

Market Conduct Annual Statement

Discussion of Centralized Data Collection

January 25, 2007

Background / Rationale

The new paradigm for market regulation is to move away from comprehensive market conduct examinations as the fundamental – and largely only – tool for data collection and market regulation enforcement – to a system grounded in market analysis employing a variety of market regulation tools focused on particular issues and companies stemming guided by market analysis. As market analysis is the foundation of more focused, efficient and effective market regulation, the availability of relevant data and information is the prerequisite for market analysis.

The pilot program for market analysis consisted of more systematic use of some existing data sources – consumer complaints, financial information – and some data specific to market analysis needs – the market conduct annual statement. The more systematic and intensive use of existing data sources is described in the Market Analysis Handbook and the use of MCAS data has been demonstrated through the pilot program which produced some simple analyses to identify outlier companies for further questioning.

While there has been initial market analysis activity that has helped focus some market regulation activity, the limited data available has limited the scope of market analysis and, subsequently, the ability to move more towards targeted market regulation activities and away from traditional comprehensive examinations. In addition, the requirement that each state collect the same set of MCAS data has prevented the majority of states, with limited resources, from participating. Other states have not participated because the MCAS provides less data than they otherwise collect for market regulation purposes.

The obvious conclusion is that improvements in market analysis capacity and capabilities require an increase in the amount of data collected and centralized data collection to allow more states to participate.

It is useful to contrast the data available and collection methods used for financial analysis and rate analysis to those for market analysis. Financial analysts have literally thousands of data elements in the statutory annual statements and quarterly reports, some of which is countrywide data and some of which is state-specific. The data are centrally collected by the NAIC and run through a variety of standard analytics to help with financial analysis. Rate analysts have access to statistical data which includes detailed information on premiums, exposures and claims detailed geographic basis and is reported and collected centrally by designated statistical agents mostly on a countrywide basis. In some states, there is state-specific reporting. But, again, the number of data elements runs into the hundreds.

Data for market analysis currently consists of complaint data and about 20 data elements collected individually by states. To put market analysis on a par with financial and rate analysis – the amount of data available for analysis must be dramatically increased and some form of centralized data collection must be developed to enable all states to participate and to provide for efficient reporting by regulated entities.

Data Elements vs. Data Collection

Conceptually, the contents of the MCAS, which I'll use as shorthand for market analysis data elements to be collected, is separate from the mechanism for collecting the data from regulated entities and compiling the data for analysis. The nexus of the two is where the amount and type of data required in the MCAS determines the resource and skill requirements of the collector. While small and medium-sized states may be able to handle small data calls like the current MCAS, the vast majority of individual states could not manage data collection on the scale of a statutory annual statement.

Consequently, the two issues should be separated and handled with procedures specific to each. For MCAS contents, there should be a procedure in place to propose, review and expose for comment and adopt the data elements to be collected and the instructions for reporting the data elements. Stated differently, there should be a procedure for adopting or changing a MCAS blank or data dictionary. An approach modeled after the Blanks TF is a starting point for discussion.

Once the MCAS blank is developed, the second procedure is employed for centralized collection and compilation. The approach for this is set out the scope of data collection and the performance requirement in a request for proposal that allows organizations to compete for the data collection assignment. The 12/1/04 CEJ proposal sets out the basic framework. Attached are some documents to provide a flavor of what such an RFP might look like.

Summary vs. Detailed (Individual Application, Policy, Claim) Data

A threshold issue is what type and detail of data to collect and the decision should be made right up front that detailed data – meaning individual application, policy and claim data – is required and that summarized data is inadequate for the purposes of market analysis.

Summary data refers to compilations of experience according to pre-selected categories, as in the current MCAS. The analysis of summary data is limited to questions specifically related to the summary categories used. There is no ability to analyze categories of data simultaneously or to examine sub-groupings within a category. Stated differently, there no ability to employ modern data mining analytics on summary data. In addition, changes to summary data reporting requirements are expensive because they involve reprogramming the software to produce the summary reports and may involve the collection of additional data elements that were not necessary to collect with the prior

data elements. Summary data collection also limits the ability to perform data quality review on the data because data errors may be masked in compilations. Looking at the MCAS, for example, there is no ability to examine claim outcomes other than the time frames specified so there is no ability to examine what happens at the edges of the categories. This leads to another problem – company behavior conditioned by the types of analyses produced by the summary categories.

The alternative is detailed data collection at the individual application, policy and claim level. For item, there are a series of characteristics. Attached is the Texas Private Passenger Auto Stat Plan which can be used to illustrate both summary and detailed reporting.

The benefits of detailed reporting are many. Foremost is the ability to employ multivariate, data mining analytics – data analysis that looks at multiple data items simultaneously. This is precisely the approach taken by insurers for developing sales, underwriting, rating, risk classification and claims settlement strategies – data mining techniques applied to massive detailed databases. Regulators should obviously be employing the same advanced techniques to focus market regulation activities in the same way that insurers use the techniques to focus their sales, pricing and claims settlement activities. Data mining of detailed data would enable regulators to identify issues of concern before they become headlines.

Another critical feature of detailed data is the ability to answer questions that have not been previously imagined. With summary data, the summary categories are established to answer specific questions. When a new question arises, the summary data are useless. With detailed data, a new question can generally be answered by a different analysis of existing data. Even in the situation where additional data are required, the new reporting consists of adding data elements to an existing data framework, as opposed to creating entirely new summary categories. This means greater efficiency and cost effectiveness over time.

Another benefit of detailed reporting is the elimination of special data calls for information not contained in summary reporting categories – an efficiency and effectiveness tool. The effectiveness comes into play because data routinely reported pursuant to data dictionary or stat plan are much more reliable than data reported on a one-time basis.

Another benefit of detailed reporting is the ability to employ more detailed data quality review to ensure that data reported accurately reflect actual experience.

Finally, there is a point where the detail desired of summary reporting requires more records than necessary for detailed reporting. Consider a company with 200,000 auto insurance policies, which would be 200,000 records in a detailed reporting framework. For summary reporting, let's assume we want the experience broken out by coverage, ZIP Code, driver classification, and just other rating factors, each with 8 categories. For

example, a rating category based on miles driven or credit score with 4 surcharges and 4 discounts. If we assume six major coverages, 200 zip codes, 50 driver classifications (fewer than most companies use) and just the three additional rating factors, we get $6 \times 200 \times 8 \times 8 \times 8$ or 614,400 records. Once you get to any level of summary detail, detailed reporting is more efficient.

Utilization of Existing Databases

There are detailed databases currently in use which are the result of centralized collection of detailed data. These include:

- Statistical data for property casualty personal lines of insurance collected by advisory organizations / statistical agents. The Insurance Services Office collects detailed data – more detailed than the individual policy or claim – for insurers representing about 25% to 40% of the market, depending on the line of insurance and state. PCI and NISS are also statistical agents for personal lines, but use summary reporting. The Texas Auto Stat Plan gives a flavor of what the ISO stat plan is like. ISO utilizes a countrywide statistical plan with some state exceptions. Of course the ISO stat plan is tied into ISO's forms and loss costs. Nevertheless, ISO could provide policy and claim detail data for certain insurers and while those data are currently used to develop loss costs and rate filings, the data could be mined for market analysis purposes. The PCI and NISS summary data, while more detailed than the MCAS, are relatively limited. While the ISO database has detailed premium, exposure and claim information, there is no application data.
- Claims Databases for property casualty fraud detection. ISO reports that its Claimsearch database now contains records of 500 million claims covering 95% of insurers. If fraud investigators – as well as insurers – have access to these data, so should market analysts. A review of the Claimsearch data layout would determine if the data were suitable for market analysis.
- CLUE and A-Plus claims databases used for underwriting. While these databases have claim level detail, they are probably limited in the amount of detail included and, consequently, of limited use for market analysis. A review of the data elements included in the databases would determine the usefulness of the data for market analysis. Access to these data – if collected by a credit bureau or group like Choice Point which is not licensed as an advisory organization – may be more problematic than access to data from, say, ISO.
- On the life and health side, the Medical Insurance Board, MIB collects application and claim data, providing the same types of services to insurers for underwriting and fraud detection as ISO and ChoicePoint and others on the property casualty side. Finding out the detail on their database elements is a starting point for considering a life / health MCAS data set.

To the extent possible, the MCAS should rely on existing data collection, if relevant and accessible and sufficiently comprehensive. The MCAS could then supplement other data collection with additional complementary data elements. If the existing sources are not suitable or available, then a comprehensive MCAS can be built from the ground up. Even under the best case scenarios, we know there are no existing databases for property casualty application data. Even if the proposal is to have the MCAS stand alone and be comprehensive, regulators should be able to explain why existing data collection systems and products were not used and MCAS data reporting for the same data items should be made as consistent as possible with reporting of those data items to the other systems to ease the burden on reporting companies.