Bering Sea/Aleutian Island Crab Economic Data Report
Center for Independent Experts Review Meeting

August 23-25, 2011

Panel Chair's Summary Report

November 2011

Submitted by:
Christopher M. Anderson, Ph.D.
Associate Professor
Environmental & Natural Resource Economics
University of Rhode Island

Executive Summary

The rationalization of the BSAI crab fisheries transformed the economic structure of the industry, allocating quota-based privileges both to harvesters and to processors. Economic objectives of the program were to improve safety and increase rents by reducing derby fishing, and to assure that there was not a significant and disruptive shift in the distribution of rents between harvesters and processors, or among communities linked to Bering Sea fisheries. To not only track the effectiveness of the novel measures implemented in pursuit of these goals but also facilitate outcome transparency and economic analysis as the program evolves, the Council commissioned the development of an Economic Data Report (crab EDR) that collects from harvesters and processors the most comprehensive set of variable cost data currently tracked in any US fishery.

The Center for Independent Experts Panel (the Panel) was convened by NOAA in support of the Council's evaluation of the crab EDR process and the data it had generated to date. The Panel reviewed the EDR questionnaires, Council motions and staff reports, meeting proceedings, NOAA memos and reports, and scholarly articles prior to a three-day public meeting on the crab EDR. The materials documented the development of the EDR, and included extensive supporting information on the processes undertaken to understand the quality of the data being generated. This included results of three metadata processes: a records audit (both random and outlier) to ensure reported values are supported; a statistical sensitivity process to assess how aggregate data quality measures have changed over time; and a consultative process with the industry (PNCIAC) to understand difficulties in reporting and credibility gaps in the data.

While many EDR variables were reliable and credible, metadata reflected that many others were not. In some cases, EDR questions were not interpreted consistently by respondents. Often, consistent or accurate reporting was hampered by requesting highly disaggregated data (e.g., by fishery, species or quota type) when it is not naturally generated or tracked at that level of detail, leading respondents to expend considerable effort to apportion known values across strata in ultimately inconsistent ways. Respondents also believed some EDR data was available from other sources. These factors contribute to a reporting burden far in excess of the expectations that were initially established, and what all parties feel is reasonable.

The Panel associated these challenges with the pilot nature of the crab EDR. Fisheries economics lacks a commonly accepted process that dictates the flow of information and guidance among scientific and research staff and contractors, the SSC and other advisory bodies, and the Council itself. This is in contrast to the more developed and formalized process for stock assessment science across regional councils, including codification of specific biological reference metrics in the Magnuson-Stevens Act and the transparency and peer review that is provided by Plan Team discussions, in addition to expertise on the SSC. This process assures quality while limiting political influence on science, and assures key baseline data are available through successive management changes.

There was a consensus among the Panel that the data that has been collected through the EDR is suitable for use by analysts sophisticated enough to manage limitations inherent in any field data: the EDR is being developed through a process consistent with the best available scientific practices, and represents the best available scientific information on the economics of the BSAI crab fisheries.

There was also a strong consensus on the Panel that some elements of the EDR are not yet achieving their intent, and that information contained in the metadata can be leveraged to achieve measured improvements in data quality that enhance the value of the resulting analysis to the Council in evaluating progress against the rationalization's stated objectives. To advance the goal of improving data quality quickly and efficiently, the Panel offers general and specific recommendations for each stage of the EDR process:

- Additional efforts to reduce reporting burden by aligning reporting variables and levels in EDRs with those commonly tracked by fishery participants, as well as eliminating duplicative reporting by linking or pre-populating the EDR based on other reporting databases. (However, this is not to suggest that simply because a variable is currently reported elsewhere that the incumbent form should be the surviving reporting venue.)
- Institute a formal role in the process for economic scientists who will analyze the data to establish scientifically based standards for the content, precision and accuracy necessary to conduct the analysis desired by the Council.
- Conducting more extensive interactive or cognitive style interviews with those who complete the EDRs at each revision to ensure question statements lead to reporting consistent with designer intent. Survey experts may be needed to advise in the design and interpretation of these processes.
- Reducing redundancy in reporting across fisheries by integrating with EDR collection in other fisheries.
- Increasing consistency and transparency of necessarily apportioned variables by collecting at levels of aggregation where minimal respondent apportionment is necessary, allowing analysts to do the apportionment, or providing industry input into methods of apportionment that are sensible.
- Further development of the infrastructure of the EDR database to integrate entry error checking and to allow analysts more transparent access to data stored in different tables.
- Explore whether communities and quota holders may be better survey frames for variables in which they have more direct knowledge and interest in accurate measurement.

As these recommendations are implemented, the Panel foresees value in the continued application and evolution of the metadata generation process. As the EDR itself improves and stabilizes, some current metadata components may no longer be necessary, while needs may emerge for components targeted at providing additional key quality information to support analysts.

Background and Scope

The rationalization of the BSAI crab fisheries transformed the economic structure of the industry, allocating quota-based privileges both to harvesters and to processors. Economic objectives of the program were to improve safety and increase rents by reducing derby fishing, and to assure that there was not a significant and disruptive shift in the distribution of rents between harvesters and processors, or among communities linked to Bering Sea fisheries. To not only track the effectiveness of the novel measures implemented in pursuit of these goals but also facilitate outcome transparency and economic analysis as the program evolves, the Council commissioned the development of an Economic Data Report (crab EDR) that collects from harvesters and processors the most comprehensive set of variable cost data currently tracked in any US fishery.

The Center for Independent Experts Panel (the Panel) was convened by NOAA in support of the Council's evaluation of the crab EDR process and the data it had generated to date. The Panel reviewed the EDR questionnaires, Council motions and staff reports, meeting proceedings, NOAA memos and reports, and scholarly articles prior to a three-day public meeting on the crab EDR. The materials documented the development of the EDR, and included extensive supporting information on the processes undertaken to understand the quality of the data being generated. This included results of three metadata processes: a records audit (both random and outlier) to ensure reported values are supported; a statistical sensitivity process to assess how aggregate data quality measures have changed over time; and a consultative process with the industry (PNCIAC) to understand difficulties in reporting and credibility gaps in the data.

To the knowledge of the Panel and participants, this is only the second CIE review panel meeting convened to deal with a social science process; the prior case was a much different calibration exercise. That this review was held reflects both the significance of this particular task within NOAA Fisheries' social science research program, and the difficulty of the challenges that are being addressed. The Panel found this task to be broader than simply evaluating the development of the BSAI Crab EDR against a set of well-established standards, but rather about evaluating processes for developing data collection and assessment procedures for comprehensive fishery economic data. The individual panelists each prepared and offered comments based on the standards in their individual fields of expertise. This report offers some discussion that synthesizes context, issues, conclusions and recommendations across the panelists' expertise, and without the strictures of the enumerated terms of reference.

Following extensive materials review and three days of meetings, the overall picture that emerged is one of a data collection enterprise that is still evolving toward meeting all of its goals. Despite not being finalized, there is tremendous value in many of the variables that have been collected, but others have been problematic. However, there is underway a process that the Panel found to be uniquely comprehensive in its ability to identify variables with issues on several dimensions of quality, including accuracy, precision, unbiasedness, credibility, and minimal reporting errors. It is essential that this enterprise be pushed to develop quickly, so that there is data of sufficient quantity and quality to support tracking and analysis of progress toward the current or future goals of the

Council and national fisheries policy, and so that the lessons from this effort can be translated to other US fisheries.

Structure of the Meeting

The meeting took place over three days at the Alaska Fisheries Science Center's facility at Sand Point in Seattle. It was attended by the panel; the NOAA staff involved in designing the EDR and managing its implementation; the contractors performing EDR collection and database administration; representatives from the Northwest Science Center who are developing a similar economic data instrument; accountants for vessels (who actually complete vessel EDRs); accountants for processors (who actually complete processor EDRs); and representatives of crab industry harvester and processor organizations.

The first day of the meeting was dominated by a series of presentations by NOAA staff and contractors. The presentations gave an overview of the EDR, its goals, its development process, and the process of developing the metadata on the quality of individual variables. The presentations served as a basis for a discussion, including extensive questions from the panel and questions, comments and reactions from the audience. The second day focused on a discussion of the issues that were raised during the first day, driven by the Panel but with extensive involvement of both NOAA staff and the audience. On the third day the Panel met in executive session, arrived at a consensus understanding of the process and issues, and an outline of the issues related to each element of the panelists' Terms of Reference.

Scope of Evaluation

Throughout the discussions, the Panel and participants were conscious that the BSAI crab EDR is an important pilot for nascent economic data collection efforts throughout NOAA, and thus it was to be viewed as a product under development at three levels. On one level, the Panel was to assess the data products that have been produced by the EDR, the years of vessel and processor data collected to date. On a second level, those data and the corresponding metadata are a product of a data collection instrument and data handling procedures. This includes the procedures for administering the survey, handling and verifying the data, and generating metadata for verification and assessment. The third level arises from lacking a standard economic data process for fishery management, and the need to identify how best to coordinate the Council, NOAA, outside social scientists (and expert panels), and industry to develop an instrument and procedures for administering it. Thus, identifying best practices for working among groups, and integrating industry knowledge and scientific expertise in recognition of the interests of each party, also falls within the Terms of Reference. At each level, improvements are important as the agency-wide NOAA social science data initiative expands and lessons must be drawn to collect social science data more effectively and efficiency.

While not structured as such, much of the discussion at the meeting, and assessments and recommendations in the reports can be classified as affecting process elements in the following taxonomy:

1. The process used to develop and manage the survey
    a. The process used to identify the specific measures collected to meet the goals set by the Council
    b. The process for developing, phrasing and decomposing measures on the survey
    c. The process of working with industry to assure survey responses reflect the intended interpretation of the question
    d. The design of the database used to store the data
2. The processes for assessing the quality of the current and retrospective data collected to date
3. The quality and sufficiency of the collected data for analysis
    a. The implications of the audit process data quality metadata
    b. The implications of the regression-based data quality metadata
    c. The implications of the PNCIAC data quality metadata
    d. The set of human and computer procedures used to enter the data and minimize entry error
4. The process of engaging with industry, economic analysts and the Council to ensure EDR data support assessment of program objectives
    a. Advice on the utility of alternative quality metadata metrics
    b. Application of the quality metadata
    c. Guidance on proposals for responding to feedback on data quality and feasibility

While this list is lengthy, the emphasis reflects the focus on process represented in the Panel's Terms of Reference. The Panel was not charged with assessing any of the scientific or policy content of the published articles (e.g., Abbott et al. 2010) or reports (e.g., NPFMC 2010) that utilize the data; evaluating the quality or reliability of particular variables in the EDR; making recommendations to the Council on variables that should or should not be included; or evaluating the effectiveness of the crab rationalization program.

Not every Panelist report covers each Term of Reference or recommendation at each level, nor are they always explicit about the level at which they are drawing conclusions or making recommendations.

Specific Concerns and Issues to be Addressed

Best Available Science determination for collected data

Term of Reference 8 requests an "Explicit Determination as to whether this NMFS project presented the best available science." Even more than the other Terms of Reference, this requires consideration at each level of multi-level process underway.

There was a consensus among the Panel that the data that has been collected through the EDR is suitable for use by analysts sophisticated enough to manage limitations inherent in any field data: the EDR is being developed through a process consistent with the best available scientific practices, and represents the best available scientific information on the economics of the BSAI crab fisheries.

In the Panel's assessment of the data collected to date, in particular, the "available" criterion receives significant emphasis, and the best available science conclusion must be put in context. The data arising from this project are not only the most comprehensive and detailed for the BSAI crab rationalization, but surpass that for any other US fishery of which the Panel's fishery economists are aware. However, the general state of economic data collection in fisheries is so poor that this is not a particularly high standard (especially for processors), and it is acknowledged by all parties that sections of this data are far from ideal: there are broad categories of variables that are challenging to interpret, costly to obtain, or are less reliable than all parties desire. Therefore, despite setting the standard for firm-level fishery economic data, all parties should aspire to improve the data quality and reliability, and make the best better.

The quality of the collected data reflects the survey process that generated it. The current survey, and the process for developing it, has not at every stage adopted all best practices in establishment survey development. This is not surprising, given the timeline for development and the extent to which this project advances the state-of-the-art in fisheries firm-level primary data collection. While there were discussions with processors and harvesters during the development of the instrument, and early versions were discussed with a few representatives, the number and depth of these was insufficient. As a result of not conducting cognitive interviews at each major revision, the instrument that was used for current and retrospective data collection did not anticipate some likely foreseeable problems with certain measures; these problems are emerging now as the collected variables are of "low quality" by one or more metadata measures. However, this is a product of the constraints built into the EDR development process, where annual distribution dates impose strict deadlines and an expectation that major revisions to be EDR be reviewed by the Council, a time consuming process that, while intended provide feedback and oversight, may prevent adapting the EDR to scientific quality information as it becomes available.

While the survey instrument has some problems the panel believes could have been anticipated and partially resolved before the initial rounds of the data census, the process for identifying those problems was praised by all members of the panel as being uniquely comprehensive in any application in any field of which the panelists are aware; this is not "best available", but rather state-of-the-art science. The strength of this process is the multiple notions of quality that are evaluated for each variable--from a range of perspectives by a range of users--allowing a clear diagnosis of why each variable may rank higher or lower on key quality dimensions. Understanding why data is unreliable facilitates improvement, in the form of caveats to the users regarding the types of analyses for which the data are and are not reliable, redesign of the instrument, or reconceptualization of how to measure progress on certain program goals.

Using multiple processes for generating and applying quality metadata

To assess the quality and utility of the data collected to date, metadata has been generated through three distinct processes. Each process focuses on a different aspect of data quality. The Panelists' heterogeneous backgrounds imply that data quality means different things to different evaluators. The individual reports make explicit each member's data quality framework; the Panel views these differences as beneficial to the evaluation and did not feel the need to arrive at a consensus framework. From all perspectives, the Panel believes the multi-pronged approach NOAA and the Council has taken to analyze quality in the BSAI crab EDR program is extraordinary. At the current stage in the EDR development process, it plays an important role in facilitating refinement of the data collection procedures and instrument; once the instrument is refined, it is likely that some elements of the metadata will no longer be necessary in subsequent seasons of the EDR.

While each panelist evaluated data quality within the frameworks of their fields of expertise, much of the discussion can be organized around the following dimensions of quality: accuracy, precision, unbiasedness, low variance, low error rates, and credibility. These warrant careful discrimination, as several points in the Panel meeting arose from confounding these different notions.

For each appropriately chosen measure (i.e. variable), there is a "true value" for each respondent i, μ_i, which, for example, may represent the actual pounds of crab caught or dollars spent on fuel for crab fishing of vessel i. While the EDR's goal is to learn μ_i, in practice this is unknowable due to limits on measurement, supporting records, interpretational differences, and other factors. Thus we instead obtain a reported measure of μ_i, which will be denoted m_i. There are two corresponding notions of *variation* in a population of size N. First, true values μ_i will vary from respondent to respondent, with a population standard deviation, $\sigma = \sqrt{\frac{1}{N}\sum_i (\mu_i - \bar{\mu})^2}$, where $\bar{\mu} = \frac{1}{N}\sum_i \mu_i$ is the average of the true measures across the population of size N. Note that σ is simply a measure of the true variation in the population: we would not expect it to approach zero, since the fishery is diverse. Second, there is the standard deviation of the measures, $s = \sqrt{\frac{1}{N}\sum_i (m_i - \bar{m})^2}$. Ideally, s approaches σ and reflects that m_i captures μ_i. This measure should be high where the fleet is heterogeneous, as s is not a direct measure of data quality in the absence of information about why it is high.

However, s does play an important role in determining the utility of the data. The statistical confidence we have in summary statistics of m, like its average $\bar{m}$, is proportional to s ($\frac{s}{\sqrt{N}}$ in the case of $\bar{m}$). This measure captures all sources of variation represented in the data collection process. In conventional treatments, the focus is on sampling variation: if a large number of different samples of size N were drawn from a large population, the different samples would lead to different averages. The standard error captures the variance of the distribution of those averages. High standard errors are

obviously related to the heterogeneity of the underlying data, and within the statistical inference framework they present a higher standard for saying two measures are statistically different (e.g., that there is a statistically significant difference between two years, or between two subgroups). Being able to discriminate such differences with greater statistical confidence is one important dimension of data quality. However, due to the full census collection of EDR data, sampling error is not present, so there is no ability to increase sample size to decrease standard errors.

However, s can diverge from σ if there are other, non-sampling sources of errors in the reporting process that lead to m_i differing from μ_i, reflecting low data quality. These sources could include variation in interpretation across reporters or variation in record keeping. From an overall design perspective, this variation is not a problem in itself, but rather may be reflection of other underlying problems with the data that may or may not be problematic for certain analytical applications. The chief underlying problems are inaccuracy and biasedness.

Accuracy captures the extent to which a measure deviates from its true value. An (impossibly) accurate measure would have $m_i = \mu_i$, for all respondents, and $\overline{m} = \overline{\mu}$ in the case of aggregate measures, for which there can be separate accuracy metrics. Inaccuracy arises in two important ways. First, m_i can be a noisy measure of μ_i. In this case, if N is small, then $\overline{m}$ can be far from $\overline{\mu}$ through random errors, but with larger N random noise that averages zero may not cause $\overline{m}$ to differ from $\overline{\mu}$. Second, inaccuracy can be caused by bias. *Bias* occurs when m_i is systematically higher or lower than μ_i, for the whole population, or for a subset of the population. Bias could arise because a question is structured in a way that leads to systematic errors, or respondents strategically report higher or lower values for a variable. This is problematic because in many circumstances, it could lead to incorrect conclusions about policy effects. It is important to note, however, that differences in measures with common bias structures largely control for the bias because it is present both before and after the policy change.

Precision is a dimension of quality that is often associated with accuracy, but is importantly different. Precision captures the smallest unit at which the data recording process can distinguish between different levels of measure m_i. In the EDR, money measures are reported to the dollar and landings to the pound; they could also be reported to the thousand-dollar increment or the ton. The value of precision is the ability to detect small changes in the data: one would not be able to tell if one trip had 10 pounds more landings than another if landings were reported to the precision of tons.

Some of the quality issues raised with EDR data reflect concerns about illusory precision. Illusory precision occurs when data are reported in finer units than the measurement process that generates the data. For example, although landings weights are reported in pounds, crab weights might not be actually be correct to the pound: different loads might have different amounts of water, debris or barnacles in them when weighed, or the industrial scales used to measure weight might only be sensitive within several pounds. Thus, the weight data might appear to be low quality because it is not accurate at the precision at which it is reported. However, that does not mean it is not accurate at some

lower level of precision (e.g., hundreds or thousands of pounds) that has policy relevance. Precision can generally be increased with additional data collection effort. Thus, the instrument design challenge is reconciling the costs of collecting high precision data with the value of, and ability to, make fine distinctions, the latter also being influenced by the variance of the measure.

Low *error rate* reflects the frequency with which $m_i = \mu_i$ (or m_i is within a tolerance of μ_i). It is intended to capture the extent to which the data collection instrument, and the reporting and recording process, approach the best possible levels of statistical performance. In this context, an error is a reason that m_i might deviate from μ_i that might be thought to be within the respondent's or the agency's control. True errors might arise in several ways. There might be data entry errors (e.g., typos, illegible entries, etc.) or there might be misunderstandings about what is being measured. Problems like typos increase variance, but are purely random ex ante, and therefore do not introduce bias. Misunderstandings that are shared among groups of people can introduce bias as they affect several respondents in the sample in the same way. Finally, errors might arise through strategic misreporting, if a respondent wishes to conceal information and is willing to violate the regulation in order to do so (no claim was presented that this has happened).

Credibility reflects whether people feel the information contained in the data performs sufficiently well on the properties above to be useful for the types of analysis targeted. This can integrate information on the statistical properties of the data, but also knowledge of the reporting process and an understanding of whether the specific metric being collected is appropriate for the goal being monitored.

The Panel particularly appreciated that the multiple metadata approaches cover all of these dimensions of quality:

- The *enterprise response audit* reflects the extent to which the reported data can be supported by records and a defensible interpretation of the survey question, within the bounds of general accounting practices. An audit correction can reflect a calculation error, a record-keeping error, a misinterpretation of the question, a conscious evasion (no CIE review participant claimed this has occurred), or simply that reported data is not well substantiated by records because it is a guess, estimate or apportionment. This provides key information on the extent to which variation is due to certain forms of reporting error. Because audits can be triggered through an outlier analysis, it helps to minimize variability (s) in the data. Importantly, when high rates of misunderstanding of, or varying interpretations of, the question are identified, this means the phrasing or respondent training is insufficient, and does not discount the value of the variable once corrected.

- The *regression process* leverages the corrections reported through the audit process to determine the effect of the errors discovered on the industry-wide

pictures of the data: it fits a linear regression model to the post-audit data with a constant and the pre-audit data. If changes were statistically or economically small, then analysts would be assured that auditing did not result in meaningful changes in the data, and that aggregate measures were accurate. One advantage of this approach is that it puts each m_i in the context of all observations, the level at which most analysis is carried out. The incidence of these errors should fall over time as survey support materials improve, respondents learn through doing, and through corrections made after audits.

- The *PNCIAC process* asked the Pacific Northwest Crab Industry Advisory Committee to evaluate the quality of each variable. Quality grades of A, B and C were assigned to each variable. High scores on this metric reflect that the members of the industry committee have confidence in the data and find it to be credible. Thus, this metric is able to pick up the respondents' perceptions of the quality of the records used to answer the question, and their confidence in their interpretation of the question. In ideal application, this process would express which statistical dimensions are unsatisfactory and lead to low quality rankings, a critical starting point for improvements. As a formalized process, this is an innovative element in establishment surveys.

As an element of the process for developing the crab EDR--and fishery EDRs more generally--the value in these multiple approaches is *not* primarily in identifying data elements that are "low quality" on at least one dimension for exclusion from analysis, or elimination from the EDR. Rather, these metadata complement one-another to support precise identification of problems in the survey instrument, and *how they arise*. For example, where higher than expected population variance s might reveal high levels of misreporting or misinterpretation of a variable, the audits can identify if misreporting contributed to the problem, and the PNCIAC process can reveal whether respondents had difficulty consistently interpreting the question. Conversely, if the PNCIAC process reveals credibility concerns about a variable, the other processes can gauge whether those concerns are likely to affect conclusions once the data are analyzed. When the strengths and weaknesses of each approach are recognized, and the approaches used in balance, a comprehensive understanding of where difficulties arise can support refinement of the survey process so that future iterations of the EDR yield more satisfactory responses.

Measures that are currently assessed as high quality by all three processes are clearly suitable for analysis. However, a challenge remains in that there are current analytical needs that rely on measures that are not high quality on one or more dimension. In this case, it is important to first consider the extent to which the problem leads to misleading conclusions in the desired analysis: not all quality problems affect all forms of analysis. In particular, much of the intended analysis relies on comparing operations year-over-year, in which consistent reporting biases may allow analysts to accurately identify changes; other analysis relies on aggregate trends that may smooth interpretational variation across respondents; still other analysis is interested primarily in large changes in conditions in the fishery, and require only moderate precision data.

In addition to the strengths of each metadata process, the Panel reviewed the shortcomings and caveats of each component, suggesting care must be taken in the interpretation of the metadata, as well as the data itself. The Panel's primary concern with the enterprise audit process was that it did not initially do enough to support respondents in developing consistent question interpretations and response practices. That is, in the first rounds of audits, the auditor did not seek to understand and enforce the intended question interpretation, and provided minimal feedback. While audit coding has since improved and correction letters are now distributed, impressions of the process based on initial practices seem to persist, supported by current practices like not distributing confirmation letters to audited participants with no corrections. The Panel also notes that auditors allow differences in apportionment methods to persist with the standards of general accounting practices. Since this is not an enforcement audit, it may be possible to introduce a teaching component into the audit process (though it compromises the randomness of the pre-audit and post-audit comparison that supports the regression analysis).

The regression process does a good job of linking the effects of the audit to the changes in aggregate values within the same year, but the Panel identified a few interpretational issues. In the linear model, outliers (before or after the audit) have disproportionate influence on the estimates of slope and intercept; corrections often involved moving an outlier to a central value, or a central value to an outlier, meaning conclusions of bias or unreliability we driven primarily by outliers. The term "reliability ratio" seems to overstate the information contained in this statistic, given that only some sources of error are reflected. Further, the variable of policy interest is often an aggregate summary statistic, such as $\overline{m}$ or quantiles of m_i, and thus the most important reliability question is whether it the audit changed inferences based on the pre-audit value. The slope metric in the regression looks at the relationship between pre- and post-audit values of m_i as m increases, which is not clearly related to $\overline{m}$ (as $\overline{m}$ is invariant to where in the range of m an error falls); ascribing functional or interpretational meaning to this implies that there would be some systematic misreport occurring by firms with higher values of μ_i. Thus, whether this value-linked bias is meaningful depends on how the data is being generated and used, and has not been fully justified for this metric.

The Panel also expressed that similar regression methods could be integrated into the data quality process by identifying other relationships one would expect to be stable, in addition to pre- and post-audit values of the same variable in the same year. For example, within respondent (or respondent's accountant) models could assess consistency of apportioned variables year-over-year, where one would expect apportionment methods to be consistent within-respondent, even if not across respondents. For certain types of trend and impact analysis, within-respondent consistency might be considerably more important. For example, omitted variables, variables capturing small components of costs, or variables that are consistently biased in their estimates will affect different years in the same way, so that establishing trends is possible even if they are not perfectly precise or accurate. There might also be fairly tight bivariate or multivariate relationships (e.g., crew food expense per day to number of crew) that could be studied for validation

purposes, either as field value checks in an online database platform, for identifying targets of outlier audits, or for calculating reliability statistics.

The Panel found the data from the PNCIAC process to be both interestingly and importantly useful, and challenging to reconcile with the other metadata. While it provides insight into the process that generates the data otherwise only possible through cognitive interviews, it is also the perspective of one party that has preferences and incentives that may not be consistent with all of the objectives the Council has established for the data analysis. While the Panel felt industry representatives at the CIE Review meeting were invested in the process, it is important to acknowledge that the EDR requests information that is, in most industries--and in some parts of the businesses of the same firms being surveyed--proprietary. While industry clearly acknowledges the greater responsibility that comes from profiting from the crab resource, they also know that the division of rents in the fishery was a hard-fought political battle during rationalization, and they may not perceive it to be in their interest to have the Council and the public know the division of rents that has resulted (or to stipulate as correct or legitimate any estimates that are presented).

Thus, while most of the critiques of the EDR arising from the PNCIAC process are credible to the Panel and help to advance the process, others reveal an interest in the type of analysis that can ultimately be conducted with the data. One example is repeated insistence that the Council's stated goal of estimating quasi-rents is not attainable and should be abandoned. While the Panel agrees that it is unlikely that the bottom lines of each firm's profit-and-loss statement will be able to be recovered, the Panel believes that this is not necessary to fulfill the Council's intent to monitor the program's attainment of its goals. A second is the argument that fuel or quota expenses cannot be associated with individual fisheries consistently enough for policy purposes, while the Panel understands that very similar apportionment calculations seem to be done in order to determine crew shares on a per-trip basis. Finally, the Panel touched on several variables where fairly straightforward modifications of industry recordkeeping practices could significantly reduce reporting burden, but these practices have not been widely adopted, instead leaving accountants to perform calculations from primary records and inflating reporting time.

While challenges clearly exist to getting all data in the EDR, and some variables may indeed not be worth their collection effort, the Panel did not encounter any circumstances where the Council should consider abandoning collecting information essential to tracking progress against the rationalization program's primary goals.

Clarifying how the data will be used

The Panel encountered uncertainty, and concern, among participants about how the data will be used, and believes reducing this uncertainty will facilitate the further evolution of the EDR. While the combination of the metrics being collected and the Council's goals provide some broad guidance as to the intended purpose of each variable, that is not detailed or scientific enough for several key purposes.

First, it is necessary to specify the statistical and empirical frameworks that will be used to link individual metrics to program outcomes in order to address basic questions about data sufficiency. Sufficiency addresses whether the data is accurate and precise enough to be useful for its intended purpose. By establishing scientifically acceptable levels of error and precision necessary to support the management program's performance on Council's objectives, EDR designers can identify variables where lower cost or less burdensome collection techniques can lead to data that is adequate for its intended purpose, although it may have higher variance or lower precision than could be attainable with unbounded resources.

Second, the intended use of the data, particularly as it relates to policy outcomes, is critical to establishing its credibility. Like accuracy and precision, it is useful to think of data as being credible for a purpose. For example, tons landed may be credible for purposes of tracking biomass, but not for purposes of determining how much a processor should pay an individual harvester. By specifying more carefully the intended purpose of the data, participants in the process can acknowledge that even data with variation or less-than-accounting-level precision can be appropriately used for its intended application.

The Panel is concerned that the Council has not drawn sufficiently upon the scientific community to understand the tradeoffs among dimensions of data quality and utility. Scientific standards require that experienced practitioners in a given field of research and analysis participate in informing policymakers in matters of scientific and statistical methodology. This includes the specification of the most appropriate models, consideration of alternative models that may make valid inferences possible based on lower quality data, and the evaluation of the extent to which statistical properties of the variables relate to the statistical and economic significance of the conclusions based upon them. These tradeoffs are deceptively difficult, and require scientific expertise, because simple comparisons of summary statistics are not sufficient to associate an observed change in outcomes with a management action. For example, a summary statistic analysis of the hatchery program for the Alaskan salmon industry over the last twenty years might conclude it had stabilized landings but decreased revenue, when in fact the expansion of aquaculture has driven down prices in the global market. Reliable policy analysis requires data to support building models, and alternative models, to test competing hypotheses about *why* an economic shift occurred. In many cases, shifts in exogenous factors mean counterfactual outcomes must be constructed to scientifically link observed changes in outcomes to management programs. Since it is difficult to anticipate the nature of the endogenous and exogenous shifts, and thus which data from within the managed industry are necessary to incorporate effective controls within statistical models, it is essential that data beyond the core set of metrics for the management program be collected.

Third, there is reasonable concern among respondents about the types of policy actions that might arise from the data, and some clarity about the likely outcomes of different analytical results could focus efforts of both analysts and stakeholders in improving the

data process. For example, industry representatives expressed concern that dramatic program revisions would take place based on data that were so inaccurate as to be wrong. It may provide reassurance, or at least clarity, for the Council or NOAA to establish guidelines and expectations for the economic and statistical significance of analytical results that would result in a material reconsideration of program elements.

Relatedly, in a small industry where individual operations often represent significant presences in communities, there is concern among some respondents about maintaining sufficient confidentiality of the data. While it was explained to the Panel that standard practices are in place, and the Panel--who all have experience with confidential data--did not see them as insufficient, they were not sufficiently clear to all respondents. Again, providing greater specificity about the types of analysis and results that will be reported may provide some reassurance. However these assurances should not be made with such specificity that they constrain analysts' ability to construct alternative models, or that they limit analysis of future management changes or exogenous events.

Better aligning requested information with industry information practices

Industry representatives and their accountants both expressed concerns--and frustration--over ways in which the information requested in the EDR did not align well with internal performance metrics of the businesses the data are supposed to describe. Adapting available records to calculate requested measures (which may make little business sense out of context) has proven time consuming, eroded confidence in the EDR data because adapting business records often requires guessing and interpretation, and frustrated users because measures were calculated differently by different respondents.

Many of these problems seem to arise from one or both of two disconnects between the potential users of the data and the EDR respondents. First, the EDR was specified by the Council to request only information on the part of each harvesting or processing firm's business related to crab, and then often to break down those variables by specific crab fisheries or locations. This is not, at first glance, problematic for analysis, because most of the models economists use to assess the welfare impacts of management changes are single fishery models that look at rents from individual fisheries in isolation; improving upon this is an important research frontier that requires a comprehensive fishing business-level data collection that is currently lacking. However, many businesses do not naturally track or apportion all their costs (or revenues) among the fisheries in which they participate, or across species within fisheries. As a result, the respondents are forced to apportion costs and earnings information across fisheries in unnatural ways: costs such fuel purchased to steam from Seattle to the Bering Sea for multiple seasons; line labor costs on days in which multiple species are being processed; and annual insurance do not have a single sensible way to be divided across species-level fishing activity. While respondents made clear they try to be diligent about apportioning these costs, doing so is time consuming, and because different respondents do it differently, there can be across-respondent inconsistencies in the meaning of the variables that reduce confidence in the data and limit its utility.

The second disconnect arises from differing expectations about precision of the data. In the environmental impact statement (NMFS 2004), and during the Panel meeting, the data program objectives included identifying any relatively large effects arising from changes in management, including changes in fuel use or improvements in product quality as a result of the elimination of the race to fish; industry is also interested in being able to understand these changes. As presented, the EDRs collect dollar-precision accounting data to address questions about costs that would need to change by thousands of dollars to be economically meaningful. Thus, while different apportioning methods lead to different numbers of dollars being allocated--and thus the perception that they are "wrong", as evidenced by these variables receiving low quality ranking in the PNCIAC review--all reasonable apportioning methods likely fall within 5% or 10% of each other, which is sufficient for the analytical purposes when testing for relatively large effects with models that account for statistical variation.

The Panel discussed two options for addressing the apportioning problem that warrant further exploration; different options may be best suited to different measures. One is to collect the aggregated measure (e.g., total insurance cost), and allow data analysts to develop models for apportioning most suited to the inquiry. These models would have the benefit of being consistent across observations, and make explicit that inferences are based on assumptions about apportionment. Alternatively, the EDR could collect the aggregated measure and the respondents' estimates of the percentage (perhaps to nearest 5%, for example) of how much is attributable to crab activity.

Two categories of data that frustrated harvesters and processors because they are not part of their regular business records are those pertaining to the location of goods purchased and information on quota holdings. In each case, the management objectives being tracked relate to the sellers, not the respondents who are acquiring them. Since the sellers are likely more knowledgeable of, and invested in, analysis of this data, it may simply be better practice to request it from them directly, if these objectives are key to evaluating success of the management program.

Database system and process design

Given the size of the reporting population of BSAI crab harvesters and processers, the Panel was surprised by the scale of the database development operation that NOAA had undertaken. This clearly reflects the desire on the part of the crab EDR design team to build an infrastructure that can be leveraged in subsequent EDR efforts in other fisheries.

The Panel was impressed with the detailed level of support that AKFIN provides, and the thoroughness of their manual checks on the entered data. Although most data are submitted on paper and entered manually by AKFIN, the Panel has a high degree of confidence that the data in the electronic database meet or exceed standards for data entry error.

The Panel did express some concern about the scalability of the manual components of the operation. Looking ahead to where more fisheries, with many more participants, are

using the same database platform and process for their EDRs, the Panel sees two key opportunities for improvement. First, through using the crab EDR as a pilot, everyone involved with the database has learned that the information architecture needs to be changed to be more flexible and to better meet the needs of the users. While improvements have been identified, the Council's current review of the crab EDR, and the options on the table to significantly reduce the scale of data collection, have led to hesitation in investing in this wholesale redesign. The Panel feels that experience gained through the crab EDR is a critical asset that should be leveraged if economic outcomes are to remain a part of the objectives of fishery management, and thus clear direction needs to be given to act on this experience for the benefit of the crab fishery, and other fisheries for which EDRs will be developed in the future.

Second, the Panel feels that the database entry interface could be significantly improved to make greater use of the technology in guiding both respondents and data entry clerks in maintaining low error levels. Increasing entry efficiency and accuracy will be particularly important as the operation becomes larger. For clerks entering data from paper forms, value bounds on individual fields need to be more carefully considered; the bounds could be linked across fields as early entries indicate the scale of the operation, and thus likely ranges of values for subsequent fields. For respondents entering their own data electronically, the survey could help ensure consistency by providing prior-year values or drawing information in from other data sources (e.g., fish tickets).

Discussion

There was a consensus among the Panel members that the data collection instrument had not yet achieved its goals of capturing the economic status of harvesters and processors in a way that would allow the Council to assess progress on all the goals of the rationalization program. However, the Panel feels that the data that has been collected through the EDR is suitable for use by analysts sophisticated enough to manage limitations inherent in any field data: the EDR is being developed through a process consistent with the best available scientific practices, and represents the best available scientific information on the economics of the BSAI crab fisheries.

The Panel perceived a sense among some participants, and in some Council documents and the CIE panel charge, that there was urgency in determining--at this stage and upon available information--whether individual EDR variables have achieved their desired accuracy based on threshold levels of metadata metrics, and to discontinue collection of all of those that have not, as not providing sufficient analytical benefit to warrant collection burden and cost. It is the expert Panel's unanimous professional assessment that this sentiment is premature. Drawn from diverse backgrounds in economic survey design and data management and analysis, the Panel assures all involved that the EDR's present state of uneven data quality is a normal stage in the development of significant new data collection efforts. Supported by a uniquely comprehensive--and novel in many respects--set of data quality metrics, the Panel believes continued work, perhaps through several more iterations of the EDR instrument, is the most appropriate course of action, and is likely to yield improvements in data quality that enhance the value of the resulting

analysis to the Council in evaluating progress against the management program's stated objectives.

The Panel associates most of the challenges discussed with the pilot nature of the crab EDR. A recurring theme of the CIE meeting--which surprised panel members from industries other than fisheries, and is an ongoing struggle in fisheries economics--is that fisheries economics lacks a commonly accepted process that dictates the flow of information and guidance among scientific and research staff and contractors, the SSC and other advisory bodies, and the Council itself. This is in contrast to the more developed and formalized process for stock assessment science across regional councils, including codification of specific biological reference metrics in the Magnuson-Stevens Act and the transparency and peer review that is provided by Plan Team discussions, in addition to expertise on the SSC. This process provides important clarity in the institutional role of scientists within the Council process, ensures that iterative improvements in data and methods are achieved over time, limits political influence on science, and assures key baseline data are available through successive management changes.

While the crab EDR represents the North Pacific Council's leadership in a similarly standard process and information base for economic tracking, how best to achieve the same standards and benefits is being explored concurrent with the development of the crab EDR instrument and analysis. The Panel acknowledges that this imposes additional burdens on the BSAI crab industry, but also recognizes this responsibility arises from harvesters' exclusive access to a particularly valuable public resource, and crab processors' uncommonly protected claim to crab rents. The evolution of the crab EDR can establish process, infrastructure and institutional capacity for economic data collection and analysis that provide support to the Council in designing, evaluating and adapting economically effective management plans, not just in the crab fisheries, but in other Alaska fisheries and fisheries in other regions.

In addition to making the crab EDR a more difficult process than it might otherwise be, not having a standard economic data process significantly limits the types of policy-driven economic analysis that are currently possible. First, because there is no consistent economic data collection, it is not possible to make before and after comparisons of a management measure. As a consequence, initiating the crab EDR required firms to go back into their records to collect retrospective data; the data would have been simpler to compile--and likely more accurate--had it been collected on an ongoing basis. Further, while each management system is enacted with the hope that it will solve all the fishery's problems and be in place in perpetuity, history suggests this is unlikely. Rather, it is more likely that there will be future changes in management regimes, industry structure, technology, stock status, or competitive conditions--not specifically foreseeable--that will require analysis relying on variables that may not have been central indicators of success for the current management program.

Second, because a standard set of economic variables is not collected across fisheries, it is difficult to determine if the enabling conditions that facilitated success of a

management system in one fishery are present in another where the same change is being considered. Third, because there is no infrastructure to collect data on economically important groups whose outcomes have not been historically valued in fisheries management (e.g., crew, processors), it is costly to establish connections with the people holding this information in the event a new policy dictates that their welfare be tracked.

Avoiding these limitations in the future requires adopting an approach to economic data collection like that used in stock assessment: develop a standard data program with at least a core set of variables that are collected regularly in all fisheries of interest, regardless of the current management plan or its goals. This will provide historical baseline and cross-fishery comparison data through management and other changes, and relieve NOAA and Council staff and other participants of the burden of designing a data collection scheme in the midst of the design, assessment, and implementation of a new management program. While there may still be need to develop supplemental collections for specialized aspects of particular programs in some fisheries, establishing a regularly applied EDR process would reduce the burden of developing new regulations, and improve the extent and quality of the economic analyses that could be done in support of management decisions.

To make progress on this goal, an idea that emerged from the Panel meeting, and elaborated in Panelist Hannah's report, is to establish an independent scientific panel charged with approaching the variable specification question from a perspective broader than capturing the effects of rationalization as it has evolved within the BSAI crab fishery. By convening a working group of economic analysts to advise the Council on the use of EDR data and to work with the EDR design team, the Council can begin to build an economic science structure, analogous to that used for biological stock assessment, that insulates the scientific process of developing a sound understanding of economic dynamics and where rents are generated from the political process of how rents are allocated among participants. Thereafter, in designing new management measures, the Council could focus more on the goals and objectives of management, rather than the specific variables that are collected to assess whether or not these objectives are being achieved.

While some may have a perception that the BSAI crab fishery has been singled out as guinea pigs for an ambitious agency experiment in data collection, there is a countervailing argument that the rationalization program has produced a windfall for BSAI quota holders--both harvesters and processors--based on the current and future revenue stream from the publicly owned crab resource. Enjoying the benefits of this windfall carries with it special responsibility to the public to ensure the public's goals for the program are being met, including resource health, wealth generation, and community sustainability. While industry has clearly been forthcoming and diligent about the requested details of their operations needed to support tracking these goals, it will also inevitably entail some forbearance with the development of a new process. Going forward, the EDR instrument and variable definitions will be refined and reporting costs will decline, especially for businesses that are flexible in adapting their record keeping practices to facilitate and expedite their data collection (which does not yet appear to

have happened on a broad scale). Industry will receive the benefit of the regulatory stability that arises from high levels of outcome information and transparency in an industry with consistent policy objectives, and the public will receive the benefits of knowing their resource is generating the outcomes established by the Council.

Documents Referenced

Abbot, K., B. Garber-Yonts and J. Wilen. 2010. Employment and remuneration effects of IFQs in the Bering Sea/Aleutian Island crab fisheries. *Marine Resource Economics* 25:333-354.

North Pacific Fishery Management Council. 2010a. Five-Year Review of the Crab Rationalization Management Program for Bering Sea and Aleutian Islands Crab Fisheries. Appendix A. Social and Community Impacts. Anchorage, AK. December 28, 2010. http://www.fakr.noaa.gov/npfmc/current_issues/crab/5YearRev1210_AppxA.pdf

North Pacific Fishery Management Council (NPFMC) and National Marine Fisheries Service (NMFS). 2004a. Final Environmental Impact Statement for the Bering Sea and Aleutian Islands Crab Fisheries. August 2004. http://alaskafisheries.noaa.gov/sustainablefisheries/crab/eis/default.htm

North Pacific Fishery Management Council (NPFMC) and National Marine Fisheries Service (NMFS). 2004b. Appendix 1 to the Final EIS: Regulatory Impact Review/ Initial Regulatory Flexibility Analysis, Voluntary Three-Pie Cooperative Program for the Bering Sea And Aleutian Islands Crab Fisheries. August. Http://Www.Fakr.Noaa.Gov/Sustainablefisheries/Crab/Eis/Final/Appendix1_Rir.Pdf