External Independent Peer Review
for the Center for Independent Experts

Puerto Rico Commercial Port Sampling and Catch Validation Project

Cynthia M Jones, Ph.D.
Professor Emerita and Eminent Scholar
Old Dominion University, Norfolk, Virginia

August 2023

Executive Summary

Commonwealth of Puerto Rico Department of Natural and Environmental Resources (DRNA) has relied on self-reported landings as a basis of their commercial fisheries management. DNRA, in coordination with the National Marine Fisheries Service, hired MER Consultants to undertake port sampling to validate landings and species compositions. Port sampling of Puerto Rico's commercial fisheries provided challenges for the MER Survey Team that were well met. The design was developed after an initial pilot survey of ports throughout Puerto Rico, St Croix and St Thomas. The pilot study allowed the designers to test the efficacy of various sampling approaches and to work through logistical issues. This allowed them to establish a well-vetted spatial frame of commercial fishing ports and landing sites. Based on the pilot study, they determined that the main sampling would be done with a multi-stage cluster design. The availability of information from the pilot study allowed the Survey Team to produce a high-quality survey that resulted in low proportional standard error (PSE) for many species that is important in guiding science-based management.

Even though Hurricane Maria interrupted the survey mid-sampling, the Survey Team was able to reassess the impact of the hurricane on accessibility of landing sites and modified the spatial frame to eliminate sites that were damaged beyond use. They resumed the survey after a hiatus and were able to complete it successfully. It is a reflection of the original extensive scouting of sites that this was feasible. Moreover, the detailed descriptions of sites provide a valuable resource that should not be underestimated as a foundation for future work and to better understand fishing access.

As a large island, the physical geography and coastal currents of the four coasts differ from each other, as does access and infrastructure. Different mixtures of species and harvest methods were to be expected in such a situation. In the design phase, the Survey Team added a secondary stratum of site use (high and low) within each first level regional stratum. Use classification was based on infrastructure such as the number of ramps, parking, etc. and by direct observation. The overall design is a multi-stage cluster design, similar to that of Hege et al. 2021. The sampling unit is site by day, within use and region strata. The estimate for landings was given by equation 1 (pg. 38: Part 2). These were excellent choices to sample landings for Puerto Rico.

Total landings for individual species were estimated weekly. Using weekly expansion also is likely to result in highly variable estimates of less abundant species compared with estimating monthly or seasonally. Their design is amenable to aggregation across time. The advantage of weekly estimates lies in the ability to use model-based approaches to evaluate seasonal or cyclical factors in species landings, which they allude to but is worth further work. Estimation in the West and North regions was further complicated. A major site in the West, Puerto Real Soltero (SOL), disallowed survey agents to be on site. This precluded direct measurement of catch. Hence, landings were estimated by counting arriving boats and using a CPUE ratio estimator obtained from a nearby site to expand landings as CPUE x E (EQ 3 pg. 42: Part 2). This is a commonly used and reliable estimator for such a circumstance. They used the Horovitz-Thompson estimator, and I suggest that other estimators such as the Yates-Grundy might yield more stable variance estimates. Because the survey design in the North region was similarly complicated, the expansion of landings is more challenging too. Because the documents didn't include an equation for total landings over the survey period, I assume that landings were simply totaled ($\hat{Y}=\Sigma_j\Sigma_i\hat{y}_{ij}$). They conclude that their estimate is unbiased and that the variance is biased high. Rather than make this assumption about variance, there are alternate methods besides design-based analytic solutions to explore, such as bootstrapping, machine learning, and various new techniques other than the classic methods available from Cochran (1977), which is an excellent reference but dated.

They were able to achieve low proportional standard errors (PSE) for many species because of the large sampling size they accomplished. Comparing the survey estimates to those that were self-reported shows the limitations of accepting self-reporting as a measure of true landings. Major species, such as spiny lobster and queen conch have apparently been underreported by fishers while snappers have been overreported (Tables 36-41: Part 2). The report indicates that the expansion factor (≈2x) used by Puerto Rico misses the actual relationships between observed and self-reported landings. Based on my review, I trust that the survey provides a better estimate of true reporting, at least for the abundant species (Table 43 pg. 110: Part 2), and that the expansion factors developed through the survey be considered for use.

The commercial fishery of Puerto Rico is characterized by many smaller landing sites with less commercial infrastructure than for most of the United States commercial ports. Hence it lies between mainland commercial and artisanal fisheries. Several approaches can be used to estimate total catch in such a fishery: on-site direct observation of effort and catch, offsite estimation from list frames or a combination of these. Landings can also be estimated from expanding catch-per-trip (CPUE) by an independent measure of effort. CPUE is measured by interviewing fishers on site and observing their catch as they complete their fishing trip. Effort can be estimated independently from 1) a vantage point or device such as a traffic counter or from 2) an offsite list frame. The Survey Team chose an on-site survey and because it was conducted to high professional standards, it produced reliable data on landings and species composition.

The Survey Team produced in-depth documentation of their survey, with particularly impressive validation of species identification and documentation of potential landing sites throughout the island. Although I have recommendations that may improve the survey in the future, these should not take away from the high quality of work that was done.

Background

The Commonwealth of Puerto Rico operates under the Magnuson–Stevens Fishery Conservation and Management Act that requires fisheries be managed with the best available science through the Caribbean Fishery Management Council. Article 11 of the Commonwealth of Puerto Rico Department of Natural and Environmental Resources (DRNA) fisheries regulations state that commercial fishers must self-report their landings as a requirement of holding a license. Although there is mandatory reporting, a method of verifying these self-reported landings with dealer reports is not feasible. There is uncertainty that these self-reported landings are accurate, so MER was given a contract by the DRNA to conduct on-site port surveys to compare estimated landings with self-reported landings during August 2017 -December 2019. During the Fall 2017 Hurricane Maria caused major damage to Puerto Rico and her ports, resulting in a months-long hiatus in sampling and a reconfiguring of the sampling plan.

Puerto Rico is a large island that is characterized by distinct coastal habitats, over one hundred ports where commercial fishery landings are made and where landing sites have wide disparities in infrastructural support. There are several approaches to sample commercial fisheries landings under these conditions, through use of a list frame or with an on-site spatial frame. The list frame would consist of the names and contact information of all active commercial fishers. Such a list frame can be obtained by requiring a license, through another fisher registry, or with a snowball method. It appears that Puerto Rico has a list frame through licensure. Typically, the fisher is contacted and self-reports their number of trips and landings. Self-reporting is thought to be more reliable for obtaining trip information than for landings which might be subject to catch limits and possible inaccurate reporting. The MER reports provided to me indicate that there may be both under- and over-reporting of landings. In fisheries with larger vessels and few vessels, fishers can be met at ports to verify their self-reported landings in coordination with a list frame – one of the MER recommendations suggests this approach. However, this approach seems impractical for the entire Puerto Rico commercial fishery, which is presented as having many participants with smaller boats that would provide a challenge to verify in this manner. Another list frame approach would be to obtain purchase records from fish houses that buy the catch. This approach works well if all landings are sold to dealers but is not reliable if there is significant "basket trade" – cash sales to individuals that are not traceable. This appears to be true for Puerto Rico.

The alternative to a list frame is to sample on site with a spatial frame. This requires a frame that encompasses all of the sites where fish are commercially landed. It is a spatial-temporal frame that when sampling is done over time, consists of all landing sites over all days that the fishery is prosecuted. The challenge to conducting an on-site survey is not only that all sites must be known, but its greater expense to conduct. It requires greater logistics around travel and training of survey agents as they interact with fishers. Nonetheless, if frame coverage is complete, it provides unbiased estimates of the landings.

A widely used method to sample this type of spatial-temporal frame is to stratify the survey by region, possibly by day type (e.g., weekday versus weekend), and by site use or landings. Because the fishers landing catch each day is unknown, sites can be handled as clusters of fishers. This type of survey is a multi-stage cluster sample. An alternative approach would be to use a bus-route design for sampling sites as a multistage bus-route design with sites as clusters. The MER Survey Team (henceforth called the Survey Team) chose a two-stage cluster design for their primary sampling and ancillary sampling with the bus-route.

Before the survey could be undertaken, the Survey Team verified and documented commercial landing sites through on-site visits, descriptions and photographs. After Hurricane Maria, sites were re-verified and damaged sites eliminated from the survey. Training of survey agents was extensive, with much attention given to careful recording of proper identification of a large species complex. Lengths and photographs were taken of fish to verify species identifications. The report's four parts were very comprehensive. The Survey Team put in great effort to make a very complex design clear.

Description of the Individual Reviewer's Role in the Review Activities

This was a desk review, and I was sent the review materials listed in Appendix 1 at the very end of April 2023. I read through these materials and some supplemental papers in the ensuing two weeks in preparation for a webinar held on May 16, 2023, wherein NMFS personnel were able to answer some of the two CIE reviewers' questions. For questions that were pertinent to the Survey Team, NMFS provided written responses in subsequent days. Because of health-related intervening events, I requested and was granted an extension of the due date for this report.

Summary of Findings for each ToR

TOR1. Evaluate the statistical design and implementation of the port sampling survey to estimate commercial catch by species and the methods used to develop the design.

Port sampling of Puerto Rico's commercial fisheries provided challenges for the survey design team that were well met. The design was developed after an initial pilot survey of ports throughout Puerto Rico, St Croix and St Thomas. The pilot study allowed the designers to test the efficacy of various sampling approaches and to work through logistical issues. Sites were chosen by referencing community knowledge and through site visitations. Thus, they were able to establish a well-considered spatial frame of commercial fishing ports and landing sites. It is difficult to ascertain whether there are any frame coverage issues that will become clear over time. It appears that frame coverage issues would be minimal and easily corrected. Were sites to become active or inactive, modifications could be easily made to the spatial frame. The implementation of the survey was interrupted mid-sampling by Hurricane Maria in 2017. The Survey Team was able to reassess the impact of the hurricane on accessibility of landing sites and they modified the spatial frame to eliminate sites that were damaged beyond use. They resumed the survey after a hiatus. It is a reflection of the original extensive scouting of sites that this was feasible. Moreover, the detailed descriptions of sites provide a valuable resource that should not be underestimated for future work and to better understand fishing access.

As a large island, the physical geography and coastal currents of the four coasts differ from each other, as does access and infrastructure. Different mixtures of species and harvest methods are to be expected in such a situation. The Survey Team was aware of this and chose to create four regional strata (East, South, West, North) wherein estimates would be calculated independently and survey sampling approaches applied to best fit regional differences. When strata are chosen carefully the resultant variance is minimized by the putative homogeneity within each stratum. Typically, stratification doesn't increase total variance. Additionally, a compelling reason to stratify regionally was to better apportion survey teams and minimize travel. It also has the effect of spacing out sampling across the island. This approach was well chosen and should help logistics and decrease variance of the estimates.

In the design phase, the survey team chose to add a secondary stratum of "site use" to each primary stratum (PSU) of region. Within each first level regional stratum, a second level (SSU) was applied to partition by high and low use sites. Use classification was based on infrastructure such as the number of ramps, parking, etc., and by direct observation. High-use sites could then be sampled with higher frequency, thus with greater probability of selection compared with low-use sites --sampling proportional to size (PPS). This is a valid approach when there is a disparity in site use, as is indicated in the report, and theoretically should reduce the total variance. Sampling proportional to effort may yield greater design efficiency if effort correlates closely to landings. An approach that is used less frequently is to sample proportional to landings. Sampling proportional to landing requires historic site-specific data on landings. While not feasible for this current study, the results of the survey can be used in simulations to see if this is a reasonable method for allocating on-site sampling.

An unknown in on-site surveys of fishing sites is the number of fishers that will be encountered during a sampling event. Fishers and their catch are the elements within the second stage that are enumerated. However, there is no predetermined list of fishers who will use a specific site on a given day during the time interval of the sampling event. A viable approach to deal with this issue is to use a cluster sample, wherein effort and catch is calculated as the mean per cluster. Hence, variance will include within and between cluster components. This is the method that the Survey Team chose. Within the second stage of the design, sites were randomly chosen without replacement (pg. 38: Part 2). Note that for three regions (E, S, W) only a single site was chosen each day, precluding the ability to estimate within-day variability. For the North, the high use stratum was a single site. Thus, they sampled the high use site 2x/week as a simple random sample and the low use sites 3x/week as a cluster sample. In this instance the choice of days allocated to low use was not independent of high use. The Survey Team recognizes that this complicates their estimation of weekly mean and variance. Additionally, with many low-use sites, choosing without replacement could be problematic if site numbers were few. However, that was apparently not the case for this survey. One solution would be to use a correction factor at this level if needed.

For the West high-use stratum one important site was unavailable for on-site interviews. The survey used an alternate method for measuring effort and estimating landings at this site. Boats were counted and the CPUE from the sites within the stratum were used to expand landings from effort – catch per trip x trips. This is a standard methodology for estimating landings from count data (pg. 42-43: Part 2). No variance estimate was listed for this estimator, so I presume that they just added this estimate with the other sites. Because CPUE based estimates can have skewed distributions, the standard variance may not capture this dynamic. However, this is arguably a minor issue.

To document low use, the survey team chose to use a bus-route survey for ancillary sampling. In a bus-route survey the list of sites per route is chosen, the travel time around the route measured and the remaining time in the workday apportioned to the sites. The entire route is converted to a timeline with the beginning chosen randomly each survey day. Thus, if the random start falls within a travel segment, the agent will be at the first site into the workday. It appears that this was how the survey was conducted in Puerto Rico. It is important that the start location (time) be instituted correctly for the geometric probability to be correct, and that the direction be chosen randomly each time, backwards or forward along the route. There isn't sufficient documentation to know if this was done. Moreover, we (Robson and Jones, 1989) developed the design to provide the catch and effort estimate for the entire route, not for specific sites within a route, because of variance consideration. The report refers to the bus-route as a roving design and we so stated in our paper (Robson and Jones, 1989) that it was a *roving-type design*, but were careful to address that because the survey agent only interviews fishers after they complete their trip, that it is an *access design* in its probability structure (Robson and Jones, 1989, p.97). It is important to make this distinction because catch and effort are computed with entirely different estimators for roving and access designs.

The main survey design also had a temporal frame. Five sample days were chosen from six days, Monday through Saturday. Because sampling was done without replacement obviously and because this constituted close to a census, the survey team used the finite population factor to correct for overestimation of variance. Their results show no trend between days, so the choice of equal probability was sound. The survey period extended from 9am-5pm (pg. 39: Part 2 document) for three regions (E, S, W) choice of day was the primary sampling unit (PSU). Sundays were considered to be extremely low use. The Contractor email response (May 19, 2023) stated that no sampling was done on Sundays. However, Table 9, pg. 50: Part 2 shows that a bus-route design was used in all 4 regions in AM/PM and Sundays; Table 53 shows sites with significant night and Sunday landings, apparently from the pilot study or ancillary sampling. This narrative isn't part of the pilot project references and so it appears to be part of the full survey. I find this confusing, making it difficult to know how this sampling fits the narrative. If in fact the pilot or ancillary studies showed significant night and Sunday landings at some sites (Table 53), then the temporal frame shows undercoverage and landings will be underestimated. This issue needs much more clarification than I received if it is to be eliminated as a concern. The lack of clarity in the AM/PM sampling narrative constrains me to address this sampling scheme hypothetically. When the day length exceeds the legal working day, the day can be further divided into morning and afternoon strata, the secondary sampling unit (SSU). The SSU can be chosen randomly with replacement, meaning consecutive mornings or afternoons could be chosen. This is a commonly used and valid practice. Note that when days shorten, there can be overlap between morning and afternoon strata. When this occurs, the overlap period probability can be adjusted allowing maximization of the survey agent's day.

Night fishing was assumed to be incidental with no sampling conducted over that period, which was as stated in an emailed response to webinar questions. For most of the regions, diving was the major catch method, and this occurs primarily during the day. However, it is always judicious to have a minimal check when eliminating a time period from sampling. But note, Table 53 shows the majority of landings occurred at night for several sites in all four regions. Again, it isn't clear what changes occurred or what decisions were made to assume night fishing is incidental, which perhaps might be the case at most sites, but not for all. I don't understand the disparity between the dismissal of night sampling and the data presented in Table 53. Given that data, at least some accommodation for night sampling was warranted.

Auxiliary sampling was undertaken as part of the pilot study to determine the amount of landings at three locations on Vieques and Culebra. Two sites were considered low use and one high use. It appears that mean daily landings were low for these islands.

An unappreciated component of a well-designed survey is the care taken with logistics, including quality control and training. A major component of the survey report presented the effort to train survey agents to follow specific guidelines in approaching fishers, counting boats, conducting interviews, and recording their observations. Specific training was conducted to ensure correct identification of species including photographing catch for species verification. The use of tablet devices which recorded GPS locations guarded against "dry-labbing" where a survey agent records fictitious data without a site visit. Moreover, supervisors also visited sites to observe for themselves the sampling process and to see if agents had questions. I was impressed with the care that was taken to ensure valid data were recorded. As long as the survey design ensures random and accurate data are taken, the survey team can employ a variety of estimation techniques to evaluate landings, even with other estimation approaches beyond the standard ones that they chose.

TOR 2. Evaluate survey expansion algorithm for total catch by species.

The estimation of total catch depended on the design specifics for the regions. The general multi-stage cluster design and its estimator was straightforward for the majority of the island. Exceptions included the North and West regions. In the North, there was only one high use site and in the West one site disallowed survey agents on site. These considerations result in a multi-pronged approach to landings estimators.

The overall design is a multi-stage cluster design, similar to that of Hege et al. 2021. The sampling unit is site by day, within use and region strata. The estimate for landings was given by equation 1 (pg. 38: Part 2). Total landings for individual species were estimated weekly. Not every trip was intercepted for interview and a ratio between sampled to total trips was used as an expansion factor to estimate total landings. I was a bit surprised that landings (by species) was expanded weekly and presume that was a contractual obligation. While commercial landings can be substantial, for example with gear such as gill nets, capture methods like diving are less likely to produce prodigious landings. Using weekly expansion also is likely to result in highly variable estimates of less abundant species. Moreover, in regions with many sites random draws may not result in duplicate sites chosen within a week, thus precluding measurement of any within-site variation. As the survey team also recognizes, the choice of visiting a single site each day precludes estimation of within-day variance. I don't make this a major issue because the stratification that they've chosen makes it easy to aggregate weekly data in less variable monthly or seasonal expansions. The advantage of weekly estimates also lies in the ability to use model-based approaches to evaluate seasonal or cyclical factors in species landings.

When a site was busy, the survey agent counted completed trips and a ratio of total trips/sampled trips was used to expand landings at a given site. I was unsure if "total trips" referred only to commercial trips as it should have. If not, then the ratio of recreational to commercial trips would also need to be quantified or else sites with appreciable recreational fishing would result in an overestimate of commercial landings.

Estimation in the West and North regions was further complicated. A major site in the West, Puerto Real Soltero (SOL), disallowed survey agents to be on site. This precluded direct measurement of catch. Hence, landings were estimated by counting arriving boats and using a CPUE ratio estimator obtained from a nearby site to expand landings as CPUE x E (EQ 3 pg. 42: Part 2), a commonly used and reliable estimator for such a circumstance. The survey team went an additional step and compared the relationship between landings at SOL taken before agents were disallowed on site with landings at the adjacent site and verified linearity. When 2 other sites within Puerto Real were discovered during the survey, the same approach used for SOL was used for them. The only complication with using the CPUE to expand is that CPUE distributions tend to be long-tailed with the mean unequal to the median. Such is obvious in Figure 22 (Part 2). The survey team also adjusted the sampling probabilities in the Horovitz-Thompson estimator to account for combined previous independent sites. Given the circumstances, the survey team used one of the best approaches available. One concern for this set of sites is the statement (pg. 43: Part 2) that total counts were made inconsistently over 5 months, necessitating use of a regression predictor for total counts at 3 of the sites. Compared with the estimation of landings at other sites, I would expect the use of a linear repression to minimize the true variance at Puerto Real.

Because the survey design in the North region was similarly complicated, the expansion of landing is more challenging too. The survey team realizes that by decreasing sampling at Jarealito (a single site in the high-use stratum) and shifting that survey agent into sampling low-use sites, they have a complex and non-independent probability structure. Note the second survey agent was sampling low-use sites with the same protocol as for other regions. They conclude that their estimate is unbiased and that the variance is biased high. Rather than make this assumption, they have alternate methods besides design-based analytic solutions to explore, such as bootstrapping, machine learning, and various new techniques other than those available from Cochran (1977), which is an excellent reference but dated.

The Survey Team used equation 1 (pg. 38: Part 2) to expand observed landings to estimated weekly landings and equation 2 (pg. 38: Part 2) to estimate weekly variance. Because the documents didn't include an equation for total landings over the survey period, I assume that landings were simply totaled ($\hat{Y}=\Sigma_j\Sigma_i\hat{y}_{ij}$). They also presented the weekly estimates of total landings as a time series, but didn't fit any models to those data. That would provide another option where warranted by seasonal patterns (mindful that they only have one period). They were able to achieve low proportional standard errors (PSE) for many species because of the large sampling size they accomplished. Comparing the survey estimates to those that were self-reported shows the limitations of accepting self-reporting as a measure of true landings. Major species, such as spiny lobster and queen conch have apparently been underreported by fishers while snappers have been overreported (Tables 36-41: Part 2). The report indicates that the expansion factor (≈2x) used by Puerto Rico misses the actual relationships between observed and self-reported landings. I would recommend that the survey provides a better estimate of true reporting at least for the abundant species (Table 43 pg. 110: Part 2).

On a minor point, I would suggest a change of notation in one case: Pg 44: Part 2:

$$\bar{y}_i = y_{ij}$$

Perhaps there was a typographic error for this equation. It wasn't clear what was done exactly because this equation was a mystery to me. I assumed that the survey team just used the totals for species landed instead of the daily site mean when only one site was visited. It is simply easier to say that y_{ij} was used as a proxy when a mean was unavailable.

Because the North was characterized by many low use sites, the survey used a bus-route estimator for that component.

TOR 3. Evaluate the approach for estimating total catch in light of other approaches that utilize CPUE and estimates of total effort.

The commercial fishery of Puerto Rico is characterized by many smaller landing sites with less commercial infrastructure than for most of the United States commercial ports. Hence it lies between mainland commercial and artisanal fisheries. Several approaches can be used to estimate total catch in such a fishery: on-site direct observation of effort and catch, offsite estimation from list frames or a combination of these. The Survey Team chose an onsite method that expanded landings and trips from trips completed at landing sites. The advantage of this approach is that it is design unbiased as long as there is little to no frame undercoverage.

Landings can also be estimated from expanding catch-per-trip (a form of CPUE) by an independent measure of effort. CPUE is measured by interviewing fishers on site and observing their catch as they complete their fishing trip. Effort can be estimated independently from 1) a vantage point or from 2) an offsite list frame. The Survey Team used method 1 in the West Region at the Puerto Real sites where they lacked access to complete coverage of observed catch at all sites (discussed in TOR 2). The advantage in using a CPUE estimator is greatest when landing sites can't be accessed but can be observed, when there are overwhelming numbers of vessels landing their catch at the same time, or when there are logistic and budgetary constraints on visiting access sites in a designed sampling approach. This use of CPUE relies on assumptions of equivalence in CPUE between sites, an assumption that was tested at Puerto Real. Under circumstances wherein catches can be observed at sites, there is little advantage in using an on-site CPUE estimator in Puerto Rico commercial fisheries other than discussed above and a disadvantage to be discussed later. It doesn't provide a better estimator than the one the Survey Team used; it doesn't provide logistic or budgetary advantages for this region.

Landings can also be estimated from expanding catch-per-trip (CPUE) by an independent measure of effort. CPUE is measured by interviewing fishers on site and observing their catch as they complete their fishing trip. Effort can be estimated independently from 1) a vantage point or from 2) an offsite list frame. This type of method has been used by the NMFS in its Marine Recreational Information Program (MRIP) wherein it uses an access-point estimate of CPUE with a mail survey for the number of trips. Although MRIP is a survey of recreational fishing, it bears similarities to the artisanal component of the Puerto Rican commercial fisheries. Alternatively, commercial fisheries effort is also estimated from mandated trip reports or mail-in vessel reports in mainland commercial fisheries. To use this method, there has to be a list frame that is complete and does not have meaningful undercoverage (in this case fishers who are not included on the list). It appears that this type of list frame is not currently available for Puerto Rico and would require not only mandated licenses but enforcement of reporting requirements. Moreover, self-reported trips are subject to mis-reporting, memory bias, and fabrication unless there are methods to verify trips numbers. In many commercial fisheries, trip reports can be validated through dealer slips where purchasers report the landings from an individual boat or fisher. In Puerto Rico there appears to be partial sales to dealers, direct sales to restaurants, and also "basket" trade where catches are sold directly to consumers without reporting. For species that have reliable list frames, a CPUE method could be cost effective. From the report and discussion in the webinar, it didn't appear that this was a viable method at present. Nonetheless, laws, regulations, and enforcement to develop and implement reliable list frames has value in regulating Puerto Rican commercial fisheries.

An additional consideration in using a CPUE estimator lies in its underlying probability distribution. Unlike the access point estimate of mean landings, the distribution for CPUE is often skewed. Although CPUE can be less highly skewed for commercial fisheries compared with recreational fisheries, Figure 22: Part 2 demonstrates a long right tail for Puerto Real CPUE that may also be present at other sites. The Central Limit Theorem is a tenet of statistics and states that a distribution of means will exhibit a Gaussian probability curve. That tenet works well for access-point estimates but it problematic for CPUE-based estimates (Jones et al., 1995) because the long-right tail can persist even after mean values are calculated. This persistent skew can produce unequal confidence bands. It is unknown if this would be a problem for estimates of CPUE in Puerto Rican commercial fisheries, but it would need to be investigated.

TOR 4. Evaluate recommendations for future work as provided in the survey report. Provide any additional recommendations for future research/improvements for the survey design and expansion methods.

The Survey Team listed 10 recommendations that I will address in order before providing my additional recommendations.

Recommendation 1. Governance – This first recommendation concerns coordination between governance entities, with DRNA to develop formal objectives to obtain reliable commercial landings, specifically reliant on port sampling. As they point out, the value of coordination would be to have clear objectives that provide reliable landings and species composition data to be used by multiple programs. I agree that this is a valuable recommendation that can be achieved. It would result in potential cost savings to all entities. Moreover, it would be foundational to managing Puerto Rican fisheries sustainably as mandated by the Magnuson-Stevens Fishery Conservation and Management Act (MSA). That should also include Vieques and Culebra as they are under the same legal mandate.

The second part of this recommendation is a bit disconnected from the first part because it is a technical issue of sampling design. Whether the sampling plan extends to AM, PM, or Sunday shifts or 9 AM to 5 PM shifts should be determined by the characteristics of the fishery, labor law constraints, and budgets. These are well researched issues (Pollock et al., 1994) that are best left to the survey design teams. Pilot studies and the current study reveal the extent of landings at different locations and expose the extent to which sampling is needed.

<u>Recommendation 2</u>. Survey design – The survey design that was used proved to be well suited to the current conditions in Puerto Rico's commercial fisheries. The characterization of landing sites and ports is invaluable for future work. Multi-stage cluster sampling is widely used around the world (see Hege et al., 2021 for example) for recreational and to some extent artisanal fisheries, so the methodology is well-studied and vetted. I agree wholeheartedly that sampling proportional to effort or to landings would yield smaller variance and was surprised that wasn't done. The size of a landing site alone indicates greater use and now that the survey has provided such data it should be used to establish future proportional sampling. I also agree that simulation is an inexpensive method to configure sampling to provide minimal variance.

A two-stage cluster sampling survey is one of many methods that can be used. For example, the Survey Team used a bus-route design as an auxiliary method. We developed the bus-route survey for a region that had many low-use sites that were closely spaced. It made no sense to visit one or two sites a day. It worked quite well and provided unbiased estimates of landings with low variance. Variance was low because it was calculated using geometric probability and included all sites on the route as a totality. The drawback of the bus-route method is that it is logistically complicated and takes additional training of survey agents. Other possibilities are the use of CPUE with list frames as discussed in TOR 3.

<u>Recommendation 3</u>. Extending impact of survey program - This recommendation addresses the value of obtaining samples for biological metrics to be used in stock assessment. The survey team is correct in stating that such samples would be valuable to several programs while also cautioning that this sampling could interfere with obtaining landings and effort data. I would suggest that collection of samples for biological data be instituted as a separate program wherein fish are purchased with specific objectives for numbers at size. Quinn and Deriso (1999) discuss how such collections can be made that provides age-length, age distribution and maturity schedules with minimization of variance for each size/age bin. The Virginia Marine Resource Commission instituted these collection methods and reduced costs, and Dr. Liao developed an app that is available for use by agencies (https://vmrc-cqfe-web-app.shinyapps.io/cqfefishageingsamplesize/_w_2eb8fd71/SampleSizeUsersGuide.pdf). Although having a separate sampling program adds some cost to budgets, it provides better data. Collection of biological samples other than length during port sampling requires a considerable increase in duration of the interview for the fishers and potentially damaging the appearance and decreasing the value of their catch.

<u>Recommendation 4</u>. Rapid sampling and use of technology - The Survey Team has demonstrated that rapid sampling can be done during port sampling using new technologies. I agree that this advance will provide critical information for stock assessments and for ecological studies. Combined with separate sampling for biological hard parts (discussed in Recommendation 3), age-length and maturity at length information will permit more advanced stock assessment techniques to be used for Puerto Rican commercial fisheries to better fulfill legal mandates.

<u>Recommendation 5.</u> Quantifying effort - The Survey Team recommends that alternative methods to quantify effort may be justified and cost effective. They state that counts taken from stationary cameras might not be feasible. We used cameras in our survey of Swedish coastal fisheries and found cameras effective when positioned correctly in large marinas. However, we had to abide by strict privacy laws and there is a serious cost to validate counts of fishing boats and process video. So, I would agree that the use of cameras is not a universal answer to effort capture.

Alternate effort counts, such as trailer counts, can be used in direct expansion as they did in equation 3 (Pg 42: Part 2) where r_i would be the ratio using ancillary count data to interviews. This necessitates that there be a direct linear relationship between trailer counts and fishing boats. In fact, for the New York Great Lakes Creel Survey we used trailer counts as one measure of effort in our application of the bus-route estimator. My colleague, Dr. Aldo Steffe, (Steffe et. al., 2008) used traffic counters laid down at landing sites to count boat trailers at sites. This was a form of double counting at the access points and resulted in significant improvement to the precision and accuracy of survey counts. Because people are used to traffic counters, they experienced little notice and no vandalism.

Recommendation 6. Challenging species/special considerations - The Survey Team brings up an important point that a general survey of ports does a poor job of sampling rare species events. The Horowitz-Thompson variance estimator is notoriously bad when dealing with rare large landing events. Variance can become very large as a result. It is also difficult to obtain a reliable estimator of short-season fisheries for similar reasons. Rare species that are very infrequently landed have similar issues. The Survey Team recommends a self-reporting and verification method. They propose to use a list frame registry, for example of yellowtail snapper fishers, who hail in before they return to the landing site and are met by a survey agent to verify the catch. Even with a mandated registry, there may be little incentive for fishers to hail in. It would also require a survey agent to be ever ready to go to the landing site in time to meet the boat. I see lots of difficulties with this approach but there are few good alternatives.

Recommendation 7. Analysis of weather and lunar cycles - The Survey Team states the importance of further analyzing their data to explore if there are correlations with landings and effort to meteorologic factors. I expect that this could reveal interesting patterns, but I'm not sure how the results would be used for regulatory purposes or to improve stock assessments. Perhaps a relationship could be used as a covariate to reduce variance. I do suggest that if they have a significant number of days of bad weather with zero catch and effort that they employ zero-inflated estimation procedures.

Recommendation 8. Expansion factors, family grouping, and species compositions - If I understand correctly, one of the goals of this project was to provide a comparison of catch estimates between the survey and the self-reported landings so as to develop expansion equations that would provide more accurate estimates of landings, especially if this could be applied to historical self-reporting. Additionally, the survey provides landings by verified species. Self-reporting appears to be rife with misidentification and misreporting of species. The Survey Team went to great effort to assure that species were reported correctly including photographic identification and in the case of sharks, DNA verification. They have provided the most accurate classification of species and species groups that is currently available. Were the proportions to be stable, then an estimate could be made from self-reported landings of true species composition. It would be an interesting idea to test. The task of educating commercial fishers to correctly identify species is a good idea but there would have to be a real incentive for the fishers to take their time to do this. What would they get out of it, if not more regulation? Perhaps this could be tried with a specific fishery and incentives to see if it would be worth the cost.

Recommendation 9. Deepwater snapper (DWS) reporting validation – As I understand, to obtain a DWS permit the fisher must have 5 years of reporting an average of 1,000 lbs of snapper per year. For fishers who do not meet the average poundage, there is incentive to overreport their landings to retain their permits. The on-site survey didn't cover the full landings times for this fishery, whereby boats return after the survey agent finishes their day. Hence, it's difficult to make inferences from the survey results. The Survey Team suggests that trailer registrations be used to identify permit holders and thus validate the number of trips that were taken. This will require cooperation with the Department of Motor Vehicles and perhaps funding to cover DMV costs. Survey agents would keep a list of vehicle or trailer registrations for each site. The Team would compile a non-duplicative list which the DMV could then match to registrants. That would then have to be matched with permit holders. In the Great Lakes Angler Survey we did something similar where we matched vehicle registrations to addresses and sent survey questionnaires to elicit the economic value of that fishery. With computers, the matching is quick, but the administrative issues can be more time-consuming and difficult. This task would be narrowed if there were known landing sites for DWS.

Recommendation 10. Fishing community outreach and inclusion in the process - I agree with the Survey Team that gaining trust of the fishing community is a valuable goal. This survey has established rapport between survey agents and fishers evidenced by the overwhelming cooperation they received in examining the landings. Trust is difficult to gain and keep, especially if fishers are regulated more because of survey results. It is standard practice in regional councils to have representatives of fishing communities as members. Further inclusion can help. Transparency is beneficial when the scientific process is made intelligible to the lay person. Much easier said than done, but worth the effort.

Conclusions and Recommendations in accordance with the ToRs.

My recommendations

I have included my recommendations throughout the TORs. I have a few additional ideas to share that might be useful.

As I mentioned above, the use of traffic counts should be useful at active sites that are used almost entirely for commercial fishing. Placement of the counters near ramps works well (Steffe et. al. 2008) and improves accuracy at little additional expense. Traffic counters could provide information on trips that occur out of the temporal frame, such as nights and early and late arrivals. Of course, night counts would have to be verified, but this could be done with a few random visits. Moreover, people are so used to traffic counters that they take little notice of them, and the counters are unlikely to be vandalized. This can be confirmed with the DMV which may also use them.

The report discussed the possibility of noting the names of fishers, the vessel registration, or the boat trailer registration. Were vessel registration numbers universal and easy to read, they would provide the ability to use a mark-recapture estimate of the number of vessels fishing during the season. It could be used to populate a more accurate list frame of commercial fishers at little extra cost. Dr. Hoenig is adept at using such techniques.

Another method that might be helpful in populating a more accurate list frame of commercial fishers would be the use of "snowball" sampling (https://research.oregonstate.edu/irb/policies-and-guidance-investigators/guidance/snowball-sampling). This technique has been used in the social sciences and asks participants for the names of other people who are then contacted. Recommendation 10 emphasizes the importance of community outreach, and if such a program was undertaken then snowball sampling could be introduced as an integral part of that program. It would not be costly and would provide a formal framework for upgrading the list frame of commercial fishers.

The Horowitz-Thompson variance estimator doesn't provide stable variance estimates when low probability, rare large events are sampled. I suggest that other variance estimators be investigated. We had similar issues in our survey of New York Great Lakes and found that the Yates-Grundy estimator provided more stable estimates. I don't necessarily recommend any specific estimator because modern sampling textbooks provide newer methods for such sampling problems.

Appendix 1. Bibliography of materials provided for review

Materials Reviewed for CIE Report
PR Port Sampling - MER - Part 1 Executive Summary (FINAL v2 -sent full resolution)
PR Port Sampling - MER - Part 2 Main Body (FINAL v2 sent- compressed resolution)
PR Port Sampling - MER - Part 3 Appendices (FINAL sent - compressed resolution)
PR Port Sampling - MER - Part 4 Site Descriptions (FINAL sent -compressed resolution)
Design of a Port Sampling Program for the U.S. Caribbean Appendix 1
MER_Final_Port_Sampling_Report_Appendix_2
MER Port Sampling Site Descriptions Appendix 3
SOW MER Consultants 2017

Additional Materials in Support of the Review

Cochran, W. G. 1977. Sampling Techniques 3rd Edition. Wiley. New York.

Heydayat, A.S. and B.K. Sinha. 1991. Design and inference in Finite Population Sampling. Wiley. New York.

Jones, C.M., D.S. Robson, H.D. Lakkis and J. Kressel. 1995. "Properties of Catch Rates Used in Analysis of Angler Surveys". Transactions of the American Fisheries Society 124(6):911-928.

Levy, P.S. and Lemeshow, S.L. Sampling of Populations: Methods and Applications. 2008. 4th Ed. Wiley. New York.

Pollock, K.H., J.M. Hoenig, C.M. Jones, D.S. Robson, and C.J. Greene. 1997. "Catch rate estimation of roving and access point surveys". N. Amer. J. Fish. Management 17:11-19.

Quinn, T.J. and R.B. Deriso. 1999. Quantitative Fisheries Dynamics. Oxford University Press. 560 pp.

Robson, D. and C.M. Jones. 1989. "The theoretical basis of an access site angler survey design". Biometrics 45:83-9

Sande, H., N. Pristo, de Groote, A., Casini, M., Jones, C.M. and A. Sundelof. 2021. Frameless—finding and refining a sampling frame for surveying recreational fisheries: lessons from estimating Swedish harvest of western Baltic cod. *ICES Journal of Marine Science* 79(4):.1217–1231, https://doi.org/10.1093/icesjms/fsac0

Steffe, A.S., J. J. Murphy and D. D. Reid (2008) Supplemented Access Point Sampling Designs: A Cost-Effective Way of Improving the Accuracy and Precision of Fishing Effort and Harvest Estimates Derived from Recreational Fishing Surveys, North American Journal of Fisheries Management, 28:4, 1001-1008, DOI: 10.1577/M06-248.1

Appendix 2. A copy of the CIE Performance Work Statement

Performance Work Statement

National Oceanic and Atmospheric
Administration (NOAA) National Marine
Fisheries Service (NMFS)

Center for Independent Experts (CIE) Program External Independent Peer Review

Puerto Rico Port Sampling and Catch Validation Project

Background

The National Marine Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, Endangered Species Act, and Marine Mammal Protection Act to conserve, protect, and manage our nation's marine living resources based upon the best scientific information available (BSIA). NMFS science products, including scientific advice, are often controversial and may require timely scientific peer reviews that are strictly independent of all outside influences. A formal external process for independent expert reviews of the agency's scientific products and programs ensures their credibility. Therefore, external scientific peer reviews have been and continue to be essential to strengthening scientific quality assurance for fishery conservation and management actions.

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards[1].

Scope

The National Marine Fisheries Service is seeking a desk review to evaluate the methods, results, statistical design, implementation, catch expansion algorithm, and recommendations of a commercial port sampling fishing survey in Puerto Rico. The study was carried out by an independent consulting firm. NMFS is requesting a review of the consultant's final report.

The port sampling fishing survey utilized a stratified random design. Stratification was based upon a combination of time, space, and fishing effort (high and low use sites). An expansion algorithm was used to calculate total catch by species.

The goals and objectives specific to the review are to:

1) Evaluate the survey design for the port sampling survey to estimate commercial catch by species.
2) Evaluate survey expansion algorithm for total catch by species.
3) Suggest future research priorities to improve the existing survey design and expansion algorithm.

The specified format and contents of the individual peer review reports are found in **Annex 1**. The Terms of Reference (ToRs) for the review of the surveys are listed in **Annex 2**.

Requirements

NMFS requires two reviewers to conduct an impartial and independent peer review in accordance with this Performance Work Statement (PWS), OMB Guidelines, and the ToRs below. The reviewers shall have expertise and experience in probability survey design and the implementation of fishery independent and/or fishery dependent

[1] https://www.whitehouse.gov/wp-content/uploads/legacy_drupal_files/omb/memoranda/2005/m05-03.pdf

probability surveys. Each CIE reviewer's duties shall not exceed a maximum of 10 days to complete all work tasks of the peer review described herein.

Tasks for Reviewers

Each CIE reviewer shall complete the following tasks in accordance with the PWS and Schedule of Milestones and Deliverables herein.

1. Pre-review Background Documents: Two weeks before the peer review, the NMFS Project Contact will send by electronic mail or make available at an FTP site to the CIE reviewers all necessary background information and reports for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the CIE on where to send documents. The CIE reviewer shall read all documents in preparation for the peer review, for example:
    a. SOW MER Consultants 2017
    b. Design of a Port Sampling Program for the U.S. Caribbean Appendix 1
    c. MER Final Port Sampling Report Appendix 2
    d. MER Port Sampling Site Descriptions Appendix 3

2. Webinar: Additionally, approximately two weeks prior to the peer review, the CIE reviewers will participate in a webinar with the NMFS Project Contact and Population Evaluation Tool Subgroup members to address any clarifications that the reviewers may have regarding the ToRs or the review process. The NMFS Project Contact will provide the information for the arrangements for this webinar.

3. Desk Review: Each CIE reviewer shall conduct the independent peer review in accordance with the PWS and ToRs, and shall not serve in any other role unless specified herein. Modifications to the PWS and ToRs cannot be made during the peer review, and any PWS or ToRs modifications prior to the peer review shall be approved by the Contracting Officer's Representative (COR) and the CIE contractor. The consultant's report includes the following documents, all are to be considered in the review.
    a. PR Port Sampling – MER – Part 1 Executive Summary
    b. PR Port Sampling – MER – Part 2 Main Body
    c. PR Port Sampling – MER – Part 3 Appendices
    d. PR Port Sampling – MER – Part 4 Site Descriptions

4. Contract Deliverables - Independent CIE Peer Review Reports: Each CIE reviewer shall complete an independent peer review report in accordance with the PWS. Each CIE reviewer shall complete the independent peer review according to required format and content as described in Annex 1. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in Annex 2.

Place of Performance

Each CIE reviewer shall conduct an independent peer review as a desk review, therefore no travel is required.

Period of Performance

The period of performance shall be from the time of award through June 2023. Each reviewer's duties shall not exceed 10 days to complete all required tasks.

Schedule of Milestones and Deliverables: The contractor shall complete the tasks and deliverables in accordance with the following schedule.

Schedule	Milestones and Deliverables
Within two weeks of award	Contractor selects and confirms the two reviewers.
No later than two weeks prior to the review	Contractor provides the pre-review documents to the reviewers.
April 2023	Each reviewer conducts an independent peer review as a desk review.
Within two weeks after review	Contractor receives draft reports.
Within three weeks of receiving draft reports	Contractor submits final reports to the Government.

Applicable Performance Standards

The acceptance of the contract deliverables shall be based on three performance standards:

(1) The reports shall be completed in accordance with the required formatting and content (2) The reports shall address each ToR as specified (3) The reports shall be delivered as specified in the schedule of milestones and deliverables.

Travel

Since this is a desk review, travel is neither required nor authorized for this contract.

Restricted or Limited Use of Data

The contractors may be required to sign and adhere to a non-disclosure agreement.

NMFS Project Contact

Kevin McCarthy
Supervisory Research Fish Biologist NMFS
Southeast Fisheries Science Center 75 Virginia
Beach Drive, Miami FL 33149
Kevin.J.McCarthy@noaa.gov
305-361-4492

Annex 1: Peer Review Report Requirements

1. The report must be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether or not the science reviewed is the best scientific information available.

2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.

3. The reviewer report shall include the following appendices:
   a. Appendix 1: Bibliography of materials provided for review
   b. Appendix 2: A copy of the CIE Performance Work Statement

Annex 2: Terms of Reference for the Peer Review

Puerto Rico Port Sampling and Catch Validation Project

1. Evaluate the statistical design and implementation of the port sampling survey to estimate commercial catch by species and the methods used to develop the design.
2. Evaluate survey expansion algorithm for total catch by species.
3. Evaluate the approach for estimating total catch in light of other approaches that utilize CPUE and estimates of total effort.
4. Evaluate recommendations for future work as provided in the survey report. Provide any additional recommendations for future research/improvements for the survey design and expansion methods.