

## RMS® CCRA® Training Program Exposure Data Quality Course Exercise

### ANSWER KEY

#### Learning Objectives:

Identify data quality resolution and completeness issues for a commercial California earthquake book of business. Where possible, also identify potentially suspect data values that indicate other data accuracy categories that may be compromised and list those categories.

#### Introduction:

The year is 2018. Your company's California commercial book of business was affected by an M 7.1 earthquake on the Puente Hills Fault in Southern California on September 7, 2017. Your modeled losses based on both event scenario and footprint analyses are lower than actual. RMS has performed a month-long claims analysis for this book of business and summarized their conclusions in a report.

The following files are available to you in order to answer the subsequent questions:

- Claims Analysis Report.pdf – Company A Claims Analysis Report from RMS
- Query Results CAEQ.xls – Excel spreadsheet containing exposure data profiles on 14 separate worksheets

### PART I – Exposure Data Value Profiling

1. Listed below are the three “fitness for purpose” dependencies for any catastrophe exposed data analysis. For each category, use the information in the Excel spreadsheet and the Claims Analysis Report to provide data profiles, relative total insured values (e.g. percentages), and/or limits as metrics. One example is already provided for the peril category.

<b>Peril</b>	<p>100% of the TIV (total insured values) is exposed to the earthquake peril.</p> <p><i>31% of the TIV (total insured values) is exposed to the terrorism peril.</i></p>
<b>Line of Business</b>	<p><i>More than 99% of the TIV is Commercial/Industrial (there is a group of single family residential risks comprising &lt;1% of TIV).</i></p> <p><i>80% of limits are commercial; 20% of limits are industrial.</i></p>
<b>Hazard Region*</b>	<p><i>High Hazard = 81% of TIV (S.F./Bay Area and S. California)</i></p> <p><i>Med. Hazard = 3% of TIV (North Coastal California)</i></p> <p><i>Low Hazard = 15% of TIV (Central Valley/Sierra)</i></p>

\*Hint: Refer to TIV by County tab of spreadsheet provided.

2. Referring to the information in your answer to Question 1, and the data provided in the claims analysis report, would you focus your data quality efforts on just 2 or 3 primary data categories? If so, list which ones, and explain why. If not, explain why.

**Answer:**

*It is always best to get the best data quality possible; however, given time and market constraints, it appears that you can make some high-impact choices on where to focus your data quality efforts.*

- *Given that the majority of the book resides in a high hazard seismic region, the obvious first choice would be to focus on geocoding resolution in those regions. While the geocoding resolution is acceptable overall, there are some high value risks, nearly 73% of the low resolution locations reside in high seismic hazard regions.*
- *Construction class is another area to focus on. Fully 46% of the TIV is classified with unknown construction.*
- *Coding of secondary construction characteristics is another area of concern. It appears that default assumptions may have been made in the data and should be investigated.*

3. For each of the data categories listed below provide one data resolution or completeness issue and one data accuracy issue, and include a metric for each.

<b>Data Characteristics</b>	<b>Data Resolution or Completeness</b>	<b>Data Accuracy</b>
<b>Geocoding</b>	<i>~89% of TIV is geocoded to street (high) resolution.</i>	<i>~50% of low resolution geocoded TIV clearly have missing address information (no street number, address or ZIP Code) or have information coded in the wrong data column.</i>
<b>Construction Scheme</b>	<i>~77% of TIV has construction scheme coded to the ISO Fire scheme.</i>	<i>The ISO Fire scheme is not as appropriate for earthquake or terrorism exposed building construction classification.</i>
<b>Occupancy Class</b>	<i>Occupancy classification is highly resolved with nearly 100% of TIV coded as something other than unknown.</i>	<i>The classification ATC General Industrial would be better classified with an industrial facilities model classification. Single family and multi-family housing listed, but the only lines of business classified are commercial and industrial. Are these correct?</i>
<b>Construction Class</b>	<i>~46% of TIV is classified as unknown construction.</i>	<i>Some general industrial risks are classified as wood frame structures. This is suspect. ISO Class 13 (Mix Constr. - Ordinary EQ Res (5AA)) maps to RMS tilt-up construction. This construction is not usually 6-stories tall and is suspect.</i>

<b>Year Built</b>	<i>Over 71% of TIV has buildings classified as unknown year built with the 01/01/9999 default entry.</i>	<i>48% of TIV has year built ending in either 0 or 5 implying that assumptions may have been made.</i>
<b>Number of Stories</b>	<i>~11% of TIV is unknown.</i>	<i>~75% of portfolio TIV is unknown construction coded as 1-story. One story could be a default assumption (instead of coding as unknown).</i>
<b>Secondary Characteristics</b>	<i>All accounts have secondary characteristics coded, making this a highly resolved data element.</i>	<i>337 locations all have the same six characteristics coded as known. Could this be a default? URM chimney retrofit category may also be defaulted to Yes.</i>
<b>Policy Coding</b>	<i>Coverage values appear to be entered in hundreds of dollars for some accounts, thus decreasing how sharply the data is defined. Policy status is entered.</i>	<i>(\$13,354,000.00) in negative coverage values, potentially reducing TIV by &lt;1%. Sum of "Part of" policy terms are about double the TIV. One cancelled policy is included.</i>
<b>Treaty Coding</b>	<i>All necessary treaty fields entered (including producer ID and Cedant ID).</i>	<i>Inuring priorities are suspect Multi-line accounts with separate policies. It may be better to have separate portfolios for more accurate treaty attachment.</i>

## PART II – Processes Impacting Data Quality

4. Each of the below data accuracy categories were discussed in Unit 3 of the course document, and sample data accuracy questions were provided. Using this information from the course material, provide one example of each category from the Company A portfolio. Use the data accuracy observations in Question 3 to help guide your answers. Finally, rank them in what you believe to be their order of importance. **Note: The ranking is purely subjective and will be part of the discussion during first day of the instructor-led portion of the course.**

Rank *	Data Accuracy Categories	Example
	<b>Bias</b>	<i>All 337 locations have engineered foundation entered. Bias in number and type of secondary modifiers entered.</i>
	<b>Vintage</b>	<i>One policy flagged as cancelled. Data vintage noted in Claims Analysis Report. 24 earthquake policies are shown to have expired by 9/7/2017.</i>
	<b>Validity</b>	<i>Number of buildings always set to 1 - Invalid for aggregate value locations. Potential aggregate locations not flagged as such. No square footage entered for ITV checks. No information discussing if street address is physical location or mailing address.</i>
	<b>Consistency</b>	<i>Inconsistent entry of address information for low resolution locations. Construction scheme is entered as ISO, Fire, and ATC. No one consistent scheme used.</i>
	<b>Interpretability</b>	<i>No user notes.</i>

*\*Note: Rankings are subjective; therefore, no one “right” answer is provided here.*