The New NCCI Hazard Groups

Greg Engl, PhD, FCAS, MAAA
National Council on Compensation Insurance

CAS Reinsurance Seminar June, 2006
Workers Compensation Session

Agenda

- History of previous work
- Impact of remapping
- Methodology employed

Current Hazard Groups

HG	Number of Classes	Standard Premium	% of Total Premium
I	38	1,262,958,374	0.86%
II	428	67,150,463,296	45.58%
III	318	75,288,994,325	51.10%
IV	86	3,623,213,832	2.46%
Total	870	147,325,629,827	100.00%

Assigning Classes to HGs

- Prior NCCI Method
- California Approach
- ELF Based Method

Prior NCCI Method

- Hazardousness
- “Excess loss potential”

Hazardousness Variables

For each state, the following seven quantities were measured by class and expressed as ratios to the corresponding statewide value:

- Claim Frequency
- Indemnity Pure Premium
- Indemnity Severity
- Medical Pure Premium
- Medical Severity
- Total Pure Premium
- Serious Severity (including Medical)

California Methodology

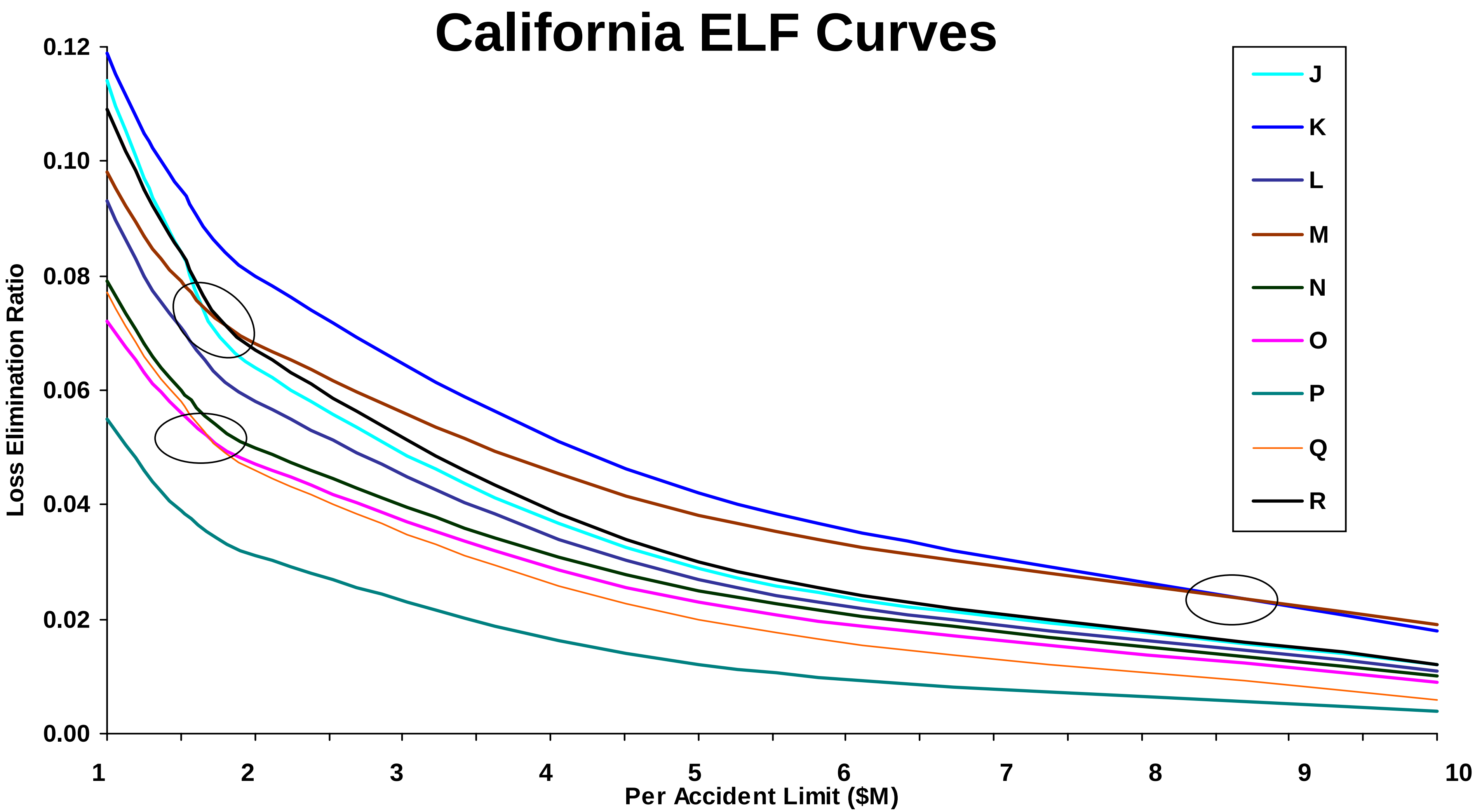
- Group classes with similar loss distributions together
- Need to precisely define ‘similar’

Crossover

Excess Loss Factors

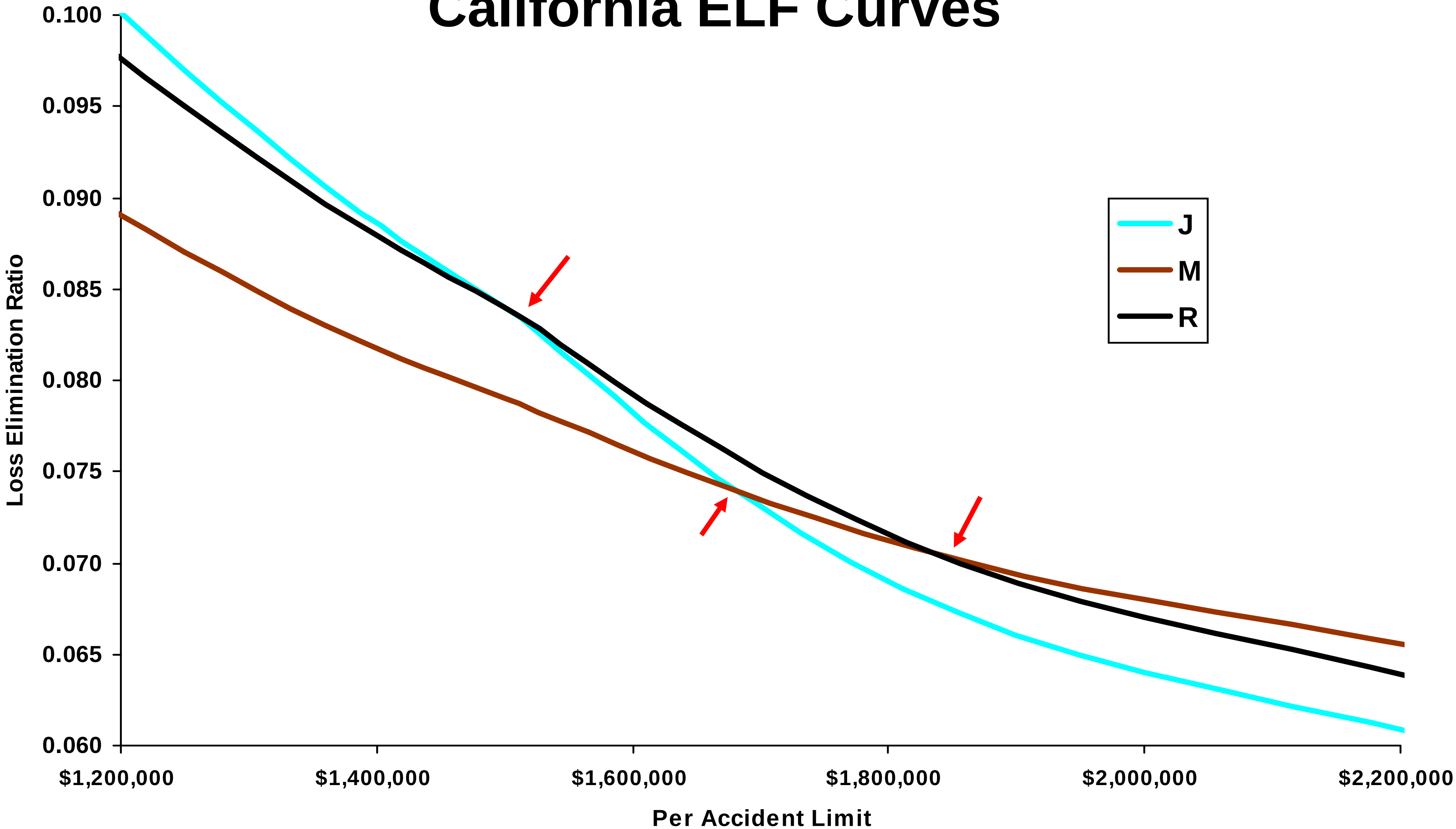
Loss Limit	Hazard Group O	Hazard Group P
25,000	0.643	0.660
600,000	0.098	0.082

Crossover


California ELF Curves
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Loss Elimination Ratio
1
2
3
4
5
6
7
8
9
10
Per Accident Limit ($M)
J
K
L
M
N
O
P
Q
R

Crossover

California ELF Curves



HG Remapping Rationale

- What are HGs used for?
- Determining ELFs

Kentucky 9/1/04 Filing

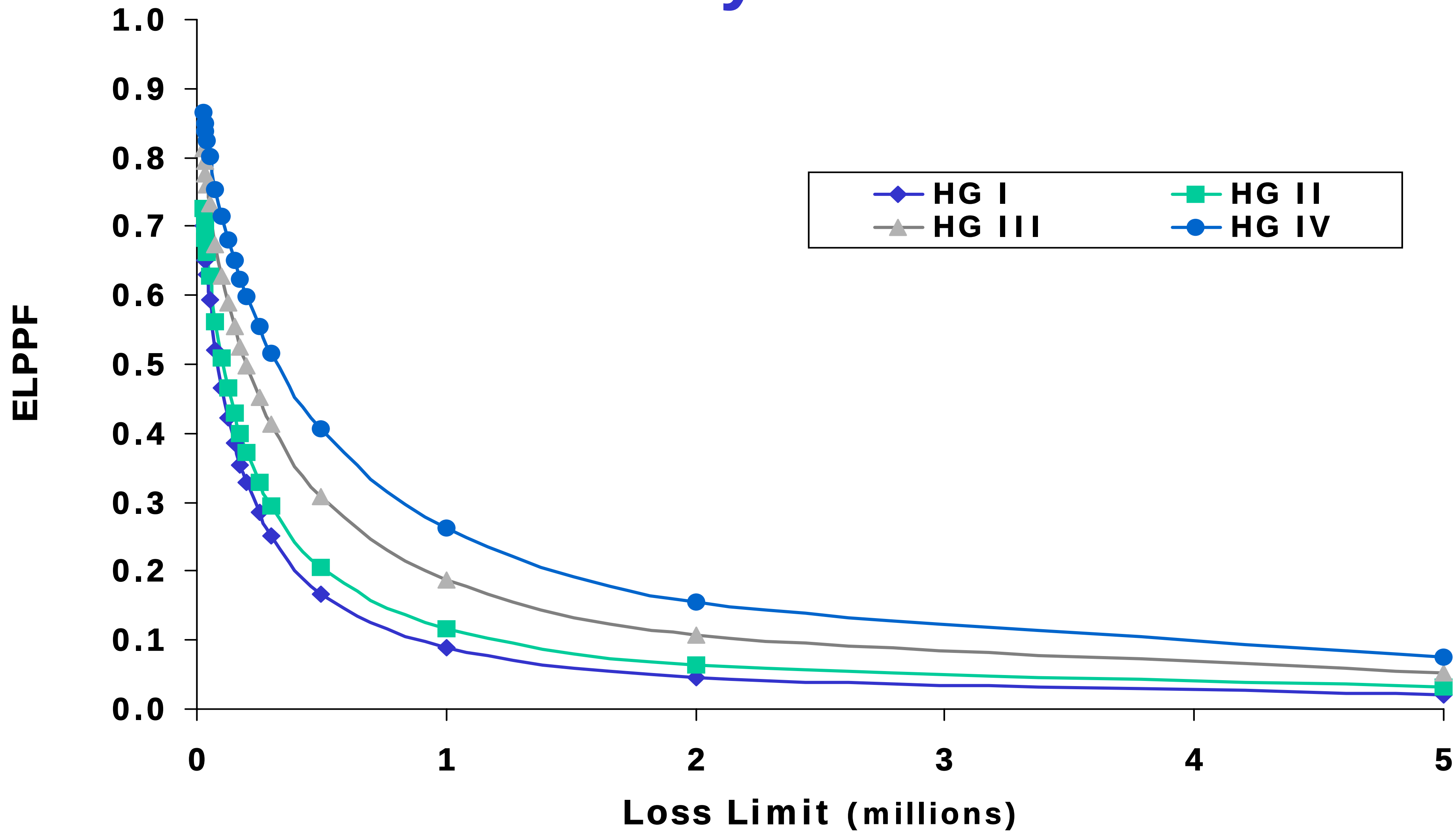
Excess Loss Pure Premium Factors

(Applicable to New and Renewal Policies)

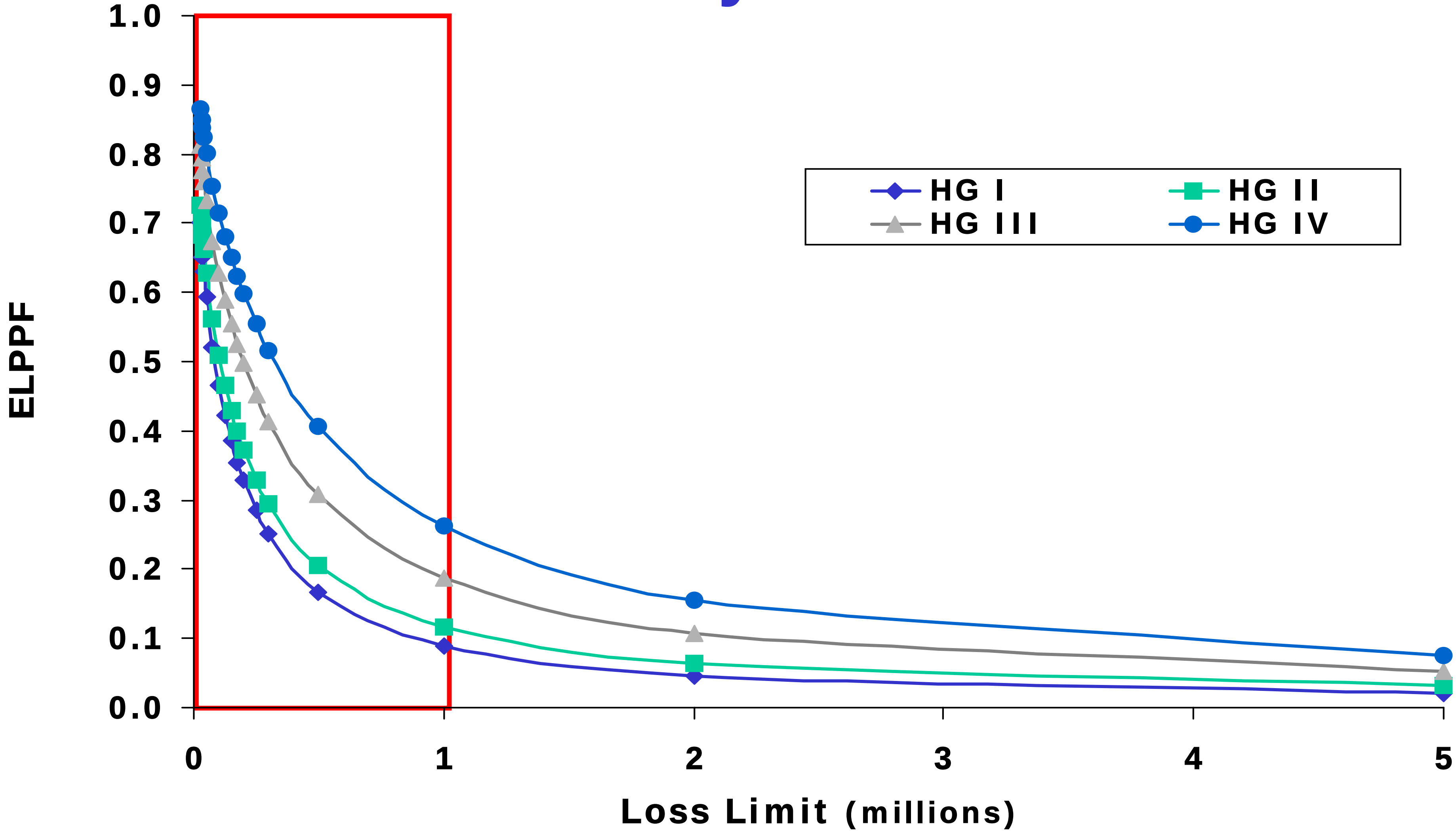
Per Accident Limitation	Hazard Groups I	II	III	IV
$25,000	0.701	0.727	0.813	0.865
$30,000	0.675	0.703	0.794	0.850*
$35,000	0.651	0.682	0.777	0.837
$40,000	0.630	0.662	0.761	0.824*
$50,000	0.593	0.628	0.732	0.802*
$75,000	0.521	0.561	0.674	0.754
$100,000	0.466	0.509	0.628	0.714*
$125,000	0.422	0.466	0.589	0.680
$150,000	0.386	0.430	0.555	0.650
$175,000	0.355	0.399	0.525	0.623
$200,000	0.328	0.373	0.498	0.598
$250,000	0.285	0.329	0.452	0.555
$300,000	0.251	0.294	0.413	0.517
$500,000	0.167	0.205	0.308	0.406
$1,000,000	0.088	0.117	0.188	0.262
$2,000,000	0.045	0.064	0.108	0.155
$5,000,000	0.020	0.032	0.053	0.075

* Also applicable to Underground Coal Mine classifications.

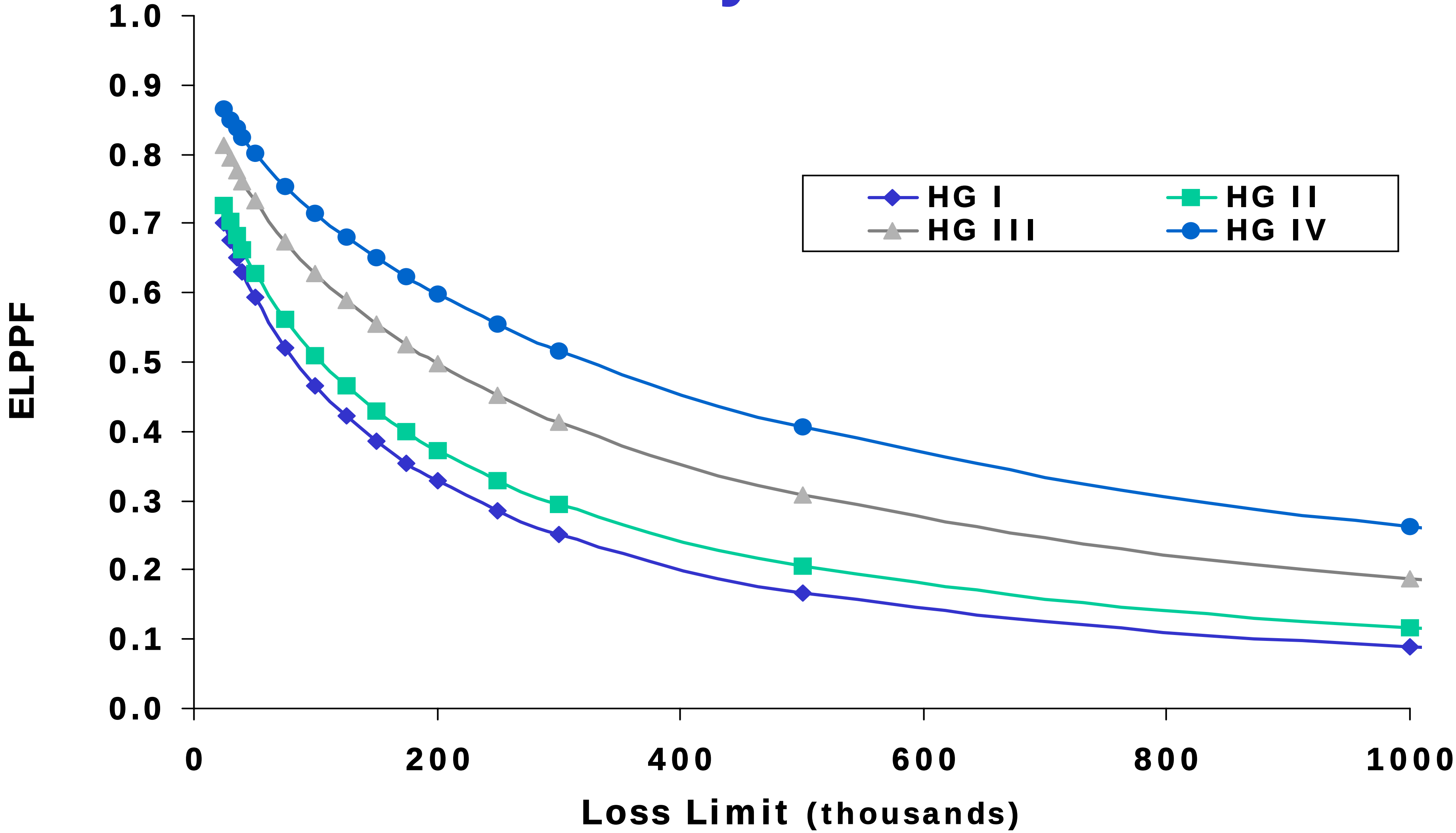
Kentucky ELPPFs



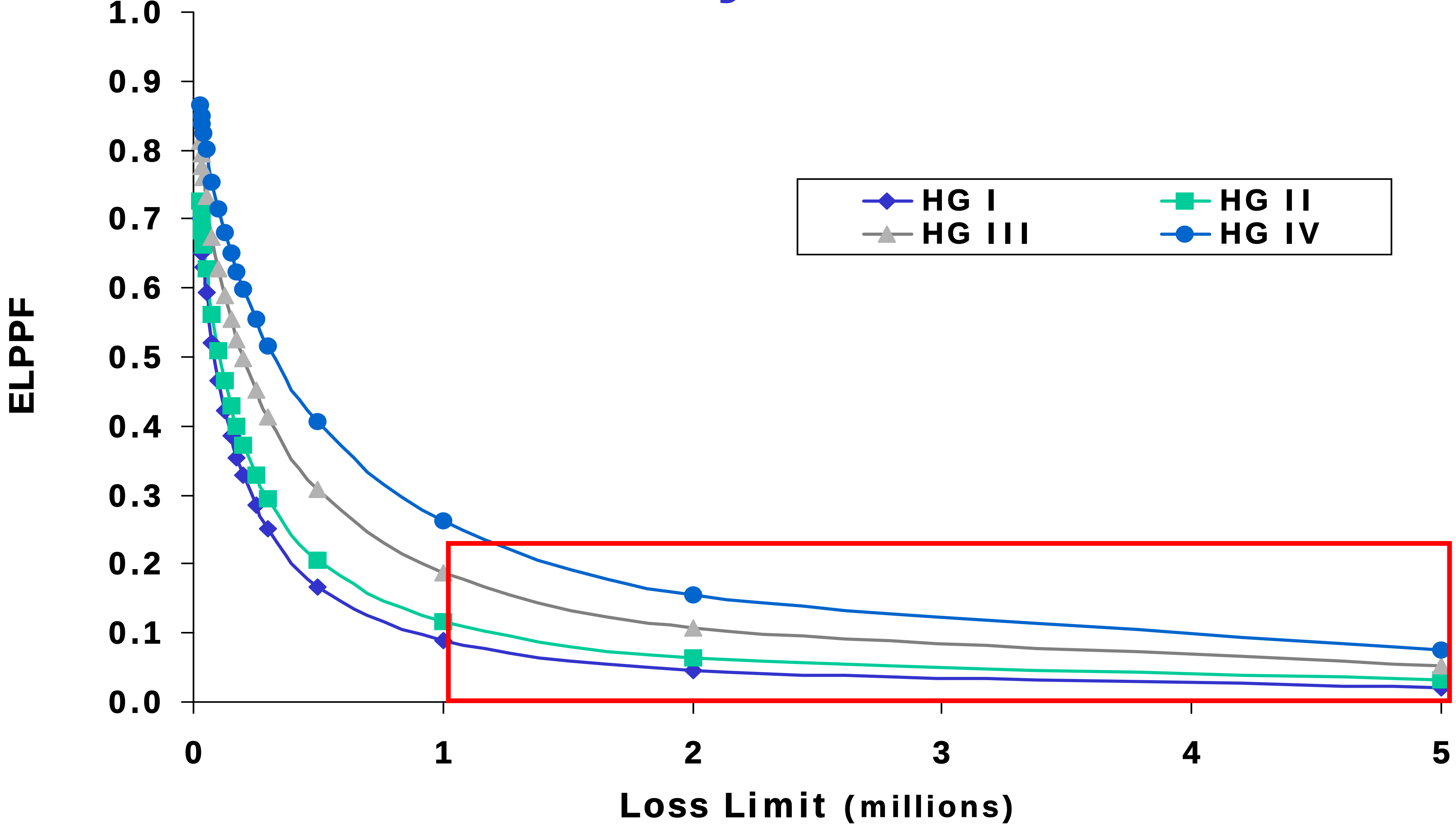
Kentucky ELPPFs



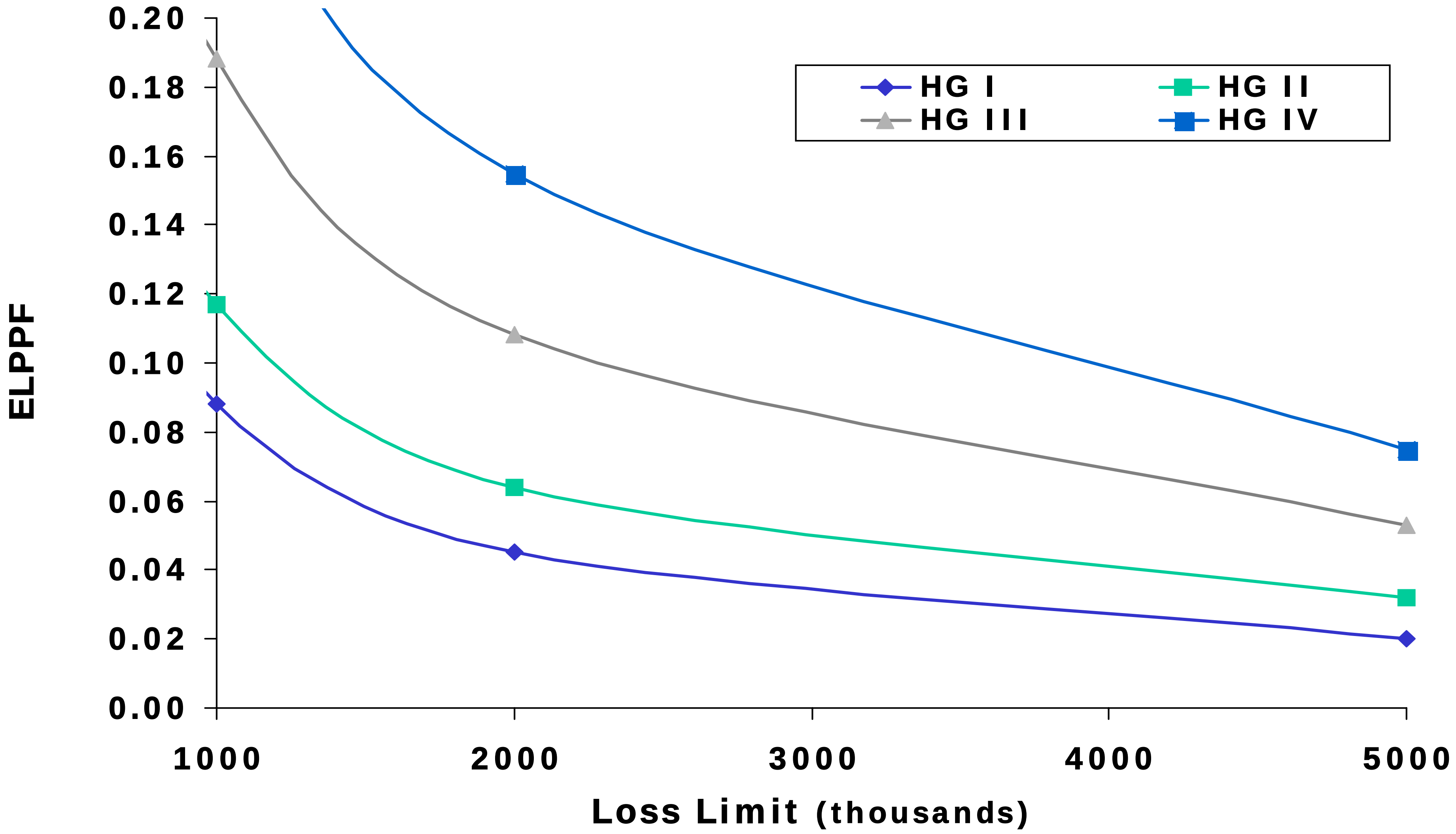
Kentucky ELPPFs



Kentucky ELPPFs



Kentucky ELPPFs



HG Remapping Approach

- Makes sense to sort classes by ELF vectors
- Class ELF vectors approximated by HG ELF vectors
- ELF curves characterize loss distribution

Excess Ratio Calculations

$$R(A) = \sum w_i R_i(A/\mu_i)$$

R_i = excess ratio function for injury type i

$$w_i = L_i / \sum L_i$$

L_i = injury type i losses

μ_i = mean injury type i loss

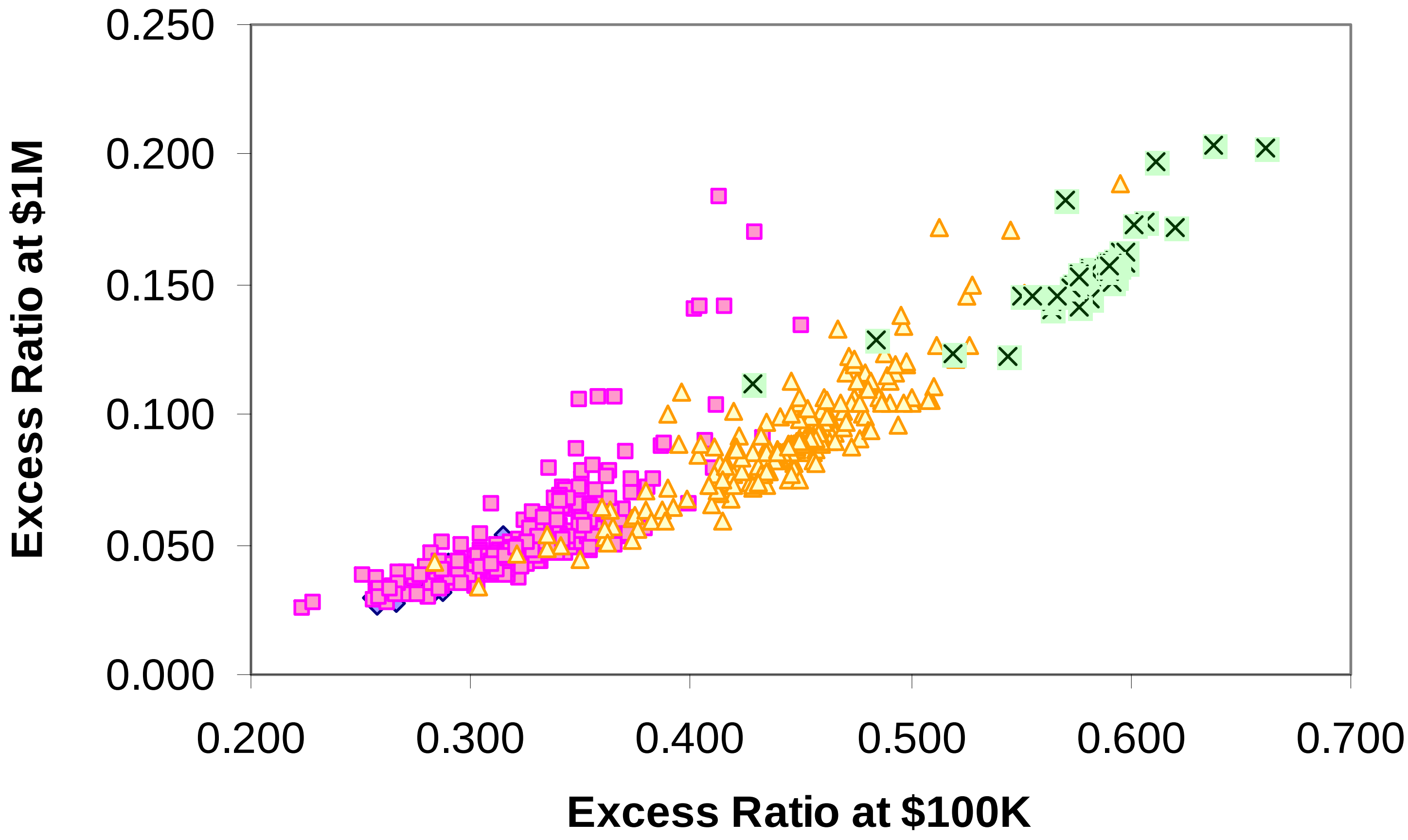
HG Remapping
Basic Data

- For each class code, c, we have a vector of ELFs:

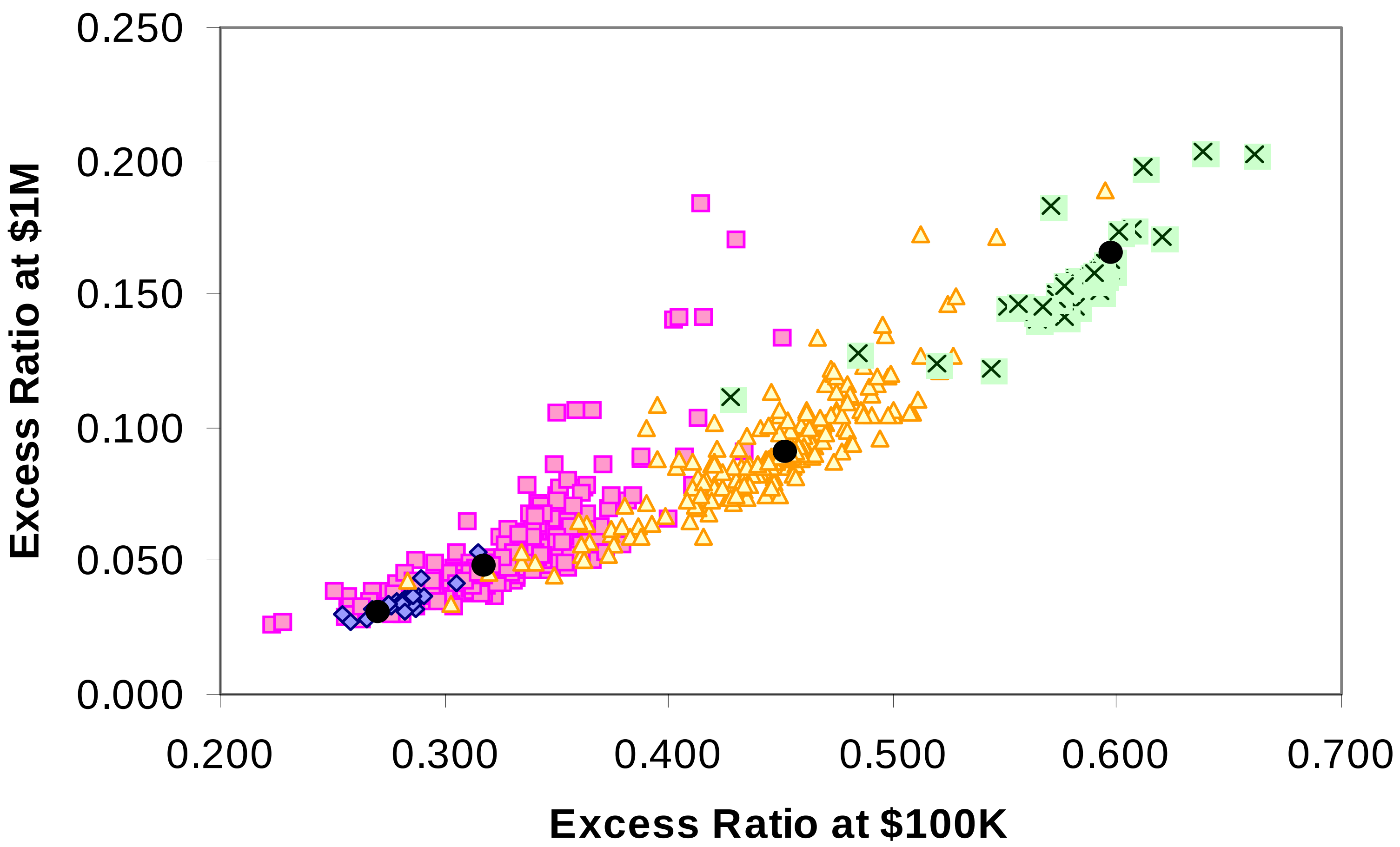
$$x_c = (x_{25K}^c, x_{30K}^c, \ldots, x_{5M}^c)$$

- Credibility weight with current HG ELF vector

Current Hazard Groups

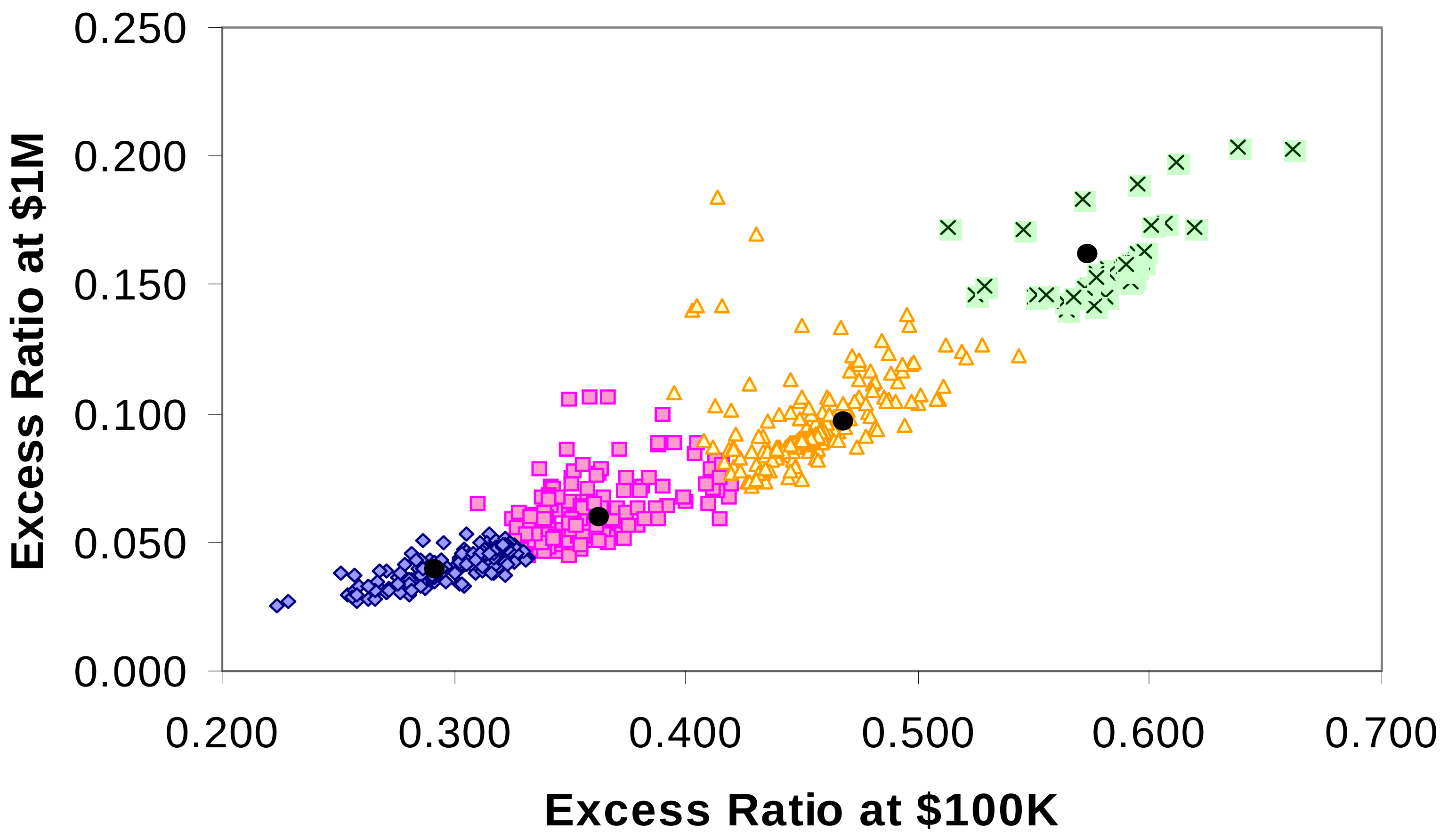


Current Hazard Groups



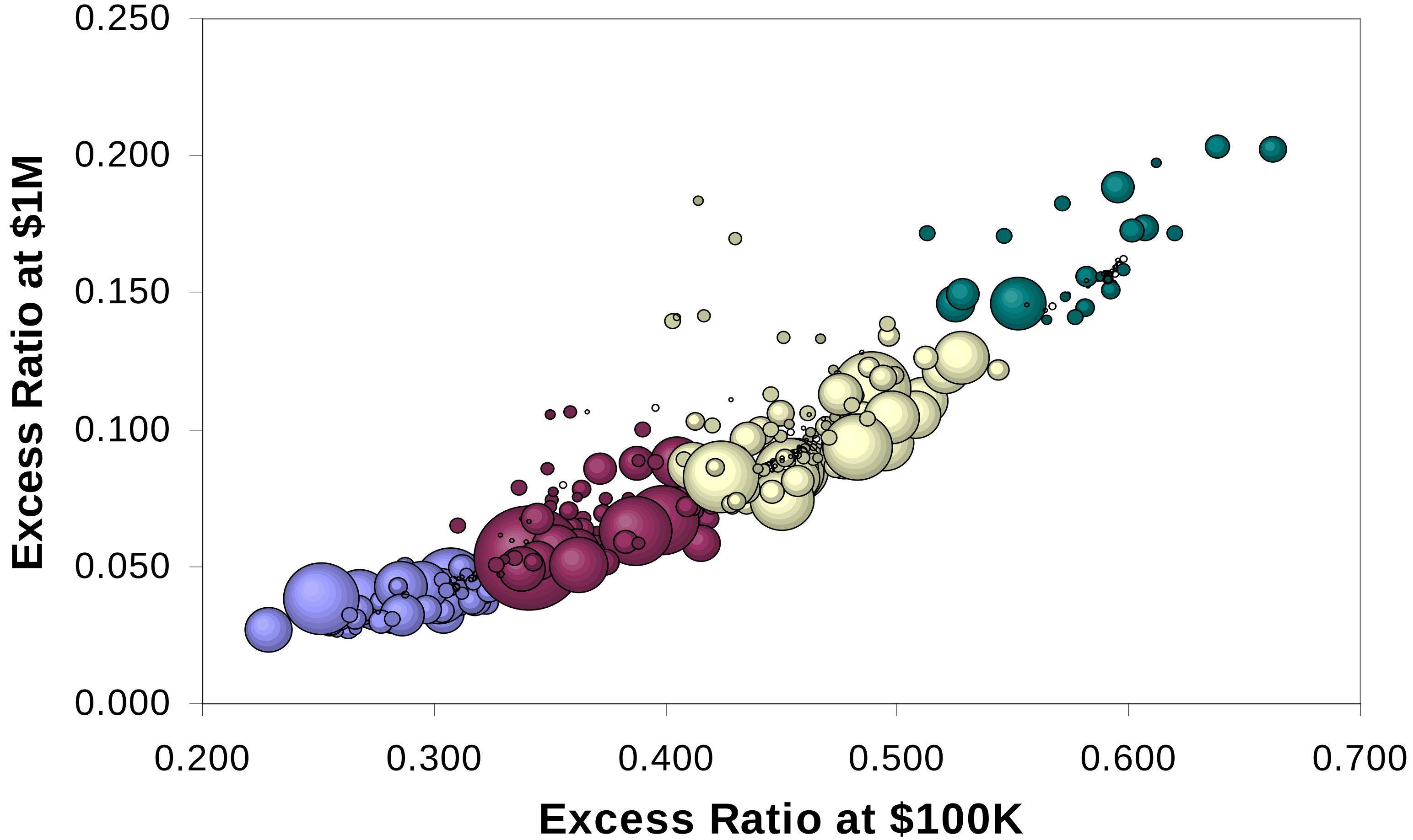
New 4 Hazard Groups

Preliminary Mapping



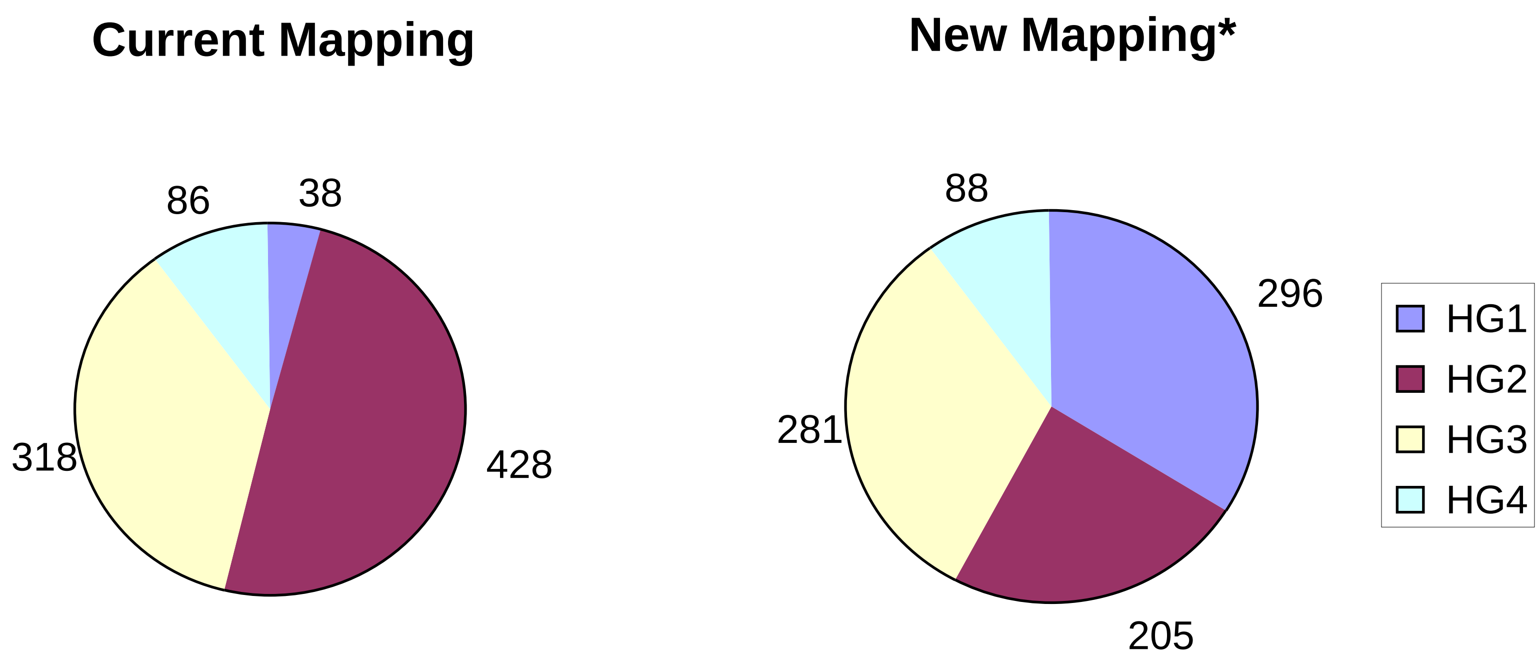
New 4 Hazard Groups

Preliminary Mapping



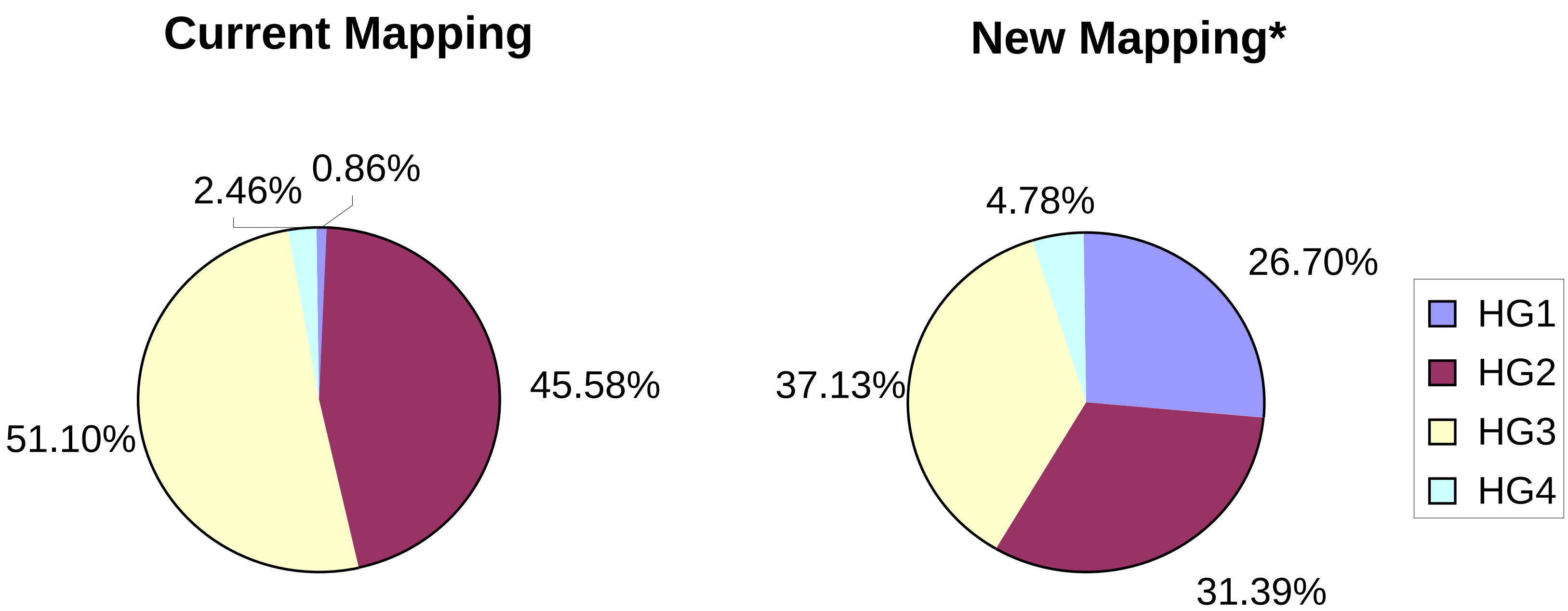
4 Hazard Group Comparison

Number of Classes per Hazard Group



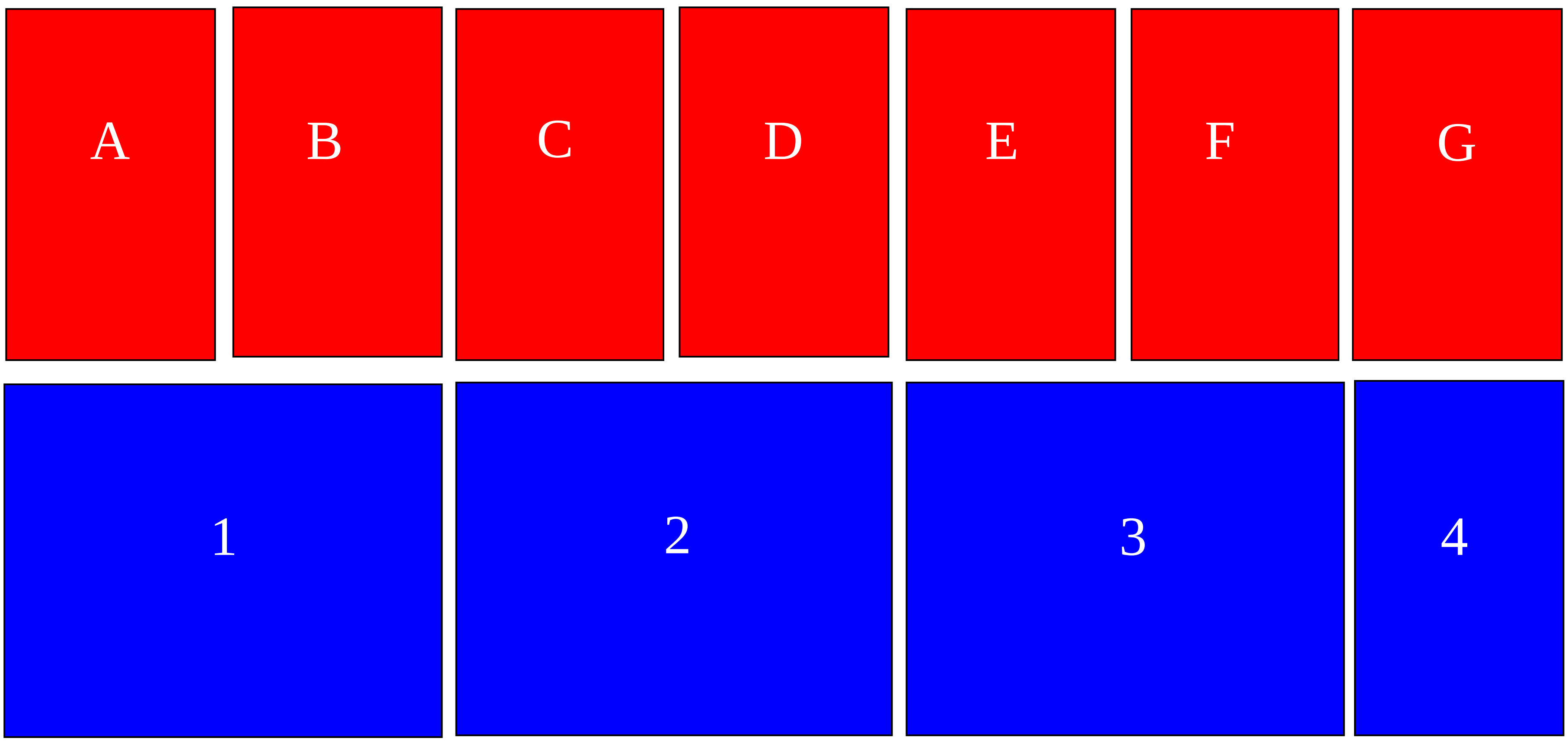
* Preliminary

4 Hazard Group Comparison
Percent of Premium Per Hazard Group



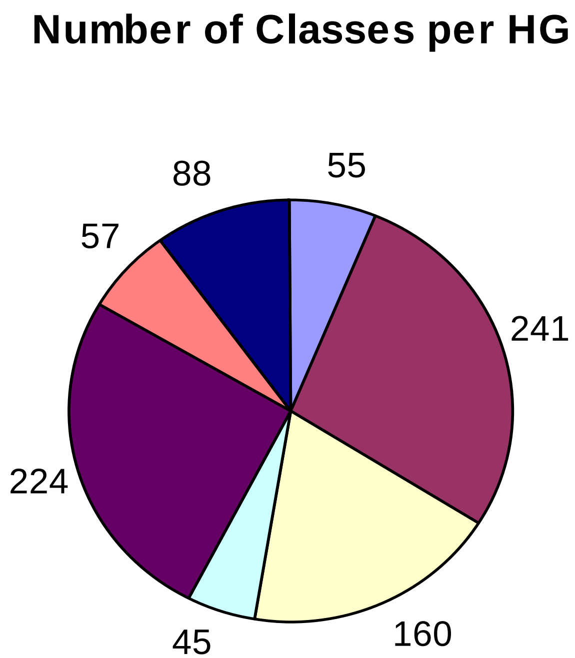
* Preliminary

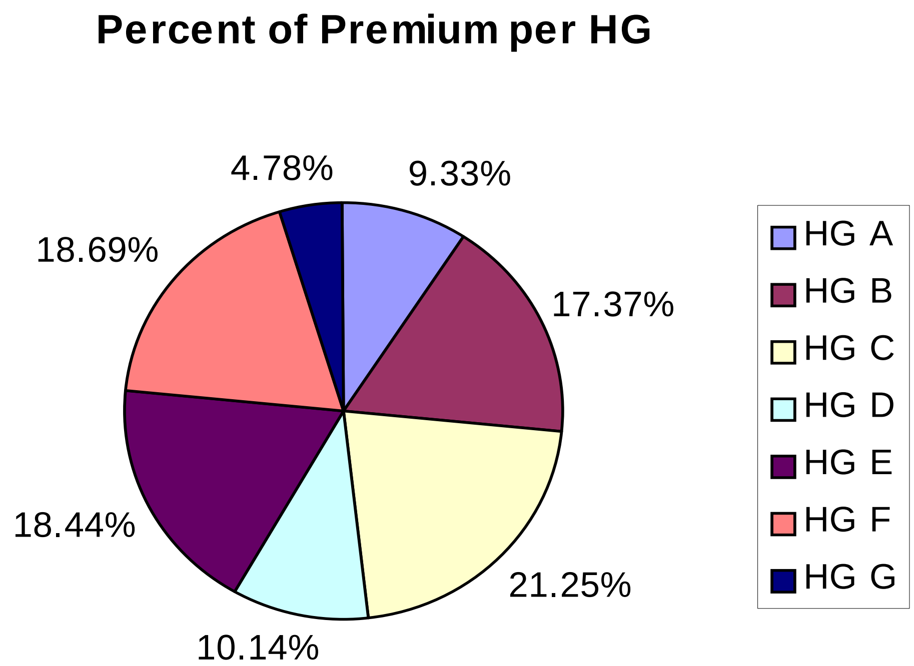
Hierarchical Collapsing of New Mapping



New Hazard Groups

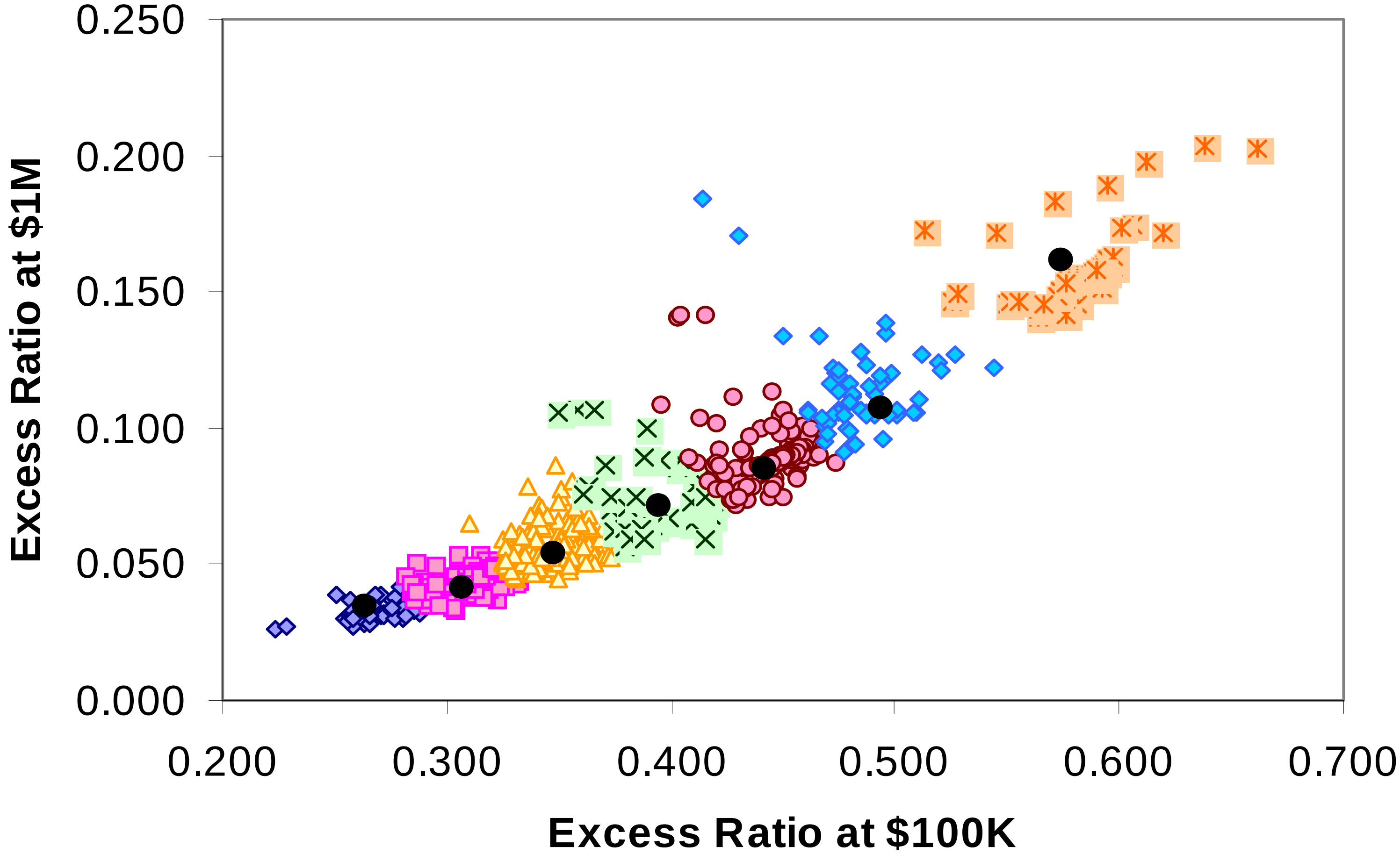
Preliminary Mapping





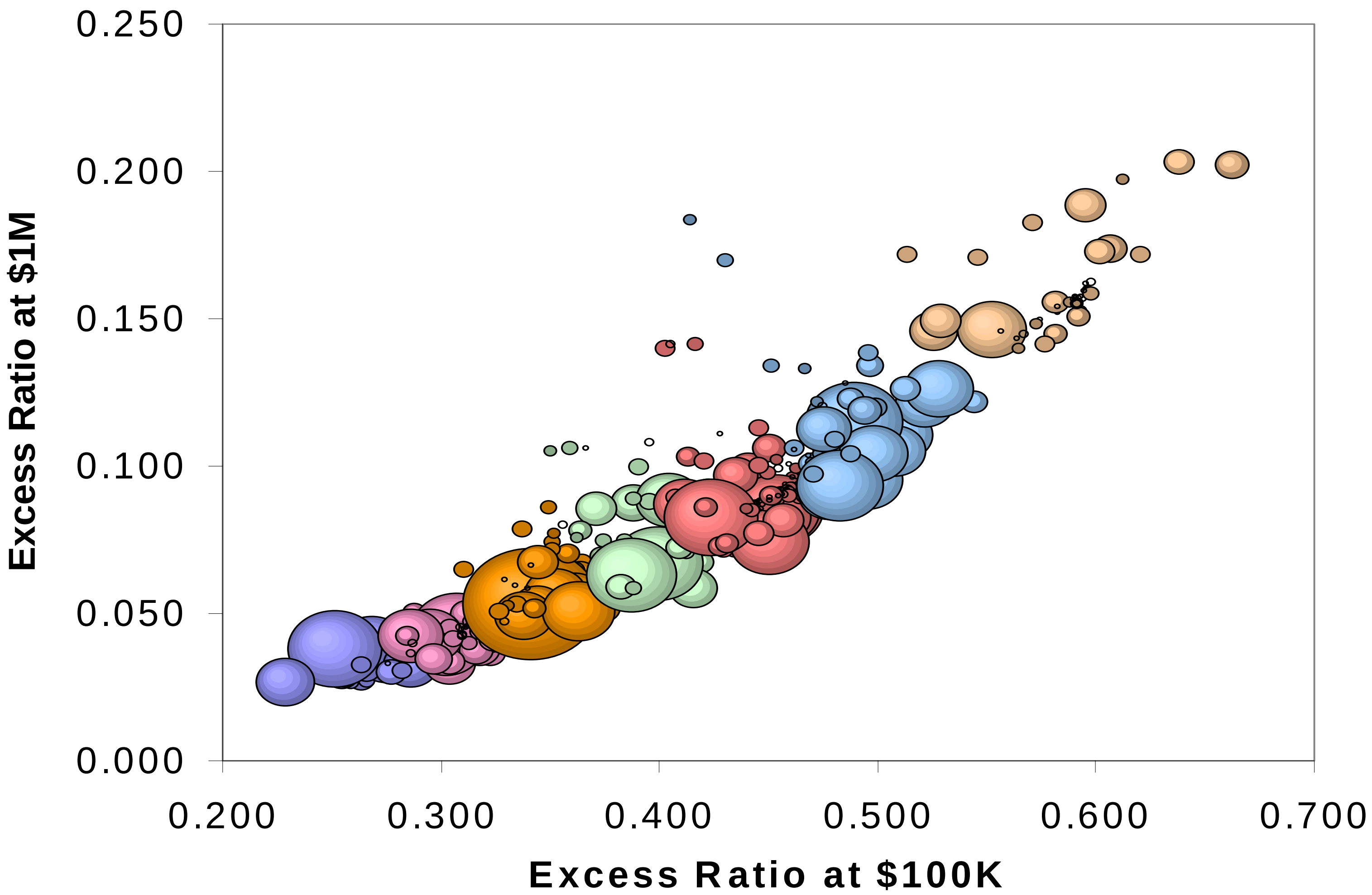
New Hazard Groups

Preliminary Mapping

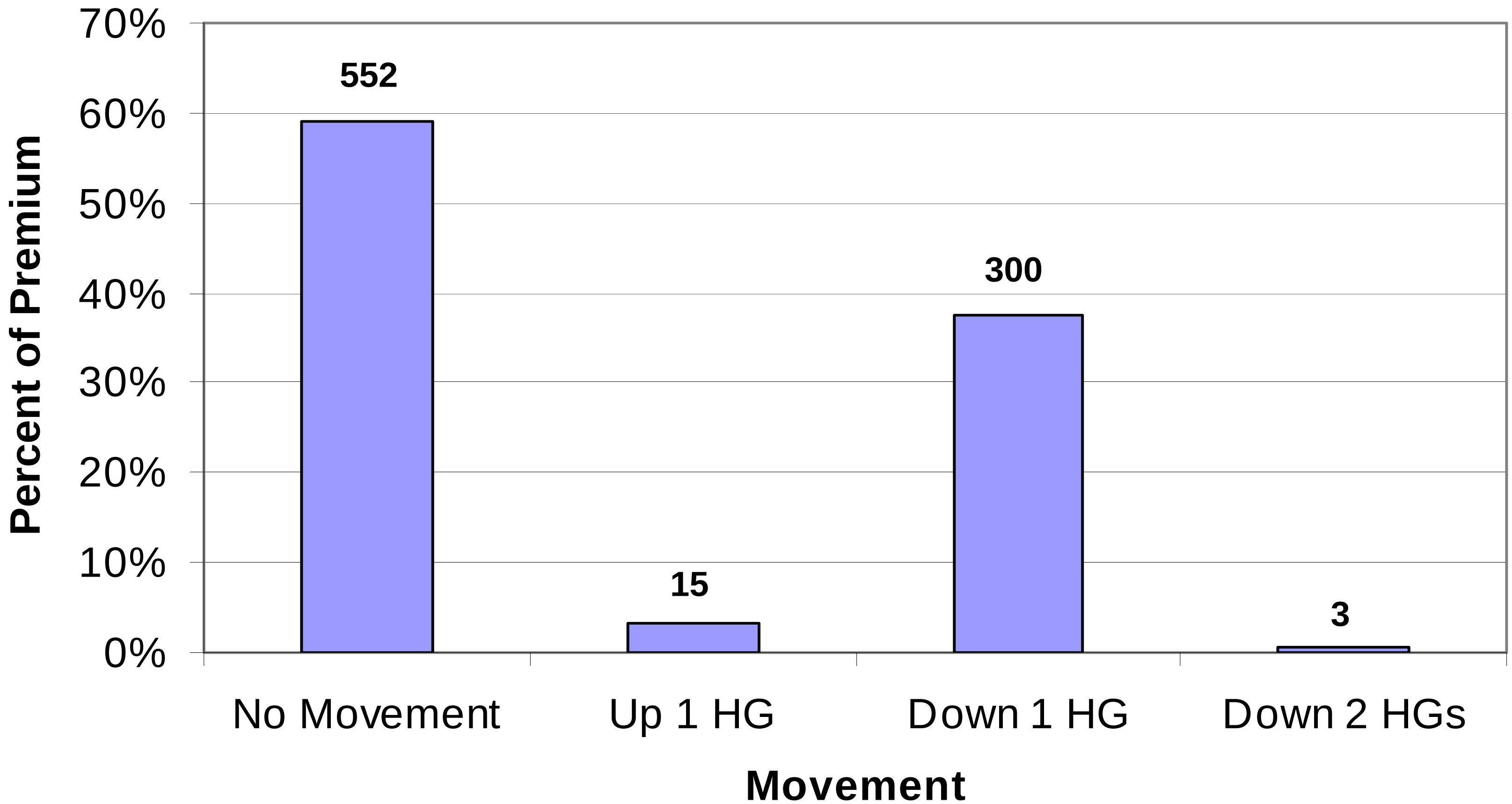


New Hazard Groups

Preliminary Mapping



Percent of Premium Moved
Current Mapping to New 4 Hazard Groups
(Based on Preliminary New Mapping)



*** Number above bar represents the number of classes in each category.**

Movement of Classes

Based on Preliminary New Mapping

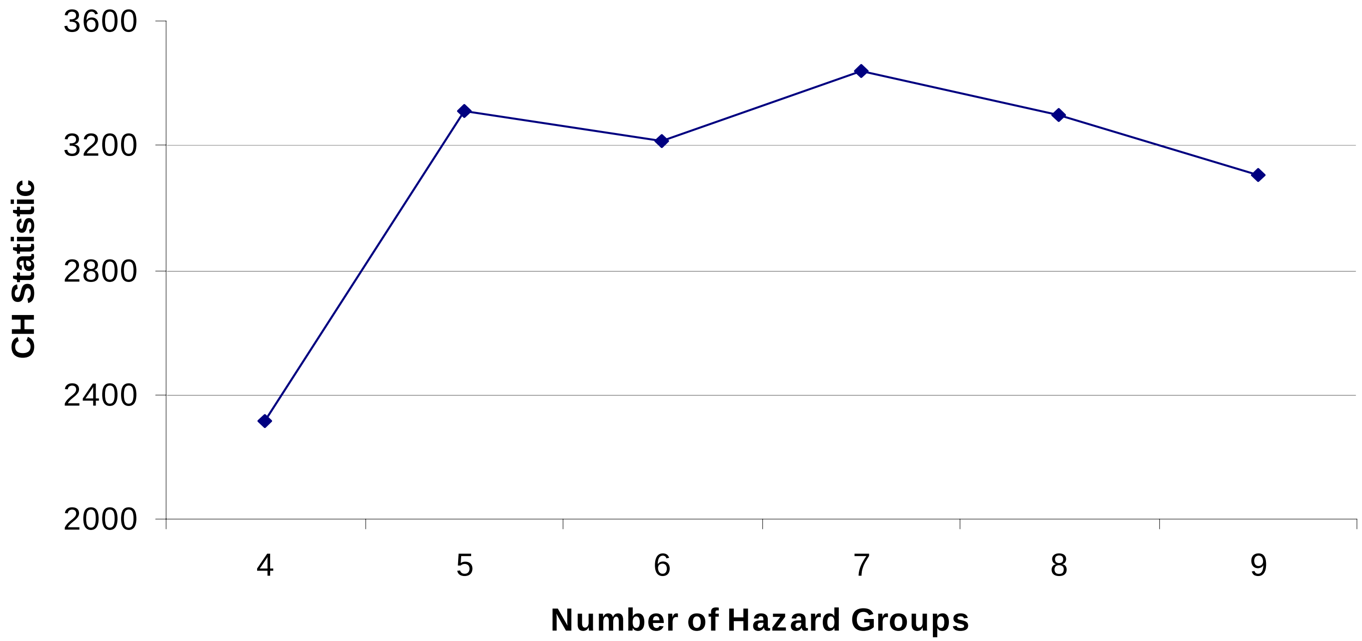
Current Mapping

Hazard Group	I	II	III	IV	Total
Number of Classes	38	428	318	86	870
% Premium	0.9%	45.6%	51.1%	2.5%	100%

New Mapping

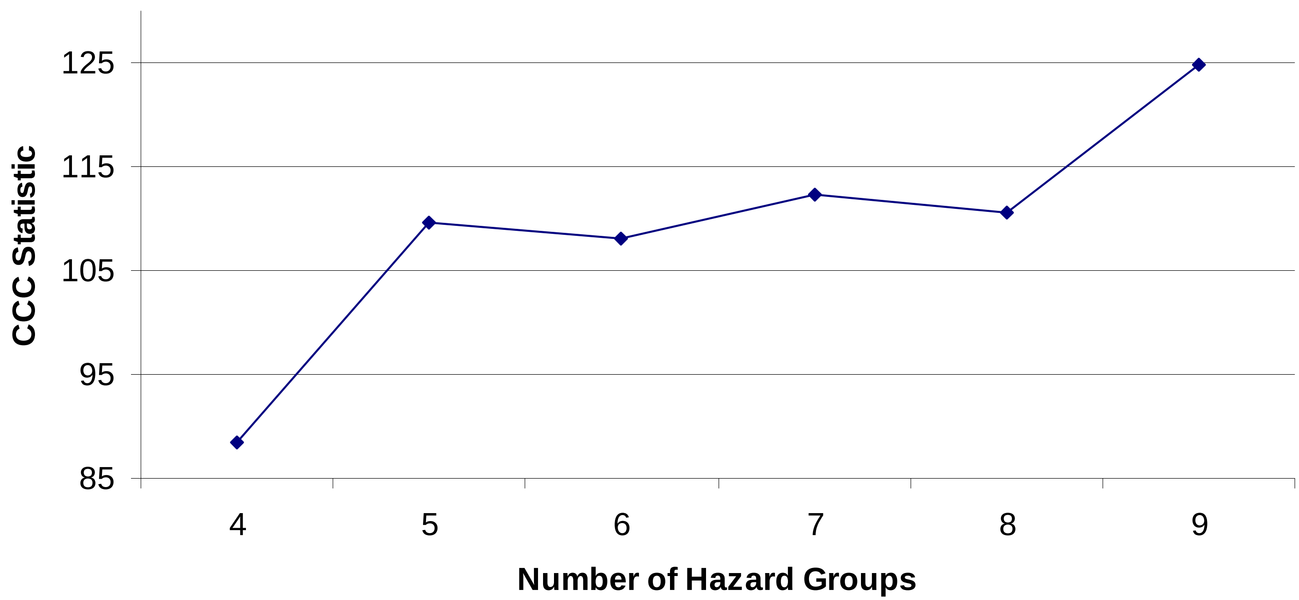
Hazard Group	I	II	III	IV	Total
1	38 0.9%	255 25.4%	3 0.5%	0 0.0%	296 26.7%
2	0 0.0%	164 19.6%	41 11.8%	0 0.0%	205 31.4%
3	0 0.0%	9 0.6%	268 36.3%	4 0.2%	281 37.1%
4	0 0.0%	0 0.0%	6 2.6%	82 2.2%	88 4.8%

Number of Hazard Groups
Calinski and Harabasz



Number of HGs	4	5	6	**7**	8	9
CH Statistic	2317	3310	3213	**3442**	3297	3102

Number of Hazard Groups
Cubic Clustering Criterion



Number of HGs	4	5	6	7	8	**9**
CCC Statistic	89	110	108	112	111	**125**

Number of Hazard Groups

Calinski and Harabasz

Number of HGs	All Classes	50% Credibility Classes	Full Credibility Classes
4	2317	793	433
5	3310	759	393
6	3213	705	450
7	**3442**	**1025**	**638**
8	3297	958	620
9	3102	915	584

Cubic Clustering Criterion

Number of HGs	All Classes	50% Credibility Classes	Full Credibility Classes
4	89	51	37
5	110	50	34
6	108	48	36
7	112	**59**	**42**
8	111	57	41
9	**125**	56	40

Three Key Ideas

- Map based on ELFs
- Compute ELFs by class
- Cluster Analysis

HG Remapping
Objective

Break C = set of all class codes, into Hazard Groups:

$$C = \bigcup_i HG_i$$

HG Remapping
Basic Data

For each class code, c, we have a vector of ELFs:

$$x_c = (x_{25K}^c, x_{30K}^c, \ldots, x_{5M}^c)$$

Using Hazard Groups

- $\bar{x}_i = \frac{1}{|HG_i|} \sum_{c \in HG_i} x_c$ (HG mean)
- approx x_c by $\bar{x}_i$ for $c \in HG_i$
- Want x_c as close as possible to $\bar{x}_i$

HG Remapping Method
k-means

Splits classes into HGs to minimize

$$\sum_{i=1}^{k} \sum_{c \in HG_i} \left\| x_c - \bar{x}_i \right\|^2$$

Optimal HGs

- % of total variance explained
- Analogous to an R-squared
- k-means maximizes this

R-squared

$$R^2 = 1 - \frac{\sum_{i=1}^{k} \sum_{c \in HG_i} \left\| x_c - \bar{x}_i \right\|^2}{\sum_{c} \left\| x_c - \bar{x} \right\|^2}$$

Optimal HGs

- Want well separated, homogeneous HGs
- Minimize within variance
- Maximize between variance

Optimal HGs

- Between variance vs. within variance
- Have one variance for each variable (ELFs at different attachment points)
- Need to consider variance-covariance matrices

Optimal HGs

Dispersion matrix of whole data set is given by

$$T = \sum_{c} \left(x_c - \overline{x}\right)^T \left(x_c - \overline{x}\right)$$

Dispersion Matrix

$$T = (n-1)\begin{bmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{12} & \hat{\sigma}_{13} & \hat{\sigma}_{14} & \hat{\sigma}_{15} \\ \hat{\sigma}_{21} & \hat{\sigma}_{22} & \hat{\sigma}_{23} & \hat{\sigma}_{24} & \hat{\sigma}_{25} \\ \hat{\sigma}_{31} & \hat{\sigma}_{32} & \hat{\sigma}_{33} & \hat{\sigma}_{34} & \hat{\sigma}_{35} \\ \hat{\sigma}_{41} & \hat{\sigma}_{42} & \hat{\sigma}_{43} & \hat{\sigma}_{44} & \hat{\sigma}_{45} \\ \hat{\sigma}_{51} & \hat{\sigma}_{52} & \hat{\sigma}_{53} & \hat{\sigma}_{54} & \hat{\sigma}_{55} \end{bmatrix}$$

Optimal HGs

Dispersion matrix of HG_i is given by

$$W_i = \sum_{c \in HG_i} \left(x_c - \bar{x}_i\right)^T \left(x_c - \bar{x}_i\right)$$

Optimal HGs

- If we let $B_i = \left| HG_i \right| \left(\bar{x}_i - \bar{x} \right)^T \left(\bar{x}_i - \bar{x} \right)$

- Then

$$\sum_{c \in HG_i} \left(x_c - \bar{x} \right)^T \left(x_c - \bar{x} \right) = B_i + W_i$$

Optimal HGs

- Pooled within group dispersion matrix

$$W = \sum_{i=1}^{k} W_i$$

- Weighted between group dispersion matrix

$$B = \sum_{i=1}^{k} B_i$$

Optimal HGs

- Between variance vs. within variance
- T = B + W
- k-means minimizes trace W

Credibility

- Compute class ELFs
- Assign a credibility to each class
- Use current HG as complement

History

- Hazard Groups were last remapped in 1993
- Prior to that, Hazard Groups were remapped in 1981
- Same credibility method used both times

Pseudo-Bühlmann

$$z = \min\left(\frac{n}{n+k} \times 1.5, 1\right)$$

where

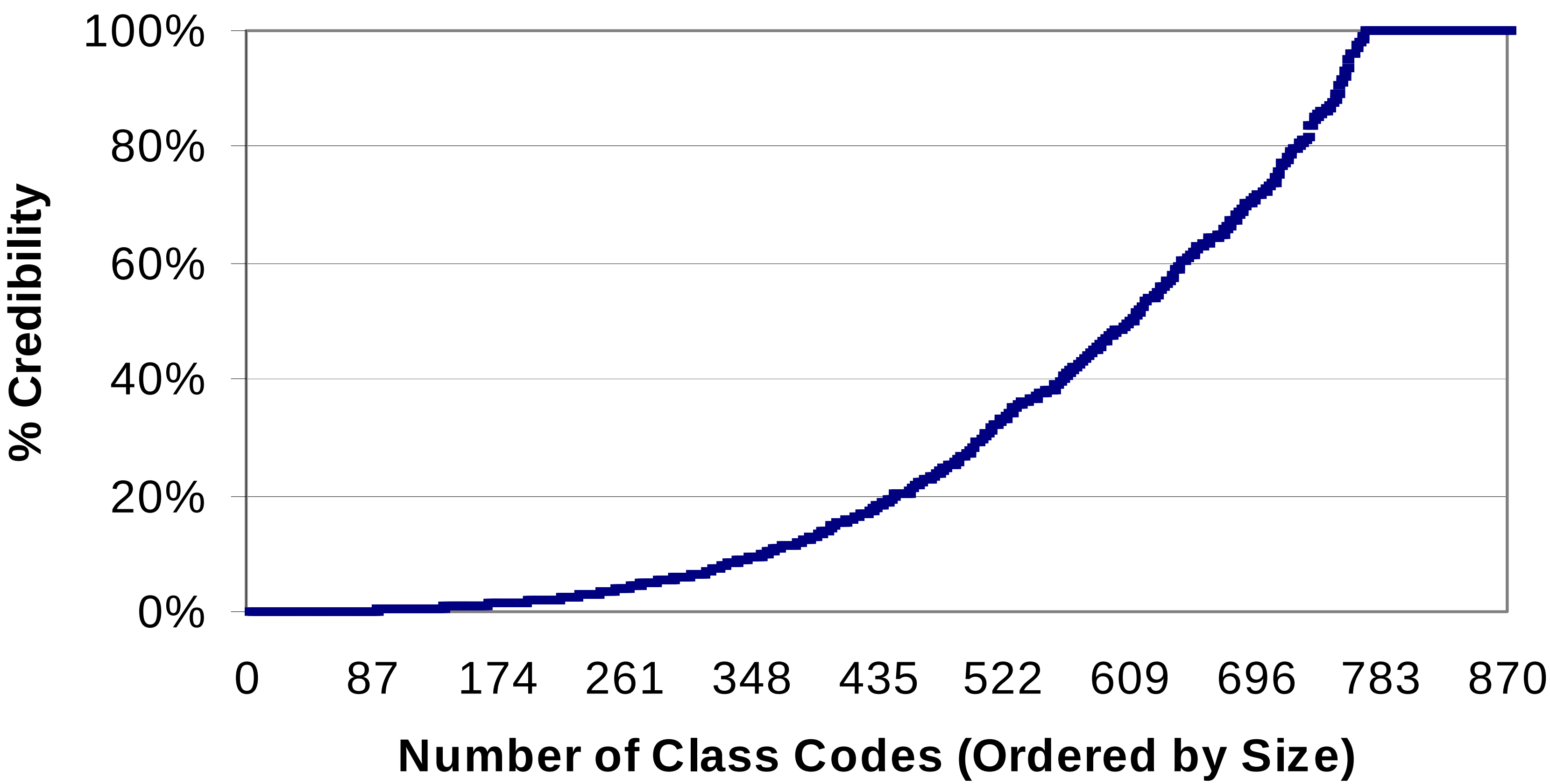
n = number of claims

$$k = \bar{n}$$

Pseudo-Bühlmann

- A class with the average number of claims gets 75% credibility.
- Classes with twice the average number of claims get full credibility.
- Based on $n/n+k$

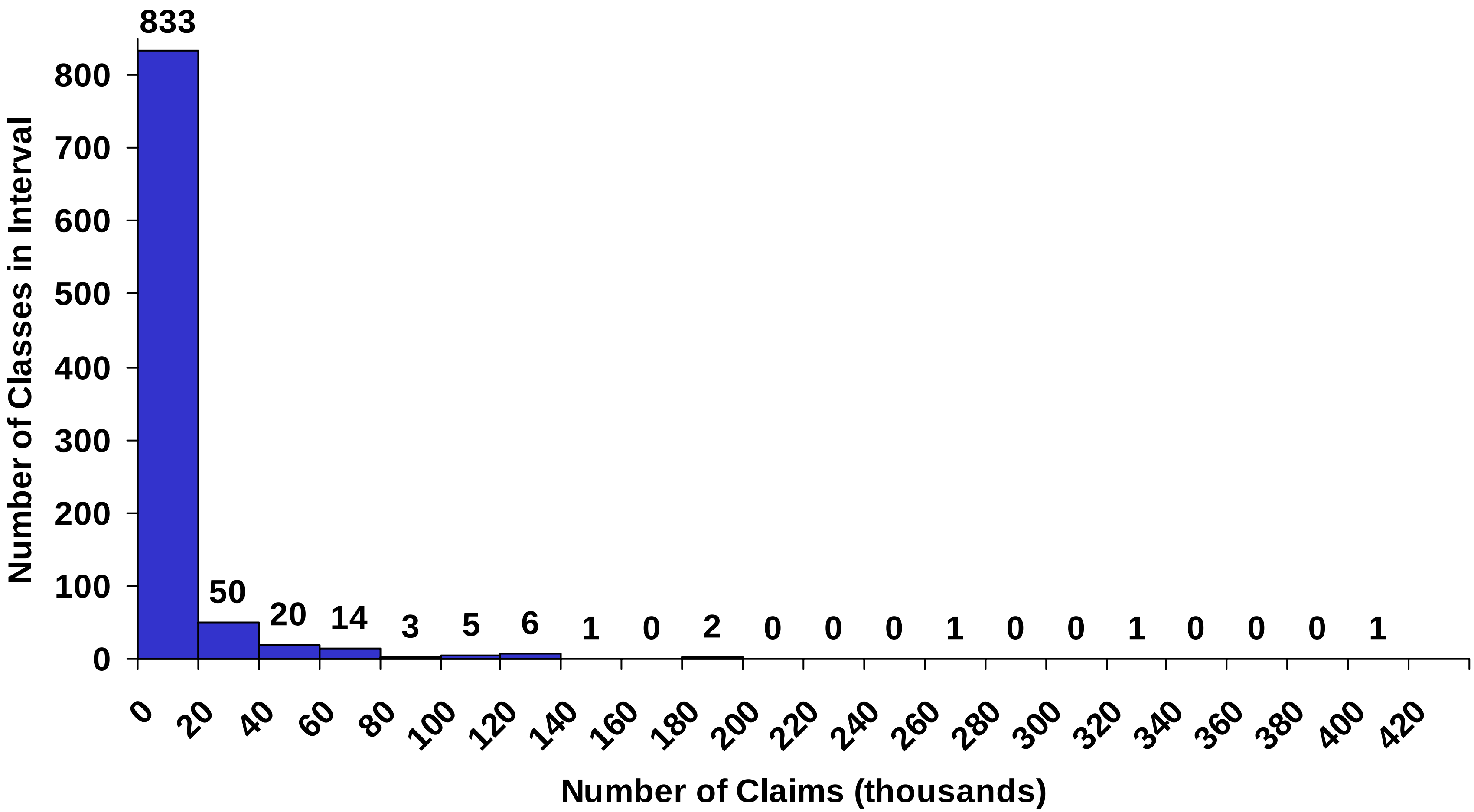
Pseudo-Bühlmann Credibility



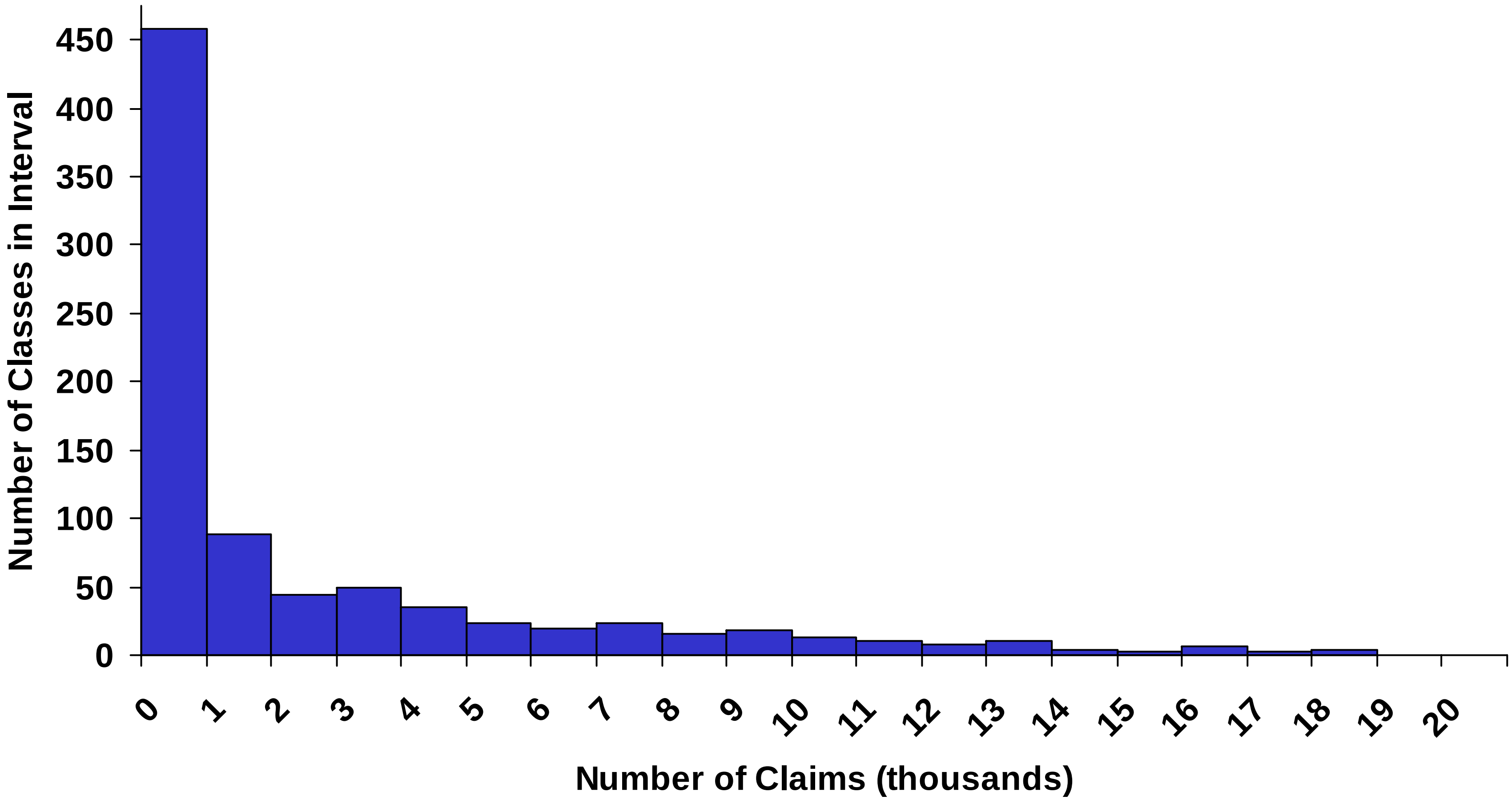
Pseudo-Bühlmann Credibility

Credibility Size Range	Claims per Year	Number of Classes	% Premium
$0\% \le Z < 10\%$	0-237	355	1.2%
$10\% \le Z < 20\%$	238-511	89	1.3%
$20\% \le Z < 30\%$	512-831	61	1.6%
$30\% \le Z < 40\%$	832-1209	56	2.7%
$40\% \le Z < 50\%$	1210-1662	46	2.5%
$50\% \le Z < 60\%$	1663-2216	34	2.5%
$60\% \le Z < 70\%$	2217-2909	46	4.8%
$70\% \le Z < 80\%$	2910-3800	35	4.3%
$80\% \le Z < 90\%$	3801-4987	29	4.0%
$90\% \le Z < 100\%$	4988-6650	18	3.2%
$Z = 100\%$	≥ 6651	**101**	**71.8%**
Total		870	100.0%

Number of Classes by Claim Count



Number of Classes by Claim Count



Number of Classes by Claim Count