

## A decision support system for liability in civil litigation: a case study from an insurance company

Wen Zhang · Andrew Dunkley · Urvi Kanabar · David Elliott · Henry P. Wynn

Received: date / Accepted: date

**Abstract** The use of statistical and AI methods in civil litigation is an area likely to expand. As with many areas of social science, the data requirements are high but complex, because of the complexity of the legal process and the nature of the causal connections. This paper looks at the early stage of the process where the initial establishment of liability acts as a legal triage which affects the route through the litigation process. A simple model is used in which the training set is the assessment of the probability of liability given hypothetical scenarios in road traffic accidents. The model is augmented by additional “weight of evidence” assessments. The model, once built, is used as a decision support system for claim handlers on a routine basis. The methods can be seen as a way of utilising a special type of expert judgment elicitation.

**Keywords** Business analytics · Decision support systems · Experts’ judgment · Legal · Insurance claims

---

Wen Zhang  
University of Essex  
E-mail: phd14wz@mail.wbs.ac.uk

Andrew Dunkley  
Herbert Smith Freehills  
E-mail: Andrew.Dunkley@hsf.com

Urvi Kanabar  
BLM LLP  
E-mail: Urvi.Kanabar@blmlaw.com

David Elliott  
BLM LLP  
E-mail: David.Elliott@blmlaw.com

Henry P. Wynn  
The London School of Economics and Political Science  
E-mail: H.Wynn@lse.ac.uk

## 1 Introduction

Civil litigation is a process in which decisions are made based on a combination of empirical data and expert judgment. For example, in a case involving physical injury, medical data will be presented together with some judgment of severity. Expert legal judgment is needed to make decisions about whether to settle, go to court, etc. There are also lawyers, claims handlers and other stakeholders involved. A decision or judgment that needs to be made and which will determine the path through the litigation process is “liability”. This determination is typically made at an early stage, not long after “first notification”.

In some cases, the facts will, in some sense, speak for themselves. That is to say, there is a straightforward series of logical steps that lead in a deterministic way to an assessment of liability. In other cases, there will be judgment represented as a probability assessment. The judgment will itself be based on various queries of the type “what if X happened?”. Note that we make a considerable play in this paper of the counterfactual “what if X did not happen?”.

This paper covers the case of insurance claims for vehicle accidents. The definition of liability in civil cases, and road accidents, in particular, depends, in the UK, very much on precedent. Liability is separated into four features: negligence, recklessness, intention, and strict liability. At one end, negligence means behaviour which essentially increases the risk of the situation. Here we may think of changing lanes (under some conditions). It may also be that liability is shared by the insured party and another driver. There are in addition important background concepts such as “duty of care”, whose breach is a necessary condition for liability.

Note that liability does not suddenly imply that a driver is responsible for all bad outcomes. In terms of the litigation, it may mean that a category of doubt has been removed. For example, if a second driver was involved, liability may not be attributed to that driver, eliminating the need for a tit-for-tat and possibly expensive legal process.

A useful example of an event,  $X$ , which will lead to an assessment probability, is “changing lanes”. One can imagine that changing lanes while driving may be a risk factor. As liability is itself a subtle concept, the probability of liability is even more so. There is both a future and a past element. Since the accident has happened, we might naturally consider that liability is something associated with the period around the time of the lane changing, or even to be part of the “intentionality” of the driver as in the phrase “behaving recklessly”.

It is this last point that drives the modelling. This is clarified by how it is intended to be used in the model. Expert judgment, in the case of an experienced lawyer, is used to capture probabilities to calibrate a model which is then used by other lawyers/claim handlers to predict/decide liability for the case in hand. This is familiar in the language of machine learning: one set of experts provides a training set to build a model that can then be used by others, routinely, for prediction.

In addition to the probabilistic model, the strength of the evidence is taken into account, for example, using a photograph of the accident site, by introducing the idea of weighting. That is to say some of the answers to a query may be supported by stronger factual evidence. The model will include an informal representation of the

weighting, which will have the effect of weighting parameter values in a model more when evidence is strong and less when it is weak.

## 2 Expert judgment

It is notable that while the related areas of machine learning and AI gather strength, satisfactorily extending the methods of classical statistics, so also do the techniques for collecting and processing expert judgment. Indeed, perhaps the greatest challenge of modern quantitative methods is to combine raw data with elicited judgment. This is an old challenge that goes deep into the foundation of the scientific method.

Despite a widespread opinion that eliciting entities such as probability is hard, carefully controlled methodologies have been successful and useful in risk analysis, safety critical areas [O'Hagan et al.(2006)] [Cooke et al.(1991)] and modern Uncertainty Quantification [M2D(2018)]. Risk matrices which plot probabilities against effect, for risk events, are a cornerstone of risk analysis, for example for capital projects. This is despite the fact that the entries are often based on guesswork even often in more formal Probabilistic Risk Analysis (PRA).

There is considerable literature on the use of probabilistic methods in forensic science [Roberts and Aitken(2014)] [Dawid et al.(2007)]. These involve a careful understanding of the background populations and odds ratios leading to the assessment of probabilities of guilt. The main system tool used is some causal diagram through which conditional probabilities are propagated. As with risk analysis the tension between objective data and judgment cannot be avoided. This is particularly apparent in court when the evidence from witnesses is weighed against physical evidence.

If there is a single most challenging issue in these causal analyses, it is that of counterfactuals. The problem can be expressed simply in terms of controlled experiments. Consider a drug trial. If a patient is administered a drug, then there does not exist a clone of the patient who does not get the drug but gets a placebo. In a court if a witness says: "I hit him back because he hit me" we cannot turn the clock back and create a situation in which the antagonist does not strike our witness. In both examples we do not have access to the counterfactual. In a laboratory physical experiment we can (almost) create a clone, by carefully controlling the test environment. One process gets an intervention while an (almost) identical process (or the same process at a later time when the intervention has been removed), does not get the intervention. Without the ability to do controlled experiments there would be no science. In the social sciences a raft of methods has been introduced to mimic controlled experiments. These are sometimes called natural experiments [Craig et al.(2017)], and measure an average effect via randomisation or matching methods. In the latter case, a person well-matched on a number of factors with the patient who gets the drug is a sort of clone [Rubin(1980)].

This paper concerns civil litigation in which judgments and insurance payouts are made without going to court. An early judgment of liability will lead to one route through the litigation process, non-liability to another.

We try to put together the idea of elicitation and counterfactuals mentioned above. A simple term for such idea would be a "counterfactual query". Work by [Abell(2004)]

has described the value of such queries, in the context of eliciting narratives, even when the counterfactual event may not have actually have occurred. We may ask our witness above: “would you still have hit him if he had not hit you”. Of course, this raises other issues such as trust and realism. Although the lack of realism or artificiality of counterfactuals has been criticised, we claim that counterfactual queries should be awarded status comparable with any elicitation of expert judgment. The fact that the question is conditional should not of itself undermine its status. Expertise must often include expertise related to causation itself. In a civil litigation context, a finding of liability by a judge must necessarily include a finding that causation is satisfied.

In this paper we have elicited judgments concerning the probability of being liable for an accident, when a preceding event has or has not occurred. This is somewhat unorthodox in the sense that liability is hard to locate during the narrative of events, but as we hope to show the querying is a most effective method for pre-screening cases.

In a physical experiment with many (possibly) causal variables it is costly, but often possible, to test with multiple combinations. Such experiments are called factorial experiments [Box et al.(2005)]. An example is a choice experiment for attributes in quality improvement, marketing and policy formation [Li et al.(2013)]. Here we carry out what is often called a “one-factor-at-a-time” experiment. It will be combined with a simple mathematical model which allows prediction in the multi-factor case. Even though querying is only one query at a time, the model predicts well into the multiple factor case. An analogy would be a drug trial with multiple drugs. When, say, two drugs are investigated but never give together, a model may still predict the effect quite well if both drugs are used at the same time unless there are serious hidden synergies (interactions). If a patient takes say ten drugs then it is likely that no clinical trial has been carried out with combinations of all ten drugs.

### 3 Workshops and background

#### 3.1 Workshops

The development of the model required a number of workshops to answer some basic questions. These established quite soon the dichotomy between a logical situation in which lawyers or claim handlers expressed a high degree of certainty and those which required more probabilistic judgment. This dichotomy is reflected in the models. The issue of changing lanes before an accident happens is used both to ask key questions and capture the basic technical facts. Importantly, two parties are involved when a road accident occurs, namely the *insured party* and the *third party*. The purpose of the model is to predict whether a party is at fault by asking both of them a series of questions. All possible results of the prediction are summarised in Table 1. For each party, there are three possible results, fault, not fault and no decision, depending on their answers to the questions.

From the case handlers’ viewpoint, when a party changed lanes certain facts will be taken into consideration when deciding whether the party is at fault: why did the party change lanes; did he/she changed lanes because of mechanical failure; was there

**Table 1** All possible prediction results

Insured	Third Party
Fault	Fault
Fault	Not Fault
Fault	No Decision
Not Fault	Fault
Not Fault	Not Fault
Not Fault	No Decision
No Decision	Fault
No Decision	Not Fault
No Decision	No Decision

some physical impairment; was the impairment due to a legitimate or illegitimate reason?

The claim handlers found it challenging to suggest weightings of facts (or evidence) when a party has changed lanes. Since the probability of fault where there is a lane change lane is thought to be higher, it was decided to use a deterministic model to predict liability without using weightings of facts or evidence. We refer to this as Model 1.

But, on the other hand, the claim handlers were able to suggest a list of evidence for the no changing lane's scenario and to provide to respective evidence "weightings". The models were divided into four cases, with Model 1 giving the deterministic case and Models 2, 3, 4 the probabilistic version. Here is a summary of the relevant queries:

- Model 1: Did the party change lanes?
- Model 2: Was the party doing the right thing in the right place?
- Model 3: Was the party driving appropriately for the conditions?
- Model 4: Should they be aware of the other vehicle?

The next step is to transform the facts into a series into questions, under each model type. These questions are representative of the facts and used to ask the users, insured party and third party (or their solicitors). Forms were filled in accordingly.

From the workshop, we understand some groups of factors are relatively important when deciding whether the party is at fault when not changing lanes. We had a discussion with these five claim handlers, who helped to determine that the relative weight of model 2, 3 and 4. How these weights enter the model is discussed in Section 5.2.

#### 4 Model

First, there is the Model 1, the deterministic model. This may also be called a rule-based model and points to the distinction, already mentioned, in this and other studies between models which are based only on logical principles and which form a tree with logic gates at each level, and probabilistic model. Second, is a basic probability model where the elicited probabilities described above provide the training set from

the model. A similar hybrid model has been proposed by [De Caigny et al.(2018)], where they argue that different models constructed on segments of the data rather than on the entire dataset lead to better predictive performance, which also applies to our problem. Third, the weight of evidence is used, in two different ways, to the parameters in the basic model.

#### 4.1 Deterministic model

In practice, lawyers/claim handlers are almost certain that changing lanes will result in being at fault, subject to some rare exceptions such as mechanical failure. Therefore, we will ask the party “did you change lanes” first, if the answer is “Yes”, we proceed through a logical tree which results in a decision: at fault or not at fault. This tree is presented in Figure 1. Depending on the party’s answer of “Yes”(‘Y’), “No”(‘N’), “Don’t know”(‘D’) to three levels of questions, the deterministic model will help to decide whether the party is at fault or not. The questions are used here as generic: “Q1” to “Q3” (due to confidentiality reasons). It should be noticed that there are only two results in the deterministic model, either at fault or not, which is different from those of the probabilistic models, where “unknown” is applicable. This is again, on the grounds that the “change lanes” questions will almost decide whether the party is at fault or not.

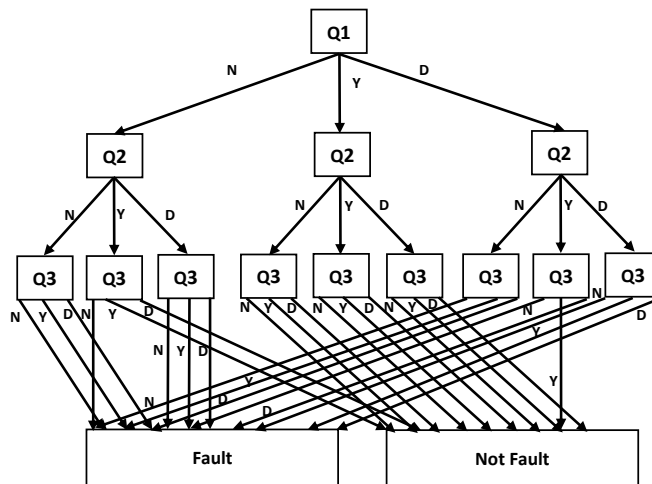


Fig. 1 Deterministic model

## 4.2 Basic probabilistic model

Given that the output of the predictive model is binary: liable/not liable, it seems natural to introduce a logistic regression model, which is very prevalent in both mainstream statistics and modern machine learning. It is used in medicine to predict mortality and in credit risk assessment. Our approach is somewhat different. We do not have a large training data set which would allow a logistic model within the Generalised Linear model category. Our model is an “empirical” logistic regression in which the raw elicited probabilities are used directly. Moreover, possessing the counter-factual allows us to think of the set up as a kind of experiment in which, as explained one factor (query) is varied at a time, with the implication that the other factors as not varied, modelled here by being set to zero.

Assume we have  $k$  factors that will affect the response determination, such as “indicate before”. We will construct a prediction model:

$$y(x) = \log \left( \frac{p(x)}{1-p(x)} \right) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_k x_k + \varepsilon \quad (1)$$

where  $p(x)$  is the probability of being at fault at  $x = (x_1, \dots, x_k)$ , where  $x_j = 1$  when query is “yes”,  $x_j = -1$  when the query  $j$  is “no” (counterfactual) and  $x_j = 0$  when the answer for a neutral situation, for  $j = 1, \dots, k$ . The response values are the elicited probabilities  $p_i$  for query  $i$ ,  $i = 1, \dots, N$ , where, typically,  $N = 3k$ ,

The predicted probability for a given query answers is given by inverting the logistic

$$\hat{p}_i = \frac{\exp(\hat{y}_i)}{1 + \exp(\hat{y}_i)},$$

where the predictor of  $y_i(x)$  is the raw least squares predictor at the  $i$ -th configuration  $x^{(i)} = (x_{1i}, \dots, x_{ki})$ :

$$\hat{y}_i = \hat{\theta}_0 + \hat{\theta}_1 x_{1i} + \dots + \hat{\theta}_k x_{ki}.$$

### 4.2.1 The role of $\theta_0$

The simplicity of the model means that the value  $\theta_0$  automatically corresponds to the neutral case when all  $x_i = 0$ . That is  $y(0, \dots, 0) = \theta_0$  so that

$$p(0, \dots, 0) = \frac{\exp(\theta_0)}{1 + \exp(\theta_0)}.$$

The orthogonality of the “design” gives

$$\hat{p}(0, \dots, 0) = \frac{\exp(\hat{\theta}_0)}{1 + \exp(\hat{\theta}_0)},$$

where

$$\hat{\theta}_0 = \frac{1}{N} \sum y(x_i),$$

the sample mean of the  $y_i$ . Again, one of the nice features of the model, which is a consequence of the orthogonality mentioned above is that this does not depend on any other parameter estimates.

There are two further suggestions. The first is to set  $\theta_0 = 0$ . In legal terms, this is to say that in the absence of information, it is natural or fair to set the probability of liability equal to one half. The second is to make a ‘‘by hand’’ adjustment. That is, take the  $\theta_0 = 0$  in the base model but add in a known adjustment post hoc; an additional calibration based on the raw data, to improve outcomes; what might be called a ‘‘tuning parameter’’.

## 5 Model with evidence strength

### 5.1 Basic weighting

A separate analysis leads to a measure of the weight of evidence. An item in this analysis may be evidence from CCTV, telematics and vehicle history. The strength of different types of evidence is different. The relative strength of evidence is collected from experienced lawyers/claim handlers and then averaged.

Assume that the insured party collected  $Q$  items of evidence in total and define the strength of the  $q$ th evidence as  $e_1^q$ ; third party collected  $P$  items of evidence in total and the strength of the  $p$ th evidence as  $e_2^p$ . The total evidence strength for the insured party and the third party is  $E_1$  and  $E_2$  respectively. The evidence weight for the insured party  $w_{in}$  is calculated as

$$w_{in} = \frac{E_1}{E_1 + E_2}, \quad (2)$$

where, as mentioned,  $E_1 = \sum_{q=1}^Q e_1^q$  and  $E_2 = \sum_{p=1}^P e_2^p$ . Combining evidence weight with  $\hat{\theta}$ , we now have a weighted models  $\hat{y}$  to predict liability

$$\hat{y}(x, w_{in}) = \hat{\theta}_0 + (\hat{\theta}_1 x_1 + \dots + \hat{\theta}_k x_k) w_{in}.$$

The probability that the insured party is at fault is then

$$\hat{p}(x, w_{in}) = \frac{\exp(\hat{y}(x, w_{in}))}{1 + \exp(\hat{y}(x, w_{in}))},$$

The probability that the third party is at fault can be calculated similarly. Although this weighting method is a somewhat ad hoc, shrinkage of an estimator towards some prior value is a familiar method in statistics. It is of some interest here that shrinkage is towards the neutral value of  $\hat{\theta}_0$ . We should also note that the linear model(s) given in (2) can be replaced by a non-linear model, which we consider deals better with small and large values. These details are omitted.



## 5.2 Sub-model weighting

In the previous subsection, a single  $w$  is applied to all the terms in the logistic model, except  $\hat{\theta}_0$ . A somewhat different kind of weighting is also applied, which reflects the relative importance of Models 2, 3 and 4. To recall, in the overall combinations of models, there are a number of initial steps. The first is to ask if the party denies changing lanes, if 'yes', the deterministic model will be used to predict. If 'no', the probability model is used. In our additional weighting, each of the Models 2, 3 and 4 is allocated a different weighting. It seemed natural to apply the weighting directly to the predicted probabilities.

Thus, for every sub-model, we calculate the probability that the party under concern is at fault the analysis in section 4.2. Define weights  $W = (W_2, W_3, W_4)$  and  $\hat{p}_i(x, W), i = 2, 3, 4$  as the prediction from Models 2, 3 and 4 respectively. The combined predicted probability is given by

$$\hat{p}(x, W) = W_2\hat{p}_2(x, W) + W_3\hat{p}_3(x, W) + W_4\hat{p}_4(x, W) \quad (3)$$

The values of the weights were elicited following further special workshops dedicated to this task. In the absence of evidence, either party (insured, third party) the probability is assigned to  $\frac{1}{2}$  which in the model is equivalent to  $\theta_0 = 0$  as explained in subsection 4.2.1.

## 6 Validation

We test the performance of the model by comparing its predictions with that of subjects. Nine claim handlers were each asked a variety of questions according to the routes to Model 1 (deterministic) or Models 2, 3, 4 (probabilistic).

The respondents were asked to decide whether the insured party was at fault (1), not at fault (-1), neutral (0), and the same for the third party. The model prediction was similar: 1 if the prediction  $\hat{p}_i > 0.5$ , -1 if  $\hat{p}_i < 0.5$  and 0 if  $\hat{p}_i = 0.5$ . In addition, the third set of data, which is referred to informally as the "truth", are decisions made by the respondents after have received a full set of data. Thus there are three sets of data (i) respondent, (ii) model and (iii) "truth". Each data set consists of 225 two-entry vectors with entries  $\pm 1$  or zero.

In assessing how well the model performs, we compare the initial respondent with the truth and the model with the truth. We denote the  $u, v$  as the prediction result for the insured party and third party, which is 1 if the predicted result is same as the truth and 0 otherwise; and  $r, m, t$  refers to the case of the respondent, model and truth and  $i$  for the  $i$ th entry. Since there are three possible prediction results, we decided against the more familiar ROC curve and F-score. Instead, we use accuracy, Hamming distance, which is the number of positions at which the corresponding symbols are different, and  $L_1$  distance to measure the performance. The four scores we have for the claim handler's prediction are

1.  $S_1$ : prediction accuracy for the insured party, calculated as  $\frac{\sum_{i=1}^{225} u_r^i}{225}$ ;

2.  $S_2$ : prediction accuracy for the third party, calculated as  $\frac{\sum_{i=1}^{225} v_r^i}{225}$ ;
3.  $S_3$ : Hamming distance;
4.  $S_4 = \sum_{i=1}^{225} (|u_r^i - u_t^i| + |v_r^i - v_t^i|)$ ,  $L_1$  distance.

The same scores are calculated for the model's prediction. For  $S_1$  and  $S_2$  we give the proportion  $S_1 = 1$  values, for  $S_3$  and  $S_4$  the average values, noting that the maximum are 2 and 4 respectively. Table 2 presents how many cases that the respondent (claim handler) and model have made a correct decision for the insured and the third party respectively. The result scores are given in Table 3.

**Table 2** Validation results

	<i>Insured</i>	<i>Thirdparty</i>
<b>Respondent</b>	205	216
<b>Model</b>	223	224

**Table 3** Validation results scores

	$S_1$	$S_2$	$S_3$	$S_4$	$1 - S_3$	$1 - S_4$
<b>Respondent</b>	0.911	0.960	0.064	0.113	0.936	0.887
<b>Model</b>	0.991	0.996	0.007	0.007	0.993	0.993

It can be seen that the model performs better than human respondents with all measures, especially the  $L_1$  distance score. This distance is of importance in practice because it measures how far the predicted result is from the truth, a large distance usually brings a high cost.

## 7 Implementation

Combined the deterministic and probabilistic model, a decision support system is built for liability identification in a road traffic accident. The system is designed to be used by a claims handler or lawyer with knowledge of the facts of the case. This section illustrates the implementation of this decision support system and some details are concealed due to confidential reasons.

A model-based decision support system typically has three parts. Working backwards, a decision environment is support by the output of a models which in turn is calibrated with data generated by some kind of observational study or experiment. There may also be places in the model where expert judgement can be included. In additional once the system is built there will be stakeholders, who are not usually those who calibrated the system who are the "users", who are essentially the decision-maker in the specialised decision environment, in this case the legal or insurance environment.

The opening action of the user is to complete a set of questions on-line via an interface. The first question is "Did the party change lanes?". Depending on the user's

answer to each question, the user will answer more than a dozen questions. These results are then stored in a database displayed where 0 to 3 denote the answer of 'No', 'Yes', 'Unknown' and 'Blank' to different questions. The data is then input into the combined model, which generates a recommendation. Both parties' probability of being at fault and final recommendation based on inputs. The result is posted back into the database, for future use, e.g., model improvement. A separate process then generates an email to the user, stating which party is at fault, or no decision, based on available information. Figure 2 is an example result email. The four main stages of the process are accompanied by user interface, which are simply given in outline in Figure 3.

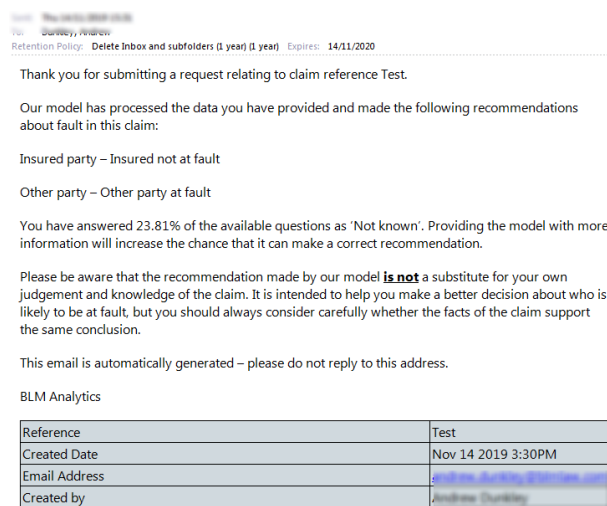
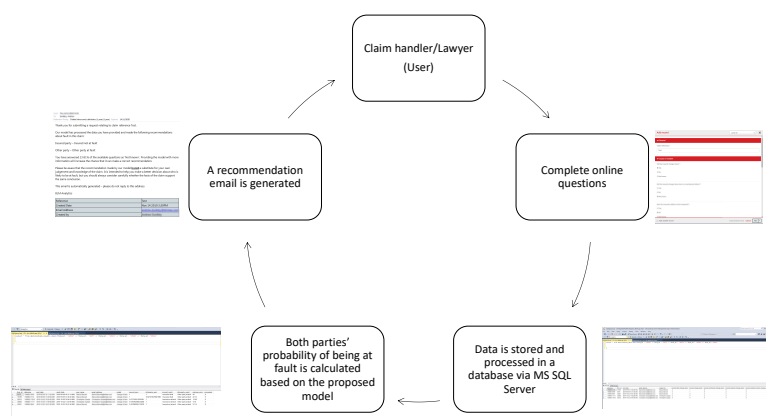


Fig. 2 Example result email

The architecture is chosen for a number of reasons. First, an online form is relatively straightforward to build and integrate into other systems; second, a simple architecture will be easy to maintain; third, the user interface can be easily accessed and understood, maximising the reach of the researched model; finally, sending the user output via email means that no systems integration is required at the end of the process, further increasing maintainability and reducing the cost of deployment.

Several key design considerations flow from the statistical model that the decision support tool had to reflect. First, the questions provided in the user interface had to reflect those in the tested model. Any material changes to the question set compared to the test set would require the model's accuracy to be reconfirmed. Also, from a user perspective, there is a fatigue effect to completing a form that is unnecessarily long. This must be set against the need to collect enough information for the model to function. To balance these consideration, further work is being carried out to ascertain where groups of questions can be aggregate into a small number, even a single question. The data based referred above will be used to do the necessary clustering



**Fig. 3** Framework of the decision support system

analysis. This enabled a reduction in the length of the user-facing questionnaire without any impact on the results generated.

## 8 Conclusion

This project can be considered as a proof of concept in a highly technical and sensitive area, where professional judgment plays a central role. The early assessment of liability is a critical part of civil litigation and was selected as being amenable to simple statistical modelling built on careful elicitation experiments, and nuanced with judgments about the quality of evidence of different types. Early trials have proved successful and should give confidence that lessons have been learned which may contribute to the growing awareness of the need for more quantitative methods, to save cost to stakeholders in an equitable and validated fashion. Future research could examine how does the selection of the base parameter  $\theta_0$  will affect the result, automatic selection of weightings while larger dataset is available and how does the system perform accordingly.

## Conflict of interest

The authors declare that they have no conflict of interest.

## References

- [Abell(2004)] Abell P (2004) Narrative explanation: an alternative to variable-centered explanation? *Annu Rev Sociol* 30:287–310

- [Box et al.(2005)] Box GE, et al. (2005) Statistics for experimenters. In: Wiley Series in Probability and Statistics, Wiley Hoboken, NJ
- [Cooke et al.(1991)] Cooke R, et al. (1991) Experts in uncertainty: opinion and subjective probability in science. Oxford University Press on Demand
- [Craig et al.(2017)] Craig P, et al. (2017) Natural experiments: an overview of methods, approaches, and contributions to public health intervention research. *Annual review of public health* 38:39–56
- [Dawid et al.(2007)] Dawid AP, et al. (2007) Object-oriented bayesian networks for complex forensic dna profiling problems. *Forensic Science International* 169(2-3):195–205
- [De Caigny et al.(2018)] De Caigny A, et al. (2018) A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research* 269(2):760–772
- [Li et al.(2013)] Li W, et al. (2013) Conjoint analysis and discrete choice experiments for quality improvement. *Journal of Quality Technology* 45(1):74–99
- [M2D(2018)] M2D (2018) M2d: Research agenda on decision making under uncertainty. URL <https://blogs.exeter.ac.uk/models2decisions/m2d-research-agenda/>
- [O’Hagan et al.(2006)] O’Hagan A, et al. (2006) Uncertain judgements: eliciting experts’ probabilities. John Wiley & Sons
- [Roberts and Aitken(2014)] Roberts P, Aitken C (2014) The logic of forensic proof: inferential reasoning in criminal evidence and forensic science. Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses, Practitioner Guide (3)
- [Rubin(1980)] Rubin DB (1980) Bias reduction using mahalanobis-metric matching. *Biometrics* pp 293–298