

Science, Service, Stewardship



New MRIP Estimation Method for Use with MRFSS Data

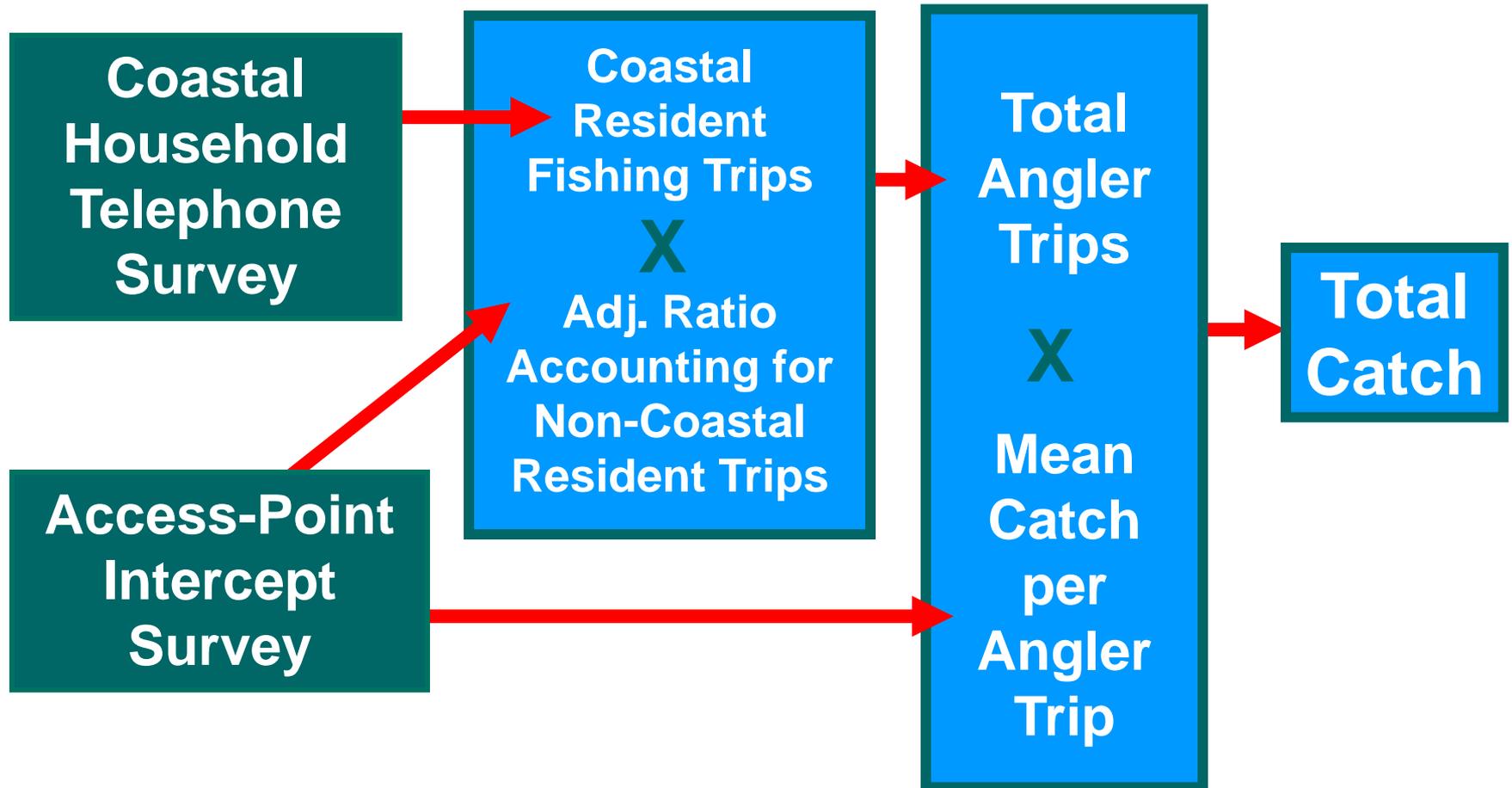
Dave Van Voorhees
Economics & Social Sciences Workshop
July 26, 2011

**NOAA
FISHERIES
SERVICE**

NOAA



MRFSS Telephone-Access Design





NRC Criticisms of MRFSS

- Estimation methods do not account for complex sampling design of the Access Point Angler Intercept Survey
 - Sampling design is stratified, multi-stage cluster sampling
 - Estimation assumes simple random sampling
- APAIS sampling design allowed flexibility to increase productivity, but ignored possible impacts on estimation
 - Flexibility increased complexity and risk of sampler errors
- Potential for bias in the estimates and estimated precision



NRC Recommendations

- Weighted Estimation:
 - Determine sample inclusion probabilities of intercepted angler fishing days.
 - Use inclusion probabilities to calculate “sampling weights”
 - Apply “sampling weights” in the estimation process.
- Eliminate flexibility in the Sampling Design:
 - Fixed design will make it easier to determine inclusion probabilities and proper sampling weights
 - Reduce the risk of sampler errors



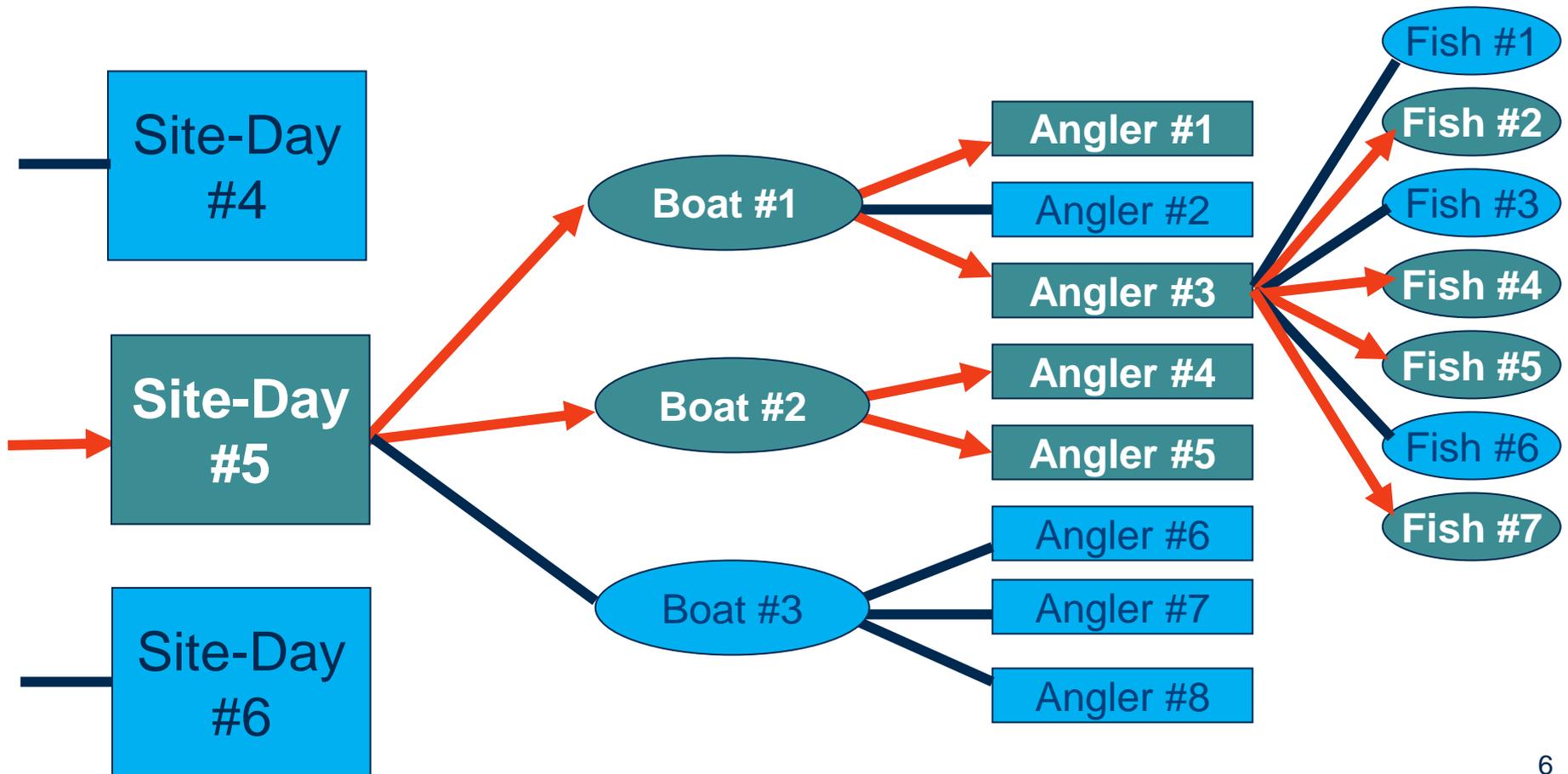
Weighted Estimation

- Sampling design must be taken into account
 - Stratification by
 - Fishing mode
 - State
 - Month
 - Day type
 - Multi-stage cluster sampling
- Inclusion probabilities must be calculated at each stage of sampling within each sampling stratum



Multi-Stage Cluster Sampling

Private Boat Angler Trips





Multi-Stage Cluster Sampling Design

- Primary stage – selection of site and day (PSU)
- Shore Fishing:
 - Secondary stage – selection of angler fishing trips (SSU)
 - Tertiary stage – selection of fish in angler’s catch (TSU)
- Private Boat or Charter Boat Fishing:
 - Secondary stage – selection of boat fishing trips (SSU)
 - Tertiary stage – selection of anglers on boat trip (TSU)
 - Quarternary stage – selection of fish in angler’s catch (QSU)

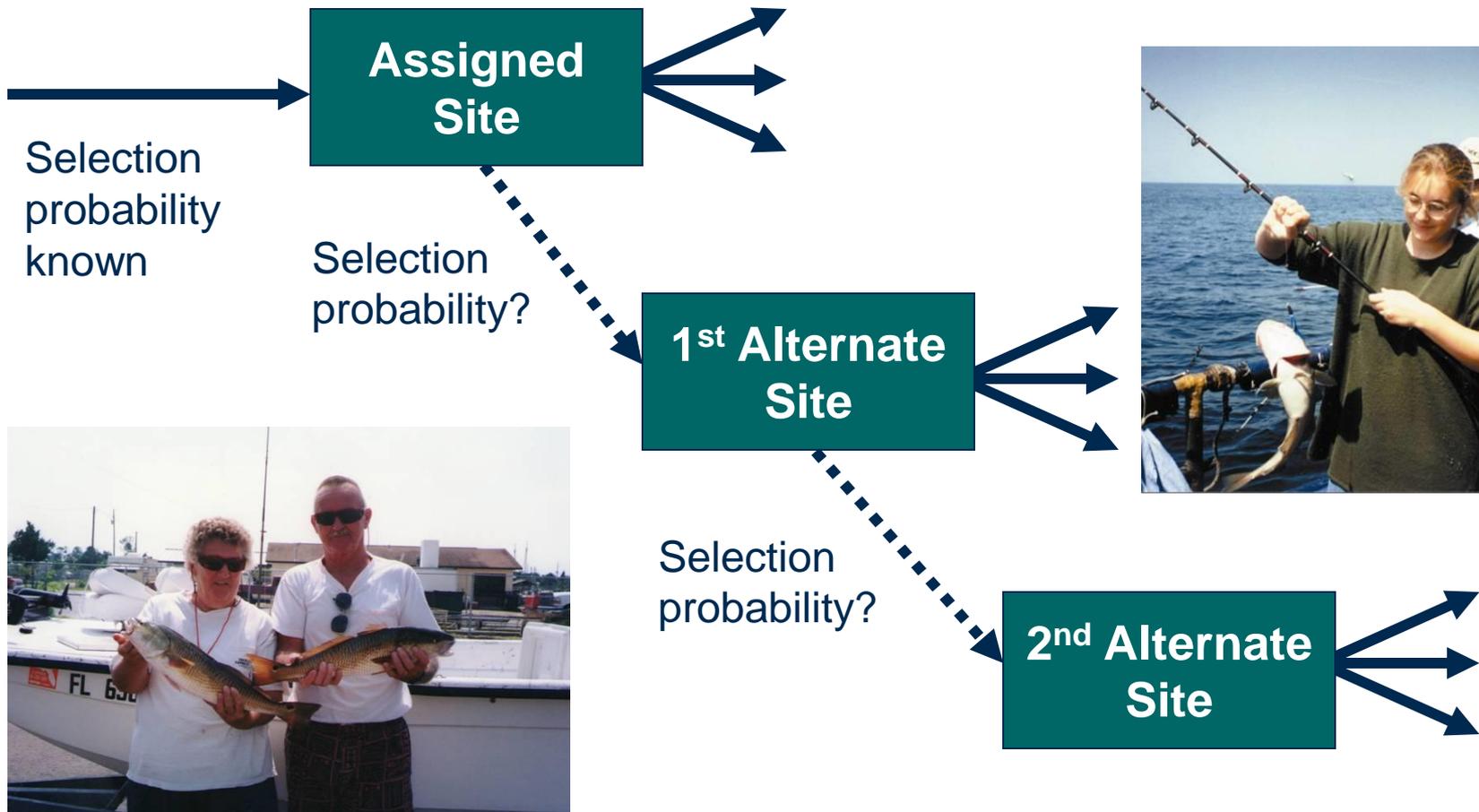


Sample Inclusion Probabilities 1st Stage: Site-Day Sampling

- Sites selected as “primary” sites
 - Unequal probability sampling
 - Probability based on estimated site fishing pressure
 - Angler trips intercepted on site-days with higher probability of selection need to be “weighted down”
 - PSU sampling weights easy to calculate
- Sites selected as “alternate” sites
 - Selection probabilities unknown, but needed to determine total probability of selection for each site
 - Needed for determining total PSU sampling weights



Alternate Site Sampling Unknown Selection Probabilities?





Sample Inclusion Probabilities 2nd Stage: Cluster Sampling

- Intercepted trips are only subset of entire cluster of returning trips during time spent on site
 - Selected subsample must represent the entire site-day
 - Selected boat or angler trips must be “weighted up”
 - SSU weight should be inverse of sampling fraction at site-day level
- Time spent on site is only a portion of the whole day
 - Time slice sample must represent the fishing trips occurring over 24 hours for the sampled site-day.
 - Need count of trips for full 24 hours to calculate the right sampling fraction.



Sample Inclusion Probabilities 3rd Stage: Cluster Sampling

- Interviewed private or charter boat anglers may only be a subset of the anglers who fished on an intercepted boat trip
 - Selected subsample of anglers must represent the entire boat trip
 - Selected angler trips need to be “weighted up”
 - TSU weight should be inverse of sampling fraction at boat trip level



MRFSS Estimation “The Old Way”

$$\hat{Y} = \sum_k y_k / n$$

$$n = \sum_h^H \sum_i^{n_h} \sum_j^{b_{hi}} n_{hij}$$



MRIP Weighted Estimation “The New Way”

$$\hat{Y} = \sum_h \frac{X_h}{X_{\bullet}} \hat{Y}_{1h}$$

Population Mean

$$= \sum_h \frac{X_h}{X_{\bullet}} \sum_i \frac{\pi_{hi}^{-1}}{\sum_i \pi_{hi}^{-1}} (X_{hi} \hat{Y}_{2hi})$$

Substitute \hat{Y}_{1h} by PSU mean

$$= \sum_h \frac{X_h}{X_{\bullet}} \sum_i \frac{\pi_{hi}^{-1}}{\sum_i \pi_{hi}^{-1}} \left(X_{hi} \left(\sum_j \frac{X_{hij}}{X_{hi\bullet}} \hat{Y}_{3hij} \right) \right)$$

Substitute \hat{Y}_{2hi} by SSU mean

$$= \sum_h \frac{X_h}{X_{\bullet}} \sum_i \frac{\pi_{hi}^{-1}}{\sum_i \pi_{hi}^{-1}} \left(X_{hi} \left(\sum_j \frac{X_{hij}}{X_{hi\bullet}} \left(\frac{\sum_k y_{hijk}}{n_{hij}} \right) \right) \right)$$

Substitute \hat{Y}_{3hij} by TSU mean



MRIP Project Team for Developing Weighted Estimation

- Jay Breidt, Ph.D. – Colorado State University
- Jean Opsomer, Ph.D. – Colorado State University
- Han-Lin Lai, Ph.D. – NMFS
- Dave Van Voorhees, Ph.D. – NMFS
- John Foster - NMFS



Challenges for Weighted Estimation: Alternate Site Sampling Weights

- Probability of site selected as alternate site?
 - Not known directly from a formal sample draw process
 - Contingent on:
 - Proximity to selected primary site
 - Activity at selected primary site
- Modeling approach used:
 - Historical frequency of alternate site visits used to model alternate site selection probabilities
 - “Pseudo-weights” approximated for alternate site-day samples



Challenges for Weighted Estimation: Boat Trip Cluster Sizes

- No counts made of boat trips missed while on site
 - Counts of missed angler trips made and recorded
 - Counts of anglers who fished together on same boat were recorded for intercepted angler trips
 - Therefore, possible to estimate mean number of angler trips per boat trip
- Estimated total count of boat trips missed based on those available counts

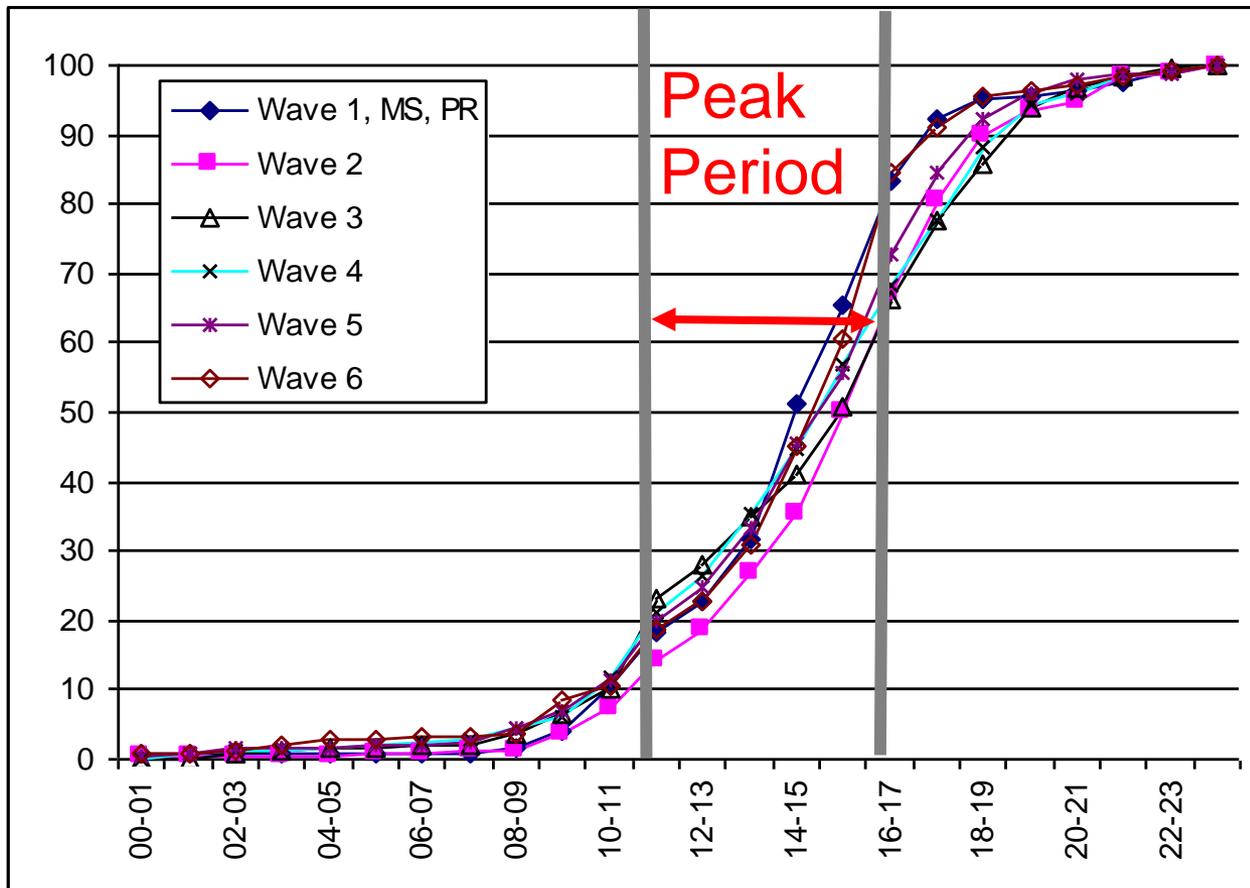


Challenges for Weighted Estimation: Time Slice Sampling

- Site-day assignments did not cover a whole day
 - Time period of sampling usually during peak period
 - Variable duration of sampled time period
 - Need to expand counts of boat trips and/or angler trips to estimate 24-hour counts
- Modeling approach used
 - Historical Telephone Survey data on reported return times of fishing trips
 - Used this data to determine appropriate expansion factors for sampled time slices



Time Slice Sampling





Challenges for Weighted Estimation: Alternate Mode Sampling

- Alternate mode angler trip intercepts
 - Opportunistic sampling not based on known probabilities for the mode
 - Difficult to know how to weight such intercepts
 - Modeling approaches considered, but too complex
 - Comprise less than 15% of the total intercepts in any given fishing mode
- Decided not to use alternate mode intercepts in the weighted estimation.



Weighted Estimation in Summary

- Selection probabilities used to weight data
 - Assigned primary site-day probabilities known
 - Alternate site probabilities approximated
- Multi-stage cluster sampling design taken into account
 - Used available data on cluster sizes at each stage
 - Expanded peak activity period counts to estimate total 24-hour counts for each sampled site-day
- Eliminated opportunistic sampling of fishing trips in other modes



Mixed Approach

Design-based and Model-based Components

- Design-based adjustments are textbook best statistical practices
 - Used selection probabilities to weight data
 - Accounted for multi-stage cluster sampling
- Model-based adjustments required novel statistical procedures
 - Estimation of alternate site probabilities
 - Expansion of peak activity counts to full day



Independent Peer Review

- Both standard and new estimation methods were subjected to rigorous review
 - Three external reviews
 - US Census Bureau
 - American Statistics Association - 2 reviewers selected by Survey Research Methods Section
- Team wrote response to external reviews and included it with the final report submitted to the MRIP Operations Team



Results of Weighted Estimation

- Point estimates of catch rates changed, and consistent direction of change observed for some species
- Point estimates of effort ratios changed, but no consistent directional patterns observed
- Estimates of the variance of point estimators increased across the board
 - The precision of catch rate estimates was over-estimated by the unweighted MRFSS approach



Potential for Bias

- Mismatch between sampling design and estimation
 - Unequal probability sampling of site-days
 - Probability proportional to site pressure
 - Unweighted estimation method
 - No weighting of data to compensate
 - High pressure sites tended to be over-represented
- Potential bias is not certain bias
 - Do angler catch rates differ between high and low pressure sites?
 - Do angler fishing targets differ between high and low pressure sites?



Sample Size and Clustering

- The sample size is the number of site-days (PSUs) included in the sample
- The sample size is **not** the number of angler trips intercepted within the selected site-day
 - Catch or other characteristics of angler trips tend to be similar within the same site-day cluster
 - One random draw of 30 angler trips from the same site-day is not as informative as getting 1 trip from each of 30 different randomly selected site-days.

Science, Service, Stewardship



Implementation of Weighted Estimation Method

Dave Van Voorhees
Economics and Social Sciences Workshop
July 26, 2011

**NOAA
FISHERIES
SERVICE**

NOAA



Implementation Plan

- QC and preparation of legacy MRFSS data:
 - Checking of new key data elements
 - Correction of identifiable errors
- Preparation of new data structures:
 - Integration of key elements across different datasets
- Preparation and testing of estimation programs
 - Implement and test new estimation components
- Development of comparison tools:
 - “New” MRIP estimates vs. “old” MRFSS estimates



New Key Data Elements

- New estimation method uses data elements not previously used in the estimation process.
 - Site selection probabilities
 - Based on site pressures stored in Master Site Register
 - Stored in deliverable site-day assignment draw files
 - Time slice sampled at each site
 - Stored in assignment summary files
 - Counts of missed angler trips at each site
 - Stored in assignment summary files



Integration of Key Data Elements

- Integration of Key Intercept Survey Datasets
 - Interview data files (I1-I6)
 - Type 1 data (I1) – angler and trip data
 - Type 2 data (I2) – unobserved catch data
 - Type 3 data (I3) – observed catch data
 - Type 4 data (I4) – linkage of mixed group catches
 - Type 6 data (I6) – linkage of anglers on same boat trip
 - Assignment summary data file (IA):
 - Summary data for each site visited
 - Site-day assignment file:
 - Listing of primary site-day assignments
 - Master site register:
 - Site-day sample frame with pressure estimates



MRIP Re-Estimation Project

Streamlining of Estimation

- Sequencing of New Estimation Programs
 - Telephone survey estimates
 - Design-based
 - Intercept survey estimates
 - Model-based components
 - Design-based components
 - Combined estimates of effort and catch



MRIP Re-Estimation Project

Tools for Reviewing New vs. Old

- Programs to compare estimates
 - New weighted estimates vs. prior MRFSS estimates
 - New weighted estimates vs. new unweighted estimates
- Webtool for comparisons
 - Facilitates general review process
 - Facilitates examination of specific changes

Science, Service, Stewardship



New Access Point Sampling Design

Dave Van Voorhees
Economics & Social Sciences Workshop
July 26, 2011

**NOAA
FISHERIES
SERVICE**



New Access Point Sampling Design Pilot Study in North Carolina

- Sampling frame improvements:
 - Standardized site cluster units
 - Unit = Cluster of proximate 1-3 sites
 - Total pressure of cluster unit used for PPS sampling
 - Selection probability for each cluster unit is known
 - PSU sampling weight easily calculated
 - Order of sites within selected cluster randomized and fixed for a given interviewing assignment
 - No “alternate site” or “alternate mode” interviews



New Access Point Sampling Design Pilot Study in North Carolina

- **Sampling stratified by time of day**
 - 6-hour time blocks
 - Covering all time periods
 - Including fishing trips ending at night
 - Including fishing trips ending off-peak during daytime
 - Consistent timeframe for sampling
 - No need to expand cluster counts to estimate 24-hour counts



New Access Point Sampling Design Pilot Study in North Carolina

- Maximizing number of site-days sampled
 - Quota-based sampling of angler trips eliminated
 - Set number of site-day assignments to be completed
- Sampling will be set at boat trip level in boat modes
 - Eliminates a stage of sampling
 - Still possible to subsample angler trips if needed for other purposes



New Access Point Sampling Design Pilot Study in North Carolina

- Estimation is totally design-based
- No need for model-based estimation components