NORTH AMERICAN AVALANCHE DANGER SCALE: DO BACKCOUNTRY FORECASTERS APPLY IT CONSISTENTLY?

Brian Lazar[1]*, Simon Trautman[2], Mike Cooperstein[1], Ethan Greene[1], Karl Birkeland[2]

[1]Colorado Avalanche Information Center, Boulder, CO, USA
[2]USDA Forest Service National Avalanche Center, Bozeman, MT USA

ABSTRACT: The North American Avalanche Danger Scale is a tool used by backcountry avalanche forecasters to communicate the potential for avalanches to cause harm or injury to backcountry travelers. Danger ratings are the most basic component of the public forecast, providing the foundation for more nuanced descriptions of avalanche conditions. In 2010, the United States, Canada, and New Zealand adopted a consistent, five-tiered danger scale. Although widely used, we do not know how consistently the danger scale is applied both within and between avalanche forecasting operations. To address this question, we developed ten scenarios capturing a variety of avalanche conditions at the mountain range scale. We derived the scenarios from real avalanche forecasts issued by various avalanche centers throughout North America. Avalanche forecasters in the United States, Canada, and New Zealand reviewed each scenario and assigned a single danger rating for the forecast period. Results indicate that although most respondents choose ratings within one step of each other, individual forecasters can arrive at different conclusions when presented with identical information. Additionally, it appears that there are regional and/or cultural differences in how forecasters assign danger ratings.

KEYWORDS: avalanche danger scale, avalanche danger ratings, avalanche forecasting

1. INTRODUCTION

Backcountry avalanche advisories are the cornerstone of public avalanche safety. In the United States alone, more than 1.5 million users accessed avalanche advisories a total of ~6.5 million times during the 2015/16 season (www.fsavalanche.org). In addition, educators increasingly use them as a foundation to introduce and implement planning and decision-making strategies for those looking to venture into avalanche terrain (Haegeli, 2010; Zacharias et al., 2015; KBYG, 2015). These public safety products are constructed around the North American Avalanche Danger Scale, and require forecasters to come to a conclusion on the regional avalanche hazard, and then communicate this hazard to their audience. The simplest and most fundamental information in a public avalanche advisory is the danger rating. This discrete piece of information alone has a major influence on decision-making among backcountry recreationalists (Jamieson et al., 2009). As a result, how forecasters assign danger a rating has a big impact on public safety, perhaps the biggest single impact for public safety.

The five-level avalanche danger scale was first developed in Europe in 1993. It was slightly modified and introduced to North America the following year, though each country employed somewhat different descriptors for the danger levels. To address this inconsistency, a group of Canadian and U.S. avalanche professionals began developing a single system for North America in 2005. By 2010, this effort produced a revised North American Public Avalanche Danger Scale (Statham et al., 2010b), and an accompanying conceptual model of avalanche hazard (Statham et al., submitted).This system was first implemented across North America for the 2010-11 snow season, and for the 2011 season in New Zealand.

The 2010 danger scale fulfilled two purposes. First, it provided an updated and consistent public communication tool. Second, it served as a guidance document and reference for forecasters (Statham et al., 2010b). When combined with the conceptual model of avalanche hazard (Statham et al., submitted), a basic framework exists to promote consistency in the application of the danger scale.

LaChapelle (1980) posited that although avalanche forecasters employ different inductive reasoning paths to arrive at an avalanche hazard assessment, they typically come to the same conclusion. Avalanche forecasters have now been

---

* *Corresponding author address:* Brian Lazar, Colorado Avalanche Information Center, 325 Broadway WS1, Boulder, CO 80305;
tel: 303-499-9650; fax: 303-499-5018;
email: brian.lazar@state.co.us

using the North American Public Avalanche Danger Scale in three countries for five snow seasons. The primary question this paper addresses is: "Do experienced avalanche forecasters assign the same danger rating when presented with identical information?"

2. METHODS

2.1 *Scenarios*

We developed ten scenarios capturing a variety of avalanche conditions and problems at the mountain range scale (Tbl 1). The scenarios were derived from real avalanche forecasts issued by various avalanche centers throughout North America. To normalize the locations of the scenarios, we omitted specific location names, and replaced them with fictional place names, such as Snowy Pass, Snowy Valley, or Stormy Peak. We noted that these names did not refer to actual locations. We presented each scenario with consistent components and structure: weather information, snowpack information, and avalanche occurrence data.

Weather information included a short weather summary and a table of prior (last 48 hours) and forecast (next 24 hours) weather. Snowpack information included a summary of past and current snowpack conditions, as well as a baseline snow profile of mid-level elevations (we assume that this information provides some seasonal context). Avalanche information included a textual description of recent avalanche activity, a table listing the number of reported natural and triggered avalanches and their size ranges over the previous seven days, and images of recent avalanches. You can view the scenarios here: https://goo.gl/forms/G95PhtBQNj11Cs4B3

2.2 *Danger Ratings*

A total of 68 avalanche forecasters in the United States (n=43), Canada (n=14), and New Zealand (n=11) reviewed and completed these scenarios. We told them it was 6:00 AM, and they were forecasting for the next 24-hour period. We asked them to issue a single danger rating for each scenario, choosing the highest danger rating they thought would be reached in the next 24-hour period.

Tbl. 1: Summary of forecast scenarios

#	*Avalanche Problem(s)*	*Forecast danger rating*	*Brief description*
S1	Persistent Slab	CON(3)	Moderate loading on a sensitive pwl
S2	Deep, Persistent Slab	CON(3)	Incremental loading on deep pwl
S3	Wet Slab/ Loose Wet	HIGH(4)	Rain on substantial new snow
S4	Deep, Persistent Slab	CON(3)	Substantial rapid loading on deep pwl
S5	Persistent Slab	LOW(1)	Worrisome snow structure with many pwl, but only small storm snow avalanches reported
S6	Deep, Persistent Slab	CON(3)	Incremental loading on an old pwl, and skier triggered slides increasing in size
S7	Loose Wet	MOD(2)	Melt-induced loose wet activity with small slides hitting roads in town
S8	Wind Slab/ Storm Slab	MOD(2)	12" (30cm) of new snow last 24 hrs followed by strong winds
S9	None	LOW(1)	Little snow in previous two weeks, one D1 avalanche a week ago
S10	Persistent Slab	HIGH(4)	Massive loading event on a snow structure with several reactive pwl

Note: pwl = persistent weak layer

2.3 *Data Collection*

We collected responses through an on-line survey, and we restricted respondents to forecasters currently employed to issue backcountry avalanche advisories for the public. In addition to danger ratings we also recorded the associated forecasting operation for each respondent, and the top three environmental factors most influential in foercasters' danger rating assessments.

2.4 *Statistical Comparison*

Since our data are categorical, we used non-parametric tests for our analyses. To test for differences between the distributions of danger ratings for each country, we first used the Kruskal-Wallis test (Kruskal and Wallis, 1952) to assess whether the results from the three countries for a given scenario likely originated from the same distribution. For the five scenarios with a $p < 0.05$ we then applied the Fisher Exact test (Daniel, 1990) to evaluate the differences between countries. We also did this to a sixth scenario (S3) that did not meet the requirements of the Kruskal-Wallis test but visually appeared to have different distributions by country.

For the Fisher Exact test we chose $\alpha < 0.05$ as our level of significance. Thus, we consider $p < 0.05$ to be good evidence that the distributions differ, while higher values suggest the populations are not significantly different. In this work we sampled more than two thirds (68%) of our population (86% in the US, 41% in Canada, and 69% in New Zealand), so it is not necessary for us to be overly conservative with our level of significance. As such, we feel that the $\alpha < 0.05$ is appropriate for our analyses.

3. RESULTS

3.1 *Danger Ratings by Scenario*

Most forecasters assigned danger ratings similar to one another and to the original forecasted danger level. Scenario 9 had the highest universal agreement with 88% of the respondents choosing LOW danger and the remaining 12% choosing MODERATE. For all other scenarios, the most frequently selected danger rating accounted for 40 to 63% of total responses. The two most frequently selected danger ratings combined accounted for 73 to 98% of total responses. So, most forecasters ended up assigning one of two adjacent danger ratings.

Despite this general agreement, results also indicate that it is not uncommon for forecasters to assign different danger ratings when presented with identical weather, snowpack, and avalanche information (Figs. 1 and 2). Additionally, in some cases forecasters from the same operation were two danger ratings apart. We surveyed forecasters representing 21 discrete operations. Of the 13 operations with three or more respondents, 10 of them (77%) had at least one scenario with a spread of three danger ratings. All seven operations with four or more respondents had at least one scenario with a spread of three danger ratings.

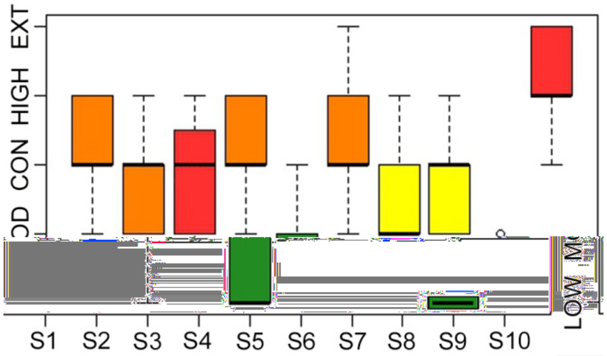

Fig. 1: Summary of danger ratings for all respondents (n=68). The box represents the interquartile range, the dark horizontal line marks the median value, and the whiskers represent the range excluding outliers. The circle indicates an outlier (defined as more than 1.5 times greater than the upper quartile). The color corresponds to the actual danger rating assigned on the day from which the scenario was derived.

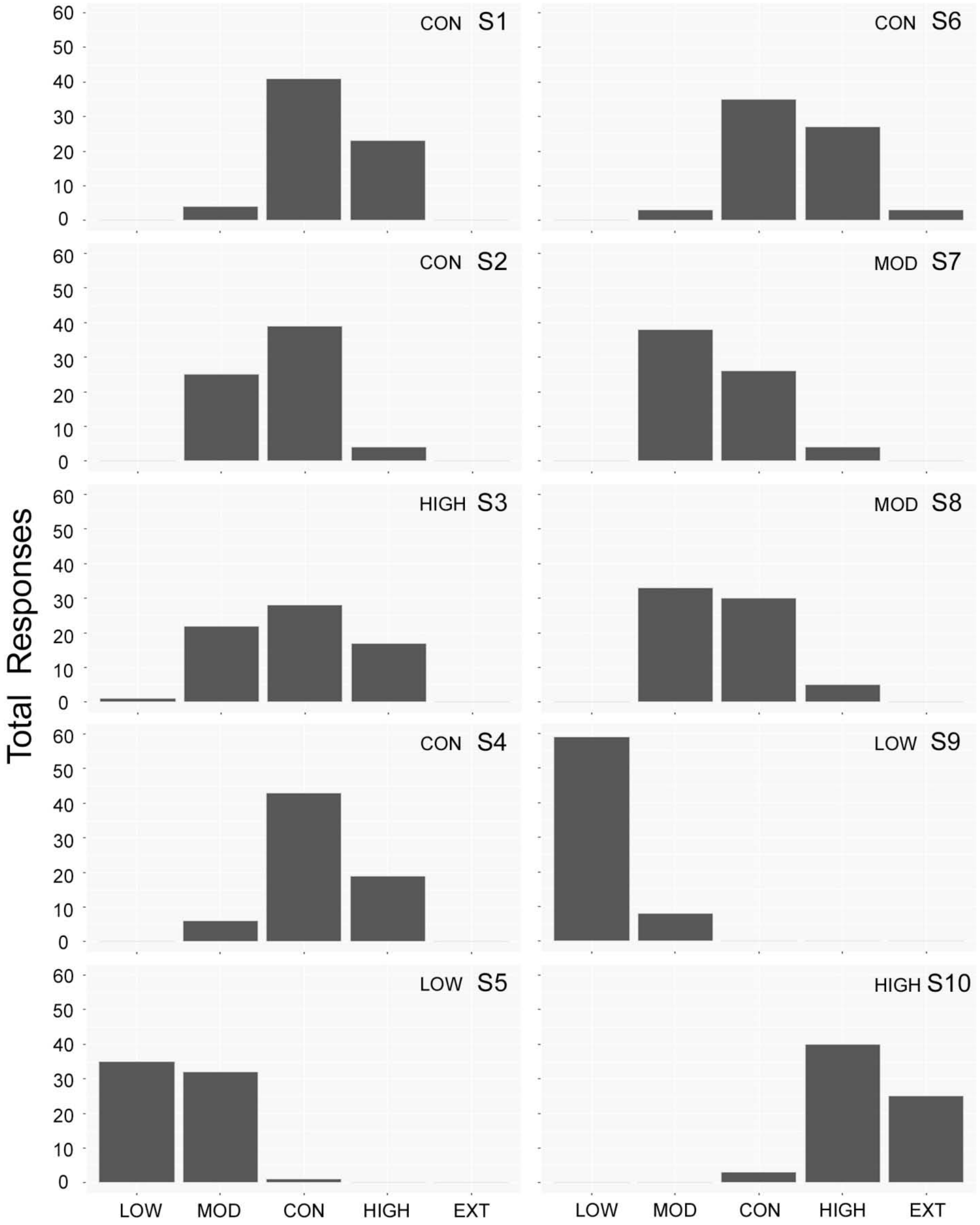

Fig. 2: Total number of responses by danger rating for each scenario. The assigned danger rating for the actual scenario is listed in the upper right corner of each panel. S1= scenario 1 through S10=scenario 10, respectively.

The data collected from this survey show the following (Figs. 1 and 2):

- No scenario had a single danger rating for all respondents.
- Most scenarios (90%) had a spread of at least three danger ratings.
- One scenario (S9) had a spread of only two danger ratings. Most respondents (88%) assigned a LOW danger rating.
- Two scenarios (S3 and S6) had a spread of four danger ratings.
- The spread of danger ratings did not appear to be affected by where on the danger scale spectrum the scenario lies (Fig. 1).
- Three scenarios (S1, S4, and S10) skew towards a higher danger rating. All three of these scenarios had actual danger ratings of CONSIDERABLE or HIGH.
- Four scenarios (S2, S5, S7, and S8) skew towards a lower danger rating. Three of these four scenarios had actual danger ratings of LOW or MODERATE.
- In nine of the ten scenarios (all but S3), the most commonly selected danger matched the actual danger rating.

3.2 Danger Ratings by Country

A visual comparison suggests differences in danger rating by country, with U.S. forecasters generally rating the avalanche danger lower than Canada and New Zealand (Fig. 3). Our initial Kruskal Wallis test showed that five of our scenarios had p values < 0.05 (Scenarios S1, S2, S6, S7, and S10). Scenarios S3 did not meet the assumption of equal variance required by the Kruskal-Wallace test. However, a visual assessment of this scenarios suggested that there might be a difference in distributions between countries.

Applying the Fischer Exact test to assess differences between countries showed statistically significant ($p < 0.05$) differences between the distributions of danger ratings by countries existed in all six scenarios we tested: S1, S2, S3, S6, S7, and S10 (Tbl. 3, full table of results presented in Appendix A).

- Canada and New Zealand are only statistically different in S2 and S3.
- The U.S. differs from Canada in five of the scenarios.
- The U.S. differs from New Zealand in three scenarios.
- Most differences between countries (4 of 5, or 80%) occur for scenarios originally assigned CONSIDERABLE or above.

Tbl. 3: Summary of results from the Fisher Exact test of danger rating by country. US=United States, CA=Canada, and NZ=New Zealand.

#	*Summary of Fisher Exact Results*
S1	US differs from CA and NZ
S2	US, CA and NZ are all different
S3	US differs from CA, and CA differs from NZ
S4	US, CA, and NZ show no differences
S5	US, CA, and NZ show no differences
S6	US differs from NZ
S7	US differs from CA
S8	US, CA, and NZ show no differences
S9	US, CA, and NZ show no differences
S10	US differs from CA

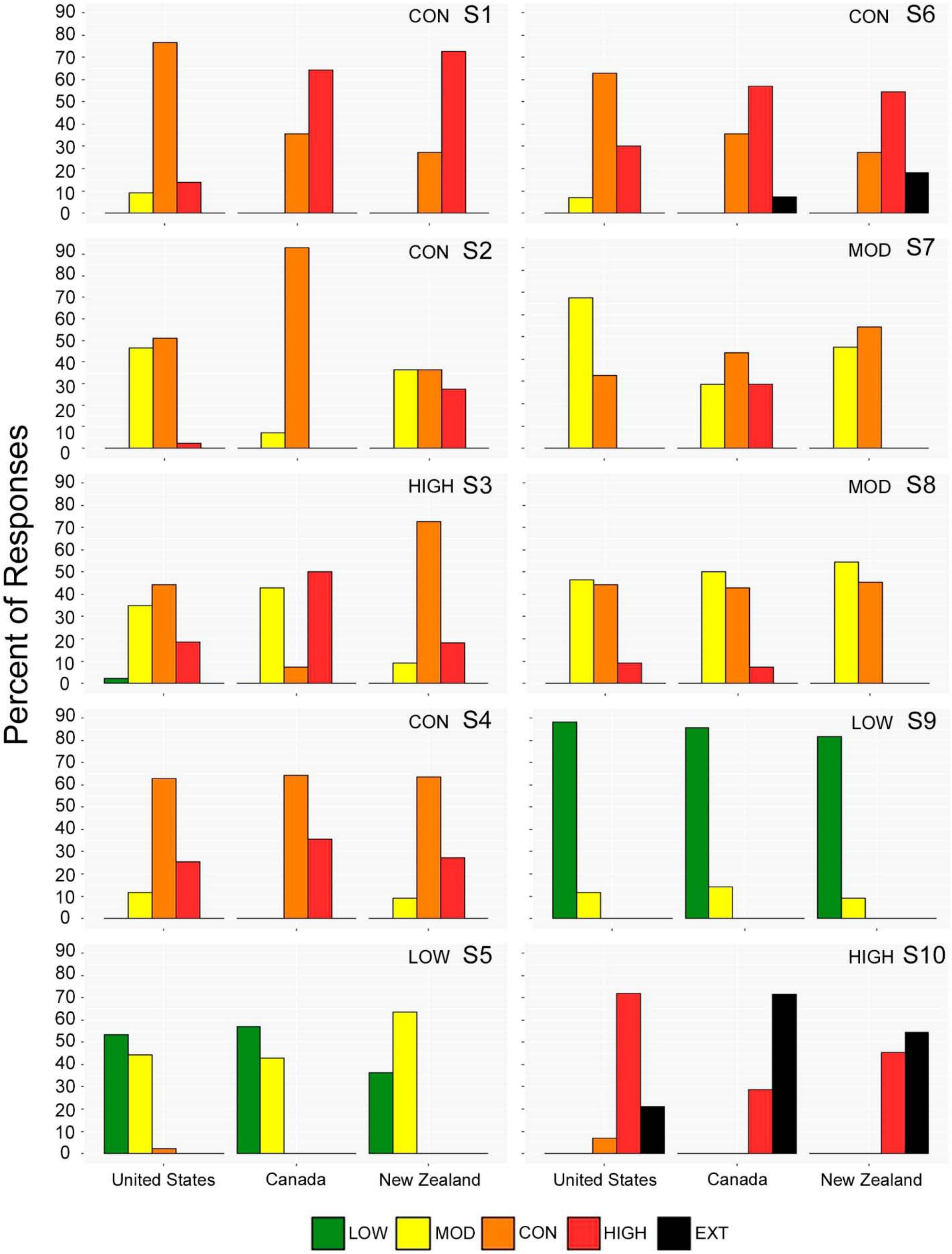

Fig. 3: Percentage of danger rating responses by country for each scenario. S1= scenario 1 through S10=scenario 10.

4. DISCUSSION

4.1 Danger Ratings

Are forecasters applying the danger scale consistently? In nine of ten scenarios the most commonly assigned rating matched the actual danger. In addition, most respondents chose one of the two most selected ratings in each scenario. Although these results are generally encouraging, the five step scale is built around distinct travel advice (by level) and a one-step difference delivers a very different message to the public on how to manage their risk in the field.

Our results suggest that individual forecasters can arrive at different conclusions when presented with identical information. Interestingly, the scenarios with the largest spread of responses (S3 and S6) replicate conditions that are traditionally problematic for forecasters: wetting previously dry snow, and incremental loading of a deep persistent weak layer.

In addition, and similar to other findings (Greene et al., 2006), it appears that there are regional and/or cultural differences in how forecasters assign danger ratings. Because the information presented to the forecasters in these scenarios cannot perfectly mimic actual forecasting conditions, the results are not conclusive. However, data suggests that the U.S. tends to assign lower danger ratings than Canada and New Zealand (Fig. 3).

There is more consistency in danger ratings between countries when the actual scenario was rated LOW or MODERATE. Where significant differences existed between countries, the danger was rated CONSIDERABLE or above in four of the five scenarios. In all of these cases it appears that the U.S. leans toward lower ratings. U.S. forecasters are also much less likely to assign an EXTREME danger rating than their peers in Canada or New Zealand (Fig. 3).

Inconsistencies in assigned danger ratings suggest that forecasters (and public safety messaging) may benefit from guidance beyond the descriptors contained in the danger scale and conceptual model. If consistent application of the danger scale between operations is desirable, the importance of training and calibration between forecasters on an operational, national, and potentially international scale is clear.

5. CONCLUSIONS

This exercise provides valuable insight into how forecasters from the U.S., Canada, and New Zealand use and apply the North American Avalanche Danger Scale. Encouragingly, most forecasters assign danger ratings within one step of one another. However, the fact that forecasters can arrive at different danger ratings when supplied with identical information highlights the need for discussion and calibration between team members. Discussions within a highly functioning team not only improve the quality of the forecast, but also minimize inconsistencies within an operation.

Based on our scenarios, there appear to be differences in the way the US, CA, and NZ apply danger ratings. Our data suggest that US forecasters are generally more likely to assign lower danger ratings and are less likely to use a rating of EXTREME than their commonwealth counterparts. Inconsistencies between operations would also likely be reduced from consistent inter-operational guidance and/or training.

There are clearly some inherent limitations to our study. When reading these scenarios, the forecast is obviously not integrated through time and therefore forecasters cannot minimize uncertainty through iteration; indeed, LaChapelle (1980) discusses in detail the necessity for continuously monitoring the snowpack throughout the season. Many forecasters find that leaving their forecast areas for even a few days in the middle of the season creates forecasting challenges. In our case we used snowpack descriptions and profiles as an imperfect and incomplete proxy for prior knowledge, but this really only provides a small sliver of the information that forecasters typically have available about the current season. In addition, our scenarios require forecasters to work alone rather than in a team where team members can bounce ideas and information off of each other to come up with a better assessment of the current conditions.

Despite its limitations, this study provides valuable insights into local and regional differences in the application of the avalanche danger scale. Our dataset contains much more information, such as the environmental factors that forecasters weighed most heavily in their decisions. Further analyses planned for these data are likely to provide additional insights into how avalanche forecasters arrive as specific danger ratings in their avalanche assessments.

ACKNOWLEDGEMENTS

We would like to acknowledge Colin Zacharias for thoughtful conversations and input on constructing this study. The idea for this exercise derived in part from similar exercises he has run at professional training workshops. We would also like to thank all the forecasters who participated in this survey. Without your willingness to take the time to thoughtfully complete the exercise, this study would not have been possible.

REFERENCES

Daniel, W. W., 1990: *Applied Nonparametric Statistics*. PWS-Kent, Boston, 120-125.

Greene, E., T. Wiesinger, K. W. Birkeland, C. Coleou, A. Jones, and G. Statham, 2006: Fatal avalanche accidents and forecasted danger levels: Patterns in the United States, Canada, Switzerland and France. Proceedings of the *International Snow Science Workshop*, Telluride, CO, 640-649.

Haegeli, P. (2010). Avaluator V2.0—avalanche accident prevention card. Revelstoke, BC: Canadian Avalanche Centre.

Jamieson, J. B., Haegeli, P., & Schweizer, J. (2009). Field observations for estimating the local avalanche danger in the Columbia Mountains of Canada. Cold Regions Science and Technology, 58(1–2), 84–91.

Know Before You Go, 2015. Kbyg.org

Kruskal; Wallis (1952). "Use of ranks in one-criterion variance analysis". *Journal of the American Statistical Association*. **47** (260): 583–621

LaChapelle, E. R. (1980). The Fundamental Processes in Conventional Avalanche Forecasting. Journal of Glaciology,26(94), 75–84.

Miller, R. G. Jr., 1991: *Simultaneous Statistical Inference.* Springer-Verlag, New York, pp. 299.

Statham, G., Haegeli, P., Birkeland, K., Greene, E., Israelson, C., Tremper, B., Stethem, C., McMahon, B., White, B., Kelly, J. 2010b. The North American public avalanche danger scale. Proceedings of the 2010 International Snow Science Workshop in Squaw Valley, California, USA.

Statham, G., Haegeli, P., Birkeland, K., Greene, E., Israelson, C., Tremper, B., Stethem, C., McMahon, B., White, B., Kelly, J. (Under review). The Conceptual Model of Avalanche Hazard. In preparation for Natural Hazards.

Zacharias, C., Pritchett, B., Lazar, B., Murphy, T., & Schwartz, H. (Eds.). (2015).AIARE 1 Avalanche Course: Decision Making in Avalanche Terrain - Student Manual. Gunnison, CO: American Institute for Avalanche Research and Education.

Appendix A – Fisher Exact Test Results by Scenario

Scenario 1

	NZ	CA	US
NZ	---	0.99	< 0.01
CA	0.99	---	< 0.01
US	< 0.01	< 0.01	---

Scenario 2

	NZ	CA	US
NZ	---	< 0.01	0.038
CA	< 0.01	---	< 0.02
US	0.038	< 0.02	---

Scenario 3

	NZ	CA	US
NZ	---	<0.01	0.32
CA	<0.01	---	0.021
US	0.32	0.021	---

Scenario 4

	NZ	CA	US
NZ	---	0.65	0.93
CA	0.65	---	0.51
US	0.93	0.51	---

Scenario 5

	NZ	CA	US
NZ	---	0.42	0.46
CA	0.42	---	0.99
US	0.46	0.99	---

Scenario 6

	NZ	CA	US
NZ	---	0.72	<0.02
CA	0.72	---	0.062
US	<0.02	0.062	---

Scenario 7

	NZ	CA	US
NZ	---	0.2	0.29
CA	0.2	---	< 0.01
US	0.29	< 0.01	---

Scenario 8

	NZ	CA	US
NZ	---	0.99	0.78
CA	0.99	---	0.99
US	0.78	0.99	---

Scenario 9

	NZ	CA	US
NZ	---	0.99	0.99
CA	0.99	---	0.99
US	0.99	0.99	---

Scenario 10

	NZ	CA	US
NZ	---	0.43	0.95
CA	0.43	---	< 0.01
US	0.95	< 0.01	---