An overview of Paired Comparison Methods

Professor Johan Rene van Dorp, D.Sc.
http://www.seas.gwu.edu/~dorpjr/

WHY USE OF EXPERT JUDGEMENT?

- Risk Management Analysis and Risk Assessment inherently deal with low probability – high consequence events. Hence, not much data is available to assess accident probabilities in traditional statistical ways.

- Data sources may not have been constructed with a Risk Management Analysis in mind leading to an incomplete data descriptions.

- Data sources can be fraught with problems (missing data, or the same event appears multiple times in separate accident data bases making integration of data sources a costly and difficult effort).

ONLY USE EXPERT JUDGMENT WHEN DATA IS MISSING OR INCOMPLETE

EXPERT JUDGMENT ELICITATION PROCEDURE

STRUCTURED APPROACH TO CAPTURING A SUBJECT MATTER EXPERTS KNOWLEDGE BASE AND CONVERT HIS/HER KNOWLEDGE BASE INTO QUANTITATIVE ASSESSMENTS.

MODELERS SKILLED IN **DECOMPOSITION, DIAGNOSTICS AND AGGREGATION** OF ASSESSMENTS

NORMATIVE

EXPERTS

SUBSTANTIVE

KNOWLEDGABLE ABOUT **THE SUBJECT MATTER** AND HAVE EXTENSIVE EXPERIENCE

ELICITATION PROCESS = **MULTIPLE CYCLES** (AT LEAST 2)

1. DECOMPOSITION OF EVENT OF INTEREST TO A MEANINGFULL LEVEL FOR **SUBSTANTIVE EXPERT**.

2. ELICITATION OF JUDGMENT OF SUBSTANTIVE EXPERT FACILITATED BY **NORMATIVE EXPERT**.

3. **DIAGNOSTICS** + **AGGREGATION** OF SUBJECT MATTER JUDGMENTS BY NORMATIVE EXPERT.

EXPERT JUDGMENT ELICITATION PRINCIPLES

Source: "*Experts in Uncertainty*" by Roger M. Cooke (1991)

Reproducibility: It must be possible for scientific peers to review and if necessary **reproduce** all calculations. This entails that calculation models must be fully specified and the expert judgment ingredient data must be made available.

Accountability: The source of expert judgment must be identified (not necessarily by name, but certainly by profession and level of expertise).

Empirical Control: Expert probability assessments must in principle be susceptible to empirical control.

Neutrality: The method for combining/evaluating expert judgments should encourage experts to state true opinions.

Fairness: All experts are treated equally, prior to processing the results of observation.

PRACTICAL EXPERT JUDGMENT ELICITATION GUIDELINES

Source: "*Experts in Uncertainty*" by Roger M. Cooke (1991)

1. The questions must be clear.
2. Prepare an attractive format for the questions and a graphic format for the answers.
3. Perform a dry run.
4. An analyst must be present during the elicitation.
5. Prepare a brief explanation of the elicitation format, and of the model for processing the responses.
6. Avoid coaching.
7. The elicitation session should not exceed 1 – 2 hours.

EXPERT JUDGMENT ELICITATION PROCEDURES

Direct Procedures: Ask for Probabilities\Measures of Central Tendency\Measures of Variability Directly.

Problems with Direct Procedures:

• People are not born with a natural feel for low probabilities in the order of e.g. 1e-6. Only when a high number of instances of a particular event are observed by an expert one could consider these procedures. (Typically, not the case in Risk Analysis).

• People may not understand what a probability is and hence asking for them becomes problematic.

Indirect Procedures: A variety of Paired Comparisons Techniques are available e.g.: Bradley Terry Paired Comparison, The Analytical Hierarchy Process (AHP), Bayesian Paired Comparison for relative accident probability assessment. (More suitable for the domain of Risk Management Analysis).

Example AHP Type Paired Comparison Question:

Please compare the two locations in terms of the percentage of time that vessels operate in restricted visibility (i.e. vessels are required to use their fog signal) in the specified quarter.

THIRD QUARTER: July - August - September

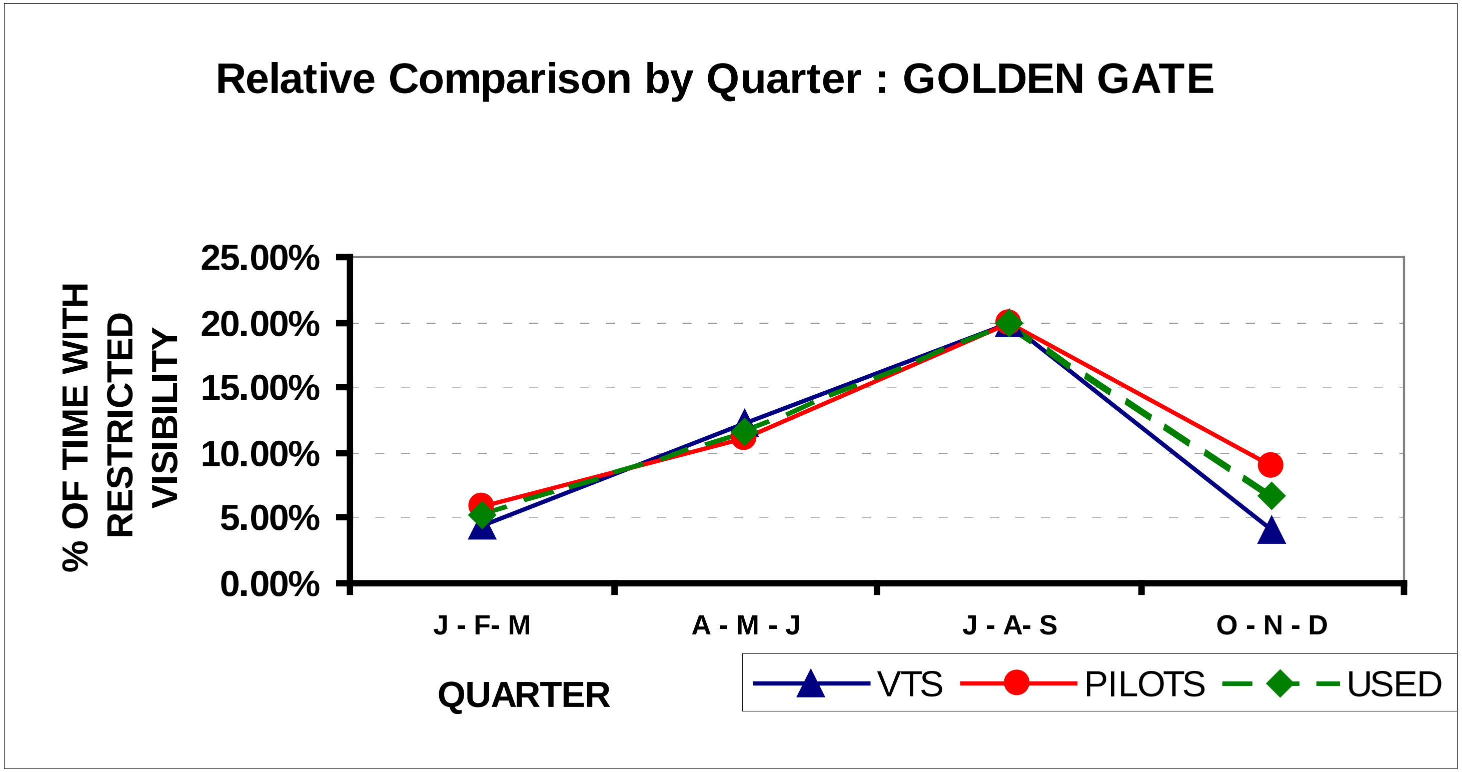
Location **Golden Gate** **Left Hand Side More**	Location **San Pablo Bay** **Right Hand Side More**

9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9

1 Same amount of time
3 Three times more
5 Five times more
7 Seven times more
9 Nine times or more

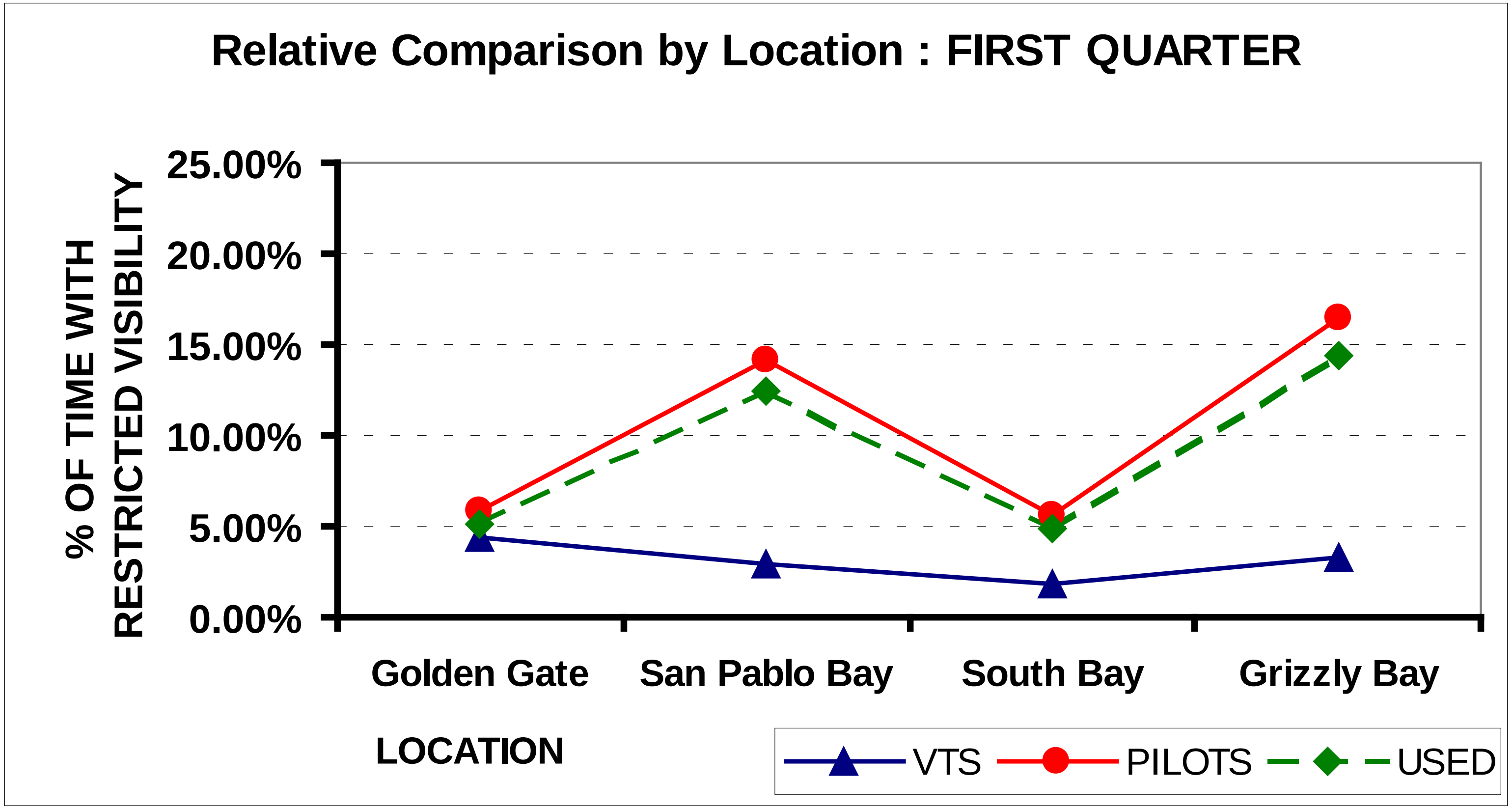
Used to determine relative frequency of Fog-Conditions in San Francisco Bay over different locations

Example AHP - Results:



Notice that a remarkable agreement is observed

Example AHP - Results:



Notice that a remarkable disagreement is observed.

Example of AHP - Diagnostics:

If an expert says it is **5 times more likely** to have bad visibility in Golden Gate than in San Pablo Bay and it is **2 times more likely** to have bad visibility in San Pablo Bay than South Bay, he/she should respond it is **10 times more likely** to have bad visibility in in Golden Gate than South-Bay, to be consistent.

Given **the number of items** one is comparing, one can determine the distribution of **a consistency index** (equal to zero if the expert is perfectly consistent) of an expert responding at random. **Larger values** of the consistency index indicate **a lesser consistency**.

For every subject matter expert we can calculate his/her consistency index and evaluate the probability that the consistency index of a random expert is worse. **You would want this probability to be high to accept the responses of the subject matter expert!**

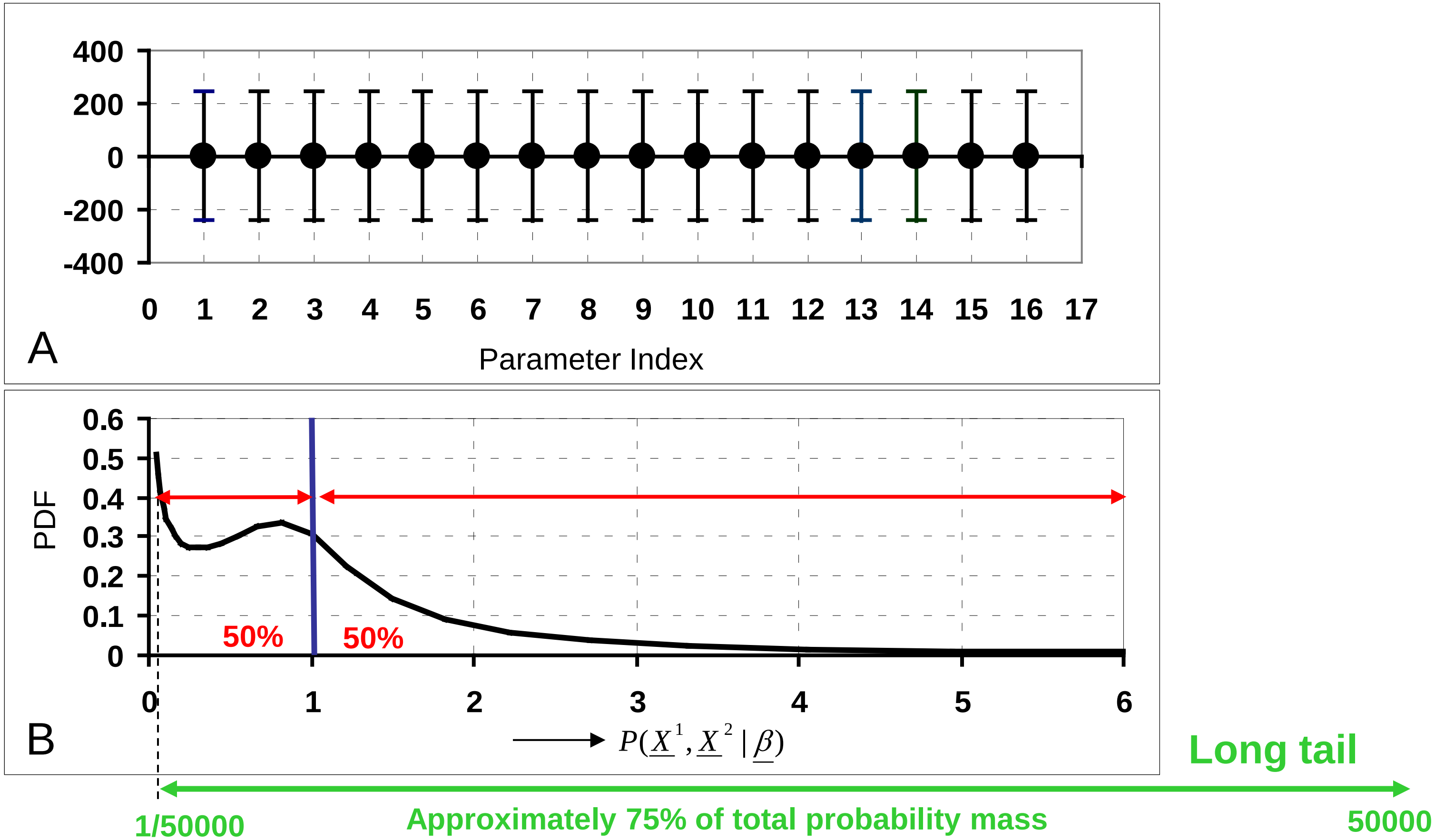
Example Bayesian Relative Accident Probability Paired Comparison Question:

Question: 32 48

Situation 1	**Attribute**	**Situation 2**
Super	**Ferry Class**	-
SEA-BAI	**Ferry Route**	-
Naval Vessel	**1st Interacting Vessel**	-
Crossing the bow	**Traffic Scenario 1st Vessel**	-
1 to 5 miles	**Traffic Proximity 1st Vessel**	-
Deep Draft	**2nd Interacting Vessel**	-
Crossing the bow	**Traffic Scenario 2nd Vessel**	-
1 to 5 miles	**Traffic Proximity 2nd Vessel**	-
more than 0.5 mile	**Visibility**	less than 0.5 mile
Along Ferry	**Wind Direction**	-
40 knots	**Wind Speed**	-
	9 8 7 6 5 4 3 2 1 2 3 4 5 6 7 8 9	
Situation 1 is worse	**<===================X====================>**	**Situation 2 is worse**

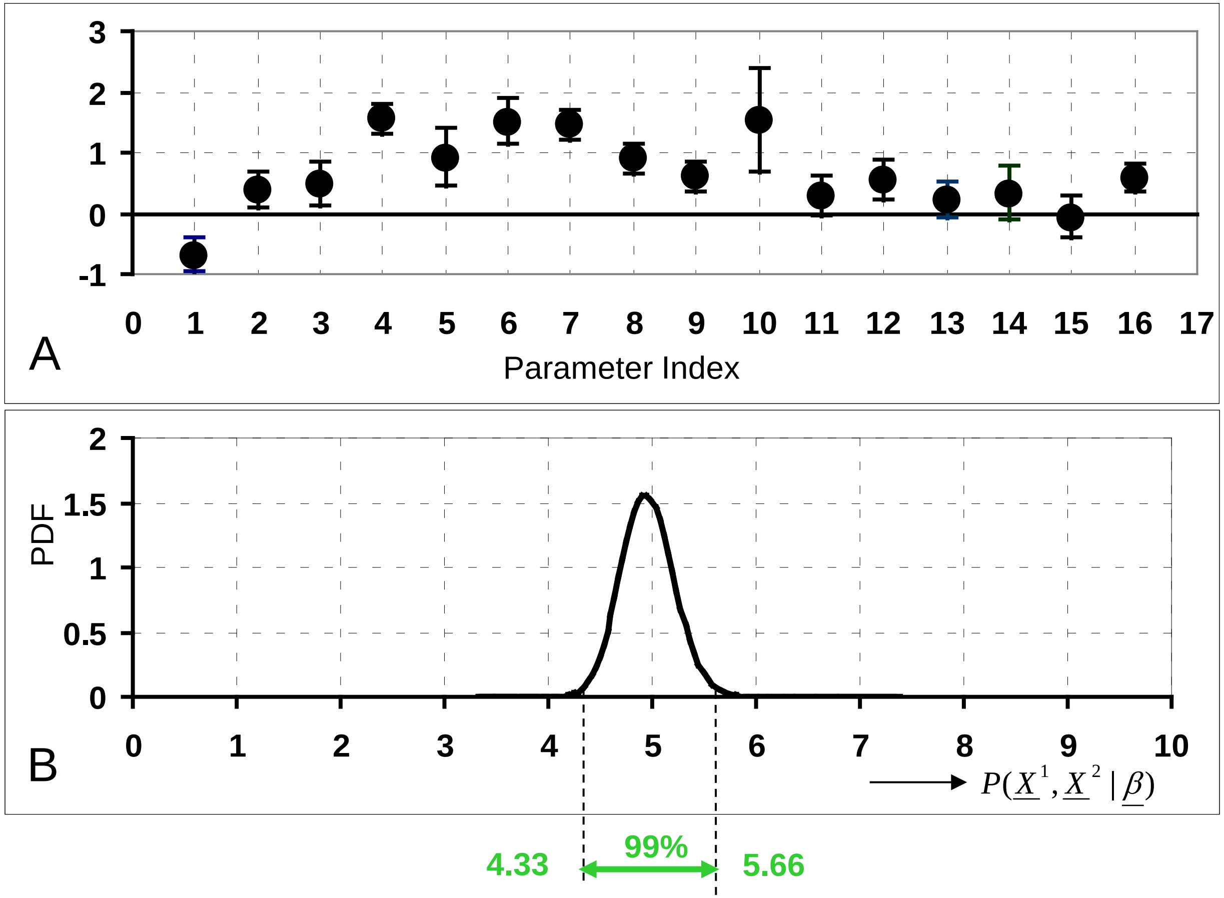
Used to determine the relative probability of a Ferry-Collision as a function of a number of situational attributes

Example Bayesian Relative Accident Probability Results:



Prior specification based on no preference and expert responding at random to the questionnaire

Example Bayesian Relative Accident Probability Results:



Posterior results obtain after updating with expert judgment.
You would like convergence of the distribution as demonstrated above.

A More Detailed Look at The Bradley Terry Paired Comparison Method

Professor Johan Rene van Dorp, D.Sc.
http://www.seas.gwu.edu/~dorpjr/

Bradley-Terry Paired Comparison Model

- It is assumed that we have n objects each having associated with it a scale value $V_i, i = 1, \cdots, n$.

- When an expert is asked if he/she prefers object i over object j it is assumed that:

$$\Pr(Expert's\ response\ is\ i > j) = \frac{V_i}{V_i + V_j}$$

- Since the values V_i are only determined up to a multiplicative constant, one may assume:

$$\sum_{i=1}^{n} V_i = 1, V_i > 0$$

Example Bradley-Terry Paired Comparison Question

An example question is as follows. Let the two vessel types you are comparing be: Container Vessel and Bulk Carrier. Next, you are asked assuming **the same traffic scenario for both vessel types** to indicate the vessel type for which you would be more concerned for a collision to occur and you are asked to indicate your answer in the following format.

Container	<--	=	-->	Bulk carrier	?

If you are **equally** concerned about a Container Vessel and Bulk Carrier you answer:

Container	<--	X	-->	Bulk carrier	?

If you are **more** concerned about a Container Vessel compared to a Bulk Carrier you answer:

Container	X--	=	-->	Bulk carrier	?

If you are **less** concerned about a Container Vessel compared to a Bulk Carrier you answer:

Container	<--	=	X>	Bulk carrier	?

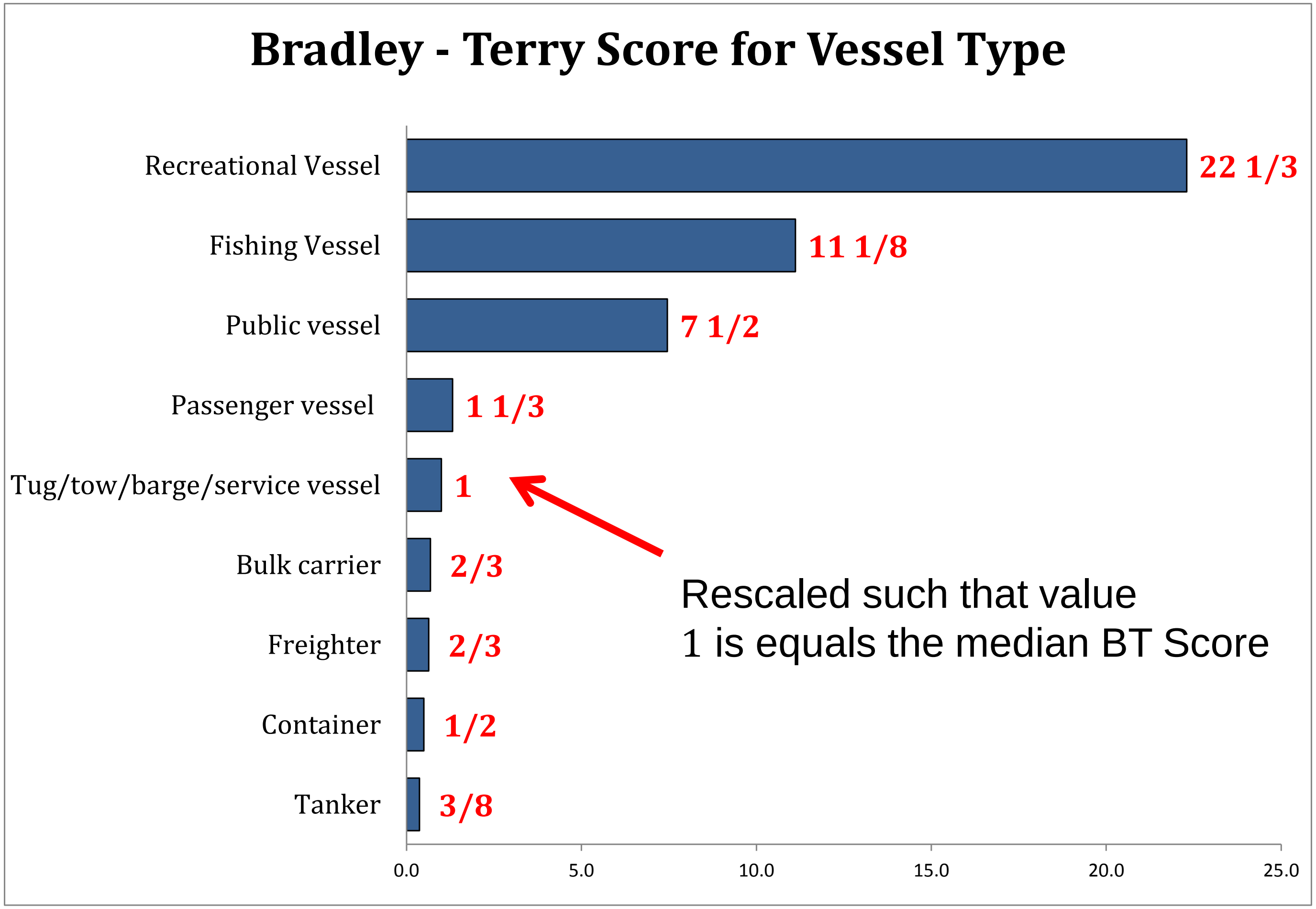
If you cannot answer this question, you answer:

Container	<--	=	-->	Bulk carrier	X

THE QUESTIONAIRE STARTS ON THE NEXT PAGE

Used to determine a measure a quantitative scale for Vessel Type to be used later in Accident Probability Model with Covariates

Example Bradley-Terry Paired Comparison Scores



Used as a quantitative scale for Vessel Type
to be used later in Accident Probability Model with Covariates

Bradley-Terry Diagnostics

Based on circular triads: A circular triad occurs when an expert says that the "**Container Vessel**" is worse than "**Bulk Carrier**", The "**Bulk Carrier**" is worse than the "**Tanker**", but next responds that the "**Tanker**" is worse than the "**Container Vessel**"

Given **the number of items** that are being compared, one can determine how many circular triads one can expect for an expert responding completely at random.

One would want **a subject matter expert** to have a statistically significant less number of circular triads than an expert responding at random. If this is not the case, one may ignore his/her responses.

Bradley-Terry also contains a diagnostic procedure to determine if agreement amongst subject matter experts is **statistically significant.**

Bradley-Terry Paired Comparison Model

- Recall, it was assumed that we have n objects each having associated with it a scale value $V_i, i = 1, \cdots, n$.
- When an expert is asked if he/she prefers object i over object j it is assumed that:

$$\Pr(Expert's\ response\ is\ i > j) = \frac{V_i}{V_i + V_j}$$

- What happens when the scale value V_i are all the same?

Thus, equals scale values is equivalent **to a series of independent coin tosses of a fair coin,** modelling a randomly Responding expert with no preference structure amongst the objects

Bradley-Terry Paired Comparison Simulation

Simulated response of a randomly responding expert comparing 9 objects

	1	2	3	4	5	6	7	8	9	a(i)	$(a(i)\text{-}a\_bar)^2$
1	0	1	1	0	0	0	0	1	0	3	1
2	0	0	0	0	0	1	1	0	1	3	1
3	0	1	0	0	1	0	0	0	0	2	4
4	1	1	1	0	0	0	1	0	1	5	1
5	1	1	0	1	0	0	1	0	1	5	1
6	1	0	1	1	1	0	0	1	0	5	1
7	1	0	1	0	0	1	0	1	1	5	1
8	0	1	1	1	1	0	0	0	1	5	1
9	1	0	1	0	0	1	0	0	0	3	1

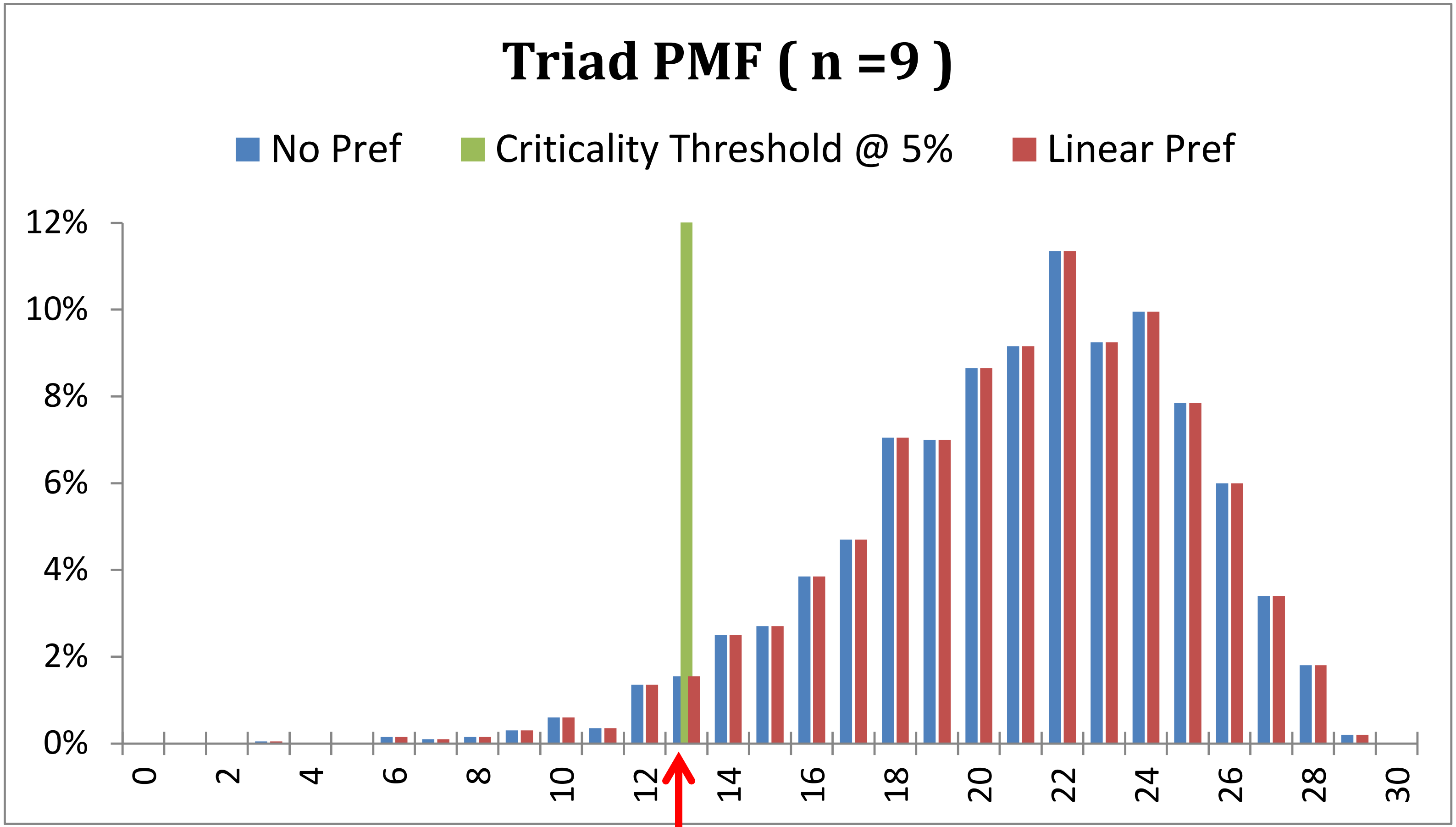
$a(i)$ = # times item i is preferred over the other objects $\quad \overline{a} = \frac{1}{n}\sum_{i=1}^{n} a(i)$

Number of circular triads in a binary matrix of binary choices - Kendall and Smith (1940):

$$C = \frac{n}{12}(n^2 - 1) - \frac{1}{2}\sum_{i=1}^{n}(a(i) - \overline{a})^2 \quad \text{and} \quad \overline{a} = \frac{1}{2}(n-1)$$

Bradley-Terry Paired Comparison Triad Distribution

Triad Distribution randomly responding expert comparing 9 objects



Criticality Threshold :13

Example Bradley-Terry Paired Comparison Expert Responses

A LOT OF DATA ENTRY

Comparisons for the Expert: E3 ...

Nr.	Id	1	2	3	4	5	6	7	8	9
1	Tanker	.	>	>	>	=	=	>	<	<
2	Container	<	.	=	=	=	<	<	<	<
3	Freighter	[illegible]	=	.	=	=	<	<	<	<
4	Bulk Carrier	<	[illegible]	=	.	=	<	<	<	<
5	Tug /tow/barge	=	[illegible]	=	[illegible]	.	<	[illegible]	[illegible]	<
6	Passenger Vess	=	>	>	[illegible]	>	.	<	>	[illegible]
7	Public Vessel	<	>	>	>	[illegible]	[illegible]	.	=	[illegible]
8	Fishing Vessel	>	>	>	>	>	<	[illegible]	[illegible]	<
9	Recreational V	>	>	>	>	>	>	<	>	.

Comparisons for the Expert: E1...

Nr.	Id	1	2	3	4	5	6	7	8	9
1	Tanker	.	=	<	<	<	>	<	<	<
2	Container	=	.	<	<	>	<	=	<	<
3	Freighter	>	>	.	<	>	>	>	<	>
4	Bulk Carrier	>	>	>	.	>	>	>	<	>
5	Tug /tow/barge	>	<	<	<	.	>	<	<	<
6	Passenger Vess	<	>	<	<	<	.	<	<	<
7	Public Vessel	>	=	<	<	>	>	.	<	>
8	Fishing Vessel	>	>	>	>	>	>	>	.	=
9	Recreational V	>	>	<	<	>	>	<	=	.

Comparisons for the Expert: E6 ...

Nr.	Id	1	2	3	4	5	6	7	8	9
1	Tanker	.	=	=	>	<	<	<	<	<
2	Container	=	.	=	>	<	<	<	<	<
3	Freighter	=	=	.	=	<	<	<	<	<
4	Bulk Carrier	<	<	=	.	<	=	<	<	<
5	Tug /tow/barge	>	>	>	>	.	<	<	<	<
6	Passenger Vess	>	>	>	=	>	.	<	<	<
7	Public Vessel	>	>	>	>	>	>	.	<	<
8	Fishing Vessel	>	>	>	>	>	>	>	.	<
9	Recreational V	>	>	>	>	>	>	>	>	.

Comparisons for the Expert: E6 ...

Nr.	Id	1	2	3	4	5	6	7	8	9
1	Tanker	.	=	<	=	<	=	<	<	<
2	Container	=	.	<	=	<	=	<	<	<
3	Freighter	>	>	.	>	>	>	<	<	<
4	Bulk Carrier	=	=	<	.	<	=	<	<	<
5	Tug /tow/barge	>	>	<	>	.	>	<	<	<
6	Passenger Vess	=	=	<	=	<	.	<	<	<
7	Public Vessel	>	>	>	>	>	>	.	<	=
8	Fishing Vessel	>	>	>	>	>	>	>	.	<
9	Recreational V	>	>	>	>	>	>	=	>	.

Comparisons for the Expert: E1...

Nr.	Id	1	2	3	4	5	6	7	8	9
1	Tanker	.	[illegible]	=	[illegible]	<	<	>	<	<
2	Container	=	.	=	[illegible]	<	[illegible]	<	<	<
3	Freighter	=	[illegible]	[illegible]	=	[illegible]	[illegible]	<	[illegible]	<
4	Bulk Carrier	=	=	=	[illegible]	[illegible]	[illegible]	<	[illegible]	[illegible]
5	Tug /tow/barge	>	>	>	>	.	[illegible]	<	[illegible]	<
6	Passenger Vess	>	>	>	=	<	.	[illegible]	<	[illegible]
7	Public Vessel	<	>	>	>	>	>	.	[illegible]	=
8	Fishing Vessel	>	>	>	>	>	>	<	.	<
9	Recreational V	>	>	>	>	>	>	=	>	.

Comparisons for the Expert: E8 ...

Nr.	Id	1	2	3	4	5	6	7	8	9
1	Tanker	.	=	=	>	<	<	<	<	<
2	Container	=	.	=	>	<	<	<	<	<
3	Freighter	=	=	.	=	<	<	<	<	<
4	Bulk Carrier	<	<	=	.	<	=	<	<	<
5	Tug /tow/barge	>	>	>	>	.	<	<	<	<
6	Passenger Vess	>	>	>	=	>	.	<	<	<
7	Public Vessel	>	>	>	>	>	>	.	<	<
8	Fishing Vessel	>	>	>	>	>	>	>	.	<
9	Recreational V	>	>	>	>	>	>	>	>	.

Comparisons for the Expert: E9 ...

Nr.	Id	1	2	3	4	5	6	7	8	9
1	Tanker	.	<	<	<	<	>	<	<	<
2	Container	>	.	<	<	<	<	<	<	<
3	Freighter	>	>	.	<	<	>	<	<	<
4	Bulk Carrier	>	>	>	.	<	>	>	<	<
5	Tug /tow/barge	>	>	>	>	.	>	<	>	<
6	Passenger Vess	<	>	<	<	<	.	<	<	<
7	Public Vessel	>	>	>	<	>	>	.	<	<
8	Fishing Vessel	>	>	>	>	<	>	>	.	<
9	Recreational V	>	>	>	>	>	>	>	>	.

Comparisons for the Expert: E5 ...

Nr.	Id	1	2	3	4	5	6	7	8	9
1	Tanker	.	=	=	=	<	<	<	<	<
2	Container	=	.	=	=	<	<	<	<	<
3	Freighter	=	=	.	=	<	<	<	<	<
4	Bulk Carrier	=	=	=	.	<	<	<	<	<
5	Tug /tow/barge	>	>	>	>	.	<	<	<	<
6	Passenger Vess	>	>	>	>	>	.	<	<	=
7	Public Vessel	>	>	>	>	>	>	.	>	<
8	Fishing Vessel	>	>	>	>	>	>	<	.	<
9	Recreational V	>	>	>	>	>	=	>	>	.

Comparisons for the Expert: [illegible]

Nr.	Id	1	2	3	4	5	6	7	8	9
1	Tanker	.	[illegible]	=	[illegible]	=	[illegible]	<	[illegible]	<
2	Container	=	.	=	[illegible]	=	[illegible]	<	<	<
3	Freighter	=	=	.	=	=	[illegible]	<	<	<
4	Bulk Carrier	=	=	=	.	=	=	<	<	<
5	Tug /tow/barge	=	=	=	=	.	=	<	<	<
6	Passenger Vess	=	=	=	=	=	.	<	<	<
7	Public Vessel	>	>	>	>	>	>	.	=	=
8	Fishing Vessel	>	>	>	>	>	>	=	.	=
9	Recreational V	>	>	>	>	>	>	=	=	.

Results Bradley-Terry Paired Comparison – Vessel Type

EXPERT	# Triads	13
1	11	1
2	1	1
3	3	1
4	9	1
5	10	1
6	2	1
7	6	1
8	6	1
9	5	1
		9

a_{ij} = # times item i is preferred over item j by 9 experts

	1	2	3	4	5	6	7	8	9
1	4.5	4.5	3.5	5	1	3.5	2	0	0
2	4.5	4.5	3	4.5	2	1	0.5	0	0
3	5.5	6	4.5	4	3	3.5	1	0	1
4	4	4.5	5	4.5	2	4.5	2	0	1
5	8	7	6	7	4.5	4.5	0	1	0
6	5.5	8	5.5	4.5	4.5	4.5	0	1	0.5
7	7	8.5	8	7	9	9	4.5	3	3.5
8	9	9	9	9	8	8	6	4.5	1
9	9	9	8	8	9	8.5	5.5	8	4.5

Bradley-Terry Paired Comparison – Scale Estimation

	1	2	3	4	5	6	[illegible]	[illegible]	[illegible]
1	4.5	4.5	3.5	5	1	3	[illegible]	[illegible]	[illegible]
2	4.5	4.5	3	4.5	2	1	0.5	0	0
3	5.5	6	4.5	4	3	3.5	1	0	1
4	4	4.5	5	4.5	2	4.5	2	0	1
5	8	7	6	7	4.5	4.5	0	1	0
6	5.5	8	5.5	4.5	4.5	4.5	0	1	0.5
7	7	8.5	8	7	9	9	4.5	3	3.5
8	9	9	9	9	8	8	6	4.5	1
9	9	9	8	8	9	8.5	5.5	8	4.5

a_{ij} = # times item i is preferred over item j by 9 experts

Maximum Likelihood Estimation

$$\max : (p_{ij})^{a_{ij}}$$

$$\text{subject to : } p_{ij} = \frac{V_i}{V_i + V_j}$$

$$\sum_{i=1}^{n} V_i = 1,\ V_i > 0$$

Bradley-Terry Paired Comparison – Scale Estimation

$\hat{p}_{ij} = \frac{a_{ij}}{E}$ = proportion item i is preferred over item j by e experts, $E = 9$

	1	2	3	4	5				
1	0.50	0.50	0.39	0.56	0.11				
2	0.50	0.50	0.33	0.50	0.22				
3	0.61	0.67	0.50	0.44	0.33				
4	0.44	0.50	0.56	0.50	0.22	0.50	0.22	0.00	0.11
5	0.89	0.78	0.67	0.78	0.50	0.50	0.00	0.11	0.00
6	0.61	0.89	0.61	0.50	0.50	0.50	0.00	0.11	0.06
7	0.78	0.94	0.89	0.78	1.00	1.00	0.50	0.33	0.39
8	1.00	1.00	1.00	1.00	0.89	0.89	0.67	0.50	0.11
9	1.00	1.00	0.89	0.89	1.00	0.94	0.61	0.89	0.50

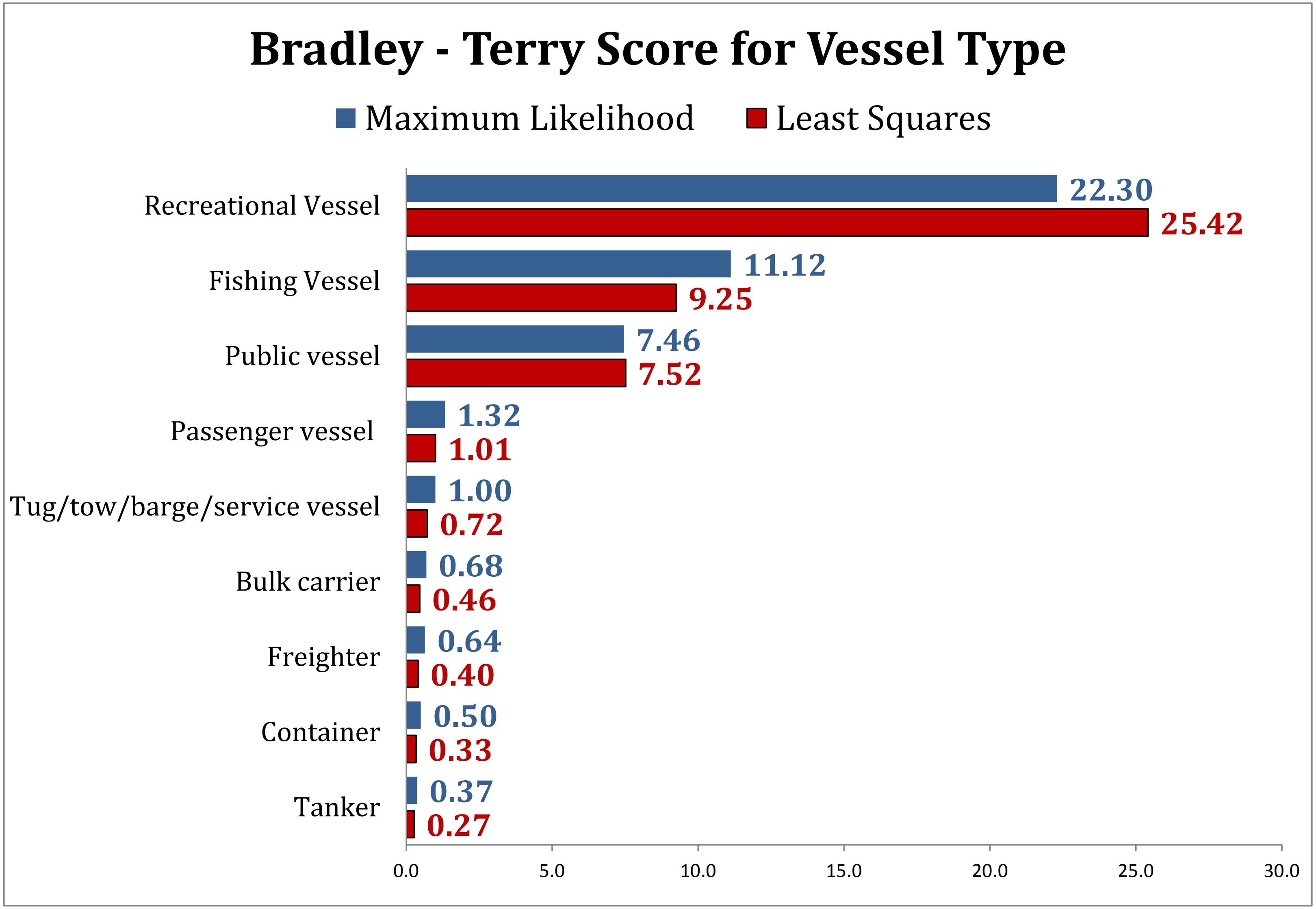
Method of Least Squares

$$\max : \sum_{i=1}^{n}\sum_{j=1}^{n}(p_{ij} - \widehat{p}_{ij})^2$$

$$\text{subject to} : \; p_{ij} = \frac{V_i}{V_i + V_j}$$

$$\sum_{i=1}^{n} V_i = 1, \; V_i > 0$$

Bradley-Terry Paired Comparison – Scale Estimation



Bradley-Terry Paired Comparison – Expert Agreement?

Coefficient of Agreement:

$$u = \frac{\sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \binom{a_{ij}}{2}}{\frac{1}{2}\binom{e}{2}\binom{n}{2}} - 1, 0 \leq u \leq 1$$

Kendal (1962) showed:

$$u' = \frac{2}{e-2} \times \binom{e}{2}\binom{n}{2}\left[u + \frac{1}{e-2}\right]$$

$$u' \sim \chi^2_{\nu}, \nu = \binom{n}{2} \times \frac{e(e-1)}{(e-2)^2}$$

Bradley-Terry Paired Comparison – Expert Agreement?

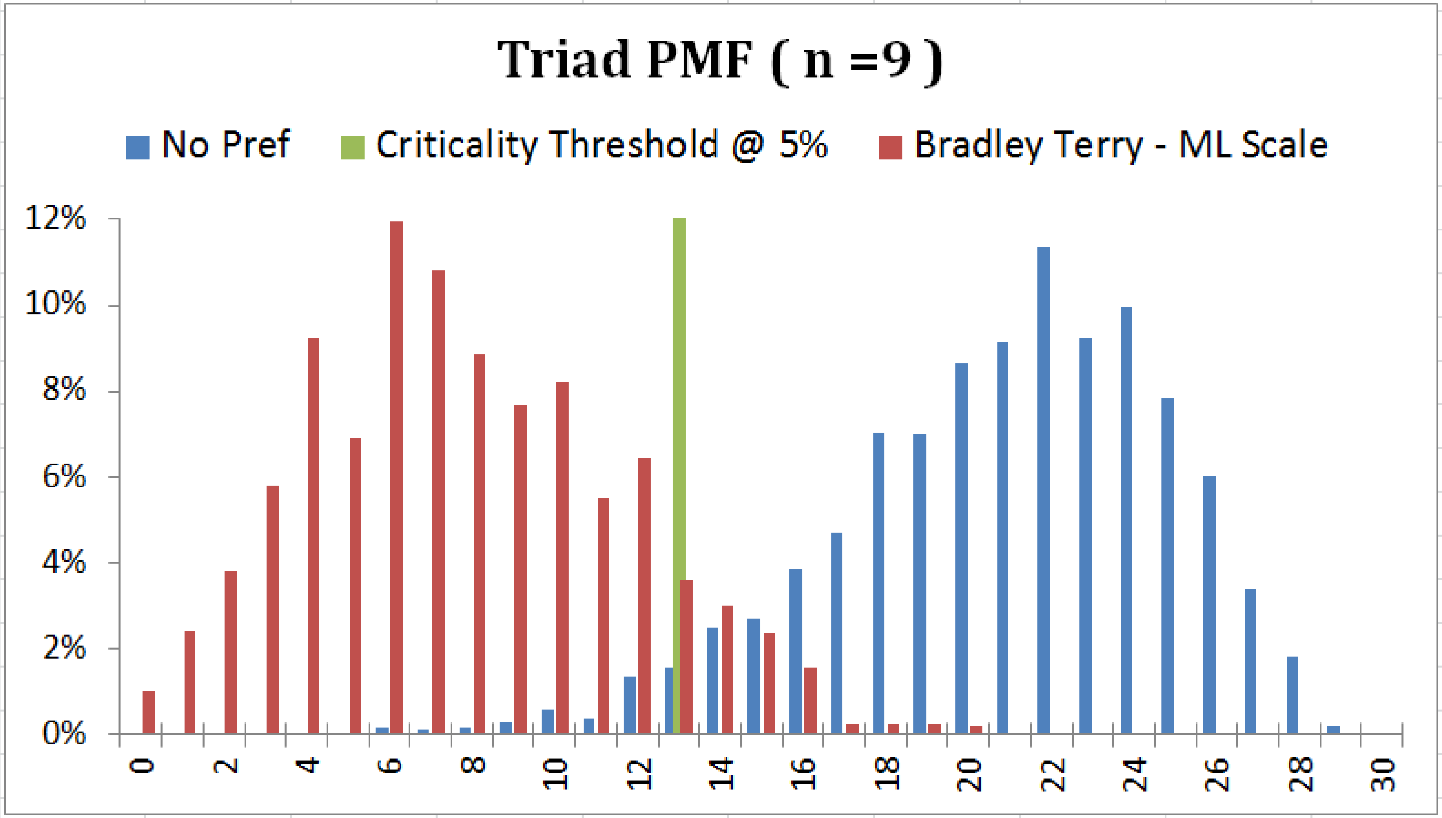
	1	2	3	4	5	6	7	8	9
1	4.5	4.5	3.5	5	1	3.5	2	0	0
2	4.5	4.5	3	4.5	2	1	0.5	0	0
3	5.5	6	4.5	4	3	3.5	1	0	1
4	4	4.5	5	4.5	2	4.5	2	0	1
5	8	7	6	7	4.5	4.5	0	1	0
6	5.5	8	5.5	4.5	4.5	4.5	0	1	0.5
7	7	8.5	8	7	9	9	4.5	3	3.5
8	9	9	9	9	8	8	6	4.5	1
9	9	9	8	8	9	8.5	5.5	8	4.5

	1	2	3	4	5	6	7	8	9
1	7.9	7.9	4.4	10.0	0.0	4.4	1.0	0.0	0.0
2	7.9	7.9	3.0	7.9	1.0	0.0	0.0	0.0	0.0
3	12.4	15.0	7.9	6.0	3.0	4.4	0.0	0.0	0.0
4	6.0	7.9	10.0	7.9	1.0	7.9	1.0	0.0	0.0
5	28.0	21.0	15.0	21.0	7.9	7.9	0.0	0.0	0.0
6	12.4	28.0	12.4	7.9	7.9	7.9	0.0	0.0	0.0
7	21.0	31.9	28.0	21.0	36.0	36.0	7.9	3.0	4.4
8	36.0	36.0	36.0	36.0	28.0	28.0	15.0	7.9	0.0
9	36.0	36.0	28.0	28.0	36.0	31.9	12.4	28.0	7.9

Number of Experts	9
Number of Objects	9
Coefficient of Agreement u	**0.413**
Transformed Coef. Agree. u'	205.898
Degrees of Freedom ν	52.898
Significance Level α	0.050
Criticality Threshold	69.83216
H0	**Agreement due to Chance**
H1	**Agreement not due to Chance**
Results	**Reject H0**

Bradley-Terry Paired Comparison – Expert Agreement?

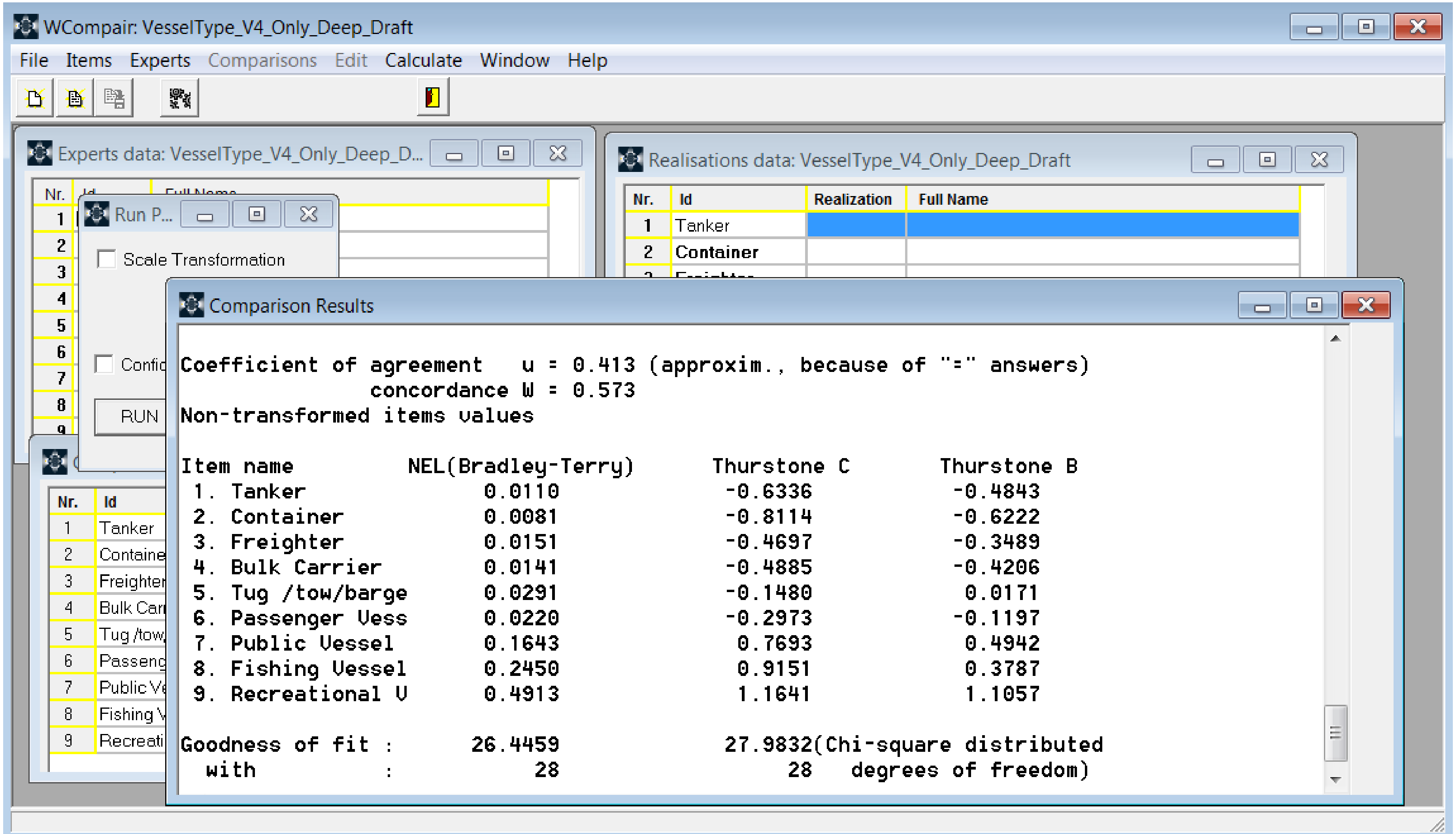
Iteration	**2000**		
n	9		
Significance Level	5%	Pr(H0 = False | H0 = True)	H0
Statistical Power	92.2%	Pr(H0 = False | H0 = False)	H1



i	1	2	3	4	5	6	7	8	9
V_i	3/8	1/2	2/3	2/3	1	1 1/3	7 1/2	11 1/8	22 1/3

Bradley-Terry Paired Comparison – Expert Agreement?

WCOMPAIR: A program that performs B-T Analysis (and Thurstone)



CLOSING REMARKS USE OF EXPERT JUDGEMENT

- **The combination or aggregation** of several expert judgments is an active research area. Should we give **a lesser consistent expert a lesser weight** when combining the judgment of several experts?

- **The quantification of uncertainty in the expert judgment** is also an active research area. Certainly, the uncertainty in the expert judgment results should preferably be conveyed to the decision maker. For example, knowing that **an average accident probability equals 1e-6** is **not particularly informative** when it may range anywhere from 1e-10 to 0.90 (for example).

- I hope you agree after this presentation that for the **use of subject matter experts in risk assessment/management to be a reasonable and trusted data source**, and one should go **beyond expert judgment elicitation techniques** that do <u>not</u> include **diagnostic analyses.**