State of the Las Cienegas National Conservation Area: Gila Topminnow population status and trends 1989-2005

Gita Bodner[1], Jeff Simms[2], and Dave Gori[1]

July 2007

[1]The Nature Conservancy in Arizona; [2]Bureau of Land Management, Tucson Field Office

The Nature Conservancy
Protecting nature. Preserving life.™

Executive Summary

The Las Cienegas National Conservation Area (LCNCA) supports three of the rarest ecosystems in the American Southwest: riparian, aquatic, and semidesert grassland; is home to 6 endangered species; and has 2 eligible wild and scenic river segments. In response to its exceptional biological, cultural, and scenic values, the Bureau of Land Management (BLM) has committed to implementing a science-based adaptive management program at Las Cienegas.

To this end, the LCNCA Resource Management Plan (RMP) sets out numerous clear, measurable management objectives; however, when the RMP was finalized in 2003, it was unclear whether the existing monitoring program could adequately measure progress towards objectives. BLM entered into a cooperative agreement with The Nature Conservancy in 2004 to evaluate the current condition of resources on LCNCA and to review existing monitoring protocols with respect to management objectives. Results are being summarized in a comprehensive "State of the NCA" report, with sections produced during each phase of this project. The first phase addressed adaptive management needs of the upland watershed (see reports on grassland status and trends and on the Biological Planning process, http://www.azconservation.org). This project's second phase has focused on stream systems and includes this report on the endangered Gila topminnow as well as an upcoming report on riparian and aquatic habitats. Pending funding for a third phase of the project, future chapters will address additional special status upland and riparian target species.

The Gila topminnow is a particularly important conservation target because Las Cienegas supports the best remaining wild population of this endangered fish in the US. The BLM has monitored fish and their habitats here annually since 1989. Monitoring is done to screen for disease and non-native fishes, track topminnow population trends, and help managers evaluate how their actions affect the fish. This report present results from 15 years of BLM fish data, tests the monitoring program's ability to detect important changes, and offers options for improving the program's efficiency and effectiveness.

Gila topminnow continue to be present in both upper and lower perennial reaches of Cienega Creek, and remain abundant in some sites. Remarkably, the creek and its tributaries appear to have remained free of the non-native fishes and crayfish that have devastated many other streams in the region. However, topminnow abundance in the upper reach monitoring sites has dropped to less than two percent of its 1989 levels. Lower reach sites have not suffered such extreme drops, but highly variable fish counts make it difficult to reliably determine whether topminnow populations in this lower reach have been stable, decreasing, or even increasing through time.

To give managers more reliable and timely information about topminnow population trends, we tested a suite of possible variations on the existing monitoring protocol. We looked for modifications that would enable managers to detect trends that double or halve the population over a ten-year span (the time frame specified by the topminnow Recovery Plan as well as a common length for management cycles). We preferred changes that could detect these trends in eight years in order to give managers some response time within the ten-year cycle. We also looked for changes that would preserve the ability to compare new results

with the existing long-term data set and could be implemented with the same or less effort now being invested in topminnow monitoring.

Overall, options that offer the best tradeoffs between effort and results are as follows:

- Increasing sampling stations in the lower reach from 4 to 16 will improve detection of population changes in what has become the species stronghold, with only a modest increase in effort. Locating new sampling stations randomly will enable results to be applied to the reach as a whole, while keeping some of the existing stations will provide continuity with past data sets.
- Reducing uninformative habitat measurements will enable staff to sample more efficiently.
- Refocusing upper reach monitoring on screening for exotic species and tracking topminnow distributions will reduce effort in this reach while still enabling early detection of exotics and revealing major topminnow changes.
- Shifting topminnow trend sampling to every-other-year will substantially reduce effort and will enable staff to satisfy other needs in alternate years, e.g. RMP and endangered species compliance requirements for Gila chub that are currently not being addressed.

We estimate these changes would increase the required topminnow field time from 5 days (current average) to 7 days per year. However, allowing 3 days of preparation and follow-up data analysis per field season year, *staff time commitments for topminnow monitoring per two years would decrease from 16 days (current) to 10 days.*

With these modifications, BLM's topminnow monitoring at Las Cienegas would meet the monitoring standards of the species' Recovery Plan and augment some of the other aspects of BLM's aquatic monitoring and management program that already improve upon Recovery Plan recommendations such as tracking major threats to the species (exotic species, loss of perennial water flow) and actively mitigating them. Extrapolating BLM's approach across the topminnow's range would greatly enhance chances of species' recovery and downlisting.

Causes of the topminnow declines observed in the upper reaches of Cienega Creek are unclear but may include drought and habitat changes from lack of disturbance. Ecological models and management experiments are needed to shed light on these or other potential causes and help guide appropriate management response.

Acknowledgments

We would like to thank Doug Duncan, Bob Steidl, Rob Marshall, and Karen Simms for generously sharing their insights, experience and knowledge with us throughout the project. Their continued willingness to answer our questions and share their perspectives greatly assisted us with the analyses presented in this report. Thanks also to the many volunteers who helped with fish sampling over the past 17 years. Scott Bonar, Doug Duncan, Rob Marshall, Phil Rosen, Bob Steidl, and Dave Weedman generously agreed to review the manuscript and provided many thoughtful comments; the report has greatly benefited from their time and effort. Finally, we would like to thank the BLM, National Park Service, and The Nature Conservancy for their generous and continued financial support of the project.

Table of Contents

Introduction

The ecological and administrative setting

Gila topminnow (*Poecilopsis occidentalis*, abbreviated as POOC) is a warm-water desert fish from the tropical guppy family Poeciliidae. These small (less than 45mm), short-lived fish have the potential for rapid reproduction and can reach high densities (Constanz 1973; Minckley 1973; Schoenherr 1977; Simms and Simms 1992). The species was once widespread throughout the Gila River system of Arizona, New Mexico, and Sonora (Meffe et al. 1983), and was reportedly "one of the commonest fishes in the lower Colorado River drainage basin" before 1940 (Hubbs and Miller 1941). After dramatic declines in distribution, the species was federally listed as endangered in the US part of its range in 1967 (USDI 1967). Primary ongoing threats to the species are predation by non-native fishes (especially *Gambusia* mosquitofish), habitat alteration and destruction (especially loss of permanent water), and extreme flooding events (Meffe et al. 1983; Weedman 1999). Gila topminnow at Las Cienegas represent the best remaining wild population of this species in the US (Meffe et al. 1983; Minckley 1999). This population is considered Level 1, highest priority for protection, by the draft revised Gila Topminnow Recovery Plan (Weedman 1999, hereafter referred to as the Recovery Plan), and is one of only 14 naturally occurring Level 1 populations in the country. At the time of the Recovery Plan's publication, only three of these 14 —Cienega Creek, Cottonwood Springs, and Monkey Springs—were free from non-native fishes, which prey upon and compete with the topminnow. Of these sites, Cienega Creek provides over 13 km of occupied habitat, while Monkey and Cottonwood Springs each provide less than 100 m (lengths of occupied habitat are from the 1998 Recovery Plan; these lengths may have changed somewhat since then, but the relative size of these habitats remains roughly the same). BLM staff estimate that Cienega Creek was home to more than two million topminnow when BLM acquired the property (Simms and Simms 1992).

The Bureau of Land Management (BLM) has monitored these fish and their habitat within the Las Cienegas National Conservation Area (LCNCA) since 1989, and plans to continue such monitoring indefinitely. This monitoring was begun with the intent to track population trends and to screen for the presence of disease and non-native fishes. Consultation with the U.S. Fish and Wildlife Service (USFWS) over activities in and around the creek produced a Biological Opinion that now requires such monitoring, though does not specify details of the monitoring protocols (USFWS 2002). BLM would like to satisfy this basic monitoring requirement as efficiently as possible, but also wants to obtain more detailed information about population status and trends to inform decisions about how to best manage aquatic species and habitat at this important site. The Nature Conservancy and BLM are engaged in a cost-share agreement to analyze past fish count data, review monitoring protocols, and evaluate options for improving the efficiency and effectiveness of the fish monitoring program. This report focuses on trends observed from 1989 to 2005.

One of the Recovery Plan's three criteria for topminnow downlisting or delisting is that "protocols for population, habitat and genetic monitoring are developed, funded, and started" (Weedman 1999). The Recovery Plan further specifies that "protocol[s] must be sufficient to detect changes in population size and habitat quality." This report evaluates how well the existing monitoring program satisfies this Recovery Plan criterion at Las Cienegas. Appendix C analyzes possible modifications that could improve the program's ability to detect population changes, and discusses how such modifications might be implemented without radically increasing monitoring effort.

Monitoring and Statistical Methods

Fish sampling

Fish sampling has been conducted every fall at Las Cienegas since 1989, skipping three years (1991, 2000, and 2001) and including one spring sampling event in 1990. Pools (with glides considered a type of pool in this paper) were seined at nine fixed sampling stations, four in the upper reach (headwaters to Springwater Canyon), four in the lower reach (Fresno Gap to the Narrows), and one in the Mattie Canyon tributary (see Figure 1). These sites were chosen by BLM and the Arizona Game and Fish Department (AZGF) specialists according to accessibility, appropriate microhabitat structure for the sampling methods used (e.g., marshy areas were avoided because seining tends to be ineffective in such terrain, and a sense of being typical of high-quality fish habitat in their respective reaches. These sites were set up in 1989 as part of a statewide AZGF sampling array (with some 100 stations) that was established to track changes in fish species distributions across the state, but was discontinued over 10 years ago (Young and Lopez 1995).

Because stream features may change after flood scour and deposition events, a seinable pool or glide close to the original 1989 sampling location was chosen each year for sampling. This excluded areas that were too deep to seine effectively or were obstructed by debris or vegetation. Where more than one seinable pool was available near the fixed point, the one in which visual surveillance showed the most fish activity was sampled. This was done consistently to improve fish catch and to increase the likelihood of detecting any exotic fishes that might be present. The length of the seine pass was limited to less than ten times the width of the stream, and areas where obstructions were likely to reduce effectiveness of seining were avoided. Most seine passes therefore sampled only a portion of their respective pools. With the exception of depletion sampling (see below), only one seine pass was done in each pool regardless of length to avoid double-counting individuals or allowing remaining fish to move into the area already sampled. A summary of sampling methods used by field personnel is found in Appendix A.

All fish captured during seining were identified and counted. From 1989-1995, pools were blocked and seined several times in succession (depletion sampling); later years were sampled with a single pass. To ensure comparable data between years, our analysis considers only counts from the first seine haul for each sampling date. Individuals of the larger and longer-lived species (*Gila intermedia* and *Agosia chrysogaster*) were recorded as adults or juveniles based on body length threshold values, and were measured individually in some years. Most fish under 20mm in length slip through the 3 mm (1/8”) nets used here. Because *Poeciliopsis occidentalis* mature at approximately 25mm in length and grow from newborn to adult within a few months, all individuals caught were counted in a single subadult/adult size class. Individual topminnow were also measured in a few years. Measuring fishes was discontinued because it was time consuming and BLM did not expect to use length data in their abundance analyses.

Each count is considered an index of fish abundance in the sampled pool. Because the length and width of the pool segments sampled were also measured at the time of sampling, and because only one first-pass haul was done in each pool, fish count data can be analyzed as counts per unit area or as counts per pool. Both qualify as a measure of count per unit effort (CPUE). Our analyses assume that this index has a constant linear relationship with the sampled sites' true topminnow density; Appendix C suggests a method for testing this assumption.

Analyses presented here focus on Gila topminnow. While some (mostly juvenile) Gila chub were caught in these samples, these seining methods may not capture Gila chub effectively enough to provide a consistent index of chub density. The methods used here are tailored to capturing slow-moving fish in relatively shallow waters (topminnow can also be sampled by electrofishing, but the high voltage required to stun the small fish is apt to injure or kill many larger fish, including chub; J. Simms, pers comm. 2006). Gila chub often occur in pools too deep to seine effectively. In pools with undercut banks or abundant root mats like those found in Cienega Creek, chub fairly readily escape nets such as those used here (e.g. Nelson 1993). Sampling tailored to capturing chub and similar fast-moving, secretive fishes generally employs either passive trapping, electroshocking, and/or repeated seining of stream segments that have been isolated by block nets to catch fleeing fish (e.g. Griffith and Tiersch 1989; Schultz and Bonar 2004). Longfin dace are considered intermediate between topminnow and chub in the ability of single-pass seining to provide an adequate index of population densities (Jeff Simms, pers. comm.. 2006). Data for dace densities are presented in Aquatic Habitat Status and Trends report (in prep).

In addition to annual seining of index sites, an extensive fish survey was conducted in 1990 in conjunction with an aquatic habitat inventory. A rough population estimate of two million topminnow in Cienega Creek was generated based on depletion estimates for several pool and glide macrohabitats (Simms and Simms 1992). Average numbers of fish per square meter was multiplied by the area of pool and glide habitat (topminnow were rarely found outside these macrohabitat types) for the entire creek from the Narrows to the Head waters, minus the mostly dry habitat between the upper and lower perennial reaches.

In addition, each macrohabitat in the approximately 15 km of stream was scored for apparent density of fish. Fish sightings were categorized by species and apparent abundance per 10m of habitat (topminnow and dace: abundant: > 100fish seen; moderate: 10-100 seen; scarce: 1-9 seen; none: no fish seen; chub: a: > 20seen; m: 5-20 seen; s: 1-4 seen; n: no fish seen). Neither extensive depletion estimates nor visual surveys have been repeated.

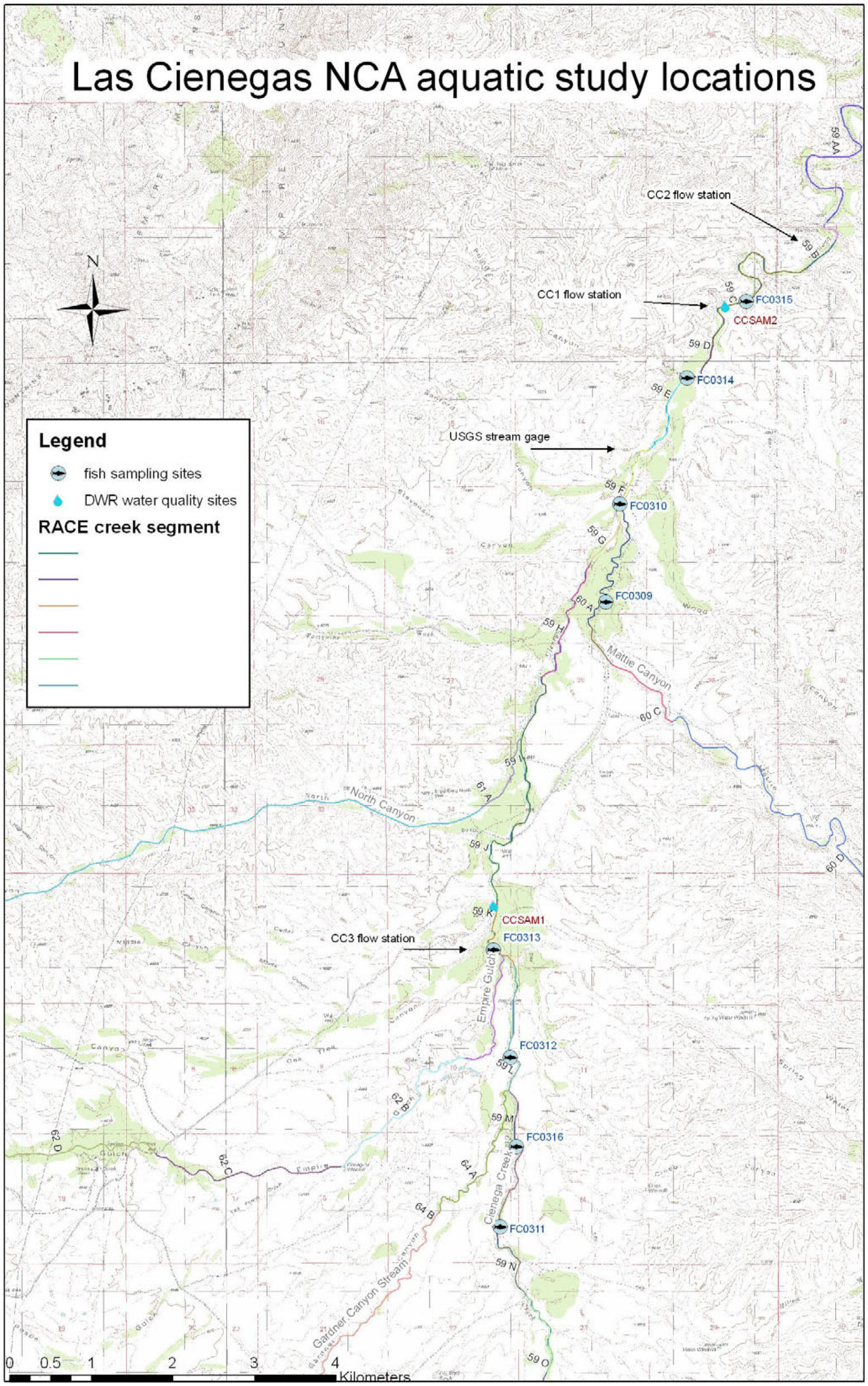


Figure 1. Cienega Creek and major tributaries within the LCNCA, with stream segment numbers and locations of fish sampling sites, Arizona Department of Water Resources water quality sampling stations, and current and past streamflow measurement points.

Population trend analysis

To determine whether fish numbers in the creek have been stable, declining, or increasing through time, analyses for population trends are often performed as linear regressions of abundance or density (or an index of these) against time. The slope of this regression line reflects the trend the population index has been experiencing during the course of monitoring, and the regression's r^2 estimates how much of the population's changes can be explained by this trend through time rather than by year-to-year fluctuations or sampling error. While trend data can be analyzed in other ways (e.g. Thomas 1996; Rasmussen et al. 2001), linear regression is one of the most commonly used, straightforward, and easily interpretable methods, and has the added benefit of well-developed methods for power analysis (Gibbs 1995; Gerrodette 1987, 1993).

In this analysis, we regressed counts from each sampling station against year; these counts were standardized by the area sampled (producing a count per unit effort) and then log-transformed (see below). A common alternative to using each count separately, as we did here, is to average values from all sites for each year and use this set of yearly averages for the regression against time. If the same number of points are sampled in every year, the estimates should be identical to those generated with separate sites. We chose to leave sites separately because not all sites were sampled in every year; keeping sites separate gives years in which more sites were sampled a stronger influence on the regression. These two approaches, however, yield different error estimates and are not always interchangeable (see Appendix C).

Statisticians recommend log-transforming survey counts for many trend analyses (Hayes and Steidl 1997; Gotelli and Ellison 2004; Zar 1999), for three main reasons: First, the slope of a regression of log-counts vs. time represents an estimate of percent population change per time period (per year, in this case). Second, many population surveys show variances that are positively correlated with abundance (e.g., the more fish you catch the more variable those catch numbers are), which can interfere with statistical analyses. Log-transforming such data helps solve this problem by equalizing variances. Third, survey data often fit a log-normal distribution. Log-transforming such data produces an approximately normal distribution. Since many statistical analyses assume normally-distributed data, this transformation can help survey data meet model assumptions and therefore make analyses more appropriate and results more reliable.

When transforming data to a log scale, zero counts have no log value and therefore appear as missing data in the log scale. Adding one to each count before log-transforming them prevents loss of this data (Sokal and Rohlf 1995; Gotelli and Ellison 2004); the original zero counts then become zeroes on the log scale as well (i.e., $\log(0+1) = 0$).

We chose to scale counts by area seined to transform raw numbers into an index of counts per unit effort. This scaling makes our results more consistent with other fish studies and reflects the fact that changing the length of seine hauls typically does change the number of fish caught (AZGFD 2004). Counts per square meter were deemed inappropriate for use with log transformation because dividing counts into such small units makes each unit's numbers fairly low; adding a small constant (usually one) to each of these before log-transforming the data changes these counts relatively more than adding the same adjustment factor to larger counts. We chose to use counts per $50m^2$ when log-transforming data because this area is close to the average size of pool samples (pools themselves were larger but only a portion of each pool was seined) and therefore log-transforming these values would have approximately the same effect as using whole seine hauls.

Confidence intervals around trend estimates

With endangered species such as Gila topminnow, it is important to evaluate the ability of a monitoring program to detect meaningful trends. A regression analysis gives an estimate of change through time, but cannot say how confident we should be in that estimate, nor whether the study is likely to have missed an important trend.

Calculating confidence intervals around trend (slope) estimates provides the following information: The width of the confidence interval reflects the monitoring program's precision. The confidence interval itself represents a bracketed estimate of the amount of change observed, which essentially says "with 90% certainty (or some other specified level), we can say the population changes have been between X percent and Y percent per year." Any confidence interval that includes a zero slope (no change) means the trend is not statistically significant at the stated alpha level (here 10%). Lastly, the estimated amount of change observed can be compared directly to the amount of change thought to be biologically meaningful (or some other pre-determined effect size). If the confidence interval includes zero but also reaches the biologically meaningful threshold, one must conclude that the study may well have missed an important trend. Similarly, if the study did find a significant trend but the confidence interval did not reach the biologically meaningful threshold, one can conclude that the amount of change observed is probably not important to the species (Hayes and Steidl 1997).

We accept that setting a "biologically meaningful" threshold for the amount of change to be detected is somewhat arbitrary at this time, but propose a working target of being able to detect at least a doubling or halving of the population within the span of this study (17 years) or within ten years (for future studies) (See Appendix C for additional justification). An overall loss of half the population translates to an annual decline of 13% for five years, 7% for ten years, and 4% for 17 years (assuming exponential change). An overall doubling translates to an annual increase of 15% for five years, 8% for ten years, and 4% for 17 years (also assuming exponential change).

Retrospective power analysis provides a slightly different approach to the same problem by estimating how likely one was to have detected a change of a particular magnitude if that change in fact occurred. However, several practitioners point out pitfalls and misuses of this approach and recommend calculating confidence intervals instead (Steidl et al 1997; Thomas 1997; Steidl and Thomas 2001; Gotelli 2004).

We used a 90% confidence interval for these calculations, which leaves a 10% probability that the true amount of change is above or below our bracketed estimates. Details of how to calculate confidence intervals around trend estimates are presented in Appendix B, using data from this study as an example.

Results

Gila topminnow continue to be present in both upper and lower perennial reaches of Cienega Creek managed by BLM, and remain abundant in some sites. The creek and its tributaries appear to have remained free of non-native fishes.

Topminnow populations have been more ephemeral in Cienega Creek's tributaries than in the main creek. Fall surveys last documented topminnow in Mattie Canyon in 1999. This tributary site was not sampled in 2000, 2001, 2004, or 2005, but surveys in 2002 and 2003 found only longfin dace at the site. In June of 2006, however, knowledgeable biologists observed abundant topminnow in the Mattie Canyon headwaters spring and in several pools downstream towards the tributary's confluence with Cienega Creek (K. Simms, Bureau of Land Management, and D. Gori, The Nature Conservancy, pers. comm. June 20, 2006). In 2001, BLM received approval from USFWS to begin moving topminnow from Cienega Creek into a segment of the Empire Gulch tributary that is separated from the main creek by approximately two and a half miles of intermittent streambed and another mile of interrupted perennial pools (USFWS 2001). Topminnows have been transplanted from Cienega Creek to upper Empire Gulch several times since, most recently in April of 2006. Persistence of these transplants has been tenuous; fish sampling in Empire Gulch in October 2005 found few remaining topminnow. Seining in October of 2006 found slightly more topminnow but still low fish densities, suggesting that the 2006 transplants have survived but have not reproduced in large numbers.

Over all sites and all years from 1989-2005, Gila topminnow catch averaged 328.5 fish per seine haul or 8.5 fish per m^2 (see Appendix D). Pools seined averaged 49.4 m^2 overall, being significantly smaller in the upper reach than lower reach (42.6 m^2 and 55.9 m^2 respectively, t-test, $t = 2.1$, two-tailed $P = 0.036$). Differences in area seined reflect, in part, the creek's diverse habitat structure, with upper reach pools often being more isolated from the surrounding aquatic habitat than pools in the continuous-flow system of the lower reach. Size of pools correlates poorly with numbers of fish caught (lower reach $r^2 = 0.038$, $P = 0.169$; upper reach $r^2 = 0.019$, $P = 0.348$).

Population trends in Gila Topminnow

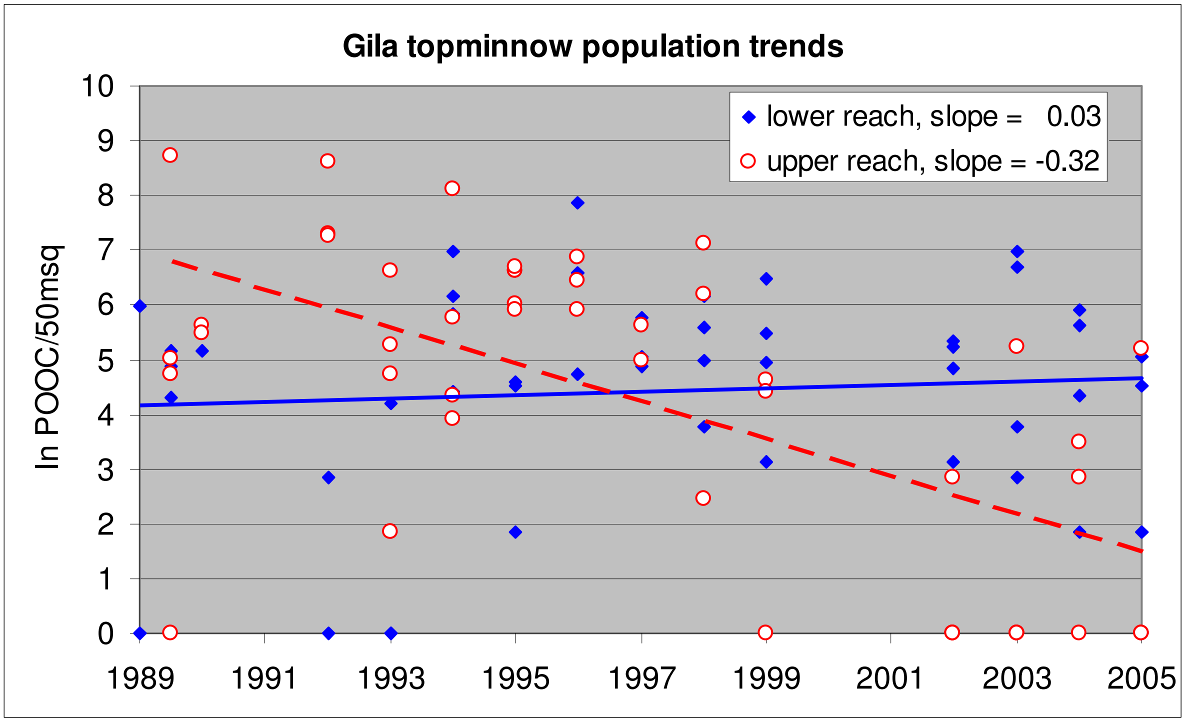
Over the 16 year monitoring period, Gila Topminnow populations in the creek as a whole (giving equal weight to all sites in all years) have declined by 15.6% per year (90% confidence interval: 8% to 23% decline).. This trend is highly significant, although it explains only about a tenth of the variation in fish abundance (Table 1).

Separating samples into reaches, however, reveals that trends differ strikingly between the upper and lower reaches (see Figure 1). Populations in the lower reach remained relatively stable over this 16-year period, with no trend in numbers over time and no samples without fish of some kind, but an increasing number of occurrences of samples without topminnow that were formerly productive. In contrast, topminnow numbers in the upper reach declined dramatically over the same period. Seven of the samples taken after 2001 had no fish of any species. The downward trend is still clear when we consider only samples with fish (hereafter referred to as occupied samples), though the estimated decline and amount of variation explained by year are both lower.

Table 1: Gila topminnow population trends between 1989 and 2005 expressed as an annual percent, as estimated by the least-squares linear regression of log(topminnow per 50m^2) by year. Standard error $\hat{\sigma}_k$ of slope estimates, coefficient of determination, r^2, and associated *P*-values are also given. Summary statistics for Gila topminnow population trends between 1989 and 2005, regression results. Probability values are the likelihood that the observed trend between fish numbers and year is caused by chance alone.

	Slope estimate parameters			ANOVA, ln(POOC/50m^2) vs. year				
	Annual change	$\hat{\sigma}_k$	r^2	Source	DF	Sum of Squares	Mean Square	Probabilities
Whole creek	**-16%**	**0.045**	**0.104**	Regression	1	64.1	64.1	F Ratio = 12.095
				Residual	104	551.2	5.3	*P* = 0.0007
Lower reach	**+3%**	**0.055**	**0.005**	Regression	1	0.8	0.8	F Ratio = 0.223
				Residual	49	183.6	3.7	*P* = 0.638
Upper Reach all samples	**-32%**	**0.060**	**0.374**	Regression	1	132.4	132.4	F Ratio = 28.056
				Residual	47	221.9	4.7	*P* = 0.0001
Upper Reach occupied samples	**-19%**	**0.067**	**0.172**	Regression	1	60.1	60.1	F Ratio = 8.313
				Residual	40	165.8	4.1	*P* = 0.0006

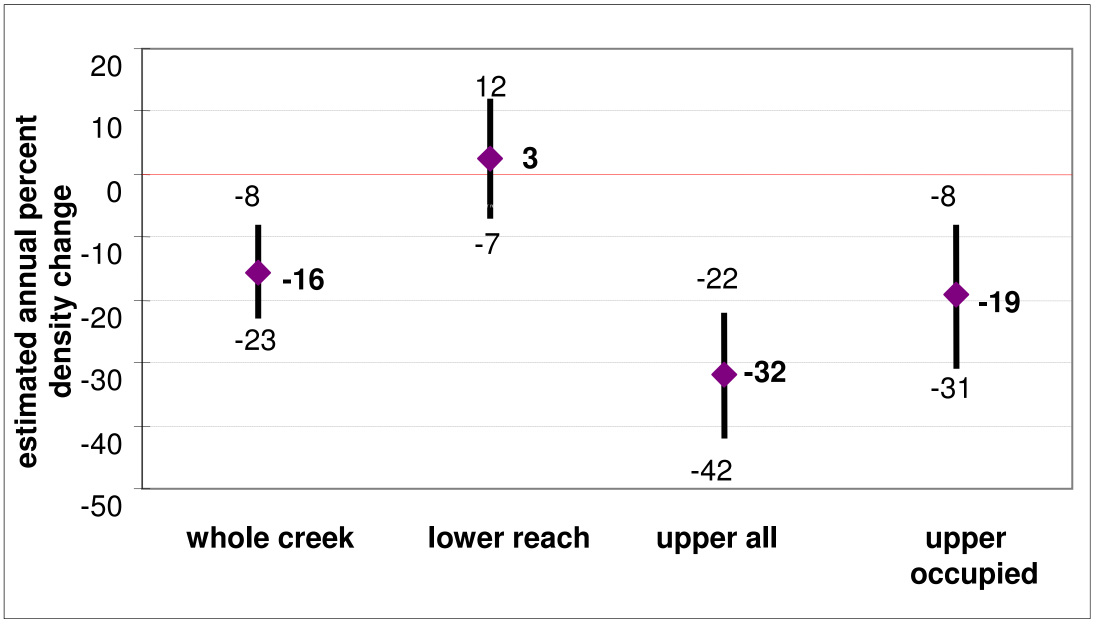
Figure 2. Gila Topminnow population trends, 1989 to 2005, with regression lines for upper (dashed line) and lower (solid line) reaches of Cienega Creek on the Las Cienegas National Conservation Area. Points represent annual counts of topminnow at individual sampling sites, scaled by area sampled and log-transformed.



Confidence intervals and ability to detect change

Calculating confidence limits for slope estimates shows the precision of our actual trend estimates (see Figure 2). As expected, using within-year averages to estimate trends produces very similar slope estimates, though confidence intervals are slightly wider (see Discussion). Confidence limits for the whole creek and for the upper reach leave little doubt that their Gila topminnow populations have dropped by more than half over the 16 years of this study, and indicate that these fish may well have dropped to less than one tenth of their former numbers. Note, however, that while the 90% confidence limits for the lower reach include 0 (trend is not statistically significant), they also include both +5% (population may have doubled) and -5% (population may have dropped by half). In other words, our data for the lower reach cannot reliably distinguish between a stable population, one that has increased dramatically, and one that has declined precipitously.

Figure 3. Regression slope estimates, in bold, with two-tailed 90% confidence intervals (□ = 0.1) for ln(POOC/50m2) vs. year.



Discussion

Trends in fish numbers

Considering all samples together, topminnow in Cienega Creek have experienced declines that are both statistically significant and biologically meaningful. Nevertheless, topminnow remain present throughout the creek and appear to have expanded back into the sampling site in Mattie Canyon in 2006. Mattie Canyon was severely impacted by excess sediment resulting from the October 2000 failure of a grade control structure located a short distance upstream. Visual observations shortly after the structure failed revealed little flowing habitat and no pool or glide habitats. Sampling had to be curtailed for several years due to poor conditions.

While topminnow numbers in the lower reach have clearly fared much better than those in the upper reach, we cannot be fully confident that they are, in fact, stable. The estimated trend is not significantly different from zero (stable), but the 90% confidence interval around this trend estimate is quite broad. This confidence interval tells us that our data is sufficient to conclude that as a best-case scenario, the population has grown by up to 12% per year (increasing the population six fold over the 16 year study span), and as a worst case scenario, this population may have been declining by up to 7% per year (losing two thirds of the population over the study span). Either of these extremes is biologically significant, and the fact that we cannot distinguish between a very large increase, a large decrease, and no change at all is problematic for managers. This broad confidence interval shows the importance of increasing the monitoring program's precision in the future (see Appendix C).The upper reach has clearly experienced declines that are both statistically significant and biologically meaningful. Using all samples for our analyses gave a worst-case scenario of a 42% decline per year and a best-case scenario of a 22% decline per year. These declines would have been considered biologically meaningful if they had been observed over just a 2-3 year period, but this analysis suggests they have continued for much longer. Excluding samples with no fish from the trend analyses showed somewhat more gradual declines of 8% to 31% per year over 16 years. Even the best-case scenario still amounts to a biologically meaningful change in just ten years.

Three upper reach sampling sites have gone anoxic or dried out during one or more years. We analyzed upper reach trends both with and without these no-fish samples to test the hypothesis that overall fish numbers in this reach were stable but that extirpation of fish at some sampling sites created the *appearance* of a decline overall. Population declines were evident in both analyses. We can now conclude that fish numbers clearly have declined throughout the upper reach. Nevertheless, there is still some uncertainly as to how well these sampling sites represent conditions overall. Given this fish's ability spread out from refuge populations and recolonize habitats where it has been temporarily extirpated, adding some consistent presence-absence sampling of a larger number of sites (as has been done informally) might contribute important information about population health, and would at least help gauge how well the existing sampling sites do, in fact, reflect dynamics in the reach as a whole.

Our analyses use density as an index of population size. Changes in area of available habitat could de-couple these two measures. An increase in available habitat, for example, can enable existing populations to spread out and thereby reduce the density at any given sampling site; however, this effect is likely to be short lived (a few months) as the reproductive rate of the species is extremely high (Minckley 1973). BLM has recently begun mapping the extent of perennial surface water each year, showing changes in amount of available habitat (see Threats Monitoring below). In the case presented here, decreases in fish density coincided with a loss of perennial stream area, which suggests that density declines do reflect overall population declines. Interpreting future density changes should be coupled with examining amounts and patterns of change in stream area. If, for example, loss of stream area coincides with increases in density, one will have to consider the possibility that existing fish are simply being concentrated in shrinking pools. There are limits, however, to how far these weak-swimming fish can move as pools dry up; many will simply be stranded and die. Evidence of recent drying can also provide an explanation for a reduction in stream length occupied and, therefore, overall population size.

To convert density measures into an estimate (rather than an index) of population size for the whole creek, density measurements would have to be collected from a random array of sites, and would have to be combined with a measure of available habitat area. Neither of these conditions are met for existing data, but could be met with modified monitoring protocols described below. Some depletion sampling might also be needed to translate first-pass measures of relative density into absolute population estimates.

Methods of analysis

Exploration of analytical alternatives convinced us that two of our analytical decisions—using fish counts per unit area rather than counts per seine haul, and using data from each sampling site separately rather than reducing these into yearly averages of all sites—had little effect on results. Preliminary tests using counts per seine haul produced trend estimates that were very similar to the estimates presented above that used the area-scaled measure, with broadly overlapping confidence intervals.

As we expected, slope estimates and confidence limits around these estimates differ only slightly between regressions that use each sampling event separately (presented above) vs. analyses that reduce multiple sampling sites into yearly averages (compare Figure 2 with the following 90% confidence intervals around annual change estimates: Whole creek: -6% to -25%; Lower reach: +16 to -9%; Upper reach all samples: -19% to -43%; Upper reach occupied samples: -8% to -31%). This yearly average approach reduces sampling variation but also reduces the degrees of freedom. Unless one implements complex weighting schemes during the regression, it also weights each year equally. During this study, there were several years in which some of the sites were not sampled. By using each site's yearly values separately rather than averaging sites within years, we allowed the regression to be weighted more heavily by years in which more sampling was done. The similarity between these results and those presented earlier suggests that having skipped some sites in some years had little effect on trend estimates.

The subjective selection of fixed sampling sites and of pools within sites each presents some difficulties for interpreting results, though they may well have reduced sampling variability and have clearly increased the likelihood of detecting exotics. Subjective selection reduces the scope of inference of a study to the sampled sites themselves (Gotelli and Ellison 2004). With a site of this size, the object of monitoring should be to know what is happening in the creek as a whole, or in its main reaches, not just in a handful of sampling spots. Counts from such selective sites may act as consistent indices of these larger populations, or they may not. This relationship has not yet been tested here. However, field observations by fisheries professionals at Las Cienegas do suggest that patterns documented at these sites appear to be consistent with changes observed in the rest of the creek (Jeff Simms, pers comm.. 2005; Phil Rosen, pers. comm.. 2007). Managers may be content to assume this untested relationship and continue to use these sites as uncalibrated indices of fish populations throughout the creek. However, to truly extrapolate trends to the creek as a whole, one would need to document the numerical relationship between these index sites and a set of randomly located sampling sites (e.g. Larson et al. 2001).

Sampling an array of randomly-located sites in multiple habitats would also enable a statistically valid estimation of the topminnow population in the creek as a whole, with less effort than was required for the 1992 population estimate. However, this study and many subsequent observations show that the vast majority of topminnow are found in pools (with glides as a subset of pools). Pairing an estimate of the percent of the creek composed of pools with density measurements of topminnows in a random set of pools would approximate an estimate of topminnow populations throughout the creek.

The assumption that fish numbers from the index sites are representative of numbers in the rest of the stream reach is more tenuous in the upper reach where several pools have gone dry or otherwise lost fish entirely. Because BLM has only four sites in this reach, losing fish in one or two of these dramatically changes results. Casual observations of pools near the index sites show that some have also lost fish. It is unclear, however, what proportion of habitats in the reach have suffered such losses, and whether this is comparable to the proportion of index sites that have lost fish.

The broad confidence intervals around trend estimates demonstrate that this monitoring program would greatly benefit from improved precision. Such increased precision can be achieved by reducing sampling variability, sampling sites more often, lengthening the study span, and/or adding more sampling sites. These alternatives are explored in detail in Appendix C. Adding new sites into an existing array of subjectively selected sites may also provide an opportunity to test these assumed relationships between index sites and the rest of the creek.

Monitoring design

This monitoring design clearly satisfies some key elements of the Recovery Plan's delisting criterion for monitoring. Most importantly, protocols for population and habitat monitoring have been successfully implemented here over 17 years. Nevertheless, data from the lower reach cannot be used to distinguish between population increase, stability, or decrease since 1989 shows that it does not entirely meet the criterion of being "sufficient to detect changes in population size and habitat quality." Appendix C explores ways to improve power to detect population changes. Sufficiency of habitat monitoring to detect change is discussed in the Aquatic Habitat report (Bodner et al. in prep). Genetics of this population have been sampled but not monitored here. The genetic monitoring called for in the Recovery Plan is meant to screen for loss of heterozygosity due to small founder populations, which so far has not been a concern in Cienega Creek's extensive natural population. Across the species' range, genetic monitoring can also reveal changes in distribution of distinct genetic lineages.

At Las Cienegas, this will be of little concern unless managers have reason to suspect that the site's naturally-occurring fish may have been "contaminated" by fish from genetically-distinct populations.

The Recovery Plan considers conservation of the Gila topminnow across the US portion of its range, not just at Las Cienegas. To this end, the Recovery Plan emphasizes that since "several natural resource agencies are involved in Gila topminnow monitoring, a standardized monitoring protocol must be developed and implemented by the agencies." The AZ Game and Fish Department is the other main entity monitoring topminnow populations. AZGFD attempts to visit all occupied topminnow habitats each year and uses a variety of capture methods to determine what species are present (e.g. Weedman et al. 1997a, b; Voeltz and Bettaso 2003), amounting to a network of "spot-check" surveys (AZGFD 2004). This widespread monitoring may enable managers to detect changes in numbers or persistence times of extant populations, but is not well suited to detecting abundance changes within a population. Las Cienega's much more intensive monitoring is only marginally successful at detecting topminnow population trends, which suggests that spot-check monitoring is unlikely to detect even very large changes in abundance or density at other sites. If the goal of standardizing methods is to meet the Recovery Plan monitoring criterion across the topminnow's range, any attempts at such cross-agency standardization should view Las Cienegas trend monitoring as a minimum model, and should conduct power analyses similar to those presented in Appendix C but targeting changes across multiple sites (e.g. Urquhart et al. 1998; Allison 2000; Larsen et al. 2001) to ensure that overall monitoring effort is sufficient to detect important changes. Detecting trends across multiple sites will generally require fewer samples per site than are needed to detect trends at a single site. If the goal of standardizing monitoring is to help make recovery real across the species' range, such attempts should consider adding some threats monitoring as the LCNCA has.

Threats monitoring, possible causes of trends at Las Cienegas, and follow-up research questions

Monitoring population numbers and habitat parameters are not enough to ensure survival of the species. In reviewing monitoring for 181 endangered and threatened species that have active recovery plans, Campbell et al. (2002) found that most recovery plans and their associated monitoring programs "do not adequately address the specific threats affecting species." The overall monitoring and management program at Las Cienegas improves upon mandates of the Recovery Plan in this regard. LCNCA monitoring already tracks some of these threats, management is actively mitigating some threats, and plans are in place to improve our ability to track, respond to, and reduce threats.

The two greatest known threats to topminnow populations across the species' range are contamination by invasive fish and drying up of aquatic habitats, with extreme flooding as an additional factor that has caused extirpation in many sites where introduced populations of topminnow were established (Meffe et al. 1983; Brooks 1985; Weedman 1999; Minckley 1999). Existing LCNCA fish monitoring tracks invasive fish presence. BLM removal of stream impoundments and seasonal drying of stock ponds has already reduced risks of contamination. Implementation of a watershed re-inventory of current and potential sources for invasive species (currently being developed), and mitigation of threats from these sources (tasks in the LCNCA RMP), will further reduce risks of contamination. The newly formalized summer wet-dry monitoring tracks changes in availability of surface water, as does the USGS streamgage. A more detailed plan to track groundwater levels will help address the groundwater component that generates changes in baseflows. Work with the surrounding communities to limit impacts of groundwater use will be key to reducing the threat of losing surface flows. The USGS stream gage tracks changes in extreme flooding. Monitoring upland shrub, grass, and bare ground cover on the NCA helps track changes in some of the components that contribute to flooding risk (Gori and Schussman 2005; Simms et al. 2006), as does monitoring of riparian vegetation and bank stability. If successful, actions to reduce shrub cover (in favor of grasses) and to otherwise reduce bare ground will reduce the risks of extreme flooding. Ongoing BLM actions to stabilize headcuts and prevent stream capture by an eroded tributary also address this threat.

In the absence of exotic fishes, the specific causes of these topminnow declines at Las Cienegas are not clear, but several main factors—and interactions among these factors—have been proposed:

- Flooding;
- Extreme winter cold temperatures;
- Reduced streamflow;
- Decline in water quality at some sites, particularly low dissolved oxygen;
- Physical habitat structure changes, with less shallow back-water habitat available; and
- Changes in primary productivity and foodweb dynamics with increased tree cover.

Data on several of these factors are presented in the LCNCA Aquatic Habitat Status and Trends report (Bodner and Gori, in prep.). Here we report conclusions from habitat analyses and speculate on how these factors may influence topminnow density trends.

Flooding and winter cold temperatures are periodic disturbances that are presumed to have happened throughout the topminnow's history; we have no evidence that the frequency or intensity of these disturbances has changed in recent years.

The region has been in a pronounced drought for much of the last ten years, which has impacted and even extirpated many Gila topminnow populations throughout the region (Duncan 2006). Early LCNCA streamflow data are too spotty to enable direct correlations between fish numbers and local streamflow parameters, but a variety of data suggests that Cienega Creek flows have declined. Some formerly perennial segments of Cienega Creek have dried up since 1989, affecting up to 40% of the stream's wet length. Within the NCA, the upper reach consistently has lower flows than the lower reach, and runs at levels that may put it near a lower threshold for habitability by fish. The fact that declines have been sharper in the upper reach is consistent with this drought threshold hypothesis. Apart from areas of the creek that have dried completely, drought may drive topminnow numbers down via several mechanisms.

Reduced streamflow can cause or compound water quality problems by limiting a stream's ability to buffer, dilute or flush out chemical changes. We suspect water quality changes—particularly low dissolved oxygen (DO)—have caused some of the fish declines in the upper reach. Recent years' readings have shown DO concentrations of less than 2 parts per million (ppm) in several upper reach pools (compared with 6 ppm or more in lower reach sites). Low DO here is presumably generated by natural causes including proliferation and death of diatoms that leave an oily film on the water surface and deposition of leaf litter that consumes oxygen as it decays. Such low concentrations are lethal to some fish and cause sub-lethal damage to many others (Lowe et. al. 1967). When topminnow are exposed to low levels of oxygen, they congregate near the surface so that they can use the well oxygenated microhabitat in the top few centimeters at the water's surface where gas exchange is occurring. This behavior was demonstrated in the laboratory for other desert fishes (Lowe et al. 1967). A cap of surface scum will interfere with gas exchange, leaving the topminnow exposed low oxygen conditions throughout the water column precluding survival.

Gila topminnow, however, are known for their tolerance of high temperatures and low oxygen (Meffe et al. 1983). An October 2003 survey of 15 upper reach pools found DO ranging from 5.2 to 0.9 ppm. Visual topminnow surveys found no apparent relationship between DO readings and fish presence at these sites; in fact, the highest density of topminnow were observed at the site with lowest DO. Fish appeared healthy and active under these conditions. However, instantaneous DO readings do not necessarily reflect the oxygen stresses fish have been exposed to over time; even pools that report "normal" DO levels at the time of fall fish sampling may have experienced critically low levels during previous summer months. High levels of dissolved carbon dioxide (CO_2) have also been recorded in some of the stagnant pools with no fish, up to 29 mg/l (site FC0311, 2005). High CO_2 concentrations can be toxic to fish, especially in combination with low DO (Everhart et al. 1981).

Streamflow declines have subjected index pools to varying degrees of isolation during this study. Lower reach sites were all connected by flowing water during the June 2006 flow survey and presumably have remained so throughout the study. In the upper reach, the uppermost site FC0311 has become isolated except during flood events and has had no fish since 2002. The lowermost site FC0313 has been dry in the summer and fall since 2004. The two middle sites have remained wet and are presumed to have retained at least some of their connectivity with the rest of the stream. Isolation would prevent topminnow from leaving declining habitats and thereby increase mortality, and would keep fish from recolonizing a pool from adjacent areas when conditions improved.

Cienega Creek has undergone dramatic changes in riparian and aquatic habitat since 1989. Repeat photographs show changes from a broad sunny waterway with banks bare or sparsely covered with herbaceous growth in 1990, to a narrow channel with densely vegetated banks and a full riparian tree overstory by 2006. BLM aquatic macrohabitat surveys in 1990 and 2000 show a narrowing channel (except marsh habitats), with pools deepening throughout the creek. Total pool area in the lower reach quadrupled, replacing large amounts of riffle and marsh. Average overstory canopy cover increased in every habitat category, from 8% to 42% throughout the creek.

Such changes—narrower channel, deeper pools, increased canopy cover, more undercut banks, more woody cover, etc—are generally considered desirable changes for fisheries (Everhart et al. 1981). They may not, however, all be good for topminnow. Without an independent understanding of topminnow habitat needs, it is difficult to relate any of the observed habitat changes with changes in fish populations. A 1990 study found that topminnow used pool, glide, and backwater habitats disproportionally to their availability and preferred sandy substrates (Simms and Simms 1992). The study's authors comment that these associations represent a preliminary step towards understanding topminnow habitat preferences, and note that "a knowledge of basic habitat factors which are essential to maintenance of the topminnow population at Cienega Creek is necessary to guide future management actions" (Simms and Simms, 1992). This statement still holds true.

The increase in tree canopy may exacerbate drought effects by transpiring larger amounts of water. However, recent studies in other systems have questioned the common conception that proliferation of riparian trees increases water loss from the system by demonstrating that increased canopy cover can lower surface temperatures and reduce evaporation rates enough to offset any increases in transpiration (Leenhoutts et al. 2006). The net impact of these two effects is unknown at Cienega Creek.

Fish experts speculate that growth in the canopy cover of riparian trees may have triggered cascading changes in foodweb dynamics (Jeff Simms, pers. comm. 2005; Phil Rosen, pers. comm.. 2006). They note dense concentrations of aquatic invertebrates in the floating mats of macrophyte algae that thrived in sunny parts of the stream during the 1980's and early 1990's. In addition to causing oxygen depletion from decaying leaves, they argue that the shading of the stream may have caused a shift from a highly productive primary production foodweb to a less productive foodweb based largely on detritus. Even where oxygen levels are adequate for topminnow, they argue, such a trophic shift would reduce food availability for topminnow and thereby limit the fish densities that the stream could support.

Teasing apart complex chains of causality is a job for targeted research rather than general monitoring. Cienega Creek is an ideal place to research some of these habitat and streamflow relationships. Any improvements in our understanding of what drives topminnow population fluctuations would improve our prospects for helping this endangered fish recover in Cienega Creek and elsewhere. Other stream systems that have been studied often show increases in populations of native fishes after large floods, which they have attributed to differential flushing of non-native species and/or to pulses in reproduction of natives initiated by flooding (Meffe 1985; Schultz et al. 2003). Cienega Creek offers a unique opportunity to study responses of Gila topminnow, Gila chub, and longfin dace to streamflow fluctuations in the absence of confounding effects of competition with non-natives.

Managers have more influence on some of these factors than others. Vegetation changes in and around the creek itself are likely to be the most responsive to direct management, along with changes in physical habitat structure that are often associated with changes in stream bank vegetation. Research on how these factors affect topminnow may be the most immediately relevant for implementing adaptive management. Nevertheless, understanding the effects of factors that managers have little control over (e.g. droughts and floods) may improve our ability to detect and respond to changes caused by the factors over which managers exercise more influence.

Possible modifications to the monitoring program

Our ultimate goal in reviewing the Gila topminnow monitoring program is to identify modifications that could give managers more reliable and relevant information without requiring an appreciable increase in resources. Appendix C tests the ability of protocols with a range of sample sizes and sampling frequencies to detect population halving or doubling in ten and eight year time spans. Considering results of these tests along with factors discussed above, options that appear to offer the best tradeoffs between effort and results are:

- Shifting topminnow trend sampling to every-other-year and using alternate years to satisfy other needs, e.g. Gila chub and/or aquatic habitat monitoring;
- Increasing sampling stations in the lower reach from four to either 11 or 16 (for ten-year or eight-year spans, respectively) permanent stations. New sites should be randomly located throughout the lower reach's perennial length. To maintain continuity with past datasets, we recommend keeping three of the existing stations and incorporating them into the random-site array;
- In cases where the previous protocol would have passed over a station's closest pool in favor of one showing more topminnow activity, avoid the tradeoff between detecting exotics and extrapolating topminnow trends by sampling both pools and analyzing data separately;
- Reducing the number of habitat measurements taken at new stations during fish sampling until research demonstrates importance of particular habitat parameters to topminnow in this system;
- Making use of other existing threats monitoring, e.g. using annual mapping of perennial flow to document the amount of potential topminnow habitat gained or lost through time and help interpret fish density results;
- Reducing sampling effort in the upper reach by focusing on screening for exotic species and tracking changes in topminnow distribution through this reach, with the option to resume trend monitoring when topminnow again occupy more of the available habitat. Mattie Canyon and Empire Gulch could receive similar annual "spot-check" monitoring, with the option of occasionally conducting more intensive sampling to check for low level persistence.

Details and justifications for these recommendations are discussed below.

Time spans: Ten year spans are a common default management cycle and are used for aspects of the topminnow recovery plan. Detecting change in eight years allows for two years of management response within the same cycle. Other time-span considerations are presented in Appendix C.

Alternate-year sampling: Monitoring topminnow every other year opens several attractive options for managers. Top on the list is the option of sampling for Gila chub in the off-years. The Las Cienegas Resource Management Plan has explicit chub monitoring requirements that are not being met and chub density targets that are not being measured. Chub are best sampled by electrofishing and/or passive trapping, neither of which is effective at catching topminnow (at least with the same voltages or trap types best suited for chub). Chub sampling has been done opportunistically over the years, but managers have found it difficult to fit both types of fish sampling into the same field season. Listing of the chub as endangered in 2005 raises the importance of satisfying these requirements. Adding regular sampling with different techniques also improves the overall ability of the LCNCA fish monitoring to satisfy its original purpose of screening for exotic fishes and crayfish, since there are several non-natives in the region that are more likely to be caught with the methods used for chub than with the methods used for topminnow. Limits on staff time have also forced BLM to drop its annual topminnow sampling during years scheduled for extensive surveying of aquatic habitat (1990 and 2000). This periodic habitat monitoring includes measurements of pool and fish cover parameters that have specific objectives in the RMP. Skipping years between topminnow sampling would add flexibility of conducting habitat monitoring in response to catastrophic events such as fires or flooding. It is worth noting that because points near the ends of regressions have more influence on the slope than points in the middle (Larsen et al 2001), any necessary skipping of years that are scheduled for topminnow sampling should be done in the middle of the management cycle rather than near the beginning or near the end. Also, our power simulations all included sampling on the last year of the monitoring span (year 8 or 10), i.e. in both of the last 2 years in what was otherwise alternate-year sampling. If future sampling sticks to a strict alternate-year schedule, adequate power would be expected in 9 or 11 years, respectively.

Increasing lower-reach sampling: From a manager's perspective, interpreting and responding to changes will be very different whether one has found declines throughout the creek vs. stability in one reach and more drastic decline in another. For this reason, we believe it is worth continuing to treat the upper and lower reaches separately. The lower reach appears to now be functioning as the fish's stronghold in this system and might therefore be considered the most important area in which to understand topminnow population dynamics. With alternate-year sampling, our simulations predict that 11 or 16 samples will be needed to achieve adequate power to detect population halving or doubling over ten and eight year spans, respectively. In terms of effort required to sample this many sites, choosing the larger number of sites (and expecting meaningful results within a shorter span) is likely to reduce per-site effort somewhat because sites will be closer together and travel time between them will be reduced. Obviously, adding even more stations would enable managers to detect trends more quickly and/or at less extreme rates, but this comes at the expense of additional effort required each year.

Randomizing station locations: The current set of subjectively-placed sampling stations functions well as index sites for spot-checks (AZGFD 2004). However, trend detection across an area larger than the sampling sites themselves technically requires sites to be randomly placed. Adding sites brings an opportunity to redistribute sampling and improve our ability to extrapolate from individual sites to trends across the creek as a whole. We propose that new sites be randomly-located throughout the lower reach's perennial length. Some stratifying may be in order to make sure sites adequately represent the diversity of habitats present in this reach.

The challenge of transitioning from one sampling protocol to another without compromising the comparability of old and new data is one of the main obstacles to standardizing or modernizing monitoring programs (Bonar and Hubert 2002). One common approach is to collect data in both new and old protocols for several years to determine whether results are consistent among protocols (e.g. Larson et al 2001) and/or to generate a correction factor for comparing these data sets. This is unavoidable if collecting techniques are radically changed. However, we are proposing to keep the same basic seining protocol and just expand it to additional sites (with one exception; see below). Furthermore, we see value in keeping at least some of the long-standing sampling stations so long as they do not fundamentally undermine our ability to extrapolate to the rest of the creek. The AZGFD's attempt to standardize monitoring presents a rule-of-thumb recommendation that for trend studies in areas with pre-existing fixed stations, 20% of the total number of sites sampled be composed of these existing stations (dropping a random subset of such stations if there are too many) and 80% be randomly sited. With 11 or 16 total lower reach sites, this would retain two or three of the existing stations. Integrating these existing stations into a network of randomly-placed sites could be done by selecting more random sites than will be needed and then dropping (at random) one of the two sites closest to each of the retained stations. This approach to switching from subjectively-located sites designed for spot-checks to randomly-located sites designed for trend detection offers the benefit of maintaining our ability to compare densities in long-standing sites across time, without having to collect additional data during a transitional period. This approach does relax the strict statistical rules of random placement but we do not believe the effect will substantially alter the monitoring design's ability to accurately represent dynamics of the creek as a whole. In any case, this move towards random sampling is a big statistical improvement over the existing subjectively-sited locations in terms of ability to extrapolate to larger spatial scales.

Expanded sampling offers the opportunity to attain better coverage of the lower reach as a whole. No sampling stations are currently located downstream of the Narrows road crossing. Fish habitat in this unsampled area appears to be rather different from the rest of the lower reach, with a string of large pools that become isolated during low flow times. Observations in June 2006 recorded abundant topminnow in several of these pools. If these new sites show different dynamics than the existing sites, BLM staff analyzing fish data in the future might choose to exclude them from some analyses of long-term trends. However, future data from all sites will likely give a more balanced picture of dynamics in the reach as a whole.

New randomly-placed sites could be kept as fixed stations or could be placed anew each year. We prefer having fixed stations because they enable researchers to factor out site-specific variation when looking for trends, and make it easier to co-locate other monitoring work (e.g. channel cross section measurements, riparian vegetation sampling). Relocating sites each year , however, enables researchers to cover a broader set of sites. Cienega Creek is small enough relative to the number of samples to be taken that we do not believe the expanded coverage would compensate for the loss of year-to-year site comparability.

Standardizing pool selection: We grappled with whether to retain the past practice of choosing among nearby seinable pools on the basis of which one shows the most obvious topminnow activity. This choice almost certainly improved the likelihood of detecting exotics and may well have reduced variation in fish counts in the past. Continuing this non-standardized element of response design, however, compromises the ability to extrapolate trends to the rest of the creek and to compare average fish densities measured here with those encountered elsewhere. Discontinuing this choice and standardizing pool selection (i.e., sampling the seinable pool closest to the original sampling point), however, may compromise the continuity of long term data. It is not clear how severe this compromise would be since there is no record of how often this choice was invoked or how much fish density differed between the pool chosen and the closest seinable pool. We propose a transitional period in which these tradeoffs can be evaluated for existing index sites. In cases where the original criteria would have passed over the closest seinable pool in favor of one with more obvious fish activity, we recommend sampling both pools. Double-sampling would not be necessary at newly established sites since they have no data collected under the original protocol and are being established specifically for detecting density trends. Once several years of data are available for analysis, means and variance can be compared among sets of pools chosen under each protocol. If these do not differ, subjectively selected pools can be dropped from both the sampling protocol and trend analyses. If they do differ, a correction factor can be calculated or double sampling can be continued.

We propose that other aspects of the sampling protocol be left unchanged. Retaining the original response design's directive to choose a pool close to the original sampling point provides the flexibility needed to respond to changes in channel morphology without having to change general site placement each year. Because obstacles such as thick emergent vegetation can make seining ineffective and thereby introduce a large amount of sampling error, we recommend keeping the practice of selecting among nearby pools on the basis of suitability for seining, with clarification of the parameters within which the choice is made (e.g. distance upstream and/or downstream of the sampling point).

Streamlining habitat measurements: BLM's aquatic habitat monitoring has come in two forms: measurements taken across long stretches of the creek (in 1990 and 2000), and measurements taken during yearly fish sampling and applied only to areas seined. The first set has been analyzed extensively (Simms et al. 2006; see also Aquatic Habitat Status and Trends, in prep). With the exception of length and width measurements used to determine the area seined, however, measurements taken during fish sampling have never been digitized or analyzed. These habitat measurements sometimes take as much time as the seining itself.

In the absence of evidence that the existing suite of habitat measurements taken during fish monitoring helps explain topminnow dynamics, we propose that some of this effort might be more effectively spent on sampling fish in additional stations. Some of the parameters (e.g. length of undercut banks and amount of woody debris) are probably more relevant to chub than to topminnow, and some parameters that may well be more relevant to topminnow (e.g. area of shallow or backwater habitat) are not being measured. Better understanding the relationship between topminnow dynamics and various habitat parameters—pushed by Simms and Simms in 1992— is clearly overdue. Any factor demonstrated to have a significant influence on topminnow populations and that can be readily measured during monitoring will account for some of the overall variation in year-to-year or site-to-site fluctuations in topminnow densities and will therefore improve ability to detect trends, as well as provide direct guidance to managers. Determining such relationships, however, is a job better suited to research than monitoring. BLM and partners are currently investigating opportunities to conduct such research. In the meanwhile, we propose that BLM spend less effort collecting habitat data during topminnow sampling by streamlining these measurements and/or dropping some altogether.

Re-focusing upper reach sampling: Now that topminnow populations have declined in the upper reach, boosting sampling intensity enough to track trends here may not be the best use of monitoring resources. Detecting change becomes more difficult as populations decrease (Taylor and Gerrodette 1993). The topminnow's disappearance from some stations and declining density at other stations mean that one would have to sample much more intensively in this reach to reliably detect additional declines or recovery. Nevertheless, early detection of exotics in this reach is as important as ever, and the reach still has the potential to support large topminnow populations in the future.

Shifting the focus of sampling in the upper reach could satisfy the goals of detecting exotics and tracking topminnow recovery potential with substantially less effort than would be required to reliably detect abundance trends. The upper reach could be treated as a spot-check area, with sampling discontinued at the two upper reach sites that have not had fish for the past several years (FC0313, now dry in most summer and fall months, and FC0311, anoxic and fish-free since 2002). The remaining two sites (FC0316 and FC0312) could easily be sampled in one day, and additional opportunistic seining of nearby habitats could be added to this same day's work to improve early detection of non-native fishes. A visual survey along the upper reach's flowing waters would enable managers to track the distribution of topminnow in this reach, which is arguably more important than absolute population numbers to fish recovery in what has apparently become marginal habitat.

Resources required: We feel that with these changes, topminnow monitoring could achieve its major monitoring goals—screening for exotic species, tracking fish distributions throughout Cienega Creek and its major tributaries, and reliably detecting large abundance trends in the fish's lower reach stronghold—without increasing total monitoring effort.

Shifting topminnow monitoring to every-other-year by itself would cut in half the effort expended on sampling. Adding more stations will take up some of these savings, but even doubling the number of stations (from the current 9 total to 16 stations in the lower reach and 2 in the upper) is unlikely to double the number of total sampling days in a given year. From 1992 to 2002 (years for which the original datasheets were immediately available), sampling took place in an average of 4.5 days per year. Two to three-person teams sampled an average of 1.7 stations per day but several times completed four in a day. However, existing sites are widely dispersed and much of the sampling cost is travel time; adding sites between exiting stations would enable field crews to access more in one day. It seems reasonable to expect that with sites closer together, crews could sample three to four sites per day, which would complete the 16 lower-reach station in four to six days. Dropping some habitat measurements would make crews even more efficient. Adding one day for the upper reach gives a total of five to seven sampling days every-other-year—only marginally more than is now being spent every year. Furthermore, BLM staff spend considerable time organizing each year's survey; this effort would be cut roughly in half by sampling in alternate years. For topminnow monitoring, these modifications represent a reduction in total effort while generating better power to detect change. Using alternate years for sampling Gila chub would undoubtedly use up the effort saved on topminnow and might require a modest increase in effort overall, but would measure additional RMP objectives and satisfy endangered species compliance requirements that are currently not being addressed.

Follow-up: Our power analyses used means and variance of past samples to predict how many samples would be required to achieve our desired sensitivity to change. If either means or variances of future samples are very different from those of past samples, the protocol's ability to detect change may also differ from the predictions we present here. Once several rounds of data have been collected under the modified protocol, we recommend examining sample means and variances. If they are outside the bounds used in Appendix C analyses, running the new data through this type of power analysis may provide useful guidance for additional modifications.

Re-running power analyses may also prove useful if monitoring goals change. For instance, if managers decide they want to detect smaller population changes or detect change over a different time span (e.g. to correspond to particular management actions), analyses like those used here can illustrate the cost-benefit tradeoff for protocol changes.

Vital Signs and NPS' Inventory and Monitoring Program

To generate reliable data needed to manage and maintain critical resources, the National Park Service (NPS) established a nationwide Vital Signs Inventory and Monitoring Program composed of 32 park networks grouped by proximity and ecological similarity. The Sonoran Desert Network (SODN) includes national parks and monuments in the Sonoran Desert and the higher-elevation "sky island" region to the east where LCNCA is located. SODN is partnering with professionals from universities, agencies and non-profit organizations to expand its park ecological monitoring framework to other sites in the region and to assist with the development of statistically robust monitoring protocols. Thus, we are collaborating with SODN in our efforts to review, make recommendations, and assist BLM with the development of an ecosystem monitoring plan for LCNCA with the idea that the monitoring parameters and protocols we develop can be broadly applied to other sites (including NPS sites) that share ecosystems, target species, and stresses with LCNCA.

Annual fall fish surveys at Las Cienegas address several aspects of the fish community dynamics vital sign. Surveys were originally designed to provide early detection of exotic species. Fortunately, none have been found as of October 2006, so tracking changes in abundance or distribution of non-natives has not been a part of Las Cienegas fish monitoring. Fall seine sampling provides an index of species diversity, distribution, catch per unit effort, relative abundance, and age structure for all three native species present, and provides opportunities to observe their physical health status.

References cited

Allison, L. J. 2000. Power analysis for long-term monitoring of fishes in selected waters of the Gila River Basin, Arizona. Arizona Game and Fish Department Nongame Branch, Wildlife Management Division. Phoenix, AZ.

Arizona Game and Fish Department. 2004. Standard Fish Sampling Protocol for State of Arizona Waters: Guidelines for the collection, analysis, and interpretation of fisheries data. AZGFD report.

Brooks, J. E. 1985. Factors effecting the success of Gila topminnow (*Poecilliopsis o. occidentalis)* introductions on four Arizona National Forests. Report to the US Fish and Wildlife Service, Office of Endangered Species, Albuquerque, NM. Arizona Game and Fish Department, Phoenix AZ.

Campbell, S. P., Clark, J. A., Crampton, L. H., Guerry, A. D., Hatch, L. T., Hosseini, P. R., Lawler, J. J. & O'Connor, R. J. 2002. An Assessment of Monitoring Efforts in Endangered Species Recovery Plans. Ecological Applications 12[3], 674-681.

Constanz, G.D. 1973. Reproductive effort in *Poecilliopsis occidentalis* (Poecilidae). Southwestern Naturalist 19: 47-52.

Di Stefano, J. 2003. How much power is enough? Against development of an arbitrary convention for statistical power calculations. Functional Ecology 17: 707-709.

Douglas K. Duncan, 2006. Native fish conservation and climate variability in Southeastern Arizona. Paper presented to the Sixth Conference on Research and Resource Management in the Southwestern Deserts: Borders, Boundaries and Time Scales, Tucson AZ, May 2006.

Droege, S., Cyr, A. & Larivee, J. 1998. Checklists: An Under-Used Tool for the Inventory and Monitoring of Plants and Animals. Conservation Biology 12[5], 1134-1138.

Everhart, W. H., A.W. Eipper, W. D. Youngs. 1981. Principles of Fishery Science. Comstock Press, London. 360p.

Gerrodette, T. 1987. A Power Analysis for Detecting Trends. Ecology 68[5], 1364-1372.

Gerrodette, Tim. 1993. TRENDS: Software for a power analysis of linear regression. Wildlife Society Bulletin 21:515-516.

Gibbs, J. E. 1995. *MONITOR* Software for power analysis in population monitoring programs. USGS Pautuxet Wildlife Research Center. htttp://www.mbr-pwrc.usgs.gov/software/monitor.html. Site used September 2006-January 2007.

Gibbs, J. E., S. Droege, and P. Engle. 1998. Monitoring populations of plants and animals. BioScience 48(11): 935-940.

Gori, D., and H. Schussman. 2005. State of the Las Cienegas National Conservation Area. Part 1. Condition and trend of the Desert Grassland and watershed. Prepared by The Nature Conservancy in Arizona. 63 pp. Available at www.azconservation.org/eco_man.html.

Gotelli, N. J. & Ellison, A. M. 2004. A primer of ecological statistics. Sunderland, Mass: Sinauer Associates Publishers.

Griffith, J. S. & Tiersch, T. R. 1989. Ecology of Fishes in Redfield Canyon, Arizona, With Emphasis on Gila-Robusta-Intermedia. *Southwestern Naturalist* 34: 131-134.

Gryska, Andrew D., Wayne A. Hubert, and Kenneth G. Gerow. 1997. Use of power analysis in developing monitoring protocols for the endangered Kendall Warm Springs dace. North American Journal of Fisheries Management 17:1005-1009.

Hawkins, C.P., J.L. Kershner, P.A. Bisson, M.D. Bryant, L.M. Decker, S.V. Gregory, D.A. McCullough, C.K. Overton, G.H. Reeves, R.J. Steedman, and M. K. Young. 1993. A Hierarchical approach to classifying stream habitat features. Fisheries 18(6):3-11.

Hayes, J. P. & Steidl, R. J. 1997. Statistical Power Analysis and Amphibian Population Trends. Conservation Biology 11[1], 273-275.

Hedrick, P. W. & Parker, K. M. 1998. MHC Variation in the Endangered Gila Topminnow. Evolution 52[1], 194-199.

Jassby, A. D. & Powell, T. M. 1990. Detecting Changes in Ecological Time Series. Ecology 71[6], 2044-2052.

Larson, David P., Thomas M. Kinkaid, Steven E. Jacobs, and N. Scott Urquhart. 2001. Designs for evaluating local and regional scale trends. Bioscience 51(12: 1069-1078.

Leenhouts, J. M., Stromberg, J. C., and R. L. Scott. 2006. Hydrologic requirements of and consumptive ground-water use by riparian vegetation along the San Pedro River, Arizona. Reston, Va: U.S. Geological Survey.

Lowe, Charles H., D. S. Hinds, and E. A. Halpern. 1967. Experimental catastrophic selection and tolerances to low oxygen concentration in native Arizona freshwater fishes. Ecology 48(6): 1013-1017. Mapstone, Bruce D. 1995. Scalable Decision Rules for Environmental Impact Studies: Effect Size, Type I, and Type II Errors. Ecological Applications 5[2], 401-410.

McCain M., Fuller D., Decker L., and K. Overton. 1989. Stream Habitat Classification and inventory procedures for northern California. Fish Habitat Relations Program, U.S. Forest Service, Eureka CA.

Meffe, G. K. 1985. Predation and Species Replacement in American Southwestern Fishes - a Case-Study. Southwestern Naturalist 30[2], 173-187.

Meffe, G. K., Hendrickson, D. A., Minckley, W. L. & Rinne, J. N. 1983. Factors Resulting in Decline of the Endangered Sonoran Topminnow Poeciliopsis-Occidentalis (Atheriniformes, Poeciliidae) in the United-States. *Biological Conservation* 25: 135-159.

Mills, L. Scott, Stephen G. Hayes, Calib Baldwin, Michael J. Wisdom, John Citta, David J. Mattson, Kerry Murphy. 1996. Factors Leading to Different Viability Predictions for a Grizzly Bear Data Set. Conservation Biology 10 (3): 863–873.

Minckley, W.L. 1973. Fishes of Arizona. Sims Printing Co, Phoenix AZ. 293pp.

Minckley, W.L. and M.R. Sommerfeld. 1979. Resource inventory for the Gila River complex, eastern Arizona. Final report for U.S. BLM. Contract YA-512-CT6-216, Arizona State Univ., Tempe.

Minckley, W. L. 1999. Ecological Review and Management Recommendations for Recovery of the Endangered Gila Topminnow. *Great Basin Naturalist* 59: 230-244.

Morrison, L. W. 2007. Assessing the Reliability of Ecological Monitoring Data: Power Analysis and Alternative Approaches. 27: 83-91.

Nelsom, Benjamin T. 1993. Spawning characteristics of Gila chub in Cienega Creek, Arizona. Report for the USDI Bureau of Land Management, Tucson Resource Area, Arizona.

Parker, K. M., Sheffer, R. J. & Hedrick, P. W. 1999. Molecular Variation and Evolutionarily Significant Units in the Endangered Gila Topminnow. *Conservation Biology* 13: 108-116.

Parkinson, E.A., J. Berkowitz, and C.J. Bull. 1988. Sample size requirements for detecting changes in some fisheries statistics from small trout lakes. North American Journal of Fisheries Management 8:181-190.

Peterman, Randall M. 1990. The Importance of Reporting Statistical Power: The Forest Decline and Acidic Deposition Example. Ecology 71[5], 2024-2027.

Platts, W.S., Megahan W.F. and G. W. Minshall. 1983. Methods for evaluating stream, riparian and biotic conditions. U.S. Forest Service, General Technical Report INT-138.

Reed, Michael J., and Andrew R. Blaustein 1997. Biologically significant population declines and statistical power. Conservation Biology 11(1): 281-282.

Rice, C.G., J. Rohlman, J. Beecham, and S. Pozzanghera. 2001. Power analysis of bait station surveys in Idaho and Washington. Ursus 12: 227-236.

Reed, J. M. & Blaustein, A. R. 1997. Biologically Significant Population Declines and Statistical Power. Conservation Biology 11[1], 281-282.

Richter, B. D. A. T. W. J. L. M. K. L. 2006. A collaborative and adaptive process for developing environmental flow recommendations. *River Research and Applications* 22 : 297-318.

Rohlf, F. J., Sokal, R. R. & Sokal, R. R. 1994. Statistical tables. San Francisco: Freeman.

Schoenherr, A. A. 1977. Density Dependent and Density Independent Regulation of Reproduction in the Gila Topminnow, Poeciliopsis Occidentalis (Baird and Girard). Ecology 58[2], 438-444.

Schultz, Andrew A., and Scott A. Bonar. 2004. Selected aspects of the natural history of the Gila chub, 2002-2004 progress report. Arizona Cooperative Fish and Wildlife Research Unit, Fisheries Research Report 04-04.

Simms, J.R. and K.M. Simms, 1992. What constitutes high quality habitat for Gila topminnow (*Poeciliopsis o. occidentalis)?* An overview of habitat parameters supporting a robust population at Cienega Creek, Pima County, AZ. Abstract, Proceedings of the Desert Fishes Council 22-23: 82.

Simms. K., G. Drennen, B. Cooper, G. Bodner, and D. Gori. 2006. Report on the Biological Planning Process for livestock management at the Las Cienegas National Conservation Area: Fall 2005 upland monitoring results and adaptive management by the Bureau of Land Management, Technical Resource Team, and Rangeland Resource Team. Report prepared by the BLM and The Nature Conservancy in Arizona. 29 pp. Available at www.azconservation.org/eco_man.html.

Sokal, R.R., and F.J. Rohlf. 1995. *Biometry.* 3rd edition. W. H. Freeman, New York. 887pp.

Steidl, R. J. & Anthony, R. G. 1996. Responses of Bald Eagles to Human Activity During the Summer in Interior Alaska. Ecological Applications 6[2], 482-491.

Steidl, R. J. 2001. Practical and statistical considerations for designing population monitoring programs. Pages 284-288 in R. Field, R. J. Warren, H. Okarma, and P. R. Sievert, editors. *Wildlife, land and people: priorities for the 21st century. Proceedings of the Second International Wildlife Management Congress.* The Wildlife Society, Bethesda, Maryland, USA.

Taylor, B. L. & Gerrodette, T. 1993. The Uses of Statistical Power in Conservation Biology: The Vaquita and Northern Spotted Owl. Conservation Biology 7[3], 489-500.

Taylor, B. L. 1995. The reliability of using population viability analysis for risk classification of species. Conservation Biology 9: 551-558.

Tharme, R. E. 2003. A global perspective on environmental flow assessment: emerging trends in the development and application of environmental flow methodologies for rivers. *River Research and Applications* 19: 397-441.

Thomas, L. 1997. Retrospective Power Analysis. Conservation Biology 11[1], 276-280.

Thompson, P. M., Wilson, B., Grellier, K. & Hammond, P. S. 2000. Combining Power Analysis and Population Viability Analysis to Compare Traditional and Precautionary Approaches to Conservation of Coastal Cetaceans. Conservation Biology 14[5], 1253-1263.

U.S. Department of the Interior. 1967. Native fish and wildlife. Endangered Species. Federal Register 32(48):4001.

Urquhart, N. S., W. S. Overton, and D. S. Birkes. 1993. Comparing sampling designs for the monitoring of ecological status and trends: impact of temporal patterns. Pages 71-85 in V. Barnet and K. F. Turkman, editors. Statistics for the Environment. John Wiley and Sons, London, UK.

Urquhart, N.S., S.G. Paulsen, and D.P. Larsen. 1998. Monitoring for policy-relevant regional trends over time. Ecological Applications 8:246-257.

U.S. Department of Interior, Bureau of Land Management. 1994. Arizona statewide wild and scenic river legislative EIS - Final. BLM State Office, Phonix, AZ. Report BLM/AZ/PL-95/002+4333. 551 pp.

U.S. Fish and Wildlife Service. 2001. Biological and conference opinion summary, Effects of the proposed Gila topminnow reestablishment in Empire Gulch in Pima County, Arizona. USFWS Arizona Ecological Services Field Office.

U.S. Fish and Wildlife Service. 2002. Biological And Conference Opinion, Las Cienegas National Conservation Area Resource Management Plan. Arizona Ecological Services Field Office. Report # AESO/SE 2-21-02-F-162, October 4, 2002.

Voeltz, J. B., and R. H. Bettaso. 2003. Status of the Gila topminnow and desert pupfish in Arizona. Ariz. Game and Fish Dept. Technical Report 226, Phoenix. 124pp.

Weedman, D. A. & Young, K. L. 1997a. Status of the Gila topminnow and desert pupfish in Arizona. Technical Report 118, Arizona Game and Fish Department, Phoenix Arizona.

Weedman, D. A., Girmendonk, A. L., and K. L. Young. 1997b. Gila topminnow sites in Arizona: Provisional extirpation report for the 1996-97 field season. Technical Report 116, Arizona Game and Fish Department, Phoenix Arizona.

Weedman, D. A. 1999. Draft Gila topminnow, *Poeciliopsis occidentalis occidentalis*, revised recovery plan. Prepared by Arizona Game and Fish Department for U.S. Fish and Wildlife Service, Albuquerque, NM.

White, G.D., D.R. Anderson, K.P. Burnham and D.L Oatis. 1982. Capture-recapture and removal methods for sampling closed populations. LA-8787-NERP, Los Alamos National Laboratory, Los Alamos, NM

Young, K.L. and M.A. Lopez. 1995. Fall fish count summary: 1988-1994. Tech. Report 81. Ariz. Game and Fish Dpt., Phoenix, AZ

Zar, J. 1999. *Biostatistical Analysis*. 4th Ed. Prentice Hall. ISBN

Appendix A: Fall Fish Count Methodology, Cienega Creek

Site locations: There are nine AGFD "Fall Fish Count" locations on Cienega Creek. Seining locations (stations) for topminnow were less than 500 meters long.

FC0309	Mattie Cyn Headwaters	T18S,R17E,SEC.26,NW,NE
FC0310A	Fresno Gap	T18S,R17E,SEC.23, NE,NE
FC0310B	Cold water Exclosure	T18S,R17E,SEC.23,NE,NE
FC0311	Headwaters	T18S,R17E,SEC.15,SE
FC0312	below Gardener Cyn	T19S,R17E,SEC.10,SE,NE
FC0313	Below Spring Water Cyn	T19S,R17E,SEC.3 SE, NE
FC0314	Pump Canyon	T18S,R17E,SEC.13,NW
FC0315	Above Fresno Cyn	T18S,R17E,SEC.12, E½
FC0316	Above Gardener Cyn	T19S,R17E,SEC.15,NE,NE

Depletion Seining: a single pool or glide is blocked with long seines (1/8" mesh) that are staked into the stream bed. Habitats with a large numbers of fish visible to the observer and channel characteristics favorable to seining are selected over those with no or few fish visible or are not easily seined (non random sampling). The float-line is then elevated above the water surface to discourage fish from jumping over the top. The nets are set in a manner to avoid scaring fish into or out of the depletion area. All debris and vegetation that might interfere with seining is then removed. Seining is repeated until a haul produces < 10% of the number of topminnow captured in the most populated previous haul (White et al. 1982). Each seine haul is carefully conducted to minimize fish escape and is repeated using the same rate of drag and technique to maximize uniformity. Each seine haul is conducted with the brail at waters' edge with collectors walking vigorously through edge cover to scare topminnow to the center of the water column. The number of fish collected on each haul is recorded. The lead line is carefully dragged up onto shore to minimize fish escape under the net. The smooth sand/silt bottom of Cienega Creek is ideal for maintaining constant leadline contact with the gentle bottom contours.

This type of sampling has been conducted from 1989 to 1995. In the past fish have been classified and recorded by sex. Topminnow are enumerated as one age class due to the difficulty of measuring small fish. Longfin dace are classified by age/size class (>45mm = adult) and any Gila chub captured are measured (total length) and classified to age/size class (>90mm = adult).

All collections are screened for the presence of western mosquito fish or other exotic species, and for signs of disease.

Seining Single Pass: Several habitats are selected. Habitats with a large numbers of fish visible to the observer and channel characteristics favorable to seining are selected over those with few or no fish visible or in which physical obstacles or very deep pools make them unfavorable for seining (non random sampling). Each habitat (pool or glide) is seined in the portion where topminnow abundance is greatest; this is done to capture the site's school as effectively as possible (topminnow are non-territorial and aggregate when disturbed). All topminnow are counted and screened for the presence of mosquitofish. Longfin dace are categorized as to age class (adult vs. juvenile) and counted; Gila chub encountered are measured (total length).

Depletion Shocking: Gila chub are surveyed using backpack electroshocking gear in two or three habitats. Pools with a depth greater than 2 feet are sampled by three-pass depletion sampling (Platts et al. 19983) with each pass being of similar effort (real time shocking seconds). Block netting is rarely done due to the secretive nature of the fish which gives it a predisposition of hiding under cover rather than fleeing when disturbed. All fish are measured (total length) and released. Presence of parasites and reproductive coloration and condition is noted. The pass sampling provides a good measure of the population size when electroshocking (Minckley and Sommerfeld. 1979).

Characterizing Habitats Sampled During Topminnow Surveys: The same macrohabitats are seined annually. However, streams often change character over time. Therefore conditions are comparable but rarely the same from year to year or site to site. The documentation of habitat changes may be useful for distinguishing causes for large changes in the Gila topminnow population.

Only the portion of a macrohabitat that is sampled for fish has its habitat characteristics measured, using the methods described below for aquatic habitat inventories.

Characterizing Habitat For Inventory and Monitoring: Macrohabitats are delineated using the classification in McCain et al (1989) and Hawkins et al (1993) with the addition of marsh and sheet-flow macrohabitat types. The following information is recorded by the habitat or portion thereof sampled: length, average channel width, mean water depth, maximum depth, over-hanging cover by shrubs/tall-grass (in ft^2), floating vegetation (in ft^2), and debris (in ft^2), emergent cover (in ft^2), undercut bank (linear ft.), overstory canopy cover (ocular estimate by two or more observers), and substrate (dominant three types of the following (boulder, cobble, pebble, gravel, sand, and silt cover).

Average width was taken at a single location in the middle of a habitat when the habitat was small and uniform. If the habitat was large (length greater than 10 widths or banks irregular, then three or more width measurements were taken and averaged. Average depth was estimated by taking depth measurements at ¼, ½, and ¾ the distance across the wetted width and averaging these. When the banks were shear, we used a standard averaging procedure (ie., adding three values and dividing the total by three); when the banks sloped into the water gradually, the summed three depth values were divided by four better represent overall depth (Platts et al. 1983). Average depth was estimated in one representative location in short habitats and at three or more locations for long habitats. The average of multiple average depth estimates was recorded as the final average depth.

Percent overstory was estimated by two or more observes. Each estimate was generally within 10% for each observer and the average of the estimates was recorded. The results are presented in the following five categories: 0%, 1-25%, 26-50%, 51-75%, 76-100%.

Length and widths were measured by a recently calibrated HAGLOF DME 201 to the nearest 0.1 foot (3 cm). Depths were measured using a surveyor's rod to the nearest 0.1 foot (3 cm).

Appendix B: Confidence interval calculations for fish trends

The generalized equation for confidence intervals around the slope parameter is:

$$\hat{\kappa} \pm \sqrt{F_{crit(\alpha,1,n-2)}} \times \hat{\sigma}_{\hat{\kappa}} \qquad \text{(Thomas 1997, equation A15)}$$

To calculate such a confidence intervals around a slope estimate, one needs:

- An estimate of slope, $\hat{\kappa}$, from the regression of log-count vs. time;
- Choice of desired confidence level; the confidence interval is calculated as $1\text{-}\alpha$, where α is the rate of Type I error one deems acceptable. For a 90% confidence interval like that used here, $\alpha = 0.1$ or 10%;
- The F-ratio value $F_{crit(\alpha,1,n-2)}$ for the combination of α and the study's sample size, n; the F_{crit} value is available from any standard set of statistical tables;
- Estimate of the standard error of the slope estimate, $\hat{\sigma}_{\hat{\kappa}}$.

  Most statistical programs report $\hat{\sigma}_{\hat{\kappa}}$ directly. If a statistical package does not report $\hat{\sigma}_{\hat{\kappa}}$, one can calculate it as $\hat{\sigma}_{\hat{\kappa}} = \sqrt{MSE / SS_x}$. Mean Square Error, MSE, is reported in the ANOVA table when one performs the original regression, or can be calculated from the data. Sums of Squares for X, SS_x, is part of the regression process but is often not reported in an ANOVA table, It may be calculated directly, as $SS_x = \sum (\overline{X} - X)^2$ where X is the time measure, e.g. year, for each data point, and $\overline{X}$ is the average of all time measures.

It is worth noting that some practitioners may be accustomed to a slightly different equation for slope estimate confidence intervals. Gotelli and Ellison (2004) use the somewhat less general equation $\hat{\kappa} \pm t_{crit(\alpha,\, n\,-\,2)} \times \hat{\sigma}_{\hat{\kappa}}$, based on the t distribution rather than the f distribution. For the same denominator degrees of freedom, the square root of f_{crit} is approximately equal to t_{crit}.

The example below demonstrates how confidence intervals around trend estimates were calculated for this study. While the F distribution can be found in most publications of statistical tables, e.g. Sokal and Rohlf 1994, we include an abbreviated version below to facilitate similar confidence interval calculations. F_{crit} values for *df*'s not specified here can be interpolated between numbers given below.

Note that the initial slope estimate can be done with separate counts for each site, or with multi-site yearly averages. Confidence intervals will be similar; using yearly averages gives a smaller standard error around the slope estimate but a larger F_{crit} (fewer degrees of freedom).

Whole creek:

Regression of log topminnow counts per $50m^2$ gave us a slope estimate $\hat{\kappa}$ of -0.156, with the standard error of this estimate $\hat{\sigma}_{\hat{\kappa}} = 0.045$.

The F ratio is based on comparing the variance of the regression to the variance of the error term. Regressions of this sort have a single numerator degree of freedom. With 105 samples, this study has 103 denominator (error) degrees of freedom. We have chosen to accept a 10% Type I error rate (i.e., $\alpha = 0.1$), so we are constructing a 90% confidence interval. Statistical tables tell us that $F_{crit(0.1,1,103)}$) is approximately 2.76.

Substituting estimated values into the equation above gives us:

$$-0.156 \pm \sqrt{2.76} \times 0.045$$

The upper limit to our 90% confidence interval for topminnow population change in the whole creek is therefore -0.081 (an 8% annual decline), and the lower limit is -0.231 (a 23% annual decline).

If our statistical package had not reported $\hat{\sigma}_{\hat{k}}$ directly, we could have calculated is as $\hat{\sigma}_{\hat{k}} = \sqrt{MSE / SS_x}$, with MSE = 5.3 (from ANOVA output), and $SS_x = \sum (\overline{X} - X)^2$ where $\overline{X}$ is 1996.6 (the average of year from all data points) and X is the year in which each data point is taken. In this case, SS_x is 2643.

Abbreviated F distribution for $\alpha = 0.1$, numerator *df* = 1.

Study *df*	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Fcrit	39.86	8.53	5.54	4.54	4.06	3.78	3.60	3.46	3.36	3.29	3.23	3.18
Study *df*	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
Fcrit	3.14	3.10	3.07	3.05	3.03	3.01	2.99	2.97	2.96	2.95	2.94	2.93
Study *df*	**25**	**26**	**27**	**28**	**29**	**30**	**40**	**60**	**120**	**inf**		
Fcrit	2.92	2.91	2.90	2.89	2.89	2.88	2.84	2.79	2.75	2.71		

Appendix C: Power analyses and evaluation of possible sampling design modifications

Introduction and Methods

Detecting population trends

In a general sense, a study's "power" is how well it is able to detect real change. Such power depends on three key factors: how populations vary across space and time, the amount of change one wants to detect, and the study design. Power analysis for tracking trends looks at the population fluctuations and sampling error observed by a monitoring project, and estimates how often the project's sampling design will actually detect a change if that change is, in fact, happening. This type of analysis is critical to interpreting results from wildlife studies, but has only recently become widespread. Reviews of past wildlife and fisheries studies show that many are inadequate to reliably detect important changes (Peterman 1990; Reed and Blaustein 1997; Maxwell and Jennings 2005), even declines large enough to drive populations extinct within the span of the study (Taylor and Gerrodette 1993) or to undermine future recovery from resource depletion (Maxwell and Jennings 2005). Without examining a study's power, one often cannot tell whether a population appears to be stable because it truly is stable, or because the study is too weak to pick up on real changes.

Doing a power analysis requires that one specify the minimum amount of change that one wants to be able to detect (the effect size), or viewed conversely, the maximum amount one is willing to let go undetected. Effect size targets can be set according to criteria such as objectives from management or recovery plans, biological thresholds that have been observed or hypothesized (e.g. from state-and-transition models or population viability analyses), or more-or-less arbitrary benchmarks. Such targets should clearly be outside the range of normal year-to-year variations, either in rate or in duration.

Neither the Las Cienegas National Conservation Area Resource Management Plan nor the USFWS Gila Topminnow Recovery Plan have specific population size or density objectives for the Gila topminnow. No population viability analyses have been published for this species nor are any thresholds of population response apparent from existing topminnow research, though both merit further investigation. We accept that setting a "biologically meaningful" threshold for the amount of change to be detected is somewhat arbitrary at this time, but propose a working target of being able to detect at least an overall doubling or halving of the population. When monitoring began, BLM's initial population estimates for topminnow in this creek exceeded two million (Simms and Simms 1992). These are short-lived animals with high reproductive capacities (Minckley 1999). With such a large initial population and the potential for rapid increases, liberal rate-of-change thresholds are justifiable unless the overall population dips below an alarmingly low total (e.g. the effective population size of 500 -1000 cited bySoule [1980]; Franklin and Frankham [1998]; but see Lande [1995]'s suggested increase to 5000). Because topminnow populations are known to fluctuate wildly with changes in stream conditions (Meffe et al. 1983), we propose focusing on trends that continue for several years. Trends associated with factors that have caused rapid and persistent declines elsewhere (e.g. introduction of non-native fishes, severe erosion, or stream dewatering) would be an exception and we suggest that these be seen as important for management as soon as they are detected.

Opportunities to make monitoring programs more effective and/or efficient: Ability to detect change in fish populations depends on how fish populations vary across space and time, the relationship of the index being measured to the population itself, and the sampling design (which includes number and spatial distribution of samples, sampling frequency, and specific sampling protocols). The first of these is beyond our control, though understanding it can help us optimize our sampling design. The second for now remains an assumption, that counts per square meter of seine haul is an accurate representation of topminnow density in this creek; any modification that can tighten the relationship between numbers of fish caught and true fish density (i.e., reduce sampling error) will improve our power to detect change. However, there may not be much room to improve on what experienced professionals have already implemented here. Our clearest target for improving power of this monitoring program, then, lies in sampling design.

Radical changes in sampling design can reduce the value of a long term study by making it difficult to compare recent data with information collected early on. The least invasive changes to an ongoing monitoring program such as this lie in changing the number of samples taken, size of sample (in this case, length of seine haul), and sampling frequency. How sampling sites are distributed across a landscape is clearly of interest as well, though changing this parameter may require some additional data collection to bridge new and old data sets. We start by evaluating changes in number of samples, then consider sampling frequency. Options for other possible changes are discussed here but not analyzed in detail.

Time spans: For monitoring information to be used in making management decisions, the monitoring program needs to provide results in the same time frame as relevant decisions are or need to be made. However, LCNCA has no formal schedule for topminnow-related management cycles (unlike grassland management which requires that grazing decision be made on an annual basis and therefore requires that monitoring be able to detect year-to-year changes as well as longer trends). Conversations with BLM staff suggest that many such management goals that do not have set schedules also end up being reviewed approximately every ten years (J. Simms, pers. comm.. 2005). The topminnow Recovery Plan, however, does set out some specific time frames; the main one states that populations will not be considered secure unless landowners or land managers “have demonstrated over a period of not less than ten years adequate capability to protect and manage a viable population,” and have enough statutory authority and regulations to enforce protection strategies (Weedman 1999). The USFWS is tasked with reviewing status of all endangered species every five years (Endangered Species Act, Section 4(c)(2)) and producing more detailed analysis of progress towards recovery for individual species every two years (ESA Section 4(f)(3)). In the past, resource constraints have delayed and/or limited the depth of many such reviews; having reliable information on population trends every ten years would substantially increase the rigor with which these reviews could be conducted.

Based on time spans stipulated in the Recovery Plan and on general agreement among agency personnel about default lengths of management cycles, we consider a ten year span as a benchmark for being able to detect important change in Gila topminnow populations. From a manager's perspective, finding out about a change before the end of a decision-making period greatly increases one's ability to respond to such change effectively, e.g. to try to halt a decline or expand on an improvement. To allow for two years of response time within the same ten-year cycle, we also examine power to detect trends over an eight-year span. Clearly there is value in being able to detect trends even sooner, so we also present a different approach to power analysis that can be used to compare any two sampling periods.

Acceptable error levels: Power analysis requires that one specify acceptable levels of α (Type I error, claiming a change where none exists) and β (Type II error, failing to see a change where there is one), where power = $1- \beta$. Traditional ecology studies set $\alpha = 0.05$. Most such studies have not considered power; those that do typically set $\beta = 0.2$ (i.e., power = 0.80; di Stefano 2003). This ratio of α to β implies that researchers consider a false alarm to be four times as harmful as missing a real change, though the researchers themselves may not recognize this implication. Authors have begun pushing for more conscious decisions setting α and β, including proposing guidelines for how to weigh relative costs (in a broad sense, costs can include harm to a species as well as economic expense) associated with Type I and Type II errors and even providing mechanisms for treating α and β as variables (e.g. Mapstone 1995; Di Stefano 2003). Some researchers now set $\alpha > \beta$ when they perceive the harm caused by failing to detect a real change is likely to be more severe than the harm caused by having to investigate a false alarm, as in the case of an endangered species with small population size (e.g. Grska 1997; Gibbs 1998). When relative cost of each error is the same, or when one lacks information about costs, several authors recommend setting $\alpha = \beta$; this has become fairly common in wildlife studies (e.g. Sandin and Johnson 2000; Rice et al. 2001; Sims et al. 2006). We take this last approach, and choose $\alpha = \beta = 0.2$ (accepting a two-in-ten chance of perceiving a change where there is none and an eight-in-ten chance of correctly identifying a change if there is one).

Sufficiency criteria: To judge alternative sampling designs, we began with the same criteria for adequate power discussed above: each design should be able to detect a halving or doubling of the population index over a given time span, with power of at least 80% and $\alpha \leq 0.2$. If populations are changing exponentially, as they appear to have been over the past 16 years (see above), this overall change would translate to a 7% annual decline or 8% annual increase over a ten-year span. We refer to this as the preferred target.

Testing for differences between two population measures

The simplest possible trend analysis involves comparing average catch per unit effort (CPUE) from two different sampling periods. One can easily estimate the number of samples (n) needed to determine if a certain percent change in CPUE has occurred using equations from Parkinson et al. 1988:

$$n = \frac{100^2 k \left(\frac{s}{\bar{x}} \right)^2}{A^2}$$

where n is sample size, s is the sample standard deviation, $\bar{x}$ is the sample mean (which is also the population estimate), and A is the difference in means that one wants to be able to detect (the effect size, expressed as a percent; to detect a 50% difference, A = 50). The constant *k* is a factor of the Type I and Type II error rates one has chosen to accept, and can be looked up in statistical charts in Parkinson et al. 1988 or in other statistical tables. For comparison with other sample size calculators, note that $s/\bar{x}$ here is the coefficient of variation, CV. Both the mean and standard deviation should be among sampling sites within one sampling period, e.g. among all LCNCA fish count sites in one year. This is the procedure advocated by Arizona Game and Fish Department for estimating necessary sample size for trend studies (AZGFD 2004). Note that this procedure is for comparing means from two sampling events; it can be used with any two points of a multi-year study, but does not take advantage of information from intervening or surrounding years.

We used this procedure to estimate how many samples would be needed to detect a 50% change in POOC/m^2, at two levels of false-change error ($\alpha = 0.1$ and $\alpha = 0.2$). For estimates of sample variation for each reach, we averaged the within-year CV (i.e., the among-sample standard deviation for each year, divided by that year's mean count) over all years.

Simulation methods for detecting multiple-year trends in population indices

Incorporating additional years of data can enable analyses to differentiate between year-to-year fluctuations and ongoing trends, but requires more complex tools for power analyses. Most standard statistical software packages have some capacity to do power analysis, but few are designed to be used with multi-year trend data. Of those that do, the program MONITOR (Gibbs 1995) combined the best features for this analysis: its structure closely mimics the type of trend analysis we would like to perform, it can accommodate many design choices including skipping years between sampling, and it is freely available online with a user-friendly interface (htttp://www.mbr-pwrc.usgs.gov/software/monitor.html). This last factor makes it useable by a wide range of wildlife professionals.

MONITOR takes user-input estimates of a population index (e.g. a measurement of density or abundance) and variance in that index (in time and/or space; see below) to generate simulated populations, and subjects each of these simulated populations to a range of annual increases and decreases for a given number of years (the monitoring span). The program then samples from these simulated populations according to designs set by the user, and calculates the percentage of these sampling runs that have succeeded in detecting actual trends of various magnitudes. This percentage represents the estimated power of each monitoring design to detect a given amount of change over the specified time frame, assuming that the original abundance and variance estimates used accurately represent the population being sampled.

This analysis used many simulation runs, each with a unique combination of parameters. We set up two time scenarios, one with a monitoring span of ten years and another with a span of eight years. For each of these, we ran a set of simulations sampling every year and a separate set with sampling done every other year. Each combination of monitoring span and sampling interval was run separately for the upper reach (n = 4 stations; FC0311, FC0312, FC0313, and FC0316), the lower reach (n = 4 stations; FC0310A, FC0310B, FC0314, and FC0315) and the creek as a whole (n = 9 stations: the 8 above plus FC0309 in Mattie Canyon just upstream of its confluence with Cienega Creek). For each combination of reach*monitoring span*sampling interval, we ran two sets of simulations, one with low estimates of variance on density and one with high variance estimates (see below). Each set began with the number of sampling stations in use now and added more stations until power to detect a 50% overall decline over the monitoring span exceeded 80% ($\alpha = 0.2$). Several additional scenarios were run to explore sensitivity to changes in a variety of parameters.

Density and variance estimates were taken from BLM's 16-year fall fish count dataset (Table 1). For density estimates, we used multi-year mean counts for each sampling station. Scenarios that tested effects of adding more stations modeled the additional stations with an average of mean counts from all stations in the reach being examined. All stations were weighted equally in the main simulation sets; additional simulations run with weight equal to mean density, which may better represent overall trends by emphasizing changes at sites with higher densities (Gibbs 1995), gave similar results. Additional analyses explored the effects of using only recent years' data (2002-2005) to calculate these means and variances (see Discussion).

How to best estimate sampling variance for power analysis is a matter of considerable debate (Gibbs 1995, Gerrodette 1993, Larsen et al 2001). To get around some of this debate, we used a conservative approach to estimating variance (i.e., one that errs on the side of high estimates), calculated a confidence interval around this estimate, and ran a separate set of power analyses with the upper and the lower bounds of this confidence interval. We used the residual Mean Squared Error estimate from a linear regression of fish counts vs. time for all years sampled as our primary variance estimate since it includes all sample variation in fish counts not attributable to trends. 95% confidence intervals around these estimates of variance were calculated using multipliers for the shortest unbiased confidence limits for the variance (Sokal and Rohlf 1995). These multipliers are based on the X^2 distribution and are given in standard statistical tables (e.g. Table O in Sokal and Rohlf 1994), where *df* = total number of samples n minus1 for each analysis, e.g. 105 for the whole creek and 50 for the lower reach.

Table 1. Values used in main sets of simulation runs. MONITOR simulation inputs are strings of station mean, reach variance, and weight (1) for each simulated station. Station means were calculated across all years of data for each station, overall means were from all years and all stations of each reach. Variance estimates come from regressions of POOC/50m^2 versus year. Each simulation run used individual station means up to the number of samples actually taken in this monitoring program and overall means to model additional stations, and the same variance estimate for all stations (upper or lower confidence limit RMSE's).

Reach	Mean POOC/50m^2, stations (overall)	Variance		
		RMSE from regression (model *df*)	Lower 95%CI	Upper 95%CI
Whole creek	437, 158, 265, 210, 198, 1454, 374, 272, 247 (423)	866 (105)	759	1002
Lower reach	437, 158, 265, 210 (274)	432 (50)	359	533
Upper reach	1454, 374, 272, 247 (604)	1135 (48)	924	1439

Results

Two-event comparisons of CPUE:

The number of samples needed to detect a 50% difference in CPUE between two time periods is strikingly large, with up to 62 samples needed (whole creek, $\alpha = 0.2$, power = 0.8) for each sampling period (Table 2). This great effort required reflects the fact that this study system has large spatial variation and sampling error. Any changes in protocol that could significantly reduce sampling error are likely to pay for themselves many times over by raising statistical power. Results presented in Table 2 use topminnow caught per square meter as the CPUE; results are virtually identical if the analysis is done with counts per seine haul rather than counts per square meter. It is worth noting here that mean and median yearly CVs were very close, so half of the years had larger CVs than the ones used in these sample size calculations. In other words, these numbers would be inadequate somewhat less than half the time (when comparing two years, lower-than-average variance in one year may compensate for higher-than-average variance in the second); using the 75th percentile CV's suggests that many more samples would be needed if we want to be assured of having adequate power in most years (to detect population halving, n=104, 70, and 130 for whole creek, lower reach, and upper reach respectively, using the 75th percentile values of 16 years of annual CVs).

Table 2. Number of annual samples (i.e., seine hauls) needed to detect a halving (50% decline) or doubling (100% increase) in Gila topminnow per square meter in Cienega Creek between two sampling dates. The coefficient of variation (CV) was calculated for each year as the standard deviation of fish densities among sites divided by the mean of all sites, and were then averaged among all years.

Reach	Average within-year CV, 1989 to 2005	Number of samples needed to detect CPUE changes, $\alpha = 0.2$, power = 0.8	
		halving	doubling
Whole creek	1.31	62	15
Lower reach	1.02	38	9
Upper reach	1.10	43	11

Multi-year trend simulations:

Compared with two-event comparisons of CPUE, running regressions for trends with data from additional years radically reduces the number of samples predicted to be necessary to meet our power goals.

Compared with existing sampling designs, Figures 1 through 3 show how adding a moderate number of sampling stations improves our power to detect change across all scenarios, though the value of adding one more sample is greater in some scenarios than others. Increases were easier to detect (i.e., required fewer samples to reach each threshold of power) than decreases. As expected, simulation runs using lower variance values produced better power than comparable runs using higher variance estimates, though this difference is not large; to be conservative, we focus the rest of our discussion on simulations that use the higher variance estimate. Reducing sampling frequency to every-other-year also reduced power somewhat. Shortening the monitoring span by two years reduced power more dramatically than either changing the variance used or the sampling frequency. The notably lower power of eight-year vs. ten-year spans for a given number of stations holds whether one compares their ability to detect the same annual change or the same overall change (e.g. 7% annual decline over ten years vs. 9% annual decline over eight years).

Treating the whole creek as one unit: For the creek as a whole, simulations that used the same number of yearly samples currently being taken at Las Cienegas (9) did meet our original target of detecting a halving or doubling of the population over a ten year span—but only when the variance input was low (Figure 1a, top line). To be conservative and allow for the possibility that variances are as high as the upper bounds we estimated, these simulations suggest BLM would need to add two stations to meet this same target power and sensitivity. Adding just two additional samples, for a total of 13, increased our simulations' power enough that sampling every other year still achieved adequate power to detect halving of the population (see Figure 1 and Table 3). To detect population doubling, nine samples gave adequate power with yearly or every other year sampling.

Table 3. Minimum number of samples that simulations needed to detect specified amounts of change in Gila topminnow populations in Cienega Creek, with a power > 80% and $\alpha = 0.2$. These values are taken from graphs in Figures 1-3. Values shown here are from simulations using upper variance estimates; those in parentheses are for sampling every other year. The center column's bold values represent our original targets: adequate power to detect an overall halving or doubling of the population over ten year span with yearly sampling. The column to the right shows how additional effort can detect the same annual change two years sooner.

	Existing sampling stations	7% annual decline over a 10-year span	7% annual decline over an 8-year span
Whole creek	9	**11** (13)	17 (20)
Lower reach	4	**8** (11)	14 (16)
Upper reach	4	**9** (11)*	15 (20)*
Overall decline before change is likely to be detected		52% Target sensitivity	44% Target sensitivity, faster detection

*Using only recent years' data for simulations shows that many more samples will now be required to detect these change in the upper reach; see Discussion.

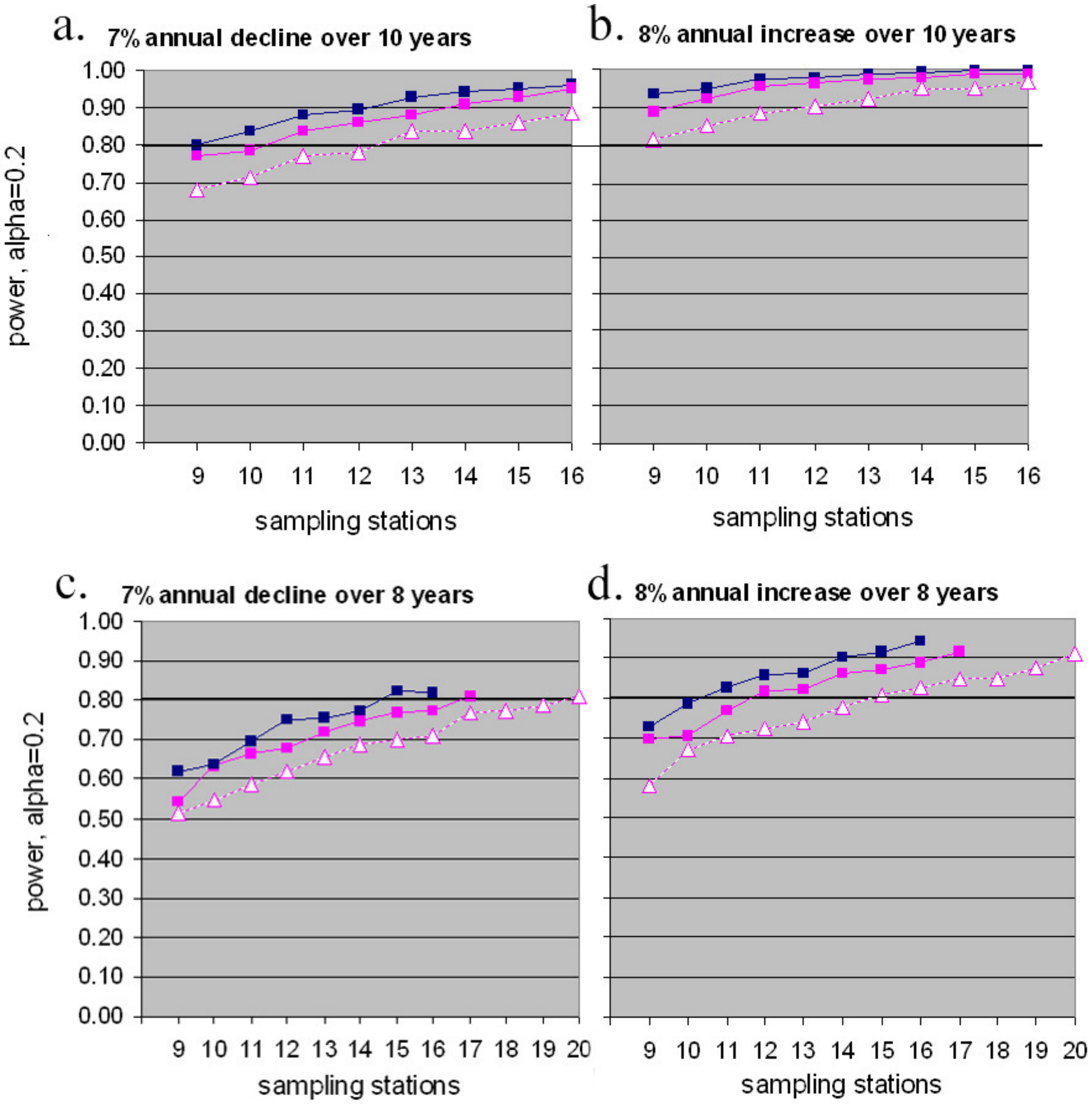


Figure 1. Whole creek, simulated power to detect specified annual changes over ten year (a-b) and eight year (c-d) time spans with increasing numbers of sampling stations. Over ten years, 7% annual exponential decline and 8% annual increase translate to an overall halving or doubling of the population respectively. Our threshold for adequate power is 0.8, or 80%. The left-most point of each curve represents the number of sites currently being sampled at Las Cienegas. Blue lines represent runs using low variances, pink lines represent runs using high variances, filled symbols represent yearly sampling and open symbols represent every-other-year sampling.

With nine sampling stations, an eight-year study is predicted to miss a halving of the population one out of every two times such a dramatic decline occurs (see Figure 1c). Over this shorter time span, BLM would need add 8 sampling stations to reach the target power for detecting 7% annual declines with yearly sampling, or 11 stations for sampling every other year (to 17 and 20 total samples respectively). Simulations achieved adequate power to detect 8% annual population increases with five fewer samples than required to detect decreases under each scenario. **Viewing lower and upper reaches separately:**

When the lower and upper reaches were simulated separately, eight and nine stations, respectively, are needed to detect a loss of half the population with acceptable power over a ten year span, sampling every year (Figures 2a and 3a). Adding three and two additional stations (totaling 11 stations each) would enable sampling to be done every other year and still achieve adequate power.

Shortening to an eight year span again dramatically reduces power (Figures 2c and 3c). Lower and upper reaches would need 14 and 15 stations for adequate power to detect 7% annual declines with yearly sampling, and 16 and 20 stations if sampling every other year. Again, all of these numbers are more than adequate for detecting target population increases (Figures 2d and 3d).

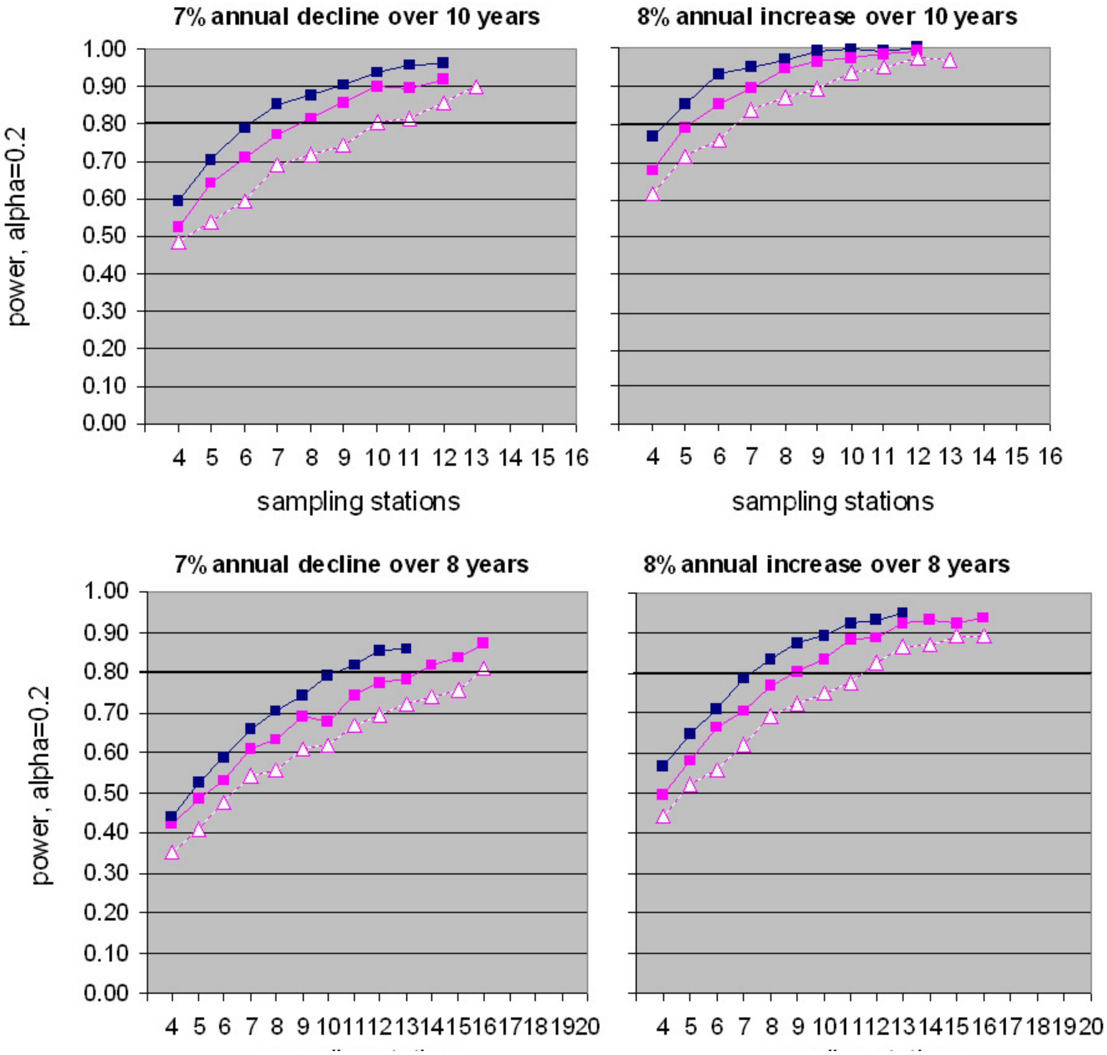


Figure 2. Lower reach, simulated power to detect specified annual changes in topminnow density over ten year (a-b) and eight year (c-d) time spans with increasing numbers of sampling stations. Over ten years, 7% annual decline and 8% annual increase translate to an overall halving or doubling of the population, respectively. Our threshold for adequate power is 0.8, or 80%. Blue lines represent runs using low variances, pink lines represent runs using high variances, filled symbols represent yearly sampling and open symbols represent every-other-year sampling.

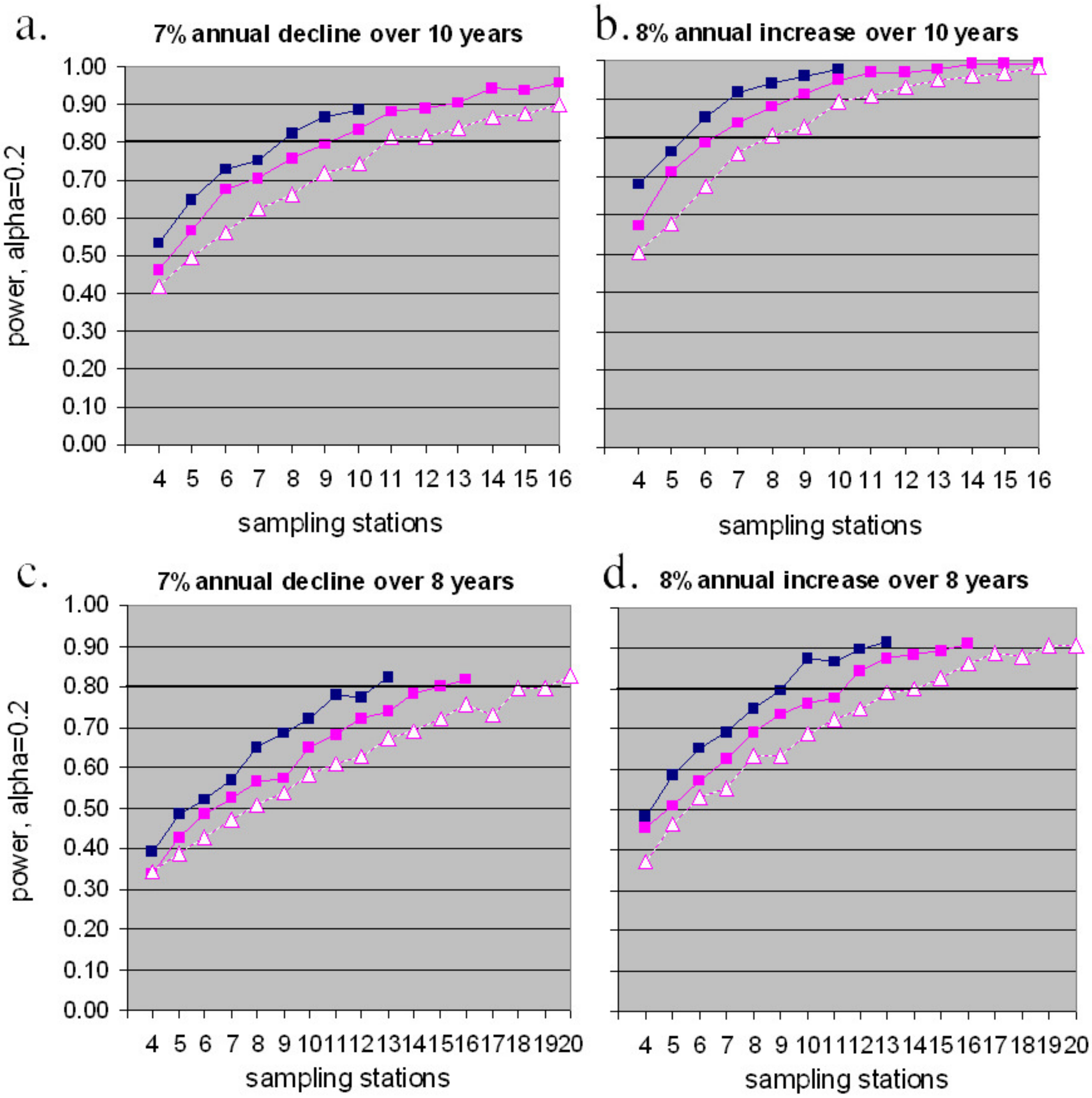


Figure 3. Upper creek, simulated power to detect specified annual changes in topminnow density over ten year (a-b) and eight year (c-d) spans with increasing numbers of sampling stations. Over ten years, 7% annual decline and 8% annual increase translate to an overall halving or doubling of the population, respectively. Our threshold for adequate power is 0.8, or 80%. Blue lines represent runs using low variances, pink lines represent runs using high variances, filled symbols represent yearly sampling and open symbols represent every-other-year sampling.

Discussion

The existing literature on survey design and analysis does not always distinguish clearly between approaches that look for multi-year trends and those that compare means from two time periods. Most data could be analyzed either way, but optimizing survey design for one type of data analysis may over- or under-shoot the amount of sampling needed to get sufficient power with the other. Examples in the AZGFD fish sampling manual use the sample-size equation for comparing two means (given above) to calculate sample sizes needed to detect a particular amount of change over a five or a ten year time period. These examples state that sampling will be done each year, but the analytical procedure used to estimate necessary sample size implies that you are actually only using two years of data to test for change (usually the first and last years). Neither the AZGFD manual nor Parkinson et al. discuss other ways to analyze data for trends when you do have multi-year data (AZGFD 2004 cites but does not discuss Urquhart et al.1998, which presents sophisticated tools for analyzing multi-year trend data). This is relevant in part because comparing two mean counts does not attempt to distinguish how much of the difference in mean count between two sets of measurements is part of an ongoing multi-year trend vs. how much is year-to-year fluctuations or other sources of variation. Using multiple years' data can make testing for change more powerful because it can factor out some of the year-to-year fluctuations that may mask trends. Even if one samples only at the end points of the time frame of interest (62 samples at year 0 and 62 at year 10), this approach still represents considerably more overall sampling effort than the option presented above of sampling 10 stations every other year for ten years (60 samples total) and using a regression of counts vs. time to determine trends. Unless there is an urgent need to have short-term data or to test specific year-to-year hypotheses, we suggest that monitoring effort is better spent tracking trends over multiple years. Sampling repeatedly over multiple years also provides more frequent screening for exotics, disease, and other problems, and enables managers to react to such problems more quickly. Nevertheless, the two-event-comparison analysis may provide useful guidance on sample sizes needed for research projects. Standards for acceptable error in research are generally higher than those used here, however, so sample size calculations will need to use levels of α and β deemed appropriate for particular projects.

Partitioning variance, choosing software, and estimating abundance for trend power analysis

The main dilemmas we had approaching this analysis were how to calculate variance estimates, whether to use analytical or simulation tests, and which years' data to use in generating inputs for these tests. Various practitioners of power analysis for trend studies rely on different theoretical ways of estimating variance (Gibbs 1995; Gerrodette 1993; Larsen et al 2001). The contrast between MONITOR and another prominent publicly available trend-detection power analysis software called TRENDS (Gerrodette 1993) is a good example. TRENDS is designed to use variance measures that reflect spatial variation within sampling periods (usually years) and MONITOR is designed to use variance measures that reflect short-term temporal variation within each site sampled. In reality, however, many practitioners converge on using RMSE from a regression or ANOVA because this measure incorporates all sources of variation not explained by one's model (i.e., all the variation not explained by year in a simple regression like we used above, or the variation left over once other effects are taken into consideration by more elaborate ANOVA or ANCOVA models) (Nickerson and Brunell 1997; Rice et al. 2001). We tried a variety of ways of calculating variance estimates from these fish counts, including de-meaned standard deviations (Gibbs 1995) and multi-factor ANOVAs with pool area and site as factors. Most of these estimates were quite close to one another, and virtually all fell within the 95% confidence limits around the variance estimates produced by simple linear regression against year. The exception was variance measures calculated from yearly averages; averaging of sites within each year eliminates the spatial component of overall variation and is therefore lower than other measures. In practice, MONITOR simulations showed fairly low sensitivity to the exact magnitude of variance values we used (Figures 1-3), even though upper and lower variance inputs differed by as much as 50%. This is reassuring, and suggests that the debate on precisely how variances ought to be estimated for trend power analyses is, in this case, more theoretical than practical. Note, however, that variance estimates from multi-site yearly averages do not consider spatial variation among sites and are therefore not appropriate for MONITOR simulations (they give unrealistically high power estimates).

Approaches by Urqhart, Larsen, and others (Urquhart et al. 1998; Larson et al. 2001) take a more explicit approach to partitioning variance and discuss how each of these sources of variance dictate how changing each aspect of monitoring design may or may not increase power. At Las Cienegas, each site's year-to-year change is largely independent of change shown at other sites (i.e., variation is not concordant among sites). In our lower reach data, among-year residual error is the largest source of variation, followed by year*site interaction variation, with site-specific and year variation explaining considerably less of the overall variation in fish counts. For the upper reach, residual error is still the largest source of variation, followed closely by year, then with interaction variation and site-specific variation explaining more of the overall variation in fish counts than in the lower reach. This low level of concordant variation allows us to better detect trends by visiting more sites, since the effect of concordant year variation cannot be reduced by adding sites of visits to a survey (Larsen et al. 2001). These authors have written code for the statistical program "R" that enables users to specify how sample variance is partitioned among all these sources. Because our data was taken once a year, however, we cannot separate pure within-year sampling error from year-to-year fluctuations that are independent of sites.

An obvious follow up step for this project would be to separate these sources of variation further by doing a round of repeat sampling (re-sampling, say, a week later to minimize effects of the first sampling on subsequent fish behavior without introducing seasonal variation), run a set of analyses with this more complex R code, and compare results with those generated by MONITOR. Repeat sampling and depletion sampling would also test the assumption that counts from a single-pass seine haul are a consistent index of true fish density in a given pool.

In choosing among power calculation software, TRENDS relies on analytical calculations while MONITOR is based on simulations. We have no a priori reason to prefer one approach over the other. In practice, we chose MONITOR because it allowed us to explicitly change more details of sampling design including multiple samples per year and irregular survey schedules. With our data, TRENDS' analytical results tended to be slightly more conservative (i.e., require a few more samples) than simulation results from MONITOR. The reasons for this are not clear, since they use such different approaches. TRENDS appears to be more sensitive to changes in CV than MONITOR is. It is important to note that TRENDS is designed to use variance measures that exclude spatial variation (e.g., variance among yearly averages). The appropriate comparison among outputs is therefore between TRENDS analyses that use variance calculated from yearly average counts, and MONITOR analyses that use variance calculated from counts of each site for each year.

We expected MONITOR simulations to be somewhat sensitive to changes in the abundance estimates used. For the main set of analyses presented above, we used mean and variance inputs calculated from all years of monitoring data. Since topminnow numbers have declined, however, we thought using data from only recent years might give a more realistic picture of power for future sampling years. We ran an extra set of analyses (10-year span, yearly sampling) for the lower and upper reaches using only data from 2002-2005. As with the main analyses, density values for each station were calculated as the mean of that station (POOC/50m$^{2)}$ from all years considered. Variance estimates were again RMSE's from regressions of year vs. counts for all stations. Instead of using high and low confidence intervals around these RMSE's, however, we used the RMSE's. Using fewer years in the regression gives the model fewer degrees of freedom, which in turn greatly broadens the confidence interval around the variance estimate. We used RMSE's directly because our main goal was to see whether the recent years' lower abundances would change our power simulation results, not to compare a narrow confidence interval to a broad one.

For the lower reach, simulations using recent years' data gave results that were very similar to those generated with all years' data (from Figure 2). The nine stations predicted to give adequate power by our main analyses (Table 2) still gave adequate power with recent years' inputs. For the upper reach, however, using only recent years' data was slightly more complicated. Stations FC0311 and FC0313 had no fish from 2002-2005, but MONITOR cannot run with zeros as input values. We ran one set of analyses using counts and variances from all four stations but replacing 0 mean counts with 0.5 before running them through MONITOR, and a separate set using data from occupied sites only (FC0312 and FC0316). The first approach required eleven stations to achieve adequate power, and the second required eight, compared with the nine required by our main analyses (Table 2). These upper reach results, however, are perhaps a bit misleading. Abundance inputs for additional sampling stations are modeled as the mean of all other sites, so neither upper reach approach included the near-zero counts that might be expected in a reach in which two of the four existing sites have had no fish for the past several years. Modeling half the additional stations has having 54 fish per 50m^2 (the mean of the two occupied sites from 2002-2005) and half as having mean of 0.5 fish per 50m^2 required 15 stations for adequate power. For every-other-year sampling over an eight-year span, more than 30 sampling stations would be necessary.

Appendix D: Raw Data

Sampling site	1989	1990	1990	1992	1993	1994	1995	1996	1997	1998	1999	2002	2003	2004	2005
	fall	spring	fall	fall	fall	fall	fall	fall	fall	fall	fall	fall	fall	fall	fall
	topminnow caught														
FC0309	-	-	-	-	0	-	2	-	1199	-	3	0	0	-	-
FC0310A	429	44	77	-	126	1965	1095	2325	239	571	193	21	54	8	2
FC0310B	1	-	-	-	3	312	45	698	122	211	32	292	17	58	10
FC0311	320	1667	92	1931	402	1853	2431	-	-	968	62	0	0	0	0
FC0312	736	0	95	1385	247	101	226	86	-	-	-	0	0	10	2
FC0313	61	75	-	1340	9	24	800	705	143	10	0	0	0	-	-
FC0314	-	70	-	18	0	45	74	506	168	174	585	325	2194	480	129
FC0315	-	109	-	3	0	364	12	116	-	60	131	148	622	252	155
FC0316	-	38	-	-	105	20	564	559	210	399	49	11	104	10	82
	Sampled area, m²														
FC0309	-	-	-	-	18.2	-	17.7	-	50.7	-	31.2	44.3	29.7	-	-
FC0310A	53.7	30.0	22.4	-	100.8	90.5	73.9	44.7	37.9	59.4	68.0	44.3	62.6	83.5	83.5
FC0310B	23.9	-	-	-	55.2	45.3	24.8	48.8	39.5	40.1	71.3	71.2	46.7	38.4	97.6
FC0311	14.0	13.6	16.8	17.7	26.8	27.5	151.7	-	-	39.0	39.6	41.7	41.7	44.9	47.5
FC0312	23.8	26.8	20.1	49.2	63.7	15.8	27.9	12.0	-	-	-	30.4	36.0	16.3	45.1
FC0313	28.8	33.5	-	45.5	89.0	16.0	107.3	36.2	50.1	48.1	84.5	161.4	160.3	-	-
FC0314	-	20.5	-	55.0	48.4	27.4	38.5	40.1	66.0	59.9	45.5	131.3	103.8	87.0	72.2
FC0315	-	41.2	-	66.1	50.3	39.1	148.3	51.1	-	68.3	27.8	41.0	38.0	33.7	49.6
FC0316	-	12.5	-	-	46.5	20.0	37.9	44.2	37.9	40.1	23.7	36.4	28.3	28.3	23.4