PM-223-U5-4A

PROJECT MEMORANDUM

U.S. DEPARTMENT OF TRANSPORTATION
RESEARCH AND SPECIAL PROGRAMS ADMINISTRATION
TRANSPORTATION SYSTEMS CENTER
KENDALL SQUARE
CAMBRIDGE, MA 02142

LITERATURE REVIEW ON INDUCED EXPOSURE MODELS, TASK 2 HS-270

PETER MENGERT

FEBRUARY 1982

REVISED JUNE 1982

SPONSOR: SANDRA SMITH, NRD-31

This document contains information subject to change. This is considered an informal, technical document for working level communication and dissemination of preliminary information within the cited project. Distribution is effected by and the responsibility of the TSC Program Manager.

APPROVED FOR DISTRIBUTION: Edwin J. Roberts

Branch Chief

T. Hausler

Division Chief

INTRODUCTION

Sections 1, 2 and 3 of this report describe the development of induced exposure models, together with a discussion of questions of validity. These Sections focus on the most important and relevant results from the literature, while Appendix A contains brief reviews of all the papers, including those covered in more detail in sections 1 thru 3. Appendices B, C, and D contain technical discussions of various technical points relevant to the report. Numbers in parentheses after authors' name, or otherwise identified as reference numbers, refer to the reference list at the end of this report (after Appendix D).

Induced exposure models have been proposed to derive exposure information from accident information. Induced exposure models generally seek to estimate relative exposure. Specifically if road users are categorized into several groups labled by $i = 1, . . ., N$ then induced exposure models are for the purpose of estimating e_i, $i=1, ..., N$ where e_i represents the relative exposure of the ith road user group. The total exposure for the ith group can be written $E_i = e_i E$ where E is the total exposure over all user groups. This assumes that the e_i's sum to one over all user groups so that e_i represents the proportion of the total exposure experienced by members of group i. The road user groups referred to here are classifications by driver and vehicle characteristics. Thus the road user groups in question could be determined by a set of driver age categories or vehicle size categories or a combination of driver age and vehicle characteristics, etc.

The key question to be addressed is whether the estimates of relative exposure from certain induced exposure models are accurate. If direct comparisons are to be made between estimated and measured exposure then the appropriate measured exposure must be selected. The measure VMT is usually chosen to measure the driving exposure of road user groups. Induced exposure estimates might be more directly comparable to some other measure of driving exposure or quantity of driving (or perhaps measure of dangerous situations encountered) but at this stage of development it will be useful to compare the induced exposure estimates to VMT since VMT is the available and generally used measure of exposure. All direct comparisons to date have been with VMT estimates. Consequently one may take the e_i as estimates of relative VMT. Of course if some other measure of quantity of driving is available the comparison of it with induced exposure estimates will be of interest.

SECTION 1: THORPE'S INTRODUCTION OF INDUCED EXPOSURE

The first paper on induced exposure was that of Thorpe (1). Thorpe developed an exposure estimate from accident data and used it as an intermediate quantity to derive a "relative accident likelihood" or R.A.L. in Thorpe's terminology.

In his article the R.A.L. was the quantity of most interest to Thorpe and its properties were studied using Australian accident experience. Thorpe noted the agreement of the R.A.L. with intuition and general experience.

From the point of view of general induced exposure models the key product of Thorpe's development was an estimate of relative exposure by driver group obtained solely from the proportion of single car accidents and the proportion of two car accidents which each group accounted for.

Thorpe compared the R.A.L. to a similar ratio with relative proportion of licensed drivers in the group as the substitute for an exposure measure and concluded that the latter measure agreed well with the R.A.L. These observations were of great interest to Thorpe but are not very satisfactory for present day use since the number of licensed drivers in a group is not a satisfactory measure of the relative exposure of the group.

In general, Thorpe's original article only contained very indirect evidence as to the empirical validity of his model. Although his model yields an induced exposure estimate, Thorpe did not compare this estimate with any other estimate of exposure.

Thorpe noted the importance and difficulty of getting relative exposure estimates for road user groups and this provided the motivation for his work. Incidentally Thorpe makes it clear that he equates exposure to VMT as a practical matter.

Thorpe's model will be discussed further in relation to other models but certain main points may be made now.

The main result in terms of induced exposure models may be stated as follows:

Let D be the proportion of two car accident involvements pertaining to a group of drivers (or other road users). Let S be the proportion of single car accidents which pertain to the group and let E be the proportion of total exposure pertaining to the group. Then D and S are known from accident data while E is unknown, hard to estimate yet of great interest. Thorpe's model states that E is well approximated by the following expression:

$$E = 2D - S$$

In order to derive this model, Thorpe introduced a number of clearly stated hypotheses. In this brief review only these key hypotheses will be stated:

1. The proportion of "not responsible" involvements pertaining to each group is equal to the proportion of total exposure pertaining to that group. (Thorpe's assumption (e).)

2. The proportion of "responsible" involvements in two car collisions pertaining to each group is equal to the proportion of single car accidents pertaining to that group. (Thorpe's assumption (d).)

The proportion referred to in the first hypotheses is E and that in the second hypotheses is S so the two hypotheses together then lead to the equation leading to the induced exposure model.

$$D = 1/2\ (E+S)$$

The second hypothesis (Thorpe's assumption (d)) seems to have drawn the most criticism as will be seen in the discussion of other papers.

SECTION 2: QUASI-INDUCED EXPOSURE AND RECOGNITION OF PROBLEMS WITH THORPE'S FORMULATION

Several authors in the late 60's and early 70's studied induced exposure using models different from Thorpe's but based on similar concepts. Several papers dealt with models based on assigned responsibility (these have also been called quasi-induced exposure models). Papers by Carr (2), Hall (3), Carlson (4) and Cerrelli (9) considered induced exposure based on this approach. In each of these papers exposure was estimated (relatively) as the proportion of non-responsible involvements in two car accidents in a group. When this exposure estimate is divided into the proportion of responsible involvements a quantity called the "Relative Risk" by Carr, the "Hazard Index" by Cerrelli and the "Overinvolvement Ratio" by Carlson results.

As in Thorpe's investigation, this ratio (which is analogous to Thorpe's R.A.L.) was examined to see if its behavior as a function of age and sex and other driver and vehicle characteristics agreed with intuition and general experience. Again there was little specific attention given to the exposure estimates themselves and in particular the exposure estimates were not generally compared to estimates obtained otherwise, nor generally were any quantities materially dependent on the exposure estimate (e.g., the Relative Risk) compared to any directly comparable observation data.

An exception was the calculation by Cerelli of numerous exposure estimates. Wass reports that these were subsequently compared to exposure observations with favorable results. These results were obtained by Wass from an OECD report (this report has not yet been obtained by TSC).

Carr's paper is of particular interest because in it he offered evidence contrary to Thorpe's hypothesis (d) that the proportion of each group in the single accidents equals the proportion of the group as responsible party in two car collisions. Thorpe's hypothesis is not supported by Carr's data in which actual responsibility has been assigned in two car collisions. Responsibility was assigned to drivers given police citations or who struck stopped vehicles. Only accidents in which one driver was assigned full responsibility in this way were included in the analysis.

Carr conceded that the problem might be in the assignment of responsibility by the police. However, Thorpe had already alluded to the possibility that older drivers are involved in relatively more collisions than in single car accidents and that this might explain why the R.A.L. does not start to increase again (in the Thorpe Australian data) for elderly drivers, as expected.

Carr observed (in his Toronto data) that older drivers had a much larger proportion of responsible involvements in collision accidents than their proportion of single car accidents. This is in direct contrast to one of Thorpe's key hypotheses (assumption (d)).

In examining Ann Arbor, Michigan data Hall and Carlson agreed with the conclusion that drivers over 55 have a relatively higher representation of responsible drivers than non-responsible drivers, in agreement with Thorpe's expectation, but in disagreement with his R.A.L. estimate. This further suggests problems in Thorpe's modelling hypothesis.

Wass (15) has applied the Thorpe model and the Koornstra model (to be discussed at length in the next section) to Danish accident data. Wass concluded that the Thorpe model was similar to the Koornstra model in its conclusions. In Appendix C it is shown that the Thorpe model is related to the Koornstra model and in a sense consistent with it. Wass accepted the conclusion, based on the models he tested, that overinvolvement in collisions does not start to increase sharply at advanced ages and disputed Carr's statement that his data disagree with Thorpes R.A.L.

Apparently, the key point is that, in collision accidents, the overinvolvement ratio increases after 50 to 55 but the ratio containing single vehicle accidents in the numerator continues to decrease. This was noted by Carr and was one of his chief points. Wass points out that the single vehicle ratio computed by Carr is in fair agreement with the R.A.L. computed by Thorpe. That doesn't, however, clear up the problem that Carr's responsibility data is at odds with Thorpe's modelling hypothesis assumption (d).

In summary, Wass' observations not withstanding, there appears to be strong evidence, developed by Carr and others, that the Thorpe hypothesis in question does not hold up when examined in comparison to accident data where responsibility has been assigned. This does not close the question of the validity of the Thorpe model or of the relative accuracy of the Thorpe model compared to assigned responsibility models.

Waller, et al (1973) (Reference 13) reported on some North Carolina data which provided further confirmation of Carr's observations. This paper is discussed in more detail in Appendix A.

SECTION 3: THE KOORNSTRA MODEL

Koornstra in 1973 (References 11 and 12, see also Wass, Reference 15) introduced an induced exposure model which makes much more detailed use of accident data and which suggests much about the detailed nature of assumptions needed in induced exposure models.

The Koornstra model is rich enough in structure to suggest several generalizations and extensions.

Koornstra applied his model to some accident data for the Netherlands and concluded that the fit was rather satisfactory. There was some lack of fit in the diagonal cells which he attributed to modelling problems and suggested the model not be fit to diagonal cells. However, he had apparently misinterpreted his own model in a very simple way that led to this lack of fit. The misinterpretation disappears in Wass' book along with any problems with lack of fit with the diagonal elements. The interpretation question is discussed in Appendix B where the Wass formulation is given. The Wass formulation is most convenient. Appendix B also contains a general statement of the Koornstra model.

Koornstra further suggested that single vehicle accidents be treated as two vehicle accidents where the second vehicle belongs to a fictitious "dummy" category (Koornstra's basic model **can be** formulated initially for two car collisions only). When this assumption is made, Koornstra's model comes into the realm of Thorpe's shaky hypothesis. However the Koornstra model makes such rich use of the data that the failure of that hypothesis should have much less effect on the results than it will on those of Thorpe's model. If necessary, single vehicle accidents can be excluded from Koornstra's model. The connection with Thorpe's model when single vehicle accidents are included is rather interesting. In Appendix C it is shown that Thorpe's model is a sort of specialization of Koornstra's model in that case. A model suggested by Engel is discussed there and shown to be in a sense, intermediate between the Thorpe and Koornstra models. The sense in which Thorpe's model is a specialization of the Koornstra model (with single vehicle accidents) has to do with the fact that if certain estimates in the Koornstra model are identified with the quantity they estimate, the Thorpe model results. The imposition of certain estimates or approximations as identities is equivalent to the imposition of certain constraints.

Koornstra's model provides a rather general framework in which to view induced exposure. As noted Thorpe's model is related (and in a sense the two models are consistent) and this relation sheds more light on Thorpe's model.

Koornstra's derivation of his model is somewhat difficult to follow completely and seems to contain gaps that are not identified. Wass' derivation of Koornstra's model closely follows Koornstra's and presents the same difficulties. Nevertheless Koornstra's formulation helps clarify some aspects of induced exposure models. In particular it seems that Koornstra's formulation is so much more detailed than previous induced exposure models that it permits a much more detailed identification of the problems involved. (Note that each potential defect to be discussed may be present or absent in a particular situation.) The potential problems that may be identified are:

1. One of the most troublesome potential problems was noted by Haight (6,7) and called "imperfect mixing" by him. It relates to the fact that two user groups (driver or vehicle classes) may not have their exposure distributed identically over time or roadway types. For example, one group may get more exposure at night than another group. This leads to two separate problems:

   a. One of the groups may be exposed to more hazardous driving conditions than the other, i.e., their exposure is not of the same quality.

   b. The interactions between the groups (i.e., collisions between the groups) will be overestimated by their exposure and proneness parameters since their interaction is limited by a factor not in the model.

   It may be pointed out that the model should reflect more hazardous exposure as a greater amount of exposure, so that problem (a) would not in itself appear to prevent the model's validity. Problem (b) is a real problem to be recognized and dealt with e.g., by disaggregating (i.e., solving separately) over time and roadway type, to the extent possible, to diminish the effect of problem (b) (this practice will also mitigate problem (a)). As an example, since some groups get more exposure at night than during the day, the Koornstra model may be solved separately for day and night data. On the other hand, disaggregation may lead to data which is too thin i.e.-where the accident matrix A_{ij} (number of accidents between groups i and j) required by the Koornstra model has elements or cells which contain too few accidents. It is well known from accident studies that accident counts (like counts in many other situations) exhibit variances which are at least as large as if they were Poisson random variables and often larger. This means that a good lower bound on the standard deviation of a count is its square root. Consequently, if the square root of a cell count is substantial in comparison to the count itself the error in the count is likely to be substantial also. If the disaggregation is carried out to a sufficient degree the cell counts will be so small that they will be substantially in error and so the estimates from the Koornstra model may not correspond to the actual quantities being estimated.

2. A second problem with the Koornstra model is the fact that it models accident fault as pertaining to one party or the other or to neither party but not to both.*

An extra term can be added to the Koornstra model in an attempt to correct this error. The effect of this term is discussed in Appendix D. It turns out that if the added term is rather small the problem disappears.

----

*Note that the Koornstra model may be considered to refer to fault as an intermediate concept. This is for the purpose of developing the rationale of the model. Also the "proneness" or "cross-section" values P_i may be thought of as a measure of fault. They are outputs of the model. The inputs to the model do not include any assessment of fault. (See also Appendix B) All these statements apply as well to the Thorpe model but do not apply to the quasi-induced exposure models (e.g. Carr [2]) which do require an assessment of fault as model input.

3. A third problem is that the Koornstra model looks upon each accident situation as symmetric in the user groups in that the same potential accident situation would be as likely to occur if the roles of the user goups were interchanged: this assumption is clearly incorrect for certain combinations of vehicle groups (for example, vehicles of greatly different size). The main remedy here is to be aware of the possible problem and avoid classifications which would lead to very unsymmetric accident situations.

4. The fourth major problem identified here is the problematical assumption that the same proneness distribution for a user group applies in all situations. This assumption is an extension to the general Koornstra context of Thorpe's hypothesis which lead to difficulty. It is suggested that this may cause problems mostly in the Thorpe case, i.e., when two car accidents are considered in comparison with single car accidents and may not cause so much of a problem in other cases.

Whatever the shortcomings and drawbacks of the Koornstra formulation, it has been tested rather extensively by Wass on Danish data and found to be quite satisfactory in predicting exposure. The proneness values have also been found by Wass to agree rather well with their counterparts in observed data namely accident rates as determined by separate means. A complete test of induced exposure models requires an extensive test of the Koornstra model and some extensions of it.

APPENDIX A: BRIEF REVIEWS OF PAPERS (AND A BOOK) ON INDUCED EXPOSURE

Thorpe, John (Reference 1, 1977)

The original paper on induced exposure, is discussed at length in Section 2 above.

Carr, Brian (Reference 2, 1969).

A rather thorough analysis of an assigned responsibility model applied to Toronto data. This paper disputed one of Thorpe's major hypotheses. Responsibility was assigned to one driver, if possible, either because of a police citation or because the responsible driver's car struck a stationary vehicle. The non-responsible involvements provide a measure of exposure and the ratio of responsible to non-responsible involvement provide an overinvolvement ratio.

Hall, William K. (Reference 3, 1970)

An analysis of Ann Arbor, Michigan accident data along the same lines as Carr's. Primary conclusions include confirmation of Carr's observations that the overinvolvement ratio increases above the age of 55. Also, concluded that significant differences in overinvolvement ratios between vehicle types were difficult to detect.

Carlson, William L. (Reference 4, 1970)

Similar to the Hall paper, also analyzed Ann Arbor, Michigan data. Carlson also reported that older vehicles are overinvolved as responsible vehicles (have a relatively large overinvolvement ratio).

van der Zvaag, Donald D. (Reference 8, 1971)

This is another paper from the HIT Lab (U. Michigan Highway Safety Research Institute) group from which the Hall and Carlson papers come. It agrees in the major conclusion that the overinvolvement ratio increases for older drivers. There is also a study of trucks which concludes that trucks are "overinvolved in reportable accidents in Michigan's Oakland County."

Haight, Frank A. (References 5, 6 and 7, 1970, 1971 and 1973)
(Three closely related papers)

Haight suggested a particular model in his 1970 paper. The 1971 and 1973 papers present the 1970 model together with a revised model. The first Haight model is based on premises which are easily seen to be incorrect under certain circumstances. The second model corrects the obvious conceptual flaw but does so in a manner which does not appear to be well founded. There is no comprehensive application of the model to actual data in these papers. A comprehensive application was undertaken by Wass, who concluded that Haight's corrected model was unsatisfactory. Wass also seems to have reservations about Haight's first model since it does not agree with Koornstra's model (on the Danish data) as well as Thorpe's model does.

Haight's 1971 and 1973 papers include a good discussion of induced exposure models in general. Haight refers to models such as those studied by Carr, Cerrelli, etc. as quasi-induced exposure models since assignment of responsbility must be added to the pure accident data unlike the models of Thorpe, Haight and Koornstra. Haight provides some interesting insights into the Koornstra model.

Waller, Patricia F. et al (Reference 13, 1973)

Reported on some North Carolina data which yielded results in contradiction to Thorpe's assumption (d). In this paper an independent estimate of exposure was provided as well as an indication of responsibility in two car collisions. A comparison of quasi-induced exposure (relative proportion of innocent doubles) vs. independently estimated exposure by sex and seven age groups was available. The independently estimated exposure was obtained from drivers' estimates of their own VMT obtained from applicants for driver licences reporting to examining stations.

The agreement between relative exposure by the quasi-induced method and exposure as estimated by drivers was neither exceptionally good nor exceptionally bad.

In summary, this paper presents negative results on Thorpe's assumption (d) which was tested using accident data only; but the paper is nearly neutral on quasi-induced exposure which was tested using the independent exposure data.

Cerrelli, Ezio (Reference 9, 1972, 1973)

The Cerrelli paper investigates several related ratios making use of accident data and registration data to some extent. An interesting feature of Cerrelli's article is the development of a "Liability Index" for each driver group. The Liability Index was found to agree well in a proportional sense to insurance rates. This constitutes a comparison of the quantity "% responsible drivers" (needed in the Carr, Cerelli et al. formulations, or in the quasi-induced formulation) to an independent estimate of the same quantity. It is difficult to assess the significance of the agreement, but it constitutes one of the few direct tests of the quasi-induced exposure models. Another direct comparison of quasi-induced exposure models mentioned by Wass also involved the Cerrelli results. According to Wass, Cerrelli's exposure estimates were compared to independent estimates in an OECD report, and the results were favorable.

Joksch, Hans C. (Reference 10, 1973)

This paper provides a very complete discussion of accident and exposure data by sex from Hartford, Connecticut. It is an exceedingly careful analysis. Joksch, however, cautions that his conclusions are not to be taken too seriously as the data are thin (less than 100 total accidents). Joksch shows how to test models which assume that two car accidents:

1. happen randomly
2. require only one party to be at fault
3. require both parties to be at fault

This is an interesting analysis related to the Koornstra model and to the extension discussed in Appendix D.

Koornstra, Mattnijs J. (References 11, 12, 1973)

These are the primary papers on the Koornstra model discussed in length in Section 4. They contain a large amount of interesting procedural, empirical and speculative material.

Wass, Carsten (Reference 15, 1977)

The Wass book provides a very good review of induced exposure models, particularly the Koornstra model. The most valuable feature of the book is an examination of induced exposure models applied to Danish accident data (70,000 accidents were used). The results of exposure estimates from the Koornstra model were compared to independent data. The results of proneness predictions from the Koornstra model were compared to independent accident rate data. The comparisons provide impressive evidence of the validity of the model. Sensitivity tests of induced exposure models were conducted.

The book shows little inclination to examine the model critically, for example to probe for inadequacies, but rather takes a strong position of advocacy.

Nevertheless, the empirical studies are very impressive and the results given tend to justify Wass' claim that his study puts the model on a firm enough foundation to be a useful tool in traffic accident research. The theoretical portions of the book are also of impressive competence but seem to add little beyond the Koornstra original work. The careful presentation has cleared up the interpretation problem discussed in Appendix B. The book is privately published. Its value might be enhanced if it were published by a large research organization or a government agency.

Waller, Julian A. (Reference 4, 1976)

This paper discusses the need for exposure data.

APPENDIX B: REFORMULATION OF KOORNSTRA MODEL TO AVOID CERTAIN PROBLEMS

Koornstra's model as originally presented was:

$$X_{ij} = (P_i+P_j)e_ie_j$$

where X_{ij}, according to Koornstra, "is the number of accidents involving drivers of class i and j (doubles)". This interpretation leads to contradictions. (This was explained in an attachment to the work plan.)

Wass defined X_{ij} as the "number of road-users of groups (i) or (j) involved in collision accidents between these groups." The two definitions are equivalent if $i \neq j$ but the Wass X_{ii} is twice the Koornstra X_{ii}. This is precisely the factor of two proposed in the work plan to change the Koornstra model on the diagonal. However, it is neater to make the adjustment in X_{ij} so that the model retains the simple form. The Wass definition given above requires the correct interpretation if a single road user undergoes more than one accident wth members of a particular group. It must be undertstood that X_{ij} is the number of involvements in collision accidents which members of group i have with members of group j (if two members of a single group are in the same accident, then, of course, both parties, involvements are counted). This matrix is a little awkward to define but well worth the trouble. The matrix which corresponds to Koornstra's original definition is easy to define and will be denoted by A_{ij}:

A_{ij} = # of two car accidents in which one party belonged to group i and the other party to group j

(as noted this is the matrix which was denoted by X_{ij} by Koornstra but will now be denoted by A_{ij} here).

Then:

$$A_{ij} = X_{ij} \qquad \text{if } i \neq j$$

while

$$A_{ii} = 1/2\ X_{ii}$$

Then X_{ij} as defined by Wass and in agreement with the notation to be used here satisfies:

$$X_{ij} = (P_i + P_j) e_i e_j$$

while A_{ij} of course, satisfies a somewhat different model. Since X_{ij} satisfies the simpler model it is most convenient to formulate all discussions in terms of X_{ij}.

The Koornstra model goes on to assign an interpretation to the parameters on the right hand side of this equation. The e_i's are taken to be the relative exposure values discussed at the beginning of this project memorandum. The p_i's are called "proneness" values by Koornstra and have sometimes been referred to as "cross sections." Wass uses the term "liability" but this term was given a different meaning by Cerrelli. A proneness value according to the derivation could be thought of as a linear function of the probability of being in a "fault state" at the time of the accident i.e. in violation of safety rules or otherwise in a state produced by wrong action. The proneness value may be thought of as being a linear or at least increasing function of the corresponding probability of being at fault. This is hypothetical but would be strongly supported if the exposure estimates were found to be correct. Note that the term "proneness" as used here is quite distinct from the psychological or behavioral term. It refers to a model parameter which may be related to fault rate but does imply anything else about internal states of the drivers.

APPENDIX C: RELATION OF THE THORPE MODEL TO THE KOORNSTRA MODEL WITH COMMENTS ON DR. ENGEL'S SUGGESTION

In this appendix the relationship between the Thorpe model and the Koornstra model will be investigated. It will turn out that, from the point of view developed here, the model embodied in Engel's suggestion is in a sense intermediate between the Koornstra and Thorpe models.

Let the Koornstra model be expressed thus:

$$X_{ij} \approx \hat{X}_{ij} = K(P_i+P_j)e_ie_j$$

(the constant K at this point is arbitrary).

Special interest will center on single car accidents which in the Koornstra formulation are represented through the use of a special "fixed" category. This is actually a fictitious class to represent the "other party" in single car accidents. Thus if B_i denotes the number of single car accidents involving category i then in the Koornstra formulation this is represented by setting $X_{i0}=B_i$. Here 0 labels the fixed category so that X_{i0} represents the number of involvements of category i with the fixed category which is by definition equal to the number of single car accidents involving category i.

It will be convenient to introduce a new matrix y_{ij} which is a normalized form of X_{ij}.

Let:

$$y_{ij} = X_{ij} / \left(\sum_{i,j \neq 0} X_{ij} \right)$$

so that:

$$\sum_{\substack{i \neq 0 \\ j \neq 0}} y_{ij} = 1$$

Let:

$$y_{ij} \approx \hat{y}_{ij} = k(P_i + P_j) e_i e_j$$

and let

$$\sum_{i,j \neq 0} \hat{y}_{ij} = \sum_{i,j \neq 0} y_{ij} = 1. \qquad \text{Constraint 1}$$

This represents a constraint on the Koornstra model, namely $\sum_{i,j \neq 0} \hat{y}_{ij} = 1.$

This is a very reasonable constraint and should have very little effect on the solution.

As a matter of notation, let $t_i = p_i e_i$. Then we have

$$\hat{y}_{ij} = k(t_i e_j + t_j e_i).$$

With no effect on the model characteristics and with no loss of generality we may impose:

$$\sum_{i\neq 0} t_i = 1 \text{ and } \sum_{i\neq 0} e_i = 1$$

Together with $\sum_{ij=0} \hat{y}_{ij} = 1$ these imply that $k = 1/2$ and that $\hat{y}_{ij} = 1/2\ (t_i e_j + t_j e_i)$.

(In summary note that $\sum_{i,j\neq 0} \hat{y}_{ij} = 1$ is a modeling assumption while the assumption $\sum_{i\neq 0} t_i = 1$ has no modeling content but simply fixes an arbitrary normalization as does $\sum_{i\neq 0} e_i = 1$ which now fixes the undertermined constant in y_{ij}).

Recall that 0 labels the fixed category so that X_{i0} is the number of single car accidents involving category i. Since there are no accidents involving only category 0 we must have $y_{00} = 0$. This leads to the requirement that $p_0 \approx 0$ and that $t_0 \approx 0$.

The requirement $\hat{y}_{00} = y_{00} = 0$ was imposed explicitly by Koornstra and will be introduced as constraint 2.

$$\hat{y}_{00} = y_{00} \qquad \text{Constraint 2}$$

Since $\hat{y}_{00} = p_0 e_0^2$ this will lead to the constraint $p_0 = 0$ or equivalently $t_0 = 0$ and $t_0 = 0$ is another way of expressing constraint 2. (Note that $e_0 = 0$ involves dropping the 0 category and is thus meaningless.)

Since:

$$\hat{y}_{i0} = 1/2(t_0 e_i + t_i e_0)$$

$t_0 = 0$ leads to the equation

$$\hat{y}_{i0} = 1/2\ e_0 t_i$$

Now:

$$y_{i0} = Cb_i \text{ where } b_i = B_i / \left(\sum_j B_j\right) \text{ and C is a constant.}$$

(the sum is automatically over $j \neq 0$)

The next constraint to be imposed is equivalent to that suggested by Dr. Engel:

$$\hat{y}_{i0} = y_{i0} \qquad \text{Constraint 3}$$

Since $B_i = X_{i0} = cy_{i0}$, this leads to:

$$b_i = ce_0 t_i$$

where c is a constant and $b_i = B_i / (\sum_j B_j)$.

Since $\sum_{i \neq 0} b_i = \sum_{i \neq 0} t_i = 1$

this leads to: $t_i = b_i$

which is another way of expressing this constraint.

The final constraint to be imposed will lead to a derivation of Thorpe's model:

$$\sum_{j \neq 0} \hat{y}_{ij} = \sum_{j \neq 0} y_{ij} \qquad \text{Constraint 4}$$

Let:

$$R_i = \sum_{j \neq 0} y_{ij}.$$

Then since $\sum_i R_i = 1$ it follows that R_i is the portion of two car

accident involvements had by category i. Then:

$$R_i = \sum_{j \neq 0} y_{ij} = \sum_{j \neq 0} \hat{y}_{ij} = 1/2 \sum_{j \neq 0} (e_i t_j + t_i e_j) = 1/2\ e_i + 1/2\ b_i$$

or

$$e_i = 2R_i - b_i$$

and this is Thorpe's model.

Review and discussion of the constraints:

$$\sum_{i,j \neq 0} \hat{y}_{ij} = \sum_{i,j \neq 0} y_{ij} \qquad \text{Constraint 1}$$

$$\hat{y}_{00} = y_{00} \qquad \text{Constraint 2}$$

$$\hat{y}_{i0} = y_{i0} \qquad \text{Constraint 3}$$

$$\sum_{j \neq 0} \hat{y}_{ij} = \sum_{j \neq 0} y_{ij} \qquad \text{Constraint 4}$$

All these constraints are on the t_i's and e_i's saying in effect that some combination of them is exactly equal to the corresponding combination of y_{ij}.

Since it is assumed that $\hat{y}_{ij}$ approximates y_{ij} and Koornstra's model is only good to the extent this assumption is true it is reasonable to constrain that certain aspects of this model fit exactly leaving the possibility of poor fit in regard to other aspects. In general the unconstrained model is probably to be preferred for overall accuracy, but the constrained models insure giving more importance to certain aspects of the data, in particular single car accidents. Furthermore, they are easier to solve.

In addition, the Thorpe model has the recommendation that it has been observed in several studies and has a longer history than any of the induced exposure models it being the oldest. Also it has the advantages of requiring much less data than is contained in the X_{ij} matrix and instead requires only b_i and R_i. This demonstration of the connection between the models is gratifying since in reading the literature one may get the impression that there is very little connection. On the contrary: if the Koornstra model holds exactly then the Thorpe model holds exactly. If the models are only approximations (as they are) then the Koornstra model appears to be an efficient extension of the Thorpe model. Engel's model appears to be intermediate, retaining the focus on single car accidents in the Thorpe model and the ease of solution but sacrificing the lack of necessity of complex data.

APPENDIX D: NOTES ON SOME POSSIBLE GENERALIZATIONS OF THE KOORNSTRA MODEL

It was noted in the attachment to the work plan that the derivation of the Koornstra model would, if made more general, require two extra terms which were deleted by Koornstra using certain arguments. A more complete version would be:

$$X_{ij} = (p_i + p_i + \alpha p_i p_j + \beta) e_i e_j \qquad D1$$

If the α term can be ignored it is then an easy matter to transform the equation so that the β term disappears. However, if the α term remains then the situation is more complicated. Consequently, this appendix considers the more general model and makes some comments on it. One further generalization may be introduced namely:

$$X_{ij} = (p_i + p_j (1 + \gamma) + \alpha p_i p_j + \beta) e_i e_j \qquad D2$$

Then γ is 0 in a symmetric situation and γ is not 0 if the X_{ij} matrix is not symmetric as for example if X_{ij} represents involvements with i as responsible party and j as non-responsible party. The initial discussion will focus on the symmetric case, however.

The question is this: if the true model is:

$$\hat{X}'_{ij} \approx X_{ij} = (p'_i + p'_j + \alpha p'_i p'_j + \beta) e'_i e'_j \qquad \text{D3}$$

and instead one fits a model of the form

$$\hat{X}_{ij} = (p_i + p_j) e_i e_j \qquad \text{D4}$$

then how accurate are e_i and e_j as estimates of e'_i and e'_j?

Often in discussing the Koornstra model it is useful to introduce the notation (first introduced by Koornstra):

$$t_i \overset{\text{def}}{=} p_i e_i \qquad \text{D5}$$

In this notation the "true" model becomes:

$$\hat{X}'_{ij} \approx X_{ij} = (t'_i e'_j + t'_j e'_i + \alpha t'_i t'_j + \beta e'_i e'_j) \qquad \text{D6}$$

while the model to be fit is $\hat{X}_{ij} = t_i e_j + t_j e_i$ D7

A natural question to ask is if $\hat{X}_{ij}$ may actually represent a form $\hat{X}'_{ij}$, i.e., is it possible to find t_i, e_j, such that $\hat{X}'_{ij} = \hat{X}_{ij}$.

It may be shown that if

$$t_i = 1/(1+rs)(t'_i + re'_i)$$

and D8

$$e_i = e'_i + st'_i$$

then: $\hat{X}_{ij} = \hat{X}'_{ij}$

providing: $\alpha = 2s/(1+rs)$ and $\beta = 2r/(1+rs)$ D9

In other words if r and s can be chosen such that α and β (which occur in D6) are given by equation D9 and if t_i and e_i are given by equation D8 then $\hat{X}'_{ij} = \hat{X}_{ij}$ where $\hat{X}'_{ij}$ is given by equation D6 and $\hat{X}_{ij}$ is given by equation D7. This is purely a matter of formal algebra, and the equality is easy to establish. The question which remains is whether α and β (both > 0) can be expressed in this form (D9). The answer to this is that so long as $\alpha\beta \leq 1$ then α and β can be expressed in this form and $\hat{X}_{ij}$ of the form (D7) can equal $\hat{X}'_{ij}$ of the form (D6) so long as t_i and e_i are given by (D8). If $\alpha\beta > 1$ then no such equivalence is possible and a model of the form (D7) will not fit well to data which satisfies (D6). The proof that r and s can be found such that α and β are given by equations D9 when and only when $\alpha\beta \leq 1$ is straightforward. (In the above r, s, α, $\beta > 0$)
When such a fit is possible then $e_i = (1 + sp_i)e_i$ which means

that the proneness p_i influences the estimated exposure e_i. Large proneness values will make the exposure look larger. Another way to look at this is through the equation:

$$p_i' = (p_i - r)/(1-sp_i)$$

The effect of this transformation can be summarized simply (since $r \geq 0$ and $s \geq 0$):

if $p_i/p_j > 1$ then $p_i'/p_j' > p_i/p_j$

which means that the variability of proneness with group is underestimated by the p_i's as determined from (D7). It may be suggested that this inequality is somewhat reassuring: the estimation is conservative; where it errs, it errs in the direction of underestimating the differences between groups in proneness (proneness may be thought of as related to probability of being responsible in an accident situation). Where differences in groups are indicated these differences are not an artifact of this particular shortcoming of the model, instead the opposite is true: the model shortcoming attenuates true differences.

The case where $\alpha\beta > 1$ should lead to lack of model fit. This can be detected and a decision of whether to fit a more complicated model such as expressed by D1 can be made.

Where model fit is good, however, the ambiguity is unavoidable and the most that can be said is that the proneness ratios implicit in the model as fit probably underestimate the true proneness ratios.

To summarize the situation: If the true model may be expressed:

$$\hat{X}'_{ij} \approx X_{ij} = (p'_i + p'_j + \alpha p'_i p'_j + \beta) e_i e_j$$

while the model to be fit is expressed:

$$\hat{X}_{ij} = (p_i + p_j) e_i e_j$$

Then if $\alpha\beta \leq 1$ a good fit is possible such that $\hat{X}_{ij} = \hat{X}'_{ij}$ and if:

$p_i / p_j > 1$ then $p'_i / p'_j > 1$

and $p'_i / p'_j > p_i / p_j$

i.e., the p_i's are correctly ordered but they don't show as much variability as they should (i.e., as the p'_i's do). (The conclusion is that estimated proneness will be an increasing function of true proneness with less variation.)

To understand the situation when $\alpha\beta > 1$ one may return to the general symmetric model

$$X_{ij} = t_i e_j + e_i t_j + \alpha t_i t_j + \beta e_i e_j \qquad \text{D10}$$

which is equivalent to D1. It is not hard to see that X_{ij} is a rank 2 matrix since acting on any vector it produces a linear combination of the e and t vectors. Consequently X_{ij} may be brought into the form:

$$\hat{X}_{ij} = \lambda_1 U_i U_j + \lambda_2 V_i V_j \qquad \text{D11}$$

Koornstra proposed an eigen value/eigen vector analysis of the X_{ij} matrix to produce least squares estimates of t_i and e_i. Koornstra noted that if

$$U_i = 1/\sqrt{2}\,(t_i+e_i) \qquad \text{and} \qquad V_i = 1/\sqrt{2}\,(t_i-e_i) \qquad \text{D12}$$

then the ordinary Koornstra model (D4) or (D7) may be expressed

$$\hat{X}_{ij} = U_i U_j - V_i V_j \qquad \text{D13}$$

Koornstra's proposed method for a least square solution will work in the more general case. The two principal eigen vectors and eigen values of X_{ij} are used in any case to construct $\hat{X}_{ij}$ according to D11. If the original Koornstra model holds then the two principal eigen values will be of opposite sign with the positive value larger in magnitude and all other eigen values will be approximately zero. In situations in which the second largest eigen value is also positive then the more general model (D1) still holds.

Consequently if the Koornstra eigen value/vector technique is used the model fit will automatically be to the more general model (if appropriate). However, as already noted the exposure and proneness estimate may be subject to the distortion indicated and this will not be detectable. Koornstra did not discuss the more general model and so, of course, did not observe its connection to his solution method.

Next, consider a non-symmetric model:

$$\hat{X}_{ij} = (p_i + (1+\gamma)p_j)e_i e_j$$

If $\gamma = 0$ this reduces to the ordinary Koornstra model. If X_{ij} is symmetric then $\gamma = 0$ will give the best fit. If X_{ij} is not symmetric then the asymmetry has been introduced to reflect the asymmetry of the accident situation, e.g., because responsibility has been assigned.

The assymetric model would combine the best features of the Koornstra model with the extra information made use of by quasi-induced exposure models. This could end up being the best type of induced exposure model when the information for its implementation is available. This model can be fit using maximum likelihood. It is not immediately clear how the eigen value and eigen vector solution technique can be used for the non-symmetric model, but it would appear that such an analysis would still be possible.In the validation context described in the Work Plan responsibility data will not be available so this type of non-symmetric model could not be tested.

Whatever model is used, a maximum likelihood solution technique can be used. It should be pointed out that the maximum likelihood method has an advantage which might be of some value in this context, namely that only positive solutions will be found if there are positive counts in each cell since the log likelihood function will blow up (to negative infinity) for zero estimates and so starting from a positive estimate a negative estimate will never be reached.

REFERENCE LIST

1. Thorpe, John D., "Calculating Relative Involvement Rates in Accidents Without Determining Exposure," Traffic Safety Research Review, March 1967.

2. Carr, Brian, "A Statistical Analysis of Rural Ontario Traffic Accidents Using Induced Exposure Data," Accident Analysis and Prevention, Vol. 1, Pergamon Press, 1969.

3. Hall, William K., "An Empirical Analysis of Accident Data Using Induced Exposure," HIT Lab Report, September 1970.

4. Carlson, William L., "Induced Exposure Revisited," HIT Lab Report, November 1970.

5. Haight, Frank A., "A Crude Framework for Bypassing Exposure," Journal of Safety Research, Vol. 2, 1970.

6. ______________, "Indirect Methods for Measuring Exposure Factors as Related to the Incidence of Motor Vehicle Traffic Accidents," Contract No. DOT-HS-041-0-049, September 1971.

7. ______________, "Induced Exposure," Accident Analysis and Prevention, Vol. 5, Pergamon Press, 1973.

8. van der Zvaag, Donald D., "Induced Exposure as a Tool to Determine Passenger Car and Truck Involvement in Accidents," HIT Lab Report, January 1971.

9. Cerrelli, Ezio C., "Driver Exposure - The Indirect Approach for Obtaining Relative Measures," DOT-HS-820, 179, March 1972. Also a shortened version in Accident Analysis and Prevention, Vol. 5, 1973.

10. Joksch, Hans C., "A Pilot Study of Observed and Induced Exposure to Traffic Accidents," Accident Analysis and Prevention, Vol. 5, Pergamon Press, 1973.

11. Koornstra, Matthijs J., "A Model for Estimation of Collective Exposure and Proness from Accident Data," Accident Analysis and Prevention, Vol. 5, Pergamon Press, 1973.

12. ______________, "Empirical Results on the Exposure Proneness Model," acompanies previous article in same issue.

13. Waller, Patricia F., Donald W. Reinfurt, Jean L. Freeman, and Peter B. Inrey, "Methods for Measuring Exposure to Automobile Accidents," presented at the 101st Annual Meeting of the American Public Health Association, November 1973.

14. Waller, Julian A., "Issues Concerning Measurement of the Population at Risk in Crashes," presentation at Second National Formum on Traffic Records Systems, St. Louis, March 1976.

15. Wass, Carsten, Traffic Accident Exposure and Liability, Allerod, Denmark, 1977 (privately published), printed in Denmark by ROC, Rungstead has been available from author at: Ligustervangen 57, DK-3450, Allerod.

16. OECD (1974): Research on Accidents Involving Young Drivers. Final Report from Research Group 58 (This is the reference as given in the Wass book cited just above. This report has not been obtained.)