derive an overall index score. The three indices are derived from answers to 6 to 37 questions. Most of the constituent questions are activity interference questions. An examination of the original German text determined that the scale for the Munich aircraft study (GER-034) is based on a multi-item index, not on a verbal scale as stated in the synthesis article. The Munich survey's index was derived by first combining the 37 questions into 11 variables and then computing the index score from a weighting of these 11 variables (Deutsche Forschungsgemeinschaft, 1974).

The meaning of the 2-item index in Table 1 (DEN-200) is somewhat clearer; a respondent is considered to be highly annoyed if the respondent either: (1) volunteers that the noise source is a problem, OR (2) answers "very much annoyed" on a verbal scale after giving certain responses to several screening questions. As a result of accepting this index a respondent who volunteers noise as a problem but who expresses only a little annoyance is classified as being "highly annoyed" in the synthesis article. The scale thus includes some of the same characteristics as the ranking question that is described under the next heading. The meaning of the index is further complicated by the fact that three different screening questions affect whether an individual is asked about the degree of annoyance.

From the composite scores on these multi-item indices it is not possible to determine how the respondents answered each question. This makes it impossible to directly equate these answers with the answers to the other scales.

Ranking scale One survey (FRA-041) uses an entirely different scale. Respondents rank-order 10 aspects of the neighborhood. If road traffic noise is the biggest problem then a respondent is counted as highly annoyed. This approach means that the answer depends not

only on the level of annoyance with traffic noise (the annoyance that is measured in the other scales) but also on the annoyance with any other aspects of the neighborhood. A respondent who is basically satisfied with all aspects of the neighborhood and who finds traffic noise only a little annoying could be scored as "highly annoyed" using this scale. The synthesis publications include this as a clustering survey, but note that this is an "... entirely different kind of scale, not an annoyance scale at all."(Schultz, 1978: 381)

2.3. Division point for "high" annoyance

In order to attempt to compare the answers to the different types of annoyance questions the synthesis has dichotomized each scale and labeled the upper part "high" annoyance. This division point is based on the synthesis article authors' judgement. The basis for this judgement is indicated for each data set under the heading "definition of high annoyance".

For the verbal scales the word considered to denote "high" annoyance for each survey is provided. Words such as "very", "very much", and "considerably" are chosen. The other words in a list influenced the choice of a particular word in at least one instance. The word "considerably" is assumed to identify "high" annoyance in two surveys (CAN-168, CAN-121), but to identify "low" annoyance in another (AUS-210) which also includes the word "highly " as a response alternative. The number of words from which the respondent chooses do not affect the division point. The "% of scale" column shows that from 20% (1 of 5) to 40% (2 of 5) of the offered words may be counted as "high" annoyance.

For the numeric scales and the multi-item indices the division between high and other annoyance is chosen so that top 27% to 30% of the scale points are counted as "high" annoyance. This means that the verbal labels for end points must be considered to be functionally equivalent. The phrase "extremely" in English (USA-203) is thus equivalent to the German word "unerträgliche Störung" (unbearably disturbing) (SWI-053).

2.4. Data for estimating DNL

The synthesis uses Day-Night Average Sound Level (DNL) as the index of noise exposure. The first of the three columns under "Noise data reported in original survey" in Table 1 lists the noise metric that was evaluated to provide the value of DNL. The value of DNL for 12 of the 29 data sets was directly provided by the original survey or was presented separately for each of the DNL noise periods so that DNL could be directly calculated. For two other data sets (CAN-121, BEL-122) the only difference from a direct estimate of DNL was the use of an 8-hour rather than 9-hour period for the nighttime weighting. While some other metrics are based on measures of equivalent-energy indices, the metrics also include NNI, L_{50}, Average peak Perceived Noise Level and the number of aircraft over 80 L_{PN} dB.

The two additional columns under the "Noise data" heading indicate the length of the daily time period for which noise estimates were available and whether separate estimates were provided for daytime and nighttime noise environments. In one data set (BEL-137) the noise

data are available for the daytime and separately for either the night or evening but not both. In one other (SWE-035) the estimate of DNL is based on NEF and CNR, metrics that incorporate a nighttime weighting.

For the remaining 13 data sets, the authors estimated DNL with only a single indicator for the noise environment and thus without any indicator of the relative importance of daytime and nighttime noise. In seven of these data sets the single number represented the noise during the entire 24-hour period (e.g. 24-hour L_{eq}, 24-hour L_{50}), while in the 6 remaining data sets the noise data referred to only the daytime. In all 13 data sets the relative importance of daytime and nighttime noise environments had to be assumed to be the same at all study sites and had to be assumed to be similar to that experienced in another survey (sometimes in another country) and, in some cases, from another type of noise source. The amount of nighttime tramway noise in Gothenburg (SWE-165) was predicted from the amount of daytime tramway noise using information about the relative amount of daytime and nighttime road traffic noise in several cities in Sweden and the United States.

Analyses would be needed before determining whether the absence of information about nighttime noise could affect estimates of DNL. The most extreme range of assumptions might range from assuming that there is no nighttime noise to assuming that the daytime and nighttime hourly equivalent continuous noise levels (Leq) are equal. This extreme assumption would generate a 6.6 dB difference in DNL for two sites with equal daytime noise levels. The additional 10-decibel nighttime weighting in DNL thus accentuates the effect of variations in estimated nighttime noise levels. DNL is thus relatively sensitive to errors in estimating nighttime noise environments. For 24-hour Leq, on the other hand, the difference between an equal and no nighttime noise environment would be only two decibels.

2.5. Errors and misunderstanding in synthesis data

This review was not conducted with the objective of identifying errors in the data. This review does not suggest a revised dose/response relationship. In the course of the review, however, a few major errors and a larger number of minor inaccuracies have been identified. These are noted in this report for the purpose of alerting future researchers to these details. Most of these errors are either explicitly noted or corrected by the information in Appendices A and B or in Table 1 or Table 2. The primary data from 3 data sets (CAN-121, UKD-071, USA-338) could not be checked because the data came from unpublished communications to the synthesis authors. This section describes and categorizes some errors. The significance of the errors for the estimate of the overall dose/response relationship is discussed later in this review.

Errors in recording annoyance data: The most important recording error is that the annoyance percentages for the 12 data points for the Swiss Aircraft Survey (SWI-053) are totals that have been cumulated over the lower noise levels, not the percentages that are annoyed at a

particular noise level. Thus the 1991 synthesis indicates that at the highest noise level (65 L_{dn}) 100% of the respondents were highly annoyed, when in fact 63.6% were highly annoyed.

The data points recorded for one survey (BEL-122) do not match the data presented in the referenced source. No explanation for this discrepancy is apparent.

Other annoyance transcription errors are smaller. Four of the 5 points in the French aircraft survey (FRA-016) are off by 2 to 5%, one of the data points in the Danish Railroad survey (DEN-200) had roughly 25%, not 13.5% annoyed, and one of the Canadian Aircraft (CAN-168) points is off by 0.5%. Occasionally a data point in the original publication does not appear in the synthesis (BEL-137, GER-034, SWI-053, UKD-008, CAN-168-road traffic).

Errors in recording noise data: The synthesis authors mistook the total noise levels for the road traffic noise levels in one mixed source survey (SWE-165) and, as a result, over-estimated the road traffic noise levels by 1 to 10 Db for five of the six data points for the data set. Two other surveys have one point each with noise levels that differ by 1 to 5 dB from the published values (USA-082, DEN-200).

Incorrect descriptions of annoyance scales: Several of the scales are considerably different than represented in the synthesis. The Munich aircraft survey (GER-034) is based on a multi-item index (11 items based on 37 questions) rather than on a five-point verbal scale. This is of some importance since this survey provided many of the observations at the highest noise levels.

A number of the descriptions of the annoyance scales do not agree in details with the original publications. Three of the surveys ask about neighborhood or community noise generally rather than about street traffic noise (CAN-121, DEN-075, USA-102). The two Swiss data sets from one survey (SWI-053) use an 11-point numeric scale with the quite strong label of "unbearably disturbing" ("unerträgliche Störung", "gêne insupportable") where as the synthesis represents the top three points with a quoted label of "very annoyed" ("Stark Störung"). The scoring of two of the five items in the multi-point Heathrow annoyance scales are not correctly described (UKD-008, UKD-024).

Incorrect descriptions of study design: The departures from accurate descriptions of study designs are not of great importance for the synthesis article conclusions. Seven of the counts of total numbers of respondents have been corrected as indicated by a "C" in the fourth column of Table 1. The corrections are usually minor and sometimes reflect the difference between the total number of responses in a survey and the number that were included in the plotted data points. The synthesis article describes all data points as being study areas, but as is indicated in Table 1, at least 13 represent intervals on the noise metric and include respondents from more than one area.

3. THE SYNTHESIS METHODOLOGY

The implemented methodology defines the synthesis and determines what conclusions can be drawn from the synthesis. This section examines the overall purposes of the synthesis, reviews some of the requirements for an objective comparative analysis, lists some conditions that are often controlled in noise/annoyance studies, compares the planned with the implemented methodology, identifies other eligible studies, and discusses the use of a "high" annoyance indicator. The statistical analysis techniques are evaluated later in Section 4: Weaknesses of the analysis methods.

3.1. Primary objective of the synthesis

The first synthesis paper presented the results of the synthesis as "... the best currently available relationship for predicting community annoyance due to transportation noise of all kinds." (Schultz, 1978: 377) The 1978 synthesis was planned to meet at least two goals: (1) to provide policy makers with a tool that could be widely accepted as a basis for predicting responses to noise and for planning, and (2) to determine "... whether or not a single relationship between noise exposure and annoyance can be found that is valid for all types of noise." (Schultz, 1978: 379)

Since 1978 the synthesis results have been widely accepted for planning purposes and the accompanying noise/annoyance curve has become a standard of comparison for new surveys' results. The wide acceptance shows that the synthesis has been found to be useful.

The fact that the curve has been useful and widely accepted does not, of course, prove that it is necessarily accurate or that it was derived using rigorous, objective scientific standards. At least two authors' publications have suggested that there were important weaknesses in the methodology (Kryter, 1982a; Kryter, 1982b; Kryter, 1983; Griffiths, 1983). The author of the 1978 synthesis defended many aspects (Schultz, 1979; Schultz, 1982).

3.2. Standard requirements for a comparative study

Dissatisfaction with traditional literature reviews has lead to the development of quantitative methods for evaluating many studies on the same topic. The term "meta-analysis" has come to be used to describe a collection of broad methodological concerns and a range of statistical techniques for combining evidence across empirical studies (Rosenthal, 1984; Wolf, 1986; Wachter and Straf, 1990). Some of the general methodological guidelines that have been developed for meta-analytic studies are applicable to the development of this dose/response relationship.

The noise-annoyance synthesis goals differ from most meta-analysis goals in at least one important respect. Most meta-analyses have either attempted to determine whether there is any relationship between two variables, or attempted to measure any such relationship using indicators that are not dependent upon the scales that are used to measure the variables. The synthesis, however, attempts to present the results in the units of the scales in which the variables are measured (i.e. DNL dB(A) for noise and percentage expressing "high annoyance" for annoyance). This imposes some additional requirements that have been described by Fields and Walker (1980; 1982a).

Some of the most important requirements for creating an objective, methodologically satisfactory summary of the results from noise studies are the following:

1. The population of eligible studies is clearly defined. This implies that:
   a. The criteria for excluding studies must be clearly defined and uniformly applied.
   b. The common features of the selected studies must be accurately stated.
2. All eligible studies or a probability sample of eligible studies is included.
3. The two variables being analyzed (noise and annoyance) are uniformly defined.
4. A common "synthesis" estimate is developed from the studies by weighting the individual studies on the basis of their accuracy or other indicator of importance.
5. Factors that could introduce differences between studies are examined by either:
   a. Setting a population definition that is uniform
   b. Standardizing all studies to the same nominal conditions
   c. Systematically reporting the values of variables that might create differences between studies
   d. Discussing the probable impact of the remaining unexamined variables
6. Decisions about all aspects of the study methodology are independent of the study results.

3.3. Factors to consider in comparing survey results

The methods followed for any particular noise/annoyance survey are generally rigidly standardized to ensure that differences in methodology do not create artificial differences between respondents. The same attention to a standardized approach should reduce unwanted errors that could inflate the differences between different studies. Table 3 lists 29 factors of varying degrees of importance that are often considered in designing surveys and comparing survey results.

Two relations between these factors and the synthesis are identified in the second and third columns of Table 3. The second column gives examples of the range of conditions that are present in the synthesized studies. The third column identifies four factors that have been systematically normalized in the synthesis and one factor that was addressed in some, but not all, studies.

Table 3 thus indicates that attempts were made to control for 5 of the 29 listed factors. The noise source and language are systematically documented for most studies. There is no evidence that a systematic assessment was made to determine whether the remaining 22 factors are constant. Even though a 5-dB correction was made for a measurement at the facade of the dwellings in one study (DEN-075), no attempt was made to apply a correction for reflective surfaces in other studies. For example, the English railway and road traffic measurements that were made at 1 meter from the facade were not corrected for comparisons to aircraft noise. Approximately a 2.5 dB correction would have been appropriate.

Although Table 3 contains a long list of factors and additional factors could be added, this does not imply that comparisons must be restricted to some small number of surveys that have followed identical methodologies. Table 3 does not take the additional step of evaluating the practical importance of ignoring these factors. Some may have little or no impact. Others, such as microphone placement for surface transportation, may be of considerable importance. Other analyses would be needed to determine their probable importance. One possible practical approach to the problem would be to carefully asses major factors and ignore other factors that could not be expected to create systematic differences between surveys of more than a specified number of decibels.

3.4. An evaluation of the 1978 definition of the synthesis design

An important aspect of all meta-analytic study designs is the method that is used for including or excluding studies from a synthesis. For noise/annoyance surveys there are many more studies (over 200 by 1991) than are included in the synthesis. In the 1991 synthesis, the criteria for including studies are explicitly listed. In the 1978 synthesis, the overall criteria were not addressed, but the issue of selection is present in the analyses that included 11 data sets while excluding 9 other "non-clustering" data sets from the final synthesis.

The 1978 article appears to indicate that a three-step procedure was followed: (1) the noise/annoyance relationship was plotted for all data sets; (2) on the basis of the noise/annoyance relationships, the surveys were divided into those that agreed and those that disagreed; (3) it was reported that those that agreed "... presented the subjective response data in such a way that a consistent choice could be made of who were 'highly annoyed..."'(Schultz, 1978: 379). This is an important finding if all the data sets that disagreed had annoyance scales that were not comparable.

The 1978 synthesis concluded: "The degree to which these curves agree with one another was surprising and impressive, particularly since the noise ratings and interview methods were, in some cases, quite different". (Schultz, 1978: 379) Whether or not this agreement is important depends upon the synthesis study methods. If these surveys clearly differ from the non-clustering surveys in their study methodology, then the agreement is important. If on the other hand, this is only a loose *ex-post-facto* explanation of a possible difference then, the agreement is less important and may simply be the result of rationalizing the differences between the two groups.

While the 1978 synthesis represents a great deal of diligent, carefully documented work by the author, it appears that on a critical aspect of the methodology there is a substantial weakness and lack of sensitivity to the possibility that the outcome of the study may have biased the interpretation of the methodology. One indication of this weakness is the fact that the First Heathrow survey (UKD-008) is represented by two data sets in the 1978 analysis that differ only in the cutting point that is used for the same annoyance scale. One data set is counted as providing support for the consistency in annoyance responses when it is tabulated as a "clustering survey" while the other division of the annoyance scale is considered to be a "non-clustering" survey. A rigorous approach would give each survey only one chance of being included.

It seems likely that after dividing the surveys into two groups on the basis of their results, the author diligently and successfully searched for weaknesses in the non-clustering surveys, but did not equally diligently search for similar weaknesses in the clustering surveys. It seems likely that a sufficiently diligent search could reveal some questionable aspect of any survey that could be grounds for its rejection. The following aspects of the 1978 synthesis study suggest that the "clustering surveys" were not critically examined to determine whether they had comparable methodologies:

1. The Paris street survey (FRA-041) is retained as a clustering survey even though the author states that this is an "...entirely different kind of scale, not an annoyance scale at all." (Schultz, 1978: 381)

2. All three of the Heathrow data sets (one is not a clustering data set) are based on multi-item questionnaire indices with numbers that can not be directly interpreted. It seems quite likely that if the results had not happened to cluster with the other surveys that they could have been easily dismissed as too methodologically dissimilar in the same way that the two TRACOR data sets (surveys USA-022, USA-032, USA-044) were discarded.

3. If the Munich survey (GER-034) had not clustered, the authors might have more carefully read the Munich report and discovered that it also had a multi-item scale that was too dissimilar to expect it to cluster.

The discussion of the evidence against two Swedish road traffic surveys (SWE-021, SWE-142) suggests that they might have been included in the synthesis if their results had clustered. The article appears to accept the Swedish surveys' annoyance questions. The surveys are finally dismissed because there is "...sufficient question about the determination of the noise exposure...". However, the noise exposure estimation techniques used in the other surveys are not examined. The validity of the surveys is also questioned because of the comments of an anonymous reviewer. Reviewers' comments are not cited for any other surveys.

3.5. An evaluation of the 1991 definition of the synthesis design

The 1991 synthesis article offers five explicitly numbered criteria and two additional criteria for selecting and thus characterizing studies. These seven criteria are listed in the first column of Table 4. These criteria have been examined with some care. To determine whether the criteria have been applied uniformly, the wording of the criteria have been examined in three ways: (1) by determining whether the criteria are precisely worded so that an objective determination could be made of which surveys meet the criteria, (2) by determining whether all of the data sets included in the synthesis met the criteria, and (3) by determining whether there were other surveys that met the criteria. The latter topic is addressed in the next subsection.

The results of the examination of the wording are reported in the second column of Table 4. It is clear that each of the first five criteria contain subjective elements that do not provide an objective basis for describing the studies.

The data sets in the synthesis are consistent with the last two of the seven criteria in Table 4. No laboratory studies are included (Criterion #6). No impulsive noise studies are included (Criterion #7).

The characteristics of the studies do not appear to be consistent with the first five of the seven criteria. The third column of Table 4, identifies the data sets that contradict each of the stated criteria. For criterion #5 (the definition of high annoyance), the third column also highlights some aspects of the criterion that do not contradict the criterion but do amplify the complex definition provided in the synthesis article.

Although the inconsistencies between the criteria and the study characteristics are self-evident for most criteria, a short explanation is useful for others.

The first criterion requires that one questionnaire item inquire about long-term annoyance. For three studies, however, this of little significance. The third column of Table 4 identifies three studies in which that single item about long-term annoyance is a part of indices consisting of from 6 to 37 items. For the second criterion, it is noted that a "preference" for noise measurements did not eliminate one survey with no noise measurements or at least 6 other surveys that had no noise measurement data for nighttime exposure.

For the third criterion, it is noted that even though nighttime noise must be weighted with a 10-decibel penalty in DNL, 13 of the data sets could not isolate the nighttime data to give it a separate weighting. For these 13 studies the synthesis was also unable to rely on any other type of local data (i.e. data about number of events) that might have been used to estimate the nighttime noise environment. Nighttime estimates were instead made by estimating that the nighttime distribution would be about the same as that found in a survey from another setting.

The fourth criterion, sample size, seems to not have been implemented. Since all of the synthesis analyses treat each data point equally, the only relevant sample size is the number of interviews supporting each data point. As is noted in the third column of Table 3, one data set includes data points with only a single interview and seven data sets include data points with fewer than 20 interviews.

The fifth criterion, definition of high annoyance, is much more complex and requires more subjective judgements by the synthesis authors than is suggested by the short statement of the criterion that is printed in the 1991 synthesis definition. The 1991 synthesis definition refers the reader to the 1978 synthesis article to find the actual definition. The 1991 article itself contains a statement only about the "identification of numbers of respondents describing themselves as 'highly annoyed'..."

In fact only one of the 29 data sets can be directly classified using the shortened 1991 definition because only one of the 29 data sets actually presents the word "highly" for the respondents to choose. The classification of the remaining 28 data sets involves a more extended definition with additional subjective judgements. Creating the extended definitions involved the subjective judgements that people use the description "highly annoyed" in the same way they would use the following: the word "very", a score of 7 on a 11-point numeric scale, the answers to 6 questions in a multi-question index, the placement of a tick mark at 109 millimeters on an unmarked 155 millimeter line, and the ranking of traffic noise as the worst aspect of a neighborhood. For some data sets a further subjective judgement concerns whether such words as "very", "considerably", or "mycket" [SWEDISH] are equivalent to "highly".

When the data are examined carefully, it is seen that applying subjective definitions raised additional problems. The word "considerably" is assumed to be equivalent to high annoyance in two data sets (both CAN-168 data sets), but is not assumed to be equivalent to high annoyance in a data set in which the word "highly" also appears in the scale (AUL-210). Sixty percent of numerical scale is considered to be equivalent to high annoyance, but 60 percent of a verbal scale is not equivalent to high annoyance. For the Danish railroad study (DEN-200) the authors' did not realize that a high annoyance rating was given to any respondent who volunteered railway noise was a problem, regardless of the annoyance score on a direct question.

The information summarized in Table 4 indicates that five of the seven study definition criteria used in the 1991 article are either not consistently applied or do not have a clear, objective meaning. The synthesis is not based on an objective definition of the study population.

3.6. Identification of other surveys that would appear to be eligible

The up-dated synthesis was published in 1991. Over 250 surveys' results had been published by 1985. Many of these surveys included measurements of annoyance and noise exposure

that would have been comparable to those included in the synthesis. In some cases the information is included in available publications. In most instances the data could only be obtained, as it was for at least six of the synthesis data sets, through communication with the original researchers.

A cursory examination of some of those 250 surveys identified 20 data sets that are publicly available. Of these there are 14 data sets for which the results had been published by 1986 in a form that would have permitted comparison with the 26 surveys that were included in the synthesis. The survey identification numbers for these 14 data sets are the following: AUL-227, AUL-264, CAN-120, FRA-092, FRA-197, NET-153, SWI-133, SWI-158, SWI-173, SWI-180, UKD-052, UKD-242, UKD-243, UKD-309 (Survey identification numbers come from Fields' (1991) catalog of surveys). At least six additional data sets would have involved considerably greater effort to analyze but are available through the ESRC data archive at the University of Essex, England (USA-022, USA-032, USA-054, UKD-130, UKD-148, UKD-157).

The existence of these 20 additional, accessible surveys makes it clear that the synthesis does not include all of the available data. This in turn points to the fact that clear criteria were not enunciated in the synthesis methodology for determining which studies should be included or excluded.

3.7. Choice of "highly annoyed" as an impact descriptor

To measure the impact of noise in the synthesis article, each individual is classified as "highly annoyed" or "not highly annoyed". This particular classification has been popular for policy purposes but has not been shown to be superior to other descriptors on objective, scientific grounds.

There are at least three non-scientific, policy-related reasons for the popularity of this descriptor. First, the term "highly annoyed" seems to have a clear, common-sense meaning. Secondly, the "high" annoyance descriptor seems to directly refute the suspicion that noise annoyance is frivolous and unimportant. Third, the descriptor directs attention to only the higher noise levels and thus defines the problem as a relatively small, manageable public policy problem. With a less severe descriptor the noise problems would appear to be more widespread.

The 1978 synthesis article offers conjectures about several scientific grounds for selecting a high annoyance measure. However, no objective evidence is offered and measurement theory suggests that most of the conjectures would not be upheld if they were tested.

People's annoyance responses can be thought of as being arrayed along a continuum from no annoyance to the very greatest degree of annoyance. The most fundamental, widely accepted theories of attitude measurement agree that the greatest precision in locating an individual along such a continuum can be achieved by using multi-point scales (preferably based on

answers to several questions) to place individuals along this continuum. The dichotomous scale, such as the "high annoyance" measure simply uses a two-category scale. This is somewhat analogous to throwing away a measuring tape for measuring people's heights and simply dichotomously classifying every individual as "high" or "not high". For the purely scientific purpose of reliably and accurately measuring annoyance, a multi-point scale will be more accurate than a two-point scale and a multi-item index will be expected to be superior to a single-item index.

Empirical evidence for the superiority of a multi-point scale for annoyance measurement comes from an article in which the noise/annoyance correlation was compared for annoyance questions that were either dichotomized at the "high" annoyance point or were represented by scores that represented the full range of the questions (from 4 to 9 points). In 12 of the 13 comparisons the multi-point scale was more closely correlated with noise level (Fields, 1980).

The 1978 synthesis article conjectures that the high annoyance dichotomization would be less affected by non-acoustical variables and more closely related to noise level than other annoyance measures. No empirical evidence is offered for this position. The previous paragraph showed that a multi-point scale has a higher relationship with noise level than a dichotomous scale. An examination of a graph of the relationship between noise level and different dichotomizations of annoyance shows that, in at least one survey, a less severe dichotomization is more sensitive to changes in noise levels, especially low noise levels, than is a more severe dichotomization (Fields and Hall, 1987: Figure 3.6).

Other arguments for the use of the "highly annoyed" dichotomization are influenced by the fact that "high annoyance" is confined to high noise levels. Such arguments thus prejudge the issue of whether annoyance exists at lower noise levels. It is argued that at high transportation noise levels residents may hear all of the transportation noise events inside their homes where as at low noise levels residents may only hear some of the transportation noise events, those that occur during periods when interior noise levels are reduced. The relevance of this for the choice of an annoyance measure is unclear. Long-term residents realize that an unpleasant outside noise environment exists even when it is masked by interior noise events. In fact residents may be annoyed precisely because they can hear the exterior noise at times when the otherwise distracting interior noise levels are low.

If it were accepted that some type of dichotomous measure is needed, there may be small technical differences between more or less severe annoyance dichotomizations. The high annoyance dichotomization, for example, might be somewhat less sensitive to the use of the annoyance screening questions that are present in some, but not all, social surveys. However, any slight differences between the scales in technical characteristics are not important enough to outweigh the policy implications of choosing a scale that rates a small, rather than a large, number of people as being impacted by noise.

3.8. Illustrating the extent of weaknesses in data sets

This section has identified a number of weaknesses in the data. Evaluating the impact of these weaknesses on the estimate of a dose/response relationship would involve data gathering that has not been performed for this report. Some insight into the extent of these weaknesses can be gained without additional data.

After examining the comments in the last column of Table 1, eleven of the 29 data sets have been identified that have especially severe weaknesses. The eleven data sets have been marked with a "(P)" in the "Comments" column of Table 1. In all cases the examination of the data indicate that the data do not represent the relationship between a measured noise level and a direct rating of high annoyance with the measured noise. Three data sets' annoyance questions did not specify a noise source (DEN-075, CAN-121, USA-102). Two data sets reported the highest ranked disturbance in neighborhoods without indicating the degree of annoyance. Two data sets related the annoyance with a specified target noise to the total noise exposure rather than to the noise exposure from the target noise (SWE-165, USA-301). One data set related annoyance "since this time last year" to the noise level measured in the past week even though the noise environment had changed at some time between the past week to three months (USA-203). One data set consists of the cumulative percentages of highly annoyed respondents at each noise level (SWI-053) with the result that the reported value of 100 percent for the highest noise level simply expresses the truism that all of the highly annoyed respondents live at or below the highest noise level. One data set simply classifies the top half of the respondents as "highly annoyed" without regard to the meaning of the annoyance questions (GER-034). One traffic noise data set does not rate the total traffic noise, but instead gives respondents three chances to give a highly annoyed response by rating three separate components of traffic noise (CAN-168).

These eleven data sets have been analyzed separately. These data sets account for 38 percent of the data sets and 41 percent (185) of the data points. The 185 points are identified with open circles in Figure 3. The points are not randomly distributed. They account for all of the points with the 11 highest annoyance scores (all above approximately 73 dB), for about three-quarters of the data points above 80 dB and for about 80 percent of the data points at which more than 60 percent are highly annoyed.

The data and notes in Table 1 indicate that the data problems are not confined to the eleven data sets identified in Table 1 and Figure 3. A strong case could be made for excluding additional points and data sets. The values for the data points in one of the remaining data sets does not match the published data source (DEN-122). Eighty percent of the annoyance measures in two of the remaining data sets consist of activity disturbance items (UKD-008, UKD-024). Measurement locations for surface transportation sources are generally unknown. There are at least 20 small data transfer errors. Two additional data sets could not be examined to determine whether they contained errors (UKD-071, USA-338).

4. WEAKNESS OF THE ANALYSIS METHODS

The fundamental strength and appeal of the synthesis is that the analysis provides a simple, apparently meaningful, measure of impact at each noise level. Policy makers have found this impact measure to be more useful than many of the measures that had previously been provided in many of the frequently used, more sophisticated analysis techniques such as path analysis, multiple regression or analyses of variance. They have also found a single-item "percentage-impacted" measure to be a more useful measure of impact than that provided by more sophisticated, multi-item annoyance scales. Although the synthesis is built on this fundamental, useful analysis approach, some serious weaknesses are evident in the details of the analysis techniques that were used.

4.1. Comparability of "high" annoyance indicators

A key element of the analysis methodology is the judgement that the definitions of high annoyance in the different data sets are equivalent. The 1978 synthesis article contains the following expression of confidence that the judgement can be made: "... a sensible person ought to be able to locate with useful accuracy the points on *all* the scales corresponding to the same degree of annoyance." (Schultz, 1978: 377) No empirical evidence is offered to support this judgement. Four considerations suggest that comparable definitions of high annoyance have not been established in the synthesis article.

An examination of the difficulty of the task: A review of the annoyance scales presented in Table 2 indicates that the task of judging equivalent annoyance cutting points is too complex and uncertain to be based on simple common sense. There are fundamental differences in the types of scales; verbal, numeric, multi-item indices, or ranking. There are differences in the location that is mentioned in the question; "at home" (implying inside the house), or "in the neighborhood" (seeming to include positions that are nearer to surface transportation noise sources). There are differences in the words that are used to describe high annoyance: "very", "very much", "considerably" [English], "beaucoup" [French], or "bastante" [Spanish].

The synthesis authors' own presentation reveals their own ambivalence about their ability to make these judgements. In the first synthesis article the scales are described by stating that "....the noise ratings and interview methods were, in some cases, quite different" (Schultz, 1978: 379). In the second synthesis article the authors abandon their previous confidence that only the meaning of the verbal scale labels is important. In the section on "Bias errors in definitions of high annoyance" the authors use the number of verbal scale points to assess the similarity of the annoyance scale definitions. In Table 4 it was noted that the word "considerably" is judged to represent a different degree of annoyance in different surveys.

The experience of social scientists with opinion question wording suggests that the task of choosing equivalent questions is too difficult to be accomplished on purely *a priori* grounds. Social science journals and books present frequent examples where small, seemingly innocuous differences in question wording are associated with substantial differences in responses (Schuman and Presser, 1981).

Empirical evidence that tests some judgements: In the first synthesis article it is stated that high annoyance is equivalent to "very annoyed" and that, in the First Heathrow survey, the top 2 points on a 7-point multi-item index were also equivalent to high annoyance. The synthesis article argues that the original authors of the survey were clearly wrong in choosing the top 3.5 points to represent high annoyance (Schultz, 1978: 391). The synthesis article expression of confidence in this judgement as a "sensible person" could in this case be tested because another publication for the same survey, that was apparently not consulted, presented the results for both types of scales. These are reproduced in Table 5 and show that the "very much" annoyed alternative (19 percent annoyed) more closely matched the original author's decision to count the top three points (21 percent annoyed), than it does the synthesis author's decision to count the top 2 points (11 percent annoyed).

In the first synthesis article the seemingly reasonable assertion is made that inserting the word "ever" into an activity disturbance question would obviously increase the reports of the activity disturbance (Schultz, 1978: 391). This proposition was tested in an experiment in the British Railway Survey (UKD-116) in which carefully balanced halves of the sample were either asked "...does the noise from the trains make it hard to hear what is being said on T.V.?" or asked "...has the noise of the trains ever made it hard to hear something that was said on T.V.?" [The underlining was deliberately included in the original, text for emphasis.] This seemingly important wording difference did not generate different answers from respondents (Garnsworthy, 1977: 51). Although, as noted earlier, opinion research literature has emphasized the instances in which *a priori* judgments underestimate the effects of question wording; this example shows that *a priori* judgments can sometimes overestimate the impact of question wording.

These two examples show that "common sense" is not sufficient to determine the relationship between alternative annoyance measures. Empirical evidence for transfer functions between different annoyance questions is needed.

A lack of agreement with the subjective judgement of other noise experts: Most survey reports have been written without their authors committing themselves to any particular definition of an important degree of annoyance. Several of the authors of the original reports that appeared in the synthesis articles selected less severe cutting points than were used for the synthesis. McKennell conducted a rather elaborate analysis of the first Heathrow survey before concluding that the top 3.5 points "... appears to be a reasonable interpretation of serious annoyance." (McKennell, 1963: Summary, p.2) The synthesis counted only the top 2 points. Bullen and Hede considered the problem carefully and concluded that their measure of "seriously affected" from a 23-item index is "...a more appropriate measure, both for

descriptive and for legislative purposes.... The 'highly annoyed' measure appears to exclude some people whose degree of reaction would normally be termed 'high'"(Hede and Bullen, 1982: 94). Much of the public debate between Schultz and Kryter (Kryter, 1982a; Kryter, 1982b; Kryter, 1983) concerned the correct definition of high annoyance.

Weaknesses and errors in applying the synthesis definitions: The evidence for the fifth synthesis study criteria in Table 4, showed that the synthesis definitions of high annoyance contain important subjective elements and that in applying these definitions some inconsistencies and errors have occurred. Table 1 listed other errors in interpreting the annoyance scales that would affect the calculation of the percentage that are highly annoyed.

Summary: For professional opinion researchers this entire debate about the "sensible person's" judgement is completely miscast. In fact, there is no true objective definition of "high annoyance". All that can be scientifically determined is whether scales measure the same degree of annoyance. That statement can only be determined on an empirical basis by comparing the answers of the same population to alternative annoyance measures.

4.2. Method of weighting data points

The synthesis papers give each data point the same weight in the analysis despite the fact that the data points differ greatly in the numbers of interviews they represent (from 1 to at least 1,186 in Table 1) and the fact that the studies differ considerably in the number of data points that they contribute to the analysis (from 2 to 42). No justification is offered for this approach in the synthesis papers. The relative size of the points is approximately indicated in Figure 2 where the area of each data point is approximately proportional to the numbers of interviews. The figure shows that the points with the highest noise levels and the most extreme annoyance responses tend to be some of the smallest data points.

A more appropriate conventional approach would be to give equal weight to every interview and to analyze a data set consisting of as many data points as there were individual interviews. This approach would therefore weight each synthesis "data point" by one estimate of its accuracy, its number of interviews. This approach has the secondary advantage of being suitable for evaluating the effects of individuals' characteristics on their annoyance responses. The approach also has the advantage for logistic regression of not requiring the introduction of approximations to represent the annoyance score when zero percent or 100 percent of the interviews at the synthesis data point are highly annoyed. The analysis in Appendix C shows that this approximation can affect estimates of logistic regression coefficients.

A number of other possible alternatives are preferable to the approach taken in the synthesis article. The interviews within each study could be weighted equally, but the synthesis could be formed by giving equal weight to each study. A more complex, acceptable procedure would be to weight each study by the accuracy of its estimate of the dose/response relationship.

4.3. An important technical flaw in the calculation of sampling errors and inductive statistics

The 1991 synthesis article contains estimates of sampling errors. The most prominent estimates are the 95% confidence intervals in the graphs. All of these estimates are incorrect because they are based on simple random sampling assumptions. Simple random sampling assumptions are clearly inappropriate for noise surveys because interviews are clustered into neighborhoods. An analysis of seven surveys, including four of the synthesis data sets, shows a consistent tendency, even after controlling for noise level, for more agreement within neighborhoods than would be expected on a chance basis (Fields, 1983). A National Academy of Sciences panel supported this position when criticizing one community noise survey by stating that "..clustering increases the standard error of estimate.." (Committee..., 1976). Thus the survey samples should not be considered to be simple random samples of the population and the simple random sampling formula used in this article are not appropriate.

The incorrectly calculated sampling errors in the synthesis article underestimate the actual sampling errors. For the "category" type data points, this is because the calculations ignore the between-site differences in average annoyance scores. For the "area" type data points, the sampling errors in the article are for the error in determining the percentage annoyed in a particular neighborhood and not at the noise level generally.

Even intermediate statistical texts now offer a variety of techniques for more accurately estimating sampling errors. Taylor-series expansion methods could be used for the simple statistics calculated here (Kalton, 1983). For the more complex statistics often used in noise surveys, an appropriate choice would be one of the pseudo-replication techniques (balanced repeated replication, the jackknife, or the bootstrap) (Lee, et al., 1989).

4.4. Measuring uncertainty in the dose/response relationship

In the section labeled "Uncertainty of measurements of percentages of respondents highly annoyed", the 1991 synthesis article uses an inappropriate method for measuring the accuracy of the estimate of the dose/response relationship. Table III of the synthesis article measures the accuracy of each survey by the average of the estimated confidence intervals for each data point. The article then concludes that "since this represents the fundamental level of precision of measurement on the ordinate of the dosage-response relationship, it is unproductive to seek explanations for smaller differences among potential fitting functions for these data."(Fidell, Barber and Schultz, 1991: 232) This measure is obviously inappropriate because, among other things, it does not take into account the number of data points or total number of interviews in a study. Using this criteria a survey with 100 interviews concentrated at a single site would be more accurate than a survey with 10,000 interviews spread over 101 sites.

The actual uncertainty of the estimates should be measured by confidence intervals for what ever combined statistic is of interest. The precision of the estimate of the overall dose/response relationship might be represented by the 95% confidence interval of the percentage who are estimated from a regression analysis to be annoyed at each noise level. When calculated correctly this would account for the number of interviews, number of sites and the noise level at which the prediction was made.

The fundamental question that the synthesis article is attempting to address in this section is whether the differences between alternative curves (for example between a cubic and quadratic form) are statistically significant. This might be tested in a number of different ways. One appropriate test would be to test for the difference between the multiple correlation coefficients for the two curves. Such a test would account for the number of interviews, number of sites, distribution of interviews over different noise levels and number of surveys. Other steps would also be needed. For example, the pattern of the residuals from the regressions should be examined. Whether or not the surveys all come from the same population should be tested. Of course, as was noted above, the test would need to use techniques that take into account the clustered sample design.

4.5. Assessment of annoyance definition differences between 1978 and 1991

In the section "Bias errors in definitions of high annoyance," the 1991 synthesis addresses the possibility that the average division point for high annoyance may have changed from 1978 to 1991. It is stated that the average division point for the 12 clustering data sets in 1978 was about 27 to 29% and the average division point for the 1991 surveys was 31.4%. The analysis is weakened by the method of weighting the data points, introduces a largely irrelevant example and then implicitly accept assumptions that undermine the logic of the synthesis.

The analysis computes an average that weights each survey (not each data point) equally. Since the synthesis curve weights each data point equally, the analysis of average differences should also assign weights based on data points. The synthesis publication appears to dismiss the possible impact of any differences in definitions of annoyance by stating that a change in the definition for only one survey (only 20 data points) does not the alter the average for the total analysis (453 data points). This is largely irrelevant since the issue concerns the differences between the large number of data points in the 1978 data (161 data points) and the 1991 data (292 data points).

The logic of the synthesis authors' analysis in this section, in effect, confirms the near impossibility of comparing such divergent scales on purely *a priori* grounds. The most fundamental assumption of the synthesis is that as "sensible" people the authors have been "able to locate with useful accuracy the points on *all* the scales corresponding to the same degree of annoyance" (Schultz, 1978: 377). A part of this leap of faith was made by assuming that only the words, not the number of scale points, is important for verbal scales. In the "Bias errors..." section, however, this article of faith is dropped and the proportion of

verbal scale points is considered. Once the number of verbal scale points is admitted as a criterion for defining annoyance, much of the apparent similarity in the synthesis definitions disappears. The percent of the verbal scale points that are counted as "high annoyance" varies from 20% (DEN-200) to 40% (USA-203), unlike the percentage of the numeric scale points that varies from only 27% to 30%. This in turn highlights the fact that a 40% criterion would also now make eligible two data sets that were excluded in 1978 as "non-clustering" surveys (the two TRACOR data sets that used multi-item indices and had cutting points of 40%).

To seriously consider differences in annoyance definitions between 1978 and 1991, it would be necessary to confront the large amount of diversity in the many facets of the annoyance scales within each of the years. With the existing information about differences between annoyance scales there is no basis for conducting such an analysis or, in effect, for carrying out the synthesis.

4.6. Assessment of errors in estimating noise exposure

The synthesis publication concludes that differences between curves of 3 dB can not be precisely estimated if each study site includes residents who are spread over a $\pm$ 2.5 dB range. Consideration of more general theories of the place of errors in statistics (Johnston, 1972) indicates that while errors in noise measurement can be important, the synthesis article's analysis on this issue is largely irrelevant.

The critical problem caused by errors in estimating noise exposures is in estimating the average exposure of the residents at each site. This is not a function of the spread of the residents' exposure at a site, but rather of the accuracy with which the noise measurement program was able to estimate the exposure at even the very best location in each site. The synthesis article correctly points out that information that would make it possible to evaluate these errors is almost never available from these surveys. The information is not available for the survey that was cited as an example (USA-203). Several of the factors that are relevant to assessing the impact of noise measurement errors in these types of community surveys have been briefly discussed previously (Fields, 1982). The same discussion provides examples under which the effect of such errors could be an underestimation of the slope of dose/response relationships by 10% to 90%. (Fields, 1982: 610).

The synthesis article's discussion of the ranges of noise levels within sites is largely irrelevant. A $\pm$ 2.5 dB range in exposures within sites certainly does not automatically impose a 3 dB limit on the accuracy of the estimate of the dose/response relationship. If the average of the respondents' noise levels within the, for example, $\pm$ 2.5 dB range is accurately estimated then the only effect of the range in exposures is to slightly increase the variance of the annoyance scores. With large enough sample sizes a $\pm$ 2.5 dB range of exposures within sites does not preclude drawing conclusions about 3 dB or less differences between curves.

The synthesis does not systematically address noise estimation issues that could create errors in estimating the average noise exposure at different sites. Microphone positions for surface transportation are not systematically addressed, even though it appears that some noise measurements are at the edge of roadways, some are at a distance of 1 meter from facades, and some others have been corrected or measured under conditions that approximate free-field conditions. Measurements at 1 meter from a facade can be 2.5 dB higher than free-field conditions. The effects of measuring at the road side might be greater. The synthesis also does not address the issues raised by Kryter concerning estimating acoustical energy from peak noise levels (1982a: 1230). The previous discussion by Schultz on these issues stated that there were inconsistencies in Kryter's methods, but did not address the more basic problem of estimating DNL from peak noise levels (Schultz, 1982: 1246)

4.7. Important issues that are not addressed

The 1991 synthesis article does not address several important issues.

The article does not revisit a primary conclusion from the 1978 synthesis; the conclusion that there is quite strong agreement between the different surveys. If this analysis had been repeated in 1991, it seems likely that there would no longer seem to be a high level of agreement. The differences between the surveys are now considerably larger. In fact the Swedish tramway survey results are more extreme than any of even the "non-clustering" results in 1978.

The dose-response relationships in Figure 4, Figure 5 indicate that the surveys provide quite different estimates of the dose/response relationships. The article does not formally test whether there are statistically significant differences between the individual surveys. This would provide the important test of whether the synthesis relationship can be legitimately used to represent the dose/response relationship in all surveys or even subgroups of particular transportation sources.

5. CONCLUSIONS

The 1991 synthesis article provides a readily available dose/response relationship that will probably be used as a baseline for comparisons for administrative and research purposes. After examining the synthesis data and analysis methods six conclusions have been reached:

1 *The best objective description of the synthesis curve is that the synthesis curve represents one research group's judgements about the evidence from 29 data sets (26 surveys) on the relationship between a high degree of annoyance and Day-Night Average Sound Level (DNL).*

Since no other researchers have attempted to combine the data from large numbers of data sets, the synthesis curve is the best available estimate of the relationship between DNL and the proportion of the population classified as having a vaguely defined high degree of annoyance. If it is necessary to estimate the dose/response relationship for a particular annoyance question or survey condition that is present in one of the constituent surveys, a single constituent survey provides a better estimate of a dose/response relationship than does the vaguely defined dose/response relationship defined by the synthesis.

2 *The synthesis curve does not have a clear, definite meaning because it is not based on objective, reproducible study techniques. The curve is NOT a measurement of the relationship between DNL and the percent of the population that would describe itself as "highly annoyed".*

The synthesis articles do not describe a methodology in enough detail for others to reliably repeat the procedures for selecting surveys, defining high annoyance, defining data points, or estimating DNL. The respondents for 28 of the 29 data sets were not asked if they were "highly annoyed".

3 *The inaccuracies and ambiguities in the two synthesis publications are sufficiently important that the synthesis articles should not be relied upon for information about the individual surveys or for a description of the synthesis methodology.*

The examination of the study methodology in Section 3 showed that the published synopsis of the synthesis methodology is not consistent with the methods that were employed. There are a few errors that might have a major effect on the synthesis curve (for example, the misinterpretations of the scores for half of the data points above 80 DNL). There are many smaller errors that could have a lesser impact on the overall dose/response curve, but that might be important for specific purposes. For example, four of the five data points in one survey are off by 2 to 5 percentage points.

This review has not attempted to provide a comprehensive review of each of the constituent studies' methodologies. A systematic review would require additional analyses.

4 *The number and importance of the unexamined differences between the surveys' annoyance scales, noise estimation procedures and study conditions are too great to provide an accurate comparison of the dose/response relationships in different surveys.*

The data can not, for example, support accurate studies of the differences between reactions to different types of noise sources. The synthesis articles describe some, but not all, of the important differences between the data sets. The articles do not systematically normalize the data to standard noise measurement positions for surface transportation noise sources. There are at least four fundamentally different types of annoyance scales. Additional differences between annoyance questions relate to the wording of the questions, the use of screening questions, the number and wording of the annoyance answer alternatives, the language of the survey, and the location (at home or in the neighborhood). At least 13 estimates of DNL were made without information about the relative exposure from nighttime and daytime noise. There is no indication that a systematic assessment was made to determine whether the studies differed on 22 of the major and minor noise measurement, annoyance measurement or methodological differences that are described in Table 3.

5 *Departures from standard methodological principles for selecting study data mean that the results do not provide information about the degree of similarity in dose/response relationships.*

A fundamental methodological requirement is that the selection of studies must be based on objective criteria that are independent of the studies' outcomes. For these syntheses, however, the methods and criteria for choosing the 26 surveys are not described. Standards are proposed for excluding data sets, but the standards are ignored when the results of the dose/response analysis fit the synthesis curve from other surveys. The 1978 synthesis appears to have considered whether or not study results "clustered" before deciding whether or not to include them in the synthesis. There is no clear statement of the criteria that were used to determine why these 26 surveys were included (sometimes only after personally soliciting the required information from authors) from among the more than 200 social surveys that have been conducted around the world. The methodological weaknesses in excluding and including studies in the synthesis thus make it impossible to judge whether the similarity of the results is due to a fundamental similarity in reactions to noise or is due to the selection of surveys with similar dose/response relationships.

6 *The effect of the methodological weaknesses is not known. It is not known whether a more rigorous methodology would give a different estimate of the relationship between noise level and empirically equivalent measures of high annoyance.*

The present report has described weaknesses in the synthesis methodology. The report has provided enough evidence to demonstrate that the weaknesses affect the interpretation of the synthesis findings. The report has not, however, provided enough evidence to demonstrate the extent to which the synthesis dose/response relationship differs from any particular, empirically defined dose/response relationship (e.g. the relationship between estimates of DNL at the noisiest side of a dwelling and the percentage of the population that states that it is "very much" annoyed on a four-point verbal annoyance scale.) The report has not determined whether a more rigorous analysis would provide a substantially different estimate of the dose/response relationship.

7 *The single, most difficult obstacle to an objective synthesis is the absence of transfer functions for linking divergent annoyance scales.*

The impacts of most of the acoustical and situational factors in Table 3 could either be standardized on the basis of more thorough assessments of existing documents or evaluated with existing knowledge. There is no published empirical evidence for relating the divergent annoyance scales. Additional secondary and primary research is needed to obtain this evidence.

6. SUGGESTIONS FOR FUTURE SYNTHESES OF NOISE SURVEY RESULTS

The two synthesis articles reviewed in this report represented major efforts on the part of the authors. The weaknesses in the final synthesis are partly due to the synthesis methodology but more fundamentally to the lack of comparability in the constituent surveys' methodologies and reporting procedures. These types of problems are typical of the problems faced in meta-analyses of many subjects. Many of the solutions are, however, specific to noise surveys. Many of the solutions that relate to analysis methods have been described in more detail in previous papers (Fields and Walker, 1980; Fields and Walker, 1982a). Suggested improvements are grouped under five headings in the remainder of this section.

6.1. Research to provide an empirical basis for relating divergent annoyance questions

The primary weakness in comparing dose/response functions has been the absence of an empirical basis for locating equivalent annoyance questions in different questionnaires. Three types of special research projects could help to solve this problem.

1 Reanalyses of archived community reaction data sets should be analyzed to determine the empirical relationship between some of the existing scales. Several single surveys have included two or more types of annoyance scales in a single survey. Reanalyses of these surveys could provide transfer functions that would in turn provide an empirical basis for normalizing the results from different surveys.

2 New community surveys should be designed to develop transfer functions between pairs of annoyance scales that have been used in previous surveys. At least some of these surveys could be relatively economical if they were based on telephone interviews around airports where long-term average noise exposures were well known.

3 A new research effort should be begun to develop standardized magnitude estimation scales that are relatively transportable across different languages and situations. Such multi-point scales would provide readily interpretable meanings and would be more accurate than simple dichotomous scales.

6.2. Improvements in future standard community surveys

Standard community surveys could easily be modified to provide more comparable data. Changes need to be made in both the survey designs and in the reporting procedures.

1 Community noise surveys should include pairs of questions that have been used in previous surveys. Analyses of these pairs of questions could provide transfer functions

that would in turn provide an empirical basis for linking the results from different surveys.

2 Community noise-annoyance questionnaires should include the annoyance question that has been most widely used in noise surveys as well as any specialized questions that may seem to be more suitable for the specific survey. The most widely used survey question is the following:

> "Does the noise from the ..... bother or annoy you very much, moderately, a little or not at all."

While every question has its weakness and strengths, this one is probably as strong as any other and has the overwhelming virtue of having been administered to over 11,000 people over a time span of more than 25 years.

3 The full annoyance question, together with any screening questions, should be reported in all publications in both the language(s) in which it was administered and in the language of the publication.

4 Detailed survey reports should include a table with the number of respondents at each noise level that give the each answer on the annoyance scales. For small numbers of study locations, the day, evening and nighttime noise levels (L_{eq}) should be reported for each location. For large numbers of locations where the data are collapsed into noise categories, the 24 hour equivalent continuous sound level (L_{eq}) should be reported and, at the very least, a transfer function for estimating DNL and the daytime noise level (L_{eq}) should be provided.

5 Reports should at least discuss the factors that may affect measurements of noise/annoyance relationships. Twenty-nine of these factors were listed in Table 3.

6.3. Improvements in designing a synthesis

A synthesis should be planned to follow the principles that have been come to be accepted as standard for meta-analyses of social science data. The following general standards should be followed:

1 The population of studies and the methods of selecting studies from within that population should be precisely described.

2 The criteria for inclusion of a study should be objectively described.

3 Studies should only be included if an empirical basis is available for normalizing the annoyance measures to a common base.

4 The observer's position for noise environment estimates should be specified.

5 Estimates of noise exposure for unmeasured periods (i.e. night or evening periods for some studies) should be based on study-specific or site-specific data about noise source activity periods during that time

6 The effects of important, uncontrolled variables should be acknowledged in discussions. If possible, their possible impact should be represented by a range of estimates for the dose/response relationships.

6.4. Improvements in documenting synthesis reports

The two synthesis reports were notable for usually including both the original language and English translations of all questionnaire items. They also described the procedures for converting the original study's noise data into DNL. Their usually practice of specifying page numbers for all references is also to be commended. A better documented synthesis would enforce greater discipline on the synthesis process, provide proof of the accuracy of the synthesis work and provide information that could be useful for additional research. Some of this documentation can not be located in published documents and would need to be obtained directly from contacts with the authors of the original research. The following additions to the documentation would be especially useful:

1 The structure of any screening questions (with the possible exception of questions about audibility) should be presented by directly quoting both the screening question and the actual annoyance question.

2 The number of interviews should be reported for each data point.

3 The following minimal list of characteristics should be systematically documented for every survey. Simply reporting the information when it is easily obtained or appears to be unusual is not adequate:
   (1) The relationship between the nominal noise measurement position and reflecting surfaces.
   (2) The nominal position of the noise measurement relative to the respondent's dwelling and the location of the surface transportation noise.
   (3) The extent to which the measurement procedures excluded noise that was not from the studied noise source.
   (4) The hours that delimit each time period of the original study's noise metric.
   (5) The operating conditions that the noise metric is presumed to encompass (i.e. weekdays and weekend traffic conditions for all sources; worst case or "standard" operating conditions for aircraft; only through train or maintenance operations for railways)

4 The synthesis publication should, at the very least, contain a short discussion of the possible impact of all of the factors outlined in Table 3. As a result of the discussion the reader should have some insight into whether the absence of some information on an individual survey basis could affect the outcome of the synthesis.

6.5. Improvements in analysis techniques

Analysis techniques are needed that weight data points by their relative importance and that describe the precision of the estimates.

1 Data points should be weighted by numbers of interviews

2 When possible, a data set should be analyzed that preserves the original structure of the sample. This will be either individual or study-site specific data.

3 Sampling errors should be computed with a method that accounts for the complexity of the sample design. For individual surveys the study sites should be treated as primary selection units. For combining the results from a large number of surveys, the survey should be treated as the primary selection unit.

4 The precision of the estimates should be represented by the 95% confidence intervals for the predicted regression line.

5 Tests should conducted to determine whether all surveys are drawn from the same population.

6 The dose/response relationship for a dichotomous variable should be, at a minimum, modeled with either a logit or probit model.

REFERENCES

Alexandre, Ariel: 1970. Prévision de la Gêne due Autour des Aéroports et Perspectives sur les Moyens d'y Remédier. (Prediction of Annoyance Due to Noise Around Airports and Speculations on the Means for Controlling It). Anthropologie Appliquée, Doc. A.A. 28/70.

Alexandre, Ariel: 1973. Decision Criteria Based on Spatio-temporal Comparisons of Surveys on Aircraft Noise. Proceedings of the International Congress on Noise as a Public Health Problem, Dubrovnik, Yugoslavia, May 13-18 1973, pp. 619-626. USEPA 550-9-73-008. U.S. Environmental Protection Agency, Washington D.C.

Andersen, T.V.; Kühl, K.; and Relster, E.: 1983. Reactions to Railway Noise in Denmark. J. Sound Vib., vol. 87, no. 2, pp. 311-314.

Aubree, D.: 1973. Enquête Acoustique et Sociologique Permettant de Définir une Echelle des la Gêne Eprouvée par l'Homme dans son Logement du Fait des Bruits de Train. Centre Scientifique et Technique du Bâtiment, Paris, June 1973. Translation Available as: Acoustical and Sociological Survey to Define a Scale of Annoyance Felt by People in Their Homes Due to the Noise of Railroad Trains. Technical Information Report No. 88. Bolt Beranek and Newman, Cambridge, Massachusetts.

Aubree, D.: 1975. La Gêne due au Bruit des Trains. (Annoyance Due to Train Noise). EN-SH-75.2. CSTB, Nantes.

Aubree, D.; Auzou, J.; and Rapin, M.: 1971. Etude de la Gêne due au Trafic Automobile Urbain: Compte Rendu Scientifique. (Study of Annoyance due to Urban Automobile Traffic: Scientific Report). D.G.R.S.T. No. 68-01-389. CSTB, Paris. Translations available as: Study of Annoyance Due to Urban Automobile Traffic: Scientific Report. NASA TM-77514, 1984; Study of Annoyance Due to Urban Automobile Traffic. Annex 4: A Catalog of the Characteristics of Noise at Different Measuring Points. NASA TM-77516, 1984; Study of Annoyance Due to Urban Automobile Traffic. Annex 5: Sociological Study. NASA TM-77515, 1984.

Baldwin, E.; and Fidell, S.: 1982. Evaluation of Noise Exposure and Community Response Due to Temporary Reinstitution of Night Landings at Westchester Country Airport, Spring 1982. BBN Rep. 5083. Bolt Beranek and Newman, Cambridge, Massachusetts.

Bennett, R.L.; and Pearsons, K.S.: 1981. Handbook of Aircraft Noise Metrics. NASA Contractor Report CR-3406. National Aeronautics and Space Administration, Washington D.C.

Borsky, P.N.: 1983. Integration of Multiple Aircraft Noise Exposures Over Time by Residents Living Near U.S. Air Force Bases. In Noise as a Public Health Problem, Proceedings of the Fourth International Congress on Noise as a Public Health Problem, pp. 1049-1060. Centro Ricerche e Studi Amplifon, Milano, Italy.

Borsky, P.N.: 1985. Unpublished data provided to the authors of the synthesis by Stanley Harris of Armstrong Aerospace Medical Research Laboratory, Wright Patterson, AFB, Ohio.

Centre Scientifique et Technique du Bâtiment: 1968. La Gêne Causée Par le Bruit Autour des Aéroports: Rapport de Fin D'Etude. C.S.T.B., Paris, March 1, 1968. Translation available as: The Annoyance Caused by Noise Around Airports: Final Report. NASA TM-75784.

Committee on Community Reactions to the Concorde: 1976. Review of FAA Concorde Survey Plans. National Research Council of the Assembly of Behavioral and Social Sciences of the National Academy of Sciences, Washington D.C.

Connor, W.K.; and Patterson, H.P.: 1972. Community Reaction to Aircraft Noise Around Smaller City Airports. NASA CR-2104.

Deutsche Forschungsgemeinschaft: 1974. DFG-Forschungsbericht Fluglärmwirkungen: Eine Interdisziplinäre Untersuchung über die Auswirkungen des Fluglärms auf den Menschen. (DFG-Research-Report Aircraft Noise: An Interdisciplinary Investigation of the Effect of Aircraft Noise on People). Harald Boldt Verlag KG, Boppard.

Federal Interagency Committee on Noise (FICON): 1992. Federal Agency Review of Selected Airport Noise Analysis Issues. Washington, D.C.: FICON.

Fidell, S.: 1978. Nationwide Urban Noise Survey. J. Acoust. Soc. Am., vol. 64, no. 1, pp. 198-206.

Fidell, S.; Barber, D.S.; and Schultz, T.J.: 1989. Updating a Dosage-effect Relationship for the Prevalence of Annoyance due to General Transportation Noise. Report HSD-TR-89-009 (BBN Report 6835). Human Systems Division, Brooks Air Force Base, Texas.

Fidell, S.; Barber, D.S. and Schultz, T.J.: 1991. Updating a Dosage-effect Relationship for the Prevalence of Annoyance due to General Transportation Noise. J. Acoust. Soc. Am. vol 89. pp. 221-233.

Fidell, S.; Green, D.M.; Schultz, T.J.; and Pearsons, Karl S.: 1987. A Strategy for Understanding Noise-Induced Annoyance. BBN Report 6337. BBN Laboratories Incorporated, Canoga Park, CA.

Fidell, S.; Horonjeff, R.; Mills, J.; Baldwin, E.; Teffeteller, S.; and Pearsons, K.S.: 1985. Aircraft Noise Annoyance at Three Joint Air Carrier and General Aviation Airports. J.

Acoust. Soc. Am., vol. 77, no. 3, pp. 1054-1068.

Fidell, S.; and Jones, G.: 1975. Effects of Cessation of Late-Night Flights on an Airport Community. J. Sound Vib., vol. 42, no. 4, pp. 441-427.

Fidell, S.; Schultz, T.; and Green, D.M.: 1988. A Theoretical Interpretation of the Prevalence Rate of Noise-induced Annoyance in Residential Populations. J. Acoust. Soc. Am. vol 84. pp. 2109-2113.

Fields, J.M.: 1980. A Program to Support the Full Utilization of Data from Existing Social Surveys of Environmental Noise. Proceedings of Inter-Noise 80. pp. 937-940.

Fields, J.M.: 1981. A Catalog of Social Surveys of Residents' Reactions to Environmental Noise (1943-1980). NASA TM-83187. NASA, Washington D.C.

Fields, J.M.: 1982. Effects of Errors in Specifying Noise Environments on Results from Community Response Surveys, Proceedings of Inter-Noise 82, pp. 609-612.

Fields, J.M.: 1983. Variability in Individuals' Responses to Noise: Community Differences. Proceedings of Inter-Noise 83, pp. 965-968.

Fields, J.M.: 1991. An Updated Catalog of 318 Social Surveys of Residents' Reactions to Environmental Noise (1943-1989). NASA TM-187553. NASA, Washington D.C.

Fields, J.M.; and Hall, F.L.: 1987. Community Effects of Noise. Chapter 3 in Nelson, P.M. Handbook of Transportation Noise. Butterworths, London, England.

Fields, J.M.; and Walker, J.G.: 1980. Comparing Reactions to Transportation Noises From Different Surveys: a Railway Noise vs. Aircraft and Road Traffic Comparison. In Noise as a Public Health Problem, Proceedings of the Third International Congress, pp. 580-587. ASHA Report 10. American Speech-Language-Hearing Association, Rockville, Maryland.

Fields, J.M.; and Walker, J.G.: 1982a. Comparing the Relationships Between Noise Level and Annoyance in Different Surveys: A Railway Noise vs. Aircraft and Road Traffic Comparison. J. Sound Vib., vol. 81, pp. 51-80.

Fields, J.M.; and Walker, J.G.: 1982b. The Response to Railway Noise in Residential Areas in Great Britain. J. Sound Vib., vol. 85, pp. 177-255.

Fog, Hans; and Jonsson, Erland: 1968. Traffic Noise in Residential Areas. Report 36 E/1986. The National Swedish Institute for Building Research, Stockholm.

Garnsworthy, John: 1977. A Study of Question Order and Wording Experiments. MSc. Dissertation. Dept. of Social Statistics, University of Southampton, England.

Graf, P.; Meier, H.P.; and Müller, R.: 1974a. Sozio-psychologische Fluglärmuntersuchung im Gebiet der drei Schweizer Flughäfen, Zürich, Genf, Basel. Bern, June 1973. Translation available as: Sociopsychological Airplane Noise Investigation in the Districts of Three Swiss Airports: Zurich, Geneva, and Basel. NASA TM-75787.

Graf, P.; Meier, H.P.; and Müller, R.: 1974b. Enquêtes socio-psychologiques sur le bruit des avions:dans le périmètre des trois aéroports suisses de Zurich, Genéve et Bâle. (Sociopsychological Airplane Noise Investigation in the Districts of Three Swiss Airports: Zurich, Geneva, and Bern.) Berne.

Grandjean, E.; Graf, P.; Lauber, A.; Meier, H.P.; and Mueller, R.: 1973. A Survey of Aircraft in Switzerland. Proceedings of the International Congress on Noise as a Public Health Problem, Dubrovnik, Yugoslavia, May 13-18 1973, pp. 745-659. USEPA 550-9-73-008. U.S. Environmental Protection Agency, Washington D.C.

Griffiths, I.D.: 1983. Review of Community Response to Noise. In Noise as a Public Health Problem, Proceedings of the Fourth International Congress on Noise as a Public Health Problem, pp. 1131-1048. Centro Ricerche e Studi Amplifon, Milano, Italy.

Hall, F.L.; Birnie, S.E.; Taylor, S. M.; and Palmer, J.E.: 1981. Direct Comparison of Community Response to Road Traffic Noise and to Aircraft Noise. J. Acoust. Soc. Am., vol. 70, no. 6, pp. 1690-1698.

Hall, F.L.; and Taylor, S.M.: 1977. Predicting Community Response to Road Traffic Noise. J. Sound Vib., vol. 52, no. 3, pp. 387-399.

Hede, A.J.; and Bullen, R.B.: 1982. Aircraft Noise in Australia: A Survey of Community Reaction. NAL Report 88. National Acoustic Laboratories, Commonwealth Dept. of Health, Canberra, Australia.

Information..: 1974. Information on Levels of Environmental Noise Requisite to Protect Public Health and Welfare with an Adequate Margin of Safety. U.S. Environmental Protection Agency, Arlington, VA.

Johnston, J.: 1972. Econometric Methods. McGraw-Hill, London.

Kalton, Graham: 1983. Introduction to Survey Sampling. Sage Publications, London.

Kragh, Jørgen: 1977. Analyse af Sammenhængen mellem Reaktioner på Vejtrafikstøj og Fysiske Mål for Støjen. (Analysis of the Correlation Between Reactions to Road Traffic Noise and Physical Indices of the Noise). Report no. 5. Acoustical Laboratory, Danish Academy of Technical Sciences, Lyngby.

Kryter, Karl D.: 1982a. Community Annoyance from Aircraft and Ground Vehicle Noise. J.

Acoust. Soc. of America, vol. 72, pp. 1222-1241.

Kryter, Karl D.: 1982b. Rebuttal by Karl D. Kryter to comments by T.J. Schultz. J. Acoust. Soc. of America, vol. 72, pp. 1253-1257.

Kryter, Karl D.: 1983. Response of K.D. Kryter to modified comments by T.J.Schultz on K.D. Kryter's paper, "Community annoyance from aircraft and ground vehicle noise"[J. Acoust. Soc. Am. 72, 1243-1252 (1982)]. J. Acoust. Soc. of America, vol 73, pp. 1066-1068.

Lamure, C.; and Bacelon, M.: 1967. La Gêne due au Bruit de la Circulation Automobile: Une Enquête Auprès de Riverains d'Autoroutes. Cahiers du Centre Scientifique et Technique du Bâtiment, no. 88. C.S.T.B., Paris. Translation available as: The Nuisance Due to the Noise of Automobile Traffic: An Investigation in the Neighborhoods of Freeways. NASA TM-75812.

Langdon, F.J.: 1976. Noise Nuisance Caused by Road Traffic in Residential Areas: Parts I, II. J. Sound Vib., vol. 47, no. 2, pp. 243-282.

Lee, Sul Lee; Forthofer, Ronald N.; and Lorimor, Ronald J.: 1989. Analyzing Complex Survey Data. Sage Publications, London.

McKennell, A.C.: 1963. Aircraft Noise Annoyance Around London (Heathrow) Airport. S.S.337. The Government Social Survey, Central Office of Information, London.

MIL Research: 1971. Second Survey of Aircraft Noise Annoyance Around London (Heathrow) Airport. Her Majesty's Stationery Office, London.

Myncke, H.; Cops, A.; Steenackers, P.; Bruyninckx, W.; Gambart, R.; and Verleysen, P.: 1977. Studie van het Verkeerslawaai in Steden en de Hinder Ervan voor de Bevolking. (Study of Urban Traffic Noise and the Resulting Annoyance Felt by the Population). Laboratorium voor Akoestiek en Warmtegeleiding, K. U.-Leuven, Ministerie van Volksgezondheid en van Het Gezin, Dienst Milieuhinder.

Reaktioner ... : 1982. Reaktioner på togstøj: En Analyse af Interviews Blandt Beboere Langs Danske Jernbaner. Miljø-Projekter, Miljøstyrelsen, Copenhagen.

Reaktioner..: 1982. Reaktioner på Togstøj: En Analyse af Interviews Blandt Beboere Langs Danske Jernbaner. Miljø-Projekter, Miljøstyrelsen, Copenhagen, February 1982.

Relster, Else: 1975. Traffic Noise-Annoyance, The Psychological Effect of Traffic Noise in Housing Areas. Polyteknisk Forlag, Lyngby, Denmark.

Rohrmann, B.: 1989. Personal communication.

Rohrmann, B.; Schümer, R.; Schümer-Kohrs, A.; Guski, R.; and Finke, H.-O.: 1973. An Interdisciplinary Study on the Effects of Aircraft Noise on Man. Proceedings of the International Congress on Noise as a Public Health Problem, Dubrovnik, Yugoslavia, May 13-18, pp. 765-776. USEPA 550-9-73-008. U.S. Environmental Protection Agency, Washington D.C.

Rosenthal, R.: 1984. Meta-analytic Procedures for Social Research. Sage Publications, Beverly Hills, CA.

Rylander, R.; Björkman, M.; Åhrlin, U.; and Sörensen, S.: 1977. Tramway Noise in City Traffic. J. Sound Vib., vol. 51, no. 3, pp. 353-358.

Rylander, R.; Sörensen, S.; and Kajland, A.: 1972. Annoyance Reactions from Aircraft Noise Exposure. J. Sound Vib., vol. 24, no. 4, pp. 419-444.

Schomer, Paul D.: 1983. A Survey of Community Attitudes Towards Noise Near a General Aviation Airport. J. Acoust. Soc. Am., vol. 74, no. 6, pp. 1773-1781.

Schultz, T.J.: 1978. Synthesis of Social Surveys on Noise Annoyance. J. Acoust. Soc. of Am., vol. 64, pp. 377-405.

Schultz, T.J.: 1979. Erratum: "Synthesis of social surveys on noise annoyance: [J. Acoust. Soc. Am. 64, 377-405 (1978)]. J Acoust. Soc. Am., vol. 651, p. 849.

Schultz, T.J.: 1982. Comments on K.D. Kryter's paper "Community annoyance from aircraft and ground vehicle noise". J. Acoust. Soc. Am., vol. 72, pp. 1243-1252.

Schuman, H.; and Presser, S.: 1981. Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context. Academic Press, New York.

Sörensen, S.; and Hammar, N.: 1983. Annoyance Reactions Due to Railway Noise. J. Sound Vib., vol. 87, no. 2, pp. 315-319.

Wachter, K.W. and Straf, M.L. (Eds.): 1990. The Future of Meta-analysis. Russell Sage Foundation, New York.

Wilson, Alan: 1963. Noise: Final Report. Cmnd. 2056. Her Majesty's Stationery Office, London.

Wolf, F.M.: 1986. Meta-analysis: Quantitative Methods for Research Synthesis. Sage Publications, Beverly Hills, CA.

TABLES

Table 1: Characteristics of 29 data sets

Data set definition				Data point description			Annoyance measurement information (Also see Table 2)			Noise data reported in original survey			Comments: (Appendices A and B expand on these comments.) (P=Problem data set, identified in Figure 3)
Data set title (Publication cited in synthesis)	Synthesis year (Page in article)	Catalog ID (See: Fields, 1991)	Number of individual responses	Type of data point: Area Category Phase	Number of data points	Data point sizes range from ..⇒..to..	Type of question: Verbal Numeric Index Ranking	Definition of high annoyance: Basis: % or word	Definition of high annoyance: % of scale	Noise metric	Number of hours in metric	Is night-time noise isolated?	
"Highly annoyed" Verbal Scale													
Australian Aircraft (Hede and Bullen, 1982)	1991 (p. 222)	AUL-210	2718	Category	42	40 ⇒ 112 (C)	Verbal-5	Highly	20%	DNL	24	✓	Hede and Bullen believe that this definition of high annoyance is too severe.
Single Verbal 3-Point Scale (Single annoyance question)													
Danish Street (Relster, 1975)	1978 (p. 400)	DEN-075	960 (Not checked)	Area	28	25 ⇒ 60 (approx.)	Verbal-3	Meget (Much)	33%	L_{eq} for each period (9-hour night)	24	✓ (9 hr night)	The survey question asks about noise generally, not traffic noise. The published DNL was reduced by 5 dB because measurements were made at the facade. (P)
Single Verbal 4-Point Scale (Single annoyance question)													
French Aircraft (Alexandre, 1970)	1978 (p. 392)	FRA-016	2000	Category	6	100 ⇒ 500 (approx)	Verbal-4	Beaucoup (A great deal)	25%	R (Isopsophic index) PNDB + Number	24	No	Annoyance scores incorrect by 2% to 5% for 5 of 6 points.
British Railroad (Fields and Walker, 1982)	1991 (p. 225)	UKD-116	1399	Category	11	48 ⇒ 220	Verbal-4	Very much	25%	DNL	24	✓	None
Joined Verbal 1+3-Point Scale (Initial annoyance screening question followed by 3-point verbal rating scale)													
Swedish Aircraft (Rylander et al., 1972)	1978 (p. 397)	SWE-035	1662 (C)	Area	17	86 ⇒ 156	Verbal-1+3	Mycket (Very)	25%	NEF & CNR(primary path), CNR_s(all-paths)	24	No	Copies of the three original language questionnaires have not been located.
Tramway/Traffic Comparison (Tramway only) (Rylander, et al., 1977)	1991 (p. 225)	SWE-165	464	Area	6	73 ⇒ 91	Verbal-1+3	Mycket (Very)	25%	L_{eq}	24	No	The day/night division for tram noise is based on road traffic data from other USA and Swedish cities.

Data set definition				Data point description			Annoyance measurement information (Also see Table 2)			Noise data reported in original survey			Comments: (Appendices A and B expand on these comments.) (P=Problem data set, identified in Figure 3)
Data set title (Publication cited in synthesis)	Synthesis year (Page in article)	Catalog ID (See: Fields, 1991)	Number of individual responses	Type of data point: Area Category Phase	Number of data points	Data point sizes range from ..⇒..to..	Type of question: Verbal Numeric Index Ranking	Definition of high annoyance: Basis: % or word	Definition of high annoyance: % of scale	Noise metric	Number of hours in metric	Is night-time noise isolated?	
Tramway/Traffic Comparison (Traffic only) (Rylander, et al., 1977)	1991 (p. 225)	SWE-165	464	Area	6	73 ⇒ 91	Verbal-1+3	Mycket (Very)	25%	L_{eq}	24	No	The synthesis used the combined noise exposure not the road noise exposure. (P)
Swedish Railroad (Sörensen and Hammar, 1983)	1991 (p. 226)	SWE-228	1125 (Not checked)	Area	15	50 ⇒ 100 (approx)	Verbal-1+3	Mycket (Very)	25%	L_{eq}	24	No	None
Single Verbal 5-Point (Single annoyance question)													
Burbank Airport (Fidell et al., 1985)	1991 (p. 223)	USA-203	5041 (Multiple responses)	4 Area × 5 Phase	20	x=250	Verbal-5	Very, Extremely	40%	DNL	24	✓	Short-term noise measurements in a changing environment do not match the survey question. Respondents with missing data are assumed to not be annoyed. (P)
Los Angeles Airport (Fidell and Jones, 1975)	1978 (not described)	USA-082	Not Reported (N=1417 in catalog)	Area	2	(Not in publication)	Verbal-5	Very, Extremely	40%	DNL	24	✓	One of the two data points should be reduced by 5 dB.
USA 24 Site (Fidell, 1978)	1978 (p. 399)	USA-102	2037 (C)	Area	24	65 ⇒ 137	Verbal-5	Very, Extremely	40%	DNL	24	✓	The survey question asks about "neighborhood noise" not traffic noise. (P)
Joined Verbal 1+4-Point Scale (Initial annoyance screening question followed by 4-point verbal rating scale)													
Orange Country Airport (Fidell et al., 1985)	1991 (p. 224)	USA-204	3103 (Multiple responses)	3 Area × 4 Phase	12	195 ⇒ 336	Verbal-1+4	Very, Extremely	40%	DNL	24	✓	None
Westchester Airport (Fidell et al., 1985)	1991 (p. 228)	USA-301	1416 (C)	4 Area × 2 Phase	8	104 ⇒ 258	Verbal-1+4	Very, Extremely	40%	DNL	24	✓	The noise data are for all community noise. Respondents rated a short period: "Winter" for May interviews, "Spring" for June interviews. (P)
Decatur Airport (Schomer, 1983)	1991 (p. 225)	USA-250	231	Category	4	29 ⇒ 99	Verbal-1+4	Very much, Extremely	40%	DNL	24	✓	The synthesis article raises a non-existent question wording problem.

Data set definition				Data point description			Annoyance measurement information (Also see Table 2)			Noise data reported in original survey			Comments: (Appendices A and B expand on these comments.) (P=Problem data set, identified in Figure 3)
Data set title (Publication cited in synthesis)	Synthesis year (Page in article)	Catalog ID (See: Fields, 1991)	Number of individual responses	Type of data point: Area Category Phase	Number of data points	Data point sizes range from ..⇒..to..	Type of question: Verbal Numeric Index Ranking	Definition of high annoyance: Basis: % or word	Definition of high annoyance: % of scale	Noise metric	Number of hours in metric	Is night-time noise isolated?	
Bipolar Verbal 5/9-Point Scale (The 9-point scale has 5 that are not positive)													
Canadian Road (Hall and Taylor, 1977)	1978 (p. 402)	CAN-121 (1976 part)	410 (Not checked)	Area	14	22 ⇒ 31 (approx.)	Verbal-B-5/9 bipolar 5 or 9 points	Considerably, Extremely	40% (2/5) or 22% (2/9)	L_{eq} (16-hour day), L_{eq} (8-hour night)	24	✓ (8-hour night)	The survey question asks about "overall noise", not traffic noise. (P)
Aircraft/Traffic Comparison (A/C ONLY) (Hall et al., 1981)	1991 (p. 222)	CAN-168	673	Category	9	11 ⇒ 181 (C)	Verbal-B-5/9 bipolar 5 or 9 points	Considerably, Extremely	40% (2/5) or 22% (2/9)	DNL	24	✓	One data point is off by 0.6%.
Aircraft/Traffic Comparison (TRAFFIC ONLY) (Hall et al., 1981)	1991 (p. 222)	CAN-168	638 (C)	Category	12	13 ⇒ 158	Index-3 or Verbal-B-5/9 bipolar 5 or 9 points (highest of three sources)	Considerably, Extremely	40% (2/5) or 22% (2/9)	DNL	24	✓	Three of the published data points are not in the synthesis. Annoyance is defined by the highest annoyance expressed on any of three questions. (P)
Numeric Scales (7 to 11 points)													
London Traffic (Langdon, 1976)	1978 (p. 399)	UKD-071	1539 (Not checked)	Area	24	$\bar{x}$≈55	Numeric-7	%	29%	DNL	24	✓	Measurements are made near the facade but not normalized in the synthesis.
French Railroad (Aubree, 1975)	1978 (p. 398)	FRA-063	344 (C)	Category	5	10 ⇒ 146	Numeric-7	%	29%	L_{eq}	24	No	No nighttime data and the synthesis assumes there is none, but respondents report nighttime traffic.
U.S.Airbase (Borsky, 1985)	1991 (p. 227)	USA-338 [Not in catalog]	874 (Not checked)	Not known	25	27 ⇒ 45	Numeric-10	%	30%	DNL	24	✓	Annoyance question refers to the short period ("past few weeks"). The synthesis data that come from a private letter disagree with a published article (Borsky, 1983).

Data set definition				Data point description			Annoyance measurement information (Also see Table 2)			Noise data reported in original survey			Comments: (Appendices A and B expand on these comments.) (P=Problem data set, identified in Figure 3)
Data set title (Publication cited in synthesis)	Synthesis year (Page in article)	Catalog ID (See: Fields, 1991)	Number of individual responses	Type of data point: Area Category Phase	Number of data points	Data point sizes range from ..⇒..to..	Type of question: Verbal Numeric Index Ranking	Definition of high annoyance: Basis: % or word	Definition of high annoyance: % of scale	Noise metric	Number of hours in metric	Is night-time noise isolated?	
Swiss Road (ROAD ONLY) (Grandjean et al., 1973; Graf, et al., 1974a)	1978 (p. 396)	SWI-053	945	Category	6	59 ⇒ 333	Numeric-11	%	27%	L_{50} (06 - 18)	16	No	End points of scales are more extreme than described in synthesis article. Measurements on 2nd floor balcony railings.
Swiss Aircraft (AIRCRAFT ONLY) (Grandjean et al., 1973)	1978 (p. 399)	SWI-053	3711 (C)	Category	12	11 ⇒ 585	Numeric-11	%	27%	NNI (16 hour)	16	No	Inappropriate cumulative percentages were plotted. One data point not entered. (See SWI-053-Road for annoyance label.) (P)
Numeric - Marking position on unmarked 155 millimeter line													
Antwerp Street (Myncke et al., 1977)	1978 (p. 401)	BEL-122	Unknown (1279 in survey) (Not checked)	Area	31	9 ⇒ 85 (approx.)	Numeric-line marking	%	30%	L_{eq} (16-hour day), L_{eq} (8-hour night)	24	✓ (all but 2 areas) (8-hour night)	The data in the synthesis do not match the original report. Noise measurements were made at the sidewalk.
Brussels Street (Myncke et al., 1977)	1978 (p. 401)	BEL-137	434 (C)	Area	23	5 ⇒ 30	Numeric-line marking	%	30%	L_{eq} (16-hour day or either day + 4-hour evening or 8-hour night)	12, 16 or 20 hours	✓ (about 1/2 of areas) (8-hour night)	One data point not entered. Noise measurements were made at the sidewalk.
Multiple-Item Index of 6 to 37 questions													
First Heathrow Airport (McKennell, 1963)	1978 (p. 391)	UKD-008	1731	Category	10	11 ⇒ 512	Index-6	%	29%	PNDB (Max) N (>80 PNDB) 16 hr	16	No	One data point not entered. Empirical data and McKennell's opinions support a less severe definition of "serious annoyance".
Second Heathrow Airport (MIL Research, 1971)	1978 (p. 393)	UKD-024	4678	10 Category × 2 Noise descriptors	20	68 ⇒ 1186	Index-6	%	29%	NNI (16-hour)	16	No	Empirical data from UKD-008 support a less severe definition of "serious annoyance". Responses of those within 10 miles of Heathrow are double counted.

Data set definition				Data point description			Annoyance measurement information (Also see Table 2)			Noise data reported in original survey			Comments: (Appendices A and B expand on these comments.) (P=Problem data set, identified in Figure 3)
Data set title (Publication cited in synthesis)	Synthesis year (Page in article)	Catalog ID (See: Fields, 1991)	Number of individual responses	Type of data point: Area Category Phase	Number of data points	Data point sizes range from ..⇒..to..	Type of question: Verbal Numeric Index Ranking	Definition of high annoyance: Basis: % or word	Definition of high annoyance: % of scale	Noise metric	Number of hours in metric	Is night-time noise isolated?	
Munich Aircraft (Rohrmann et al., 1973; Deutsche Forschungsgemeinschaft, 1974)	1978 (p. 396)	GER-034	660 (Not checked)	Area	27	$\bar{x}$≈20	Index-11	%	(See Appendix A)	FB1	16	No	The synthesis annoyance measure is the median annoyance scored on a 37-item index. Three published data points are not entered. (P)
Multiple-Item Index of 2 Questions													
Danish Railroad (Andersen et al., 1983)	1991 (p. 228)	DEN-200	583 (C)	Category	26	1 ⇒ 55	Index-2 (Volunteer or Verbal-4)	Meget (very much) or volunteer train noise as unsatisfactory aspect of neighborhood	25% or volunteer	L_{eq}	24	No	One annoyance score is about 11 % too low. One noise level is 1 dB high. The annoyance scale scores a volunteered dissatisfaction as high annoyance. (P)
Ranking Scale													
Paris Street (Aubree et al., 1971)	1978 (p. 395)	FRA-041	700 (Not checked)	Category	8	$\bar{x}$≈45	Ranking-10	(RANK)	rank 1st	L_{50}	24	No	Annoyance is measured relative to any other disturbances (P)

Table 2: Annoyance questions used in 29 clustering data sets

Data set definition			Annoyance measurement information				
Data set title (Publication cited in synthesis)	Synthesis year (Page in article)	Catalog ID (See: Fields, 1991)	Type of question: Verbal Numeric Index Ranking	Wording of question ([... text] identifies the text that follows each alternative answer in the question.)	Answers (The synthesis definition of "high" annoyance is double underlined.) (R=This scale is in reversed order, high annoyance came first in the questionnaire.)	Definition of high annoyance:	
						Basis: % or word	% of scale
"Highly annoyed" Verbal Scale							
Australian Aircraft (Hede and Bullen, 1982)	1991 (p. 222)	AUL-210	Verbal-5	Q.63 How would you describe your general feelings about the aircraft noise in this neighborhood? Would you say you are: [..annoyed]	Not at all.., Slightly.., Moderately.., Considerably.., Highly.. (R)	Highly	20%
Single Verbal 3-Point Scale (Single annoyance question)							
Danish Street (Relster, 1975)	1978 (p. 400)	DEN-075	Verbal-3	(Follows questions about the amount of noise from children, neighbors, traffic, aircraft, and other sources.) [DANISH] Q.29 Do you feel much, a little, or not at all annoyed by the noise you are exposed to?	Ikke spor, Lidt, Meget (Not at all, A little, Much) (R)	Meget (Much)	33%
Single Verbal 4-Point Scale (Single annoyance question)							
French Aircraft (Alexandre, 1970)	1978 (p. 392)	FRA-016	Verbal-4	[FRENCH] Does airplane noise bother you..	Pas du tout, Un peu, Assez, Beaucoup (Not at all, Somewhat, Quite a bit, a great deal) (R)	Beaucoup (A great deal)	25%
British Railroad (Fields and Walker, 1982)	1991 (p. 225)	UKD-116	Verbal-4	Q.17.b Does the noise of the trains bother or annoy you:	Not at all, A little, Moderately, Very much (R)	Very much	25%
Joined Verbal 1+3-Point Scale (Initial annoyance screening question followed by 3-point verbal rating scale)							
Swedish Aircraft (Rylander et al., 1972)	1978 (p. 397)	SWE-035	Verbal-1+3	[SWEDISH, NORWEGIAN, DANISH] In your home are you annoyed by aircraft noise? [IF YES ASK] How annoyed are you? [H] [Wording based on author's memory]	NEJ + Inte särskilt mycket, Ganska mycket, mycket (NO + Not very, rather, very)	Very	25%
Tramway/Traffic Comparison (Tramway only) (Rylander, et al., 1977)	1991 (p. 225)	SWE-165	Verbal-1+3	[SWEDISH] Q.23.a In your home are you annoyed by the noise from trams? [IF YES ASK] Q.24.a How annoyed are you by trams? [H]	NEJ + Inte särskilt mycket, Ganska mycket, mycket (NO + Not very, rather, very)	Mycket (Very)	25%

Data set definition			Annoyance measurement information				
Data set title (*Publication cited* in synthesis)	Synthesis year (Page in article)	Catalog ID (See: Fields, 1991)	Type of question: Verbal Numeric Index Ranking	Wording of question ([... text] identifies the text that follows each alternative answer in the question.)	Answers (The synthesis definition of "high" annoyance is double underlined.) (R=This scale is in reversed order, high annoyance came first in the questionnaire.)	Definition of high annoyance:	
						Basis: % or word	% of scale
Tramway/Traffic Comparison (Traffic only) (Rylander, et al., 1977)	1991 (p. 225)	SWE-165	Verbal-1+3	[SWEDISH] Q.23.b *In your home are you annoyed by the noise from other traffic?* [IF YES ASK] Q.24.b How annoyed are you by other traffic? [H]	NEJ + Inte särskilt mycket, Ganska mycket, Mycket (NO + Not very, rather, very)	Mycket (Very)	25%
Swedish Railroad (Sörensen and Hammar, 1983)	1991 (p. 226)	SWE-228	Verbal-1+3	[SWEDISH] Q.24 Are you annoyed by railway noise? [IF YES ASK] How much are you annoyed? [H]	NEJ + Inte särskilt mycket, Ganska mycket, Mycket (NO + Not very, rather, very)	Mycket (Very)	25%
Single Verbal 5-Point (Single annoyance question)							
Burbank Airport (Fidell et al., 1985)	1991 (p. 223)	USA-203	Verbal-5	Q.4 While you've been at home over the past YEAR (since this time last year), would you say that you've been [... annoyed by aircraft noise, ..]?	Not at all.., Slightly.., Moderately.., Very.., Extremely..	Very, Extremely	40%
Los Angeles Airport (Fidell and Jones, 1975)	1978 (not described)	USA-082	Verbal-5	[ENGLISH AND SPANISH] Q.1 Are you ever annoyed by aircraft noise in your neighborhood? (ALLOW A FREE RESPONSE; DO NOT RECORD REPLY) I see. Now I need an answer that I can compare with the answers that other people give me. Would you say you've been [...annoyed,...]	Not at all.., Slightly.., Moderately.., Very.., Extremely.. SPANISH VERSION Definitivamente no, Un poco, Moderadamente, Bastante, Extremamente	Very, Extremely	40%
USA 24 Site (Fidell, 1978)	1978 (p. 399)	USA-102	Verbal-5	Q.14 How ANNOYING was the noise in your neighborhood over the past year?	Not at all, Slightly, Moderately, Very, Extremely	Very, Extremely	40%
Joined Verbal 1+4-Point Scale (Initial annoyance screening question followed by 4-point verbal rating scale)							
Orange Country Airport (Fidell et al., 1985)	1991 (p. 224)	USA-204	Verbal-1+4	Q.4 While you've been at home over the past year, since (last season of year), have you been bothered or annoyed by the noise of large airliners? IF YES, ASK: Would you say you've been slightly annoyed by noise from large airliners, moderately annoyed, very annoyed, or extremely annoyed by noise from large airliners?	NO + Slightly.., Moderately.., Very.., Extremely..	Very, Extremely	40%
Westchester Airport (Fidell et al., 1985)	1991 (p. 228)	USA-301	Verbal-1+4	Q.4. And how about this past (Fall and Winter, since October../..Spring, since early May) have you been bothered or annoyed by noise from airplanes while you've been at home during these months? [The wording of the direct question following this screening question was not reported.]	NO + Slightly.., Moderately.., Very.., Extremely..	Very, Extremely	40%
Decatur Airport (Schomer, 1983)	1991 (p. 225)	USA-250	Verbal-1+4	Q.7.e In general, taking everything into consideration, does the noise from airplanes ever bother or annoy you? If yes f. Overall how annoyed are you by the noise from airplanes? Would you say...	NO + Slightly, Moderately, Very much, Extremely (R)	Very much, Extremely	40%

Data set definition			Annoyance measurement information				
Data set title (Publication cited in synthesis)	Synthesis year (Page in article)	Catalog ID (See: Fields, 1991)	Type of question: Verbal Numeric Index Ranking	Wording of question ([... text] identifies the text that follows each alternative answer in the question.)	Answers (The synthesis definition of "high" annoyance is double underlined.) (R=This scale is in reversed order, high annoyance came first in the questionnaire.)	Definition of high annoyance: Basis: % or word	Definition of high annoyance: % of scale
Bipolar Verbal 5/9-Point Scale (The 9-point scale has 5 that are not positive)							
Canadian Road (Hall and Taylor, 1977)	1978 (p. 402)	CAN-121 (1976 part)	Verbal-B-5/9 bipolar 5 or 9 points	Q.6 How would you rate the overall noise..[..agreeable.. or ..disturbing]?	Extremely.., Considerably.., Moderately.., Slightly.., Neutral, Slightly.., Moderately.., Considerably.., Extremely..	Considerably, Extremely	40% (2/5) or 22% (2/9)
Aircraft/Traffic Comparison (A/C ONLY) (Hall et al., 1981)	1991 (p. 222)	CAN-168	Verbal-B-5/9 bipolar 5 or 9 points	Q.2.b. How do you rate each of the sounds you have mentioned ..Aircraft.. [..agreeable.. or ..disturbing]?	Extremely.., Considerably.., Moderately.., Slightly.., Neutral, Slightly.., Moderately.., Considerably.., Extremely..	Considerably, Extremely	40% (2/5) or 22% (2/9)
Aircraft/Traffic Comparison (TRAFFIC ONLY) (Hall et al., 1981)	1991 (p. 222)	CAN-168	Index-3 or Verbal-B-5/9 bipolar 5 or 9 points (highest of three sources)	Q.2.b. How do you rate each of the sounds you have mentioned..Main road traffic noise..Local traffic noise...Trucks.. [..agreeable.. or ..disturbing]?	Extremely.., Considerably.., Moderately.., Slightly.., Neutral, Slightly.., Moderately.., Considerably.., Extremely..	Considerably, Extremely	40% (2/5) or 22% (2/9)
Numeric Scales (7 to 11 points)							
London Traffic (Langdon, 1976)	1978 (p. 399)	UKD-071	Numeric-7	Q.15 I want to ask you how you feel about traffic noise here where you live. Looking at this card would you tell me which number best represents how you feel?	Definitely satisfactory 1 2 3 4 5 6 7 Definitely unsatisfactory	%	29%
French Railroad (Aubree, 1975)	1978 (p. 398)	FRA-063	Numeric-7	[FRENCH] Q.57 From a general point of view, the train noise, in your opinion, is:	Très supportable 1 2 3 4 5 6 7 Tout à fait intolérable (Quite acceptable ... Completely intolerable)	%	29%
U.S.Airbase (Borsky, 1985)	1991 (p. 227)	USA-338 [Not in catalog. No published report.]	Numeric-10	Q.11 Now, in general, taking everything into consideration, how much did the noise from airplanes flying by disturb, bother, or annoy you during the last few weeks?	Not at all 0 1 2 3 4 5 6 7 8 9 Extremely	%	30%

Data set definition			Annoyance measurement information				
Data set title (Publication cited in synthesis)	Synthesis year (Page in article)	Catalog ID (See: Fields, 1991)	Type of question: Verbal Numeric Index Ranking	Wording of question ([... text] identifies the text that follows each alternative answer in the question.)	Answers (The synthesis definition of "high" annoyance is double underlined.) (R=This scale is in reversed order, high annoyance came first in the questionnaire.)	Definition of high annoyance:	
						Basis: % or word	% of scale
Swiss Road (ROAD ONLY) (Grandjean et al., 1973; Graf, et al., 1974a)	1978 (p. 396)	SWI-053	Numeric-11	[GERMAN] Q.45 Let's assume that this would be a thermometer with which it is possible to measure how much street traffic noise disturbs you at home. The number 10 means that you find street traffic noise unbearable, the number 0 that it doesn't disturb you at all. Please tell me the number that applies to you.	Überhaupt keine Störung 0 1 2 3 4 5 6 7 8 9 10 Unerträgliche Störung (Not at all disturbing ... Unbearably disturbing)	%	27%
Swiss Aircraft (AIRCRAFT ONLY) (Grandjean et al., 1973)	1978 (p. 399)	SWI-053	Numeric-11	[GERMAN, FRENCH, ITALIAN] Now, lets apply the thermometer to airplane noise.	IN GERMAN Überhaupt keine Störung 0 1 2 3 4 5 6 7 8 9 10 Unerträgliche Störung IN FRENCH Absolument aucune gêne....Gêne insupportable (Not at all disturbing ... Unbearably disturbing) (R)	%	27%
Numeric - Marking position on unmarked 155 millimeter line							
Antwerp Street (Myncke et al., 1977)	1978 (p. 401)	BEL-122	Numeric-line marking	[FLEMISH] (Respondents were instructed to place a mark on a line to indicate their degree of annoyance.) Q.5 We would like to know your opinion of the traffic noise that you hear in your residence during the day.	Helemaal niet storend _____ Zeer storend (Not disturbing at all ... Very disturbing) [The respondent placed a mark at 109 millimeters or further on the 155 millimeter line.]	%	30%
Brussels Street (Myncke et al., 1977)	1978 (p. 401)	BEL-137	Numeric-line marking	[FLEMISH, FRENCH] (Respondents were instructed to place a mark on a line to indicate their degree of annoyance.) Q.5 We would like to know your opinion of the traffic noise that you hear in your residence during the day.	IN FLEMISH Helemaal niet storend _____ Zeer storend IN FRENCH Ne pas génant du tout .... Très gênant (Not disturbing at all ... Very disturbing) [The respondent placed a mark at 109 millimeters or further on the 155 millimeter line.]	%	30%
Multiple-Item Index of 6 to 37 questions							
First Heathrow Airport (McKennell, 1963)	1978 (p. 391)	UKD-008	Index-6	One overall question and five activity interference questions about aircraft noise (wake up, radio/tv, vibrate, conversation, other)	(This multi-item index is described in Appendix A.)	%	29%

Data set definition			Annoyance measurement information				
Data set title (Publication cited in synthesis)	Synthesis year (Page in article)	Catalog ID (See: Fields, 1991)	Type of question: Verbal Numeric Index Ranking	Wording of question ([... text] identifies the text that follows each alternative answer in the question.)	Answers (The synthesis definition of "high" annoyance is double underlined.) (R=This scale is in reversed order, high annoyance came first in the questionnaire.)	Definition of high annoyance:	
						Basis: % or word	% of scale
Second Heathrow Airport (MIL Research, 1971)	1978 (p. 393)	UKD-024	Index-6	One overall question and five activity interference questions about aircraft noise (wake up, radio/tv, vibrate, conversation, other)	(This is the same as the multi-item index used in UKD-008 and is described in Appendix A.)	%	29%
Munich Aircraft (Rohrmann et al., 1973; Deutsche Forschungsgemeinschaft, 1974)	1978 (p. 396)	GER-034	Index-11	[GERMAN] Thirty-seven questions are combined into 11 variables that are used to calculate a factor score.	(This multi-item index is described in Appendix A.)	%	Not noted
Multiple-Item Index of 2 Questions							
Danish Railroad (Andersen et al., 1983)	1991 (p. 228)	DEN-200	Index-2 (Volunteer or Verbal-4)	[DANISH] Q.7 Is there something which especially makes you dissatisfied by living here in this part of the town (quarter)? [IF "train noise" is volunteered then count as highly annoyed] OR [If respond that it is the train passbys which are annoying or that trains are at least a little noisy, respondents were asked:] Q.25 How very much do you feel annoyed from the train noise?[H]	[mention train noise] [OR AFTER SCREENING QUESTIONS] Meget, Noget, Lidt, Næsten ikke, Slet ikke (Not at all, Only slightly, A little, Somewhat, very much)	Meget (very much) or mention train noise	25% or volunteer
Ranking Scale							
Paris Street (Aubree et al., 1971)	1978 (p. 395)	FRA-041	Ranking-10	(Q.4 Asks about "satisfied living in this part of town") Q. 6 Would you please rank the following items going from the most satisfactory to the least satisfactory? the amusements, ...the noise from the street.. [10 items]	RANKING QUESTION (Respondents do not rate the noise but only place it in order relative to other local nuisances.)	(RANK)	rank 1st

Table 3: Factors to consider in comparing data sets

	Factor	Example of range of observed conditions (Study identification number)	Extent normalize
Noise index calculation procedures			
1	Frequency weighting	●dB(A) (USA-203) ●PNdB (UKD-008)	All [dB(A)]
2	Method for combining different numbers, levels and durations of noise events	●Peak levels + log15(number) (UKD-042) ●L_{eq} (SWE-165)	All [DNL]
3	Cut-off for inclusion of low level noise events	●80 PndB (UKD-008) ●None (UKD-116)	
4	Position of nominal observer (Relevant for surface transportation)	●Sidewalk (BEL-122) ●At house on noisiest side (UKD-116)	
5	Position of nominal observer relative to possible reflective surfaces of dwelling	●One-meter from facade (UKD-116) ●Away from reflecting surfaces (UKD-024)	Some
6	Extent to which non-source noises are excluded	●Attended recordings with non-traffic noise removed during analysis (CAN-168) ●Unattended recordings (USA-102)	
Noise source operating conditions			
7	Day of week	●All days (UKD-116) ●Weekday (UKD-071)	
8	Type of operation (flight path and landing/take-off for aircraft)	●Loudest mode (UKD-008) ●Average (AUL-210)	
9	Inclusion of "abnormal" transportation operations (e.g. aircraft ground operations, railway maintenance, stationery road delivery vehicles)	●Excluded (UKD-024) ●Include reverse thrust (AUL-210)	
Other noise measurement issues			
10	Random error variance in estimating the long-term noise environment	Unknown	
11	Method of interpolation to dwellings without direct measurements	●Topographic model (UKD-116) ●None (USA-102)	
Human response measurement factors			
12	Type of annoyance scale	●Verbal (FRA-016) ●Numeric (UKD-071) ●Multi-item index (UKD-008) ●Ranking (FRA-041)	All (*A priori* not empirical basis)
13	Length of period to integrate annoyance	●Indefinite (UKD-116) ●Two months (USA-301)	
14	Other aspects of question-stem wording	●"In your home" (SWE-165) ●"In your neighborhood" (USA-082)	

	Factor	Example of range of observed conditions (Study identification number)	Extent normalize
15	Structure and wording of alternatives	●Very (USA-204) ●Tout à fait intolérable (Completely intolerable)(FRA-063)	
16	Language of administration	●Danish ●English ●Flemish ●French ●German ●Italian ●Norwegian ●Spanish ●Swedish	
17	Location of questions in questionnaire	●First question (USA-082) ●63rd question (AUL-210)	
18	Presented purpose of questionnaire	●Aircraft noise (USA-082) ●Environmental survey (UKD-116)	
19	Mode of administration	●Personal (UKD-116) ●Self-completion during interviewer visit (BEL-122)	
Respondent selection procedures			
20	Sampling of areas	●National probability sample (UKD-116) ●Purposively selected area (UKD-071)	
21	Sampling of individuals within areas & households	●Probability sample (UKD-116) ●First contact (USA-203)	
22	Call-back policies for households	●Five call backs (UKD-116) ●Request volunteers by mail (BEL-122)	
Annoyance moderating conditions			
23	Season of survey	●Summer (UKD-024) ●Winter (FRA-016)	
24	Year of survey	●1961 (UKD-008) ●1982 (USA-301)	
25	Noise source	●Aircraft ●Neighborhood ●Road traffic ●Railway ●Tramway	
26	Ambient noise levels and competing noise sources	●Multiple sources (CAN-168) ●Single source (UKD-071)	
27	Timing of noise events	●Much nighttime noise in some sites (UKD-116) ●Assume almost no nighttime (UKD-008)	All [DNL]
28	History of noise exposure and experience with noise source	●Long term experience (UKD-116) ●Recent change (USA-203)	
29	Acoustical characteristics of dwellings	Unknown	

Table 4: Criteria for synthesis data sets in 1991

Stated selection criterion	Subjective aspects of definition	Synthesis studies that do not meet criterion or that redefine the criterion
#1 "at least one questionnaire item had to inquire directly about long-term annoyance *per se*, rather than activity interference or other noise effects from which annoyance might arguably be inferred"	"long-term" "annoyance might arguably be inferred"	●1 study asks about "past few weeks" (USA-338) ●3 studies use multi-item indices dominated by activity questions (UKD-008, UKD-024, GER-034)
#2 "The noise source under study had to be a transportation noise source, and actual acoustic measurements of noise exposure were strongly preferred"	"strongly preferred"	●3 studies ask about neighborhood, not transportation, noise (CAN-121, DEN-075, USA-102) ●1 study did not measure exposure (CAN-168) ●6 or more data sets have no noise data on nighttime exposure (FRA-063, GER-034, SWI-053/air, SWI-053/road, UKD-008, UKD-024)
#3 "acoustic measurements if not reported in units of Day-Night Average Sound Level (DNL), had to be convertible into such units with reasonable confidence"	"reasonable confidence"	● 13 data sets that do not isolate nighttime noise do not allow nighttime noise to be separately weighted for DNL (DEN-200, FRA-016, FRA-041, FRA-063, GER-034, SWE-035, SWE-165/Road, SWE-165/Tram, SWE-228, SWI-053/road, SWI-053/air, UKD-008, UKD-024)
#4 "sample sizes had to be adequate for estimating prevalence of annoyance with reasonable precision"	"reasonable precision"	● 1 data set includes 3 data points based on only single interviews (DEN-200) ● 7 additional data sets have at least one data point based on fewer than 20 individual responses (BEL-122, BEL-137, CAN-168/Air, CAN-168/Road, FRA-063, SWI-053/Air, UKD-008)

Stated selection criterion	Subjective aspects of definition	Synthesis studies that do not meet criterion or that redefine the criterion
#5 "the scale used for quantification of annoyance had to permit identification of numbers of respondents describing themselves as "highly annoyed" in a manner comparable to that devised by Schultz". [The method devised by Schultz can be found in the 1978 synthesis publication (Schultz, 1978: 381)] "...We count as 'highly annoyed' those people who claim to be highly annoyed, when presented with annoyance scales whose steps are named, and those people who respond on the upper 27-29% of the annoyance scales if the steps (except for the extremes) are un-named. The Paris street survey is counted as before."	"claim to be highly annoyed"	●3 studies have multi-item indices that were not presented as numeric scales to respondents (UKD-008, UKD-024, GER-034) ●15 of the 16 verbal scales do not use the word "highly" (All but AUL-210) ●1 study uses volunteering that train noise creates any degree of dissatisfaction with living in an area (DEN-200) OTHER RELEVANT, BUT NOT CONTRADICTORY, LIMITATIONS ON INTERPRETING "HIGHLY ANNOYED" ●1 study equates high annoyance to choosing traffic noise as the least satisfactory of ten neighborhood attributes, regardless of degree of annoyance (FRA-041) ●3 data sets differ in their classification of "considerably". 2 equate it to high annoyance (CAN-168/road, CAN-168/air). 1 equates it to not high annoyance (AUL-210) ●8 data sets' verbal scales include 40% (2/5) of the points as highly annoyed.
#6 "...excluded from present consideration were laboratory studies of noise induced annoyance..."	✓	✓
#7 "...excluded from present consideration were ....field studies of community reactions to impulsive noise sources (i.e. gunfire, blasting, helicopters, sonic booms, etc.)..."	✓	✓

Table 5: Distribution of sample on two scales

Answers to 4-point verbal scale			Scores on Multi-item index		
Verbal label for scale point	Percentage	Cumulative %	Numeric score for scale point	Percentage	Cumulative %
Very much	19%	19%	6	2%	2%
			5	9%	11%
			4	9%	20%
Moderately	27%	46%	3	13%	33%
			2	16%	49%
A little	25%	66%	1	20%	69%
Not at all	34%	100%	0	31%	100%
TOTAL	100%		TOTAL	100%	

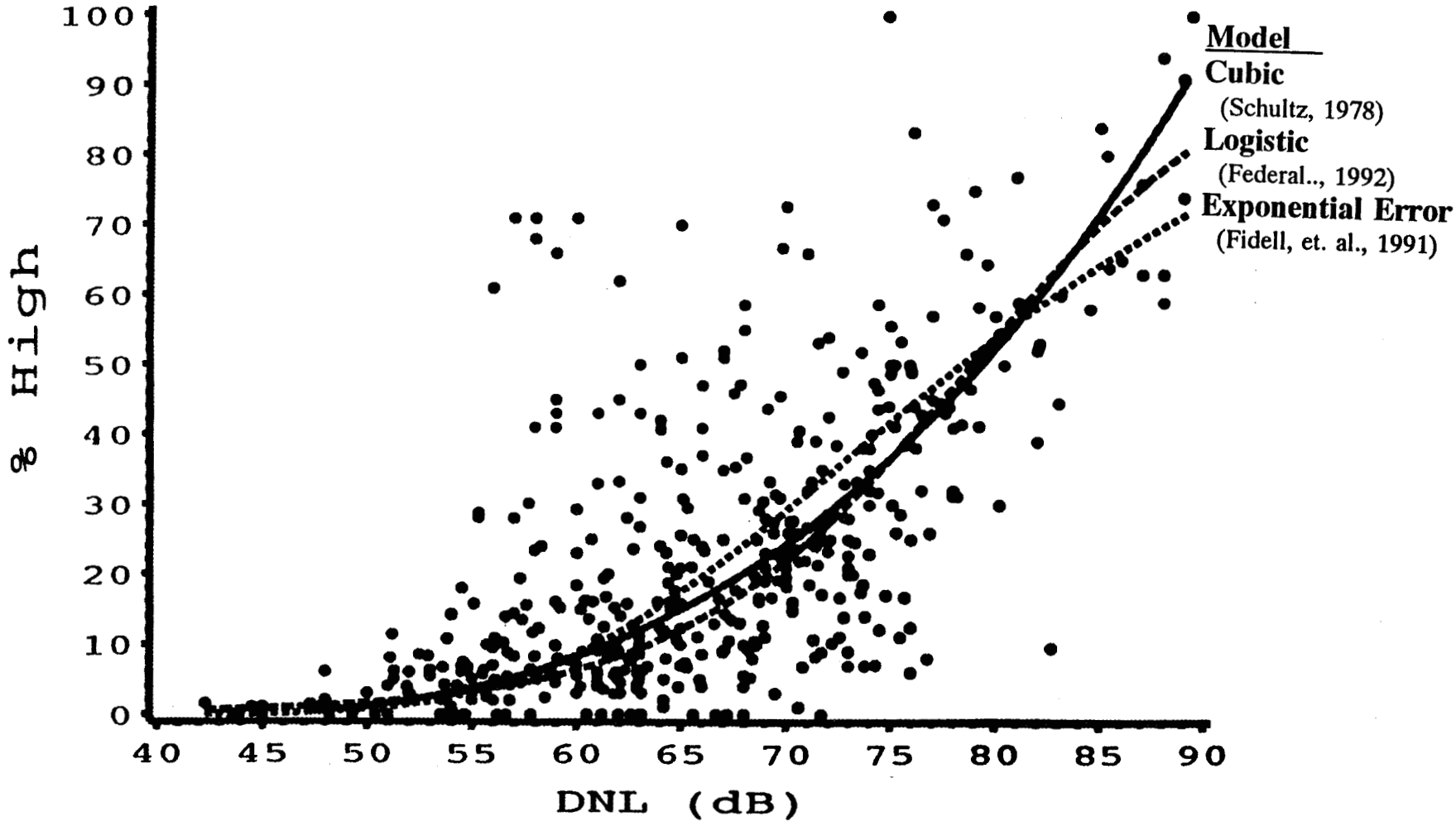
Source: First Heathrow Survey (McKennell, 1963: Summary, p.2)

Table 6: Logistic regression coefficients for the regression of noise annoyance ("high" annoyance score) on noise level under varying conditions:

Identification	Number of data sets included	Unit of analysis (n=number of respondents in group)			Logistic regression coefficients	
		Unit	Weight for each respondent	Score for 0/100% groups	Intercept	Slope
(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Standard baseline	23	Individual	1	0/1	-7.7087	0.0966
B. Grouped approximation	23	Group	1/n	(.01/99.99)	-11.1884	0.1408
C. Group- equal weights	23	Group	1	(.01/99.99)	-10.0707	0.1278
D. 0=(.0001/99.9999)	23	Group	1/n	(.0001/99.9999)	-15.6512	0.2014
E. Individual (no 0/100% groups	23	Individual	n	(exclude)	-7.4976	0.0938
F. Group (no 0/100% groups)	23	Group	n	(exclude)	-9.1196	0.1149
G. Individual (no 0/100% groups, 1/n weight)	23	Individual	1/n	(exclude)	-8.0143	0.0993
H. Group (no 0/100%, 1/n weight)	23	Group	1/n	(exclude)	-8.8742	0.1098
I. Group 29	29	Group	1/n	(.01/99.99)	-10.3959	0.1304

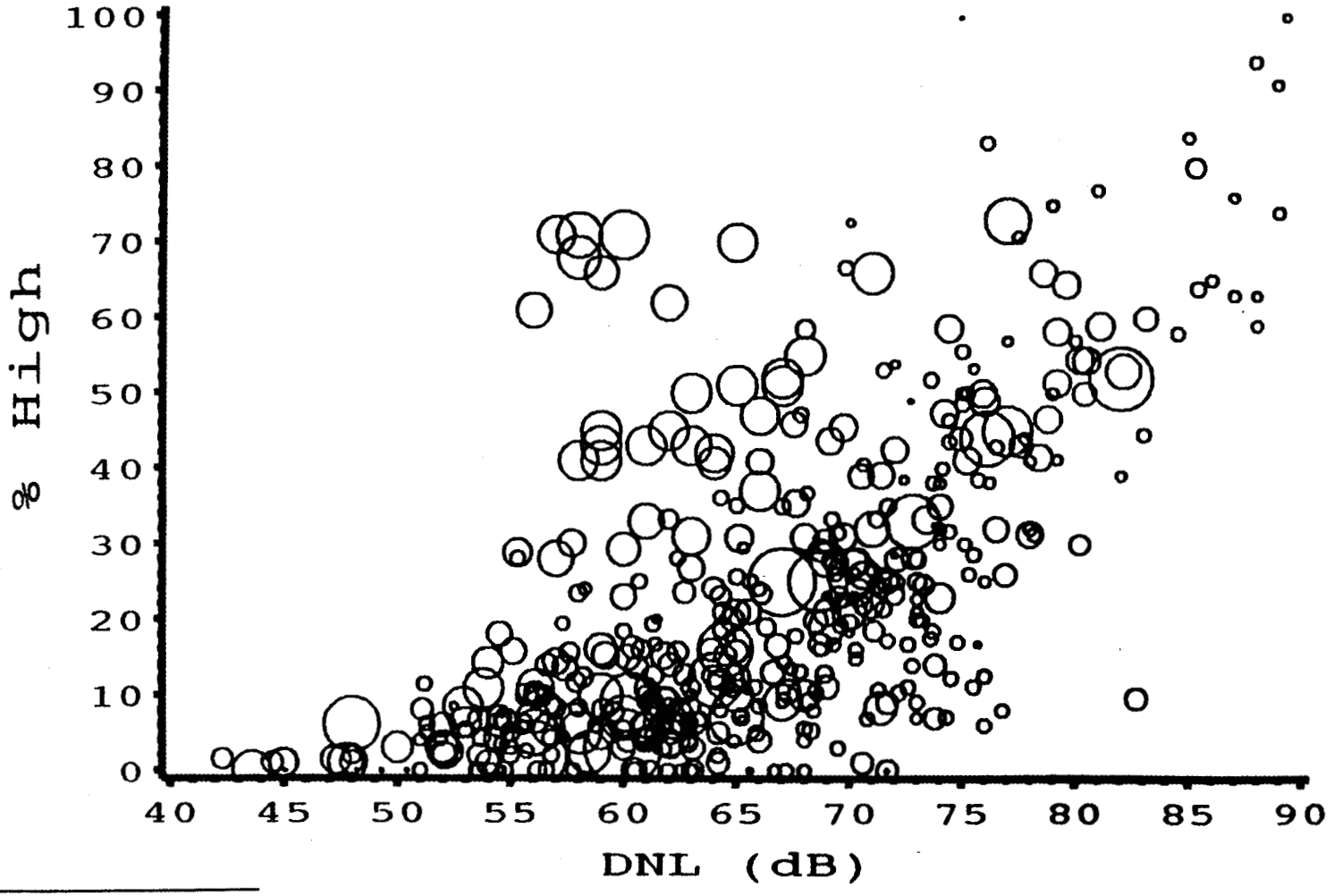
FIGURES

Figure 1: Three alternative curves for summarizing the noise/annoyance data points



(Source: Values for the plotted data points come from Appendix A of Fidell, Barber, and Schultz (1989).) Since the lower limit for DNL is 40 Ldn in these figures, the two data points from the single survey (UKD-116) with data below 40 Ldn (29.2, 37.3 Ldn) are not plotted. None of the respondents at those points were classified as highly annoyed.

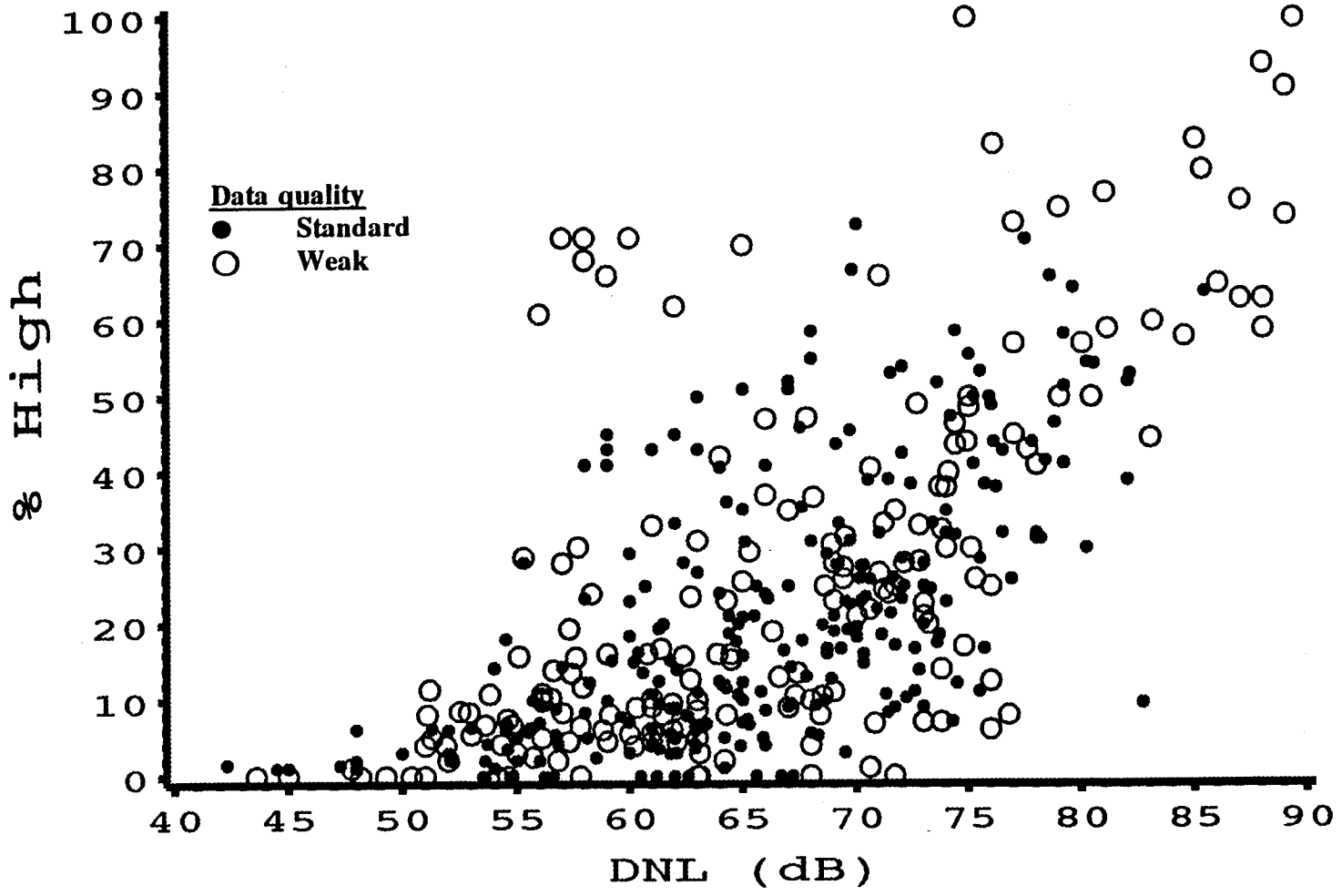
Figure 2: Variations in numbers of interviews for data points in the 1991 synthesis



(Source: See Figure 1)

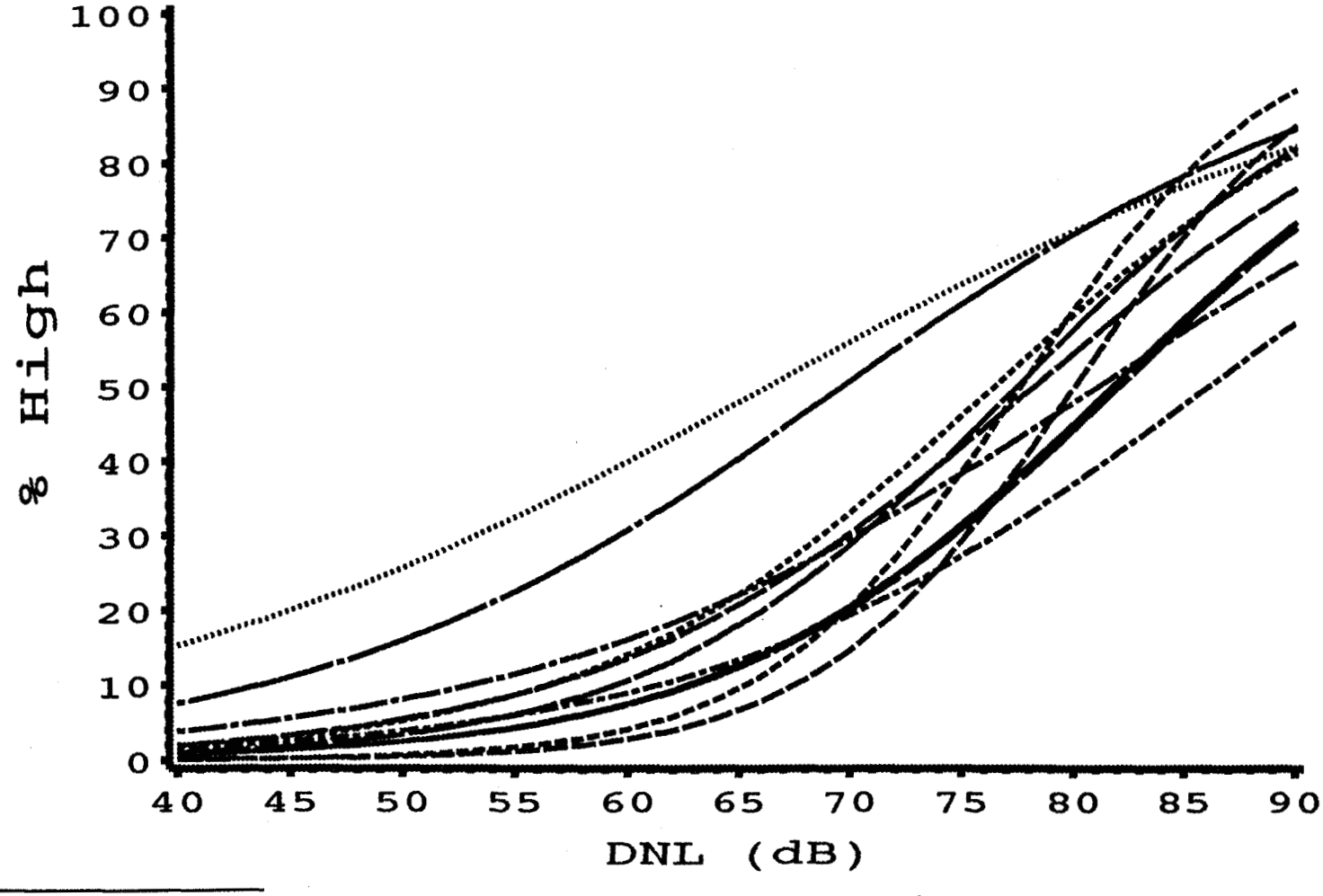
The area of each circle is proportional to the number of interviews at the data point. If the number of interviews for a particular data point is unknown, then the survey's interviews are divided evenly between the data points. Readers should be aware that visual impressions tend to underestimate differences between circles' areas.

* **Figure 3:** Annoyance by DNL for two levels of data quality



(Source: See Figure 1)
The open circles identify the 185 points from the eleven studies with the weakest data.

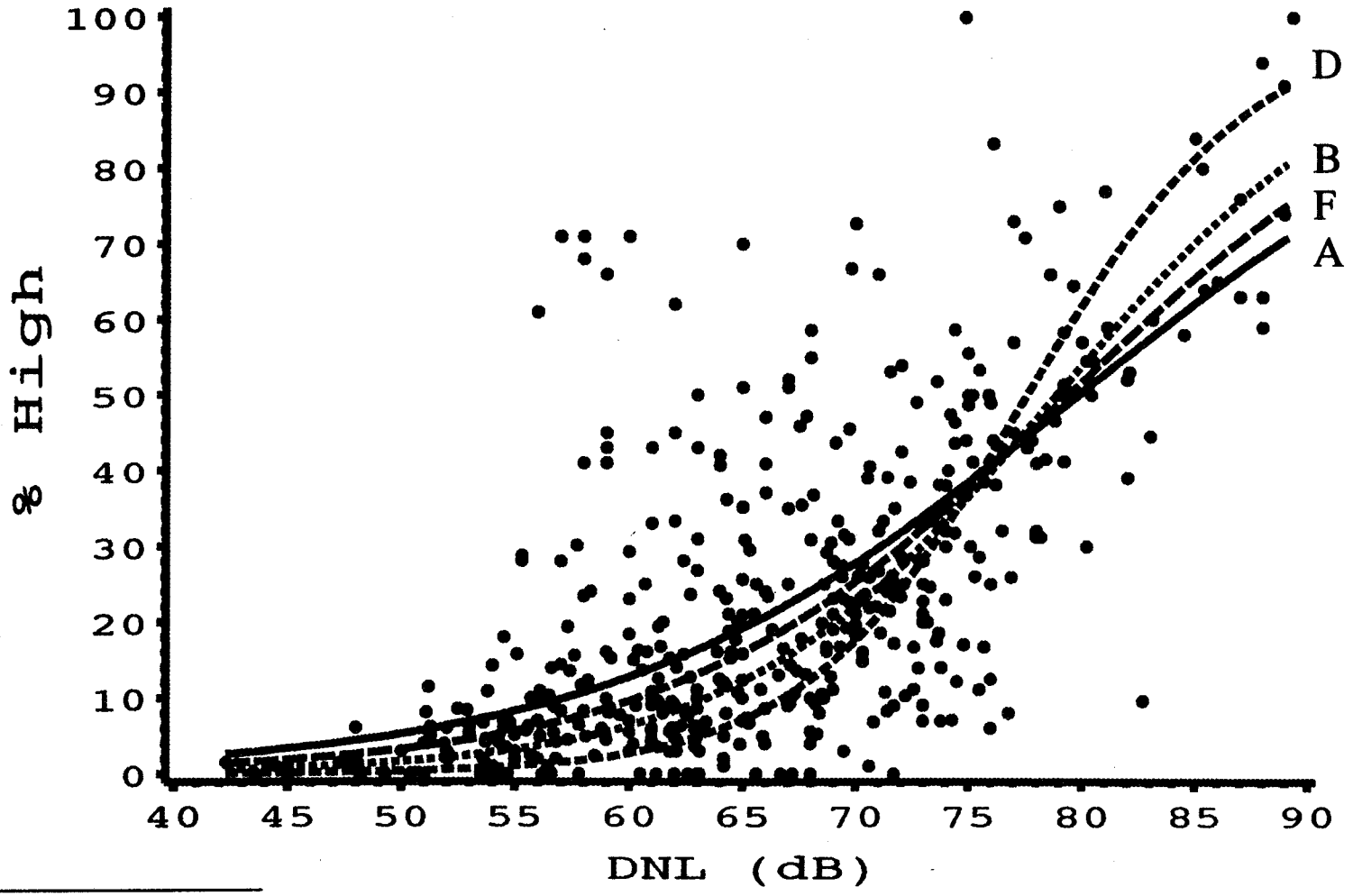
* **Figure 4:** Annoyance by DNL for 11 annoyance scales



* (Source: See Figure 1)

The logistic regression curves for eleven annoyance scales are plotted for 18 data sets. The logistic regression curves are based on analyses in which each individual interview is represented as a separate data point. The eleven weak data sets that were identified in the previous figure are not included. The 11 scales are defined by the type of scale and number of scale points. Each scale represents the data from one to three surveys. Observations are available for only a restricted range of noise levels for each data set. The curves in this figure thus represent extrapolations beyond the observed data.

* **Figure 5:** The effect of assumptions on estimates of curves from logistic regressions



* (Source: See Figure 1)

The logistic regression curves in the figure match four analyses which are identified by letter in Column 1 of Table 6 in Appendix C. The standard, individual respondent analysis (Line A) is represented by the solid line and has the shallowest slope. All of the other approximations have steeper slopes. Line B represents the relationship which was previously published in a report (Federal...., 1992: 3.6).

APPENDIX A: FULL WORDING FOR SELECTED QUESTIONS

This appendix includes the wording for the annoyance questions that could not be presented with in the space available in Table 2. This includes questions in the original language version of all non-English questionnaires. With the exception of screening questions for hearing a noise source, all screening questions are presented.

BEL-122, BEL-137: ANTWERP AND BRUSSELS STREET SURVEYS

5 We would like to know your opinion of the traffic noise that you hear in your residence during the day. Not disturbing at all.....Very disturbing [Respondents placed a mark on a 155 millimeter line to indicate their degree of annoyance.]

Original Flemish Version (Both studies) (Myncke, et al., 1977: Vol. 5-page 8, Vol. 6-page 6, Vol 11-page 4)
5. Wij hadden graag uw opinie gekend over het verkeerslawaai dat U hoort wanneer U overdag in uw woning is ["bent" replaced "is" in Brussels]? Helemaal niet storend _____ Zeer storend"

BEL-137: BRUSSELS STREET SURVEY (French version)

5 We would like to know your opinion of the traffic noise that you hear in your residence during the day. Not disturbing at all.....Very disturbing [Respondents placed a mark on a 155 millimeter line to indicate their degree of annoyance.]

Original French Version (for those interviews in French) (Myncke, et al., 1977: Vol. 11, page 12)
5. Nous voudrions bien connaître votre opinion sur le bruit du trafic que vous entendez chez vous pendant la journée. Ne pas génant du tout _____ Très gênant

CAN-121: CANADIAN ROAD

[The question had an introduction in 1976 but not in 1975.]
[1975 Version]
6. How would [sic] rate the overall noise in this neighbourhood?

[1976 Version]
4. Considering all you have mentioned, how would you rate the overall noise?

DEN-075: DANISH STREET SURVEY

[This question is preceded by four questions of the form "Is there much, a little or no noise from... neighbors (Q.25), ...traffic (Q.26), ...aircraft (Q.27), ...other types of noise (Q.28).]
29. Do you feel much, a little or not at all annoyed by the noise you are exposed to?
Yes, much annoyed
Yes, a little annoyed
No, not annoyed
Don't know, not at home

Original Danish Version (Kragh, 1977: 49)
29. Føler De Dem meget, lidt eller ikke spor generet af den støj som De unsættes for?
Ja, meget generet
Ja, lidt generet
Nej, ikke generet
Ved, ikke/uoplyst

[Note: Although both DEN-075 and DEN-200 use "meget", one source (DEN-075) translates it as "much" and the other (DEN-200) as "very much".]

DEN-200: DANISH RAILROAD SURVEY

Respondents are counted as highly annoyed if they either:

A. volunteer train noise as something that makes them especially dissatisfied (Question 7) after volunteering that something makes them especially dissatisfied (Q.5) and that it is noise that is dissatisfying (Q.6)

OR

B. say they that they are very much annoyed (Q.25) after answering the preceding filter questions either:

1. annoyed by train noise (Q.25) and then that it is either the passbys, acceleration or braking that is annoying (Q.23)

OR

2. annoyed by train noise (Q.25) and then that it is only other aspects of the train noise which are annoying (i.e. neither the passbys, acceleration or braking that is annoying (Q.23)) and then that the trains are at least a little noisy (Q.24)

OR

3. not annoyed (Q.21), but the trains are at least a little noisy (Q.24)

[Of the 146 who were counted as meeting this "high annoyance" criteria, a total of 120 respondents volunteered they were dissatisfied on the open question (Q.7) (Reaktioner på togstøj, 1982: 44). Thus though only 26 were classified as "high" on only the basis of the closed question (Q.25), it is not known how many of the 120 volunteering dissatisfaction (Q.7) also answered they were very much annoyed on the closed question (Q.25)]

5. Is there something, that especially makes you dissatisfied with living in this part of town (quarter)?
   yes
   no skip to Q.8
6. What is it, that especially makes you dissatisfied?
   noise
   other things .. go to Q.8
7. Which type of noise are you thinking of?
   [Multiple answers are allowed]
   car noise
   neighbors' noise
   noise from children playing
   aircraft noise
   train noise
   factory noise
   moped noise
   other things, describe _____

. . . . . .

21 Do you feel annoyed by railway noise?
   yes
   no .. go to question 24

22 Do you feel annoyed both inside and outside your house?
   yes
   no, only indoors
   no, only outdoors

23 What is it that is annoying?
[Multiple answers allowed]
Noise from driving ... go to Q.25
Noise from acceleration ... go to Q.25
Noise from braking ... go to Q.25
Noise from maintenance in the night
Noise from maintenance in the day
Dust (metal-dust)
Vibration (shaking)
Smell
Interference with TV-picture
24 Are the trains noisy?
yes, very much
yes, little
no ... go to question 34
25 How much do you feel annoyed from the train noise?
[Show card B]
very much
somewhat
a little
only slightly
not at all

Original Danish Version (Reaktioner.., 1982: 118, 122, 123)
5.ER DER NOGET, SOM DE ISÆR ER UTILFREDS MED VED AT BO HER I KVARTERET?
ja
nej gå til spm. 8
6. HVAD ER DET ISÆR DE ER UTILFREDS MED?
støj
andet... gå til spm. 8
7. HVILKEN FORM FOR STØJ TÆNKER DE PÅ?
[Gerne flere svar]
bilstøj
nabostøj
støj fra legende børn
flystøj
togstøj
fabriksstøj
knallertstøj
anden støj, beskriv
. . . . . .
21 FØLER DE DEM GENERET AF JERNBANETRAFIKKEN?
ja
nej gå til spm. 24
22 FØLER DE DEM GENERET BÅDE INDENDØRS OG UDENDØRS?
ja
nej, kun indendørs
nej, kun udendørs

23 HAVD ER DET, DER GENERER?
[Gerne flere svar]
STØJ FRA KØRSEL gå til spm. 25
STØJ FRA ACCELERATION gå til spm. 25
STØJ FRA BREMSNING gå til spm. 25
STØJ FRA VEDLIGEHOLDELSESARBEJDE OM NATTEN
STØJ FRA VEDLIGEHOLDELSESARBEJDE OM DAGEN
STØV (METALSTØV)
VIBRATIONER (RYSTELSER)
LUGT
FORSTYRRELSE AF TV-BILLEDE

24 STØJER TOGENE?
ja, meget
ja, lidt
nej gå til spm. 34

25 HVOR MEGET FØLER DE DEM GENERET AF TOGSTØJEN?
[Vis kort B]
meget
noget
lidt
næsten ikke
slet ikke gå til spm. 34

FRA-016: FRENCH AIRCRAFT SURVEY

17. Does airplane noise bother you not at all, somewhat, quite a bit, or a great deal?

Original French Version (Centre.. 1968: 113)
17. Est-ce que le bruit des avions vous gêné pas du tout, un peu, assez, beaucoup?

FRA-041: PARIS STREET SURVEY

6. Would you please rank the following items going from the most satisfactory to the least satisfactory? a-the amusements, b-the proximity to place of work, c-the public transport, d-the noise from the street, e-the noise in the building, f-the schools, g-the neighbors, h-the stores, i-the public services (local government, Post office), j-the doctors and pharmacists.

Original French Version (Aubree, Auzou, Rapin, 1971: Vol. 5, page 483)
6. Voulez-vous ranger la liste que voici en allant de ce qui vous satisfait le plus à ce qui vous satisfait le moins? (présenter les fiches bristol) a-les distractions la proximité du lieu de travail, b-les transports publics, c-le bruit de la rue, d-le bruit de l'immeuble, e-les écoles et les lycées, f-les voisins, g-les magasins, h-les services publics (mairies, i-postes), j-les médecins et pharmaciensa"

FRA-063: FRENCH RAILROAD SURVEY

57. From a general point of view, the train noise, in your opinion, is Quite acceptable 1 2 3 4 5 6 7 Completely intolerable

Original French Version (Aubree, 1975: 91)
57. D'un point de vue général le bruit des trains est à votre avis Très supportable 1 2 3 4 5 6 7 Tout à fait intolérable

GER-034: MUNICH AIRCRAFT SURVEY

The annoyance measure is not a five-point scale but rather a division of an index (labeled R1U in the text) that is the factor score for the first unrotated factor for 11 reaction variables. (Deutsche Forschungsgemeinschaft, 1974:Table 4.7). The scale was divided at the median score, i.e. the score at that divided the sample respondents in to equal groups (Deutsche Forschungsgemeinschaft, 1974: 139). The scale value for this division point is not reported. The eleven variables and the subjects of the 37 constituent questions that are used to compute the factor score are the following: (1) Q#44-amount of day during which aircraft noise is heard; (2) Q#46-loudness of aircraft noise relative to other noises; (3) Q#66-whether aircraft noise is or is not bearable; (4) Q#18.4,.6,.7,.8 Extent respondent is content with four conditions (area of residence, area conditions affecting health, quietness in area, possibilities for recreation); (5) Q#48.x1,x2 vibration of walls, rattle of objects; (6) Q#48.y1,y8 headaches, earaches due to aircraft noise; (7) Q#48.x4, x5, y1,y6 - aircraft interferes with radio sound, TV picture or makes raise voice or increase radio or TV volume; (8) Q#48.y2, y3, y5, y7, y8 - aircraft interfere in respect to relaxing, startling, going to sleep, reading, waking up at night; (9) Q#53.1, .2, .3, .4 install double glazing, sound proofing, ventilation fans or use earplugs (10) Q#53.6, .7, .8, .10, .11 complain by letter, telephone, personal visit, attend gathering, join group (11) Q#62.1, .2, .3., .4, .6., .7, .9, .10, .11 extent to which agree with statements about aircraft noise relating to anger, thinking about complaining, accommodating, nervousness, being bothering, ruining evening, wanting to move, making people more upset than is necessary. The specific questions that created the 11 variables were identified in a personal letter by a participant in the study (Dr. Rohrmann).

SWE-035: SWEDISH AIRCRAFT SURVEY

No copies of the questionnaires in any of the three languages used in SWE-035 exist. One of the authors (Dr. Rylander) reported in a personal communication that the wording of these questionnaires was identical to that used in other Swedish surveys such as SWE-165 and SWE-228.

SWE-165: SWEDISH TRAMWAY/TRAFFIC SURVEY

In the following questions I will ask you about different sources of nuisances in your living environment.
I will start with giving you a number of annoyance sources that might be present in a "living area". For each of them state if you notice them in your home.

	Do not notice	Do notice	Are you annoyed (a) or disturbed (b) Yes No	How much are you annoyed? (Code 2)	How often are you annoyed? (Code 1)
13. Noise from neighbors			(a)		
14. Aircraft noise			(a)		
15. Industrial noise			(a)		
16. Exhaust from vehicles			(b)		
17. Dust or soot from industries			(b)		
18. Odors from industries			(b)		
Are there any other sources of nuisance in this living area? If so: what?					
19.					
20.					
21.					

[NOTE: The alternatives for Code 2 are: 1. Very annoyed 2. Rather annoyed 3. Not very annoyed]
If IP[respondent] names traffic noise or/and tram noise go directly to question 22.
[NOTE: The author believes the question number is a misprint and should have been question 23.]
22. In your home do you notice the noise from

S) Trams	T) Other traffic	
1 No	1 No	
2 Yes	2 Yes	→What types of traffic do you mostly notice? .....

[Note: A box of skip instructions is given here]
23. In your home are you annoyed by the noise from

S) Trams	T) Other traffic
1 No	1 No
2 Yes	2 Yes

24. How annoyed are you by (Code 2)

S) Trams	T) Other traffic
1 2 3	1 2 3

Original Swedish Version

I det följand avsnittet vill jagzz fråga om olägenhetskällor i Er bostadsmiljö.

Jag börjar med att räkna upp ett antal olägenhetskällor som kan förekomma i ett bostadsområde. För var och en av dem skall Ni ange om Ni lägger märke till dem i Er bostad.

	Lägger ej märke till	Lägger märke till	Störs (a) eller besväras (b) Ni? Ja Nej	Hur mycket störs Ni? (Kod 2)	Hur ofta störs Ni? (Kod 1)
13. Buller från grannar			(a)		
14. Flygbuller			(a)		
15. Industribuller			(a)		
16. Avgaser från motorfordon			(b)		
17. Damm eller sot från industrier			(b)		
18. Lukt från någon industri			(b)		
Finns det ytterligare några olägenhetskällor i detta bostadsområde? I så fall vilken/vilka?					
19.					
20.					
21.					

[NOTE: The alternatives for Code 2 are: 1.Mycket 2. Ganska mycket 3. Inte särskilt mycket]

Om IP nämner trafikbuller och/eller spårvagnsbuller gå direkt till fråga 22.

22. Lägger Ni i Er bostad märke till buller från

S) Spårvagnar T) Övrig trafik
0 Nej 0 Nej
1 Ja 1 Ja→Vilket eller vilka fordon lägger Ni märke till mest?
.....

23. Störs Ni i Er bostad av..?

S) Spårvagnsbuller T) Övrigt trafikbuller
0 Nej 0 Nej
1 Ja 1 Ja

24. Hur mycket störs Ni av (Kod 1)?

S) Spårvagnsbuller T) Övrigt trafikbuller
1 2 3 1 2 3

SWE-228: SWEDISH RAILROAD SURVEY

I will start with giving you a number of annoyance sources that might be present in a "living area". For each of them state if you notice them in your home.

	Do not notice	Do notice	Are you annoyed (a) or disturbed (b) Yes No	How much are you annoyed? (Code 2)	How often are you annoyed? (Code 1)
14. Noise from neighbors			(a)		
15. Traffic noise			(a)		
16. Aircraft noise			(a)		
17. Industrial noise			(a)		
18. Exhaust from vehicles			(b)		
19. Dust or soot from industries			(b)		
20. Odors from industries			(b)		
Are there any other sources of nuisance in this living area? If so: what?					
21.					
22.					
23.					
If IP[respondent] names train noise go directly to question					
24. Railway noise			(a)		

[NOTE: The alternatives for Code 2 are: 1. Very annoyed 2. Rather annoyed 3. Not very annoyed]

Original Swedish Version

Jag börjar med att räkna upp ett antal olägenhetskällor som kan förekomma i ett bostadsområde. För var och en av dem skall Ni ange om Ni lägger märke till dem i Er bostad.

	Lägger ej märke till	Lägger märke till	Störs (a) eller besväras (b) Ni? Ja Nej	Hur mycket störs Ni? (Kod 2)	Hur ofta störs Ni? (Kod 1)
14. Buller från grannar			(a)		
15. Trafikbuller			(a)		
16. Flygbuller			(a)		
17. Industribuller			(a)		
18. Avgaser från motorfordon			(b)		
19. Damm eller sot från industrier			(b)		
20. Lukt från någon industri			(b)		
Finns det ytterligare några olägenhetskällor i detta bostadsområde? I så fall vilken/vilka?					
21.					
22.					
23.					
Om IP nämner tågbuller gå direkt till fråga					
24. Tågbuller			(a)		

[NOTE: The alternatives on Code 2 are: 1.Mycket 2. Ganska mycket 3. Inte särskilt mycket]

SWI-053: SWISS AIRCRAFT/ROAD SURVEY

Note: The questionnaire was presented in three languages: German, French and Italian. Only the German version is reproduced in the available reports. The French labels for the annoyance scale come from a short French version of the report (Graf, Meier, Müller, 1974b: 13). No information could be located about the wording of the Italian version.

45 Let's assume that this would be a thermometer with which it is possible to measure how much street traffic noise disturbs you at home. The number 10 means that you find street traffic noise unbearable, the number 0 that it doesn't disturb you at all. Please tell me the number that applies to you.

Not at all disturbing 0 1 2 3 4 5 6 7 8 9 10 Unbearably disturbing

45 D. Now, lets apply the thermometer to airplane noise.

Original German Version (Graf, Meier, Müller, 1974a:49, 52)

45. NEHMEN WIR AN, DIES WAERE EIN THERMOMETER, MIT DEM MAN MESSEN KANN, WIE STARK SIE DAHEIM DURCH DEN STRASSENVERKEHRSLAERM GESTOERT WERDEN. DIE ZAHL 10 BEDEUTET, DASS DER STRASSENVERKEHRSLAERM SIE UNERTRAEGLICH STOERT, DIE ZAHL 0, DASS ER SIE UEBERHAUPT KEIN BISSCHEN STOERT. SAGEN SIE MIR EINFACH DIE ZAHL, DIE AUF SIE ZUTRIFFT.
Überhaupt keine Störung 0 1 2 3 4 5 6 7 8 9 10 Unerträgliche Störung

45 D. JETZT NEHMEN WIR DAS THERMOMETER NOCH FUER DEN FLUGLAERN

UKD-008: FIRST HEATHROW AIRPORT SURVEY

13B (1) Does the noise of the aircraft bother or annoy you: Very much, Moderately, A Little, or Not at all.

14B Do the aircraft ever.....[PROMPT ALTERNATIVES AT LEAST ONCE]
When they ... how annoyed does this make you feel?

i. Startle you
ii. Keep you from going to sleep
iii. Wake you up?
iv. Interfere with listening to TV or radio?
v. Make the TV picture flicker?
vi. Make the house vibrate or shake?
vii. Interfere with conversation?
viii. Disturb your rest or relaxation?
ix. Interfere with or disturb any other activity?
x. Bother, annoyed or disturb you in any other way?

[The response categories for all items are: Very, Moderate, A little, Not at all.]

The 7-point index was obtained by giving one point for reactions on the following six items. (McKennell, 1963: Appendix J, Page 8)

Item	Categories scored 1 (All other responses scored zero)
13B(1) (Overall aircraft rating)	Very much, moderately, a little
14Biii. Wake up	Very, moderately
iv. TV or radio listening	Very, moderately, a little
iv. House vibrate	Very, moderately, a little
vii. Interfere conversation	Very, moderately
(Highest score of 2 questions) ix. Other activity x. Other disturbance	Very, moderately, a little

USA-082: LOS ANGELES AIRPORT SURVEY

Original English Version

1 Are you ever annoyed by aircraft noise in your neighborhood?
(ALLOW A FREE RESPONSE; DO NOT RECORD REPLY)
I see. Now I need an answer that I can compare with the answers that other people give me. Would you say you've been not at all annoyed, slightly annoyed, moderately annoyed, very annoyed or extremely annoyed?

Some of the respondents in Los Angeles were interviewed in Spanish.

Spanish Version (Copy of questionnaire supplied by author)

1 Se siente usted molestado por el ruido de aviones en su vecindario?
(ALLOW A FREE RESPONSE; DO NOT RECORD REPLY)
Ya veo, ahora necesito una respuesta que yo pueda comparar con las respuestas de otras personas. Dirá usted que ... Definitivamente No, Un Poco, Moderadamente, Bastante, Extremamente.

APPENDIX B: GUIDE TO INFORMATION ABOUT SURVEYS

This appendix begins by identifying several sources of information that can be consulted to obtain or confirm information about the 29 synthesis data sets. Most of the appendix consists of notes on corrections to the synthesis publication descriptions. The appendix concludes with a small amount of information about several of the "non-clustering" surveys that were described in the 1978 synthesis but not included in the 29 synthesis data sets.

A bibliography of published sources of information for 28 of the 29 data sets is included in a catalog of noise surveys (Fields, 1991). The single survey that was not included in the catalog has only a single publication (Borsky, 1985). When, as in most cases, the exact wording of the annoyance question is available in a publication, the publication is cited in Appendix A. Unless otherwise noted, the source of information in Appendix B is the publication that is referred to in the synthesis articles.

The synthesis articles provide an accessible source of information about many of the surveys. Another report presents the values of DNL and percentage annoyed that are plotted in the synthesis articles (Fidell, Barber, and Schultz, 1989: Appendix A). The present review has attempted to supplement and correct any information that was found to be incorrect in the two synthesis articles, but has not prepared a list of corrections to the data presented in the 1989 report.

A number of corrections have been routinely made in the presentation of the information and question wordings in Tables 1 and 2 and in Appendix A. The corrections to numbers of individual responses and the ranges in sizes of data points are annotated with a "(C)" in Table 1 so that the reader will know where differences from the synthesis articles should be expected. Most minor corrections, however, have not been highlighted.

This appendix provides supplementary information that expands upon some of the briefly noted points raised in the last column of Table 1. This appendix also presents any other information that corrects survey descriptions from the synthesis articles. Most of these corrections are not of general interest and not important for most purposes. Most are presented here solely to aid future researchers who may not be able to consult the original publications or who may be puzzled by differences between the original survey publications and the descriptions presented in the synthesis articles.

Most of the information for these corrections comes from publications. The wordings for a small number of questions were obtained from unpublished questionnaires that were directly obtained researchers. The few instances where the information came from direct communications with authors are noted.

AUL-210 AUSTRALIAN AIRCRAFT

The noise exposure estimates were developed from an extensive calibration procedure for individual aircraft types for each runway that was based on corrections to nominal aircraft noise contours and airport flight records (Hede and Bullen, 1982: 36-61). This is a more complex procedure than the two-week measurement program described in the synthesis article.

Hede and Bullen believe that this single-item annoyance scale is not a good measure of being "seriously affected" and that the 1991 synthesis article's definition of "'highly annoyed'...would seem to exclude some people whose degree of reaction would normally be termed 'high'" (Hede and Bullen, 1982: 92-94).

BEL-122 ANTWERP STREET SURVEY

Interviews were conducted from May to October of 1975 (not 1974) (Myncke, Cops, and Gambart, 1977: 169). The data in the synthesis do not match the cited source (Myncke, et al., 1977: Vol. V, Page, 103, Table 32, Q#5) nor any other report that has been located on this survey. The 1978 synthesis stated that 40 data points were analyzed, but there appear to be 35 in the 1978 figure (Schultz,

1978: Figure 41) and only 31 are included in the list of data points that were reported for the 1991 synthesis (Fidell, Barber, and Schultz, 1989: Table A-28). The total number of interviews in the survey (not all of which appear to have been included in the synthesis) was 1319 which, when combined with the Brussels survey (BEL-137), is less than the 2,062 interviews reported in the synthesis. Noise measurements were made at the curbside at a height of 1.2 meters (Myncke, Cops, Gambart, and Steenackers, 1979:12).

BEL-137 BRUSSELS STREET SURVEY

Interviews were conducted in May to October of 1976 (not 1974) (Myncke, Cops, and Gambart, 1977: 169) The total number of interviews in the survey (not all of which appear to have been included in the synthesis) was 494 which, when combined with the Antwerp survey (BEL-122) is less than the 2,062 interviews reported in the synthesis article. One of the sites in the original report (Site # 18) was not presented in the synthesis article. Measurements were made at the curbside at a height of 1.2 meters.

CAN-121 CANADIAN ROAD

The data for the synthesis come from a two-year study of several noise sources. The questionnaire was changed slightly between years, but in both years referred to the "overall noise" not to the noise from a specified noise source. The 1978 synthesis article is based on a personal letter that provided data from sites that it is presumed were expected to be most relevant for a study of traffic noise. The author of the letter (Dr. Hall) no longer has a record of the data. It has not proved to be possible to check the synthesis data against any of the published information in any of the 13 published reports on the survey. The synthesis article citation for this study has been corrected in the list of references (Hall and Taylor, 1977).

CAN-168 AIRCRAFT/TRAFFIC COMPARISON (AIRCRAFT ONLY)

The percent highly annoyed for 61-63 Ldn aircraft is 33.3 not 33.9. (Hall, Birnie, Taylor and Palmer, 1981: 1693).

CAN-168 AIRCRAFT/TRAFFIC COMPARISON (TRAFFIC ONLY)

A respondent is reported to be highly annoyed if the respondent describes either of three traffic noise sources as "considerably" or "extremely disturbing". This measure differs from other studies in that a respondent rates components separately and a respondent has three chances to respond with a high annoyance score. Three road traffic sites with 35 interviews are excluded from the road traffic but not the aircraft data set for the synthesis. This is the reason that the number of interviews is 638, not 673 (Hall, Birnie, Taylor, and Palmer, 1981: 1693).

DEN-075 DANISH STREET

The questionnaire does not contain a direct traffic noise annoyance question. After examining the document that is cited in the 1978 synthesis article and discussing the document with the author of the survey publications (Dr. Relster) it has been concluded that the synthesis is based on a question that does not specify the source of the noise.

DEN-200 DANISH RAILROAD

The data come from Figure 1 of a source not cited in the synthesis (Reaktioner.., 1982: 46), not from the cited source. Seven of the questionnaire items are relevant for determining whether respondents were classified as "highly annoyed" for the synthesis, not a single five-point scale as presented in the 1991 synthesis article. The basis for many of these classifications may have been responses to an open question about anything that made them dissatisfied with the part of town. More information is

provided in Appendix A. At least two points were incorrectly copied from the original data set. The point with 7.5 % annoyed is at 53.6 Ldn not 54.6. The point at 67.4 Ldn has approximately 25% annoyed not 13.5%.

FRA-016 FRENCH AIRCRAFT

The annoyance scores for five of the six points are off by from 2% to 4%.

FRA-041 PARIS STREET

This annoyance scale is completely unlike any other annoyance measure used in the synthesis. Respondents only place street noise in a ranking relative to other items in the neighborhood. Thus if a respondent can be classified as "highly annoyed" in the synthesis even if there is no item in the neighborhood, even road traffic noise, that is even moderately annoying.

FRA-063 FRENCH RAILROAD

The article suggests that there is nighttime railway noise, but the synthesis states there is none. The French publication reports that railway noise is mentioned by 31% as an explanation for the main reason for being awakened at night. (Aubree, 1975: 30, Q.51, Q.52). The original publication does not report whether or not nighttime noise was measured.

GER-034 MUNICH AIRCRAFT

In the 1978 synthesis the annoyance scale is described as a five-point verbal scale. The annoyance measure was actually a dichotomous variable created by dividing an index (labeled R1U in the original report) that is the factor score for the first unrotated factor for 11 reaction variables. (Forschungsgemeinschaft, 1974:Table 4.7). The annoyance measure is divided at the median score for the survey population and thus not necessarily at a measure indicating high annoyance. The reason for excluding three of the original publication's 30 data points (Forschungsgemeinschaft, 1974: Figure 3-19) in the synthesis is not discussed in the 1978 synthesis.

SWE-035 SWEDISH AIRCRAFT

The data for this survey came from Sweden, Norway and Denmark and thus was administered in three languages. Two of the authors' of the original report have attempted to locate a copy of the original questionnaire but have not been successful (Dr. Rylander and Dr. Sörensen). Dr. Rylander reported in a personal communication that the wording of these questionnaires was identical to that used for the four-point scales in later Swedish surveys such as SWE-165 and SWE-228.

SWE-165 TRAMWAY/TRAFFIC COMPARISON

No information is used on the relative amount of daytime and nighttime traffic. The nighttime traffic is predicted for tramway as well as road traffic on the basis of road traffic prediction equations used in the United States and Sweden. The noise levels used for traffic noise in the synthesis are the levels for all traffic (including tram noise) rather than for only the motor traffic (that includes cars and trucks but not trams).

SWI-053 SWISS ROAD AND AIRCRAFT

The synthesis publication incorrectly states that the end points of the scales are not labeled. In fact the labels are "stört kein bisschen" (not at all disturbed) and "stört unerträglich" (unbearably or intolerably disturbed). The terms "keine störung" and "stark störung" that are quoted in the 1978 synthesis

publication are the Swiss author's characterization of the sections of the scale and not the words that were used in the questionnaire.

SWI-053 SWISS ROAD (AIRCRAFT ONLY)

The annoyance scores are incorrect for these data points. The scores that are plotted are the cumulative percentages (Grandjean et al., 1974: 115) not the actual percentage annoyed at each noise level (Grandjean et al., 1974: 114). Neither of the synthesis articles includes the data point from the lowest noise level. The two data points for the highest noise levels are included in the 1991 but not the 1978 synthesis. The total number of interviews included in the 1978 synthesis (3644) and the 1991 synthesis (3717) are thus fewer than the actual total of 3940. These totals all exceed the total of 2995 reported in the 1978 synthesis that were based on the mistaken assumption that the aircraft noise survey had not been completed in Bern.

UKD-008 FIRST HEATHROW AIRPORT

The exact division of the questionnaire items within the annoyance index was correctly described earlier in Appendix A on the basis of the information presented in the primary report on the survey (McKennell, 1963: J-9) and the descriptions in the second Heathrow survey publication (MIL, 1971:159). The report that is cited in the synthesis (Noise: Final Report) was prepared under Wilson. The author is not McKennell. The data point with 11 interviews is not included in the synthesis (Wilson, 1963: 207). McKennell suggests that a score of 3.5 (not 5) "..appears to be a reasonable interpretation of serious annoyance. This score corresponds to the average reaction at the highest noise level, and is the point at which aircraft noise becomes singled out as the greatest inconvenience of living in the area, and at which 50 percent of the informants report disturbances to sleep and rest." (McKennell, 1963: 2). The data in Table 5 of the present report shows that the division point used in the synthesis article is more severe than that implied by a four-point verbal scale.

UKD-024 SECOND HEATHROW AIRPORT

Some 3104 of the 4678 respondents' answers to this survey are plotted twice because the data for ten noise zones is plotted from a Table P.2 that provides answers for the entire study area (MIL, 1971: 190) as well as from Table P.4 that provides answers for only the lesser "ten-mile" survey area (MIL, 1971: 192). This survey's annoyance index was also used in the previous Heathrow survey (UKD-008) and thus observations about the annoyance score are relevant.

USA-082 LOS ANGELES AIRPORT

The author of the original study (Dr. Fidell) reports that the noise lower of the two data points is at DNL 62 (within the DNL 60 to 65 DNL contour) rather than at DNL 67.

USA-102 USA 24 SITE

Neither the questionnaire nor the description of the study sample suggest that this is a study of road traffic noise. The questionnaire asks about "neighborhood" noise. The study publication explicitly states that the study chose sites to avoid "..significant airport and highway noise exposure..." and was "...not directed at any single noise source...". Instead the intent was to study "...the sorts of noise exposure experienced by virtually all urban and suburban residents--the daily din of local street traffic, neighborhood and household noises, building equipment, mechanical and industrial noise sources, and so forth." (Fidell, 1978: 198). It is assumed that both the personal and the telephone interviews were included in the results.

USA-203 BURBANK AIRPORT

Two aspects of this study are noted; the scoring of missing data and the changing nature of the noise environment. The annoyance percentages appear to have been calculated by assuming that those with missing data are not annoyed (Fidell, Horonjeff, Teffeteller and Pearsons, 1981). This is a study of reactions to a changing aircraft noise environment that was a consequence of runway repairs that modified aircraft operations. The noise data that were used in the analysis were all for the previous week. Eleven of the 20 data points are based on noise environments which had changed by from approximately five to 17 decibels in the past 3 months. Respondents were asked about the aircraft noise "... over the past YEAR (since this time last year),..". The measured noise data and the question to respondents are clearly not congruent for at least half of the data points. All responses could be affected by knowledge about changing operations.

USA-204 ORANGE COUNTRY AIRPORT

Estimates of noise levels are based on a method that corrected noise contours for surface area but not necessarily for the distribution of the population.

USA-250 DECATUR AIRPORT

The annoyance question uses a standard close-ended, annoyance question and not, as the 1991 synthesis article suggests, an open-ended question. The definition of the annoyance measure was confirmed through personal contact with the author of the study (Dr. Schomer).

USA-301 WESTCHESTER AIRPORT

The noise data is for all community noise at the microphone locations and not only for aircraft noise. This is consistent with the lack of a relationship between the measured noise levels and the closeness of the measurement points to the airport (Baldwin and Fidell, 1982: 13, 40). The response data in the synthesis is for only aircraft noise.

USA-338 U.S. AIRBASE

The synthesis reports that there are 874 respondents on the basis of data supplied in a personal letter from Mr. Stanley Harris. The researcher who conducted the study reports that there were 942 interviews and states that he disagrees with an analysis that has been conducted by Mr. Harris that involved dropping some data points that were considered to be "outliers" (Borsky, 1983). Neither the author, nor the organization that sponsored the study can locate a copy of the complete report on the study (personal communication with Mr. Harris and Mr. Borsky). The study was conducted in 1981 and perhaps 1982 around seven Air Force Bases in the United States (Dover, Mather, George, March, Marana, Carswell, Pope) with personal interviews.

INFORMATION ABOUT "NON-CLUSTERING" SURVEYS

The 1978 synthesis article contained discussions about several "non-clustering" data sets. These data sets have not been systematically examined during this review. The remainder of this appendix presents information to supplement the information in the 1978 synthesis article.

FRA-019 French Expressways

A total of about 345, not 400, interviews were used in the analysis. The annoyance question was worded as follows according to the text of an article (the questionnaire was not located):

57. Finally are you bothered by the freeway not at all, a little, some, a lot?

Original French Version
57. Finalement êtes-vous gênés, pas, peu, assez beaucoup par l'autoroute? (Lamure and Bacelon, 1976: 6)

SWE-021 1968 Stockholm/Gothenburg Street (Schultz, 1978: 395)

Annoyance is measured with an index that combines two questions: one on frequency of annoyance and one on degree of annoyance. The authors collapse this information into 6 categories and then assign scores of from 0 to 11 to each category based on a laboratory study. (Fog and Jonsson, 1968: 32-34)

USA-022, USA-032, USA-044 US Aircraft Surveys (TRACOR) (Schultz, 1978: 398)

The synthesis article does not accurately describe the annoyance index. Twelve 6-point activity interference, annoyance items were combined. Each one had been scored zero if the disturbance was not experienced and from one, if it was "not at all" annoying, to five if it was "extremely" annoying (Connor and Patterson, 1972: 128). The first 10 were weighted equally: relaxing/resting inside, relaxing outside, conversation, telephone conversation, listening to records/tapes, reading or concentration, eating. Two television items were given each given half the normal weight: "listening to radio/TV" "watching TV". Three nighttime items were each given 1/3 the standard weight: children sleeping/napping, going to sleep, late sleep. Connor and Patterson (1972: 24) suggest a criterion for high annoyance for this survey that disagrees with the 1978 synthesis article judgment for this survey.

APPENDIX C: THE EFFECTS OF INTRODUCING CERTAIN APPROXIMATIONS ON ESTIMATES OF LOGISTIC REGRESSION COEFFICIENTS

Logistic regression models estimate probabilities of observing dichotomous events. Direct estimates of logistic regression coefficients are thus derived from individual data in which each observation is scored as either a success (1) or failure (0). This appendix considers the impact of departures from this standard procedure.

Data from noise annoyance surveys are considered in which each respondent is classified as having either a "high" or not "high" degree of annoyance. These data are frequently published in an aggregated form in which each aggregate is characterized by the percentage highly annoyed, the average noise level, and, usually, the number of respondents. Although these data provide all of the information that is needed to conduct a standard analysis based on dichotomous individual responses, an alternative approximate method is sometimes used in which a transformation of the percentage highly annoyed is regressed on noise level in a linear regression analysis. In such instances, the analyst may also ignore the number of individuals in each group by assigning an equal weight to each group of individuals (rather than to each individual). The dependent variable in such an analysis is the logarithm of the ratio of the percent annoyed divided by the percent not annoyed. This method is labeled the "group approximation method" in this note. One reason for considering this to be an approximation is that the dependent variable (the logarithm of the odds) is undefined for groups in which 0 percent or 100 percent of the individuals in the group are annoyed.

Table 6 provides estimates of the two coefficients of a logistic regression equation under different conditions as the analysis methods and criteria for excluding cases from the data set are changed. Figure 5 plots four of the logistic curves that are considered. The most important statistic for comparing these analyses is the slope parameter, the last column in the table (Column 7). This slope is directly related to the slope of the derivative of the sigmoidal curve as it passes through the point at which the probability of observing an event is 0.5. The parameter thus determines how rapidly the slope of the curve changes around the point of inflection of the curve.

The first line in Table 6 presents the results for a standard logistic regression analysis. The columns describe this baseline analysis as being based on 23 data sets (Column 2) for which the units of analysis are individuals' responses (Column 3) and each individual is given the same weight (Column 4). If all of the individuals in the group agree, then the individuals simply receive scores of 0 or 1, just as for other groups (Column 5). The slope coefficient from this analysis is 0.09662 (Column 7) and the locational parameter (Intercept in Column 6) is 7.70868. The resulting logistic relationship appears as the line with the most shallow slope (Line A) in Figure 5.

The second line in Table 6 presents a coefficient based on one grouped approximation method. This is the analysis method that was used for the previously reported analysis of 23 of the synthesis data sets (Federal..., 1992: 3-5). This analysis is based on the same data as the first analysis, but uses an approximation in which a group of responses is the unit of analysis (Column 3), the individuals' responses are given a weight that is the inverse of the group size (i.e. each group is given the same weight) (Column 4), and groups in which there is complete agreement can not be represented by their true scores of 0% or 100%, but are instead represented by the fraction 0.01/99.99, if all individuals had a score of 0, or 99.99/0.01, if all individuals had a score of 1 (Column 5). This representation is necessary

because the grouped approximation method is based on a linear least squares regression analysis in which the dependent variable is the log of the odds ratio of the proportion of successes divided by the proportion of failures. The slope coefficient for this approximation is 0.14081* or approximately 45% greater than that for the standard analysis. This logistic relationship appears as the next-to-the highest slope (Line B) in Figure 5. The remainder of the table examines the reasons why this grouped analysis coefficient differs from the standard coefficient.

The next line (C) differs from the previous grouped analysis by weighting each individual's response equally. The resulting coefficient of 0.12784 is shallower and thus closer to the standard coefficient that also weights each response equally.

The next line (D) differs from the original grouped analysis in line B by using a more extreme approximation for the log odds ratio for 0% and 100% annoyance groups. In this analysis, the odds are assumed to be 0.0001/99.9999 and 99.9999/0.0001. The coefficient of 0.20143 is now twice that of the standard coefficient from the standard, individual analysis.

The analyses in the next four lines of Table 6 examine the possibility that the scoring of these 0% and 100% groups may explain the difference between the standard individual and grouped analyses. All four of these analyses are again based on the 23 data sets.

The comparisons of the first of these two pairs of analyses indicates that the scoring of the extreme 0%/100% groups can explain much of the discrepancy. In lines E and F in which both analysis techniques weight all individuals equally, the slope based on the grouped regression approximation is now 0.11491 (Line F in Figure 5), about 22% higher than the baseline individual analysis. In lines G and H in which both analysis techniques weight all individuals by the inverse of the group size, the coefficients are even closer. The grouped analysis estimate of 0.109476 is about 10% higher than that from the standard individual method.

It should be noted that for this particular data set, the grouped regression estimates are sensitive to the alternative assumptions. The highest estimate (0.20143 in line D) being over 85% higher than that of the lowest (0.10976 in line H). For the standard individual analysis method, however, the highest estimate is only 6% higher than that of the lowest.

For this particular data set much of the instability appears to derive from the way in which the linear regression analysis must incorporate an approximation for the groups in which all members of the group have the same score.

The last line of Table 6 addresses a separate issue: the effect of following particular rules for excluding data sets. The data that were considered for these analyses included the 29 data sets that are reported by Fidell, Barber and Schultz (1991). The slope coefficient for this full

---

*In the previously published analysis the value of the slope coefficient is the same (0.141) but the value of the intercept is slightly different in the second decimal point (11.18). (Federal..., 1992: 3.5) This slight discrepancy has not been explained.

set of data from a grouped regression analysis is 0.13038. All of the previous analyses in this table are based on only 23 of these 29 data sets. The 23 data sets were selected by excluding the six data sets with 95% confidence intervals that included the possibility of a negative relationship with noise level. By excluding these data sets, the one analysis that otherwise used exactly the same analysis methods (the grouped analysis in line B) created a slightly steeper slope (0.14081). The reason for the direction of this discrepancy is readily apparent. If the true relationship between noise level and annoyance is slightly positive, then small sample surveys with unstable estimates will include some samples with extremely high slopes and other samples with extremely low slopes, but the average of all of these small surveys would still give the correct, slightly positive slope. When however, as was the case in line B, only the small surveys with shallow slopes are rejected, then the overall estimate is biased by the remaining, equally inaccurate, surveys which happened to have extremely high slopes.

APPENDIX D: ALTERNATIVE MODELS FOR SUMMARIZING THE SYNTHESIS DATA

This appendix describes the three functions that are graphed in Figure 1.

Three of the best-known summaries of some of the synthesis data sets are presented in Figure 1. The original analysis of the 12 clustering data sets in the 1978 synthesis article is represented by the solid line in Figure 1 and is described by the following cubic equation (Schultz, 1978: 382):

$$P = 0.8533 \times Ldn - 0.0401 \times Ldn^2 + 0.00047 \times Ldn^3 \quad (1)$$

An analysis of 23 of the 29 data sets is represented by the heavy dashed line and by the following logistic relationship (Federal.., 1992: 3-5)

$$P = \frac{100}{1 + e^{-(-11.13 + 0.14Ldn)}} \quad (2)$$

An analysis of all 29 data sets is represented by the more broken dashed line. This relationship is based on an exponential error model proposed by Fidell, Schultz and Green (1988) with the empirically estimated parameter of 21.9.*

$$P = \frac{100}{e^{-\left(\frac{10^{21.9/10}}{(10^{Ldn/10})^{0.3}}\right)}} \quad (3)$$

* Table III in Fidell, Schultz, and Green (1991) provides values of 72.5, 73.0 and 73.1 for the parameter D^* from three mutually exclusive subsets of the 29 surveys. All three estimates give, when rounded, the same estimate of A^*=21.9 for the parameter in the equation above. The authors of the article give separate values for different transportation sources rather than a single overall value (Fidell, Schultz, and Green, 1991: 241).

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 1994	Contractor Report

4. TITLE AND SUBTITLE
A Review of an Updated Synthesis of Noise/Annoyance Relationships

5. FUNDING NUMBERS
Contract NAS1-19061
WU 538-03-15-01

6. AUTHOR(S)
James M. Fields

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Georgia Institute of Technology
GTRI/ASTL
Atlanta, GA 30332-0420

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
National Aeronautics and Space Administration
Langley Research Center
Hampton, VA 23681-0001

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
NASA CR-194950

11. SUPPLEMENTARY NOTES
Langley Technical Monitor: Kevin P. Shepherd
Final Report

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Unclassified — Unlimited
Subject Category — 71

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

This report evaluates the evidence supporting the "dosage-effect relationship for the prevalence of annoyance due to general transportation noise" that was originally presented in 1978 (Schultz, T. J.: Synthesis of Social Surveys on Noise Annoyance, J. Acoust. Soc. of Am., vol. 64, pp. 377-405) and more recently updated in 1991 (Fidell, S.; Barber, D. S.; and Schultz, T. J.: Updating a Dosage-effect Relationship for the Prevalence of Annoyance due to General Transportation Noise, J. Acoust. Soc. Am., vol. 89, pp. 221-233). The synthesis methods and many of the constituent studies' reports are examined.

The weaknesses of the synthesis are identified and found to be partly due to the synthesis methodology, but fundamentally to the lack of comparability in the constituent studies' methodologies and reporting procedures. Several suggested improvements to survey, analysis, and synthesis techniques are proposed.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Environmental Noise; Aircraft Noise; Social Surveys	91
	16. PRICE CODE: A05

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	