

# Effectiveness of marine reserves for large-scale fisheries management

Martin D. Smith, Junjie Zhang, and Felicia C. Coleman

**Abstract:** As more no-take marine reserves are established, the importance of evaluating effectiveness retrospectively is growing. This paper adapts methods from program evaluation to quantify the effects of establishing a marine reserve on fisheries using fishery-dependent data. The approach analyzes the effects of a policy at the individual vessel level and accommodates the coarse spatial resolution of fishing logbooks. It illuminates implicit assumptions in previous retrospective analyses of marine reserves that are unlikely to hold for large-scale fisheries. We illustrate the empirical model with an application to the Gulf of Mexico reef-fish fishery. Isolating the effects of reserves requires a full accounting of multiple gear production technologies, heterogeneity in vessel captain skill, spatial heterogeneity of fish stocks, seasonal patterns in abundance, the effects of coexisting management policies, and the possibility that the harvest sector anticipates reserve establishment. We find that the effect of two recently established marine reserves on catch is negative and trending downward, though the reserves have only been in place for 4.5 years.

**Résumé :** À mesure que les réserves marines sans récolte deviennent de plus en plus nombreuses, il est d'autant plus important de pouvoir évaluer leur efficacité de façon rétrospective. Notre travail adapte des méthodes d'évaluation de programmes afin de quantifier, à l'aide de données reliées à la pêche, les effets de l'établissement d'une réserve marine sur les pêches commerciales. Notre méthodologie analyse les effets des politiques au niveau du navire individuel et s'accommode de la résolution spatiale grossière obtenue à partir des livres de bord de pêche. Elle précise certaines des présuppositions implicites dans les analyses rétrospectives antérieures de réserves marines, qui ne valent vraisemblablement pas pour les pêches à grande échelle. Nous illustrons notre modèle empirique en l'appliquant à la pêche commerciale sur les récifs du golfe du Mexique. Pour isoler les effets des réserves, il faut tenir un compte complet des technologies multiples de production des engins de pêche, de l'hétérogénéité des compétences des capitaines de bateau, de l'hétérogénéité spatiale des stocks de poissons, des patrons saisonniers d'abondance, des effets de politiques de gestion parallèles et de la possibilité que le milieu de la pêche ait prévu l'établissement de la réserve. Nous trouvons que les effets de deux réserves marines récemment établies sur les récoltes sont négatives et ont tendance à diminuer, bien que ces réserves ne soient en place que depuis 4,5 années.

[Traduit par la Rédaction]

## Introduction

The dominant paradigm in fisheries policy is undergoing a transformation from one that views regional fish stocks as single entities towards one that increasingly acknowledges both the importance of these stocks as integral components of ecosystems and the importance of spatial heterogeneity and spatial processes in management (Ecological Society of America (ESA) 1998; National Marine Fisheries Service (NMFS) 1999). This transformation has led to intense interest in establishing marine reserves around the world (ESA 2003). Marine reserves or no-take areas, a particular type of marine protected area in which exploitation is prohibited, have long been an important part of marine conservation.

Considerable attention in recent years has focused on using reserves explicitly to manage commercial fishery yields (Bohnsack 1998; Carr and Reed 1993; National Research Council 2001). As more reserves are established and empirical data become available, it is critical to evaluate their performance and continuously assess the overall potential of reserves to address sustainability and production problems in fisheries management.

The theoretical basis for establishing marine reserves for fisheries management is straightforward. A marine reserve, by eliminating fishing pressure in a particular area, allows biomass in the closed area to rebuild, through both growth and reproduction, and eventually spill out of the reserve into the surrounding open areas (Polacheck 1990). For many spe-

Received 1 April 2005. Accepted 15 September 2005. Published on the NRC Research Press Web site at <http://cjfas.nrc.ca> on 7 December 2005.  
J18635

**M.D. Smith.**<sup>1</sup> Nicholas School of the Environment and Earth Sciences and Department of Economics, Duke University, Box 90328, Durham, NC 27708, USA.

**J. Zhang.** Nicholas School of the Environment and Earth Sciences, Duke University, Box 90328, Durham, NC 27708, USA.

**F.C. Coleman.** Department of Biological Science, Florida State University, Tallahassee, FL 32306-1100, USA.

<sup>1</sup>Corresponding author (e-mail: [marsmith@duke.edu](mailto:marsmith@duke.edu)).

cies of fish, larger individuals generate more offspring than their smaller counterparts, leading to greater reproductive success (Bohnsack 1998; Koops et al. 2004). There is also a growing body of evidence that older fish have more viable offspring than younger fish (Berkeley et al. 2004a, 2004b). Thus, a reserve could produce significant biological returns (Smith 2004) such that long-term gains from the reserve outweigh losses resulting from the reduction in fishing area (Roberts and Polunin 2001), as long as the reserve captures source rather than sink areas (Crowder et al. 2000). The reserve design literature asks questions about the ecological, oceanographic, and economic circumstances under which reserves may benefit fisheries through increased or stabilized yields (Gerber et al. 2003; Smith and Wilen 2003; Sanchirico 2004). While these papers suggest *ex ante* when we might expect a reserve to generate catch increases, they provide little guidance on how to measure the short-run and long-run changes in catch *ex post*.

In contrast, we ask how fisheries scientists would know if reserves are effective once they are established. Previous researchers have acknowledged that retrospective analyses are important but that few such analyses exist for marine reserves (Hilborn et al. 2004; Sale et al. 2005). What our study highlights is that responses of the harvest sector cast doubt on the ability of researchers to implement a before–after control–impact pair (BACIP) marine reserve experimental design.

We view a marine reserve as a policy treatment that may or may not affect fishing in the broader zone in which it is placed. The goal then is to measure the treatment effect. However, fishing pressure outside the reserve is not experimentally controlled, and this lack of control complicates the problem of measuring the treatment effect. For management of fisheries yields, the treatment site is an area near a reserve where economic benefits following from enhanced productivity could materialize for fisheries, whereas the reference (or control) site is an area far from the reserve where the fishery is presumably unaffected by the establishment of a reserve. When increasing catch is a goal of management, the definition of a reference site is obscured by the fact that fishing pressure at that site will not necessarily be the same before and after reserve formation. As a consequence, non-experimental methods are necessary for retrospective analyses, and a marine reserve for fisheries management is more appropriately analyzed as a social program rather than as a randomized experimental trial.

When a marine reserve is established and there are changes in measured fishery outcomes, determining the effectiveness of the reserve amounts to determining which of these changes are attributable to the policy and which are attributable to other factors. For judging a reserve, Gell and Roberts (2003, p. 453) suggest a “‘gold standard’ of higher overall catches with reserves than without, in spite of a reduced fishing area”. Though we agree with the sentiment of this gold standard, we seek to make it more explicit so that reserve evaluations are robust to a behavioral response of the fishing fleet. Ideally, determining the effectiveness of a reserve for fisheries management would test whether the reserve causes harvestable biomass to increase. This means that harvestable biomass outside the reserve needs to increase enough to offset what is closed off within the reserve.

Because stocks are unobservable (latent) in fishery-dependent data, we use catch as an indicator of harvestable biomass.

Higher overall catches could indicate an effective reserve for fisheries management, but higher overall catches are neither necessary nor sufficient for effectiveness. Observed higher overall catch rates do not necessarily mean that the reserve is effective, and observed lower catch rates do not necessarily mean that it is a failure; conditional catches are what matter, and the researcher must control for possible selection effects at treated and reference sites. For instance, higher total catches could reflect a change in fleet composition to higher skilled fishermen or larger boats near the reserve. This case would not indicate an effective reserve in our model. Lower total catches could mean the opposite and suggest that the reserve is not necessarily a failure in our model.

To be more specific, evaluating effectiveness requires answering the following hypothetical question: did a particular vessel deploying a particular type and quantity of gear catch more in an area near the reserve after formation of the reserve than it would have caught had the reserve never been established? We submit that a positive answer to this question should be the gold standard for evaluating whether a reserve generates fishery benefits. Addressing this type of hypothetical question is central in program evaluation of environmental policy (Benjamin and Coglianese 2005), but this literature has not yet been applied to fisheries policy.

The intuition for our proposed gold standard is that if nothing changes in the fishery except the policy change of forming a reserve, then fishermen would be catching more with the policy than without it. Of course, we cannot hold everything constant in a physical sense in the way that we would hold features of the system constant in a randomized experiment, nor can we directly observe the counterfactual case of what would have happened had the reserve not been established. In practice, we build up the counterfactual statistically by analyzing whether a particular vessel conditional on a range of factors would catch more near the reserve (in the treatment area) vs. far from the reserve (in the reference area). Our model does have aggregative properties in the sense that total catches from a reserve with a positive treatment effect would increase if everything else stays constant. That is, if there are no behavioral adjustments or selection effects, our model reaches the same conclusion about the effectiveness of a reserve as previous models that focus on overall catches.

To add to intuition, consider a hypothetical case from labor economics in which a voluntary job-training program is initiated at the onset of a recession. We would like to test whether the program decreases unemployment. Suppose further that the actual unemployment rate increases after the program for individuals enrolled in the training program (the treatment group) and for individuals not enrolled (the control group), but the increase is greater for the control group. If we simply measure total unemployment before and after the program, we would conclude that the program is a failure. If we compare the increase in unemployment between treated and control groups, we conclude that the policy is a success. However, this conclusion could be spurious if people who enrolled in the training program were different from those

who did not enroll; enrollees might have more education (an observable characteristic) or simply be more motivated (an unobservable characteristic). This is the selection problem that plagues BACIP for marine reserve design.

To isolate the policy effects of marine reserves for large-scale fisheries management, our approach is to consider whether individual fishery participants receive “treatment” depending on whether they fish near or far from the newly established reserve. This allows us to consider how measuring the treatment effect of a marine reserve depends on the composition of the treatment and comparison groups, i.e., the individuals fishing near and far from a reserve. We develop a descriptive panel data model with an application to two recently established marine reserves in the Gulf of Mexico reef-fish fishery. By estimating an individual-level model, incorporating heterogeneity, and conditioning on the complexity of the pre-existing fishery management system, our model is more general than other retrospective analyses of marine reserve formation. Specifically, we show that isolating the effects of marine reserves on fisheries requires knowledge of the spatial scope of reserve spillovers, as well as a full accounting of multiple gear production technologies, heterogeneity in vessel captain skill, spatial heterogeneity of the fish stocks, seasonal patterns in abundance, the effects of coexisting management policies, and the possibility that the harvest sector anticipates the establishment of a reserve. All of these factors confound the problem of identifying BACIP for marine reserve design. Even if researchers are not interested in the fishery-related effects of reserves per se and they only seek to measure the ecological effects, the impact of the harvest sector on reference sites cannot be ignored. Our empirical results do not provide a definitive answer as to whether the marine reserves that we analyze have benefited fisheries, but they illustrate a number of important empirical issues that have not been addressed in previous studies. Our model ultimately highlights the types of data that future researchers will need to collect to analyze the effects of establishing reserves retrospectively.

Below we describe our empirical setting: the Gulf of Mexico reef-fish fishery. We then develop a descriptive panel data model to test the fishery effects of a reserve. We show that our model nests the before–after retrospective analyses of catch per unit effort (CPUE) that are currently in the literature. We subsequently present results of our model using data from Gulf of Mexico fisheries logbooks. Finally, we discuss our results and suggest directions for future research.

## Materials and methods

### Empirical setting — the Gulf of Mexico reef-fish complex

The Gulf of Mexico reef-fish complex presents managers with enormous challenges. There are 62 reef-fish species commercially harvested, including 11 grouper species and 10 snapper species, with a wide range of gears. The most common gears are hook and line (including handlines, electric bandit reels, buoy gear, and conventional rod and reel), bottom long line, and traps (although traps are being phased out completely over the next couple of years). Existing management includes limited entry (there are currently approximately 1200 federal commercial Gulf of Mexico reef-fish

permit holders), size limits, trip limits, season closures, quota management, and more recently, marine reserves. Only a fraction of permitted vessels regularly engage in commercial reef-fish fishing. For example, 25% of vessels account for 75% of fishing trips over the 1993–2002 period.

Economically important species such as gag (*Mycteroperca microlepis*), scamp (*M. phenax*), and red grouper (*Epinephelus morio*) could gain some long-run biological benefits from marine reserves. These species are long-lived, slow-growing protogynous hermaphrodites (Coleman et al. 2000). Protogynous hermaphrodites mature first as females and then transform to males later in life. Because fishing tends to select for larger individuals, it tends to select for males. This selection reduces male-to-female sex ratios (Coleman et al. 1996; McGovern et al. 1998), and some traditional fisheries management tools, particularly size limits, may exacerbate this effect. For Gulf of Mexico gag, the percentage of males from the 1970s to the 1990s has declined from 17% to 2% (Coleman et al. 2000). The same decline does not appear in red groupers, which do not aggregate to spawn (Coleman et al. 1996).

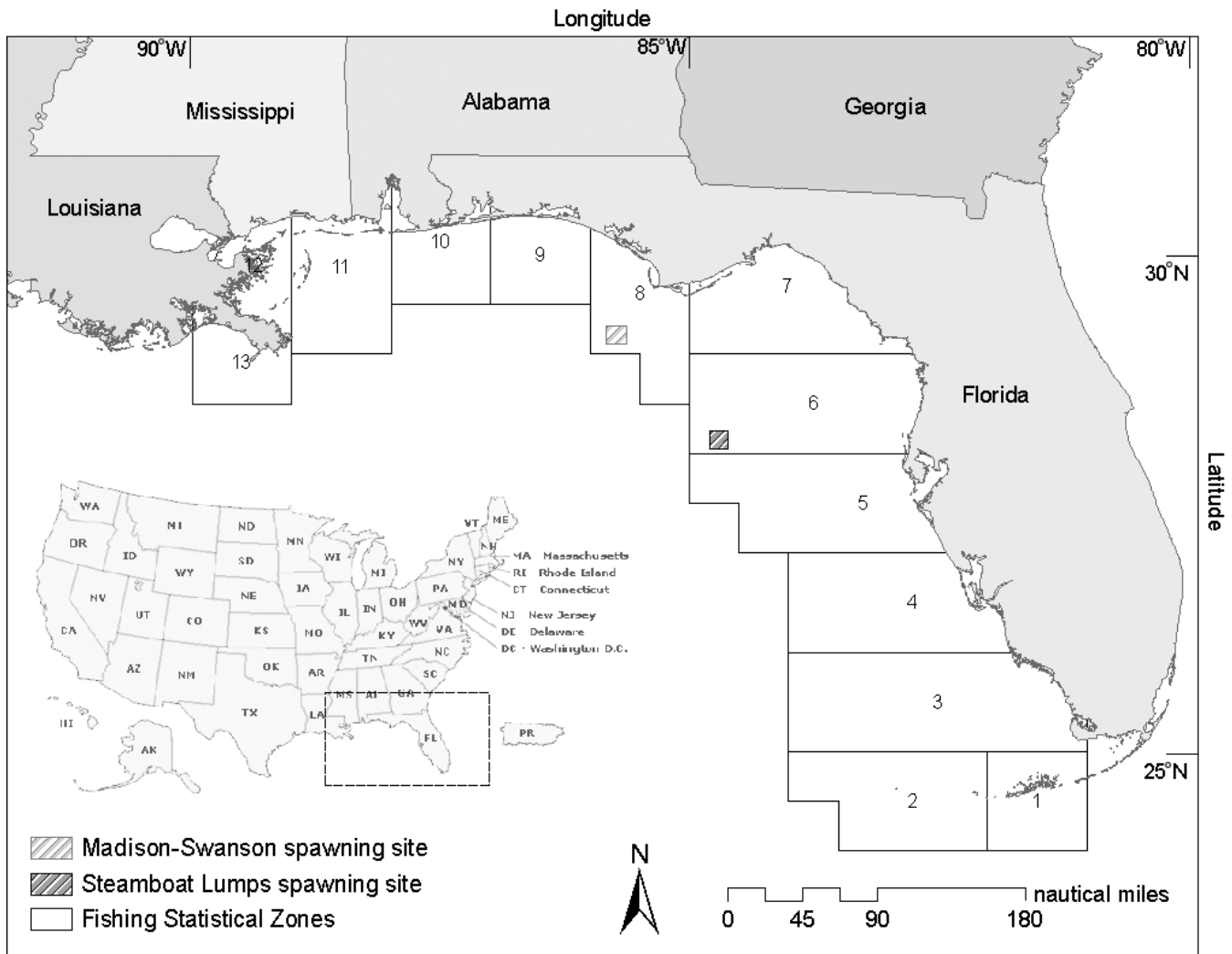
The two marine reserves in the Gulf of Mexico that we study (Madison–Swanson Marine Reserve and Steamboat Lumps Marine Reserve) went into effect in June 2000 to address concerns about this skewed sex ratio. They were announced to the fishing community 1 year before being established. They were authorized as experimental reserves with a sunset of 4 years and were recently reauthorized for an additional 6 years, based exclusively on biological data within and outside of each reserve and anecdotal information about the fishery outside each reserve. To date, there have been no systematic evaluations of these reserves as management tools, and understanding their performance will be critical for future reauthorizations.

The reserves are located in two of the 13 distinct NMFS fisheries statistical zones in the Gulf of Mexico (Fig. 1). These two zones represent the heart of the gag fishery in the Gulf of Mexico. Madison Swanson is located in zone 8, comprising 115 square nautical miles (n.mi<sup>2</sup>) of this 9570 n.mi<sup>2</sup> zone, whereas Steamboat Lumps is located in zone 6, comprising 104 n.mi<sup>2</sup> of this 8100 n.mi<sup>2</sup> zone. Each reserve captures 1.2% and 1.3%, respectively, of the total area in the NMFS statistical zone within which they occur. There is, thus, substantial fishable area left open within each of the statistical zones, as well as elsewhere on the West Florida Shelf. However, percentage area is misleading in a fisheries context, because not all ocean bottom is alike. The reserves in question are located in deep water along the continental shelf edge and contain distinct patch reef formations where reef fish aggregate in general and many grouper species aggregate to spawn (Koenig et al. 2000).

There is not a one-to-one relationship between NMFS fishing zones and what we might reasonably deem a patch from a bioeconomic perspective. The spatial resolution of the logbook data is extremely coarse, which is typical, and the NMFS fishing zones are spatial aggregates of the underlying biological and economic processes. Herein lies one of the more challenging empirical problems that we address in our model of individual vessel catch.

Complete fishing logbook data exist for all reef species in the Gulf of Mexico from 1993 through 2004. Thus, there are

**Fig. 1.** Gulf of Mexico reef no-take marine reserves established in June 2000. Marine reserves are contained within National Marine Fisheries Service statistical zones.



substantial data before and after the policy change to assess reserve performance. We cut the data after October 2004 because some logbook records for 2004 may not have been submitted or processed in the most recent database that we received in February 2005. This leaves us with 142 months of data.

**A panel data model of marine reserves for fisheries management**

Before–after comparisons of catch rates most closely parallel the program evaluation literature in economics. A reserve is a policy treatment. The general area that contains the reserve is thus a treatment area (the area in which reserve effects are most likely to be realized), and the goal is to measure the treatment effect against some reference (an area remote enough to be unaffected by the reserve). We considered the two NMFS statistical zones that contain marine reserves (zones 6 and 8) as treatment zones and the remaining 11 as reference zones. Implicitly, our panel model assumes that spillovers and larval export effects materialize exclusively in those statistical zones containing the marine reserves.

To make the hidden assumptions of retrospective analyses more transparent and inclusive, we propose a more general model that (i) accounts for exogenous effects on catch typically absent from other models and (ii) accounts for selection bias using selection on unobservables (Heckman and Hotz 1989). Let *i* index individual fishermen or fishing vessels, *j* index fishing zone, *k* index fishing gear type, *m* index month of the year, and *t* index time (in days). Denote aggregate reef fish catch as *h*, fishing effort as *E* — where effort is defined specific to the gear type (details are available from the authors) — and biomass as *X*, and define the catchability parameter as *q* and a fishery production function exponent as  $\alpha$ . An individual’s catch on a given day in a particular area is

$$(1) \quad h_{i,j,k,t} = q_{i,j,k,m}(E_{i,j,k,t})^{\alpha} X_{j,t}$$

This form allows the curvature of the catch function to depend on gear type and allows for heterogeneity in catchability across individual vessels, fishing areas, and gear types. It also allows catchability to vary by month, which may be particularly important for species that aggregate seasonally to spawn.

We do not observe the fish stock directly and, as in other retrospective analyses, must make an assumption about how it evolves over time in each location. In the absence of a policy treatment, we assume that there is an initial stock and a trend that are both spatially specific:

$$(2) \quad X_{j,t} = \tilde{X}_j \exp\{r_{j,t}t\}$$

where  $r$  is the rate of change, which could be positive or negative depending on bioeconomic conditions. Now suppose that at  $t = T$ , a marine reserve is introduced. We define  $Z$  as an indicator variable that denotes when a reserve is in effect such that

$$(3) \quad Z_{j,t} = \begin{cases} 1 & \text{if there is a reserve in } j \text{ and } t \geq T \\ 0 & \text{if no reserve or } t < T \end{cases}$$

In a large-scale commercial fishery, we typically do not observe fishing activity with fine spatial resolution. For many federally managed fisheries in the United States, we observe fishing at the scale of a National Marine Fisheries Service statistical zone, which can be on the scale of 10 000 n.mi<sup>2</sup> (Fig. 1). Thus, we consider the placement of a marine reserve within the spatial context of these statistical zones.

If a reserve generates biomass through either spillovers or larval export, it might appear to stimulate growth in the zone that contains the reserve. However, because the reserve also closes off part of the fishing area, it might appear to affect the production function for fishermen. Whether this effect reduces catch in a measurable way is an empirical question that can be tested with our model by introducing  $\phi_1$ , a parameter that scales catch. Under the null hypothesis of no instantaneous effect,  $\phi_1 = 1$ . Another important empirical question is whether the reserve stimulates catch in the surrounding zone. This effect could only happen dynamically if stocks rebuild within the reserve and spill into open fishing areas through adult emigration or larval export that repopulates open fishing areas. We approximate this effect with a parameter  $\phi_2$  that introduces a change in the trend in eq. 2. Under the null hypothesis that the reserve has no effect,  $\phi_2 = 0$ . Substituting the reserve effects with eqs. 3 and 2 into eq. 1, we have

$$(4) \quad h_{i,j,k,t} = \phi_1^{Z_{j,t}} q_{i,j,k,m}(E_{i,j,k,t})^{\alpha_k} \tilde{X}_j \exp[(r_{j,t} + \phi_2 Z_{j,t})t]$$

In essence, this model allows the reserve to influence biomass in the larger zone within which the reserve is embedded, but scaling the harvest function only permits harvesting in the open portion of the zone. This setup is consistent with many logbook data sets for which no-take fishery reserves do not correspond to an entire fishing zone.

At this juncture, it is important to reconsider how the model in eq. 4 relates back to theory of reserve design based on metapopulations. In our model, each fishing zone represents an entire metapopulation comprised of many patches. Recall that the motivation here is that we do not observe fishing activity at the subzone level, i.e., at the patch level. Thus, we have no means by which to parameterize dispersal among patches and only can generate qualitative predictions about the metapopulation as a whole. A reserve closes down one or several patches, but the zone is large enough to contain all of the relevant patches. This is a limitation of our ap-

proach but one that can be dealt with in the future as more spatially explicit data are collected.

A challenge for empirical implementation of our model is that policies typically do not go into effect for some period after they are announced. We might expect that fishing vessels would react to the announcement of a reserve at  $t = \tau$  that will not go into effect until  $t = T$ . We define an instantaneous announcement effect ( $\rho_1$ ), a trend announcement effect ( $\rho_2$ ), and the indicator variable  $A$

$$(5) \quad A_{j,t} = \begin{cases} 1 & \text{if reserve announced in } j \text{ and } \tau \leq t < T \\ 0 & \text{otherwise} \end{cases}$$

Adding announcement effects into eq. 4, we have

$$(6) \quad h_{i,j,k,t} = \rho_1^{A_{j,t}} \exp[\rho_2 A_{j,t} t] \phi_1^{Z_{j,t}} q_{i,j,k,m}(E_{i,j,k,t})^{\alpha_k} \times \tilde{X}_j \exp[(r_{j,t} + \phi_2 Z_{j,t})t]$$

Another challenge in evaluating reserves for large-scale fisheries management is that reserves typically do not exist independent of other policies. If none of these other policies changes in the sample period, then estimating a model like the one in eq. 6 would be sufficient. However, if other policies do change, then they must be accounted for. Isolating the effects of reserves is theoretically still possible, because although marine reserves affect those individuals operating closest to them, most other management approaches, such as size limits, trip limits, and seasonal restrictions, affect the entire fishery. This allows us to specify an instantaneous effect, denoted by  $\mu_1$ , and a trend, denoted by  $\mu_2$ . Because the timing of the other policy may or may not coincide with the reserve, we define a separate indicator variable  $I$  for a policy change at  $t_0$ :

$$(7) \quad I_t = \begin{cases} 1 & \text{if "other" policy is in effect} \\ 0 & \text{otherwise} \end{cases}$$

Adding other policies into eq. 6 gives rise to

$$(8) \quad h_{i,j,k,t} = \rho_1^{A_{j,t}} \exp[\rho_2 A_{j,t} t] \mu_1^{I_t} \phi_1^{Z_{j,t}} q_{i,j,k,m}(E_{i,j,k,t})^{\alpha_k} \times \tilde{X}_j \exp[(r_{j,t} + \phi_2 Z_{j,t} + \mu_2 I_t)t]$$

Even in the absence of other policies, we might include a fishery-wide indicator variable as in eq. 7 because it is still possible that the model otherwise would attribute an exogenous region-wide oceanic or biological event to the reserve treatment. This effect would come through as dampening individual zone-specific trends in the nontreated zones but picking up the shock directly in the treated zones. Thus, including this term is a difference-in-differences approach. See Benneer and Coglianese (2005) for a nontechnical discussion of program evaluation methods for environmental policy. We are measuring the difference between reference and treatment groups for the difference in catch before and after the policy. To correctly identify the treatment effect, this approach requires that the posttreatment trends in the absence of the policy would have been parallel.

The model in eq. 8 is quite general but suffers from a curse of dimensionality because the catchability coefficient is individual-, zone-, gear-, and month-specific. To reduce some dimensionality, we make the following simplification:

$$(9) \quad q_{i,j,k,m} = q_i q_m q_{k,m} q_{j,m} q_{j,k}$$

We exclude a separate zone-specific catchability effect because it would not be separately identified from the baseline zone-specific stock  $\tilde{X}_j$ . Substituting eq. 9 into eq. 8 leads to

$$(10) \quad h_{i,j,k,t} = \rho_1^{A_{j,t}} \exp[\rho_2 A_{j,t} t] \mu_1^{I_t} \phi_1^{Z_{j,t}} q_i q_m q_{k,m} q_{j,m} q_{j,k} \\ \times (E_{i,j,k,t})^{\alpha_k} \tilde{X}_j \exp[(r_{j,t} + \phi_2 Z_{j,t} + \mu_2 I_t) t]$$

For the model above, we observe  $h$ ,  $E$ ,  $A$ ,  $I$ ,  $Z$ , and  $t$ . Everything else is a parameter to be estimated.

To derive an estimating equation, we take the natural log of eq. 10 and append an additive random shock  $u_{i,j,k,t}$ :

$$(11) \quad \ln(h_{i,j,k,t}) = \ln(\rho_1) A_{j,t} + \ln(\mu_1) I_t + \ln(\phi_1) Z_{j,t} \\ + \ln(q_i) + \ln(q_m) + \ln(q_{k,m}) + \ln(q_{j,m}) + \ln(q_{j,k}) \\ + \alpha_k \ln(E_{i,j,k,t}) + \ln(\tilde{X}_j) + r_{j,t} t + \rho_2 A_{j,t} t \\ + \phi_2 Z_{j,t} t + \mu_2 I_t t + u_{i,j,k,t}$$

This form is linear in all of the parameters. We are interested in comparing observed outcomes with the counterfactual case of a fishing vessel fishing in the same place at the same time without the reserve treatment. Because we cannot observe the counterfactual, we rely on data from fishing in reference sites. This introduces potential selection bias because individual effort, gear type, and zone choice might be correlated with the individual-specific error. To control for selection, we use selection on unobservables through fixed effects (Heckman and Hotz 1989) such that  $u_{i,j,k,t} = \eta_i + \varepsilon_{i,j,k,t}$ . Thus, we cannot separately identify the selection effect from the individual-specific catchability coefficient ( $\ln(q_i)$  in the estimating equation), but lack of identification in this case does not inhibit our ability to draw inferences from the rest of the parameter estimates. The model is similar to the dynamic panel data model in the program evaluation literature used to study enterprise zones (Papke 1994). Assuming that the shocks are independently and identically distributed and  $\text{cov}(\ln(E_{i,j,i}), \varepsilon_{i,j,k,t}) = 0$ , our fixed effects model can be estimated by demeaning the data and then using ordinary least squares. Note that estimating random effects would not involve demeaning the data and thus would not remove selection bias in this case. The total number of parameters depends on the number of fishing zones and gear types and unique combinations of month, gear, and zone that are identifiable.

We are now ready to compare our model with CPUE before–after analyses from the literature. We first address the curvature of the production function. Unless fishing effort before reserve establishment is identical, analyzing CPUE is equivalent to restricting all of the  $\alpha$ s to 1. If the true  $\alpha < 1$ , then a decrease in effort postreserve will produce an increase in CPUE, even if the reserve is not generating any harvest benefits. Consider, for example, that Russ et al. (2004) find a total effort decline of 46% from 1986 to 1998–2001 and a 50% increase in CPUE over the same period. If  $\alpha = 0.5$ , then this effort decline would translate into a 36% increase in CPUE even if fishable biomass did not change. Thus, when there is a behavioral change postreserve, a change in CPUE may be falsely attributed to the policy. Moreover, by restricting the  $\alpha$ s, other studies implicitly assume that all fishing gears are the same. This may be a reasonable as-

sumption in studies of small-scale artisanal fisheries, but inferences for large-scale commercial fisheries need to account for gear differences.

Beyond generality of the production function, the model in eq. 11 allows analysts to distinguish habitat and other background effects from reserve effects. Including zone-specific effects  $\tilde{X}_j$  and trends  $r_j$  captures the influence of habitat that is independent of the reserve effects. Gell and Roberts (2003) raise this issue in discussing BACIP and point out the difficulty in resolving this issue with data sets on reserves. What we are suggesting is that the richness of fishery logbook data allow for a parameterization that can separate these effects if there is sufficient spatial resolution. Also, zone-specific trends in our model control for the possibility that apparent gains following a spatial closure are actually artifacts of upward trends in catch that were developing before reserve establishment.

Our model allows for several types of heterogeneity that previous studies assume do not matter. Perhaps most important here is heterogeneity across individuals. If individual fishing skill varies or vessel catching power varies in a way that is not controlled for by observables (e.g., captain catch history, vessel length, or vessel horsepower), changes in CPUE over time could simply reflect a change in the composition of the fishing fleet. Another source of heterogeneity is seasonality in catchability. CPUE studies can address this by always making before–after comparisons at the same time of year, but modeling seasonality directly allows for use of all observations rather than excluding before (after) observations that cannot be matched with after (before) observations in the same month. By interacting gear and month, our model also allows for particular gear types having more catching power than others when there are spawning aggregations.

Finally, our model traces out the full dynamic path of reserve impacts, not just the change at some arbitrary point in the future. This allows for before–after comparisons at each time step for whether a reserve has generated net benefits. For example, if the initial loss from establishing a reserve is large and rebuilding takes a long time, simple before–after comparisons may conclude that the reserve has not benefited fisheries when it simply has not had enough time. Our model allows us to examine the postreserve trend and the initial change.

## Results

We estimate models for all reef-fish species aggregated together. There are eight commonly used gears in the data set, though not all of these catchability effects can be estimated because not all gears are deployed in every zone in every month. Certain policies target particular species or groups of species and therefore particular fishermen, but some of the reef-fish fishery participants are licensed to fish for all reef fish. Our empirical setting hence has overlapping policy layers. For instance, coincident with forming the two marine reserves in June 2000, the Gulf of Mexico Fisheries Management Council (GMFMC) increased the size limit for gag and black grouper (*Mycteroperca bonaci*) to 61 cm. Our fishery-wide postreserve shift and trend account for this effect, but it is important to note that we cannot distinguish it

for causal inference from other coincident fishery-wide changes. Also, the GMFMC instituted a 1-month commercial grouper fishery closure from 15 February to 15 March each year beginning in 2001. We include a grouper closure dummy in the regressions to account for this effect, as well as month before and month after dummies to account for the possibility that the onset of the seasonal closure stimulates effort immediately before and after its occurrence.

Because red snapper is perhaps the most valuable reef-fish species in Gulf of Mexico and not all reef-fish vessels have equal access to snapper, we controlled for the effects of red snapper regulations on targeting strategies. We include a dummy variable for class 1 snapper vessels, which are allowed a trip limit of 909 kg. We also include a dummy to indicate when the snapper season is open and an interactive term for class 1 with snapper openings. Historical data for 1993–2000 snapper openings are from Waters (2001), and more recent data are provided directly by James R. Waters (National Marine Fisheries Service, 101 Pivers Island, Beaufort, NC 28516, USA).

We include two additional covariates to account for variable aspects of fishing intensity that are not controlled for by fixed effects or by gear-specific effort: (i) number of crew and (ii) number of vessels fishing in the zone. We would expect an increase in the number of crew to increase catch conditional on other factors. We would expect the number of vessels to be negative if there are crowding effects and positive if it captures seasonal effects, i.e., vessels tracking seasonal changes in abundance across zones. In this sense, number of vessels may proxy for temporal resolution of spawning aggregations that is finer than monthly dummies and month–zone interactions.

We report two sets of results (Table 1). The first is the full model including parameters that are not statistically significant. The full model for reef fish has 345 covariates. These are in addition to the 3173 vessel-specific fixed effects that are removed from the estimation through demeaning. The second set uses backward selection to eliminate statistically insignificant results based on  $F$  tests. After backwards selection, 290 covariates remain. Because there are so many regressors, our tables only report summaries of the gear–month, gear–zone, and zone–month dummies. Full results are available from the authors.

There are 162 634 actual trips contained in the logbook data. Because our time index is daily, we allocate multiday trips evenly across the days over which the trips span. Effort and catch for multizone trips are allocated evenly across zones. These factors inflate the number of observations to 404 564 trip days. All of the gear-specific  $\alpha$ s are statistically significant. Most importantly, using the backward selection model, all eight coefficients are significantly less than 1 ( $t$  statistics, 404, 425, 355, 193, 65, 40, 50, and 28 with corresponding  $p$  values all  $< 0.0001$ ). This means that any effort reduction resulting from a policy change would increase CPUE even in the absence of a causal effect on fishable biomass. These results provide strong evidence that drawing causal inferences about the effectiveness of marine reserves for fisheries management from fishery-dependent CPUE data is flawed.

The zone-specific effects are important (Table 1). Zone-specific constants are jointly significant ( $F_{[12,404\ 564]} = 13.90$ ,

$p$  value  $< 0.001$ ), and zone-specific trends are jointly significant ( $F_{[13,404\ 564]} = 120.69$ ,  $p$  value  $< 0.001$ ). Trends in individual zones varied, with some exhibiting positive trends and others, negative trends. In zones 6 and 8, where the reserves are located, trends are negative. Clearly, results omitting these trends would understate the benefits of reserves (or overstate the costs).

Zone–gear and zone–month interactions are also jointly significant, suggesting the need for multivariate regression models that control for interactions of harvest technology with spatial and temporal factors. Simple before–after comparisons are unlikely to provide unbiased estimates of reserve treatment effects because excluded covariates are unlikely to be orthogonal to the included ones. In simple before–after comparisons, there is no way to control for fishing effort and oceanographic conditions outside of the reserve.

The snapper variables are all positive and significant. When the snapper season is open, conditional catch of reef fish overall increases. This is not surprising for two reasons. First, for boats without a snapper permit, there is potentially less fleet competition for the other reef species. Second, for boats with a snapper permit, there is one less constraint. Catch would thus increase as long as substitute target species are not easier to catch. The positive effect of the class 1 dummy simply suggests that, on average, these vessels catch more than the rest of the fleet. Positive interaction terms mean that this margin over the rest of the fleet is larger during snapper season.

To quantify announcement effects and treatment effects, it is necessary to evaluate the estimates in Table 1 at the relevant time indices. We use the parameters from the backwards-selection model. The instantaneous net announcement effect is zero, whereas the trend is positive. After 1 year, the net effect is a 19% increase in conditional catch. Because the reserve was not yet established, the upward trend cannot be attributed to the reserve itself. Whether or not fishing vessels anticipated the formation of reserves by adjusting their behavior is difficult to answer because it ultimately requires information about perceptions that is not in the logbook data. It may be possible to explore this issue further in a survey. Biologically, the upward trend may be attributable to the arrival of a strong recruitment class that temporally coincides with the reserve announcement and spatially coincides with the reserve zones. It is impossible to test this hypothesis with fishery-dependent data.

Quantifying the net treatment effect requires evaluating the treatment estimates at particular time indices but also requires a temporal baseline. We chose the last day of the announcement period as the relevant baseline. Thus, the instantaneous effect (in logs) of treatment is  $(\phi_1 + \phi_2 T) - (\mu_1 + \mu_2(T - 1))$ . The resulting instantaneous effect is a 4% decrease in conditional catch. This seems plausible given that the two marine reserves, while closing off only a small percentage of the fishing zones, represent prime habitat for certain reef species.

By the end of October 2004, the net treatment effect was –14% in reef-fish catch, declining over the time since the reserve went into effect. The reason for the negative trend inferred by reserves is not immediately intuitive. If we attach a purely biological interpretation, then it appears that reserves

**Table 1.** Fixed effects regressions.

	Full model				Backwards-selection model			
	Parameter estimate	Standard error	<i>t</i> statistic	<i>p</i> value	Parameter estimate	Standard error	<i>F</i> statistic	<i>p</i> value
Announcement shift	-1.12965	0.22047	-5.12	<0.0001	-1.12393	0.22043	26.00	<0.0001
Announcement trend	0.00047693	0.00008665	5.50	<0.0001	0.0004754	0.00008662	30.12	<0.0001
Reserve instantaneous effect	0.33688	0.04008	8.41	<0.0001	0.337	0.02559	173.36	<0.0001
Reserve trend effect	-0.00007532	0.00001273	-5.92	<0.0001	-0.0000743	0.00000832	79.78	<0.0001
Postreserve common shift	0.00262	0.03146	0.08	0.9336	Not significant			
Postreserve common trend	0.00001973	0.00000985	2.00	0.0452	0.00002037	0.00000272	56.20	<0.0001
Month before starting in 2001	0.02399	0.0094	2.55	0.0107	0.02366	0.00936	6.38	0.0115
Grouper closure period	0.08118	0.01267	6.41	<0.0001	0.08103	0.01261	41.29	<0.0001
Month after starting in 2001	-0.01588	0.00921	-1.72	0.0848	-0.01556	0.00917	2.88	0.0897
Snapper open period	0.03516	0.00414	8.50	<0.0001	0.03515	0.00411	73.23	<0.0001
Class 1	0.26757	0.01442	18.56	<0.0001	0.26745	0.0144	344.91	<0.0001
Class 1 × snapper open	0.10089	0.01866	5.41	<0.0001	0.10215	0.01861	30.13	<0.0001
Number of vessels	0.00928	0.00024398	38.02	<0.0001	0.00929	0.00024282	1463.31	<0.0001
Crew number	0.02493	0.00166	15.06	<0.0001	0.02497	0.00165	227.67	<0.0001
Alpha_gear1	0.02556	0.00242	10.57	<0.0001	0.0255	0.00241	111.49	<0.0001
Alpha_gear2	0.15814	0.00205	77.05	<0.0001	0.15779	0.00198	6361.85	<0.0001
Alpha_gear3	0.34608	0.00186	186.30	<0.0001	0.34625	0.00184	35316.10	<0.0001
Alpha_gear4	0.33602	0.00357	94.13	<0.0001	0.33656	0.00344	9555.31	<0.0001
Alpha_gear5	0.57417	0.00705	81.43	<0.0001	0.577	0.00652	7826.06	<0.0001
Alpha_gear6	0.14012	0.02163	6.48	<0.0001	0.13797	0.02145	41.39	<0.0001
Alpha_gear7	0.53519	0.01003	53.36	<0.0001	0.5316	0.00932	3256.07	<0.0001
Alpha_gear8	0.25759	0.03837	6.71	<0.0001	0.24783	0.02684	85.29	<0.0001
Area_constant_1	-0.07279	0.23525	-0.31	0.757	Not significant			
Area_constant_2	1.1115	0.22255	4.99	<0.0001	0.88867	0.03388	688.05	<0.0001
Area_constant_3	0.70739	0.26163	2.70	0.0069	0.6592	0.03595	336.23	<0.0001
Area_constant_4	2.50875	0.32923	7.62	<0.0001	2.28874	0.05695	1615.36	<0.0001
Area_constant_5	1.28692	0.20171	6.38	<0.0001	1.30026	0.11626	125.09	<0.0001
Area_constant_6	1.02204	0.16757	6.10	<0.0001	1.03754	0.03312	981.38	<0.0001
Area_constant_7	1.52062	0.23538	6.46	<0.0001	1.50972	0.15981	89.25	<0.0001
Area_constant_8	0.07875	0.55888	0.14	0.8879	0.11346	0.0398	8.13	0.0044
Area_constant_9	0.28447	0.20833	1.37	0.1721	0.26185	0.04365	35.99	<0.0001
Area_constant_10	0.34602	0.18158	1.91	0.0567	0.4835	0.02403	404.75	<0.0001
Area_constant_11	0.83074	0.15393	5.40	<0.0001	0.81904	0.04402	346.16	<0.0001
Area_constant_12	-1.40726	0.36933	-3.81	0.0001	-0.69245	0.0719	92.75	<0.0001
Area_constant_13	Restricted for identification				Restricted for identification			
Area_trend_1	0.00001031	0.00000672	1.53	0.1249	0.00001104	0.0000064	2.97	0.0846
Area_trend_2	-0.00001846	0.00000631	-2.92	0.0034	-0.00001876	0.0000062	9.14	0.0025
Area_trend_3	0.0000953	0.00000766	12.44	<0.0001	0.00009541	0.00000759	157.90	<0.0001
Area_trend_4	0.00003339	0.00000651	5.13	<0.0001	0.00003348	0.00000646	26.89	<0.0001
Area_trend_5	0.00000305	0.00000525	0.58	0.5612	Not significant			
Area_trend_6	-0.00003304	0.00000484	-6.83	<0.0001	-0.00003435	0.0000043	63.87	<0.0001
Area_trend_7	-0.000086	0.00000543	-15.84	<0.0001	-0.00008725	0.00000501	303.25	<0.0001
Area_trend_8	-0.00008624	0.000007	-12.31	<0.0001	-0.00008736	0.00000665	172.38	<0.0001
Area_trend_9	-0.00018827	0.00000778	-24.21	<0.0001	-0.00018964	0.0000075	638.48	<0.0001
Area_trend_10	-0.0000506	0.00000696	-7.27	<0.0001	-0.00005037	0.00000691	53.14	<0.0001
Area_trend_11	-0.00002695	0.00000896	-3.01	0.0026	-0.00002679	0.00000892	9.03	0.0027
Area_trend_12	0.00018906	0.00001651	11.45	<0.0001	0.00019157	0.00001544	153.93	<0.0001
Area_trend_13	0.00012788	0.00000775	16.49	<0.0001	0.00012771	0.0000071	323.39	<0.0001
Monthly_dummy_1	-0.37276	0.21012	-1.77	0.076	-0.4119	0.03014	186.75	<0.0001
Monthly_dummy_2	0.0829	0.18867	0.44	0.6604	Not significant			
Monthly_dummy_3	-0.05645	7688.33495	0.00	1	-0.05931	0.02978	3.97	0.0464
Monthly_dummy_4	0.0431	4555.99963	0.00	1	0.05155	0.01205	18.31	<0.0001
Monthly_dummy_5	-0.81154	0.22869	-3.55	0.0004	-0.37297	0.0581	41.21	<0.0001
Monthly_dummy_6	0.18991	0.18757	1.01	0.3113	0.12925	0.03263	15.69	<0.0001
Monthly_dummy_7	0.47826	0.19575	2.44	0.0146	0.39394	0.0223	311.95	<0.0001



**Table 1** (concluded).

	Full model				Backwards-selection model			
	Parameter estimate	Standard error	<i>t</i> statistic	<i>p</i> value	Parameter estimate	Standard error	<i>F</i> statistic	<i>p</i> value
Monthly_dummy_8	0.33685	0.1928	1.75	0.0806	0.29855	0.03032	96.96	<0.0001
Monthly_dummy_9	0.11447	0.20371	0.56	0.5742	0.07087	0.03229	4.82	0.0282
Monthly_dummy_10	-0.06814	0.20734	-0.33	0.7424	-0.16204	0.04051	16.00	<0.0001
Monthly_dummy_11	-0.20152	0.19539	-1.03	0.3024	-0.16491	0.02421	46.40	<0.0001
Monthly_dummy_12	Restricted for identification				Restricted for identification			
79 gear-month interactions	Jointly significant $F_{[79,404\ 564]} = 44.28, p \text{ value} < 0.0001$				61 of 79 significant at 10% level or better			
75 gear-zone interactions	Jointly Significant $F_{[75,404\ 564]} = 119.23, p \text{ value} < 0.0001$				63 of 75 significant at 10% level or better			
132 month-zone interactions	Jointly significant $F_{[132,404\ 564]} = 19.56, p \text{ value} < 0.0001$				112 of 132 significant at 10% level or better			
Observations	404 564				404 564			

**Note:** The dependent variable is the natural logarithm of individual daily vessel catch. Trend variables are scaled based on the number of days from the beginning of the sample period.

cause declines in fish biomass available for capture, all else being equal (something we strongly doubt could be true). The most parsimonious explanation, however, is that the time frame of the study was sufficient to reveal a gradual dampening effect of lost fishing area but not long enough to reveal catch benefits. As a result, the net treatment effect is trending downward in-sample. This means that the adult spillover from the closed areas is not generating a net benefit to fisheries. This could change over time, once fish in the reserve generate offspring that enter the fishery. For instance, by the latter part of 2004, the cutoff time for this study, gag offspring generated during the first year of reserve protection (2000) would be just reaching the minimum size limit making them available to the fishery. Thus the negative trend we see now is based on recruitment of a single year class. As new data become available, these models can be rerun to examine whether this treatment effect changes over time.

## Discussion

As we transition towards a holistic view of the marine environment and the services that it provides, it is essential to learn how spatially explicit management tools such as marine reserves function. However, before we can truly assess the performance of marine reserves, fisheries biologists, economists, and managers must be able to answer the following question: If marine reserves enhance fisheries yields, how would we know? In this paper, we shed some light on this question by illuminating the hidden assumptions in previous retrospective analyses. We show by example that these assumptions are unlikely to hold where management includes marine reserves for large-scale fisheries management. Our results suggest proceeding cautiously with interpreting retrospective analyses that do not control for curvature in fishing production technologies, coexistence of multiple policy layers, heterogeneity across vessels, space, and time, and perhaps, most importantly, selection effects of fishing vessels in treatment and reference groups.

To move towards retrospective analyses that more convincingly draw causal inference, we propose a vessel-based standard for judging the fishery effects of marine reserves. The effectiveness of a marine reserve can only be assessed relative to a valid counterfactual case that is most reasonably defined at the level of individual fishing vessels. We illustrate an empirical method for constructing that counterfactual case. Unless the same people with the same vessels are fishing at the same level of intensity before and after forming a reserve in the reference sites, the search for BACIP is futile. Instead, fisheries researchers should use fishery-independent information on BACIP to inform the design of retrospective analyses like the one that we present.

Our approach in this paper is part of an emerging literature on program evaluation of environmental policy. This literature seeks to isolate policy effects on economic metrics, as well as on physical environmental outputs using non-experimental data. For example, Gray and Shadebegian (2003) analyze the effects of environmental regulations on plant-level productivity and show that controlling for plant-level technological differences is essential to quantifying the effects of regulations. Strikingly, Greenstone (2004) analyzes US sulfur dioxide emissions, which have declined 80% since the 1970s, and finds that the Clean Air Act was not the cause of this large decline.

Existing retrospective analyses of the effects of reserves on fisheries (McClanahan and Kaunda-Arara 1995; Roberts et al. 2001; Russ et al. 2004) do not address key issues for isolating policy effects that are standard in the program evaluation literature and instead make simplifying assumptions. These analyses neither condition on individual skill and technology differences of harvesters nor directly address the counterfactual case that we outline above, raising questions of internal validity (i.e., whether the treatment effects are measured correctly). By using CPUE as an indicator of the policy impact, they also embed an implicit assumption about the fishing production function in the analysis that may or may not be valid. Roberts et al. (2001), for instance, show that 5 years after reserve placement there is a statistically significant increase in CPUE for artisanal reef fishermen in

St. Lucia. However, without a reference site, it is not clear that the increase in CPUE can be attributed to the reserve treatment (Hilborn 2002). Suppose, for example, that an environmentally induced increase in recruitment coincided with establishment of the reserve. In this hypothetical situation, inferences about the effects of marine reserves on adjacent fisheries simply confound correlation with causation. In a long-term study of the Apo Island marine reserve, Russ et al. (2004) attempt to address the need for reference sites by comparing CPUE in a reserve before and 16–19 years after establishment with CPUE in reference areas near and far from the reserve. Are there hidden assumptions in this approach? Suppose that after reserve formation, highly skilled fishermen are disproportionately attracted to the reserve boundary and fish in close proximity to the reserve. If this occurs, drawing inferences about the reserve effectiveness from CPUE changes suffers from selection bias. Measured changes in CPUE are at least partly attributable to the fleet behavior. Thus, isolating the true policy impact requires controlling for selection.

Even if there are no selection effects and the modeling assumptions are internally valid, we are faced with the question of whether these results generalize to commercial reef fisheries on the scale practiced in the United States. Homogenous gear, for example, might be a reasonable assumption for small-scale artisanal fisheries. However, large-scale commercial fisheries generally involve heterogeneity in gear choices (Eggert and Tveteras 2004), spatial decision-making (Abrahams and Healey 1990; Mistiaen and Strand 2000; Smith 2005), and captain skill (Abrahams and Healey 1990; Pascoe and Coglán 2002). Without controlling for these factors, there are also reasons to doubt the external validity of existing retrospective studies. As such, we propose a more general model that nests existing studies, allows for differences across fishing production technologies, includes a range of spatial, temporal, and individual heterogeneity, and controls for selection effects.

Though we focus on fishery impacts, our results also question whether we can measure the pure biological effects of reserves without taking a program evaluation approach. Murawski et al. (2000) take an experimental approach to examine the impact of large-scale closed areas on the scallop fishery in the Northeast Atlantic using fishery-independent data. They find that biomass increases both within and outside the closed areas, and the relative increase is higher within the reserve. Though these authors acknowledge the importance of displaced fishing effort, the relative increase in biomass within and outside a reserve is a function of fishing effort outside the reserve, and this effort implicitly is not controlled for in the comparison. We submit that statistically controlling for effort outside a reserve is essential for measuring biological effects within a reserve because assessing these effects is based on before–after comparisons between the reserve and outside areas. Program evaluation offers a promising approach to deal with these situations.

For our particular application, we do not yet have a definitive answer on whether the two reserves that we study benefit fisheries. Because of life history characteristics of groupers and other reef-fish species, we are reasonably certain that the reserves have not been in place long enough for

all of the potential benefits to materialize. We have less certainty about the spatial scope of reserve effects outside the reserve boundaries. Our assumption that reserve effects will manifest in the surrounding NMFS statistical zones posits a reasonably large spatial scope, especially for adult spillover, but it is plausible that larval export effects of reserves are not fully contained in the NMFS statistical zones where the reserves are located. It may be that some larger subset or even the entire fishery is treated if larvae from the reserve are distributed over very large areas. If this is the case, the treatment effect is in  $\phi_1$ ,  $\phi_2$ ,  $\mu_1$ , and  $\mu_2$ , and isolating it from coincident changes in the fishery is not possible. This will always pose challenges for evaluating the effectiveness of marine reserves. Fisheries scientists cannot simultaneously measure the spatial scope of reserve spillovers and evaluate the effects of reserves on catch. Either one imposes a spatial scope and tests effectiveness, or one imposes a level of effectiveness and tests for a spatial scope. We argue that fishery-independent data should be used to determine spatial scope of spillovers, and fishery-dependent data can be reserved for testing effectiveness. Ongoing tagging studies will ultimately help to evaluate our assumption about treatment zones in the future. Other applications of our technique are likely to be fruitful if reserves have been in place longer and there is more fishery-independent information on larval transport. This may be the case for large-scale reserves in the Northeast Atlantic.

Another plausible explanation of the negative trend in catch is that the stock is trending downward because of the bio-economic cycling from open access (Wilén 1976; Bjørndal and Conrad 1987), and we simply have not observed the reserve for long enough to see the upward trend. This is consistent with a concentration of fishing pressure along the edge of each reserve, a phenomenon documented for fishing vessels in the Northwest Atlantic (Holland 2004) and seen by one of us (F.C. Coleman) around the Madison–Swanson Reserve.

Our approach has many advantages over existing retrospective analyses of fishery-dependent data. We control for gear differences, effort changes, environmental trends, and selection effects that left uncontrolled could lead to spurious conclusions about the effectiveness of marine reserves. If catch is the product of stock and effort, effort near and far from the reserve is unchanged before and after the policy, there are no environmental changes from before to after reserve formation, and individual fishing vessels are homogenous, then our approach based on individual catches aggregates to overall catches. In this sense, our approach nests that of Russ et al. (2004).

Nevertheless, our approach has important limitations. Beyond the issue of spatial scope of reserve effects discussed above, we do not control for catch composition after formation of the reserve (neither do previous studies). Vessels may substitute across low- and high-valued species within the reef-fish complex. Inability to control for this possibility partly reflects our data; separating effort on a species-by-species basis is problematic in multispecies fisheries. An interesting extension would be to explore the reserve treatment effect through revenue functions rather than catch functions and explore whether the sign of the treatment effect is the

same when defined in terms of revenue. Another limitation is that our trend variables and the number of vessels fishing in the zone, which capture aggregate effort, may imperfectly control for the dynamic effects of aggregate effort on stocks. This type of criticism could apply to any program evaluation in which there is a latent state variable and nonlinear dynamics. Indeed, within economics there is a tension between program evaluation and dynamic structural models to measure policy impacts. Still, it is essential to note that all of these limitations apply to previous retrospective analyses of marine reserve fishery impacts as well.

Naturally, estimation of a structural bioeconomic model provides an alternative way to assess the performance of reserves after their formation. This approach would simultaneously estimate parameters of the biological system, behavioral parameters of the harvest sector, and the direct policy effects. In the absence of a structural model, the combination of biological and economic dynamics unfolding over space with latent state variables makes it difficult to interpret the results of before–after analyses, even if they incorporate all of the controls that we suggest. In other words, we cannot disentangle the structural determinants of the treatment effect. Empirical bioeconomic models that are dynamic and spatially explicit in theory can overcome these challenges. In practice, they present enormous estimation challenges. These models are currently still in their infancy and will require considerable intellectual attention to be broadly useful in evaluating fisheries policy in the future.

## Acknowledgements

This research was supported by a grant from the Saltonstall–Kennedy Program, NOAA No. NAO3NMF4270086. The authors thank John Poffenberger (NMFS) for providing reef-fish logbook data for the Gulf of Mexico and Jim Waters (NMFS) for providing historical information on snapper closures. For helpful comments, the authors thank Lori Snyder Benneer, Jim Sanchirico, Jim Wilen, two anonymous referees, and participants at the 2005 Allied Social Science Association Annual Meetings in Philadelphia, Pennsylvania.

## References

- Abrahams, M.V., and Healey, M.C. 1990. Variation in the competitive abilities of fishermen and its influence on the spatial distribution of the British Columbia salmon troll fleet. *Can. J. Fish. Aquat. Sci.* **47**: 1116–1121.
- Benneer, L.S., and Coglianese, C. 2005. Measuring progress: program evaluation of environmental policies. *Environment*, **47**(2): 22–39.
- Berkeley, S.A., Chapman, C., and Sogard, S.M. 2004a. Maternal age as a determinant of larval growth and survival in a marine fish, *Sebastes melanops*. *Ecology*, **85**: 1258–1264.
- Berkeley, S.A., Hixon, M.A., Larson, R.J., and Love, M.S. 2004b. Fisheries sustainability via protection of age structure and spatial distribution of fish populations. *Fisheries*, **29**: 23–32.
- Bjorndal, T., and Conrad, J.M. 1987. The dynamics of an open access fishery. *Can. J. Econ.* **20**: 74–85.
- Bohnsack, J.A. 1998. Application of marine reserves to reef fisheries management. *Aust. J. Ecol.* **23**: 298–304.
- Carr, M.H., and Reed, D.C. 1993. Conceptual issues relevant to marine reserve harvest refuges: examples from temperate reef fishes. *Can. J. Fish. Aquat. Sci.* **50**: 2019–2028.
- Coleman, F.C., Koenig, C.C., and Collins, L.A. 1996. Reproductive styles of shallow-water grouper (Pisces: Serranidae) in the eastern Gulf of Mexico and the consequences of fishing spawning aggregations. *Environ. Biol. Fishes*, **47**: 129–141.
- Coleman, F.C., Koenig, C.C., Huntsman, G.R., Musick, J.A., Eklund, A.M., McGovern, J.C., Chapman, R.W., Sedberry, G.R., and Grimes, C.B. 2000. Long-lived reef fishes: the grouper–snapper complex. *Fisheries*, **25**: 14–21.
- Crowder, L.B., Lyman, S.J., Figueria, W.F., and Priddy, J. 2000. Source–sink population dynamics and the problem of siting marine reserves. *Bull. Mar. Sci.* **66**: 799–820.
- Ecological Society of America. 1998. Ecosystem management for sustainable marine fisheries. *Ecol. Appl.* **8**: S1–S174.
- Ecological Society of America. 2003. The science of marine reserves. *Ecol. Appl.* **13**: S1–S228.
- Eggert, H., and Tveteras, R. 2004. Stochastic production and heterogeneous risk preferences: commercial fishers' gear choices. *Am. J. Agric. Econ.* **86**: 199–212.
- Gell, F., and Roberts, C.M. 2003. Benefits beyond boundaries: the fishery effects of marine reserves. *Trends Ecol. Evol.* **18**: 448–455.
- Gerber, L.R., Botsford, L.W., Hastings, A., Possingham, H.P., Gaines, S.D., Palumbi, S.R., and Andelman, S. 2003. Population models for marine reserve design: a retrospective and prospective synthesis. *Ecol. Appl.* **13**: S47–S64.
- Gray, W.B., and Shadbegian, R.J. 2003. Plant vintage, technology, and environmental innovation. *J. Environ. Econ. Manag.* **46**: 384–402.
- Greenstone, M. 2004. Did the Clean Air Act cause the remarkable decline in sulfur dioxide concentrations? *J. Environ. Econ. Manag.* **47**: 585–611.
- Heckman, J.J., and Hotz, V.J. 1989. Choosing among alternative nonexperimental methods for estimating the impact of social programs: the case of manpower training. *J. Am. Stat. Assoc.* **84**: 862–874.
- Hilborn, R. 2002. Marine reserves and fisheries management. *Science (Washington, D.C.)*, **295**: 1233–1235.
- Hilborn, R., Stokes, K., Maguire, J.-J., Smith, T., Botsford, L.W., Mangel, M., Orensanz, J., Parma, A., Rice, J., Bell, J., Cochrane, K.L., Garcia, S., Hall, S.J., Kirkwood, G.P., Sainsbury, K., Stefansson, G., and Walters, C. 2004. When can marine reserves improve fisheries management? *Ocean Coastal Manag.* **47**: 197–205.
- Holland, D.S. 2004. Spatial fishery rights and marine zoning: a discussion with reference to management of marine resources in New England. *Mar. Res. Econ.* **19**: 21–40.
- Koenig, C.C., Coleman, F.C., Grimes, C.B., Fitzhugh, G.R., Scanlon, K.M., Gledhill, C.T., and Grace, M. 2000. Protection of fish spawning habitat for the conservation of warm temperate reef fish fisheries of shelf-edge reefs of Florida. *Bull. Mar. Sci.* **66**: 593–616.
- Koops, M.A., Hutchings, J.A., and McIntyre, T.M. 2004. Testing hypotheses about fecundity, body size and maternal condition in fishes. *Fish Fish.* **5**: 120–130.
- McClanahan, T.R., and Kaunda-Arara, B. 1995. Fishery recovery in a coral-reef marine park and its effects on the adjacent fishery. *Conserv. Biol.* **10**: 1187–1199.
- McGovern, J.C., Wyanski, D.M., Pashuk, O., Manooch, C.S.I., and Sedberry, G.R. 1998. Changes in the sex ratio and size at maturity of gag, *Mycteroperca microlepis*, from the Atlantic coast of

- the southeastern United States during 1976–1995. *Fish. Bull.* **96**: 797–807.
- Mistiaen, J.A., and Strand, I.E. 2000. Supply response under uncertainty with heterogeneous risk preferences: location choice in longline fishing. *Am. J. Agric. Econ.* **82**: 1184–1190.
- Murawski, S.A., Brown, R., Lai, J.L., Rago, P.J., and Hendrickson, L. 2000. Large-scale closed areas as a fishery-management tool in temperate marine systems: the Georges Bank experience. *Bull. Mar. Sci.* **66**: 759–774.
- National Marine Fisheries Service. 1999. Ecosystem-based fishery management: a report to Congress by the Ecosystem Principles Advisory Panel, April 1999. US Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Washington, D.C.
- National Research Council. 2001. Marine Protected Areas: tools for sustaining ocean ecosystems. National Academy Press, Washington, D.C.
- Papke, L.E. 1994. Tax policy and urban development: evidence from the Indiana enterprise zone program. *J. Pub. Econ.* **54**: 37–49.
- Pascoe, S., and Coglán, L. 2002. Skipper skill and fishing efficiency. *Am. J. Agric. Econ.* **84**: 585–97.
- Polacheck, T. 1990. Year around closed areas as a management tool. *Nat. Res. Model.* **4**: 327–53.
- Roberts, C.M., and Polunin, N.V.C. 2001. Are marine reserves effective in the management of reef fisheries? *Rev. Fish Biol. Fish.* **1**:65–91.
- Roberts, C.M., Bohnsack, J.A., Gell, F., Hawkins, J.P., and Goodridge, R. 2001. Effects of marine reserves on adjacent fisheries. *Science (Washington, D.C.)*, **294**: 1920–1923.
- Russ, G.R., Alcala, A.C., Maypa, A.P., Calumpong, H.P., and White, A.T. 2004. Marine reserve benefits local fisheries. *Ecol. Appl.* **14**: 597–606.
- Sale, P.F., Cowen, R.K., Danilowicz, B.S., Jones, G.P., Kritzer, J.P., Lindeman, K.C., Planes, S., Polunin, N.V.C., Russ, G.R., Sadovy, Y.J., and Steneck, R.S. 2005. Critical science gaps impede use of no-take fishery reserves. *Trends Ecol. Evol.* **20**: 74–80.
- Sanchirico, J.N. 2004. Designing a cost-effective marine reserve network: a bioeconomic metapopulation approach. *Mar. Res. Econ.* **19**: 41–65.
- Smith, M.D. 2004. Fishing yield, curvature and spatial behavior: implications for modeling marine reserves. *Nat. Res. Model.* **17**: 273–298.
- Smith, M.D. 2005. State dependence and heterogeneity in fishing location choice. *J. Environ. Econ. Manag.* **50**(2): 319–340.
- Smith, M.D., and Wilen, J.E. 2003. Economic impacts of marine reserves: the importance of spatial behavior. *J. Environ. Econ. Manag.* **46**: 183–206.
- Waters, J.R. 2001. Quota management in the commercial red snapper fishery. *Mar. Res. Econ.* **16**: 65–78.
- Wilen, J.E. 1976. Common property resources and the dynamics of overexploitation: the case of the North Pacific fur seal. University of British Columbia Resources Paper No. 3, Vancouver, British Columbia.