EVALUATION OF THE AVALUATOR DECISION-SUPPORT TOOL FOR CANADIAN ACCIDENTS: 1997-2009

Dave Gauthier*

Dave Gauthier Geoscience, Thunder Bay, CANADA

ABSTRACT: The Avaluator™ is a rule-based avalanche decision-support tool for amateur backcountry recreationists, published by the Canadian Avalanche Centre. It consists of a Trip Planner (TP) for choosing appropriate backcountry destinations, and a slope assessment tool called the 'Obvious Clues Method'© (OCM) for use in the field. Evaluating a decision aid with historic avalanche accident records is crucial for assessing its effectiveness. While the TP component of the Avaluator was examined with respect to Canadian accidents during its development, the OCM component was validated using only U.S. accident data. The goal of the current study is to provide the first evaluation of the Avaluator™ using only Canadian accident data. Significant effort was made to compile a complete record for each fatal avalanche accident that occurred in Canada in the seasons 1997 to 2009; however, missing data remain a significant challenge in the evaluation. Unfortunately, no simple and consistent treatment was available to handle missing data in the analysis. Therefore, accident prevention values were calculated under several assumptions regarding missing data to provide insights on the limits of possible values, and allow the direct comparison with values calculated from the U.S. data. The analysis showed that clue presence in Canadian accidents was not significantly different from that published in the Avaluator™, although the Avaluator™ values may be similar to the upper limit for the Canadian dataset. The main conclusion of this study is that further investigation of each accident record would reduce missing data, and allow a much more reliable evaluation.

1. INTRODUCTION

The Avaluator is a rule-based decision support tool produced and published by the Canadian Avalanche Centre (CAC). Its main component is a pocket-sized laminated card, intended for use by amateur backcountry recreationists in snow avalanche prone parts of Canada. One side of the card includes the Obvious Clues Method (OCM; Haegeli and McCammon, 2006) and is intended for use in the field to support terrain and travel decision-making. In the fall of 2010, the CAC will release an updated version of the Avaluator which will employ a new slope assessment tool.

The OCM component of the first-generation Avaluator lists seven 'obvious' factors or clues that were found to be present in many historic avalanche accidents; the user is responsible to make basic observations while travelling, and maintain a running tally of the number of 'clues' observed. At a point in the trip where some terrain or travel decision is required, users consult the OCM tool for decision-support recommendations. Normal caution (zero to two clues present), extra caution (three to four clues present), and travel not recommended (five or more clues present) advice is provided as the number of observed clues increases. Threshold values for these recommendations were determined based on the prevalence of the clues in historic avalanche accidents.

McCammon and Haegeli (2005, 2006, 2007), Haegeli and McCammon (2006), and Haegeli et al. (2006) have written extensively detailing the development of the Avaluator and the OCM and its validation against the U.S. avalanche accident database. The focus of this paper is to present the fist evaluation of the Avaluator using only Canadian accidents.

Floyer (2008) and Uttl et al (2008a) provide a thorough review of the research leading up to the development of the Avaluator product and the OCM. Floyer's (2008) study was commissioned, in part, due to a series of publications (e.g. Uttl et al 2008a, 2008b) that were critical of the Avaluator, the Obvious Clues Method, and accident prevention statistics computed from the historic database. The main criticism addressed in the current study is that related to the treatment of missing values in the original data set. See Floyer (2008) for a thorough review of the literature related to missing data in statistical analyses.

Corresponding author address: Dave Gauthier, 414 Frankwood Ave, Thunder Bay, ON, Canada P7B 4R2; tel: 807-627-8654; e-mail: davidmgauthier@gmail.com

Uttl et al (2008a, 2008b) made an argument against calculating prevention values following the casewise deletion of accident records having missing values for one or more clues, because doing so implicitly treats 'missingness' of information about a clue as random and unrelated its presence or absence. Uttl et al (2008b) attempted to replicate the accident prevention values published for the Avaluator by independently coding available accident data, and treating missing values as indicative of absence, with the latter approach justified by identical arguments to Uttl et al. (2008a). Their calculated prevention values were much lower than published for the Avaluator, as expected given their treatment of the missing data. For example, they found only 18% of accidents had greater than four clues present, versus 77% for the Avaluator. This potentially dangerous difference (Uttl et al, 2009a, 2009b) was attributed to the inappropriate treatment of missing values by the Avaluator authors

The focus of the current study is to determine the proportion of fatal avalanche accidents in Canada that occurred in the seasons 1997 to 2009 (inclusive) that *would have been prevented* if the Avaluator Obvious Clues Method had been used – and its recommendations followed – by the victims. As missing data is an intrinsic feature of the Canadian accident database, the calculation of prevention values must be preceded by a discussion of the appropriate treatment of the missing data.

2. METHODS

2.1 Data Collection

The CAC acts as the central repository for information and documents pertaining to Canadian avalanche accidents. Database records are supplemented with digital and physical background documents, which were the main data source for this study.

The background documents available included Canadian Avalanche Association (CAA) Reporting Forms, Provincial Coroner Reports, Parks Canada Rescue Reports, Avalanche Expert Reports, InfoEx submissions, Press Releases and Media Reports, Victim or Rescuer Correspondence, and other related files.

In general, data collected shortly after an accident focused on documenting the rescue and recovery, and the dimensions of the avalanche and deposit. It is often a matter of luck if the information required to determine the presence or absence of an obvious clue is contained in the background documents; however, in most cases the number of clues for which a confident assessment of presence or absence could be made increased with more detailed research.

2.2 Avaluator Obvious Clue Method Coding

The coding scheme used for this study represents a consensus, following numerous reviews and iterations, between several researchers familiar with the Avaluator and the spirit of the OCM. Table 1 outlines the fields and basic coding criteria. For each record in the database the absence or presence of each clue was coded as 'Yes' or 'No', respectively, while missing values were coded as unknown, or 'Unk, based on information available in the accident databases and background documents.

Table 1. Summary of Avaluator OCM fields and basic coding criteria.

Field	**Detail**
Avalanche	Are there signs of slab avalanche activity in the area within the last 48 hours?
Loading	Was there significant loading by snow, wind, or rain in the area within the last 48 hours?
Path	Are you in an obvious path or starting zone?
Trap	Are there gullies, trees or cliffs that would increase the consequences of being caught?
Rating	Is the danger rating considerable or higher?
Unstable	Are there signs of unstable snow, such as whumpfing, cracking, or hollow sounds?
Thaw	Has there been recent significant melting of the snow surface by sun, rain or warm air?

2.3 Treatment of Missing Values

Despite arguments to the contrary (e.g. e.g. Uttl and Kissinger, 2009), I expect that the missingness mechanism is unique to each clue and each accident, and in most cases is related to the depth and detail of the original investigation and the research of the coders. In other words, most of the missing data are likely not *unknowable,* but are simply unknown. For all clues, the missing data were reduced in the

Canadian database with further background research. For this study, I accept the limitations of the data, and present the results using the same assumptions made by previous authors in order to make valid comparisons:

1) All are missing at random, and therefore only accidents with confirmed presence or absence of all clues are included (as per the Avaluator, e.g. Haegeli and McCammon, 2006; McCammon and Haegeli, 2007);
2) All are missing not at random, and missing values indicate absence of a clue (as per Uttl et al. 2008a, 2008b, 2009b);
3) All are missing not at random, and missing values indicate the presence of a clue (alternative to Uttl et al. 2008a, 2008b, 2009b).

Note that the true or actual count of clues present for each accident with missing data must lie between the two ‘missing not at random’ distributions, and that none of these broad, simple assumptions is truly valid for the treatment of missing values in this dataset.

2.4 Calculation of Hazard Scores and Prevention Values

This study follows previous ones in presenting results in terms of *hazard scores* and *accident prevention values* (e.g. McCammon and Haegeli, 2007; Uttl et al, 2008b; Floyer, 2008). The hazard score for an individual accident is simply the tally of obvious clues that were present for an accident, and can range from zero to seven. Accident prevention values are calculated from the hazard score distribution, and are the cumulative proportion of accidents that occurred (and therefore would have been prevented) with *n* or fewer clues, where *n* can range from zero to seven. Binomial exact confidence limits are presented along with accident prevention values; these were not corrected for the known number of fatal avalanche accidents, whereas McCammon and Haegeli (2007) present corrected values. A direct comparison is therefore not valid.

Accident prevention values calculated from historic accident databases are often confused with risk reduction, especially when used to promote or validate decision support tools. It is important to clarify that neither the hazard score nor accident prevention values make reference to risk or its reduction, since an unknown number of successful backcountry decisions have been made at any clue threshold, under identical conditions to a subset of fatal accidents (McCammon, 2006).

2.5 Comparisons with the Avaluator

The hazard score distributions calculated from the Canadian dataset are compared with each other and those published for the Avaluator using the non-parametric Kolmogorov-Smirnov (K-S) test, which evaluates the cumulative frequency distribution difference through the K-S D statistic. In this study the null hypothesis - that there is no significant difference between two sample distributions - is not rejected where the K-S D statistic p-value is greater than 0.05.

3. DATA

3.1 Complete Dataset

Fatal avalanche accidents that occurred in Canada between 3 December 1996 and 14 April 2009 (i.e. seasons 1997 to 2009 inclusive) are included in the database for this study. The upcoming publication of Volume 5 of Avalanche Accidents in Canada will contain narratives and summary data for much of this interval. Data compilation and analysis resources were shared between that and the current project.

The complete set of accident records was filtered to omit from further analysis two recent accidents (2009) for which no background data or details were available, and those that occurred under conditions or during activities for which the Avaluator was not designed. The filtered dataset contains 101 accidents, all of which involved the target activities and user group of the Avaluator tool. The accidents retained in this group include those that occurred during recreational backcountry and out-of-bounds skiing and snowmobiling, mountaineering, ice-climbing, and snowshoeing and hiking.

4. RESULTS

4.1 Missing Values

Figure 1 shows the frequency of accidents with confirmed presence (Total YES), confirmed absence (Total NO), and missing values (Total Unknown) for the target audience (n = 101) dataset. The number of missing values ranges from about 12 % for the thaw clue to about 75% for the unstable snow clue. Figure 2 shows the distribution of accidents by the number of clues with confirmed presence or absence, highlighting the distribution of missing values within the dataset. There are no missing values in 19 of the 101 accidents in the target group. In 65 accidents, there are two or fewer missing values, while in 21 there are four or more missing values. The U.S. dataset had 34% of cases with no missing values.

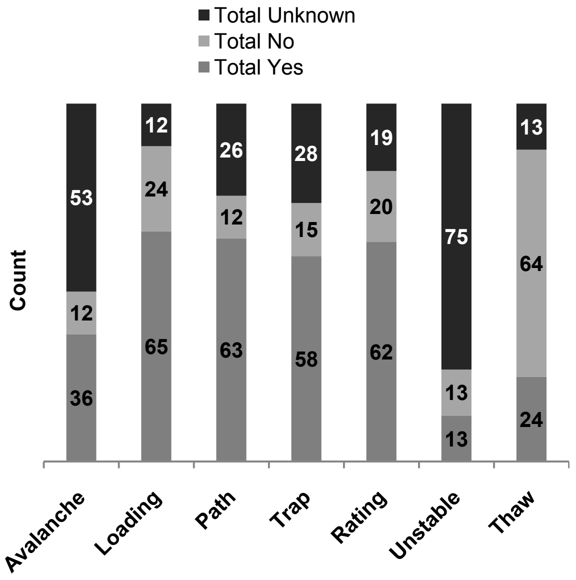

Figure 1. Frequency of accidents with confirmed presence (Total YES), confirmed absence (Total NO), and missing values (Total Unknown) for the target audience (n=101) dataset.

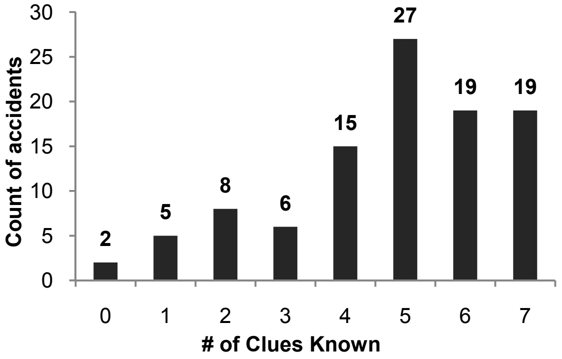

Figure 2. Distribution of accidents by the number of clues with confirmed presence or absence, highlighting the distribution of missing values within the dataset.

4.2 Hazard Scores and Prevention Values

Figure 3 shows the hazard scores (count of clues found to be present at accident) for each of the three missing value assumptions: all missing values assumed to indicate absence (*Unknown = no*, n=101; fig.3a), missing values assumed missing at random (*Unknown = delete*, n=19; fig.3b), and missing values assumed to indicate presence (*Unknown = yes*, n=101; fig.3c).

The *unknown = no* distribution had a mode of three clues present, with an average of 3.32. The hazard score of 60.4% of accidents was between two and four clues (i.e. mode +/- one clue). As expected given the assumption applied to missing values, the distribution is skewed slightly toward a lower number of clues present, with five accidents having no clues present, and only one with seven clues present.

The *unknown = delete* distribution had a mode of five clues present, with an average of 4.78. The hazard score of 84.2% of accidents was between four and six clues (i.e. mode +/- one clue). No accidents under this assumption had fewer than three clues present, and one (5.3% of accidents)

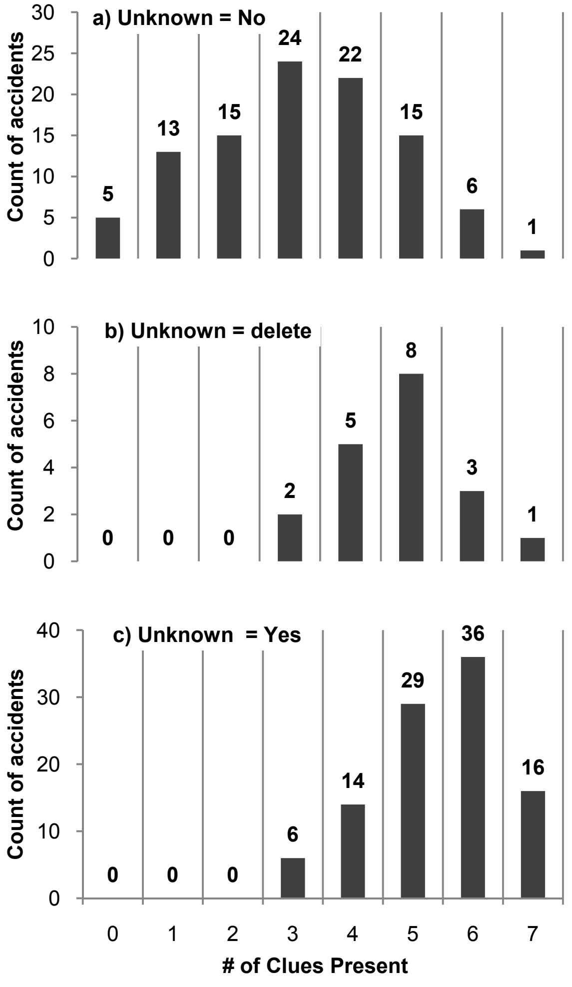

Figure 3. Hazard scores (count of clues found to be present at accident) for each of the three missing value assumptions: a) all missing values assumed to indicate absence (Unknown = no, n=101); b) missing values assumed missing at random (Unknown = delete, n=19); and c) missing values assumed to indicate presence (Unknown = yes, n=101).

had all seven clues present.

The *unknown = no* distribution had a mode of six clues present, with an average of 5.49. The hazard score of 80.2% of accidents was between five and seven clues (i.e. mode +/- one clue). As expected given the treatment of missing values for this group, the distribution skewed toward a higher number of clues present. No accidents had fewer than three clues present, while 52 had six or seven clues present.

Figure 4 shows the calculated prevention values for each of these assumptions, as well as the exact binomial confidence interval ($p < 0.05$) for each level of clue presence. The *unknown = yes* and *unknown = no* prevention values at each threshold represent upper and lower limits, respectively, for the Canadian dataset. At the six or fewer clue threshold, each missing value assumption has a prevention value of less than 15.1%; however, at the five or fewer clue threshold prevention values range from 51.4% (*unknown = yes*) to 6.9% (*unknown = no*).

4.3 Comparison with the Avaluator

Figure 5 shows the prevention values for the Avaluator and the data from the current study that was subjected to the same casewise deletion of records with missing values (n=19). The hazard score distributions of these two groups are not significantly different (K-S D = 0.2577, p = 0.162).

Figure 6 shows the prevention values for the two limiting assumptions compared to those published

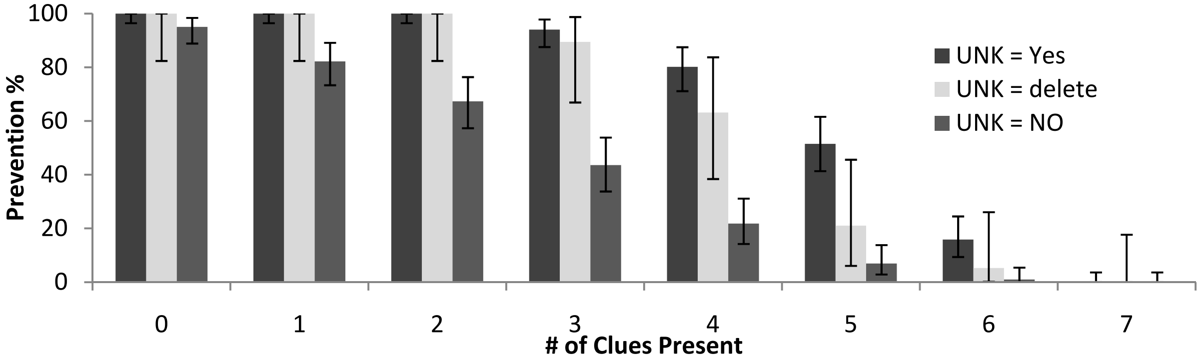

Figure 4. Accident prevention value distributions for three assumptions: all missing values assumed to indicate absence (Unknown = no, n=101); missing values assumed missing at random (Unknown = delete, n=19); and missing values assumed to indicate presence (Unknown = yes, n=101). Exact binomial confidence interval (p < 0.05) for each level of clue presence is indicated.

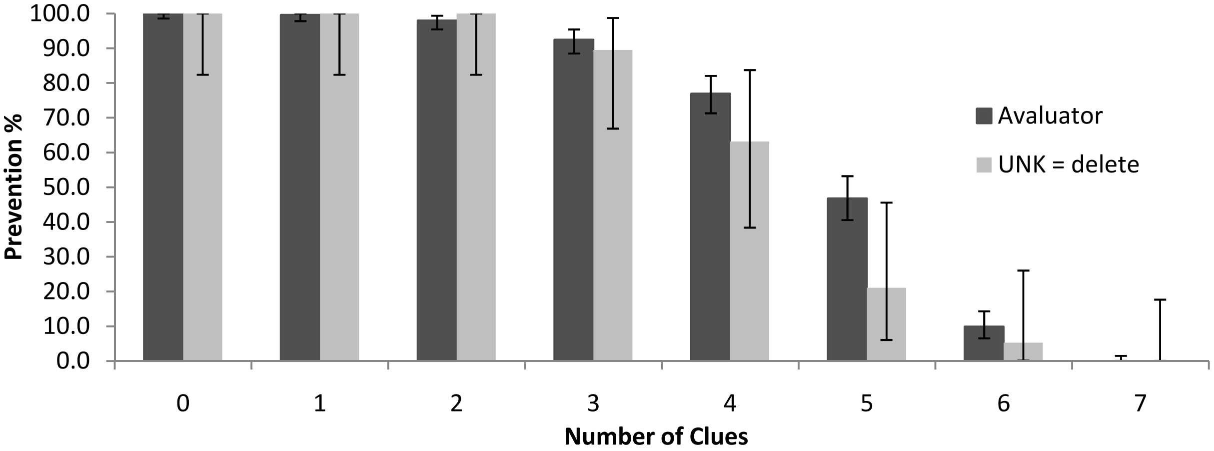

Figure 5. Accident prevention values for the Avaluator (n=252) and the data from the current study that was subjected to the same casewise deletion of records with missing values (Unknown = delete; n=19). The hazard score distributions of these two groups are not significantly different (K-S D = 0.2577, p = 0.162). Exact binomial confidence interval (p < 0.05) for each level of clue presence is indicated.

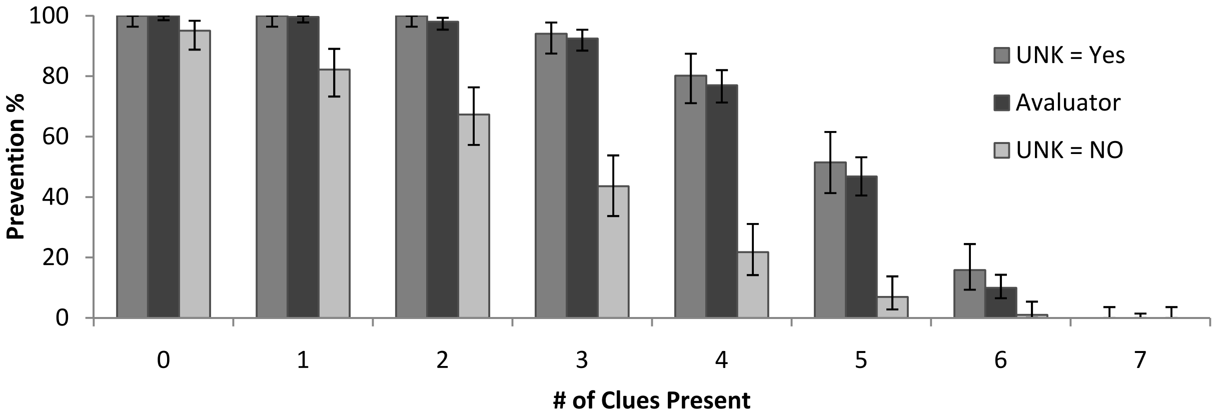

Figure 6. Accident prevention values for the Avaluator with those of the two limiting assumptions for the Canadian dataset. The unknown = yes hazard score distribution is not significantly different from that for the Avaluator (K-S D = 0.0592, p = 0.957). The unknown = no distribution is significantly different from the Avaluator (K-S D = 0.5520, p < 0.01).Exact binomial confidence interval (p < 0.05) for each level of clue presence is indicated.

for the Avaluator. The *unknown = yes* hazard score distribution is not significantly different from that for the Avaluator (K-S D = 0.0592, p = 0.957). The *unknown = no* distribution is significantly different from the Avaluator (K-S D = 0.5520, p < 0.01).

At each level of clue presence the Avaluator, prevention value lies between these end members, and is therefore within the possible range of actual prevention values for the Canadian accident dataset. Notably, the Avaluator hazard score distribution is not significantly different from the *unknown = yes* distribution. Since the *unknown = yes* is an unrealistic assumption for the Canadian dataset resulting in an overestimation of clue presence, the Avaluator may be overestimating the true number of clues present in Canadian accidents. Unfortunately, as we do not have any sound means to determine the actual clue prevalence, it is impossible to state the magnitude or significance of this possible overestimation.

5. DISCUSSION

5.1 Missing values

Each clue probably has its own missingness mechanism, and there is no simple, broad treatment applicable to all. The only common feature among the clues for knowable missing values is that more research and investigation resulted in better knowledge and fewer missing values. For many accidents in the Canadian dataset studied here, the inherently incomplete background data could only be supplemented by contacting a witness or accident investigator to help fill in the blanks. In addition, a more in depth analysis of the missing values may be able to provide some insight into the missingness mechanism and identify an appropriate treatment on a case-by-case basis. For future accidents it is likely worth the extra effort to ensure that important data is collected at the time of the accident; however, investigators would require significant foresight to predict changing priorities and contributing factors which might be identified in the future.

In this study, hazard scores and prevention values for the Canadian dataset were calculated based on the three missingness assumptions: missing values represent absence, missing values represent presence, and values are missing at random. Prevention values for the first two assumptions bracket the true value for the dataset, while the third allows direct comparison with published Avaluator prevention values.

5.2 Hazard Scores and Prevention Values for the OCM

As the purpose of this study was to calculate historic accident prevention values of the OCM for the Canadian accident dataset, the focus should be on the clue thresholds and recommended actions as presented in the Avaluator product. The Avaluator states “travel not recommended” where five, six, or seven clues are present. Therefore, the prevention value of interest is that for four or fewer clues. The results presented here showed that the ‘true’ prevention value is between 80.2%

and 21.1% for that threshold. Figure 5 shows the binomial 95% confidence limits for those values. The prevention value for the dataset subject to casewise deletion of accidents with missing values (i.e. values assumed missing at random) was 63.2%, with binomial 95% confidence interval between 38 and 84% for four or fewer clues. The large range is partly due to the low number of accidents (n=19) with a complete clue set.

Stating the accident prevention values more confidently would require a much more thorough investigation of the missingness mechanism for each clue individually. This is not possible at present given the limits of the dataset and the time available for further investigation and research; however, a focus of future efforts on this topic should be to reduce the number of missing data rather than devise a method of dealing with them.

5.3 Comparison with the Avaluator

The dataset used to calculate the published prevention values for the Avaluator (e.g. Haegeli and McCammon, 2006; McCammon and Haegeli, 2007) was subject to a casewise deletion of incomplete records or those with missing values (n = 252). As such, the most meaningful comparison with the current dataset is that which has been treated the same way (n = 19). The Avaluator states that 77% (binomial 95% confidence interval 71.3% to 82.0%) of accidents would have been prevented if travel were restricted to four or fewer clues present. This range lies within the 95% confidence interval for the current dataset (n = 19) for the same clue threshold, which had a prevention value of 63%. Furthermore, the hazard score distributions of these two groups were shown to not differ statistically. All of this suggests that while not necessarily identical, the Avaluator OCM and its recommendations apply to the Canadian accident database with a similar efficacy to that for the US dataset.

Of note is the fact that the Avaluator hazard score distribution was not significantly different from the group with missing values assumed to represent presence of a clue. This is notable both because of how similar the distributions are (K-S D = 0.0592, p = 0.957), and how similar the prevention values are. Given the fact that the *unknown* = *yes* group represents the absolute upper limit of hazard scores - and therefore prevention values - at a given threshold for the current dataset, the published Avaluator prevention values may be higher than the true value for the Canadian dataset. This would not necessarily affect the efficacy of the Avaluator OCM, since much of the difference exists in situations where travel is not recommended anyways. Alternatively, the *unknown* = *yes* assumption may be the most appropriate for this dataset, even if it is not the controlling mechanism for missingness. For example, the natural prevalence for several very well known clues favours presence (e.g. loading, trap, rating present in at least 60% of accidents, path in at least 57%) and in any scenario other than the *unknown* = *no* mechanism the total number of clues found present must increase as missing values are reduced. Until the true clue prevalence is known, any interpretations are speculative.

6. CONCLUSIONS

Any consistent missingness mechanism that may be present in the Canadian accident records is masked by the incompleteness of the database. The only effective method of dealing with missing data is to spend more time searching for background documents and consulting original sources. No other simple treatment or assumption is valid, although missing at random may be the most reasonable.

Accident prevention values calculated from the Canadian dataset were not significantly different from those published for the Avaluator when casewise deletion of missing values was used. This means that the prevention values published for the Avaluator are interchangeable with those for Canadian accidents when similar missing value assumptions are employed, and that the U.S. dataset is an acceptable proxy for Canadian accidents, just as McCammon and Haegeli (2006) found.

The hazard score distribution of the dataset used to validate the Avaluator lies within the range of *possible* values for Canadian accidents as defined by the limiting assumptions about missingness. It is not different from the assumption of *unknown* = *yes* distribution, at a much higher significance than the Canadian dataset treated with casewise deletion of missing values. This is notable because the *unknown* = *yes* distribution must be overestimating the clue prevalence and hazard scores, as it is highly unlikely that the clue was actually present in every instance of missing data. While not significantly different, the *unknown* = *delete* distribution from the Canadian dataset is less similar to the *unknown* = *yes* distribution than the Avaluator.

Given the poor understanding of the missingness mechanism masked by an incompletely researched dataset, the most important priority to determine the *true* accident prevention values for the Avaluator in Canada must be to collect complete, detailed records for past and future avalanche accidents. Only with a complete dataset can meaningful accident prevention values be calculated, after which the appropriate clue threshold-based travel recommendations may be determined.

9. ACKNOWLEDGEMENTS

Thanks to the Canadian Avalanche Centre for unrestricted access to the Canadian avalanche accident database. This research was supported by the ADFAR 2 project.

10. REFERENCES

Floyer, J., 2008. Review of the Obvious Clues Method. Report for the Canadian Avalanche Centre, Revelstoke, BC. 33pp.

Haegeli, P. and I. McCammon, 2006. Avaluator: Avalanche Accident Prevention Card and Booklet, Canadian Avalanche Centre, Revelstoke, BC.

Haegeli, P., I. McCammon, J. B. Jamieson, C. Israelson and G. Statham, 2006. The Avaluator - a Canadian rule-based avalanche decision support tool for amateur recreationalists. Proceedings: ISSW 2006, Telluride, CO, 254-263.

McCammon, I., 2006. Prediction, prevention, and the language of risk. Explaining the Avaluator to students., Avalanche.ca, 79, 28.

McCammon, I. and P. Haegeli, 2005. Comparing avalanche decision frameworks using accident data from the United States, Proceedings: ISSW 2004, Jackson Hole, WY, 502-511.

McCammon, I. and P. Haegeli, 2006. Evaluation of a rule-based decision aid for recreational travelers in avalanche terrain, Proceedings: ISSW 2006, Telluride, CO, 234-243.

McCammon, I. and P. Haegeli, 2007. An evaluation of rule-based decision tools for travel in avalanche terrain, Cold Regions Science and Technology, 47, 193-206.

Uttl, B., and K. Kissinger, 2009. Pitfalls in the analyses of accident records: the meaning of missing values. Proceedings: 2009 ISSW, Davos, Switzerland, 614-618.

Uttl, B., Uttl, J., and M. Henry, 2008a. The Avaluator Avalanche Accident Prevention Card: Facts, Fictions, Controversies. Proceedings: ISSW 2008, Whistler, BC, Canada, 445-454

Uttl, B., Henry, M., and J. Uttl, 2008b. Avaluator's Obvious Clues Prevention values: Are they replicable? Proceedings: ISSW 2008, Whistler, BC, Canada, 538-547.

Uttl, U., Kibreab, M., Kissinger, K., and J. Uttl, 2009a. Trend analysis of Canadian avalanche accidents: the Avaluator avalanche accident prevention card has not reduced the number of accidents. Proceedings: ISSW 2009, Davos, Switzerland, 610-613.

Uttl, U., Kissinger, K., Kibreab, M., and J. Uttl, 2009b. Avaluator's Obvious Clues prevention values are inflated: Evidence from Canadian avalanche accidents. Proceedings: ISSW 2009, Davos, Switzerland, 619-622.