Urban Forestry & Urban Greening 24 (2017) 109–115

Contents lists available at ScienceDirect

Urban Forestry & Urban Greening

journal homepage: www.elsevier.com/locate/ufug

Original article

Impact of assessor on tree risk assessment ratings and prescribed mitigation measures

Andrew K. Koeser[a,*], E. Thomas Smiley[b]

[a] Department of Environmental Horticulture, CLCE, IFAS, University of Florida – Gulf Coast Research and Education Center, 14625 County Road 672, Wimauma, FL 33598, United States

[b] Bartlett Tree Research Laboratory, F.A. Bartlett Tree Expert Company, 13768 Hamilton Road, Charlotte, NC 28278, United States

ARTICLE INFO

Keywords:
Arboriculture
Certification
Qualification
TRAQ
Tree hazard
Urban forestry

ABSTRACT

Tree risk assessment is an inherently human endeavor that can be influenced by risk perception, risk acceptance, and professional bias. Tree risk assessments from 296 arborists were evaluated to assess tree- and assessor-based factors that influenced ratings. Additionally, we investigated sources of variability associated with the main inputs of risk assessment – likelihood of impact ratings, likelihood of failure ratings, and consequences of failure ratings. Finally, we assessed the factors that influenced prescribed mitigation measures. Results indicate that professionals with training and industry credentials had lower risk ratings and were less likely to prescribe more active mitigation measures like tree removal. More notably, there was significant variability among raters, with the likelihood of impact and consequence of failure serving as the most variable factors in tree risk assessment.

1. Introduction

Urban trees provide economic value, serve important ecological functions, and can have a positive impact on human health and well-being (Roy et al., 2012). This wide array of benefits is typically maximized as a tree grows in size (Leibowitz, 2012). Unfortunately, the risks posed by urban trees also increase with age. Large trees have target zones that are proportional to their height (Smiley et al., 2011; Dunster et al., 2013) and the weight of large trees can cause significant injury or property damage if a structural failure occurs (Smiley et al., 2011; Dunster et al., 2013). Additionally, research has shown that tree defects like internal decay (which can predispose a tree to failure) are more prevalent in larger and older trees (Tate, 1984; Luley et al., 2009; Koeser et al., 2016a).

Arborists and urban foresters are often tasked with assessing the risk posed by either an individual tree or population of trees. While the prevention of tree-induced injury or property damage is a universal concern among urban tree care professionals, few researchers have attempted to validate or compare the risk assessment methods employed by those working in the field (Hickman et al., 1995; Rooney et al., 2005; Norris, 2007; Klein et al., 2016; Koeser et al., 2017). The tree care industry has only recently worked to establish Best Management Practices, industry standards, standardized training, and qualifications for those assessing tree risk (Ellison, 2005; Koeser, 2009; ANSI, 2011; Smiley et al., 2011; Dunster et al., 2013). The now-standard method for tree risk assessment in the United States is the ISA Best Management Practice (Smiley et al., 2011). This method combines the likelihood of tree failure with the likelihood of impact in one matrix and then combines that likelihood with the consequences of failure in a second matrix (matrices in Appendix 2).

All basic visual risk assessment methods commonly used in the United States are qualitative in nature (Koeser et al., 2016b). Even risk assessment methods that strive to be quantitative (Ellison, 2005) are forced to contend with at least one truly qualitative stumbling block – the assessment of the likelihood of failure (Ellison, 2007; Koeser et al., 2015). The reliance on one or more subjective risk rating inputs makes training critically important, as risk assessment results can vary significantly from assessor to assessor (Norris, 2007; Stewart et al., 2013).

Past research has shown the impact of industry credentials and experience on risk ratings. In determining what factors influence risk assessment ratings, Koeser et al. (2015) found that while defect severity was a driving factor for the three groups of participants who participated in the experiment (i.e., non-professional, professional, and advanced professional), advanced professionals gave more weight to target proximity than did non-professionals and professionals. This is notable, as target proximity is considered to be a key factor in risk assessment (Ellison, 2005, 2007; Koeser et al., 2015). Similarly, when comparing how arborists gauged the likelihood of failure across various levels of risk assessment (i.e., limited visual, basic, advanced), Koeser

* Corresponding author.
E-mail addresses: akoeser@ufl.edu (A.K. Koeser), tsmiley@bartlettlab.com (E.T. Smiley).

http://dx.doi.org/10.1016/j.ufug.2017.03.027
Received 19 October 2016; Received in revised form 29 March 2017; Accepted 29 March 2017
Available online 10 April 2017
1618-8667/

et al. (2017) found that arborists with the International Society of Arboriculture (ISA) Certified Arborist credential rated failure potential lower than did non-certified arborists. In another study, Klein et al. (2016) found that certified arborists rated target occupancy differently than did non-certified arborists when shown traffic count data.

For this project, we asked 296 arborists to assess risk for multiple trees as part of a nationwide series of risk training exercises. Unlike past research, this effort looked at the entire risk assessment process, identifying factors that influenced mean risk ratings and rating consistency. With nearly 300 professionals participating in this research effort, this was the largest undertaking of its kind to date. The results of this work are intended to identify potential sources of bias and variation in risk assessment to help guide industry training efforts and credential initiatives.

Table 1
Modified DOSPERT Scale used to assess each participant's perceived level of risk for hypothetical scenarios related to health/safety and finance. Participants were asked to give their gut-level assessment of risk for each of the following on a ten-point scale (1 = minimal risk, 10 = extreme risk).

Scenario	Domain
R1. Investing 5% of your annual income in a very speculative stock	Finance
R2. Driving a car without wearing a safety belt	Health/safety
R3. Driving a car without automotive insurance	Finance
R4. Sunbathing without sunscreen	Health/safety
R5. Walking home alone at night in an unsafe area of town	Health/safety
R6. Betting a day's income on the outcome of a sporting event or race	Finance

2. Materials and methods

2.1. Study sites, trees, and participants

This experiment was conducted as part of a series of tree care training program that was held at multiple locations across North America (Appendix 1). A total of 296 participants, with varying levels of education and industry credentials, were asked to complete three risk assessments using the basic (Level 2) assessment method outlined in the ISA tree risk assessment BMP (Smiley et al., 2011) and adopted as part of the ISA Tree Risk Assessment Qualification (TRAQ) program (Dunster et al., 2013). Three different trees were selected for inspection at each training site. The tree species assessed are detailed in Appendix 1.

2.2. Survey design and delivery

The survey tool was designed as a truncated version of the ISA Basic Tree Risk Assessment Form (http://www.isa-arbor.com/education/resources/basictreeriskassessmentform_firstedition.pdf) and included only the core elements required to make a risk assessment given the likelihood of impact, likelihood of failure, and consequences of failure (see Appendix 2). Relevant site, site history, and species information that may be included when using the Basic Tree Risk Assessment Form were not collected for this experiment to limit sources of variation. Definitions for advanced terminology were included as part of the survey form.

In addition to the actual risk assessment data, participants were asked to answer basic demographic questions that could potentially influence their perception and acceptance of risk (e.g., age, gender, parental status, and homeowner status) (Koeser et al., 2015). Participants were also asked to note their highest level of education, as well as any industry credentials held (see Appendix 2). Additionally, participants were asked to complete a modified Domain-Specific Risk-Taking (DOSPERT) Scale (Blais and Weber, 2006) to assess their perceived level of risk for three scenarios related to health/safety and three domain-specific scenarios related to finance (Table 1).

The training sessions and surveys were delivered by the same individuals at all study locations. Participants were instructed (verbally and within the survey form directions) to use a three-year timeframe for assessment.

2.3. Analysis

Overall risk ratings were analyzed as a mixed regression model using the nlme package in R (Pinheiro et al., 2013; R Core Team, 2013). Fixed effects included age, gender, home ownership (yes/no), status as parent (yes/no), past risk assessment experience (yes/no), average financial DOSPERT rating, average health DOSPEERT rating, the presence or absence of the ISA Certified Arborist Credential, the presence or absence of the ISA Tree Risk Assessment Qualification, and portion of the tree assessed (i.e., branches, trunk, and roots). Random effects included the tree assessed, the assessor, and the date of the training.

Variability was assessed using three tests of equal variance. Given the nature of the data (ordinal ratings), the Fligner-Killeen Test of Homogeneity of Variances was initially used as it is based on the median rather than the mean. Two more commonly used tests of equal variance (Bartlett's test of equal variance and Levene's test of equal variance) were also run for the sake of comparison. Patterns of significance for these tests of equal variance were assessed using the prop.test() function [with statistical significance for the test of equal variance = success (R Core Team, 2013)] to determine if assessments were more or less variable given the portion of the tree assessed (i.e., branches, trunk, and roots). This process was repeated to compare variability among the three key inputs: likelihood of impact, likelihood of failure, and consequences of failure.

Additionally, variance partitioning was used to assess the impact of the individual arborist and the trees being assessed on the risk rating. Three fixed effects models were fit (i.e., featuring arborist identifier and tree identifier as single and combined predictors). The resulting R^2 values from these models were inputted into the varPar() function in the modEvA package in R (Barbosa et al., 2015) for analysis and visualization.

Finally, prescribed mitigation measures (nominal outcome variables) from the open response portion of the risk assessment form were coded for consistency and modeled using multinomial regression to determine which factors influenced risk management decisions. Modeling was conducted using the multinom() function from the nnet package in R (Venables and Ripley, 2002). Where appropriate for the tests noted above, diagnostic plots were referenced to confirm that all underlying assumptions were met for the above analyses. All conclusions were made at an $\alpha = 0.05$ level of Type I error.

3. Results and discussion

3.1. Respondent characteristics

Of the 296 arborists that participated in this experiment, 275 (92.9%) were male. The average age of our participants was 40.7 (± 12.0) years. A little over half of our participants were homeowners (159, 53.4%) and slightly fewer than half of our participants were parents (129, 43.6%). Nearly 70% percent of the participants were college graduates (205, 69.3%). Of these, 25 held an advanced degree (8.4%). Three-quarters of the participants held the ISA Certified Arborist credential (222, 75%). Forty-two participants (14.2%) were ISA Board Certified Master Arborists. In comparison, a much smaller portion of the participants held the American Society of Consulting Arborists (ASCA) Registered Consulting Arborist (5, 1.7%) or Society of American Foresters Certified Forester (3, 1.0%) credentials.

Table 2
Factors that significantly impacted likelihood of impact, likelihood of failure, consequences of failure and overall risk ratings as determined using multiple regression. Non-significant factors are omitted from table.

Risk input/rating	Factor	Coefficient	Standard error	*P*-value	95% CI lower	95% CI upper
Likelihood of impact	Tree Risk Assessment Qualified – Yes	−0.1749	0.0829	0.0361	−0.3378	−0.0119
	Structure Assessed – Trunk	0.1370	0.0364	0.0002	0.0658	0.2082
Likelihood of failure	Certified Arborist – Yes	−0.1256	0.0521	0.0168	−0.2281	−0.0232
	Tree Risk Assessment Qualified – Yes	−0.1218	0.0493	0.0143	−0.2187	−0.0248
	Structure Assessed – Trunk	−0.2241	0.0362	< 0.0001	−0.2950	−0.1531
	Structure Assessed – Roots	−0.4350	0.0366	< 0.0001	−0.5066	−0.3634
Consequences of failure	Structure Assessed – Trunk	0.5997	0.0351	< 0.0001	0.5309	0.6683
	Structure Assessed – Roots	0.4514	0.0355	< 0.0001	0.3820	0.5208
Overall risk rating	Certified Arborist – Yes	−0.1407	0.0673	0.0378	−0.2730	−0.0084
	Tree Risk Assessment Qualified – Yes	−0.2797	0.0637	< 0.0001	−0.4049	−0.1543
	Financial DOSPEERT rating	0.0566	0.0204	0.0061	0.0164	0.0968
	Structure Assessed – Roots	−0.1068	0.02735	0.0001	−0.1603	−0.0533

In looking at risk-specific credentials and experience, 102 participants (34.5%) had the ISA Tree Risk Assessment Qualification (TRAQ). On average, participants with TRAQ had held that credential for 3.3 (± 3.2) years, indicating that some of our participants may have obtained the regional credential (i.e. the ISA Pacific Northwest Chapter Tree Risk Assessor Course and Exam or TRACE) that preceded the international qualification program established in 2013. Beyond credentials and advanced training, 234 (79.1%) participants stated they had risk assessment experience. On average, these participants had 10.6 (± 9.2) years of risk assessment experience. As seen in past surveys (Klein et al., 2017), the average number of trees assessed annually was high (908.9) and variable (± 2020.5), likely reflecting that responses of a handful of municipal and utility arborists/contractors who perform limited visual assessments (i.e., drive-by and/or fly-by assessments) for large populations of trees.

3.2. Factors influencing ratings

Several factors influenced likelihood of impact, likelihood of failure, and overall risk (Table 2). Participants with the ISA Tree Risk Assessment Qualification had lower ratings for likelihood of impact and likelihood of failure (Table 2). As a result, overall risk ratings (which factors in these two inputs) were also low for ISA Tree Risk Assessment Qualified arborists. Similarly, ISA Certified arborists had lower ratings for likelihood of impact and overall risk (Table 2).

Ratings for the inputs and overall risk also varied given the portion of the tree assessed. When assessing tree trunks and roots, our participants rated likelihood of failure lower compared to when they were assessing branches (Table 2). While defects or damage may have been less prevalent in the lower portions of the tree, visibility may have also come into play. Root defects are inherently difficult to detect, which may explain why likelihood of failure ratings for roots were lower than the other two portions of the tree (Table 2). This may be a source of assessor bias as data from the International Tree Failure Database (http://ftcweb.fs.fed.us/natfdb) indicate that reported tree failures are 35% root, 33% branch, and 32% trunk related.

In contrast to the pattern seen with the likelihood of failure ratings, participants inspecting trunks and roots rated the consequences of failure higher than when they were inspecting branches (Table 2). Defects or damage in the lower trunk or roots could lead to whole-tree failure, a scenario in which the entire above-ground mass of a tree could potentially come down on a target. This presents an interesting contradiction in risk assessments. The structures that are most likely to be rated high for potential damage are most likely to be rated low with regard to likelihood of failure.

Of the various demographic and personal factors assessed, only the average financial DOSPERT rating correlated with risk ratings (Table 2). While statistically significant, the factor's impact on predicted overall risk ratings is not overly dramatic. A seven-point increase in average financial DOSPERT rating (i.e., going from the lowest possible average of 1 to the highest possible average of 8) would only increase predicted overall risk ratings by 0.3962.

3.3. Variability of risk assessment rating inputs

While the impact on average ratings for tree risk assessment inputs and overall risk are important, so is the consistency of ratings (Norris, 2007). Variability in risk assessments among arborists can come from multiple sources. Some of this variability comes from factors related to the assessors – their training, experience, perceptions of risk, mental processing, tolerance of risk, and other internal influences (Ball and Watt, 2013). Additional variability may stem from ambiguities or misunderstandings related to the assessment method used (e.g., definitions, terminology, decision thresholds, scale interpretation, type of data used) (Ball and Watt, 2013).

In running our tests of equal variance, we found that there were often differences in variability for the three main inputs of tree risk assessment (Tables 3 and 4). Likelihood of impact was the most variable of the three inputs for risk assessment (Table 3). Interestingly, the

Table 3
Instances where the risk assessment inputs (i.e., likelihood of impact, likelihood of failure, and consequence of failure) were the most variable (only looking at cases where tests of equal variance were significant). This represents a total of 90 assessments (30 trees with branches, trunk, and roots assessed separately).

	Statistical test of equal variance		
	Significant Bartlett's test (n = 46)	Significant Levene's test (n = 32)	Significant Fligner-Killeen test outcomes (n = 30)
Likelihood of impact	28[a]	21	19
Likelihood of failure	2	2	2
Consequence of failure	16	9	9
Significance (*P*-value)	< 0.0001	< 0.0001	< 0.0001

[a] Interpretation: in this example there were 46 instances where the above test detected significant differences in variability among the ratings of likelihood of impact, likelihood of failure, and consequences of failure. Of these 46 instances, likelihood of impact was the most variable input 28 times, likelihood of failure was the most variable input 2 times, and consequences of failure was the most variable input 16 times.

Table 4
Number of trees (from a total $n = 90$) where at least one risk assessment input (i.e., likelihood of impact, likelihood of failure, and consequence of failure) was more variable than the others when assessing the risk associated root, trunk, or branch failure. Three tests of homogeneity of variance were used. A test of equal proportions was conducted to see if differences in variability among the risk assessment inputs were impacted by the part of the tree assessed.

	Statistical test of equal variance		
	Significant Bartlett's test outcomes ($n = 46$)	Significant Levene's test outcomes ($n = 32$)	Significant Fligner-Killeen test outcomes ($n = 30$)
Branches	11[a]	11	9
Trunk	18	13	12
Roots	17	8	9
Significance (*P*-value)	0.1822	0.3904	0.6538

[a] Interpretation: in this example there were 46 instances where the above test detected significant differences in variability among the branch, trunk, and root ratings. Of these 46 instances, branch ratings were the most variable 11 times, trunk ratings were the most variable 18 times, and root ratings were the most variable 17 times.

likelihood of failure was the least variable of the three inputs for risk assessment (Table 3). This pattern was consistent regardless of the portion of the tree inspected (i.e., branches, trunk, or roots) (Table 4).

In investigating this further, we tallied the number of instances where participant ratings of likelihood of impact, likelihood of failure, and consequences of failure were consistent (i.e. all at the same level) or split between/among two, three, or four levels for branch, trunk, and root assessments (Fig. 1). For likelihood of impact and consequences of failure, the most common scenario was a four-way split among the four levels of rating (Fig. 1). In contrast, a three-way split was the most common scenario for likelihood of failure (Fig. 1). We did not have any instances where our participants came to complete agreement when rating any of the inputs (Fig. 1).

Determining the likelihood of failure is often considered the most difficult aspect of tree risk assessment (Bellett-Travers, 2010). It can also be the most influential aspect of a tree risk assessment (Ellison, 2007; Koeser et al., 2015). As such, it is encouraging to see that this was the most consistent input in risk assessment (Table 3). Tree risk assessment efforts have traditionally been focused on identifying defects and assessing their perceived severity (Ellison, 2007; Koeser et al., 2015). As such, it seems logical that this would be the most consistent of the three inputs in the risk assessment process. Past emphasis on defects may also explain why, when modeling factors that influence input ratings, having the ISA Certified Arborist credential (established 1992) decades earlier than the Tree Risk Assessment Qualification was sufficient to impact likelihood of failure ratings, but not likelihood of impact or consequence of impact (Table 2).

3.4. Implications for practice

In contrast to likelihood of failure, likelihood of impact and consequence of impact have received less attention from researchers and the industry. While greater emphasis has been placed on these two inputs with recent training and credentialing initiatives (Smiley et al., 2011; Dunster et al., 2013), the characterization of likelihood of impact and consequences of failure remain barriers to making risk assessments more reproducible (Table 3).

Likelihood of impact and consequences of failure are more easily quantified than likelihood of failure (Ellison, 2005; Klein et al., 2016, submitted for publication). The use of traffic counting devices (or traffic count data) can make estimates of target occupancy (a critical component of likelihood of impact) more consistent and accurate

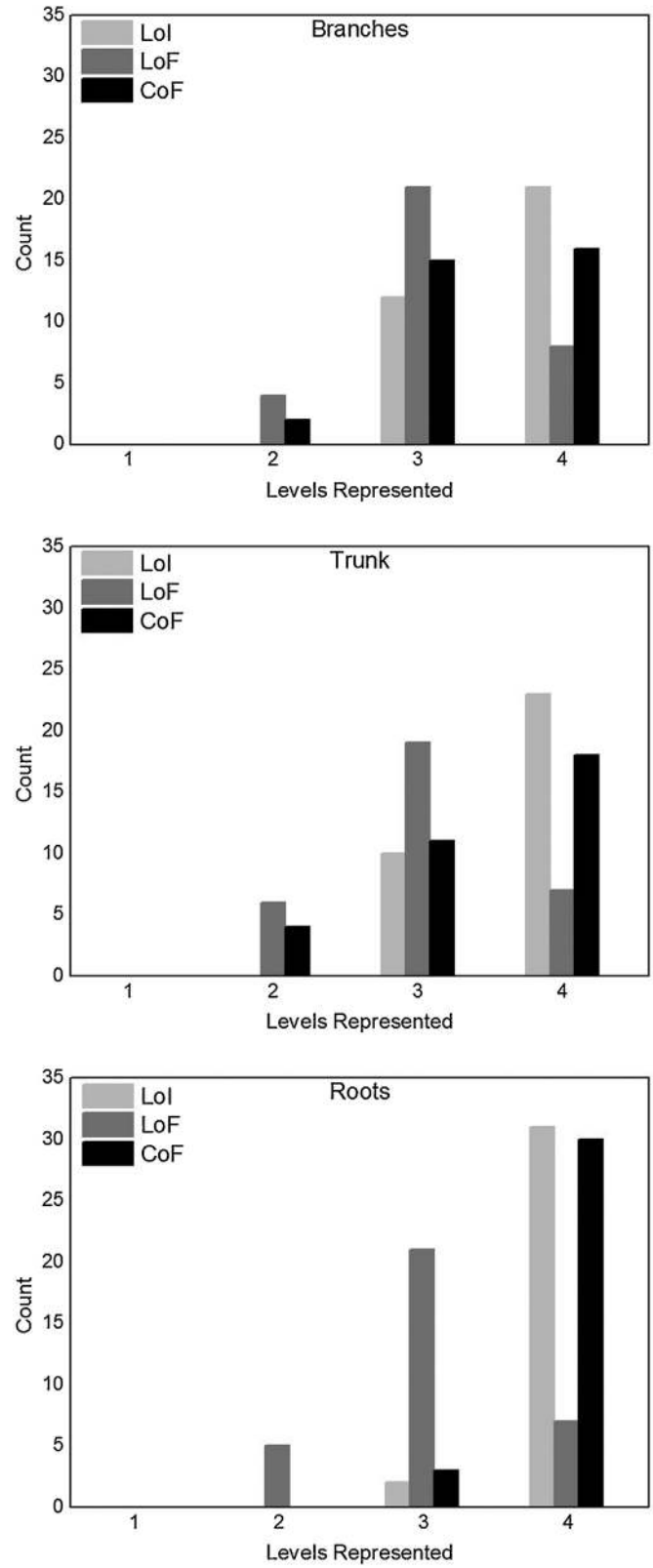

Fig. 1. Number of instances where participant ratings of likelihood of impact (LoI), likelihood of failure (LoF), and consequences of failure (CoF) were consistent or split between/among two, three, or four levels for branch, trunk, and root assessments.

(Klein et al., 2016). Similarly, consequences of failure are related to the weight of the assessed part of the tree. Estimates of weight can be obtained by measuring the size of the tree part assessed, estimating its

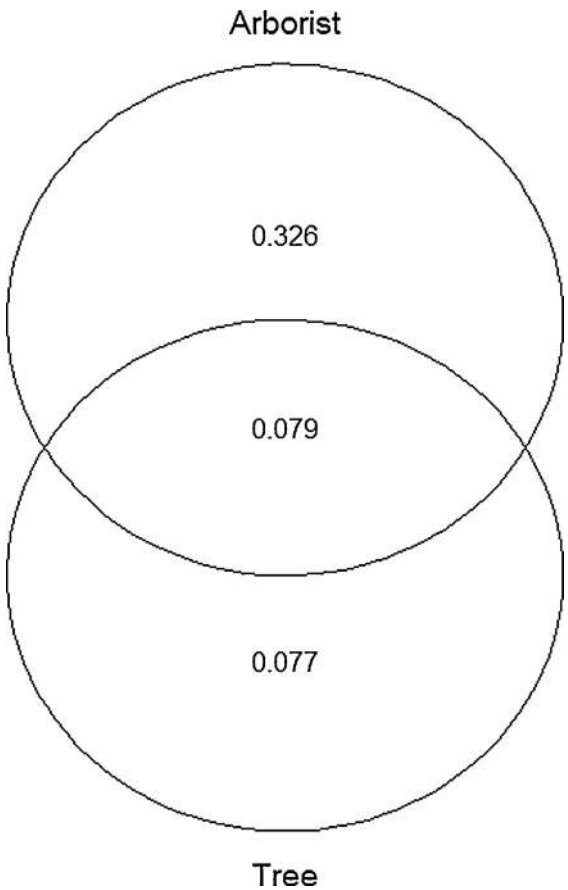

Fig. 2. Use of variance partitioning to determine the relative predictive strength of the individual arborist and tree assessed on risk ratings.

weight with a green log weight chart used for crane and rigging removal operations, and comparing the referenced value to the expected resiliency of the anticipated target.

Despite the relative ease with which these two factors can be quantified, the industry standards and best management practices currently used in the United States offer no concrete decision thresholds for determining likelihood of impact or consequences of failure (Smiley et al., 2011; Dunster et al., 2013). Additionally, the definitions used describe the levels for these two inputs are somewhat ambiguous. For example, as part of a list of scenarios that constitute "minor" consequences of failure, Dunster et al. (2013) provide the following example: *A medium-sized branch striking a deck from a moderate height*. In this scenario, the user is left to determine what "medium-sized" and "moderate height" mean – adding another potential layer of variability in risk assessments.

Without clearly-defined thresholds and terminology, risk assessment will continue to have unnecessary variability. In the risk assessment method tested here, variability arises largely from inconsistencies in the likelihood of impact and consequences of failure components. Variance partitioning for this study showed that the individual assessing the tree was over four times more important than the actual tree assessed when predicting overall risk ratings (Fig. 2). These findings support Norris's conclusion that the arborist assessing a tree is likely the most significant factor in determining risk rating (2007). If true, this sets the stage for differing assessment outcomes, legal disputes, and even public outcry (Stewart et al., 2013).

The precedent for more clearly defined industry thresholds has already been set with the training materials produced for the ISA Tree Risk Assessment Qualification. In this guide, Dunster et al. (2013) classify a range of tree defects as having improbable, possible, probable, or imminent likelihoods of failure, despite the difficulties associated with empirically supporting these assessments (Table 6). More clearly defined thresholds for likelihood of impact and consequences of failure might decrease the role assessor bias plays in overall risk determinations and would allow arborists to use technology like traffic loggers more effectively (Klein et al., 2016).

Unfortunately, consistency (while an important aspect of making risk assessment more reproducible) is not the same as accuracy. Variability in ratings means some portion of the assessments will be inaccurate. However, false precision in a risk assessment method could

Table 5
Factors influencing mitigation measures prescribed by arborists with varying industry credentials (n = 296). Non-significant predictors (i.e. gender, home ownership, average financial DOSPEERT rating, average health DOSPEERT rating, rating for likelihood of impact, and ratings for consequences of failure) are omitted from table.

Mitigation action	Factor	Coefficient	Standard error	*P*-value	Odds-ratio[a]	95% CI lower	95% CI upper
Advanced assessment	Certified Arborist – Yes	−1.2417	0.3425	0.0003	3.4615 (M)	1.7690	6.7740
	Tree Risk Assessment Qualified – Yes	−2.0133	0.3939	< 0.0001	7.4880 (M)	3.4594	16.2079
	Risk Assessment Experience – Yes	0.7884	0.3113	0.0113	2.2000 (AA)	1.1949	4.0495
	Parent – Yes	1.8280	0.3547	< 0.0001	6.2214 (AA)	3.1040	12.4684
	Likelihood of Failure Rating	0.7479	0.1807	< 0.0001	2.1125 (AA)	1.4823	3.0108
Cable	Tree Risk Assessment Qualified – Yes	−0.5730	0.2587	0.0268	1.7736 (M)	1.0680	2.9456
	Risk Assessment Experience – Yes	1.7603	0.2862	< 0.0001	5.8142 (C)	3.3181	10.1889
	Parent – Yes	1.2980	0.2967	< 0.0001	3.6620 (C)	2.0472	6.5508
	Likelihood of Failure Rating	0.3873	0.1569	0.0135	1.4730 (C)	1.0829	2.0037
Prune	Certified Arborist – Yes	−0.8934	0.2985	0.0028	2.4434 (M)	1.3612	4.3859
	Risk Assessment Experience – Yes	1.2837	0.2554	< 0.0001	3.6100 (P)	2.1878	5.9566
	Parent – Yes	1.1007	0.2794	< 0.0001	3.0063 (P)	1.7385	5.1986
	Likelihood of Failure Rating	0.4902	0.1486	0.0010	1.6326 (P)	1.2199	2.1851
Remove tree	Certified Arborist – Yes	−1.4553	0.3238	< 0.0001	4.2857 (M)	2.2721	8.0849
	Tree Risk Assessment Qualified – Yes	−0.8610	0.2874	0.0027	2.3655 (M)	1.3467	4.1554
	Risk Assessment Experience – Yes	1.9753	0.3187	< 0.0001	7.2087 (R)	3.8597	13.4636
	Parent – Yes	1.2115	0.3243	0.0002	3.3585 (R)	1.7784	6.3426
	Likelihood of Failure Rating	1.1840	0.1655	< 0.0001	3.2674 (R)	2.3622	4.4199

[a] The tendency of arborists to select a more active mitigation measure (i.e., advanced assessment, cabling, pruning, or tree removal) over the continued monitoring (i.e., the base level for each of the mitigation treatments modeled above). Negative coefficients indicate a factor was associated with an increased tendency to prescribe continued monitoring (M = continued monitoring). Positive coefficients indicate a factor was associated with an increased tendency to prescribe a more active mitigation (AA = Advanced Assessment, C = Cabling, P = Pruning, and R = Removal). As an example, International Society of Arboriculture Certified Arborists were ~2.3 times more likely to prescribe monitoring than removal. However, all participants were ~3.3 times more likely to suggest removal for each level that likelihood of failure ratings increased.

Table 6
Likelihood of failure ratings levels and definitions (Dunster et al., 2013).

Rating Level	Definition
Improbable	Failure is not likely during normal weather conditions and may not occur in many severe weather conditions with the specified time frame[a]
Possible	Failure could occur, but it is unlikely during normal weather conditions within the specified time frame
Probable	Failure may be expected under normal weather conditions within the specified time frame
Imminent	Failure has started or is likely to occur in the near future, even if there is no significant wind or increased load

[a] Definition paraphrased to match other three definitions.

create a very consistent bias that pulls the perceived level of risk away from the actual level of risk for all who subscribe to the method. Future research must validate assessments of likelihood of failure and consequences of failure (Koeser, 2009). In waiting for this external validation, arborists must continue to assess risk using BMPs based on the limited empirical data currently available and industry consensus. Care should be taken when using the latter as a justification decisions regarding risk management. Industry consensus derived from the shared experiences of many arborists will be less susceptible to bias than industry consensus stemming from a small number of persuasive experts or trainers.

3.5. Impact of arborists experience and credentials on ratings

While past and unpublished risk assessment research has often focused on factors that influence risk ratings (Norris, 2007; Klein et al., 2016; Koeser et al., 2015, submitted for publication), no studies have looked at how this ultimately impacts the tree. Our results indicate that ISA Certified Arborists were four times more likely than non-Certified Arborists to recommend retaining and monitoring the tree rather than removing it (Table 5). The impact of the ISA Tree Risk Assessment Qualification was less pronounced but still significant. TRAQ arborists were more than twice as likely to prescribe retaining and monitoring trees rather than removing them compared to non-TRAQ arborists (which included both Certified and non-Certified arborists).

When comparing more active mitigation measures (i.e., removal, cabling, pruning, and advanced assessment) to retaining and monitoring, there is strong evidence that additional training results in lower levels of perceived risk (Table 5). This corresponds with past research where experience and training tempered ratings of likelihood of failure and target occupancy (Klein et al., 2016, submitted for publication). Interestingly, arborists with children were 3–6 times more likely to opt for more active mitigation measures than those without children.

Of the three risk assessment inputs, only likelihood of failure was a significant predictor of mitigation measures (Table 5). While the significance of likelihood of failure is not surprising (mitigation measures like pruning, cabling, and advanced assessment are often tied to specific defects), the absence of any relationship between likelihood of impact and consequences of failure is notable. As research and training refine the industry's awareness of risk factors beyond defects and tree damage, likelihood of impact and consequences of failure may have greater roles in what mitigation actions are prescribed as part of a tree risk assessment.

4. Conclusion

As noted by others, the role the arborist assessing a tree plays in determining its assessed risk rating is significant. In particular, personal biases and individual perceptions of risk likely account for the variability associated with ratings of likelihood of impact and consequences of failure. In contrast, the tree care industry's historic focus on the identification of tree defects (while largely ignoring the presence or resiliency of nearby targets) may explain why it is the most consistent of the three key tree risk assessment inputs. Reducing excess variation in likelihood of impact and consequences of failure ratings will ultimately reduce cases where users risk assessment method tested arrive a differing overall risk ratings.

Consistency is not the same as accuracy and efforts to reduce variability in risk assessments must reflect our current understanding of trees, risk, and tree biomechanics. While the accuracy of likelihood of failure ratings may remain unknown for the foreseeable future, it may be possible to make other aspects of risk assessment more reproducible given training and the creation of clear industry definitions. More research is needed to provide the empirical basis for concrete and meaningful thresholds for guiding tree risk assessments.

Acknowledgements

We would like to thank the arborists of the F.A. Bartlett Tree Expert Company and the attendees of the California Tree Failure Reporting program for sharing their tree ratings and participating in this study. We also thank Sean Henry, a research technician at the Bartlett Tree Research Laboratories, for data input; Drew McLean, a Biological Scientist at the University of Florida Gulf Coast Research and Education Center, for data cleaning; and Dr. Bruce R. Fraedrich of the Bartlett Tree Research Laboratories for his contributions to experimental design and study application.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ufug.2017.03.027.

References

ANSI A300 (Part 9) – 2011 Tree Risk Assessment (American National Standard for Tree Care Operations – Tree, Shrub and Other Wood Plant Management Standard Practices (Tree Risk Assessment A. Tree Structure Assessment)) (2011) Part 9. Tree Care Industry Association.
Ball, D.J., Watt, J., 2013. Further thoughts on the utility of risk matrices. Risk Anal. 33 (11), 2068–2078.
Barbosa, A.M., Brown, J.A., Jimenez-Valverde, A., Real, R., 2015. modEvA: Model Evaluation and Analysis. R. http://R-Forge.R-project.org/projects/modeva/.
Bellett-Travers, M., 2010. A risk assessment methodology for trees in parkland based on comparative population analysis. Arboric. J. 33, 3–14.
Blais, A.R., Weber, E.U., 2006. A domain-specific risk-taking (DOSPERT) scale for adult populations. Judgm. Decis. Mak. 1, 33–47.
Dunster, J.A., Smiley, E.T., Matheny, N., Lilly, S., 2013. Tree Risk Assessment Manual. International Society of Arboriculture, Champaign.
Ellison, M.J., 2005. Quantified tree risk assessment used in the management of amenity trees. J. Arboric. 31, 57–65.
Ellison, M.J., 2007. Moving the focus from tree defects to rational risk management—a paradigm shift for tree managers. Arboric. J. 30, 137–142.
Hickman, G.W., Perry, E.J., Evans, R., 1995. Validation of a tree failure evaluation system. J. Arboric. 21, 23–34.
Klein, R.W., Koeser, A.K., Hansen, G., Escobedo, F.J., 2017. A review of urban tree risk assessment and risk perception literature. Arboric. Urban For (submitted for publication).
Klein, R.W., Koeser, A.K., Hauer, R.J., Hansen, G., Escobedo, F.J., 2016. Relationship between perceived and actual occupancy rates in urban settings. Urban For. Urban Green. 19, 194–201.
Koeser, A.K., 2009. Trees & Risk Researcher Summit White Paper. http://www.isa-arbor.com/education/resources/educ_portal_risk_whitepaper.pdf (accessed 02.02.15).
Koeser, A.K., Klein, R.W., Hasing, G., Northrop, R.J., 2015. Factors driving professional and public urban tree risk perception. Urban For. Urban Green. 14, 968–974.
Koeser, A.K., Hauer, R.J., Miesbauer, J.W., Peterson, W., 2016a. Municipal tree risk assessment: findings from a comprehensive survey of urban forest management. Arboric. J. 38 (4), 218–229. http://dx.doi.org/10.1080/03071375.2016.1221178.
Koeser, A.K., Mclean, D.C., Hasing, G., Allison, R.B., 2016b. Frequency, severity, and detectability of internal trunk decay of street tree *Quercus* spp. in Tampa, Florida, U.S. Arboric. Urban For. 42, 217–226.
Koeser, A.K., Hauer, R.J., Klein, R.W., Miesbauer, J.W., 2017. Assessment of likelihood of failure using limited visual, basic, and advanced assessment techniques. Urban For. Urban Green 24, 71–79.
Leibowitz, R., 2012. Urban tree growth and longevity: an international meeting and research symposium white paper. Arboric. Urban For. 38, 237–241.
Luley, C., Nowak, D., Greenfield, E., 2009. Frequency and severity of trunk decay in street tree maples in four New York cities. Arboric. Urban For. 35, 94–99.

Norris, M., 2007. Tree Risk Assessments – What Works – What Does Not – Can We Tell? A review of a range of existing tree risk assessment methods. Presented at the ISAAC Conference Perth 2007 1–31.
Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., R Development Core Team, 2013. nlme: Linear and Nonlinear Mixed Effects Models. pp. 1–111.
R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.
Rooney, C.J., Ryan, H.D.P., Bloniarz, D.V., Kane, B.C.P., 2005. The reliability of a windshield survey to locate hazards in roadside trees. J. Arboric. 31, 89–94.
Roy, S., Byrne, J., Pickering, C., 2012. A systematic quantitative review of urban tree benefits, costs, and assessment methods across cities in different climatic zones. Urban For. Urban Green. 11, 351–363.
Smiley, E.T., Matheny, N., Lilly, S., 2011. Best Management Practices. Tree Risk Assessment. International Society of Arboriculture, Champaign.
Stewart, M.G., O'Callaghan, D., Hartley, M., 2013. Review of QTRA and risk-based cost-benefit assessment of tree management. Arboric. Urban For. 39, 165–172.
Tate, R.L., 1984. Stem decay in Central Park. J. Arboric. 10, 307–308.
Venables, W.N., Ripley, D.B., 2002. Modern Applied Statistics with S. Springer, New York.