High risk occupations: what is the question to ask and challenges with data analysis.

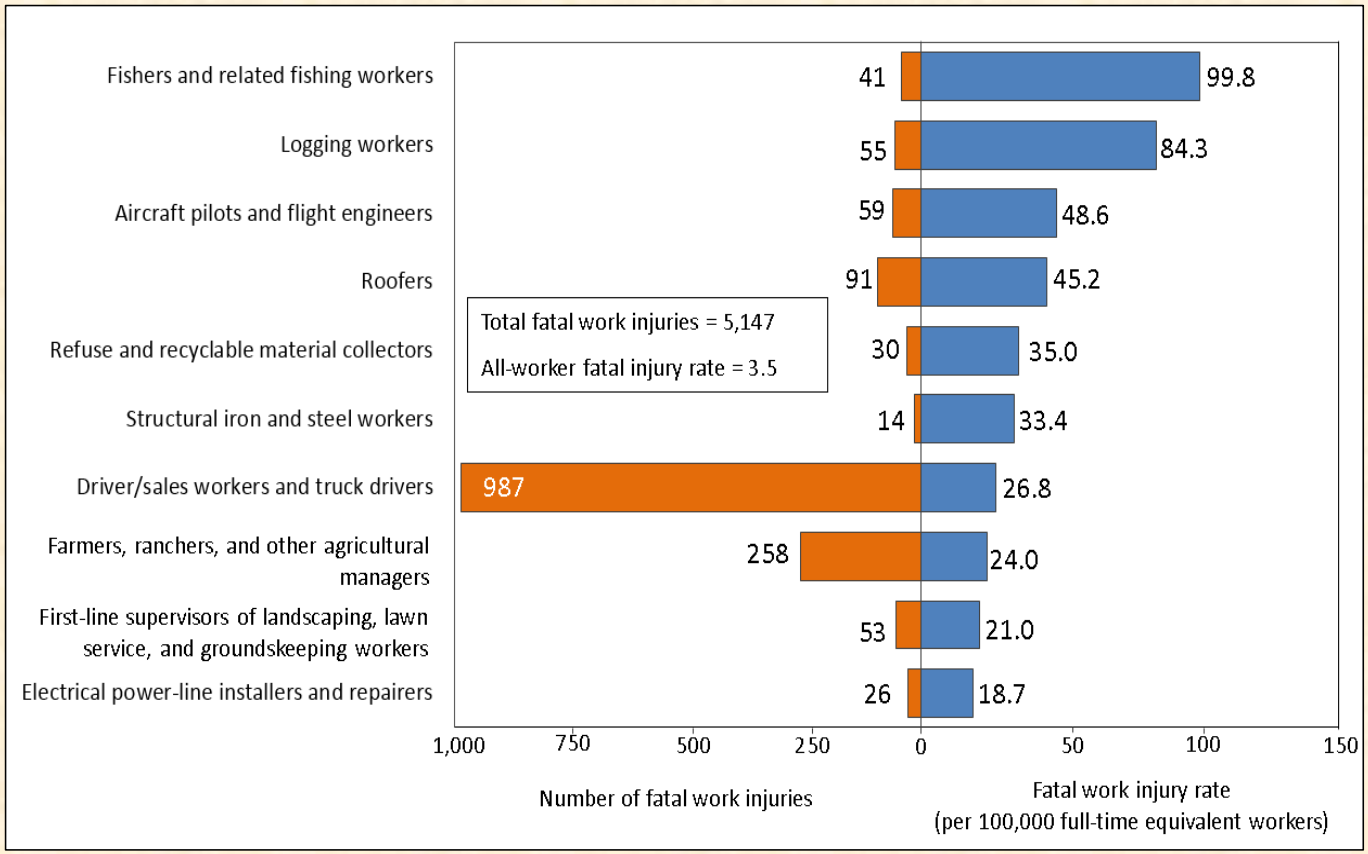
Kevin Lyons
Wes Lematta Professor in Forest Engineering
Office: Snell 311
Phone: 541-737-5630
Email: kevin.lyons@oregonstate.edu

Injury rates in logging

Figure 1.

Fatal work injury rate for forest logging workers in the United States in 2017

(Bureau of Labor Statistics, US Department of Labor, Chart 3)



Challenges to managing worker safety in logging

- Natural environment
- Continually changing locations
- Overlapping constraints
- Workers having to make important decisions that affect their safety

Manual tree falling

- In British Columbia about 3000 registered fallers, about 1500 person years of work.
- Range in fatalities per year 1 to 6 (1:1500 to 1:250 fatalities per person year)

https://www.youtube.com/watch?v=V-SwpDKkHko&t=70s

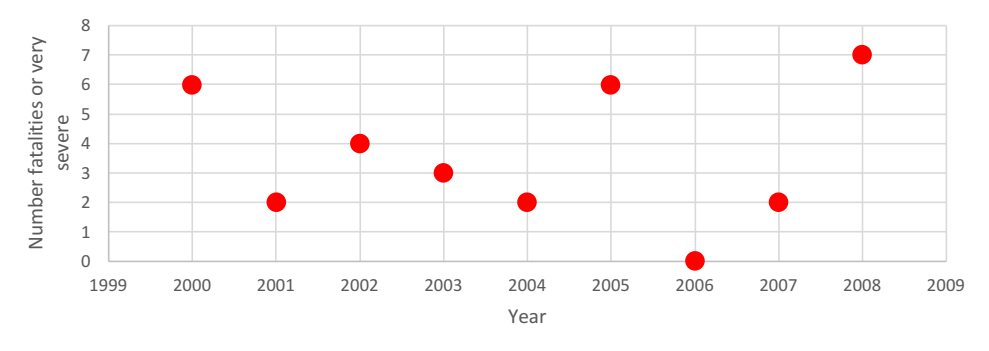
Are fatalities the metric to use in managing faller safety?

Faller serious injuries and fatalities reviewed (WorkSafeBC, 2009_B)

* 1incident was a serious injury

** both incidents were serious injuries

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008
Number incidents	6	2	4	3	2*	6	0	2**	7



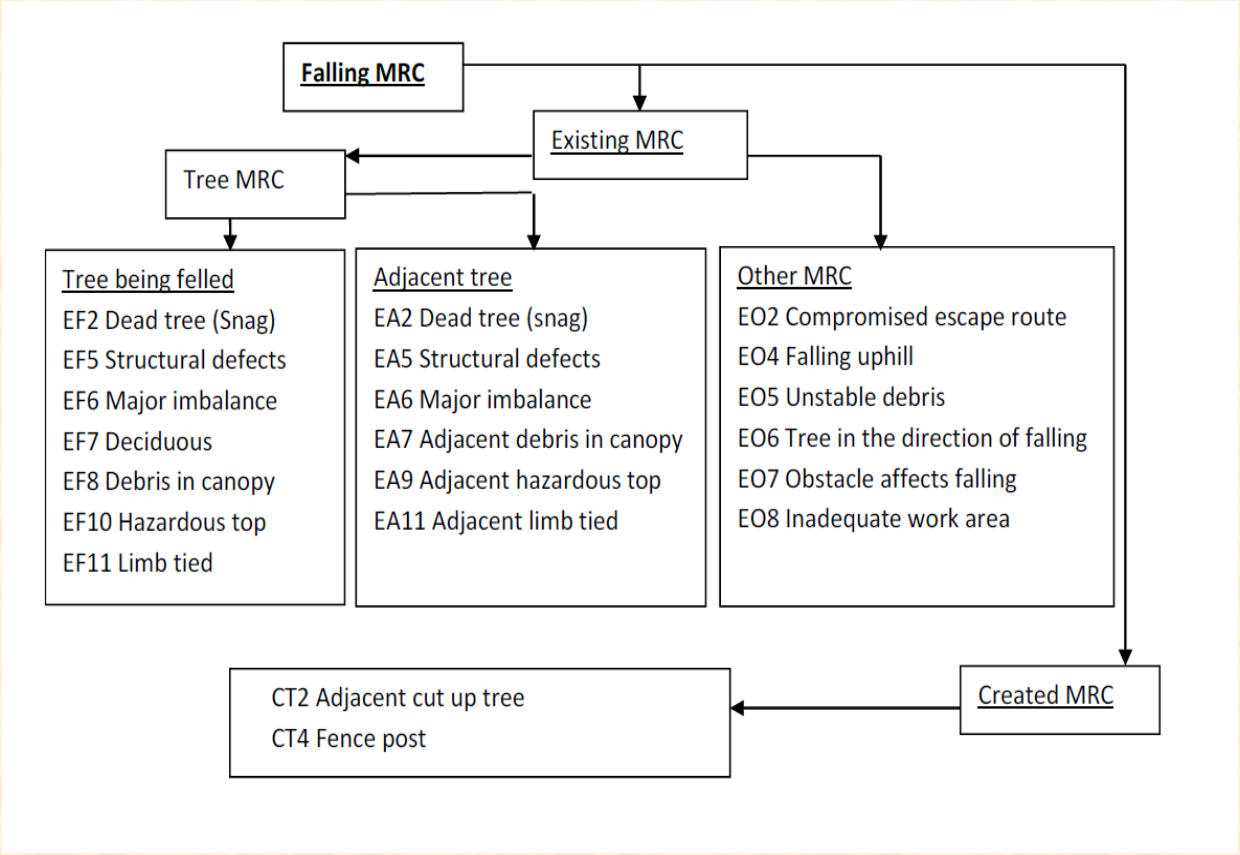
Problem with informal view of data

- In 2002 certification of commercial tree fallers was initiated in BC
- In 2004 certification became mandatory (i.e. if you were falling trees in a commercial forestry operation you had to be certified)
- Regulators viewed the drop from 2002 to 2004 as a success vindicating certification.
- When the 2005 results came out the regulators explained these away as complacency after a good year, using the 2006 results to support this.
- By 2008 the regulators finally began to listen to those arguing that certification was not having an effect on fatality results

Alternatives to incident data

- Use the concept of Antecedent and Consequence from behavior based safety management
- In falling there are general antecedents that are present for all trees (i.e. job is to fall trees) and these are not so helpful when trying to predict the occurrence of unsafe consequences.
- We developed the concept of management requiring conditions and unexpected events.

Management Requiring Conditions



Management Requiring Condition (MRC): Is a condition that requires either an action or decision by the faller before a tree can be felled.

Severity Code:

1. not an immediate threat
2. an immediate threat but the faller has existing cover or an escape route
3. an immediate threat requiring an alternate falling method

Unexpected Events

Unexpected Event (*UE*): an event that has the potential to severely injure the faller and either the faller was unaware of the possible occurrence or a planned event did not go as planned.

Severity Code:

1. within normal variation from the intended plan
2. significant variation from the intended plan but safety measures ensured the faller's safety and
3. significant variation from the intended plan and it was only chance that it did not cause a serious injury.

UET1: object falls out of the canopy
UET2: falling direction change due to the tree hitting another object
UET3: falling direction change due to wind
UET4: falling direction change due to other reasons
UET6: barber chair
UET7: tree hangs up
UET8: tree cannot be wedged over
UET9: tree in group falls early
UET14: unexpected rot resulting in the loss of control of the tree being felled
UET15: tree being felled knocks over another tree
UEB5: saw pinched
UEO2: root dislodged
UEO4: fall or trip

Advantages of MRC and UE data

- Provides information on trees where no incident occurred
- Frequency is much higher than reportable incidents
- Get detailed information about what the faller was actually seeing
- Each tree is an observation

Problems with data analysis

- Observational data not experimental
- Confounding effects
- Non-independent data

Models to use for analysis: independent data

- MLR (multiple linear regression): use for continuous response variable and independent data
- Logistic Regression: use for binary response variable and independent data

Full Model

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Regression	16	356.254	356.254	22.266	19.987	0.000000
age	1	7.361	0.136	0.136	0.122	0.726537
sex	1	0.117	1.000	1.000	0.898	0.343778
exmed	1	1.802	1.527	1.527	1.371	0.242211
CombJob	6	50.442	41.217	6.869	6.166	0.000003
children	1	3.049	9.149	9.149	8.213	0.004326
caffeinated	3	31.356	31.350	10.450	9.381	0.000005
sleptat	2	39.407	15.644	7.822	7.022	0.000978
off	1	222.719	222.719	222.719	199.928	0.000000
Error	528	588.190	588.190	1.114		
Lack-of-Fit	128	363.775	363.775	2.842	5.066	0.000000
Pure Error	400	224.415	224.415	0.561		
Total	544	944.444				

Reduced Model

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Regression	13	353.364	353.364	27.182	24.419	0.0000000
CombJob	6	52.187	45.310	7.552	6.784	0.0000006
children	1	3.120	10.381	10.381	9.326	0.0023726
caffeinated	3	31.443	31.878	10.626	9.546	0.0000038
sleptat	2	44.085	17.202	8.601	7.727	0.0004924
off	1	222.529	222.529	222.529	199.910	0.0000000
Error	531	591.080	591.080	1.113		
Lack-of-Fit	102	261.989	261.989	2.569	3.348	0.0000000
Pure Error	429	329.091	329.091	0.767		
Total	544	944.444				

Models to use for analysis: non-independent data

- LME (Linear Mixed Effects): use for continuous response variable and non-independent data
- GLMM (Generalized Linear Mixed Models): use for data with different link functions (e.g. binary response variables) and non-independent data

$$\mathbf{y}_i = \mathbf{X}_i\mathbf{b} + \mathbf{Z}_i\mathbf{u}_i + \boldsymbol{\varepsilon}_i$$

$$\mathbf{V}_i = \left(\sigma_i^2\mathbf{I}_i^{\varepsilon} + \sigma_b^2\mathbf{Z}_i\mathbf{I}^u\mathbf{Z}_i^T\right) = \left(\sigma_i^2\mathbf{I}_i^{\varepsilon} + \sigma_b^2\mathbf{Z}_i\mathbf{Z}_i^T\right)$$

In V_i the covariance is accounted for by the random effects model matrix and the inter-cluster variance.

Correlation between observations within the same cluster is greater when the inter-cluster variance is higher.

Example of LME models *Total MRC*

Model	Response	Fixed effecs									Random
1	TotalMRC	DSH	SR	Sl	SP	TR	W	R	WS	U	FallerID[a]
2	TotalMRC	DSH	SR								FallerID[a]
3	TotalMRC	DSH									FallerID[a]
4	TotalMRC		SR								FallerID[a]
5	TotalMRC	DSH	SR								FallerID[b]

[illegible] indicates the random effect for the i^{th} father follows i.i.d. $N(0, \sigma_a^2 I)$, where I is an identity matrix of dimension $q \times q$.
[illegible] indicates the random effect for the i^{th} father follows i.i.d. $N(0, \sigma_a^2 I)$, where I is an identity matrix of dimension $q \times q$

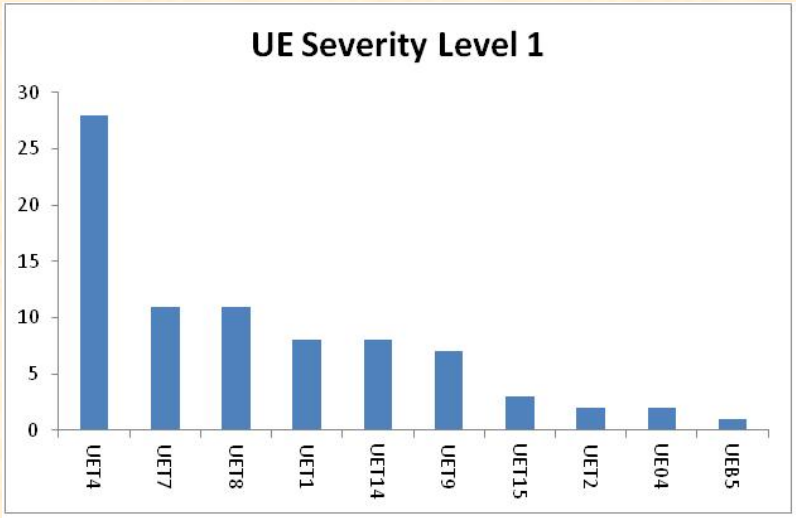
Parsimonious model

$$\log(TotalMRC) = \omega + b_1 DSH + b_2 SR + a_i$$

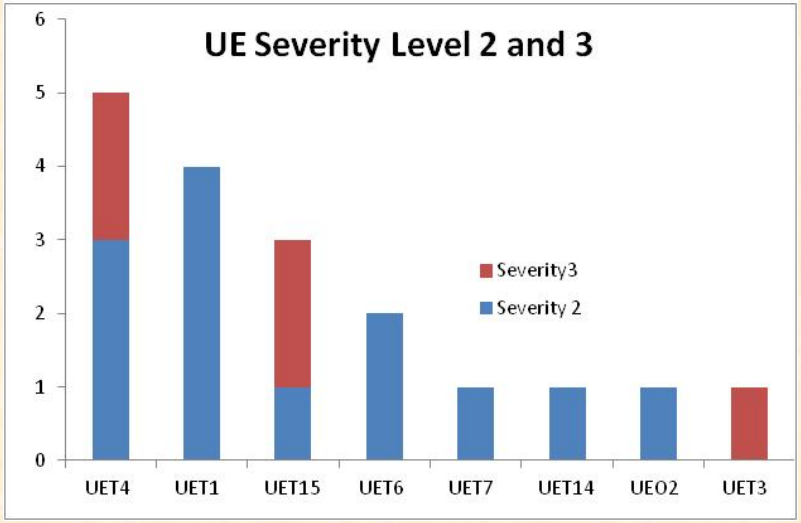
Example of GLMM models, response *UE = (0,1)*

Variable	Class Value	Estimate	Std. Error	Wald ChiSq	Prob. ChiSq
Intercept		-1.780	0.490	13.205	0.000
DSH		0.012	0.004	9.015	0.003
Slope		-0.011	0.005	3.949	0.047
Terrain	R	0.678	0.635	1.139	0.286
Terrain	G	0.224	0.324	0.480	0.488
Terrain	E	-0.693	0.251	7.602	0.006
CT2	1	0.351	0.180	3.788	0.052

Effect	Odds Ratio	C.I (Lower) α =0.1	C.I (Upper) α =0.1
DSH	1.012	1.005	1.018
CT2 1 vs 0	2.019	1.115	3.655
Slope	0.990	0.981	0.998
Terrain R vs G	1.573	0.361	6.864
Terrain R vs E	3.938	0.992	15.628
Terrain R vs B	2.426	0.571	10.302
Terrain G vs E	2.503	1.431	4.380
Terrain G vs B	1.542	0.788	3.017
Terrain E vs B	0.616	0.366	1.035



UET4 Falling direction change unknown reason
UET7 Tree hangs-up
UET8 Tree can't be wedged over
UET1 Object falls out of canopy
UET14 Loss of control, unseen rot
UET9 Tree in group falls early
UET15 Falling tree knocks over another tree
UET2 Falling direction change hit another object
UEO4 Trip or fall
UEB5 Saw pinched while bucking



UET4 Falling direction change unknown reason
UET1 Object falls out of canopy
UET15 Falling tree knocks over another tree
UET6 Barber chair
UET7 Tree hangs-up
UET14 Loss of control, unseen rot
UEO2 Roots dislodged
UET3 Falling direction change due to wind

What to do?

- Ask a question that you can actually study.
- Look for Antecedents, Behaviors, and Consequences that are observable and measureable.
- Be careful with your statistical models: confounding effects and non-independent data
- Correlation is often more useful than prediction