CONSEQUENCES OF CERTAIN DATA ERRORS IN
THE TRANSPORTATION PLANNING PROCESS

MAY 1972 - NUMBER 3

BY

R. W. ZIMMER

JHRP

JOINT HIGHWAY RESEARCH PROJECT

PURDUE UNIVERSITY AND
INDIANA STATE HIGHWAY COMMISSION

CONSEQUENCES OF CERTAIN DATA ERRORS IN THE TRANSPORTATION PLANNING PROCESS

File: 3-7-8

M. B. Scott
J. A. Spooner
N. W. Steinkamp
H. R. J. Walsh
K. B. Woods
E. J. Yoder

Digitized by the Internet Archive
in 2011 with funding from
LYRasis members and Sloan Foundation; Indiana Department of Transportation

Final Report

CONSEQUENCES OF CERTAIN DATA ERRORS IN THE
TRANSPORTATION PLANNING PROCESS

by

Ralph W. Zimmer
Graduate Instructor in Research

Joint Highway Research Project

Project No.: C-36-69H

File No.: 3-7-8

Prepared as Part of an Investigation

Conducted by

Joint Highway Research Project
Engineering Experiment Station
Purdue University

In cooperation with the
Indiana State Highway Commission
and the

U.S. Department of Transportation
Federal Highway Administration

The opinions, findings and conclusions expressed in this publication are those of the authors and not necessarily those of the Federal Highway Administration.

Purdue University
Lafayette, Indiana
May 1, 1972

ACKNOWLEDGMENTS

The author wishes to express his deep appreciation to Professor Harold L. Michael, Associate Director of the Joint Highway Research Project, for his very knowledgeable guidance of this research project and his critical review of the manuscript. Gratitude is also expressed to Professors William L. Grecco, Maurice Halstead, and Gilbert T. Satterly, Jr. for their review of the final manuscript.

Messer's Glenn E. Brokke and Jim McDonnell of the Urban Planning Division of the Federal Highway Administration in Washington, D.C., as well as several members of their respective staffs, provided advice and considerable assistance.

Vogt, Sage & Pflum, the consultant performing the Evansville transportation study, very willingly donated a substantial number of hours of their employees' time and incurred some expense in assisting the author. Although called upon to a lesser extent, the local planning staffs in both Evansville and Lafayette similarly provided substantive assistance.

The Indiana State Highway Commission permitted the use of their computer facilities for this research project and provided office facilities in Indianapolis. Additional office space in Lafayette was provided by the Lafayette Transportation Study.

The assistance and advice provided by Mr. Jay Marks of the Indiana State Highway Commission were so significant that it is doubtful that this project could have been completed successfully without them, and the author is greatly indebted to him.

Last, but certainly not least, gratitude must be expressed to the author's wife and son for their understanding sacrifice of normal family fellowship over an extended period of time.

TABLE OF CONTENTS

	Page
LIST OF TABLES.....	vi
LIST OF FIGURES.....	viii
NOMENCLATURE.....	ix
ABSTRACT.....	xvi
CHAPTER I - INTRODUCTION.....	1
CHAPTER II - OBJECTIVES AND POTENTIAL BENEFITS.....	5
Objectives.....	5
Potential Benefits.....	6
CHAPTER III - REVIEW OF LITERATURE.....	8
Code Design.....	8
Quality Control of Coding and Digitizing Operations...	10
Coding.....	10
Digitizing.....	11
Computer Editing.....	13
Directional Split Assumptions.....	15
Sensitivity of Travel Models to Errors.....	16
CHAPTER IV - BACKGROUND INFORMATION AND INITIAL DECISIONS.....	17
Study Area and Computer Battery Selections.....	17
The Evansville Regional Area Development and Transportation Study.....	18
The Greater Lafayette Area Transportation and Development Study.....	20
Selection of Computer Editing Procedures.....	21
CHAPTER V - DATA EDITING, EXPANSION, AND FACTORING.....	25
Data Editing.....	25
Development of Expansion Factors.....	27
Survey Accuracy Checks.....	28
Development and Application of Trip Factors.....	29

	Page
CHAPTER VI - TRAFFIC ASSIGNMENT.....	34
O&D Triptables.....	34
Network Loads.....	34
CHAPTER VII - TRIP DISTRIBUTION.....	40
General.....	40
Results.....	41
CHAPTER VIII - TRIP GENERATION.....	44
General.....	44
Production Equations.....	45
Attraction Equations.....	47
CHAPTER IX - EFFECTS OF DIRECTIONAL SPLIT ASSUMPTIONS...	49
General.....	49
Comparison of Network Loads.....	49
CHAPTER X - ADDITIONAL ERROR ANALYSES.....	55
Residual Incorrectly Recorded Data Errors.....	55
Key punching Errors.....	58
CHAPTER XI - SUMMARY AND CONCLUSIONS.....	61
Results of Using Various Editing Procedures.....	61
Directional Split Assumptions.....	62
Computer Program Contributions.....	62
CHAPTER XII - GUIDELINES FOR DATA PROCESSING AND EDITING	63
Notes on Form and Code Design.....	63
Guidelines for Data Processing.....	65
Guidelines for Computer Editing.....	66
CHAPTER XIII - RECOMMENDATIONS FOR ADDITIONAL RESEARCH..	68
LIST OF REFERENCES.....	69
APPENDICES	
Appendix A: Survey Forms.....	75
Appendix B: Geographic Coding Indices.....	83
Appendix C: Error Checking Specifications and Results.....	88
VITA.....	103

LIST OF TABLES

Table		Page
1.	Summary Statistics From Error Checking Process..	26
2 .	Dwelling Unit Characteristics As Related To Editing Procedure.....	28
3.	Total Screenline Crossings As Obtained From Expanded But Unfactored Survey Data.....	30
4.	Trip Factors To Compensate For Underreporting...	30
5.	Total Screenline Crossings As Obtained From Expanded And Factored Survey Data.....	32
6.	Percentage Of Total Vehicle Trips By Purpose....	33
7.	Comparison Of Procedure A To Procedure C Triptables.	35
8.	Comparison Of Procedure B To Procedure C Triptables.....	36
9.	Network Loading Statistics.....	37
10.	Comparison Of Procedure A To Procedure C Loads..	38
11.	Comparison Of Procedure B To Procedure C Loads..	39
12.	Home Based Work Trip Length Distribution Summary Statistics.....	42
13.	Home Based Shopping Trip Length Distribution Summary Statistics.....	42
14.	Ratio Of Observed To Distributed Trips By Time Increment.....	43
15.	Production Equation Variable Means.....	46
16.	Correlation Data For Home Based Production Equation Variables.....	46

Table		Page
17.	Estimated Home Based Work Attractions.....	48
18.	Estimated Home Based Shopping Attractions.....	48
19.	Comparison Of Correctly Split And 50-50 Split Network Loads.....	50
20.	Comparison Of O&D And Correctly Split Network Loads.....	52
21.	Comparison Of O&D And 50-50 Split Network Loads.	53
22.	Comparison Of O&D And P&A Network Loads.....	54
23.	Error Rates Expressed In Percentages Found In Travel Survey Data During Recoding.....	56
24.	Breakdown Of Errors In Coding Of Trip End Addresses By Cause.....	57
25.	Keypunching Errors.....	60

Appendix Table

C1.	Home Interview Survey Error Checks.....	89
C2.	Home Interview Survey Error Rates.....	91
C3.	Partial List Of Other Feasible (But Not Made) Error Checks Of Home Interview Survey Data.....	94
C4.	External Survey Error Checks.....	95
C5.	External Survey Error Rates.....	96
C6.	Truck/Taxi Survey Error Checks.....	97
C7.	Truck Survey Error Rates.....	98
C8.	Taxi Survey Error Rates.....	100
C9.	Number Of Cards Vs. Errors Per Card By Survey...	102

LIST OF FIGURES

Appendix Figure	Page
A1. Evansville Dwelling Unit Summary Card Form.....	76
A2. Evansville Internal Trip Card Form.....	77
A3. External Survey Trip Card Form.....	78
A4. Truck/Taxi Trip Card Form.....	79
A5. Truck/Taxi Trip Card Continuation Form.....	80
A6. Lafayette Dwelling Unit Summary Card Form.....	81
A7. Lafayette Internal Trip Card Form.....	82
B1. Evansville Street Address Coding Index.....	84
B2. Evansville Place-Name Coding Index.....	85
B3. Lafayette Street Address Coding Index.....	86
B4. Lafayette Place-Name Coding Index.....	87

NOMENCLATURE

ATTRACTION: The pull or attracting power of a zone, usually measured in terms of number of trip ends. In gravity model P&A terminology, the attraction zone is:

1. For home based trips, the zone corresponding to the non-home end of the trip.
2. For non-home based trips, the zone of destination.

AVERAGE DAILY TRAFFIC (ADT): The average number of vehicles passing a specified point during a 24 hour period.

BASE YEAR: The year to which all survey data are related. Usually the year in which the survey data were collected.

BMD02R: A widely used computer program for performing stepwise multiple regression developed by the University of California.

CLUSTERED TRIP DATA (OR INFORMATION): Trip data obtained through a statistical sampling process wherein the unit actually being sampled is not the trip but some associated unit such as the vehicle or dwelling unit.

CODE: (noun) A system of symbols for representing information. (verb) To reduce data to a more meaningful form for subsequent data processing.

CONSISTENCY CHECK: An error check in which the contents of two or more data fields are mutually examined to insure that they are not contradictory. The fields involved may be in the same or different records. Sometimes called a contingency check.

CORDON LINE: An imaginary line enclosing a study area.

COUNT: The actual number of vehicles passing a specified point during a stated time period as determined by human or mechanical tabulation.

DEPENDENTLY CHECK: To attempt to verify the correctness of previously obtained codes by having a second person examine the original codes and decide whether they were correct or not.

DESTINATION: The terminus of a trip.

DIGITIZATION: The process of converting data to a form which is directly readable by a computer. In the past, keypunching has been the digitization process most frequently used.

DIRECTIONAL LINK: A link in which traffic is allowed in only one direction. There may be, and usually is, a corresponding link in the opposite direction.

DIRECTIONAL SPLIT: In gravity model P&A terminology, a pair of numbers indicating the relative orientation of home based trips to and from the production (home) end. The first number is the percentage of productions which actually originate in the production zone and terminate in the attraction zone. The second number is the percentage of productions which actually terminate in the production zone and originate in the attraction zone. The two numbers sum to 100 per cent unless factoring is being performed.

DWELLING UNIT (D.U.): A room or group of rooms occupied or intended for occupancy as separate living quarters by a family or other group of persons living together or by a person living alone.

EDIT: To perform the two step process of checking data for errors and rectifying any errors which are found. Sometimes used to mean solely the checking of data for errors.

EDITING PROCEDURE: A methodology for data editing consisting of an error checking step and an error rectification step.

ERROR CHECK: A test made to insure that a given data field (or combination of data fields) conforms to predefined conditions of acceptability.

ERROR RECTIFICATION: The process whereby individual trip cards or groups of associated trip cards containing data failing error checks are either discarded or retained. If retained, methods of substituting data which will pass the error checks for those data which failed the error checks must be specified. The substituted data do not necessarily have to be "correct" data.

EXPANSION FACTOR: A multiplication factor indicating the number of units in the entire population which the given unit represents.

EXTERNAL: An adjective indicating outside of the study area.

EXTERNAL TRIPS: Those trips which cross the cordon line.

FRICTION FACTORS (F FACTORS): A set of numbers, one for each travel time increment, indicating the effects of travel time upon number of trips. Used in the gravity model.

GEOGRAPHIC CODING INDICES: Reference material similar to telephone directories used by coders in determining in which zone a given trip end lies. The two main types are the street address coding index and the place-name coding index. See Appendix B for examples.

GRAVITY MODEL: A mathematical formula that distributes trips between zones proportional to the attraction of the destination zone and inversely proportional to some function of the separation between the zones.

HOME BASED TRIPS: Trips with either end at the residence (home).

HOME BASED SHOPPING (HBS) TRIPS: Those home based trips in which the purpose at the non-home end is shopping.

HOME BASED WORK (HBW) TRIPS: Those home based trips in which the purpose at the non-home end is work.

IMPEDANCES: A series of computer records containing the resistance to travel (usually travel time) between each pair of zones. Formerly called skimmed trees.

INCORRECTLY RECORDED DATA: Data which, after passing through any or all of the human/mechanical processes of interviewing, coding, or digitizing, differs from what it should be based upon the information input to the process.

INDEPENDENTLY CHECK: To verify the correctness of previously obtained codes by assigning additional coder(s) to recode the data. These additional coders would only be exposed to the information from which the codes were derived and not to the original coding. After all coders have completed their tasks, corresponding data items would be compared to determine whether the various coders agreed or disagreed.

INTERNAL: An adjective indicating inside of the study area.

INTRAS: The number of intrazonal trips.

INTRAZONAL TRAVEL TIME: The average travel time for trips beginning and ending in the same zone, excluding terminal time.

LINK: A section of street or highway identified by the nodes at its ends. A link may be one-way or two-way.

LINKED TRIPS: Those trips remaining after application of a process whereby each sequence of trips with intermediate purposes of serve passenger and/or change mode is reduced to a single trip representing the "true" origin and destination.

LOADING (THE NETWORK): The computer process of accumulating loads upon each link in the network as trips from a triptable matrix are assigned to their associated routes.

LOADS (LINK OR NETWORK): The volume assigned on an individual link or the collection of volumes assigned on all links in a network as a result of loading the network.

LOCAL LINK: A link representing the local streets which has as one node a centroid (that point in a zone which is used to load all trips to and from that zone).

MARK SENSE READER: A mechanical device which digitizes scan sheets. The output media is a punched card.

MULTIPLE PUNCHES: The technique by which a single column on a data card is used for recording two (or more) separate data items. Normally, one data item will be a number and the second will be something such as sex which could be indicated by the presence or absence of an additional punch in row eleven or row twelve of the data card.

NETWORK: The collection of links and nodes by which the major components of the street system within the study area are defined to the computer.

NODE: A numbered, identifiable point in a network at the junction of two or more links.

NON-HOME BASED (NHB) TRIPS: Trips with neither end at the residence (home).

ORIGIN: The point at which a trip commences.

ORIGIN AND DESTINATION (O&D) DATA: A collection of data indicating the true origins and destinations of trips.

PER CENT ROOT MEAN SQUARE ERROR: A statistical measure of the differences between two sets of corresponding numeric data, one set of which is called the base data against which the other set, called the test data, is compared. If there are n data values in each set and b_i and t_i are the values of the i -th pair of base and test data, respectively, then:

$$\% \text{ R.M.S.E.} = \frac{100 \sqrt{\frac{\sum_{i=1}^n (b_i - t_i)^2}{n}}}{\frac{\sum_{i=1}^n b_i}{n}}$$

PLACE-NAME CODING INDEX: A geographic coding index consisting of a list of locations commonly known by their names and the associated zone numbers.

PROCEDURE A: As defined in this research project, the editing procedure (or the travel survey data after application of the editing procedure) whereby no error checks were made and thus no error rectification was needed.

PROCEDURE B: As defined in this research project, the editing procedure (or the travel survey data after application of the editing procedure) whereby extensive error checks were made and all samples containing errors were discarded in their entirety.

PROCEDURE C: As defined in this research project, the editing procedure (or the travel survey data after application of the editing procedure) whereby extensive error checks were performed, all possible corrections were made to the data in error, and the few remaining samples which had errors which could not be corrected were discarded in their entirety.

PRODUCTION: The generating power of a zone, usually measured in terms of number of trip ends. In gravity model P&A terminology, the production zone is:

1. For home based trips, the zone of residence.
2. For non-home based trips, the zone of origin.

PRODUCTION AND ATTRACTION (P&A) DATA: A collection of trip data or trip ends reflecting the gravity model concept of production and attraction zones rather than the true origins and destinations.

PURPOSE: The reason for making a given trip. In the travel surveys, a purpose is usually associated with each trip end. For analysis, a single trip purpose is determined based on the two trip end purposes.

QUALITY CONTROL: The techniques used to assure accuracy of results in the interviewing, coding, and/or digitizing operations.

RESTRAINED LOADS: The network loads obtained through an iterative loading process in which, at each step, the routes of travel are modified to reflect the effects of the loads on the individual links as found in the preceeding iteration(s).

SAMPLE: The individual unit (or the data associated with it) that was statistically selected for interview.

SCAN SHEET: A special form on which numeric data are coded by filling in with a pencil a series of rectangular blocks each of which is associated with an individual numeric digit (see Figure A3). A mark sense reader is then used to digitize the data.

SCREENLINE: An imaginary line, usually along physical barriers such as rivers or railroad tracks, splitting the study area into two parts.

SINGLE FIELD CHECK: An error check in which the contents of a single data field are examined without reference to the contents of any other data fields.

STATION: A location at the cordon line where external survey (roadside interview) data or count data were collected.

STREET ADDRESS CODING INDEX: A geographic coding index showing the zone numbers associated with the various ranges of house numbers on each street in the study area.

TERMINAL TIME: That time required to unpark (or park) and the associated walking time required to complete the trip exclusive of the actual vehicular travel time between the origin and the destination.

TRIP: A one-direction movement which begins at the origin at the start time, ends at the destination at the arrival time, and is conducted for a specific purpose.

TRIP CARDS: Data cards containing survey-derived trip and related information. Not every trip card represents a trip, but the data for each individual trip is punched on a single trip card.

TRIP END: Either a trip origin or a trip destination.

TRIP TABLE: A matrix in computer-readable format showing how the trips, either O&D or P&A, are distributed between each pair of zones.

UNRESTRAINED LOADS: Network loads obtained through a single loading without any iterations.

VINE: A computer record giving the path of minimum impedance from a given zone to all other zones. The path is permitted to cross over itself (e.g., three successive right turns in lieu of a single prohibited left turn are feasible).

ZONE: A portion of the study area which is delineated as such for the purpose of facilitating land use and traffic analyses.

ABSTRACT

Zimmer, Ralph Wildy. Ph.D., Purdue University, June 1972. CONSEQUENCES OF CERTAIN DATA ERRORS IN THE TRANSPORTATION PLANNING PROCESS. Major Professor: Harold L. Michael.

Using unedited travel survey data from the Evansville, Indiana, transportation study, the major objectives of this project were:

1. To classify and enumerate the number of incorrectly recorded data items in each of the sets of travel survey data (invalid or inconsistent codes, unreasonable magnitudes, omitted items, etc. are classified as incorrectly recorded data).
2. To determine the effects upon the calibrated base year traffic assignment, trip distribution, and trip generation models of following different computer editing procedures.
3. Based upon the above results, to formulate guidelines for the computer processing and editing of travel survey data.
4. To determine the error introduced by assuming a 50-50 directional split in converting triptables from origin and destination format to production and attraction format.

An editing procedure was defined as a two step process: error checking and error rectification. Error checking is the step in which the individual data fields in trip cards

are examined to insure that they conform to pre-programmed criteria of acceptability. Error rectification is the step in which some action (e.g., deletion or correction) is taken relative to the data which failed the error checks.

The results obtained when using thoroughly edited and corrected survey data were not significantly different than the results obtained using totally unedited data. Significantly different results were obtained, however, when using data which had been thoroughly edited and the incorrect data discarded.

Based on the above results, recommended guidelines for computer processing and editing of travel survey data were formulated. These call for making important single field edits but only a minimal number of consistency checks. All incorrect data in samples of clustered trip information should be rectified rather than discarded.

The review of literature disclosed a number of factors worthy of consideration in form and code design. Similarly, information on keypunching error rates was found in the literature and supplemented by data obtained during the course of the project.

An assumed 50-50 directional split for ADT trips was found to be justified.

CHAPTER I INTRODUCTION

Successful transportation planning requires accurate prediction of future travel demand and patterns of movement. It has long been recognized that such predictions, in order to be accurate, must be based upon a thorough understanding of the existing travel patterns in the given geographic area.

It has been estimated that during just the twelve year period 1958-1970, approximately two hundred million dollars were spent on urban transportation planning (7)*. A major portion of the cost of the comprehensive transportation planning process as now practiced is spent in the data collection phase using travel survey methodology which has changed little during the last twenty-five years. Data on internal trips of a personal nature by residents of the study area are collected in the home interview survey, data on trips in the study area by non-residents and on trips into and out of the study area by residents are collected in the external (roadside interview) survey, and data on truck and taxi trips are collected in the truck and taxi surveys, respectively.

Prior to the mid-1940's, transportation planners were primarily interested in determining the best locations for urban river crossings and bypasses. To provide information on which to base their decisions, roadside interview origin and destination (O&D) studies were instigated.

*The numbers in parenthesis refer to sources listed in the List of References.

In awareness of the growing transportation problems in urban areas, the United States Congress passed the Federal-Aid Highway Act of 1944 making funds available for highway projects in urban areas. Since roadside interviewing techniques were inadequate in providing the necessary data about existing travel patterns within an urban area, an interviewing procedure involving the sampling of a representative number of dwelling units within the urban area was developed cooperatively by the U.S. Bureau of Public Roads (now the Federal Highway Administration), the U.S. Bureau of the Census, and several state highway departments. Eight cities conducted home interview surveys using these procedures during 1944 (59).

Urban transportation planning, and thus travel surveys, were given added impetus by the Federal-Aid Highway Act of 1962 which basically requires all urban areas of over 50,000 population to conduct a cooperative, comprehensive, and continuing transportation planning process in order to qualify for financial participation by the Federal Government on those segments of the Federal-Aid system lying within the urban area. To date, over 200 urbanized areas are conducting such studies.

After travel survey data have been collected, they are coded and converted to computer-readable format (e.g., keypunched). The resulting number of computer card images is staggering, ranging from tens of thousands of card images for the smallest study to a few millions of card images for studies in the largest metropolises (38). These cards can and do contain a number of errors, and these might be arbitrarily classified as sampling errors, deliberate errors (e.g., manufactured data), underreporting errors (i.e., the unintentional failure of the interviewee to report all of his trips), and incorrectly recorded data errors. Substantial research has been devoted to investigating different sampling procedures and sampling errors (18,22,47), and methodology

exists and is routinely followed for identifying and correcting deliberate errors (43) and the effects of under-reporting (53).

Incorrectly recorded data errors consist of such errors as an invalid zone number or wrongly blank data field. Such errors can be attributed to unintentional erroneous disclosures by the interviewee, incorrect recording by the interviewer, incorrect transcription by the coder, and/or failures in the human/mechanical process by which the data are converted to computer-readable format. As a practical matter, many such errors could never be detected, but a significant portion of the most critical errors could be located by diligent use of special purpose computer "editing" programs. It is generally assumed that all travel survey data are thoroughly edited before being used in model development.

After the travel survey data have been committed to computer-readable format and edited, various survey accuracy checks are made to assess the accuracy and completeness of both the dwelling unit and the trip data. As a result of these checks, it is normally necessary to factor the trip data in order to compensate for the traditional under-reporting of trips.

Using these factored travel survey data and other data, the analyst is finally able to calibrate the various models used in transportation planning studies: land use, trip generation, trip distribution, modal split, and traffic assignment. Land use models are used to predict future land use, trip generation models are used to estimate the number of trips by purpose which will originate or terminate in a given geographic area (zone), trip distribution models are used to determine how many of the trips which originate in a given zone terminate in each of the other zones, modal split models determine how many of the total trips are made by each of the various major modes (e.g., auto driver, commercial bus, etc.), and traffic assignment models are used to allocate

the various interzonal trip movements to specific routes and thus predict the vehicular volume on each of the major street segments within the urban area. Trip generation, trip distribution, and traffic assignment models are always used, but land use and modal split models frequently are not used in studies involving smaller urban areas.

For over a decade, the most frequently used trip distribution model for internal-internal trips has been the gravity model. Most such trips have one end at the home and thus are treated as home based with, by definition, the production zone being the zone of residence and the attraction zone being the zone in which the other trip end lies. This results in loss of directionality, and there is no way that the corresponding origin and destination zones can be determined precisely. While there are some significant advantages in both the trip generation and trip distribution models in working with production and attraction (P&A) data rather than origin and destination (O&D) data, there is a concomitant disadvantage in that the output home based triptables must be converted from P&A format to O&D format before they can be loaded onto the network during the traffic assignment process. To do this, the analyst must specify the desired directional split (e.g., 50-50 or 55-45) by purpose.

CHAPTER II OBJECTIVES AND POTENTIAL BENEFITS

Objectives

The objectives of this study were:

1. To classify and tabulate all identifiable incorrectly recorded data errors in the unedited trip cards from a given study.
2. To determine the ultimate consequences of applying various computer editing procedures for locating and rectifying those errors.
3. To formulate, based on the above results, recommended guidelines for trip card editing procedures.
4. To determine, if possible, techniques by which the occurrence of incorrectly recorded data errors could be minimized.
5. To supplement existing knowledge about the sensitivity of the trip generation, trip distribution, and traffic assignment models to input data errors.
6. To determine the error which results from assuming a 50-50 directional split when converting home based P&A triptables to O&D triptables for a relatively small urban area.

Potential Benefits

An understanding of the nature and sources of incorrectly recorded data errors would hopefully lead to the development of techniques for the minimization of such errors.

Data on the number and type of incorrectly recorded data errors, when combined with similar data from other studies, would provide a factual basis for subsequent transportation studies to use in evaluating the quality of work being performed by their interviewers, coders, and keypunch (or equivalent) operators and/or their equipment. Unsatisfactory performance could then be corrected before the resulting errors invalidate subsequent analyses.

Quantitative data would be available for estimating the amount of error that would be introduced by following different computer editing procedures. Such data would be invaluable, for example, if data errors were suddenly discovered part way through the transportation planning process, for the analyst would have some valid basis for estimating the effects of the errors.

A factual basis would be available for the development of recommended guidelines for the computer editing of trip cards. The implementation of such guidelines would insure that the data are "adequately" edited and, perhaps, result in savings in time and money for those studies that otherwise might allocate an unnecessary amount of resources to "completely" edit the data.

Supplementary knowledge about the sensitivity of the various transportation planning models to input data errors might help the analyst in knowing how finely to "tune" each of the models. In addition, such knowledge could assist researchers in determining where best to allocate their funds and efforts.

The results of the investigation into the effects of directional split when converting home based P&A triptables

for a smaller urban area to O&D format should either substantiate the existing practice of assuming a 50-50 directional split or indicate that this practice introduces an excessive amount of error. In the latter case, the accuracy of the transportation planning process could be improved although there would be a corresponding but slight increase in cost.

CHAPTER III REVIEW OF LITERATURE

Code Design

Non-quantitative travel survey data, such as zone numbers, land use information, trip end purposes, sex, type of commodity, etc., are represented by a multi-digit (one or more) field called a code. Transportation planners generally have devoted little thought to the selection of the codes to be used other than attempting to pick codes which would facilitate computer processing.

Myers listed the characteristics of a good code and classified codes into five basic types (33). Other authors have suggested different code classifications (4,56). Not surprisingly, the two organizations which have considered general code design in some detail are the Federal Government (56) and the Bell Telephone System (4). The business industry has shown some interest in the selection of codes to represent accounts, but their interest has been largely in selecting codes which would allow sufficient expansion capabilities to avoid the necessity of assigning a new customer an out-of-sequence code.

Banking institutions in particular are concerned that, because of a human or mechanical error in entering an account number, a transaction might be "credited" to the wrong account. Such institutions have generally adopted a modulo-n check digit as the suffix to the normal account number (2,3,6,20,41,64). The value of the check digit is mathematically dependent upon the values of the preceeding digits, so a surprisingly large number of errors (e.g., all single digit substitutions and transpositions of two

adjacent digits) can be readily detected. Mod 10 and mod 11 systems are used the most frequently, but mod 9 and mod 7 systems are also in use (50). Consideration might be given to using a check digit in sample numbers and/or survey zone numbers, although such use might not be justified.

The psychological literature is full of research which might be worthy of consideration in the design of codes. In a classic paper, Yule reported substantial evidence that people have decided preferences for certain digits in reading the last digit of a scale and in reporting the last digit of their age (65). Other research indicates similar preferences among letters of the alphabet (25).

Stanhagen investigated coding errors and developed a matrix showing the probability that any given character would be incorrectly coded for another given character (48). Neisser and Weene published a similar matrix showing the number of times a given hand-printed character was interpreted as another character (34). Other research indicates the highest error rates occur when coding/copying alphabetic characters, the lowest occur when working with numeric characters, and error rates in between these two extremes occur when working with alpha-numeric codes (12,48).

Numerous researchers have considered the optimum number of characters to be grouped together for short-term memory retention such as for dialing of telephone numbers or manual transcription. They have generally concluded that groups of 3 or 4 characters are best (10,28,45). This might suggest that long survey zone numbers and sample numbers be artificially broken into smaller groups in order to minimize errors and speed coding.

Conrad (11) and Waugh (63) have shown that errors are not randomly distributed among the various characters of a multi-character group of digits subjected to short-term memory. Errors are most likely in the next-to-last character and the second-from-last character and are least likely in

the first and last characters of the group. This might suggest that codes (e.g., survey zone numbers) be designed which place the two most important digits in the first and last positions, respectively.

Quality Control of Coding and Digitizing Operations

Most transportation studies have used rather straightforward, unsophisticated quality control procedures of the coding and digitizing (e.g., keypunching) operations. Typically, coding quality control has been limited to having a second coder dependently check all or a portion of the interviews simply making whatever corrections are considered necessary. Most studies have a supervisor scan the forms for completeness and continuity. Digitizing quality control has typically been limited to 100 per cent key verification with the second operator simply making whatever corrections are necessary.

The Bureau of the Census has developed and use much more sophisticated quality control procedures (14,17,31,32,58). These generally consist of a hybrid procedure combining elements of both process control and acceptance sampling. Although they previously exclusively used dependent verification techniques, they now rely on various independent verification methods. This change came about after their own research revealed that dependent verification of coding failed to disclose from 29 to 69 per cent of the actual errors, with an average of about 50 per cent of the errors being missed (17,31,58).

Coding

Once the field interview forms reach the office, they are "coded" for subsequent conversion to machine-readable format. This coding operation entails a variety of operations ranging from simple copying of stated numeric data (e.g., number of cars) to determination of a correct zone

number or other classification and the recording of an appropriate identifying code.

The literature reveals little about the nature and magnitude of errors which occur specifically during coding, but it would seem reasonable that their nature is similar to that of digitizing errors as discussed in the next section. It would also seem reasonable to assume that the majority of the errors found during computer editing are either coding errors or should have been discovered and corrected during coding. In the one reference which was found in the literature, Fasteau et al. reported that 3.0 per cent of the industry and occupation codes in the 1960 Census were incorrect (17).

In a Rand Corporation study, Owsowitz and Sweetland suggested that coding errors occur randomly (40). For example, if there were ten possible codes and a coding error were made, they indicated that there is an equal probability that any of the other nine possible codes would have been coded even if only five of the nine were permissible entries. Stanhagen subjected this assumption to extensive research and found it to be invalid. Instead of being random, erroneous codes showed a marked tendency toward "expected values" (48).

Digitizing

Digitizing is the process of converting source documents to machine-readable format. In the past, keypunching has been almost the sole method of digitization. However, new "data entry" methods such as keyboard to tape/disc are evolving (1,5,13) so the more general term of digitizing is used in this text.

In his classic book, Brandon cited standards for keypunching speeds of 5,000 keystrokes per hour for all alphabetic punching, 4,000 keystrokes per hour for mixed alphabetic/numeric, 8,000 keystrokes per hour for large

percentage numeric with little alpha, and 10,000 keystrokes per hour for all numeric punching (8). Due to error detection stops, Brandon stated that key verification takes about 3 per cent longer than keypunching. Significantly higher and supposedly more accurate digitizing and verification speeds are claimed for some of the new keypunch replacement equipment primarily as the result of the elimination of electro-mechanical delays and the elimination of individual "punched" cards.

In comparing digitizing error rates, it is necessary to use errors per column or errors per keystroke as the base measure, for the percentage of cards in error is highly dependent upon the number of required keystrokes per card. In addition, it is normal to not consider errors by the machine operator which the operator corrected. Within this context, four researchers have reported digitizing error rates of between 0.02 and 0.06 per cent of the total keystrokes (15,21,24,29).

Almost all travel survey data are key verified to insure the accuracy of the original digitization. One consultant has reported investigating the use of unverified data, but he concluded that the "...saving in punching cost is more than offset by cost of engineers and technicians correcting the errors" (16). Key verification does not remove all digitization errors. Hinds (24) found one per cent of the original errors remained after verification, and Minton (31) has reported from 3-14 per cent of the original errors in Census data remained after verification.

Smith (46) and Carlson (9) have classified the errors which they found in data entry operations. Their results, which are very similar, show that over 50 per cent of the errors are due to single-digit substitution, about 20 per cent due to omission of a digit, about 6 per cent to insertion of a digit, and 1.5-7.3 per cent due to transposition of adjacent digits.

The in-depth study by Deming et al. (15) revealed several other items of interest. First, they found keypunching error rates increase from left-to-right across the card, thus indicating that the most important data should be placed as far left as possible. Second, they found that, within a given data field, "unusual" entries are more likely to be in error than "usual" entries. Third, given that a keypunch error has occurred, the erroneous result tends toward an expected value.

There are some significant ramifications to some of the new digitizing equipment currently on the market. In some of the equipment, there is actual computer interface between the keyboard operator and the storage media receiving the data. Thus, a certain amount of computer editing can be done concurrently with data entry. It has previously been suggested that keypunch operators could do much of the manual editing (27), and this new technology makes such a procedure highly desirable.

Computer Editing

The use of computer editing to locate incorrectly recorded data errors in travel survey data has long been advocated by the Federal Highway Administration:

"Editing trip records -- In any analysis that uses the survey data, it is always necessary to edit the information to insure that the items of data have been correctly coded and punched. This is particularly important where a great deal of time and money is to be expended in the processing and analysis of these data. If the source data are not rigidly controlled and edited, it is possible that much useless analysis will result. To avoid these costly problems and to permit smooth processing in later phases of model development, all trip records should be edited using an appropriate edit program." (52, p. IV-3)

"In performing contingency checks on the basic trip file, all of the information, or fields, that will be used in subsequent

programs should be checked for illegal codes, illogical information, etc." (55, p. III-4)

Although all transportation studies have undoubtedly done some editing of their travel survey data, most such efforts have gone undocumented so very little is known concerning the thoroughness of editing and the quantity and types of errors found. One probable reason for this lack of documentation is a natural reluctance to admit that errors were made and to reveal how they were "fixed up" (57).

Some data on editing is available in technical memorandums and other hard-to-come-by documents. For example, the Indianapolis transportation study reported error rates (per cent of cards in error) of 8.4, 4.9, 14.5, and 3.5 per cent on their dwelling unit summary cards, internal trip report cards, external survey cards, and truck/taxi survey cards, respectively (26). Based on similar data from a few other studies, the results of a questionnaire, and personal contact with a number of practitioners, it can be concluded that it is not uncommon for those studies which conduct extensive computer edits to locate one or more errors in over 30 per cent of the dwelling unit samples, 15-20 per cent of the external survey data cards, and about 15 per cent of the truck/taxi data. It should be borne in mind, however, that the percentage of errors found in travel survey data fluctuates greatly depending upon the quantity of data contained in an individual record, the quality of the preceeding interviewing, coding, and digitizing processes, the number and extent of previous manual edits, and upon the number and types of computer edits to which the data are subjected.

There is some reason to suspect that less stringent editing than that normally assumed necessary might be sufficient. In a statistical investigation involving the 1953 Industrial Census in Norway, Nordbotten concluded that a small number of edit checks would suffice and that little

would be gained by more thorough editing (36). Based on an extension of such results, the U.S. Bureau of the Census and some other organizations have generally adopted procedures whereby they automatically either "impute" another value to replace one which an edit check shows to be in error or else substitute a similar record (or portion thereof) for the record containing the data in error (19,35,37,39,42,49,62). They note, however, that "the development has not yet reached the point, nor is it likely to do so in the immediate future, of using the computer as a substitute for a trained economic statistician in handling returns that show important problems,..." (62). At least one major transportation study in the United States has used such techniques in furnishing missing travel survey data items and perhaps for replacing invalid data detected through edits.*

Directional Split Assumptions

During the early sixties, transportation studies using the gravity model method of trip distribution normally computed areawide average directional splits for each home based trip purpose and applied those percentages when converting the triptables from P&A format to O&D format. The impetus for the above procedure was provided by the Bureau of Public Roads as evidenced by the following quotation:

"These (directional split) percentages must be either assumed or preferably arrived at by an analysis of the origin-destination survey data. This analysis is rather simple in theory but it requires a considerable amount of time. Essentially all of the trip records for each trip purpose or travel mode category are sorted to determine how many of them are produced by all zones of origin and how many by all zones of destination. The required percentages are then calculated directly. These percentages, which represent

*This transportation study refused permission to be cited in a bibliographical reference.

areawide averages, are then applied to each zone" (51, p. VI-3)

About 1964, some inhouse (and undocumented) research was conducted by Marks of the Bureau of Public Roads (30). Working with 1955 Washington, D.C. origin-destination data, Marks concluded that there was no significant difference between using an assumed 50-50 directional split and the computed percentage. Based on this research, the Bureau changed its unwritten policy to accepting assumed 50-50 directional splits, and this procedure has been used almost universally since that time. This change in thinking was reflected in the wording of the 1965 gravity model manual:

"In current practice a 50-50 split is usually assumed for all home based trips."
(52, p. VI-2)

Sensitivity of Travel Models to Errors

Relatively little of a quantitative nature is known above the sensitivity of the transportation planning process as a whole to differences in input data and the manner in which such differences or errors are propagated from model to model (e.g., from trip generation to trip distribution to traffic assignment). The only two substantive studies in this area are the CONSAD report (54) and NCHRP Report 120 (23).

Both reports agree that the traffic assignment model is the least sensitive of all the models to differences in input data. The CONSAD report further states that the most sensitive of the models is the trip distribution model in which "small errors ... were greatly magnified by the operations performed within that module."

An astute observation was made in the NCHRP report:

"An early conclusion was that data obtained by survey affect the strategic plan only indirectly. Forecasts intervene between the plan and the data at every step. Under these circumstances, plans are based directly on forecasts, not on data. And the errors of forecasts are apt to be much greater than errors of data."

CHAPTER IV BACKGROUND INFORMATION AND INITIAL DECISIONS

Study Area and Computer Battery Selections

In the selection of a study area for research purposes, prime consideration had to be given to the following:

1. Availability of unedited travel survey data.
2. Availability of original interview and coding forms.
3. Availability of volume and classification counts at the screenlines and at the cordon line.
4. Availability of a coded traffic assignment network.
5. Availability of necessary independent variable data for trip generation analyses.
6. Expected cooperativeness of the local people and/or consultant who were actually involved in the original study.

On the basis of the above considerations, Evansville, Indiana, was chosen for both the investigation of editing procedures and for the investigation of the effects of directional split assumptions. In addition, it was decided to utilize unedited travel survey data from the Lafayette, Indiana, study to provide a supplementary source of information about incorrectly recorded data error rates.

For various reasons, it was decided to use the Federal Highway Administration's IBM 360 battery of urban transportation planning computer programs (60,61). The computer used was the Indiana State Highway Commission's IBM 360 Model 50 at their headquarters office in Indianapolis.

The Evansville Regional Area Development
and Transportation Study

Evansville is located in southwestern Indiana bordering the Ohio River (the internal study area lies entirely on the Indiana side of the river). A population of approximately 175,000 people reside within the cordon line in a total of 292 survey zones. Bus ridership was negligible at the time of the surveys.

Vogt, Sage and Pflum, the consultant doing the transportation study, collected the travel survey data between October of 1969 and June of 1970. Reductions of the interviewing/coding forms used in the various surveys are included in Appendix A and examples of the geographic coding indices are included in Appendix B.

In the actual study, there is a large area called the COG (Council of Government) area lying between the cordon line referred to above and an "outer cordon." While the consultant is obligated to do some special analyses within this area, no home interview data were collected from this area and all external survey interview stations were located at the "inner" cordon. Therefore, this COG area has been treated as any other external area for the purposes of this research project.

The consultant conducted a special roadside interview study at Dress Memorial Airport. However, neither this research project nor the consultant utilized this additional travel survey information to supplement the data available from the other surveys.

Unfortunately, extremely poor weather existed during a substantial portion of the period during which the home interview surveys were conducted (they commenced in January and ran through the first week of April, 1970). As a result, many residents held their travel to a minimum and thus the average daily traffic (ADT) during the survey period was significantly less than the annual average daily

traffic (AADT). In addition, there were major problems obtaining accurate volume counts during the period due primarily to the failure of the mechanical equipment to perform properly.

In the 10 per cent home interview sample, trip logs were sent in advance to the selected dwelling units. After the travel date, the interviewer went and obtained all required information. Both the interviewer and the interviewer's supervisor checked the finished forms for completeness. They were then coded for keypunching, and the coding was subsequently checked dependently by another coder. Finally, a supervisor checked the coded forms for completeness and continuity. There was a telephone call-back procedure for verifying the data turned in by the interviewers, and various statistics (e.g., trips per dwelling unit) were kept on the data submitted by the individual interviewers for comparison against standards. As a result of these checks, two interviewers were fired.

External survey data were coded on mark sense forms for digitization by the Indiana State Highway Commission mark sense reader, and the same forms were used for field interviewing. The field supervisor spot checked the completed forms as they were submitted, and they were subsequently coded in the office. Finally, the coded forms were checked by a supervisor.

In the 100 per cent sample taxi survey, one interviewer copied all of the interview data from the vehicle logs and then coded the data. The same tasks were performed by several different interviewers in the 20 per cent truck sample, with different persons coding the data than collected it. After coding, the forms from both surveys were checked dependently by other coders. Finally, a supervisor checked the forms for completeness and continuity.

The Greater Lafayette Area Transportation
and Development Study

The twin cities of Lafayette and West Lafayette are located in north central Indiana about one third of the way from Indianapolis to Chicago. A population of approximately 90,000 people reside within the cordon line in a total of 141 zones (1,611 subzones). Transit ridership is minimal.

Local personnel administered the collection of the travel survey data, and the surveys were conducted between September of 1970 and January of 1971. Reductions of the interviewing/coding forms used in the various surveys are included in Appendix A and examples of the geographic coding indices are included in Appendix B.

In the 12.5 per cent home interview sample, trip logs were delivered in advance to the selected dwelling units. After the travel date, the interviewer returned and obtained all required information. Both the interviewer and the interviewer's supervisor checked the finished forms for completeness. They were then coded for keypunching, and the coding was subsequently checked dependently by another coder. There was a 10 per cent telephone call-back procedure for verifying the data turned in by the interviewers, and statistics were kept on trips per auto and trips per dwelling unit as obtained by each interviewer for comparison against standards.

Special field interview forms were used at the roadside interview stations, and the field supervisor spot checked the completed forms as they were submitted. The trip end addresses were coded to zones directly on the interview form, and each of these codes was dependently checked by another coder. All of the data from the interview forms was then coded onto mark sense forms for digitization by the Indiana State Highway Commission mark sense reader. A random sample of the completed scan sheets was checked dependently by another coder.

In the taxi survey, one interviewer copied all of the interview data from the vehicle logs and then a second person coded the data. The same tasks were performed by several different interviewers and several different coders in the truck survey. All coding was subsequently checked dependently by another coder, and a telephone call-back procedure was used for verifying the data submitted by the interviewers in the truck survey.

Selection of Computer Editing Procedures

There are two steps to a trip card computer editing procedure, error checking and error rectification. In the first step, the data are subjected to programmed tests to insure their validity and reasonableness. These tests can be individually categorized as either single field or consistency. In the single field tests, each data field is examined by itself without regard to the content of any other data field in an attempt to determine whether its contents are possibly valid.

In the consistency checks, the contents of a given data field are considered in relation to the contents of one or more other data fields (in the same trip card or in a related trip card) to insure that they are "consistent." Thus, a trip start time of 9 AM and a corresponding trip arrive time of 8 AM would both pass the single field checks but would be rejected by a consistency check comparing trip start time to arrive time. Consistency checks are much more difficult to program than single field checks, so there is a tendency toward making only a minimal number of consistency checks. Frequently, there are many error checks which could be made which would locate some errors but would also needlessly reject valid data (e.g., a test for an "unreasonably" long trip in terms of time might needlessly reject a long distance trip made over extremely icy roads). Those entries marked by an asterisk in Table C3 fall in this category.

Because of the difficulty of handling correct data which are needlessly rejected, these potential error checks are frequently not made.

The second step of an edit procedure is error rectification. Given that there is an error, what do you do next? There are three major alternatives: correct the data in error, discard the data in error, or arbitrarily replace the bad data with some "good" data. It is usually expected that the most accurate results are obtained by correcting the data in error. Normally, this can be done by simply referring back to the original interview data and perhaps making some "reasonable" assumptions. In some cases, this might necessitate a return call to the original interviewees.

If it is decided to discard the data, there are several significant questions to be considered. Trip data from the home interview, truck, and taxi surveys are cluster sampled, and there are ramifications relative to retaining a representative sample if some or all of such data are discarded. If the whole sample is not discarded in its entirety, appropriate "correction" factors would have to be applied to the corresponding dwelling unit expansion factors in deriving the trip expansion factors to account for the fact that some individual trips (or all trips made by some persons) have been discarded. These particular considerations are not relevant for external survey data where it is the trip itself that is being sampled rather than the dwelling unit or vehicle, but one still must be concerned about maintaining a representative number of interviews by hour by direction by vehicle type at the given station.

The third alternative, that of arbitrarily replacing the bad data with "good" data, is the technique which, as described in the Review of Literature, is being used more and more by census organizations. With the exception of the one mentioned transportation study, there have been no known efforts to implement such methodology in the computer

processing of travel survey data. However, it would not be surprising to find that such techniques have been manually applied in scattered instances by some studies when it was considered expedient to do so.

For purposes of this research project, three editing procedures were investigated. Procedure A consisted of using all of the travel survey data without editing it at all. While this procedure eliminated the entire editing phase, it was recognized that there would be substantial difficulties in getting some of the computer programs to work satisfactorily with input data containing such things as blank or invalid zone numbers.

Procedure C consisted of making a quite sizable number of error checks (see the lists of error checks in Appendix C), correcting the errors if at all possible using "reasonable" judgement when necessary, and discarding in their entirety those samples containing data errors which could not be corrected. Assuming that the number of samples which would have to be discarded is minimal (this turned out to be the case), it was believed that one could a priori say that procedure C is the best editing procedure in terms of getting the most accurate results.

There was one deviation from the stated method of correcting all data in error, if possible. After investigation into the effects of simply deleting external survey trip cards containing errors as contrasted to correcting the errors, it was determined that the resulting triptable matrices would show little difference and thus the external survey trip cards containing errors were simply deleted (it should be remembered that the external survey is the only survey in which the trip data are not cluster sampled).

Procedure B data were derived by making a quite sizable number of error checks (the same error checks used in procedure C) and simply discarding in their entirety all samples containing errors. The expansion factors for the

remaining data were obtained by simply factoring the expansion factors derived in the normal manner for the entire survey data (i.e., procedure A data) by the ratio of the original number of samples to the remaining number of samples.

In both procedure B and procedure C, error checks were made on all data fields, not just those containing data used in the planning models investigated in this research project. Similarly, data records in error were considered correctable only if all data fields containing errors could be corrected.

CHAPTER V DATA EDITING, EXPANSION, AND FACTORING

Data Editing

The basic data available to this project consisted of punched but unedited travel survey data cards from both Evansville and Lafayette. The data from the home interview and truck/taxi surveys had been keypunched and key verified, while the data from the external surveys had been punched by the Indiana State Highway Commission mark sense reader from the mark sense forms coded by the two coding staffs. In order to successfully process the scan sheets, it was necessary to "fix-up" about 6 per cent of the Evansville forms and about 2 per cent of the Lafayette forms. Such action was necessary in order to replace mutilated forms, eliminate handwriting that carelessly protruded into the coding areas, or to completely erase coding marks which the coder had subsequently changed but failed to erase properly.

All of the travel survey data from both studies were subjected to a fairly comprehensive error checking process on the computer. The error checks used and the results of their application are documented in Appendix C with some of the more significant data summarized in Table 1. As previously planned, the analysis for this research from this point to completion made use of the Evansville data exclusively.

Effective with the completion of the error checking process, data corresponding to procedures A and B as discussed in the last section of Chapter IV were available. Using "reasonable" judgement where necessary, the travel

Table 1. Summary Statistics From Error Checking Process

Statistic	Survey*			
	Home Interview	External	Truck	Taxi
Percentage of Cards Having Errors	9.52	11.33	19.06	16.87
	38.41	7.07	19.07	42.77
Percentage of Samples Having Errors	29.92	---	19.37	44.74
	63.06	---	38.32	100.00
Average Number of Cards per Sample for those samples having errors	13.08	---	16.02	41.65
	11.40	---	4.40	43.25
Average Number of Cards per Sample for those samples without errors	6.41	---	5.12	13.43
	3.98	---	3.39	---

*In each cell, the data on the top line is from the Evansville Study and the data on the bottom line is from the Lafayette Study.

survey data in error were corrected and combined with the "clean" records of procedure B, thus creating a set of edited travel survey data referred to as procedure C data in this project.

The travel survey data were not linked.

Development of Expansion Factors

To account for the fact that travel survey data are collected for only a portion of the dwelling units, vehicles crossing the cordon line, trucks, and taxis, it is necessary to compute expansion factors for the trip data in order to obtain a representation of all the trips made in the study area. There are standard equations for computing these expansion factors (53).

These standard procedures were used in computing the expansion factors for the procedure A and procedure B external survey data (as explained in Chapter IV, procedure C external survey data were the same as procedure B external survey data) and for the procedure A and procedure C truck and taxi survey data. However, a special formula was used for computing the dwelling unit expansion factors for procedure A and procedure C data (the formula used was developed by the consultant performing the Evansville study and approved by the Federal Highway Administration):

$$\text{FACTOR} = \frac{10(B+S-C-E)}{B+S+X-C-D}$$

where:

- B = Number of samples initially selected
- C = Number of samples vacant, demolished, or commercial
- D = Number of samples which were refusals, no one home, or sickness
- S = Number of "supplemental" samples (these were taken when the sample address contained more households than expected)
- E = Number of "change of address" samples (these were taken when the sample address did not exist in the field).

X = Number of "extra" samples (these were taken when the number of samples selected in a given zone were less than the desired one in ten)

Expansion factors for the Procedure B home interview, truck, and taxi survey data were computed by multiplying the corresponding procedure A expansion factor by the ratio of the original number of samples to the number of samples remaining after rejecting all samples containing errors.

Survey Accuracy Checks

After expansion factors have been computed and inserted in the trip cards, it is necessary to make various standard comparisons of the expanded data from the travel surveys and corresponding estimates derived from independent data sources. These comparisons are called survey accuracy checks, and are normally grouped into dwelling unit comparisons and trip data comparisons.

For all practical purposes, identical estimates of population, total dwelling units, etc. were obtained from the expanded procedure A and procedure C data. The estimates obtained from the procedure B data did differ from those obtained from the other procedures (see Table 2), but most analysts would have accepted the results from any of the three procedures.

Table 2. Dwelling Unit Characteristics As Related To Editing Procedure

Statistic	Procedure		
	A	B	C
Persons per dwelling unit	3.04	2.76	3.02
Cars per dwelling unit	1.36	1.24	1.37
Average income per dwelling unit	\$7,810	\$7,180	\$7,820

Significant problems with the procedure B data became obvious when the trip data comparisons were made. For example, the first work trip comparison was not good for

procedure B data while it was completely satisfactory for the data from procedures A and C. The major problem, however, was that the composition of traffic showed way too many truck trips and way too few automobile trips (this occurred because of the characteristics of the rejected travel survey data as summarized in Table 1).

The primary trip data comparison is the screenline crossing analysis in which the number of survey trips crossing arbitrary lines called screenlines bisecting the area is compared against actual ground counts at the various locations. There were four screenlines in Evansville.

A tabulation of survey screenline crossings vs. ground counts by procedure is presented in Table 3. Each entry represents the sum of all crossings for all four screenlines for all vehicles for the given hour. The tabulated survey crossing data have been factored to reflect the effect of double crossings, and the estimated time of screenline crossing has been based on trip start and arrive time and the relative distance of each of the two zones from the screenline.

After examining the results of the survey accuracy checks, a prudent analyst would probably have questioned proceeding using the procedure B data. However, it was decided that, in fulfilling the objectives of this research project, it would be advantageous to continue using the data.

Development and Application of Trip Factors

It has long been recognized that there is substantial underreporting of non-work trips, and it is customary to factor the expanded trip data to compensate for the missing trips. Using prudent, justifiable logic, factors are developed by trip purpose with the objective of attempting to approximate the actual ground counts at the screenlines.

Table 3. Total Screenline Crossings As Obtained From Expanded But Unfactored Survey Data

Hour	Procedure			Ground Count
	A	B	C	
6-7 AM	18,586	17,771	18,351	18,476
7-8	33,426	30,349	33,600	34,338
8-9	20,347	17,717	20,283	21,700
9-10	15,538	14,671	15,524	19,740
10-11	16,286	16,868	16,311	21,416
11 - NOON	16,995	15,648	16,889	23,838
NOON - 1 PM	19,936	16,100	19,948	25,152
1-2	17,882	17,371	17,984	24,302
2-3	22,458	21,132	22,582	27,804
3-4	32,783	30,591	32,636	38,814
4-5	33,469	30,960	33,622	38,572
5-6	32,173	29,819	32,361	34,596
6-7	18,860	16,300	18,877	24,756
7-8	14,252	11,450	14,390	18,918
8-9	10,989	9,660	10,985	15,698
9-10 PM	10,062	8,328	10,196	13,758
TOTAL	334,042	304,735	334,539	401,878

Table 4. Trip Factors To Compensate For Underreporting

Trip Purpose	Hours	Procedure		
		A	B	C
HBW	all hours	1.00	1.10	1.00
HBS & HBO	6-8 AM	1.00	1.00	1.00
	rest of day	1.34	1.80	1.34
NHB	6-8 AM	1.00	1.00	1.00
	rest of day	1.53	2.10	1.53
Trucks	6 AM - 3 PM	1.31	1.00	1.31
	rest of day	1.96	1.00	1.96
All Others	all hours	1.00	1.00	1.00

Such an approach was followed in this research project, and the resulting factors by purpose are given in Table 4. These factors were based on an assumption that people tend to remember their first trip of the day better than subsequent trips and that, due to the high sampling rate, external and taxi trips could not be justifiably factored. Because the trip data from procedures A and C as tabulated in Table 3 were, for all practical purposes, identical, it was felt that the use of different factors for those two sets of data might introduce differences not really attributable to the actual differences in the original data. Therefore, as shown in Table 4, the same factors were used for both procedures A and C.

The effects on the screenline crossings of applying these factors are shown in Table 5. The resulting number of trips, as compared to the ground counts, is perhaps slightly higher than one would like, but since all three procedures gave results which were slightly high, it was felt that these factored data would be suitable for their intended use in this research project.

Table 6 shows the composition of total vehicle trips by mode by procedure and the effect of the factoring process on the composition.

Table 5. Total Screenline Crossings As Obtained From Expanded And Factored Survey Data

Hour	Procedure			Ground Count
	A	B	C	
6-7 AM	18,752	18,672	18,517	18,476
7-8	34,220	32,199	34,396	34,338
8-9	23,874	22,116	23,799	21,700
9-10	18,975	19,589	18,966	19,740
10-11	20,398	23,020	20,444	21,416
11 - NOON	21,833	22,677	21,719	23,838
NOON - 1 PM	25,784	24,046	25,791	25,152
1-2	22,616	24,512	22,721	24,302
2-3	27,876	29,499	28,057	27,804
3-4	40,584	41,748	40,463	38,814
4-5	40,928	40,621	41,103	38,572
5-6	38,007	39,454	38,221	34,596
6-7	23,247	23,710	23,285	24,756
7-8	18,303	17,638	18,477	18,918
8-9	13,956	14,836	13,937	15,698
9-10 PM	12,522	12,361	12,670	13,758
TOTAL	401,875	406,698	402,566	401,878

Table 5. Total Screenline Crossings As Obtained From Expanded And Factored Survey Data

Hour	Procedure			Ground Count
	A	B	C	
6-7 AM	18,752	18,672	18,517	18,476
7-8	34,220	32,199	34,396	34,338
8-9	23,874	22,116	23,799	21,700
9-10	18,975	19,589	18,966	19,740
10-11	20,398	23,020	20,444	21,416
11 - NOON	21,833	22,677	21,719	23,838
NOON - 1 PM	25,784	24,046	25,791	25,152
1-2	22,616	24,512	22,721	24,302
2-3	27,876	29,499	28,057	27,804
3-4	40,584	41,748	40,463	38,814
4-5	40,928	40,621	41,103	38,572
5-6	38,007	39,454	38,221	34,596
6-7	23,247	23,710	23,285	24,756
7-8	18,303	17,638	18,477	18,918
8-9	13,956	14,836	13,937	15,698
9-10 PM	12,522	12,361	12,670	13,758
TOTAL	401,875	406,698	402,566	401,878

Table 6. Percentage of Total Vehicle Trips By Purpose

	Before Factoring			After Factoring		
	Procedure			Procedure		
	A	B	C	A	B	C
Home Based Work (HBW)	18.15	16.05	18.07	14.40	12.73	14.33
Home Based Shop (HBS)	10.24	8.43	10.20	10.85	10.88	10.81
Home Based Other (HBO)	24.66	18.53	24.43	25.43	22.88	25.18
Non-Home Based (NHB)	21.57	16.60	21.84	25.65	24.38	25.97
Taxi	0.24	0.40	0.24	0.19	0.29	0.19
Trucks (From Truck Survey)	9.38	23.36	9.31	10.98	16.84	10.91
External Trucks	4.04	4.30	4.11	3.20	3.10	3.26
External Automobiles	11.73	12.34	11.80	9.30	8.90	9.36

CHAPTER VI TRAFFIC ASSIGNMENT

O&D Triptables

The basic input to the traffic assignment process is total purpose triptables, supposedly in O&D format, to be subsequently "loaded" onto the network. Using the data from each of the three procedures being investigated, three total purpose, O&D, ADT, vehicle triptables were built.

Since it was a priori stated that procedure C is the best in terms of accuracy of results, the triptables corresponding to procedures A and B were compared to that corresponding to procedure C. The results of these comparisons are given in Tables 7 and 8, respectively. These comparisons gave per cent root mean square error rates, the standard measure of error used in transportation planning comparisons of triptables and network loads, of 14.89 and 111.02, respectively.

Network Loads

The Evansville traffic assignment network consists of 314 zones (292 internal zones and 22 external stations), 3943 directional links, and a last node number of 2110. Each of the O&D triptables corresponding to procedures A, B, and C were loaded onto the network using vines built on the basis of observed travel time. The results of the unrestrained loading process are tabulated in Table 9, and the directional link loads are compared in Tables 10 and 11.

The per cent root mean square error of load A compared to load C is 1.53 with a maximum directional ADT volume difference of 216. The corresponding statistics for load B

TABLE 7. COMPARISON OF PROCEDURE A TO PROCEDURE C TRIPTABLES

COMPARISON OF TOT. PURPOSE VIEW. TRIPTABLES. BASE-PROCEDURE C. TEST-PROC. A

PAGE 14

TABLE 2 - MISCELLANEOUS STATISTICS
INCLUDES ALL SELECTED CELLS EXCEPT THOSE WITH BOTH ZERO BASE AND ZERO TEST VOLUMES
TOTALS FOR ALL VOLUME GROUPS

TOTAL TEST TRIPTABLES121*	519,435	MEAN TEST VOLUME*	23.20	ROOT MEAN SQUARE ERROR*	3.4437974
TOTAL BASE TRIPTABLES111*	520,990	MEAN BASE VOLUME*	23.27	PER CENT R.M.S. ERROR*	14.885074
SUM OF DIFFERENCES	-1,555	MEAN DIFFERENCE*	-0.07	VARIANCE	11.993612
SUM OF SQUARES	268633.00	NUMBER OF CELLS*	22,390	STANDARD DEVIATION*	3.4631786

RELATION OF TEST VOLUMES TO CORRESPONDING BASE VOLUMES				
NUMBER OF CELLS ON TOP LINE, SUM OF TEST VOLUMES ON MIDDLE LINE, SUM OF BASE VOLUMES ON BOTTOM LINE				
TEST VOLUME				
ZERO				
BASE				
VOLUME				
TEST EQUAL TO BASE * TEST NOT EQUAL TO BASE * TEST NOT EQUAL TO BASE*				
NON-ZERO				
TOTAL				
0/ 0.0 %	0/ 0.0 %	0/ 0.0 %	450/ 2.01 %	450/ 2.01 %
0/ 0.0 %	0/ 0.0 %	0/ 0.0 %	2,221/ 0.43 %	2,221/ 0.43 %
0/ 0.0 %	0/ 0.0 %	0/ 0.0 %	0/ 0.0 %	0/ 0.0 %

RELATION OF TEST VOLUMES TO CORRESPONDING BASE VOLUMES				
NUMBER OF CELLS ON TOP LINE, SUM OF TEST VOLUMES ON MIDDLE LINE, SUM OF BASE VOLUMES ON BOTTOM LINE				
TEST VOLUME				
ZERO				
BASE				
VOLUME				
TEST EQUAL TO BASE * TEST NOT EQUAL TO BASE * TEST NOT EQUAL TO BASE*				
NON-ZERO				
TOTAL				
0/ 0.0 %	174/ 0.78 %	13,674/ 61.07 %	8,092/ 36.14 %	21,940/ 97.99 %
0/ 0.0 %	0/ 0.0 %	301,911/ 58.12 %	215,303/ 41.85 %	517,214/ 99.97 %
0/ 0.0 %	1,531/ 0.29 %	301,911/ 57.95 %	217,548/ 41.76 %	520,460/ 100.00 %

RELATION OF TEST VOLUMES TO CORRESPONDING BASE VOLUMES				
NUMBER OF CELLS ON TOP LINE, SUM OF TEST VOLUMES ON MIDDLE LINE, SUM OF BASE VOLUMES ON BOTTOM LINE				
TEST VOLUME				
ZERO				
BASE				
VOLUME				
TEST EQUAL TO BASE * TEST NOT EQUAL TO BASE * TEST NOT EQUAL TO BASE*				
NON-ZERO				
TOTAL				
0/ 0.0 %	174/ 0.78 %	13,674/ 61.07 %	8,092/ 36.14 %	21,940/ 97.99 %
0/ 0.0 %	0/ 0.0 %	301,911/ 58.12 %	215,303/ 41.85 %	517,214/ 99.97 %
0/ 0.0 %	1,531/ 0.29 %	301,911/ 57.95 %	217,548/ 41.76 %	520,460/ 100.00 %

BREAK-DOWN BY RELATIVE MAGNITUDE OF THOSE CELLS WITH BOTH BASE AND TEST VOLUMES GREATER THAN ZERO AND NOT EQUAL

RELATIVE MAGNITUDE				
TEST GREATER THAN BASE				
TEST LESS THAN BASE				
TOTAL				
3,619/ 16.16 %	95,091/ 18.30 %	85,914/ 16.49 %		
4,473/ 19.98 %	120,252/ 23.15 %	111,634/ 25.27 %		
8,092/ 36.14 %	215,303/ 41.45 %	217,548/ 41.76 %		

MINIMUM AND MAXIMUM INDIVIDUAL CELL VOLUMES

MINIMUM AND MAXIMUM INDIVIDUAL CELL VOLUMES				
BASE				
TEST				
VOLUME				
ZERO VOLUME				
MINIMUM NON-ZERO VOLUME				
MAXIMUM VOLUME				
0	314-3079	0	312-3114	
1	1-2979	1	1-3069	
1,267	147-313	1,273	147-313	

EXTREME RELATIONSHIPS BETWEEN CORRESPONDING BASE AND TEST VOLUMES

EXTREME RELATIONSHIPS BETWEEN CORRESPONDING BASE AND TEST VOLUMES				
VOLUME DIFFERENCE				
PER CENT DEVIATION				
TEST-BASE				
VOLUME				
MAXIMUM NEGATIVE				
MAXIMUM POSITIVE				
-95	175-175	-100.00 %	CELL	
96	314-304	166.67 %	208-304	

* AS WELL AS OTHER CELLS

Table 9. Network Loading Statistics

Statistic		Procedure		
		A	B	C
Trips Loaded		482,027	495,103	483,443
Vehicle Miles	Local Links	93,545	96,856	93,754
	Other Links	1,811,593	1,821,181	1,816,706
Vehicle Hours	Local Links	7,501	7,747	7,514
	Other Links	57,755	58,215	57,902

compared to load C are 7.91 and 903.

No restrained loads were made because of concern for the accuracy of the link count data (see Chapter IV) and fear that the effects of a restraint process might confound the results directly attributable to differences in the travel survey data.

TABLE 10. COMPARISON OF PROCEDURE A TO PROCEDURE C LOADS

COMPARISON OF DIRECTIONAL LINK VOLUMES, BASE=PROCEDURE C, TEST=PROCEDURE A

PAGE 33

TABLE 2 - MISCELLANEOUS STATISTICS
INCLUDES ALL SELECTED LINKS EXCEPT THOSE WITH BOTH ZERO BASE AND ZERO TEST VOLUMES
TOTALS FOR ALL VOLUME GROUPS

TOTAL TEST TRIPSHR11	1=	8,376,173	MEAN TEST VOLUME=	2297.12	ROOT MEAN SQUARE ERROR=	34,542282
TOTAL BASE TRIPSHR11	1=	8,398,302	MEAN BASE VOLUME=	2263.08	RTM CENT A.W.S. ERROR=	1,5263367
SUM OF DIFFERENCES	=	-22,129	MEAN DIFFERENCE=	-5.06	VARIANCE	1197,9233
SUM OF SQUARES	=	4427853.0	NUMBER OF LINKS =	5,711	STANDARD DEVIATION =	34,029259

RELATION OF TEST VOLUMES TO CORRESPONDING BASE VOLUMES

NUMBER OF LINKS ON TOP LINE, SUM OF TEST VOLUMES ON MIDDLE LINE, SUM OF BASE VOLUMES ON BOTTOM LINE	TEST VOLUME		NON-ZERO		TOTAL
	TEST EQUAL TO BASE	TEST NOT EQUAL TO BASE	TEST EQUAL TO BASE	TEST NOT EQUAL TO BASE	
BASE					
NON-ZERO					
ZERO					
TOTAL					

BREAK-DOWN BY RELATIVE MAGNITUDE OF THOSE LINKS WITH BOTH BASE AND TEST VOLUMES GREATER THAN ZERO AND NOT EQUAL

RELATIVE MAGNITUDE

TEST GREATER THAN BASE

TEST LESS THAN BASE

TOTAL

MINIMUM AND MAXIMUM INDIVIDUAL LINK VOLUMES

TEST	VOLUME	LINK	VOLUME	LINK
ZERO VOLUME	0	1545-1544		
MINIMUM NON-ZERO VOLUME	1	678-680*		
MAXIMUM VOLUME	20,547	1200-1201	20,553	678-680*

EXTREME RELATIONSHIPS BETWEEN CORRESPONDING BASE AND TEST VOLUMES

VOLUME	TEST-BASE	PER CENT DEVIATION	LINK
-2795	913-916	-50.00 %	1440-1427
216	1201-1101	108.33 %	1661-1660

* AS WELL AS OTHER LINKS

TABLE II. COMPARISON OF PROCEDURE B TO PROCEDURE C LOADS

COMPARISON OF DIRECTIONAL LINK VOLUMES, BASE=PROCEDURE C, TEST=PROCEDURE B

PAGE 39

TABLE P - MISCELLANEOUS STATISTICS
INCLUDES ALL SELECTED LINKS EXCEPT THOSE WITH BOTH ZERO BASE AND ZERO TEST VOLUMES
TOTALS FOR ALL VOLUME GROUPS

TOTAL TEST TRIPS/HRI	1 = 8,483,573	MEAN TEST VOLUME =	2236.5W	RNDY MEAN SQUARE ERROR =	179,07329
TOTAL BASE TRIPS/HRI	1 = 8,308,302	MEAN BASE VOLUME =	2262.59	PER CENT R.M.S. ERROR =	7.9106686
SUM OF DIFFERENCES	1 = 85,273	MEAN DIFFERENCE =	22.98	VARIANCE =	31567.677
SUM OF SQUARES	1 = 0.118969470 09	NUMBER OF LINKS =	5,910	STANDARD DEVIATION =	177.61609

RELATION OF TEST VOLUMES TO CORRESPONDING BASE VOLUMES

NUMBER OF LINKS ON TOP LINE, SUM OF TEST VOLUMES ON MIDDLE LINE, SUM OF BASE VOLUMES ON BOTTOM LINE	TEST VOLUME		NON-ZERO		TOTAL
	TEST EQUAL TO BASE	TEST NOT EQUAL TO BASE	TEST EQUAL TO BASE	TEST NOT EQUAL TO BASE	
BASE VOLUME					
ZERO	0/ 0.0 %	0/ 0.0 %	0/ 0.0 %	0/ 0.0 %	0/ 0.0 %
NON-ZERO	32/ 0.86 %	117/ 3.15 %	100,596/ 1.19 %	8,382,977/ 98.81 %	8,382,977/ 98.81 %
TOTAL	32/ 0.86 %	117/ 3.15 %	100,596/ 1.20 %	8,297,033/ 98.79 %	8,382,977/ 98.81 %

BREAK-DOWN BY RELATIVE MAGNITUDE OF THOSE LINKS WITH BOTH BASE AND TEST VOLUMES GREATER THAN ZERO AND NOT EQUAL

RELATIVE MAGNITUDE	NUMBER OF LINKS	SUM OF TEST VOLUMES	SUM OF BASE VOLUMES
TEST GREATER THAN BASE	2,004/ 54.02 %	4,790,265/ 56.87 %	4,827,019/ 57.90 %
TEST LESS THAN BASE	1,557/ 41.97 %	3,592,612/ 42.35 %	3,770,019/ 44.89 %
TOTAL	3,561/ 95.98 %	8,382,877/ 99.22 %	8,297,033/ 98.79 %

MINIMUM AND MAXIMUM INDIVIDUAL LINK VOLUMES

	VOLUME	LINK	TEST VOLUME	LINK
ZERO VOLUME				
MINIMUM NON-ZERO VOLUME	1	678-680*	0	210A-1734*
MAXIMUM VOLUME	20,347	1200-1201	20,103	87-2059*

EXTREME RELATIONSHIPS BETWEEN CORRESPONDING BASE AND TEST VOLUMES

	VOLUME	LINK	PER CENT DEVIATION
MAXIMUM NEGATIVE	-884	1377-1378*	100*(TEST-BASE)/BASE
MAXIMUM POSITIVE	903	507-508*	PER CENT
			-100.00 %
			300.00 %

* AS WELL AS OTHER LINKS

CHAPTER VII TRIP DISTRIBUTION

General

Using base year data, it is necessary to calibrate a trip distribution model for subsequent distribution of forecast year trips. Because of the predominance of its use in distributing internal-internal trips, the gravity model was selected for application in this research project.

The objectives of this research project did not dictate the calibration of the trip distribution model for all purposes but rather called for an investigation to determine if different results of the calibration process would occur if the different sets of edited travel survey data were used as input. This investigation was restricted to the home based work and home based shopping purposes.

In preparation for applying the gravity model, the impedance dataset (commonly called "skimmed trees") was updated to account for intrazonal travel time and terminal time. P&A purpose triptables (one dataset for each of the three editing procedures being investigated) were built, and the required P&A cards punched. Using the updated impedance dataset and the three triptable datasets, the observed trip length distributions by purpose were obtained for each of the three editing procedures.

A manual comparison of the P&A data and the trip length distributions for each of the two purposes showed very little differences between corresponding data items from the procedure A and the procedure C data. There were differences between the procedure B and the procedure C data, but these appeared to be somewhat random in nature.

Results

For each of the three editing procedures and for each of the two purposes (i.e., a total of six times), the gravity model was applied once using three iterations to balance attractions. A single set of "reasonable" friction factors was used for all three home based work distributions, and a second set of "reasonable" friction factors was used for all three home based shopping distributions. The resulting effects on the trip length distributions are summarized in Tables 12 and 13, and the manually computed ratios of the number of observed trips to the number of distributed trips are summarized in Table 14.

Table 12. Home Based Work Trip Length Distribution Summary Statistics

Trips	Statistic	Procedure		
		A	B	C
Observed	Total Trips	74,794	69,729	74,664
	Total Trip Hours	16,303	15,289	16,265
	Average Trip Length (Minutes)	13.08	13.16	13.07
	Standard Deviation	5.49	5.42	5.49
	Total Intra Trips	1,466	1,425	1,477
Distributed	Total Trips	74,794	69,729	74,664
	Total Trip Hours	15,805	14,863	15,781
	Average Trip Length (Minutes)	12.68	12.79	12.68
	Standard Deviation	5.38	5.43	5.38
	Total Intra Trips	1,063	1,003	1,082

Table 13. Home Based Shopping Trip Length Distribution Summary Statistics

Trips	Statistic	Procedure		
		A	B	C
Observed	Total Trips	56,333	59,593	56,292
	Total Trip Hours	8,585	8,975	8,542
	Average Trip Length (Minutes)	9.14	9.04	9.11
	Standard Deviation	4.31	4.33	4.28
	Total Intra Trips	4,261	4,480	4,377
Distributed	Total Trips	56,333	59,593	56,292
	Total Trip Hours	8,700	9,258	8,674
	Average Trip Length (Minutes)	9.27	9.32	9.25
	Standard Deviation	4.42	4.45	4.42
	Total Intra Trips	3,432	3,467	3,495

Table 14. Ratio Of Observed To Distributed Trips By Time Increment

Time	Home Based Work Procedure			Home Based Shopping Procedure		
	A	B	C	A	B	C
1	--	--	--	--	--	--
2	--	--	--	--	--	--
3	0.88	0.75	0.86	1.24	1.28	1.25
4	1.15	1.19	1.13	0.97	1.09	0.98
5	1.05	0.96	1.06	0.90	0.90	0.90
6	0.70	0.74	0.71	0.88	1.02	0.87
7	0.92	0.92	0.94	1.21	1.18	1.22
8	0.79	0.73	0.79	0.95	0.97	0.95
9	0.94	1.02	0.93	0.98	0.89	0.97
10	0.94	0.91	0.95	0.97	0.89	0.98
11	1.01	1.06	1.02	0.97	0.97	0.97
12	0.97	0.97	0.97	1.13	1.10	1.13
13	0.97	0.87	0.95	1.01	1.19	1.01
14	1.10	1.16	1.09	0.92	1.00	0.92
15	1.07	1.06	1.07	0.83	0.65	0.85
16	1.25	1.30	1.26	0.86	0.62	0.85
17	1.15	1.23	1.16	0.89	0.85	0.86
18	1.00	0.98	1.00	1.53	1.53	1.55
19	1.11	1.07	1.10	1.53	1.09	1.53
20	1.20	1.09	1.20	1.00	0.79	0.95
21	0.96	1.09	0.97	0.63	0.34	0.67
22	1.09	1.01	1.05	0.49	0.66	0.48
23	0.93	1.08	0.89	0.68	0.69	0.71
24	1.23	1.17	1.24	1.39	1.53	1.23
25	0.97	0.98	0.98	0.62	0.96	0.61
26	0.94	0.80	0.98	0.44	0.27	0.45
27	0.90	0.64	0.86	1.33	1.02	0.97
28	0.79	0.75	0.75	0.39	0.72	0.41
29	0.90	1.07	0.99	0.00	0.00	0.00
30	0.84	0.75	0.83	1.08	2.11	1.08
31	1.27	1.66	1.22	1.41	1.64	1.41
32	1.28	0.86	1.33	1.73	3.28	1.63
33	1.40	0.93	1.40	0.00	0.00	0.00
34	1.16	1.58	1.08	0.00	0.00	0.00
35	1.75	1.74	1.64	0.00	0.00	0.00
36	3.65	2.51	3.75	0.00	0.00	0.00
37	1.28	0.76	1.28	--	--	--
38	1.67	1.00	1.74	--	0.00	--
39	0.60	0.00	0.56	--	0.00	--
40	0.95	1.52	0.91	--	--	--
41	2.94	2.64	2.94	--	--	--
42	--	0.00	--	--	--	--
43	0.00	0.00	0.00	--	--	--
44	0.00	0.00	0.00	--	--	--
45	--	--	--	--	--	--

CHAPTER VIII TRIP GENERATION

General

Using base year data, it is necessary to develop mathematical models which will predict the total number of trips produced by or attracted to a zone. Traditionally, multiple regression models with the individual observation being either a zonal total or a zonal average have been used, so this was the method selected for application and investigation in this research project.

It should be borne in mind that the objectives of this research project could be satisfied simply by making the determination that the results of the trip generation model development phase using the different sets of data corresponding to editing procedures A, B, and C would be either reasonably identical or else significantly different. It was not necessary to do all of the analyses normally done in arriving at final equations, nor was it necessary to develop all of the customary equations. The equations selected for investigation were home based work productions, home based work attractions, home based shopping productions, and home based shopping attractions.

The proportion of the total trips within the study area made for the purposes of home based work and home based shopping were given in Table 6. Because the survey trips were not linked, the proportion of home based work trips is slightly less than what studies with linked trips report.

In developing the four equations for the two purposes under investigation, some of the data were discarded before the analysis in order not to bias the results. For the

home based production equations, data from 88 zones were discarded leaving data from 204 zones (the 88 zones eliminated consisted of all those with 30 or fewer total dwelling units plus a few others deemed desirable to delete). In addition to a few other zones deleted on the basis of judgement, those zones with fewer than 50 total employees were deleted from the home based work attraction analysis and those zones with fewer than 50 retail sales employees were deleted from the home based shopping attraction analysis. This left 191 zones and 147 zones, respectively.

Production Equations

Average zonal values were used in investigating the effects of editing procedures upon the production equations. The dependent variables were vehicle trips per dwelling unit (by purpose by editing procedure), and the independent variables used were persons per dwelling unit, cars per dwelling unit, and income per dwelling unit (no labor force data were readily available). All of the variables were derived from the travel survey data and thus were dependent upon it. The means of the various variables are tabulated in Table 15, and the correlations between them are tabulated in Table 16.

BMD02R was used to build the six production equations. Typically, cars per dwelling unit was the first variable selected in all six equations. When examining the regression output for the three editing procedures, one could only conclude that the results obtained by using procedure A data and procedure C data were, for all practical purposes, identical. However, the results obtained using procedure B data did differ.

Table 15. Production Equation Variable Means

Variable	Procedure		
	A	B	C
HBW Trips/D.U.	1.315	1.235	1.314
HBS Trips/D.U.	0.983	1.027	0.975
Cars/D.U.	1.450	1.347	1.448
Persons/D.U.	3.142	2.865	3.139
Income/D.U. (in \$1,000's)	8.405	7.908	8.407

Table 16. Correlation Data For Home Based Production Equation Variables

Variable	Independent Variable*		
	Cars D.U.	Persons D.U.	Income D.U.
<u>HBW Trips</u> D.U.	0.730	0.543	0.542
	0.576	0.416	0.381
	0.738	0.547	0.553
<u>HBS Trips</u> D.U.	0.440	0.303	0.235
	0.281	0.220	0.098
	0.448	0.312	0.258
<u>Cars</u> D.U.	1.000	0.643	0.756
	1.000	0.472	0.751
	1.000	0.649	0.764
<u>Persons</u> D.U.		1.000	0.548
		1.000	0.484
		1.000	0.555
<u>Income</u> D.U.			1.000
			1.000
			1.000

*In each cell, procedure A data is on the top line, procedure B data on the middle line, and procedure C data on the bottom line.

Attraction Equations

Zonal total vehicle trips were used in investigating the effects of editing procedures upon the attraction equations. Total employment by zone, employment by category by zone, and retail sales were the independent variables available for analysis. None of these variables came from the travel survey data, so only the dependent variables differed for the various editing procedures.

BMD02R was used to build the six attraction equations. A satisfactory result was obtained with the single variable total employment in all three home based work equations, and a similar result was obtained with the single variable retail sales employment in all three home based shopping equations. The results of applying those single variable equations to zones with different numbers of employees are summarized in Tables 17 and 18, respectively.

Table 17. Estimated Home Based Work Attractions

Procedure	Number of Total Employees				
	50	400	1,000	2,000	3,000
A	78	395	939	1,845	2,752
B	78	368	866	1,696	2,525
C	78	394	936	1,839	2,743

Table 18. Estimated Home Based Shopping Attractions

Procedure	Number of Retail Sales Employees				
	50	100	200	500	1,000
A	165	349	717	1,822	3,662
B	173	371	767	1,955	3,936
C	166	349	716	1,817	3,651

Table 17. Estimated Home Based Work Attractions

Procedure	Number of Total Employees				
	50	400	1,000	2,000	3,000
A	78	395	939	1,845	2,752
B	78	368	866	1,696	2,525
C	78	394	936	1,839	2,743

Table 18. Estimated Home Based Shopping Attractions

Procedure	Number of Retail Sales Employees				
	50	100	200	500	1,000
A	165	349	717	1,822	3,662
B	173	371	767	1,955	3,936
C	166	349	716	1,817	3,651

CHAPTER IX EFFECTS OF DIRECTIONAL SPLIT ASSUMPTIONS

General

In investigating the effect of an assumed 50-50 directional split in converting triptables from P&A format to O&D format, only the "best" data (i.e., procedure C data) were used. A total purpose O&D triptable was built from trip cards, and P&A triptables were built from the same set of trip cards. The P&A triptables which were built were home based work, home based shopping, home based other, all non-home based (including those trips inventoried in the external, truck, and taxi surveys), and total purpose. All triptables consisted of ADT vehicle trips.

From the one merged P&A triptable dataset, two new total purpose triptables were built. The first of these was obtained by splitting each of the three home based purposes 50-50 and adding in the non-home based trips. The second of the new triptables was built by applying the computed areawide average directional splits (52.71-47.29, 47.81-52.19, and 54.10-45.90, respectively) to each of the three home based purposes and adding in the non-home based trips.

Comparison of Network Loads

All three total purpose triptables (O&D, 50-50 split P&A, and correctly split P&A) were separately loaded onto the network and the loads compared.

Table 19 compares the directional link volumes obtained from the 50-50 split triptable to those from the correctly split triptable. The maximum volume difference on any link in the network was 250 trips, and the per cent root mean square error was 1.79.

Table 20 compares the directional link volumes obtained from the correctly split triptable to those from the O&D triptable, and Table 21 makes the corresponding comparison using the 50-50 split triptable. The maximum volume differences on any link in the network were 302 and 290, respectively. The corresponding per cent root mean square errors were 2.35 and 2.15, respectively. Thus, the 50-50 split triptables matched O&D loads slightly better than the "correctly" split triptables (this seeming paradox is undoubtedly due to errors introduced by using areawide average directional splits).

In passing, Table 22 dramatically points out the necessity of converting P&A triptables to O&D format before making any comparisons against O&D assignment data. When a P&A triptable was loaded onto the network and compared against the loads obtained by using the corresponding O&D triptable, the maximum volume difference on any link jumped to 4,777 with a 33.09 per cent root mean square error.

TABLE 22: COMPARISON OF OBD AND PBA NETWORK LOADS

COMPARISON OF DIRECTIONAL LINK VOLUMES, BASE+OBD LOADS, TEST+PBA LOADS PAGE 33

TABLE 2 - MISCELLANEOUS STATISTICS
INCLUDES ALL SELECTED LINKS EXCEPT THOSE WITH BOTH ZERO BASE AND ZERO TEST VOLUMES
TOTALS FOR ALL VOLUME GROUPS

TOTAL TEST TRIPS/HOUR	1 = 8,406,936	MEAN TEST VOLUME =	226.02	ROOT MEAN SQUARE ERROR =	749.12573
TOTAL BASE TRIPS/HOUR	1 = 8,398,302	MEAN BASE VOLUME =	2263.69	PER CENT R.M.S. ERROR =	33.001063
SUM OF DIFFERENCES	= 8,634	MEAN DIFFERENCE =	2.33	VARIANCE =	56135.50
SUM OF SQUARES	= 0.209201360 10	NUMBER OF LINKS =	3,710	STANDARD DEVIATION =	749.22314

RELATION OF TEST VOLUMES TO CORRESPONDING BASE VOLUMES
NUMBER OF LINKS ON TOP LINE, SUM OF TEST VOLUMES ON MIDDLE LINE, SUM OF BASE VOLUMES ON BOTTOM LINE

BASE VOLUME	TEST VOLUME		NON-ZERO		TOTAL
	TEST EQUAL TO BASE	TEST NOT EQUAL TO BASE	TEST EQUAL TO BASE	TEST NOT EQUAL TO BASE	
ZERO	0/ 0.0 %	0/ 0.0 %	0/ 0.0 %	0/ 0.0 %	0/ 0.0 %
NON-ZERO	0/ 0.0 %	16/ 0.43 %	174/ 4.69 %	3,520/ 94.88 %	3,710/ 100.00 %
TOTAL	0/ 0.0 %	16/ 0.43 %	174/ 4.69 %	3,520/ 94.88 %	3,710/ 100.00 %

BREAK-DOWN BY RELATIVE MAGNITUDE OF THOSE LINKS WITH BOTH BASE AND TEST VOLUMES GREATER THAN ZERO AND NOT EQUAL
RELATIVE MAGNITUDE
TEST GREATER THAN BASE
TEST LESS THAN BASE
TOTAL

MINIMUM AND MAXIMUM INDIVIDUAL LINK VOLUMES		TEST	
VOLUME	LINK	VOLUME	LINK
ZERO VOLUME			
MINIMUM NON-ZERO VOLUME	1 678-680*	1	2104-1734*
MAXIMUM VOLUME	20,347 1900-1201	72,157	678-680*
EXTREME RELATIONSHIPS BETWEEN CORRESPONDING BASE AND TEST VOLUMES			
VOLUME DIFFERENCE		PER CENT DEVIATION	
TEST-BASE	LINK	100*TEST-BASE / BASE	LINK
VOLUME		PER CENT	
MAXIMUM NEGATIVE	-3,570 702-706	-100.00 %	93-1636*
MAXIMUM POSITIVE	4,777 706-707	213.64 %	1528-1527

* AS WELL AS OTHER LINKS

CHAPTER X ADDITIONAL ERROR ANALYSES

Residual Incorrectly Recorded Data Errors

As pointed out in Chapter I, computer error checking, no matter how thoroughly done, only detects a portion of the total number of incorrectly recorded data errors. This is because many, if not most, such errors result in legitimate but erroneous codes which pass the error checks.

In order to get a better understanding of the true number of incorrectly recorded data errors and their causes, it was decided to go back and recode a statistical sample of the original interview data and compare the recoded data with the originally coded data. It was determined that a systematic sample would be statistically valid, and thus 1 in 10 of the Lafayette home interviews, 1 in 8 of the Evansville home interviews, 1 in 10 of the Evansville external interviews, 1 in 6 of the Evansville truck samples, and 1 in 2 of the Evansville taxi samples were recoded.

Since the original coding appeared directly on the interview forms, it was not practical to independently recode the data. However, the personnel doing the recoding were strongly urged to ignore the original coding and their work was frequently spot checked to insure that they were performing adequately.

After the data were coded, they were digitized and compared against the originally coded data. Discrepancies were noted and resolved by subsequent reference to the original interview data, and, finally, error rates were computed (see Table 23). Because in a strict sense the data were not independently recoded, it cannot be asserted that

Table 23. Error Rates Expressed In Percentages Found In Travel Survey Data During Recoding

Data Field	Evansville				Lafayette
	Home Interview	External	Truck	Taxi	Home Interview
Origin Zone	1.08	4.72	0.96	6.39	1.96
Destination Zone	0.68	4.79	0.95	6.11	1.74
Origin Land Use	2.41	1.60	1.93	9.47	----
Destination Land Use	2.13	1.57	1.85	9.09	----
From Purpose	0.35	1.03	0.12	0.20	0.46
To Purpose	0.29	1.09	0.12	0.14	0.42
Start Time	1.06	0.39	0.89	1.30	2.03
Arrive Time	1.04	----	0.90	1.33	1.98
Mode	0.15	----	----	----	0.73
Other Fields	0.83	1.49	0.76	1.10	3.70

the tabulated error rates are "true" or "total" error rates. However, it is felt that they approximate the total error rate. The significant point is that the total number of incorrectly recorded data errors is a few times the number of such errors which can be found by computer error checking. Computations disclose that the errors are not randomly distributed but tend toward the most frequently occurring values.

An analysis of the causes of the various errors discloses that "procedural" errors are by far the major cause of such coding errors (for example, see Table 24).

Errors such as picking up a zone code from the wrong line of a coding index, picking up the zone code for an even numbered street address when the actual number is odd, and incorrectly using the destination zone number of the preceding trip as the origin zone number of the next trip are examples of procedural errors.

Higher error rates are found where the coder must exercise judgement, such as in the determination of a land use or commodity code. When the coders apparently thought that there was small necessity for a correct code, it frequently appeared that they were excessively careless in their code selection.

There were many occurrences of incorrect conversion of time to the 24 hour clock used in both the Evansville and Lafayette studies.

Table 24. Breakdown Of Errors In Coding Of Trip End Addresses By Causes

Evansville (except external)			Lafayette		
Bad Interview					
Information	4.9		36.3		4.2
Procedural Errors	64.2		26.1		60.1
Discrepancy in Reference Material	3.1		1.7		5.8
Digit Substitution	22.6		21.5		21.3
Transposition	2.8		3.5		5.5
Omission of Digit(s)	2.4		10.9		3.1

Substantially higher error rates were found in the coding of the external survey data on mark sense forms. This was partially due to the coders having to "manufacture" data which the field interviewers (sometimes understandably) failed to obtain, but primarily it was a function of the great difficulty in coding mark sense forms.

Almost all incorrectly recorded data errors were made by the coders. Relatively few were caused by bad interview information, and even fewer were caused by those keypunch errors which survived key verification.

Keypunching Errors

With the exception of the Evansville external data, all of the travel survey data which were recoded as described in the previous section were keypunched. Thus there was opportunity for investigating keypunch errors, and it was decided to do so.

The travel survey data were initially keypunched but not verified, after which duplicate sets of cards were mechanically made. After the original sets of cards were subsequently verified, the corresponding sets of cards were compared. The three keypunch shops involved were not told the reason for the two step process, but they may have surmised it. If so, one would hypothesize that they would have attempted to be more accurate than usual in the original keypunching in order to avoid disclosing an excessively high error rate.

Table 25 summarizes the discrepancies found between the corresponding sets of cards. The per cent of keystrokes in error is actually the per cent of keystrokes in error which were removed by key verification (a small number of errors remained after key verification, some of which showed up in the analysis of the recoded data).

Three critical errors do not show up in Table 25. Through human error or, more likely, through mechanical malfunction, three of the unverified cards contained an

invalid combination of punches in a given column. These errors prohibited the cards from even being read by the computer and thus presented serious processing problems. All of these errors were found and corrected during key verification.

Table 25. Key punching Errors

	Dwelling Unit Data		Internal Trip Data		Truck	
	Evans.	Laf.	Evans.	Laf.	Evans.	Taxi Evans.
Original Cards	786	507	6,784**	3,861	1,170	454
Extra Cards*	1	3	1,035**	23	12	1
Missed Cards	0	0	29	10	1	0
Final number of cards	785	504	5,778	3,848	1,159	453
Per Cent of Keystrokes in error	0.05	0.03	0.12	0.08	0.92	0.00

*Cards which were erroneously included.

**Includes one major set of 1,022 cards which were done over (both sets were returned to the project without any indication that there was a "duplicate" deck included).

CHAPTER XI SUMMARY AND CONCLUSIONS

Results of Using Various Editing Procedures

The analysis of the results found in this study involving the travel survey data collected in Evansville, Indiana, clearly shows that no significant differences are found in the base year traffic assignment, trip distribution, and trip generation models between a procedure using unedited data and one using thoroughly edited and corrected data. However, substantial operational difficulties are encountered in getting some of the computer programs (e.g., the programs which inserted expansion factors) to perform successfully when using totally unedited data as input.

In comparison to the results which are obtained by following either of the above editing procedures, significantly different results are found in the three travel models when an editing procedure is followed which calls for the deletion of an entire sample of clustered trip data whenever an error is found in the sample. Two possible causes for the resulting differences were observed. First, because of minor coding errors, an inordinate number of non-completed interviews was rejected thus biasing the computed expansion factors in favor of completed interviews. Second, the normal discarded sample was not representative of the entire population of samples. It typically contained more than the average number of trips and, in the case of the home interview survey, the household had higher than average number of persons, number of cars, and income.

Although only a portion of the total number of incorrectly recorded data errors are detected by computer editing, one would not expect to get significantly different results even if he were able to remove all such errors.

Substantially more trip cards failed consistency checks than failed single field error checks. However, all data errors causing operational problems in the use of the various computer programs were detected by single field error checks. The consistency checks were more difficult to implement in a computer program and, for those cards rejected, provided the greater difficulty in determining the necessary corrections.

Directional Split Assumptions

When working with unlinked ADT trips, an assumed 50-50 directional split for converting a triptable from P&A format to O&D format is justifiable. While the observed differences might be slightly greater if linked trips were used, it would be reasonable to expect that the errors introduced by an assumed 50-50 directional split would still be negligible.

Computer Program Contributions

During the conduct of this project, it was necessary to completely rewrite programs REFORM, COMPARE, and SPLIT from the Federal Highway Administration's IBM 360 battery of urban transportation planning computer programs. The corresponding source decks and documentation have been furnished to the Federal Highway Administration for incorporation in the forthcoming version of that battery. Similarly, they have been furnished minor corrections to several other computer programs, including GM and TDIST.

CHAPTER XII

GUIDELINES FOR DATA PROCESSING AND EDITING

The following paragraphs consist of notes and guidelines which reflect acknowledged good practice, the experiences of the author in earlier transportation studies, and the findings of this research.

Notes on Form and Code Design

The Indiana State Highway Commission, as well as other agencies involved in the collection of travel survey data, should consider redesigning some of its travel survey interview and coding forms with the following considerations in mind:

1. Those data fields common to both the dwelling unit summary and internal trip report forms (e.g., sample number, card number, and residence zone) should be located in the same columns on each.
2. Provision should be made on all but external interview forms for three digit trip numbers and number of trips. Possibly the size of other fields (e.g., number of persons regularly enrolled in school) also needs to be increased.
3. Since procedural coding errors were found to be the major cause of incorrectly recorded data, simplicity of form design is essential (this definitely applies to geographic coding indices as well as to travel survey forms).

Similar consideration should be given to the following points in designing the codes to be used on the forms:

1. Never use a code in which a blank field is a legitimate response (such as for an unknown land use), for there is no way that a computer program can subsequently determine whether the intended response was indeed a blank or whether another response was intended but omitted (omitted codes are a non-trivial type of coding error).

2. If possible, a code consisting of all zeros in the given field should be avoided, as it sometimes requires more complicated computer programming to be able to differentiate between a field of blanks and a field of zeros.

3. When using an IBM state-county or similar code for external zone numbers, it should be decided and stated in advance how to code responses in which only the state is indicated.

4. Serious thought should be given to using a normal 12 hour clock time with AM or PM suffix instead of a 24 hour clock time. Apparently due to unfamiliarity, many interviewer and coder errors were made in recording time to a 24 hour clock.

5. Consideration might be given to applying some of the results of the psychological research into the nature of errors discussed in the Review of Literature in an attempt to design codes which result in the least amount of incorrectly recorded data errors.
6. The use of a single card column to record two or more distinct data values through the use of "multiple punches" should be avoided.

Guidelines for Data Processing

Data digitized through a keyboard should be key verified.

When the output from the digitizing process is punch cards, the first data processing step involving the computer is to copy the punch cards onto some magnetic recording medium (normally a high density magnetic tape) which will reduce their physical volume and facilitate subsequent computer processing. It is advisable to insert a record sequence number into each output card image so that the original sequence can be determined after subsequent processing has altered the sequence.

Also, leading zeros should be inserted in at least the major and perhaps the minor sort fields in order to insure that the card image will be positioned correctly after the initial sort (this precaution is advisable even if leading zeros were supposedly coded and digitized as some will probably be missing). Finally, efficient data processing may dictate that the card layout used for coding and digitizing be altered. Computer programs are available which perform this process of copying the original data cards onto a computer-readable magnetic recording medium while simultaneously reformatting the data and inserting leading zeros and record sequence numbers.

It would be imprudent to entrust travel survey data collected at great expense to old and/or poor quality magnetic tapes (or similar storage volumes). The cost of good tapes certified at high density and full width has dropped significantly, and it would be advisable to obtain a sufficient number of such tapes prior to beginning computer processing. Even with such tapes, however, consideration should be given to having them initially certified and periodically cleaned. The certification process tests the quality of the tape but destroys any data previously recorded on it. The cleaning process simply removes lint and dust from the surface of the tape without destroying the magnetically recorded data.

In transportation studies, the number of sets of data requiring storage on a tape or similar volume increases rapidly as the study progresses through the editing, factor insertion, and model development phases. As a result, there will be temptation to rapidly scratch (i.e., effectively destroy) "old" sets of data, but sound data processing practice dictates retaining at least one and preferably two or more of the immediately previous generations of data so that, if necessitated by human and/or mechanical failure, the current generation could be recreated. Similarly, wise planning requires maintaining duplicate copies of critical sets of data and programs at a separate physical location in case the primary site is destroyed or damaged by some disaster.

Complete and accurate record keeping is a necessity, particularly since the number of sets of data will be quite large. For each individual set of data on a given storage volume, these records should include such information as the date generated, how generated, what the input data were, how many records are in the set of data, and what other sets of data have been created using this set as input and on what dates. These records should be retained even after the data have been scratched as it may be necessary at some later date to trace the ancestry of a given set of data.

Guidelines for Computer Editing

The analyses for which the travel survey data are to be used should, if possible, be determined prior to editing. The editing procedure to be used should then incorporate the minimal number of error checks consistent with the uses to be made of the data.

Assuming good quality control in the interviewing, coding, and digitizing operations, the only consistency checks which are necessary in terms of the accuracy of the resulting traffic assignment, trip distribution, and trip

generation models are those which verify that the number of total trips and the number of auto driver trips (if applicable) in a home interview, truck, or taxi sample is correct. For other analyses such as a parking study based on O&D data (44), other consistency checks might be justified. Each additional consistency check made in home interview survey data will cost, on the average, about thirty-five dollars per 10,000 trip cards plus result in the loss of additional calendar time before the data are available for analysis purposes. Similar costs and time losses could be expected with other travel survey data.

Because of problems which might otherwise result in the operation of various computer programs, it is advisable to make all of the single field error checks in those fields which contain data which will actually be used. The checks made conceivably could be limited to those which insure that no invalid characters or embedded blanks exist in the given field, but it would be best to verify that the contents of the field represent a valid number or code.

Clustered trip data failing the error checks should be either corrected or replaced by "good" data. Only rarely should the entire sample (or person) be discarded. Well thought-out procedures for compensating in the expansion factors for discarded data must be developed and applied.

After corrections have been made to travel survey data containing errors, the updated data should be recycled through the error checking process to insure that the changes made were correct and did not introduce other error failures.

If interview rates at the external stations are fairly high, the corresponding percentage of error check failures is relatively small, and the distribution of the error check failures is fairly random with respect to station, direction, hour, and vehicle type, then consideration should be given to simply discarding the records in error.

CHAPTER XIII RECOMMENDATIONS FOR ADDITIONAL RESEARCH

The results obtained in this research project and some of the reviewed literature both indicate that the "automatic" correction by the computer of travel survey data failing error checks would probably save time and money in the editing phase while not having a detrimental effect on the accuracy of the resulting models. Such a methodology should be developed and tested.

LIST OF REFERENCES

1. Beach, James H. "The Input Comes with Key-to-Tape Devices." Computing Decisions, May, 1971, pp. 38-39.
2. Bentley, Dudley F. "An Optimum System with Multiple I/O." The Computer Bulletin, 11 December, 1967, 212-213.
3. "Check Disk Verification." Data Processing, London, 8 (July-August, 1966), 19-20.
4. Bell Telephone Laboratories, Inc., Western Longbeach Department, Common Language Coding Guide, Bell Telephone Laboratories Incorporated, 1970, 51 pp.
5. Blaser, Albert E. "Key-to-Tape & Key-to-Disk Systems." Modern Data, 4 (December, 1971), 22-26.
6. Dool, R. C. "Error Detecting and Error Correcting Indexing Systems for Large Serial Numbers." Service of the International Statistical Institute, 31 March, 1961, 334-340.
7. Edwards, Richard J. "Environmental Factors in Urban Transportation Planning." remarks delivered at the 54th Annual Meeting of the American Association of State Highway Officials, in Houston, Texas, on November 10, 1970. Text distributed in Department of Transportation News, 34-DOT-70, November, 1970. 10 pp.
8. Gandon, Dick H. Management Standards for Data Processing. Princeton, New Jersey: D. Van Nostrand Company, Inc., 1963. 404 pp.
9. Carlson, Gary. "Predicting Clerical Error--in an RDP Environment." Database, 3 (February, 1973), 34-36.
10. Ockers, H. "Experimental Psychology in the Field of Telecommunication." Ergonomics, 1 October, 1963, 289-293.
11. "The Location of Figures in Alpha-Numeric Codes." Ergonomics, 6 (April, 1963), 403-404.

LIST OF REFERENCES

1. Bauch, James H. "Cut Input Costs with Key-to-Tape Devices," Computer Decisions, May, 1971, pp. 36-39.
2. Beckley, Dudley F. "An Optimum System with 'Modulus 11'," The Computer Bulletin, 11 (December, 1967), 213-215.
3. _____. "Check Digit Verification," Data Processing, London, 8 (July-August, 1966), 194-201.
4. Bell Telephone Laboratories, Inc., Common Language Department. Common Language Coding Guide, Bell Telephone Laboratories Incorporated, 1970. 51 pp.
5. Bieser, Albert H. "Key-to-Tape & Key-to-Disk Systems," Modern Data, 4 (December, 1971), 32-36.
6. Bose, R. C. "Error Detecting and Error Correcting Indexing Systems for Large Serial Numbers," Review of the International Statistical Institute, 34 (March, 1966), 334-340.
7. Bouchard, Richard J. "Environmental Factors in Urban Transportation Planning," remarks delivered at the 56th Annual Meeting of the American Association of State Highway Officials, in Houston, Texas, on November 10, 1970. Text distributed in Department of Transportation News, 54-DOT-70, November, 1970. 10 pp.
8. Brandon, Dick H. Management Standards for Data Processing. Princeton, New Jersey: D. Van Nostrand Company, Inc., 1963. 404 pp.
9. Carlson, Gary. "Predicting Clerical Error--in an EDP Environment," Datamation, 9 (February, 1963), 34-36.
10. Conrad, R. "Experimental Psychology in the Field of Telecommunications," Ergonomics, 3 (October, 1960), 289-295.
11. _____. "The Location of Figures in Alpha-Numeric Codes," Ergonomics, 5 (April, 1962), 403-406.

12. Conrad, R. and Hull, A. J. "Copying Alpha and Numeric Codes by Hand: An Experimental Study," Journal of Applied Psychology, 51 (October, 1967), 444-448.
13. "Data Entry," Government Data Systems, 1 (May/June, 1971), 21-25, 28-29, 32-33, 36-37.
14. Deming, W. Edwards and Leon Geoffrey. "On Sample Inspection in the Processing of Census Returns," Journal of the American Statistical Association, 36 (September, 1941), 351-360.
15. Deming, W. Edwards, Benjamin J. Tepping, and Leon Geoffrey. "Errors in Card Punching," Journal of the American Statistical Association, 37 (December, 1942), 525-536.
16. Edens, H. J., Chief Traffic Engineer, DeLeuw, Cather & Company of Canada, Ltd., personal letter dated December 24, 1970.
17. Fasteau, Herman H., J. Jack Ingram, and George Minton. "Control of Quality of Coding in the 1960 Censuses," Journal of the American Statistical Association, 59 (March, 1964), 120-132.
18. Fleet, Christopher R. Unpublished research investigating effect of variations in sampling rate on transportation planning process, Urban Planning Division, Bureau of Public Roads, circa 1968.
19. Freund, R. J. and H. O. Hartley. "A Procedure for Automatic Data Editing," Journal of the American Statistical Association, 62 (June, 1967), 341-352.
20. Gralen, Oliver E. "Check Digits--Does Geometry Really Do a Better Job?" Auditgram, July, 1961, pp. 16-18.
21. Hammer, Carl. "The Case of the Badly Maligned Key punch Operator," Computerworld, 4 (March 25, 1970), 12.
22. Heanue, Kevin E., Lamelle B. Hamner, and Rose M. Hall. "Adequacy of Clustered Home Interview Sampling for Calibrating a Gravity Model Trip Distribution Formula," Travel Patterns, Highway Research Record Number 88, Highway Research Board, 1965, pp. 116-136.
23. Highway Research Board. National Cooperative Highway Research Program Report 120: Data Requirements for Metropolitan Transportation Planning, Washington, D.C.: Highway Research Board, 1971. 90 pp.

24. Hinds, G. H. "The Accuracy of Data Preparation," The Computer Bulletin, 4 (June, 1960), 7-9.
25. Horton, David L. and Michael Mecherikoff. "Letter Preferences: Ranking the Alphabet," Journal of Applied Psychology, 44 (1960), 252-253.
26. Indianapolis Regional Transportation and Development Study. Travel Characteristics. Technical work paper, job 5352. February, 1966.
27. Kincaid, William H. and Charles H. Simpson. "Use the the Editors You Have!" Data Processing, 3 (August, 1961), 32-36.
28. Klemmer, E. T. "Grouping of Printed Digits for Manual Entry," Human Factors, 11 (August, 1969), 397-400.
29. _____, and Lockhead, G. R. "Productivity and Errors in Two Keying Tasks: A Field Study," Journal of Applied Psychology, 46 (December, 1962), 401-408.
30. Marks, Manuel A., Jr., Urban Planning Division, Bureau of Public Roads, private discussion, June 12, 1970.
31. Minton, George. "Inspection and Correction Error in Data Processing," Journal of the American Statistical Association, 64 (December, 1969), 1256-1275.
32. _____. "Some Decision Rules for Administrative Applications of Quality Control," Journal of Quality Technology, 2 (April, 1970), 86-98.
33. Myers, Gibbs. "Coding Business Data," Data Processing, 3 (October, 1961), 23-25.
34. Neisser, Ulric and Paul Weene. "A Note on Human Recognition of Hand-Printed Characters," Information and Control, 3 (June, 1960), 191-196.
35. Nordbotten, Svein. "Automatic Editing of Individual Statistical Observations," Proceedings of Conference of European Statisticians, Statistical Standards and Studies--No. 2. New York: United Nations, 1963.
36. _____. "Measuring the Error of Editing the Questionnaires in a Census," Journal of the American Statistical Association, 50 (June, 1955), 364-369.

37. Nordbotten, Svein. "The Efficiency of Automatic Detection and Correction of Errors in Individual Observations as Compared with Other Means for Improving the Quality of Statistics," Bulletin of the International Statistical Institute. Proceedings of the 35th Session, 1965. Vol. XLI, Book 1, Belgrade, Yugoslavia, 1966. pp. 417-441.
38. Nucci, Hubert P. "Computer Coding of Origins and Destinations," Origin and Destination Characteristics, Highway Research Record Number 191, Highway Research Board, Washington, D. C., 1967. pp. 76-102.
39. O'Reagan, Robert T. "Relative Costs of Computerized Error Inspection Plans," Journal of the American Statistical Association, 64 (December, 1969), 1245-1255.
40. Owsowitz, Sidney and Anders Sweetland. Factors Affecting Coding Errors, Research Memorandum, RM-4346-PR, prepared for U.S.A.F. Project RAND. Santa Monica, California: RAND Corporation, April, 1965. 29 pp.
41. Page, C. F. "Checking in Data Processing," Instruments & Control Systems, 37 (August, 1964), 129-131.
42. Pritzker, Leon, Jack Ogus, and Morris H. Hansen. "Computer Editing Methods--Some Applications and Results," Bulletin of the International Statistical Institute. Proceedings of the 35th Session, 1965. Vol. XLI, Book 1, Belgrade, Yugoslavia, 1966. pp. 442-472.
43. Quality Control of Origin-Destination Surveys, a collection of five papers presented at a panel on January 17, 1966, Highway Research Circular Number 49, Highway Research Board, October, 1966.
44. Schulman, Lawrence L. and Robert W. Stout. "A Parking Study Through the Use of Origin-Destination Data," Parking Analyses, Highway Research Record Number 317, Highway Research Board, 1970. pp. 14-29.
45. Severin, F. T. and M. K. Rigby. "Influence of Digit Grouping on Memory for Telephone Numbers," Journal of Applied Psychology, 47 (April, 1963), 117-119.
46. Smith, William A., Jr. "Accuracy of Manual Entries in Data-Collection Devices," Journal of Applied Psychology, 51 (August, 1967), 362-368.

47. Sosslau, Arthur B. and Glenn E. Brokke. "Appraisal of Sample Size Based on Phoenix O-D Survey Data," Traffic Origin-and-Destination Studies: Appraisal of Methods, Highway Research Board Bulletin 253, Highway Research Board, 1960, pp. 114-127.
48. Stanhagen, John Franklin, Jr. "Coding Errors in Management Information Systems," unpublished D.B.A. dissertation, University of Southern California, June, 1968.
49. Szameitat, Klaus and Hans-Joachim Zindler. "The Reduction of Errors in Statistics by Automatic Corrections," Bulletin of the International Statistical Institute, Proceedings of the 35th Session, 1965. Vol. XLI, Book 1, Belgrade, Yugoslavia, 1966. pp. 395-417.
50. Trimble, George R., Jr. and Anthony J. Penta. "Evaluation of Keyboard Data Entry Systems," Datamation, 16 (June, 1970), 93-99.
51. U.S. Bureau of Public Roads. Calibrating and Testing a Gravity Model for Any Size Urban Area. Washington, D. C.: Government Printing Office, July, 1963. 214 pp.
52. _____. Calibrating & Testing a Gravity Model for Any Size Urban Area. Washington, D. C.: Government Printing Office, 1965. 170 pp.
53. _____. Manual of Procedures for Home Interview Traffic Study--Revised Edition. October, 1954, Reprinted in 1957 as Procedure Manual 2B, Conducting a Home Interview Origin-Destination Survey. Chicago: Public Administration Service. 109 pp.
54. _____. Systematic Sensitivity Analysis of the Urban Travel Forecasting Process. Contract No. CPR-11-6005. Washington, D. C.: Bureau of Public Roads, December 15, 1968. 195 pp.
55. _____. Traffic Assignment Manual. Washington, D. C.: Government Printing Office, June, 1964. 282 pp.
56. U.S. Bureau of the Budget. Handbook for Data Standarization (Draft). Washington, D.C.: Bureau of the Budget, October 1969. 102 pp.
57. U.S. Bureau of the Census. Atlantida: A Case Study in Household Sample Surveys, Unit VII. Editing and Coding. Series ISPO 1, No. 1-1. Washington, D.C.: Bureau of the Census, 1966. 58 pp.

58. U.S. Bureau of the Census. United States Censuses of Population and Housing 1960: Quality Control of Preparatory Operations, Microfilming, and Coding. Washington, D. C.: Bureau of the Census, 1965. 57 pp.
59. U.S. Federal Highway Administration. Guidelines for Urban Travel Data Collection. Fourth Edition. Third draft under contract no. FH-11-7501. Washington, D. C.: Federal Highway Administration, 1971. 594 pp.
60. _____. Urban Transportation Planning: General Information and Introduction to System/360. Washington, D. C.: Federal Highway Administration, June, 1970. 334 pp.
61. _____. Urban Transportation Planning System 360 Program Documentation. Washington, D. C.: Federal Highway Administration, June, 1970. 614 pp.
62. Waksberg, Joseph and Leon Pritzker. "Changes in Census Methods," Journal of the American Statistical Association, 64 (December, 1969), 1141-1149.
63. Waugh, Nancy C. "Serial Position and the Memory Span," American Journal of Psychology, 73 (March, 1960), 68-80.
64. Wilson, J. G. "Notes on Geometric Weighted Check Digit Verification," Communications of the ACM, 4 (December, 1961), 551-552.
65. Yule, G. Udny. "On Reading a Scale," Journal of the Royal Statistical Society, London, 90 (1927), 570-587.

APPENDIX A
SURVEY FORMS

The following figures are photo-reproductions of the actual forms used for coding the travel survey data from the Evansville and the Lafayette transportation studies (with the exception of the Lafayette street survey, the same forms were also used for recording the original interview data). Unless otherwise specified in the caption, the same form was used in both Evansville and Lafayette.

APPENDICES

APPENDIX A SURVEY FORMS

The following figures are photo-reductions of the actual forms used for coding the travel survey data from the Evansville and the Lafayette transportation studies (with the exception of the Lafayette external survey, the same forms were also used for recording the original interview data). Unless otherwise specified in the caption, the same form was used in both Evansville and Lafayette.

TRANSPORTATION STUDY - DWELLING UNIT SUMMARY

Card Number _____
 Zone Number _____
 Census Tract Number _____
 Block Number _____
 Sample Number _____
 Travel Date _____

Preceding Number: _____
 Interview Address: _____
 Succeeding Number: _____
 Date of Travel: _____

A. Structure Type: 1. Single Detached 2. 2-4 Family Apt. 5. 21 or More Family Apt. 7. Hotel 9. Rest Home, Institution
 2. Single Attached 4. 5-20 Family Apt. 6. Rooming House, Dormitory 8. Motel 10. Other (Specify) _____

B. Do you (1.) Rent or (2.) Own this dwelling unit? _____

C. How many passenger cars are owned by persons living at this address? (Include company-owned cars) _____

D. How many persons live at this address? _____

E. How many persons 5 years of age or over live here? _____

F. How many persons are regularly employed? _____

G. How many persons are regularly enrolled in school? _____

H. How long has head of household lived at this address? _____ Years and _____ Months

I. Number of persons 5 years of age or older making trips: _____

J. Number of persons 5 years of age or older not making trips: _____

K. Number of persons 5 years of age or older with trips unknown: _____

L. Number of auto driver trips reported at this address: _____

M. Total number of trips reported at this address: _____

N. Race: 1. Caucasian 2. Negro 3. Other _____

O. Family income range: 1. 2. 3. 4. 5. 6. 7. 8. 9. (No Response) _____

P. Occupant information (complete all items below for each person, including visitors)

A		B		C		D		E		F		G		H		I		J		K		L	
1		2		3		4		5		6		7		8		9		10		11		12	
13		14		15		16		17		18		19		20		21		22		23		24	
25		26		27		28		29		30		31		32		33		34		35		36	
37		38		39		40		41		42		43		44		45		46		47		48	
49		50		51		52		53		54		55		56		57		58		59		60	
61		62		63		64		65		66		67		68		69		70		71		72	
73		74		75		76		77		78		79		80		81		82		83		84	
85		86		87		88		89		90		91		92		93		94		95		96	
97		98		99		100		101		102		103		104		105		106		107		108	
109		110		111		112		113		114		115		116		117		118		119		120	
121		122		123		124		125		126		127		128		129		130		131		132	
133		134		135		136		137		138		139		140		141		142		143		144	
145		146		147		148		149		150		151		152		153		154		155		156	
157		158		159		160		161		162		163		164		165		166		167		168	
169		170		171		172		173		174		175		176		177		178		179		180	
181		182		183		184		185		186		187		188		189		190		191		192	
193		194		195		196		197		198		199		200		201		202		203		204	
205		206		207		208		209		210		211		212		213		214		215		216	
217		218		219		220		221		222		223		224		225		226		227		228	
229		230		231		232		233		234		235		236		237		238		239		240	
241		242		243		244		245		246		247		248		249		250		251		252	
253		254		255		256		257		258		259		260		261		262		263		264	
265		266		267		268		269		270		271		272		273		274		275		276	
277		278		279		280		281		282		283		284		285		286		287		288	
289		290		291		292		293		294		295		296		297		298		299		300	
301		302		303		304		305		306		307		308		309		310		311		312	
313		314		315		316		317		318		319		320		321		322		323		324	
325		326		327		328		329		330		331		332		333		334		335		336	
337		338		339		340		341		342		343		344		345		346		347		348	
349		350		351		352		353		354		355		356		357		358		359		360	
361		362		363		364		365		366		367		368		369		370		371		372	
373		374		375		376		377		378		379		380		381		382		383		384	
385		386		387		388		389		390		391		392		393		394		395		396	
397		398		399		400		401		402		403		404		405		406		407		408	
409		410		411		412		413		414		415		416		417		418		419		420	
421		422		423		424		425		426		427		428		429		430		431		432	
433		434		435		436		437		438		439		440		441		442		443		444	
445		446		447		448		449		450		451		452		453		454		455		456	
457		458		459		460		461		462		463		464		465		466		467		468	
469		470		471		472		473		474		475		476		477		478		479		480	
481		482		483		484		485		486		487		488		489		490		491		492	
493		494		495		496		497		498		499		500		501		502		503		504	
505		506		507		508		509		510		511		512		513		514		515		516	
517		518		519		520		521		522		523		524		525		526		527		528	
529		530		531		532		533		534		535		536		537		538		539		540	
541		542		543		544		545		546		547		548		549		550		551		552	
553		554		555		556		557		558		559		560		561		562		563		564	
565		566		567		568		569		570		571		572		573		574		575		576	
577		578		579		580		581		582		583		584		585		586		587		588	
589		590		591		592		593		594		595		596		597		598		599		600	
601		602		603		604		605		606		607		608		609		610		611		612	
613		614		615		616		617		618		619		620		621		622		623		624	
625		626		627		628		629		630		631		632		633		634		635		636	
637		638		639		640		641		642		643		644		645		646		647		648	
649		650		651		652		653		654		655		656		657		658		659		660	
661		662		663		664		665		666		667		668		669		670		671		672	
673		674		675		676		677		678		679		680		681		682		683		684	
685		686		687		688		689		690		691		692		693		694		695		696	
697		698		699		700		701		702		703		704		705		706		707		708	
709		710		711		712		713		714		715		716		717		718		719		720	
721		722		723		724		725		726		727		728		729		730		731		732	
733		734		735		736		737		738		739		740		741		742		743		744	
745		746		747		748		749		750		751		752		753		754		755		756	
757		758		759		760		761		762		763		764		765		766		767		768	
769		770		771		772		773		774		775		776		777		778		779		780	
781		782		783		784		785		786		787		788		789		790		791		792	
793		794		795		796		797		798		799		800		801		802		803		804	
805		806		807		808		809		810		811		812		813		814		815		816	
817		818		819		820		821		822		823		824		825		826		827		828	
829		830		831		832		833		834		835		836		837		838		839		840	
841		842		843		844		845		846		847		848		849		850		851		852	
853		854		855		856		857		858		859		860		861		862		863		864	
865		866		867		868		869		870		871		872		873		874		875		876	
877		878		879		880		881		882		883		884		885		886		887		888	
889		890		891		892		893		894		895		896		897		898		899		900	
901		902		903		904		905		906		907		908		909		910		911		912	
913		914		915		916		917		918		919		920		921		922		923		924	
925		926		927		928		929		930		931		932		933		934		935		936	
937		938		939		940		941		942		943		944		945		946		947		948	
949		950		951		952		953		954		955		956		957		958		959		960	
961		962		963		964		965		966		967		968		969		970		971		972	
973		974		975		976		977		978		979		980		981		982		983		984	
985		986		987		988		989		990		991		992		993		994		995		996	
997		998		999		1000		1001		1002		1003		1004		1005		1006		1007		1008	
1009		1010		1011		1012		1013		1014		1015		1016		1017		1018		1019		1020	
1021		1022		1023		1024		1025		1026		1027		1028		1029		1030		1031		1032	
1033		1034		1035		1036		1037		1038		1039		1040		1041		1042		1043		1044	
1045		1046		1047		1048		1049		1050		1051		1052		1053		1054		1055		1056	
1057		1058		1059		1060		1061		1062		1063		1064		1065		1066		1067		1068	
1069		1070		1071		1072		1073		1074		1075		1076		1077		1078		1079		1080	
1081																							

INTERNAL TRIP REPORT

CARD NUMBER	<input type="text"/> 2 <input type="text"/> 7 <input type="text"/>	COST ACCOUNT	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	ZONE OF RESIDENCE	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	SAMPLE NUMBER	<input type="text"/> 1 <input type="text"/> 1 <input type="text"/> 1 <input type="text"/> 1 <input type="text"/> 1	TRAVEL DATE	<input type="text"/> 10 <input type="text"/> 17 <input type="text"/> 18 <input type="text"/> 19
-------------	--	--------------	---	-------------------	--	---------------	--	-------------	---

Page _____ of _____

[illegible]

FIGURE A2 EVANSVILLE INTERNAL TRIP CARD FORM

EXTERNAL INTERVIEW FORM

COST ACCOUNT _____ INTERVIEW BY _____

1. Date of Trip										2. Interviewer's Name									
3. Where Did This Trip Begin?										4. Where Did This Trip End?									
5. Trip Purpose (From)										6. Trip Purpose (To)									
7. Where Does Car Or LRV/Truck Belong?										8. How Many Other Stages in Area?									
9. Road Identity of Exit										10. Road Identity of Exit									

1. Date of Trip

MM	DD	YY	MM	DD	YY	MM	DD	YY	MM	DD	YY
01	01	01	01	01	01	01	01	01	01	01	01

2. Interviewer's Name

0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9

3. Where Did This Trip Begin?

0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9

4. Where Did This Trip End?

0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9

5. Trip Purpose (From)

0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9

6. Trip Purpose (To)

0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9

7. Where Does Car Or LRV/Truck Belong?

0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9

8. How Many Other Stages in Area?

0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9

9. Road Identity of Exit

0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9

10. Road Identity of Exit

0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9

FIGURE A3. EXTERNAL SURVEY TRIP CARD FORM

ADMINISTRATIVE RECORD (Confidential)				TRUCK AND TAXI TRIP SUMMARY			
Interviewer: _____				Card Number _____			
Sample Phone Number: _____				Zone Number _____			
Call-Backs ☒ Telephone Calls		Appointments		Census Tract Number _____			
Date	Time	Date	Time	Block Number _____			
1				Sample Number _____			
2				Travel Date _____			
3				Name _____			
4				Address _____			
5				Telephone Number _____			
6				Garaging Address _____			
NOTES AND COMMENTS Interviewer				Vehicle Type: 1. 2-wheeled trucks single tire			
				2. 2-wheeled dual tire and 3 or more axle single unit			
				3. 3-wheeled dual tire and 3 or more axle single unit			
				4. Truck combinations			
				5. Buses			
REASON FOR NON-INTERVIEW OR INCOMPLETE REPORT				Business-Industry of Owner _____			
				Total Trips Reported _____		Total Miles Trav. _____	
				Total Stops _____		Lic. Cap. (Weight) _____	
				Year _____ Make _____			
				State License Number _____			
Supervisor's Comment				Optional Codes _____			
				COMMENTS			
Report Completed: _____ (Date) _____ (Initial)				FOR MORE THAN TWO TRIPS PER SAMPLE USE CONTINUATION SHEETS			
Interviews Checked: _____ (Date) _____ (Initial)							
Coded By: _____ (Initial)							
Coding Checked by: _____ (Initial)							

Figure 3:

TRUCK AND TAXI TRIP REPORT

1		2		3		4		5		6		7								
Trip Number	Persons in Vehicle	Commodity	WHERE DID THIS TRIP BEGIN (Origin)	WHERE DID THIS TRIP END (Destination)	TIME OF	Start	Arrival	Purpose of trip												
								FROM TO												
43	44	45	46	47	48	49	Empty		Street Address		Street Address		AM PM		<div style="display: flex; justify-content: space-between;"> <div> 0 Some Buss 1 Pers. Use 2 Del. Del. 3 Whal. Del. 4 Miss. Del. 5 Mail Exp. 6 Constr. 7 Wld. Repsr. 8 Ser. Rider 9 Bus. Use </div> <div> 0 1 2 3 4 5 6 7 8 9 </div> </div>					
							City or Town		City or Town		AM PM									
							Land Use		Land Use		AM PM									
							43 44 45 46 47 48 49		50 51 52 53 54 55 56 57		58 59 60 61 62 63 64 65		66 67 68 69 70 71 72							
43	44	45	46	47	48	49	Empty		Street Address		Street Address		AM PM		<div style="display: flex; justify-content: space-between;"> <div> 0 Some Buss 1 Pers. Use 2 Del. Del. 3 Whal. Del. 4 Miss. Del. 5 Mail Exp. 6 Constr. 7 Wld. Repsr. 8 Ser. Rider 9 Bus. Use </div> <div> 0 1 2 3 4 5 6 7 8 9 </div> </div>					
							City or Town		City or Town		AM PM									
							Land Use		Land Use		AM PM									
							43 44 45 46 47 48 49		50 51 52 53 54 55 56 57		58 59 60 61 62 63 64 65		66 67 68 69 70 71 72							

FIGURE A4. TRUCK/TAXI TRIP CARD FORM

Figure 4

TRUCK AND TAXI TRIP REPORT CONTINUATION

Sample Number: 15 16 17 18 19 Page of

1		2		3		4		5		6		7		
Trip Number		Freight or Vehicle		Commodity		Where Did This Trip Begin (Origin)		Where Did This Trip End (Destination)		Time of		Purpose of Trip		
										Start		Arrival		
43	44	45	46	47	48	49	☐ Empty		Street Address	Street Address	50	51	52	53
							City or Town	City or Town						
							Land Use	Land Use						
50	51	52	53	54	55	56	☐ Empty		Street Address	Street Address	57	58	59	60
							City or Town	City or Town						
							Land Use	Land Use						
57	58	59	60	61	62	63	☐ Empty		Street Address	Street Address	64	65	66	67
							City or Town	City or Town						
							Land Use	Land Use						
64	65	66	67	68	69	70	☐ Empty		Street Address	Street Address	71	72	73	74
							City or Town	City or Town						
							Land Use	Land Use						
71	72	73	74	75	76	77	☐ Empty		Street Address	Street Address	78	79	80	81
							City or Town	City or Town						
							Land Use	Land Use						
78	79	80	81	82	83	84	☐ Empty		Street Address	Street Address	85	86	87	88
							City or Town	City or Town						
							Land Use	Land Use						
85	86	87	88	89	90	91	☐ Empty		Street Address	Street Address	92	93	94	95
							City or Town	City or Town						
							Land Use	Land Use						
92	93	94	95	96	97	98	☐ Empty		Street Address	Street Address	99	100	101	102
							City or Town	City or Town						
							Land Use	Land Use						

FIGURE A5. TRUCK/TAXI TRIP CARD CONTINUATION FORM

Figure 1A

INTERNAL TRIP REPORT

Page 1 of 1

CARD NUMBER	2	T	COST ACCOUNT	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740	741	742	743	744	745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800	801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820	821	822	823	824	825	826	827	828	829	830	831	832	833	834	835	836	837	838	839	840	841	842	843	844	845	846	847	848	849	850	851	852	853	854	855	856	857	858	859	860	861	862	863	864	865	866	867	868	869	870	871	872	873	874	875	876	877	878	879	880	881	882	883	884	885	886	887	888	889	890	891	892	893	894	895	896	897	898	899	900	901	902	903	904	905	906	907	908	909	910	911	912	913	914	915	916	917	918	919	920	921	922	923	924	925	926	927	928	929	930	931	932	933	934	935	936	937	938	939	940	941	942	943	944	945	946	947	948	949	950	951	952	953	954	955	956	957	958	959	960	961	962	963	964	965	966	967	968	969	970	971	972	973	974	975	976	977	978	979	980	981	982	983	984	985	986	987	988	989	990	991	992	993	994	995	996	997	998	999	1000	1001	1002	1003	1004	1005	1006	1007	1008	1009	1010	1011	1012	1013	1014	1015	1016	1017	1018	1019	1020	1021	1022	1023	10
-------------	---	---	--------------	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	----

APPENDIX B GEOGRAPHIC CODING INDICES

The following figures are photo-reductions of sample pages of the two sets of geographic coding indices actually used, one from Evansville and one from Lafayette. The street address coding index is used for determining the zone number associated with a given numeric street address (e.g., 2208 Arapahoe Drive), and the name-place coding index is used for determining the zone number associated with some location commonly known by its name (e.g., Jefferson High School).

COLLEGE	HWY	600 898	601 899	485 485
COLONIAL	AVE	200 798	201 799	234 236
COLONIAL GARDENS	RD	2700 3098	2701 3099	303 303
S COLONY	RD	700 998	701 999	342 342 242
N COLONY	RD	1900 1998	1901 1997	357 357
COLORADO	AVE	500 798	501 799	416 416
COLORADO	AVE	1000 1098	1001 1099	417 417
E COLUMBIA	ST	2 198	1 199	413 426
E COLUMBIA	ST	200 398	201 399	425 426
E COLUMBIA	ST	400 798	401 799	425 427
E COLUMBIA	ST	800 964	801 965	424 431
E COLUMBIA	ST	966 1198	967 1199	432 431
E COLUMBIA	ST	1200 1498	1201 1499	433 433
E COLUMBIA	ST	1700 1998	1701 1999	442 442
E COLUMBIA	ST	2000 2298	2001 2299	443 443
W COLUMBIA	ST	2 398	1 399	410 410
W COLUMBIA	ST	400 630	401 699	409 409
W COLUMBIA	ST	632 698	701 713	407 409
W COLUMBIA	ST	700 798	715 799	407 404
W COLUMBIA	ST	801 1198	801 1199	407 407
W COLUMBIA	ST	1200 1498	1201 1499	406 406
W COLUMBIA	ST	2000 2398	2001 2399	167 167
W COLUMBIA	ST	3000 3098	3001 3099	162 162
W COLUMBIA	ST	3900 4198	3901 4199	163 163
COMMERCE	AVE	4400 4498	4401 4499	229 229
CONCORD	BLVD	200 798	201 799	234 230
CONGRESS	AVE	2 598	1 599	344 344
N CONGRESS	AVE	200 1298	201 1299	356 356
CONLIN	AVE	1000 1198	1001 1199	465 465
CONLIN	AVE	1300 1398	1301 1399	467 467
CONLIN	AVE	1400 1998	1401 1999	466 466

FIGURE B1. EVANSVILLE STREET ADDRESS CODING INDEX

ANNS BEAUTY SALON	3102 WASH AVE	473
ANSCO INC		462
ANSWER PHONE SER INC		502
ANTHONY C ELPERS		139
ANWAY HOME PRODUCTS		162
AP CLEANING CENT SALES		221
APARTMENT VILLAGE		475
APPELS GROOMING SALON		455
AQUANESS N CHEM		348
ARAB TERMITE&PEST CONTROL	522 COURT BLDG	502
ARAB TERMITE&PEST CONTROL	1001 SOUTHFIELD RD	348
ARAMA PLASTICS INC		422
ARBETTER AGY		470
ARBUITHNOT SCH OF DANCE		468
ARRYS ROAST BEEF RESTAUR		341
ARC-LANES		346
ARCOFF CABINET SHOP		474
ARHELGERS ROOFING CO		502
ARKEYEX SALES CORP		236
ARKIA		429
ARMSTRONG BAR		167
ARMSTRONG LBR CO		465
ARMSTRONGS BAR		167
ARMY RECRUITING STA		504
ARMY RESERVE TRAIN CENTR		444
ARNOLD HE CO		502
ARNOLD PALMER CLEANERS		340
ARNOLDS BIG A BARBER SHOP		417
ARONCO-FIRELINE FIRE&SAF		432

FIGURE B2. EVANSVILLE PLACE-NAME CODING INDEX

STREET NAME	House Number		ODD		EVEN	
	FROM	TO	Zone	Bl.	Zone	Bl.
THOMPSON AVE.	601	631	202	06		
THOMPSON AVE.	—	—	202	07		
THOMPSON AVE.	—	—			202	04
THOMPSON ST.	1701	1719	201	67		
THOMPSON ST.	1700	1720			201	60
THOMPSON ST.	1801	1819	201	66		
THOMPSON ST.	1822	1822			201	61
THOMPSON ST.	1903	1921	201	65		
THOMPSON ST.	1902	1928			201	62
THOMPSON ST.	2003	2023	201	64		
THOMPSON ST.	2002	2020			201	63
THOMPSON ST.	2101	2307	205	50		
THOMPSON ST.	—	—			205	50
TINKLER ST.	205	219	200	27		
TINKLER ST.	200	224			200	26
TINKLER ST.	305	—	200	25		
TINKLER ST.	401	409	200	24		
TINKLER ST.	400	424			200	23
TINKLER ST.	—	—	202	42		
TINKLER ST.	518	534			202	40

FIGURE B3. LAFAYETTE STREET ADDRESS CODING INDEX

APPENDIX C ERROR CHECKING SPECIFICATIONS AND RESULTS

For each of the travel surveys, the various error checks made in the conduct of this research project are specified in tables which follow. A specific error number was associated with each of the error checks, and accompanying tables show the number of cards failing the given error check in both Evansville and Lafayette and the associated error rates (per cent of the total cards in which the given error check was made that were in error).

Table C1. Home Interview Survey Error Checks

<u>Error Number</u>	<u>Cause</u>
A. Single Field Checks	
1.	Invalid card number.
2.	Invalid sample number.
3.	Invalid origin zone number.
4.	Invalid destination zone number.
5.	Invalid origin land use (blanks not permitted).
6.	Invalid destination land use (blanks not permitted).
7.	Invalid from purpose.
8.	Invalid to purpose.
9.	Invalid zone of residence.
10.	Invalid person number.
11.	Invalid trip number.
12.	Invalid travel date.
13.	Invalid start time.
14.	Invalid arrive time.
15.	Invalid sex and race code.
16.	Invalid age.
17.	Invalid mode.
18.	Invalid driver's license code.
19.	Invalid number in car.
20.	Invalid kind of parking.
21.	Invalid first work flag.
22.	Invalid total number of persons living at the D.U.
23.	Invalid structure type code.
24.	Invalid own/rent flag.
25.	Invalid number of cars garaged at the D.U.
26.	Invalid number of persons regularly employed.
27.	Invalid number of persons regularly enrolled in school.
28.	Invalid interview status code.
29.	Invalid total number of trips.
30.	Invalid number of auto driver trips.
31.	Invalid number of persons over four years of age.
32.	Invalid number of persons making trips.
33.	Invalid number of persons not making trips.
34.	Invalid number of persons with unknown trips.
35.	Invalid length of residence.
36.	Invalid household income code.
37.	Invalid code for race of occupants.
38.	Invalid parking address.
39.	Invalid parking duration, rate, or fee.
B. Consistency Checks	
40.	Home coded for both from purpose and for to purpose.

Table C1 continued.

<u>Error Number</u>	<u>Cause</u>
41.	Arrival time erroneously preceeds start time.
42.	Inconsistency between from purpose, residence zone, and origin zone.
43.	Inconsistency between to purpose, residence zone, and destination zone.
44.	Both trip ends external to the study area.
45.	First work/school trip flag inconsistent with to purpose.
46.	Inconsistency between mode and kind of parking information.
47.	Inconsistency between mode and number in car information.
48.	Total number at trips less than number of auto driver trips.
49.	Number of persons over four years do not correspond.
50.	Total number of persons less than number over four years of age.
51.	Number of persons employed and in school exceeds total number.
52.	Total number of trips less than number of persons making trips.
53.	Inconsistency between status of interview and trip information.
54.	Inconsistency between mode and parking address, duration, rate, or fee information.
55.	Trips for a given person not in sequence with respect to time.
56.	Differences in tripmaker information for a given person.
57.	Trips made by a given person not numbered consecutively.
60.	Differences in travel date and/or residence zone within a given sample.
61.	Total number of trips and/or number of auto driver trips incorrect.
62.	Number of persons actually making trips less than recorded number.
63.	Incorrect number of dwelling unit summary cards for a given sample.

Table C2. Home Interview Survey Error Rates

Error Number	Evansville		Lafayette	
	Cards in Error	Per Cent in Error	Cards in Error	Per Cent in Error
1	0	0.00	0	0.00
2	42	0.08	21	0.05
3	157	0.34	341	0.89
4	18	0.04	284	0.74
5	150	0.32	---	---
6	136	0.29	---	---
7	2	0.00	39	0.10
8	1	0.00	49	0.13
9	26	0.05	167	0.39
10	1	0.00	15	0.04
11	0	0.00	24	0.06
12	110	0.21	3,741	8.63
13	89	0.19	329	0.86
14	90	0.19	424	1.11
15	8	0.02	---	---
16	8	0.02	---	---
17	11	0.02	151	0.39
18	7	0.02	---	---
19	0	0.00	3,813	9.94
20	0	0.00	0	0.00
21	24	0.05	57	0.15
22	14	0.22	5	0.10
23	10	0.16	2	0.04
24	13	0.21	4	0.08
25	11	0.18	2	0.04
26	10	0.16	2	0.04
27	10	0.16	4	0.08
28	10	0.16	14	0.28
29	10	0.16	3	0.06
30	10	0.16	3	0.06
31	11	0.18	1	0.02

Table C2. continued

Error Number	Evansville		Lafayette	
	Cards in Error	Per Cent in Error	Cards in Error	Per Cent in Error
32	11	0.18	2	0.04
33	11	0.18	2	0.04
34	10	0.16	2	0.04
35	11	0.18	17	0.34
36	20	0.32	3	0.06
37	19	0.30	---	---
38	---	---	2,770	7.22
39	---	---	2,855	7.44
40	4	0.01	20	0.05
41	102	0.22	335	0.87
42	344	0.74	461	1.20
43	271	0.59	196	0.51
44	18	0.04	12	0.03
45	18	0.04	120	0.31
46	345	0.75	9,478	24.71
47	339	0.73	4,283	11.16
48	2	0.03	7	0.14
49	180	2.87	282	5.64
50	6	0.10	7	0.14
51	152	2.42	38	0.76
52	19	0.30	29	0.58
53	24	0.38	123	2.46
54	---	---	0	0.00
55	768	1.66	851	2.22
56	504	1.09	---	---
57	878	1.90	1,087	2.83
60	1,123	2.14	694	1.60
61	285	4.54	499	9.98
62	59	0.94	141	2.82
63	44	0.70	26	0.52

Table C2 continued

D.U. Summary Cards	6,277	4,998
Internal Trip Cards	46,308	38,364
Total Cards	52,585	43,362

Table C3. Partial List of Other Feasible (But Not Made)
Error Checks of Home Interview Survey Data

1. If either the from or to purpose is serve passenger, mode must be either auto driver or motorcycle.
2. Each trip having either a from or to purpose of serve passenger (or one of two related, consecutive trips having a common trip end purpose of serve passenger) must have more than one coded for number in car.
3. The first work/school trip flag must be set on and only on the lowest number trip (for a given person) with an appropriate destination purpose.
4. If mode is walk, then to purpose must be work (or school, if appropriate).
5. Kind of parking code should be compatible with the type of parking available at the destination/parking zone.
- 6.* Kind of parking, parking rate, and parking fee information should be consistent.
- 7.* Parking fee should be reasonable.
- 8.* If driver's license code is yes, tripmaker's age should equal or exceed the minimum legal driving age.
- 9.* If driver's license code is yes, tripmaker's age should not exceed some arbitrary maximum value.
- 10.* If mode is auto driver or motorcycle, driver's license code should be yes.
- 11.* Land use at the home should be residential.
- 12.* On consecutive trips by a given person, zone number, purpose, and land use should match.
- 13.* The first trip of the day for a given person should originate either at home or external to the study area.
- 14.* The last trip of the day for a given person should terminate either at home or external to the study area.
- 15.* The highest person number in a given sample should not exceed the total number of persons living at the dwelling unit.
- 16.* Length of trip (time) should not exceed some arbitrary maximum value.
- 17.* The race code from the dwelling unit summary data should be consistent with the consensus of race codes from the internal trip report data.
- 18.* On a sequence of trips away from home for a given person, mode should not change from auto driver to some other mode or vice versa.
- 19.* If the trips made by two or more persons within a given sample parallel each other, then they should agree in particulars such as mode, start time, arrive time, land use, etc. (this particular error check would be exceedingly difficult to implement on a computer).
- 20.* Number of persons under five years of age living at a given dwelling unit should not exceed some arbitrary maximum.

* Exceptions are possible.

Table C4. External Survey Error Checks

<u>Error Number</u>	<u>Cause</u>
A. Single Field Checks	
1.	Invalid card number.
2.	Invalid station number.
3.	Invalid origin zone number.
4.	Invalid destination zone number.
5.	Invalid origin land use (blanks permitted).
6.	Invalid destination land use (blanks permitted).
7.	Invalid from purpose.
8.	Invalid to purpose.
9.	Invalid number of persons.
10.	Invalid hour of interview.
11.	Invalid direction.
12.	Invalid vehicle type code.
13.	Invalid number of stops.
14.	Invalid entry station.
15.	Invalid exit station.
16.	Invalid vehicle base code.
B. Consistency Checks	
30.	Home coded for both from purpose and for to purpose.
31.	Internal origin zone on an inbound trip.
32.	Entry station not blank on an inbound trip.
33.	Internal destination zone on an outbound trip.
34.	Exit station not blank on an outbound trip.
35.	Entry station same as interview station.
36.	Exit station same as interview station.
37.	Exit station blank on an inbound trip with an external destination.
38.	Exit station not blank on an inbound trip with an internal destination.
39.	Entry station blank on an outbound trip with an external origin.
40.	Entry station not blank on an outbound trip with an internal origin.
41.	Inconsistency between from purpose, vehicle base, direction, and origin zone.
42.	Inconsistency between to purpose, vehicle base, direction, and destination zone.
43.	Invalid hour of interview for given station.
44.	Inconsistency between vehicle type and commodity codes.

Table C5. External Survey Error Rates

Error Number	Evansville		Lafayette	
	Cards in Error	Per Cent in Error	Cards in Error	Per Cent in Error
1	0	0.00	0	0.00
2	42	0.10	232	0.47
3	809	1.99	507	1.03
4	795	1.95	312	0.63
5	52	0.13	---	----
6	61	0.15	---	----
7	166	0.41	57	0.12
8	189	0.46	56	0.11
9	199	0.49	99	0.20
10	30	0.07	49	0.10
11	65	0.16	433	0.88
12	120	0.29	102	0.21
13	665	1.63	152	0.31
14	40	0.10	253	0.51
15	36	0.09	32	0.06
16	225	0.55	80	0.16
30	131	0.32	111	0.22
31	12	0.03	143	0.29
32	34	0.08	248	0.50
33	7	0.02	69	0.14
34	29	0.07	16	0.03
35	2	0.00	20	0.04
36	2	0.00	14	0.03
37	124	0.30	200	0.40
38	24	0.06	50	0.10
39	113	0.28	161	0.33
40	11	0.03	125	0.25
41	1,241	3.05	305	0.62
42	1,002	2.46	331	0.67
43	17	0.04	8	0.02
44	---	----	445	0.90

Total Cards 40,721

49,439

Table C6. Truck/Taxi Survey Error Checks

<u>Error Number</u>	<u>Cause</u>
A. Single Field Checks	
1.	Invalid card number.
2.	Invalid sample number.
3.	Invalid origin zone number.
4.	Invalid destination zone number.
5.	Invalid origin land use (blanks not permitted).
6.	Invalid destination land use (blanks not permitted).
7.	Invalid from purpose.
8.	Invalid to purpose.
9.	Invalid garaged zone.
11.	Invalid trip number.
12.	Invalid travel date.
13.	Invalid start time.
14.	Invalid arrive time.
15.	Invalid business/industrial classification code.
16.	Invalid licensed capacity code.
17.	Invalid vehicle type code.
18.	Invalid interview status code.
19.	Invalid total trips reported.
20.	Invalid total stops.
21.	Invalid total miles traveled.
22.	Invalid commodity code.
23.	Invalid license number.
24.	Invalid number in vehicle.
25.	Invalid block number.
B. Consistency Checks	
30.	Inconsistency between total trips and interview status fields.
31.	Inconsistency between trip information and interview status field.
32.	Total stops less than total trips.
33.	Home coded for both from purpose and for to purpose.
34.	Arrival time erroneously preceeds start time.
35.	Inconsistency between from purpose, garaged zone, and origin zone.
36.	Inconsistency between to purpose, garaged zone, and destination zone.
37.	Both trip ends external to the study area.
50.	Total number of trips incorrect.
51.	Trips not numbered consecutively.
52.	Trips not in sequence with respect to time.
53.	Differences in sample information (garaged zone, travel date, etc.)

Table C7. Truck Survey Error Rates

Error Number	Evansville		Lafayette	
	Cards in Error	Per Cent in Error	Cards in Error	Per Cent in Error
1	0	0.00	0	0.00
2	0	0.00	1	0.02
3	13	0.16	79	1.44
4	9	0.11	68	1.24
5	44	0.55	---	---
6	28	0.35	---	---
7	4	0.05	6	0.11
8	5	0.06	5	0.09
9	53	0.66	44	0.80
10	0	0.00	0	0.00
11	3	0.04	1	0.02
12	12	0.15	3	0.05
13	18	0.22	16	0.29
14	19	0.24	22	0.40
15	81	1.01	101	1.84
16	100	1.25	7	0.13
17	76	0.95	50	0.91
18	0	0.00	11	0.20
19	6	0.07	228	4.16
20	159	1.98	229	4.18
21	10	0.12	228	4.16
22	11	0.14	6	0.11
23	91	1.14	38	0.69
24	0	0.00	0	0.00
25	---	---	2	0.04
30	5	0.06	51	0.93
31	5	0.06	62	1.13
32	154	1.92	22	0.40
33	1	0.01	1	0.02
34	12	0.15	13	0.24
35	125	1.56	170	3.10

Table C7 continued

Error Number	Evansville		Lafayette	
	Cards in Error	Per Cent in Error	Cards in Error	Per Cent in Error
36	125	1.56	179	3.27
37	3	0.04	4	0.07
50	19	0.24	241	4.40
51	713	8.90	104	1.90
52	114	1.42	90	1.64
53	160	2.00	2	0.04

Total Cards 8,012

5,481

Table C8. Taxi Survey Error Rates

Error Number	Evansville		Lafayette	
	Cards in Error	Per Cent in Error	Cards in Error	Per Cent in Error
1	0	0.00	0	0.00
2	0	0.00	0	0.00
3	2	0.20	2	1.16
4	2	0.20	2	1.16
5	11	1.11	---	---
6	1	0.10	---	---
7	0	0.00	0	0.00
8	0	0.00	0	0.00
9	0	0.00	0	0.00
10	0	0.00	0	0.00
11	0	0.00	0	0.00
12	64	6.46	0	0.00
13	1	0.10	12	6.94
14	3	0.30	27	15.61
15	64	6.46	0	0.00
16	76	7.68	39	22.54
17	64	6.46	0	0.00
18	0	0.00	0	0.00
19	0	0.00	0	0.00
20	0	0.00	0	0.00
21	0	0.00	0	0.00
22	0	0.00	0	0.00
23	0	0.00	0	0.00
24	0	0.00	0	0.00
25	---	---	2	1.16
30	0	0.00	0	0.00
31	0	0.00	0	0.00
32	64	6.46	0	0.00
33	0	0.00	0	0.00
34	2	0.20	1	0.58
35	0	0.00	0	0.00

Table C8 continued

Error Number	Evansville		Lafayette	
	Cards in Error	Per Cent in Error	Cards in Error	Per Cent in Error
36	0	0.00	0	0.00
37	0	0.00	0	0.00
50	1	0.10	1	0.58
51	133	13.43	0	0.00
52	68	6.87	2	1.16
53	127	12.83	0	0.00
Total Cards		990	173	

Table C9. Number Of Cards Vs. Errors Per Card By Survey

Errors Per Card	Evansville			Lafayette				
	Home Interview	External	Truck	Taxi	Home Interview	External	Truck	Taxi
0	47,576	36,107	6,485	823	26,705	45,942	4,436	99
1	3,924	3,956	1,065	37	9,616	2,685	632	61
2	831	443	351	65	2,198	682	152	12
3	195	50	81	1	2,712	74	30	1
4	38	8	22	0	295	23	142	0
5	5	7	0	0	759	9	50	0
6	1	5	1	0	908	1	33	0
7	0	36	1	0	122	6	6	0
8	0	89	2	63	38	10	0	0
9	0	13	1	1	4	2	0	0
10	1	6	0	0	1	1	0	0
11	0	0	0	0	1	3	0	0
12	0	0	0	0	0	0	0	0
13	0	1	1	0	0	1	0	0
14	0	0	1	0	0	0	0	0
15	0	0	1	0	2	0	0	0
16	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0
18	1	0	0	0	0	0	0	0
19	6	0	0	0	1	0	0	0
20	3	0	0	0	0	0	0	0
Average Errors Per Card for all cards	0.125	0.153	0.272	0.690	0.780	0.093	0.380	0.509
Average Errors Per Card for those cards in error	1.310	1.353	1.426	4.090	2.031	1.318	1.994	1.189

VITA

Ralph Wiley Hines was born April 27, 1911, in Dodge City, Kansas. He received his primary education in Lakewood, Colorado, and was graduated from Lakewood High School in 1929.

He received the Bachelor of Science in Civil Engineering degree from Purdue University in 1934 and the Master of Science in Civil Engineering degree from Purdue University in 1936.

From 1934 to 1936, he was employed by the U.S. Bureau of Public Roads. He was serving as a Highway Research Engineer in their Washington, D.C. headquarters when he was transferred to the Bureau of Highway Engineering at Purdue University.

VITA

He is a member of the Chi Rho Iota civil engineering honorary society, a member of the Tau Beta Pi engineering honorary society, a member of the American Society of Civil Engineers, an Associate member of the Institute of Traffic Engineers, a Supporting member of the Highway Research Board, a member of the National Society of Professional Engineers, and a member of the Association for Computing Machinery.

A registered professional engineer in the District of Columbia and the State of Indiana, he has been consulting in the areas of transportation planning, traffic engineering, and computer applications.

He married Gloria Lee Collins in 1938 and they have one child, Jane Patricia Hines, born in 1943.

VITA

Ralph Wildy Zimmer was born April 29, 1941, in Dodge City, Kansas. He received his primary education in Lakewood, Colorado, and was graduated from Lakewood High School in 1959.

He received the Bachelor of Science in Civil Engineering degree from Purdue University in 1964 and the Master of Science in Civil Engineering degree from Purdue University in 1965.

From 1964 to 1970, he was employed by the U.S. Bureau of Public Roads. He was serving as a Highway Research Engineer in their Washington, D.C. headquarters when he resigned to undertake additional graduate study at Purdue University.

He is a member of the Chi Epsilon civil engineering honorary society, a member of the Tau Beta Pi engineering honorary society, a Member of the American Society of Civil Engineers, an Associate Member of the Institute of Traffic Engineers, a Supporting Member of the Highway Research Board, a member of the National Society of Professional Engineers, and a member of the Association for Computing Machinery.

A registered professional engineer in the District of Columbia and the State of Indiana, he has done consulting in the areas of transportation planning, traffic engineering, and computer applications.

He married Gloria Lee Goings in 1964 and they have one child, Kurt Randolph Zimmer, born in 1969.