DEVELOPMENT OF AN INJURY PREDICTION MODEL

A study of feasability on the basis of literature data

Contribution to EEC Seminar on the Biomechanics of Impacts in Road Accidents, Session II: Lateral Impacts and Prediction Models, Brussels, 21-23 March 1983

R-83-31
T. Heijer
Leidschendam, 1983
Institute for Road Safety Research SWOV, the Netherlands

SUMMARY

In recent years SWOV, supported by the Dutch government, has striven to obtain a series of computer programmes, simulating the chain of events occurring in road accidents.

Reasons for this were partly the sometimes overwhelming costs of full-scale research, partly the conviction that the increasingly more sophisticated computer technology can provide an aide to more detailed insight into important facets of the accident process.

Up to now, two major types of models have been acquired:

- a highly sophisticated model, called VEDYAC, simulating vehicles on the road, vehicles colliding with fixed or movable obstacles or vehicle to vehicle collisions;
- the MADYMO computer programmes, simulating the behaviour of crash victims during the accident.

Both models, if applied in succession to a pre-defined type of accident, may provide a good insight into the outcome of such an accident in terms of forces, moments, accelerations and deflections as sustained by both vehicle and victim.

A major problem, however, restricting the usefulness of the models, is the translation of these output quantities into a far more relevant assessment of injury severity.

The development of an injury predicting model is an attempt to deal with this problem.

GENERAL APPROACH OF THE PROJECT

Much like individual accident reconstructions, the models VEDYAC and MADYMO will simulate a single event, in which vehicles and victims are each characterised by a specific selection of properties. The results of such simulations are only valid for that specific event and can hardly be used as indication of the hazards when for instance the properties of the vehicle are significantly different.

The Dutch government, however, which is SWOV's principal "customer", is mostly interested in a general assessment of the risks of road side circumstances. Therefore we need to find a means to generalise the model output so as to accomodate both a variety of vehicles and of occupants. Since most of the important modelling features of a large number of different vehicles are known, generalising the accident situation in this respect poses no great problem; we simply repeat the simulations for every type of vehicle, which may amount to some work but is by no means insurmountable.

Generalisation with respect to the occupants is a much more difficult problem, which is the main reason to start this project.

Basically there are two phases in this project:

a. in order to connect to the output features of the occupant model a way must be found to make valid injury predictions for a single victim;

b. thereafter a method must be developed to merge the results of a sequence of single victim predictions into a valid general prediction.

In this paper the possibilities of both phases will be discussed briefly.

MODELLING CONSIDERATIONS

Injury prediction for a single victim

The easiest way to obtain a usable prediction of injury severity would be a compilation of all applicable injury criteria like e.g. the HIC. Unfortunately, there are not many widely accepted criteria and those that do exist are generally considered unreliable due to large scatter.
Therefore, it was decided to try to develop a different model based upon existing literature and recent research. The form of this model can basically be much the same as a model consisting of separate injury criteria in the sense that it will also consist of a series of independant modules, each of which is dedicated to the prediction of injury severity for a certain part of the body.
This implies that, in order to obtain an overall indication of injury severity, the model must have some means of integrating the output of the separate modules.

Considering the input for the injury prediction, which is the output stage of the crash victim simulation (or dummy test) we mostly find fairly accurate data about forces and accelerations to which various parts of the victims body are submitted and, sometimes less accurate, deflection data. A logical way to transform these mechanical quantities into injuries would be to measure all relevant mechanical properties of the potentially affected tissues and calculate the extent of damage in much the same way as in technical constructions.
For several reasons it is virtually impossible to obtain those desired properties:
- most properties cannot be measured in a non-destructive way so measurement on living subjects is mostly out of the question;
- due to large variance in the value of individual properties it is impossible to substitute once-measured values (e.g. from a cadaver test) in every single case.

As direct measurement is very difficult, the next best is to make an estimate of the desired values as accurately as possible by means of statistical analysis of the data that are available.

As the data originate from a variety of sources, each with its own approach to research, large variance due to those inconsistencies in approach can be expected.

Since the biomechanical parameters themselves already display a considerable variance, fruitfull application of statistical techniques will require a relatively large amount of data, which is a considerable obstacle.

In fact, the situation regarding the availabililty of data seems to be the following:

- data on biomechanical research mostly consists of relations between applied violence and measured injury, often together with some anthropometric data on the subjects but rarely sufficient;
- those properties that are measured in sufficient quantity to be manipulated statistically are mostly rather superficial anthropometry like age, sex, mass and stature;
- those properties that are more directly pertinent to injury prediction are rarely measured in sufficient quantity.

The only chance to obtain usable predictions thus seems to lie in the speculation that the available data on age, sex, mass and height and some other anthropometric data like chest depth, together may account for a sufficiently large part of the inter-individual differences and so can serve as a correlating basis for other properties.

This supposes that most of the properties of the human body that govern susceptibility to injury are somehow related to the aforementioned anthropometrical parameters; those relations need not be linear and therefore the chosen method of data analysis must at least be able to afford non-linear relations within the data.

The capability to predict a certain injury does not necessarily imply a prediction of the severity of that injury; a "diagnosis" to that effect is still needed.

A common form of diagnostic tool is the Abbreviated Injury Scale (AIS) which is generally applied in both accident investigation and laboratory tests and is therefore best suited for the output description of this kind of model.

In order to implement a possibility of assessing an overall severity index, two more or less common criteria have been included in the model:
- a calculation of ISS and
- a calculation of MAIS.

The ISS divides the body into six regions, and since we do not have a more workable alternative, these regions are also adopted in the modular structure of the model.

As it is, it may already prove difficult to provide a somewhat reliable prediction for each of these regions.

Extending the predictions to a population

Up to now, only the possibilities to obtain an individual injury prediction have been discussed, which is basically very important but otherwise hardly relevant for legislative purposes. In the latter case an assessment of the risks confronting a certain population (e.g. of road users) is a far better basis for legislative decisions.

Therefore we must find a way to generalisation of the predictions.

In the first place an usable description of the considered population is needed. Theoretically this means that we must be able to derive all relevant parameters of every individual in the population from this description, in other words, we need a description of the simultaneous distribution of all parameters.

In practice we only dispose of samples of certain populations, so we can only estimate the shape of the simultaneous distribution by inter- and extrapolation, but we could compute an injury prediction for every individual in the sample in any case. The latter is only true, however, if we may assume that the pattern of violence due to a certain type of accident - the forces, accelerations etc. - will be independant of personal characteristics of the victim and so remain constant for every member of the population, which is not very likely. In most accidents the violence exerted on the victims is not only dependant upon variables pertaining to the vehicle(s) and/or outside circumstances but also such properties of the victim as mass, mass distribution and stature may play an important part.

Thus generalisation of injury prediction not only implies the knowledge and application of the aforementioned simultaneous distribution, but also means that the pattern of violence must be calculated separately for every member of the population, which surely is an arduous task if at all possible.

There are some considerations, however, that may make the generalisation feasible yet:

- model output will be defined on the AIS scale: since this scale consists of a limited amount of discrete steps it is necessary that certain groups of individuals in the population will have the same score so that group can be replaced by a single "average" individual in that group and a percentage representing the size of the group;
- also in the stages in which the violence is calculated, that is in the Application of the crash models, a certain insensitivity to small parameter variations is present.

On the basis of these considerations we may reduce the number of calculations considerably and without great loss of information, provided we take care in selecting a sample of population members and their assigned representativity.

DIFFICULTIES IN DATA ANALYSIS

Even if the number of anthropometric variables is limited to four or five it is very difficult to aggregate a sufficient number of data points to satisfy the demands of most common methods of analysis. This is best understood by means of an example. Say that we employ the variables mass, age, sex and stature, three variables of this sequence are measured on a continuous metric scale, the fourth - sex - is a nominal variable. In order to process them together we can devide the metric variables in classes; in our project for instance we divided mass in seven classes, age in six classes and stature in four classes. Together with the two classes of the sex parameter we find a total of 7x6x4x2 = 336 different combinations, all of data (but preferably more) in order to carry out a proper analysis.

Some methods allow a certain percentage of "empty cells", but generally that percentage must not be to high. As many experimental data are based in some way (e.g. towards limited age groups in cadaver experiments) or lacking one or more parameters (e.g. in accident investigation), it is clear that obtaining a sufficient number of non-empty cells is difficult. Moreover, it is likely that the empty cells will not be randomly distributed, but that they are grouped together, meaning that all data for e.g. a certain age group are missing.

Therefore we shall frequently need "educated guesses" to fill otherwise empty cells, which at the same time evokes the need for thorough evaluation of model results.

Thus the construction of such a model evolves into a recursive process of data analysis, model calibration and model evaluation.

A second difficulty often lies in the preparation of literature data for data analysis.

Since experimental data on a certain injury rarely comprise more than a very limited number of experiments, we mostly need to make a compilation of the work of several researchers. A vital condition to be able to make this compilation is, that all experiments have a number of measured parameters in common. As this condition frequently fails to be satisfied, a relatively large portion of research data has to be excluded from processing, resulting in an even more meager basis for predictions.

Another problem in data preparation constitutes the parameterisation of time series and of graphically presented data in general.
Model output of MADYMO, for instance, consists of a number of forces, moments and accelerations of different parts of the body, each supplied as a function of time.
An other example is the shape of a chest defection curve, an important input variable for MADYMO, that varies as the anthropometry varies.
If we want to include these in the data analysis, the series or shapes will have to be represented by a limited number of characteristics.
Selecting these characteristics is sometimes simple (e.g. a maximum force level) but often difficult (e.g. the complex shape of the chest deflection curve or the representation of pulse duration).
Solutions for this problem have to be found for each individual case and may require a variety of techniques some of which are:

for time series:

- peak value indication alone or together with
- rectangular, triangular or sinusoidal pulse duration approximation
- different ways of weighted integral calculations

for characteristics like force/deflection curves:

- predifined shape, varying only by a multiplication factor influencing the area under the curve
- classifying shapes by means of successive corridors.

The nature of the parameterisation may influence the choice and the possible results of the method for data-analysis.

DATA ANALYSIS: SOME POSSIBILITIES

Generally speaking many methods of analysis are based on modelling assumptions.

This means that a hypothesis about the fundamental relations of items within a set of data must be made which is then fitted to the data so as to minimise some criterion of variance.

A very simple example is a linear regression, fitting a line, plane or hyperplane by means of least squares to model the relations within a set of data points. This linear regression, although too simple for most biomechanical applications, is quite fundamental for a whole class of methods. In these methods this linear regression is not applied directly to the data, but to a set of transformed data, transformed in such a way as to obtain the best linear fit.

These methods are capable of coping with a variety of non-linear relations within a set of data and are also applicable if some of the variables are not of a continuous nature (like age: every age between certain limits is possible) but are divided into nominal classes (like sex).

It will be clear that, in order to obtain a good fit over the full range of all variables, data on (almost) all possible combinations within these ranges must be provided, either by measurement or by estimation.

Another condition for successful application is that those variables that are used as separate parameters must not be strongly interdependant. As we are not entirely free in the choice of parameters, we must ascertain the independancy of those parameters in order to leave out those variables that are redundant because they contain largely the same information as other ones.

The previously mentioned class of non-linear regression models are based on an analysis of the correlation matrix and are therefore able to reduce correlating variables to independant factors. Some of these techniques may work on the assumption that the distribution properties of all variables are <u>normal</u> or closely so. Fortunately, this is the case with most variables in biological systems, so this will not be a major obstacle.

The general lack of sufficient data, however, remains a very "tender spot" in all these procedures however sophisticated they are. It means that it will be very difficult to obtain more accurate predictions by introducing a larger number of predictive parameters.

CALIBRATION OF THE PREDICTIONS

Since relevant data on almost all types of injury are relatively scarce, we cannot expect much success from any particular method of data analysis due to the inherent number of "empty cells" mentioned before.

In order to compare and calibrate the model output we must indeed have comparable accident investigation data available. Again this poses an important problem since the level of detail in injury description, needed for suitable calibration, virtually prohibits large scale investigation, which, on the other hand is indispensable for a statistically sound calibration.

Here also the solution must lie in combining the results of worldwide in-depth studies and accident reconstructions. It is trivial that this is only possible at a certain level of standardisation in these fields.

CONCLUSIONS

In view of all the difficulties the conclusion seems almost foregone that the construction of a valid injury-predicting model is next to impossible, and, admittedly, the results as reported in our Final Report on Phase III, NL-5, (SWOV, 1981) are yet a long way from generally applicable.
Still, we feel that if we want computer modelling in this field to grow into a mature tool, these difficulties have to be solved.
Data analysis is not the real problem, but data acquisition in sufficient amount is, and only when this problem is solved better, more accurate predictions and criteria are to be expected.
Until that time, we will have to rely on "educated guesses" and continuous recalibration, which, considering the experience of many researchers in this field, can still produce workable models.
In any case, a lot of work has to be done, of which we consider these co-ordinated EEC projects a hopeful beginning.