Comparing monitoring data collected by volunteers and professionals shows that citizen scientists can detect long-term change on coral reefs

Graham E. Forrester  
*University of Rhode Island, gforrester@uri.edu*

Patricia Baily

Dennis Conetta  
*University of Rhode Island*

Linda Forrester  
*University of Rhode Island*

Elizabeth Kintzing

The University of Rhode Island Faculty have made this article openly available. Please let us know how Open Access to this research benefits you.

This is a pre-publication author manuscript of the final, published article.

Citation/Publisher Attribution

Available at: https://doi.org/10.1016/j.jnc.2015.01.002

This Article is brought to you for free and open access by the Natural Resources Science at DigitalCommons@URI. It has been accepted for inclusion in Natural Resources Science Faculty Publications by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons@etal.uri.edu.
Comparing monitoring data collected by volunteers and professionals shows that citizen scientists can detect long-term change on coral reefs

Graham Forrester\textsuperscript{a}, Patricia Baily\textsuperscript{b}, Dennis Conetta\textsuperscript{c}, Linda Forrester\textsuperscript{c}, Elizabeth Kintzing\textsuperscript{d}, Lianna Jarecki\textsuperscript{e}

\textsuperscript{a}Department of Natural Resources Science, University of Rhode Island, Kingston, RI 02881, USA

\textsuperscript{b}Ecosail, Ltd, Road Town, Tortola, British Virgin Islands

\textsuperscript{c}Department of Biological Sciences, University of Rhode Island, Kingston, RI 02881, USA

\textsuperscript{d}School of Marine Science and Ocean Engineering, Jere A. Chase Ocean Engineering Laboratory, University of New Hampshire, Durham, NH 03824 USA

\textsuperscript{e}H. Lavity Stoutt Community College, Paraquita Bay Campus, PO Box 3097, Road Town, Tortola, British Virgin Islands

Corresponding author: Graham Forrester, gforrester@uri.edu, 1-401-874-7054
Abstract

Citizen science is increasing and can complement the work of professional scientists, but the value of citizen data is often untested. We therefore compared the long-term changes to coral reefs that were detected by a professional and volunteer monitoring program, operated by University of Rhode Island (URI) staff and Reef Check volunteers respectively. Both groups monitored reefs in the British Virgin Islands from 1997-2012 but mostly monitored different sites (URI 8 sites and Reef Check 4 sites). When URI staff visited the Reef Check sites to perform a side-by-side to comparison, Reef Check fish density estimates were consistently higher than those made by URI observers but benthic indicators showed better agreement. When long-term trends were compared, the two programs detected qualitatively similar trends in the % cover of live coral and coral rubble, but temporal changes in the cover of other benthic indicators were less consistent. The URI program detected a widespread increase in parrotfish densities and a decline in snappers, whereas the Reef Check surveys detected no consistent changes in any fish density indicators. Overall, site-specific temporal trends revealed by the URI program were more often statistically significant than those from Reef Check (twice as often for benthic taxa, and five times as often for fish taxa), which implies greater precision of the scientists’ counts. Nonetheless, volunteers were able to detect important changes in benthic communities and so have a valuable role to play in assessing change on coral reefs.

Key Words: Coral reef, decline, long-term, monitoring, volunteers
Introduction

Citizen science (defined by Kruger & Shannon, 2000) has been increasing, and has the potential to complement work done by professionals (e.g., Carr, 2004; Cohn, 2008; Conrad & Hilchey, 2011; Silvertown, 2009). Non-specialist volunteers participate in many conservation-orientated projects worldwide by monitoring species and environmental conditions in various habitats. Because financial, manpower and training resources for conservation monitoring are limited, involving volunteers can compensate for these constraints and greatly increase the overall amount of information available. Professional scientists have, however, sometimes questioned the accuracy and precision of volunteer data (Boudreau & Yan, 2004; Brandon, Spyreas, Molano-Flores, Carroll, & Ellis, 2003; Crall, et al., 2010; Fore, Paulsen, & O'Laughlin, 2001; Hunter, Alabri, & van Ingen, 2013; Legg & Nagy, 2006; Nerbonne & Vondracek, 2003; Underwood & Chapman, 2002).

Arguably, citizen science monitoring is having the greatest influence on ecology by broadening the geographic scope of monitoring (Dickinson, Zuckerberg, & Bonter, 2010). These contributions are illustrated by recent marine examples, in which recreational divers helped to better define the geographic distributions of species of special significance in the Mediterranean (an endemic coral, Corallium rubrum, and sea horses, Hippocampus spp.) (Bramanti, Vielmini, Rossi, Stolfa, & Santangelo, 2011; Goffredo, Piccinetti, & Zaccanti, 2004), and helped define the spatial extent of global coral bleaching events (Hodgson, 1999; Marshall, Kleine, & Dean, 2012). In fewer cases, citizen scientists have performed long-term monitoring to reveal longitudinal trends in population status (Carr, 2004; Conrad & Hilchey, 2011). Perhaps the best examples come from ornithology, beginning with the Audubon Society’s annual Christmas bird counts, which started in 1900 and now engage 60-80,000 volunteers annually. Other major public bird monitoring surveys in the United States have developed since, including the U.S. Geological Survey Breeding Bird Survey, launched in 1966, and the Cornell Lab of Ornithology’s nest record card scheme, begun in 1965 (Dickinson, et al., 2010). The goal of biological monitoring is often to detect change over time (Boylan, Howe, Bartkowski, & Eichler, 2004), but the inability to detect ecologically significant changes is a drawback of many programs (Legg & Nagy, 2006). Several studies have compared volunteer versus professional monitoring to assess their performance (e.g., Finn, Udy, Baltais, Price, & Coles, 2010; Gillett, et al., 2012; Gollan, de Bruyn, Reid, & Wilkie, 2012; Kremen, Ullmann, & Thorp, 2011; Lovell, Hamer, Slotow, & Herbert, 2009), but few studies have compared the ability of professional and volunteer monitoring programs to detect long-term trends (for exceptions see Kallimanis, et al., 2012; Robbins, Sauer, Greenberg, & Drogege, 1989; Royle, 2004).

Collecting reliable data on coral reefs is challenging because the technical demands of working underwater on SCUBA add to the usual difficulties of data collection in the field (Gillett, et al., 2012). In addition, because coral reefs support complex species-rich communities, accurate data collection usually also requires learning to identify many species or functional groups that are used as indicators of reef status. Perhaps for this reason, previous comparisons between professionals and volunteers on coral reefs have usually focused on a single taxonomic group, such as fish (Darwall & Dulvy, 1996; Holt, Rioja-Nieto, Aaron MacNeil, Lupton, & Rahbek, 2013; Pattengill-Semmens & Semmens, 2003), sharks (Ward-Paige & Lotze, 2011), corals (Marshall, et al., 2012), or sponges (Bell, 2007) (for an exception see Mumby, Harborne, Raines, & Ridley, 1995).
Volunteer coral reef monitoring programs have been used to indicate the spatial extent of widespread impacts, such as coral bleaching events (Hodgson, 1999; Marshall, et al., 2012). Temporal change has, instead, usually been examined using professional monitoring studies (e.g. De'ath, Fabricius, Sweatman, & Puotinen, 2012; Gardner, Cote, Gill, Grant, & Watkinson, 2003; Paddock, et al., 2009), although volunteer data are included in some recent analyses (Schutte, Selig, & Bruno, 2010). These analyses of long-term change on coral reefs are often based on many short-term studies that are “stitched together” to create a region-wide picture of long-term changes, and there are relatively few extended time-series from individual sites (Bak, Nieuwland, & Meesters, 2005; Hughes, 1994). For this reason, testing whether volunteer programs can detect long-term change is of interest. Our objective was thus to compare the ability of professional and volunteer monitoring to detect long-term temporal changes in coral reef communities.

**Methods**

**Study design**

We compared two monitoring programs in the British Virgin Islands (BVI). The first is part of popular global volunteer organization (Reef Check), whose volunteers collect a comprehensive array of measurements on fish, invertebrates and structural reef properties (Hodgson, 1999, 2001). One goal of Reef Check is “to create a global network of volunteer teams trained in Reef Check's scientific methods who regularly monitor and report on reef health”. The BVI Reef Check group monitored 4 coral reef sites from 1997-2012 (Table 1, Fig. S1). We compared their results to the results of a professional monitoring program that was also based in the BVI. The professional program (hereafter referred to as the URI program) monitored 8 different BVI sites from 1992-2012 (Table 1, Fig. S1). The URI program was led by the first author and conducted by scientists with specialist training in coral reef ecology.

The two programs were designed and run independently, so our analysis involved only subsets of data from each program that were comparable. We compared sites similar in wave exposure (based on fetch distances), habitat (fringing reef slopes), and depth (10 m), and we limited the analysis to the period when the two programs ran concurrently (1997-2012). The 8 URI sites were each monitored annually from 1997-2012, but there were some early gaps in the Reef Check sampling as follows: Bronco Billy lacked data for 1997 and 2001, Diamond Reef lacked 2000, Pelican Island lacked 1998 and 2000, and Spyglass lacked 1997 and 2000.

Our primary goal was to assess whether the two programs could detect widespread temporal changes in reef communities, rather than localized changes that might occur at only one site but not others. We defined widespread changes as those detectable at the majority of sites (50% or more), so our analysis was based on the premise that if one program detected a widespread trend then it was reasonable to expect the other program to reveal a similar trend. Because the two programs monitored different sites (Table 1), our ability to compare counts made at the same time and place is limited. There were, however, seven occasions when members of the URI group visited Reef Check sites (Diamond Reef in 1999, 2001, 2004 and 2006; Pelican Island in 2004; Spyglass in 2004, and Bronco Billy in 2004). These side-by-side measurements permitted direct comparison of URI and Reef Check counts.


**Monitoring methods**

Reef check methods have described by others previously (Hill & Wilkinson, 2004; Hodgson, 1999), so are outlined briefly here. At each site, two markers were placed permanently on the reef to define a fixed 100 m long transect. A tape was laid on the reef between the two markers, and divers swam along the transect to complete a fish survey, an invertebrate survey and a benthic survey. Fish were counted first using a belt transect, during which the diver swam slowly along the tape and counted fish within an area 100m long x 5 m wide (400 m²). After the fish count, invertebrates were counted within the same area 400 m² area. The benthic survey used the linear point-intercept method (Ohlhorst, Liddell, Taylor, & Taylor, 1988). A diver then swam along the tape and, at 0.5 m intervals, the substratum under the tape was assigned to one of ten categories (Table 2). The percent cover of each substratum category was estimated as the percentage of points under which that substratum was observed. A core group of 7 regular volunteers organized and participated in most surveys, joined by a larger number of less frequent participants.

For the URI program, each site was roughly 0.6 ha in area and was surveyed annually between June and August. To estimate the density of larger fishes (visually estimated to be > 30 cm total body length), the entire site was first scanned by a single observer (the first author) as he swam slowly back and forth through the site. At the beginning of the study, and at intervals throughout, the accuracy of these body length estimates was assessed by estimating the length of pieces of PVC or plexiglass (Polunin & Roberts, 1993). During the first 5 years of the study period, the accuracy of size estimates was also tested by estimating fish body lengths visually in the field, then capturing the fish to determine their actual length (18 species, 8-34 individuals per species).

Smaller fishes (visually estimated to be < 30 cm total body length) and benthic organisms were counted using 30 m long transects, placed at haphazard locations within each site (n = 4-12 transects per site per year). First, small fishes were counted within belt transect 30 m long x 1.5 m wide. These counts were all made by a single observer (the first author), who used a T-shaped bar to determine the transect width. The benthic surveys employed the linear point-intercept method, in which a diver swam along the tape and identified the material under the tape at 0.25 m intervals (n = 120 points per transect). Scleractinian corals were identified to species, whereas other organisms were assigned to coarser groupings based on taxonomic and morphological affinity. The benthic surveys were performed by 4 observers and, to reduce observer bias, when a new observer joined the program the divers counted the same transects (n > 14 shared transects per diver) and compared counts before their data were included in the database.

**Indicators compared between programs**

Although survey methods differed between programs, there were 12 indicators that were comparable across programs and could be expressed in the same units (Table 2). The URI program collected data at a finer level of taxonomic resolution, so those data were pooled to create measurement equivalent to Reef Check indicators. For one indicator (Sea fans), the two programs collected data in different units: Reef Check observers counted the number of colonies in a 400 m² belt transect, whereas URI observers estimated % cover along 30 m line transects, so
comparison between programs was thus limited to a qualitative comparison of trends over time (Table 2).

Data analysis

When both groups monitored the same site in the same year, these were considered paired samples and we used a paired t-test to test for a difference between the programs in absolute counts.

For the long-term monitoring when the programs monitored separate sites, we used two approaches to test whether the two programs detected similar long-term changes. First, we used a simple weight-of-evidence approach. For each site, we performed a non-parametric test for a monotonic change over time (Mann-Kendall test, two-tailed test of the null hypothesis of no trend), which has the benefit of simplicity and of not assuming any specific pattern of change over time (Kendall, 1975). Each site was summarized based on whether a trend was detected (p<0.05) and, if so, the direction of the trend (increase versus decrease). A program was considered to have revealed a widespread change if significant trends were detected at least 50% of the sites, and if all trends were in the same direction (i.e. all increases or all decreases). The two programs were deemed equivalent if both revealed widespread change in the same direction.

Secondly, we used repeated measures mixed effects linear models to construct formal tests of the null hypothesis that the trends detected by the two programs were equivalent (using SPSS v.21 and SAS v. 9.2). Sites were specified as a repeated term to account for correlated observations over time at each site. For each indicator, we checked for normality and transformed the data if necessary to meet this assumption. We also examined different temporal covariance structures and the most appropriate was either the first-order autoregressive structure with heterogeneous variances or the heterogeneous compound symmetry covariance structure. These covariance structures specify that the correlation in the response from year to year is either consistent or declines exponentially with time, and that variance changes from year to year.

We modeled the effect of program (URI versus Reef Check) as a fixed effect, so the full model contained terms for the effect of program, year and the program x year interaction. Our focus was the test for a program by year interaction because a significant result indicates that the trend over time (slope) differs between programs. In the absence of a significant program by year interaction, a significant year effect indicates a change over time common to both programs. A significant program effect is of less interest because it indicates only that absolute counts differed between programs. For sea fans, which were measured in different units across programs (colony density versus % cover) we modeled each program separately. For each program, sites were still specified as a repeated term but the model otherwise contained a term for year only. For sea fans, we thus focused on whether the two programs revealed qualitatively similar trends over time (based on the sign and significance of the parameter for the year effect).
Results

Comparing counts made at the same site and time

The two groups recorded similar absolute percent cover estimates for hard coral and rubble (Fig. 1). The most obvious and striking difference between programs was that no macroalgae were recorded at the 4 Reef Check sites (Fig. 1). For three other benthic indicators there were less striking, but still statistically significant, differences in absolute percent cover estimates. URI observers recorded higher cover of soft corals and sponges than their Reef Check counterparts, but lower cover of sand (Fig. 1). For every fish indicator, Reef Check observers reported higher densities than the URI observer, and this trend was statistically significant for each group except butterfly fish (Fig. 2).

Comparing long-term trends detected by the two programs

**Hard Coral.** Significant declining trends in coral cover were identified at 6 of 8 URI sites and 2 of 4 Reef Check sites (Table 4), indicating that both programs detected a widespread decline in coral cover (Fig. 3). This conclusion is supported by the mixed linear model, which indicated a significant overall decline over time (a year effect), but no difference in the rate of decline between programs (no program x year interaction), nor a difference between programs (no program effect) (Table 5).

**Soft coral.** Soft corals changed in % cover in a way that precluded simple comparison of the two programs (Fig. 4). Soft corals showing a significant increasing trend at 5 of 7 URI sites, whereas at the 8th site (White Bay) there was a significant decline in soft coral cover (Table 4). An increasing trend in soft coral cover was detected at one Reef Check site, but no significant trends were detected at the other 3 sites (Table 4). Because the URI data revealed both positive and negative trends, our initial premise for analysis, that widespread impacts would be reflected by congruent changes across most sites, is thus not met for this indicator. It is also probably for this reason that the mixed linear model indicated no consistent trend over time (no year effect), nor a difference in trend between programs (no program x year interaction) (Table 5).

**Sponges.** Neither the Reef Check volunteers nor the URI observers recorded a widespread temporal change in sponge abundance (Table 4; Fig. S2). Changes over time were significant at 2 of the 8 URI sites, but at none of the 4 Reef Check sites (Table 4). This conclusion is also supported by the mixed linear model, which indicated no consistent trend over time (no year effect), nor a difference in trend between programs (no program x year interaction) (Table X).

**Macroalgae.** No macroalage were ever recorded by the Reef Check volunteers so we could make no further comparison between programs. Mean macroalgal cover was above 20% at all URI sites (Fig. S3), but neither the Mann-Kendall tests nor the mixed linear model detected a widespread pattern of increase or decrease over time at the URI sites (Table 4 and 5).

**Rubble.** The Mann-Kendall tests revealed a significant temporal change in rubble cover at only one of the 12 sites (Table 4). Nonetheless, the sign of the test statistic (S) was positive at 11 of 12 sites, suggesting a weak tendency towards increasing in rubble cover throughout the BVI (Fig. S4). The results of the mixed linear model were also consistent with a widespread increase that was observed by both programs; there was a significant overall increase in rubble cover over
time (a year effect), but no evidence that the rate of increase differed between programs (no program x year interaction) (Table 5).

**Sand.** The Mann-Kendall tests suggests a difference between programs in detection of a widespread trend, because sand cover increased significantly at 4 of the 8 URI sites but no significant changes were detected at the Reef Check sites. The mixed linear model, though, does not support this conclusion because there was no significant program by year interaction (Table 5). Instead, a significant year effect in the model suggests an increase in sand cover across all sites (Table 5). The differing outcome of our two analyses precludes a strong overall conclusion, but a parsimonious interpretation is that there may be a weak tendency for increasing sand cover at all sites, and that the increase is slightly more pronounced at the URI sites (Fig. S5).

**Sea fans.** Like soft corals, site-specific differences in the way sea fans changed over time precluded simple comparison of the two programs (Fig. S6). Sea fans increased in cover over time at 3 of 7 URI sites, whereas at the 8th site (White Bay) sea fan cover declined over time (Table 4). Probably because there was a mix of positive and negative trends over time, the mixed linear model for the URI sites detected no consistent change over time (no year effect). At the Reef Check sites an increasing trend in sea fan colony density was observed at 2 of 4 sites, and the mixed linear mode indicated a significant increase over time (a year effect) (Table 5).

**Butterfly fish and Grunts.** Although their densities fluctuated from year to year, butterfly fish showed a detectable decline in abundance at only 2 sites, and an increase at just 1 site (Table 4; Fig. S7). The lack of a consistent widespread change over time was reflected in the results of the mixed linear model (no year effect or year x program interaction). Grunt densities displayed a similar lack of widespread change in density over time (Fig. S8), with a significant temporal trend detected at only one site (Tables 4 and 5).

**Parrotfish.** Parrotfish were one of two fish groups for which the two programs detected differing widespread trends (Fig. 5). The Mann-Kendall tests revealed a widespread increase in parrotfish density at the URI sites, because there were significant increases at 4 of 8 sites. In contrast, a significant change was detected at only one Reef Check site (Table 4). Further support for the conclusion that the two programs detected differing trends comes from a significant program by year interaction in the mixed linear model (Table 5).

**Snappers.** The two monitoring programs also revealed differing patterns of temporal change in snapper densities. Significant declines were observed at 4 of 8 URI sites, but no changes were detected at Reef Check sites (Table 4; Fig. 6).

**Groupers and Nassau groupers.** Large groupers of all species were counted rarely by both programs (Fig. 2), and no clear temporal trends in their densities were identified (Table 4; Fig. S9).

**Discussion**

**Direct comparisons between scientists and volunteers: benthic indicators**

For scientific purposes, the value of citizen monitoring depends on the reliability of data collected. Volunteer data are most commonly assessed using one-time comparisons in which
volunteers and professionals monitor the same site(s). These comparisons are based on the assumption that professional data provides a benchmark against which to judge volunteer performance (e.g. Brandon, et al., 2003; Finn, et al., 2010; Fore, et al., 2001; Foster-Smith & Evans, 2003). Most studies compare groups of volunteers against a single scientist, but comparisons that also account for differences between scientists are more informative (Cox, Philippoff, Baumgartner, & Smith, 2012). Not surprisingly, the reported level of agreement between scientists and volunteers varies greatly, but past comparisons on coral reefs have shown that volunteers can quantify simple, but ecologically relevant, benthic indicators of reef status such as % cover of major taxa. For example, after an intensive 8-day training program in marine life identification and survey techniques, volunteers in Belize could quantify benthic groups such as corals and macroalgae quite reliably (Mumby, et al., 1995). Counts performed by different groups of volunteers were strongly correlated (indicating volunteer consistency), and volunteer counts were strongly correlated with scientist counts (indicating volunteer accuracy). Similarly, Bell presented evidence that, after classroom and in-water training, volunteer surveys of sponges using a morphological classification were congruent with scientist surveys (Bell, 2007). Reliable volunteer estimates thus depended on suitable training and cross-calibration with counts made by supervising scientists.

For benthic indicators, when URI staff and Reef Check volunteers surveyed the same sites the degree of cross-program agreement varied widely among indicators. The URI and BVI Reef Check programs were run independently, and the lack of inter-program communication about monitoring results and comparison among observers can explain some differences. For example, the striking difference between programs in estimates of macroalgal % cover clearly arose from a different interpretation of what organisms should be counted as macroalgae. Although Reef Check defines macroalgae as “all macro-algae except coralline, calcareous and turf”, the indicator is named “Nutrient Indicator Algae (NIA)”. Correspondence with the Reef Check central program about which algae were nutrient indicators, led the BVI volunteers to exclude all algae they encountered. URI observers, instead, simply counted all macro-algae they observed (most common taxa recorded were Dictyota spp. Lobophora spp. and Sargassum spp.). Other differences, such as the systematically lower soft coral % cover estimates by Reef Check volunteers are harder to explain. On the other hand, the fact that absolute % cover estimates for two important benthic indicators (hard corals and rubble) were statistically indistinguishable between programs is encouraging because it suggests that, with clear definition of what is to be counted, some benthic indicators may be comparable across programs with minimal inter-calibration.

Direct comparisons between scientists and volunteers: fish densities

Compared to surveys of benthic organisms, visual surveys of fish are subject to several additional sources of counting error because fish are mobile and react behaviourally to diver presence. Analyses show that, even for professional scientists, density estimates are heavily influenced by transect dimension (Kulbicki, et al., 2010; Sale & Sharp, 1983), observer swimming speed (Lincoln Smith, 1988), fish abundance (Bozec, Kulbicki, Laloe, Mou-Tham, & Gascuel, 2011; Kulbicki, 1998; MacNeil, et al., 2008), habitat (Edgar & Barrett, 1999), avoidance of observers by fish (Fowler, 1987; Kulbicki, 1998), diel or tidal changes in fish behaviour (Thompson & Mapstone, 2002), and water clarity (Bozec, et al., 2011). The fish indicators developed by Reef Check require observers to both identify species and visually
assess body size because, for some indicators, only fish larger than a minimum size-threshold are counted. When performing multi-species surveys of either fish or invertebrates, less experienced volunteers sometimes misidentified or overlooked rare species, which reduced their counts relative to scientists (Bernard, Gotz, Kerwath, & Wilke, 2013; Cox, et al., 2012). This effect was not apparent during our study because Reef Check volunteers recorded systematically higher fish densities than the URI observer. A more likely explanation for higher Reef Check fish counts is over-estimation of fish body sizes, so that more fish are judged to be above the size-threshold for counting. Past studies have shown that accurately judging distances and fish sizes underwater requires regular practice, which is not feasible for most volunteers (Thompson & Mapstone, 1997; Williams, Walsh, Tissot, & Hallacher, 2006).

Detection of long-term change

Our primary goal was to assess whether two monitoring programs, run independently but operating in the same area, would detect similar patterns of long-term change on coral reefs. We thus focused on widespread changes that were observed at the majority of sites, rather than localized changes that might occur at only one site but not others. Our comparison of the long-term trends detected by the two programs would be more scientifically rigorous if both programs had monitored the same sites, so that we could definitively separate program effects from differences between sites. Offsetting this drawback is the fact that long-term professional and volunteer programs have rarely been performed in close proximity. Our analysis is, thus, one of few to show that volunteer monitoring can document long-term ecological change (Boylen, et al., 2004). Most importantly, our analysis shows that the two programs often revealed similar qualitative trends (i.e. decline versus increase over time), even though absolute counts sometimes differed between programs.

Reef Check monitors consisted of a core group of experienced local divers, plus other less frequent participants, suggesting that the level of knowledge and training differed greatly among participants. The four URI benthic observers, in contrast, had a more uniform background. This, coupled with the inter-calibration among URI observers, may explain why temporal trends revealed by the URI monitoring were more likely to be statistically significant, which presumably reflects a better ratio of signal-to-noise. Nonetheless, both programs identified a clear widespread decrease in live coral cover, and weaker but apparently widespread increases in the cover of rubble and sand. The percent cover of live coral is arguably the most ecologically significant indicator of coral reef status (Gardner, et al., 2003), so it is extremely encouraging that a self-organized group of volunteers was able to document long-term change in the status of their coral reefs.

Comparisons between scientists show considerable observer-related bias and imprecision when visually censusing fish, (Thompson & Mapstone, 1997; Williams, et al., 2006), and the same is true for comparisons among volunteers, and for cross-calibrations between scientists and volunteers (Darwall & Dulvy, 1996; Halusky, Seaman, & Strawbridge, 1994). Although observer effects and other sources of counting error can be ameliorated by training and inter-calibration, we argue that achieving consistency in visual fish counts takes more training and practice than benthic surveys. We suggest this is the most likely reason why Reef Check monitoring revealed no statistically significant long-term trends in fish density, whereas the URI
monitoring program detected both increases (parrot fish) and declines (snappers) in fish abundance.

**Conclusions**

Citizen science data is generally under-represented in scientific journal articles (Ely, 2008), and this is true for meta-analyses documenting long-term change on coral reefs (e.g. Alvarez-Filip, Dulvy, Gill, Cote, & Watkinson, 2009; Ateweberhan, McClanahan, Graham, & Sheppard, 2011; John F. Bruno & Selig, 2007; J. F. Bruno, Sweatman, Precht, Selig, & Schutte, 2009; Schutte, et al., 2010). There is considerable concern about the progressive degradation of coral reefs, and long-term monitoring can reveal relationships between taxa and functional groups, and how those relationships differ across locations. These patterns can be used to generate scientific hypotheses about drivers of change (Bellwood, Hughes, Folke, & Nystrom, 2004; Hughes, et al., 2003; Pandolfi, et al., 2003). Our analysis suggests that volunteer monitoring data can detect changes over time in simple indicators of benthic community status, and supports its expanded inclusion in these scientific analyses. For conservation and management, our analysis also supports the utility of volunteer coral reef monitoring for “surveillance” purposes, where the objective is to provide early detection by volunteers of changes of environmental concern (Conrad & Hilchey, 2011). Similarly, volunteers could be used more often to assess the effectiveness of specific management interventions, such as marine protected areas, where the objective is to assess how protected sites diverge over time relative to unprotected sites (Lester, et al., 2009; Micheli, Halpern, Botsford, & Warner, 2004).

**Acknowledgements**

Financial support from the Falconwood Foundation and the J. A. Woollam Foundation is gratefully acknowledged. Thanks to Rachel Finley for help with the URI monitoring and to the many volunteers who took part in with BVI Reef Check. We also thank the Guana Island staff, Dive BVI, and UBS Dive Center for logistical support.

**References**


Table 1. Sites monitored by the two programs (plus abbreviated site names) and their geographic coordinates.

<table>
<thead>
<tr>
<th>Sites</th>
<th>Coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reef Check Program</td>
<td></td>
</tr>
<tr>
<td>Bronco Billy (BB)</td>
<td>18.29.36N 64.27.37W</td>
</tr>
<tr>
<td>Diamond Reef (DR)</td>
<td>18.27.55N 64.31.50W</td>
</tr>
<tr>
<td>Pelican Island (PI)</td>
<td>18.19.51N 64.37.38W</td>
</tr>
<tr>
<td>Spyglass (Spy)</td>
<td>18.19.27N 64.36.9W</td>
</tr>
<tr>
<td>URI program</td>
<td></td>
</tr>
<tr>
<td>Bigelow Beach (Big)</td>
<td>18.28.09N 64.33.44W</td>
</tr>
<tr>
<td>Crab Cove (Cra)</td>
<td>18.28.51N 64.34.45W</td>
</tr>
<tr>
<td>Grand Ghut (Gra)</td>
<td>18.28.48N 64.33.43W</td>
</tr>
<tr>
<td>Iguana Head (Igu)</td>
<td>18.28.23N 64.34.54W</td>
</tr>
<tr>
<td>Monkey Point (Mon)</td>
<td>18.27.46N 64.34.14W</td>
</tr>
<tr>
<td>Muskmelon Bay (Mus)</td>
<td>18.29.07N 64.34.53W</td>
</tr>
<tr>
<td>Pelican Ghut (Pel)</td>
<td>18.28.33N 64.33.27W</td>
</tr>
<tr>
<td>White Bay (Whi)</td>
<td>18.28.11N 64.34.28W</td>
</tr>
</tbody>
</table>
Table 2. Indicators compared between programs and how they were measured.

<table>
<thead>
<tr>
<th>Variable (Reef Check acronym) and description</th>
<th>URI Program</th>
<th>Reef Check Program</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Method</td>
<td>Unit</td>
</tr>
<tr>
<td></td>
<td>Unit</td>
<td>Method</td>
</tr>
<tr>
<td><strong>Hard Coral (HC):</strong> All living stony coral (Scleractinia) and fire coral (Millepora spp.)</td>
<td>point intercept transect (30 m)</td>
<td>% cover</td>
</tr>
<tr>
<td><strong>Soft coral (SC):</strong> Includes zoanthids (Zoantharia) but not anemones (Actinaria)</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td><strong>Macroalgae (NIA):</strong> All macro-algae taller than 3 cm, except coralline and calcareous species and Halimeda spp.</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td><strong>Sponges (SP):</strong> All erect and encrusting sponges (Porifera)</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td><strong>Rubble (RB):</strong> Reef rocks between 0.5 and 15 cm in diameter</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td><strong>Sand (SD):</strong> Sediment less than 0.5 cm in diameter</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td><strong>Sea fans</strong> (Gorgonacea)</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td><strong>Butterflyfish</strong> (Chaetodontidae)</td>
<td>belt transect (45 m²)</td>
<td>fish/m²</td>
</tr>
<tr>
<td><strong>Grunts</strong> (Haemulidae)</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td><strong>Parrotfish</strong> (Scaridae) &gt; 20 cm in body length</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td><strong>Snapper</strong> (Lutjanidae) &gt; 30 cm in body length</td>
<td>Scan survey (4000 m²)</td>
<td>fish/m²</td>
</tr>
<tr>
<td><strong>Nassau grouper</strong> (<em>Epinephelus striatus</em>) &gt; 30 cm in body length</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
</tbody>
</table>
**Grouper** (Serranidae) > 30 cm in body length

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
Table 3. Results of tests for a temporal trend in benthic indicators at each site. The arrow signifies a significant increasing trend (↗), a significant decreasing trend (↘), or no detectable trend (→), based on the Mann-Kendall test (see Methods for details). Full site names are given in Table 1.

<table>
<thead>
<tr>
<th>Benthic indicators</th>
<th>URI sites</th>
<th>Reef Check sites</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Big</td>
<td>Cra</td>
</tr>
<tr>
<td>Hard coral</td>
<td>→</td>
<td>→</td>
</tr>
<tr>
<td>Soft coral</td>
<td>→</td>
<td>→</td>
</tr>
<tr>
<td>Sea fans</td>
<td>↘</td>
<td>→</td>
</tr>
<tr>
<td>Sand</td>
<td>→</td>
<td>→</td>
</tr>
<tr>
<td>Sponges</td>
<td>↓</td>
<td>→</td>
</tr>
<tr>
<td>Rubble</td>
<td>→</td>
<td>→</td>
</tr>
<tr>
<td>Macroalgae</td>
<td>→</td>
<td>→</td>
</tr>
</tbody>
</table>
Table 4. Results of tests for a temporal trend in fish densities at each site. The arrow signifies a significant increasing trend (↑), a significant decreasing trend (↓), or no detectable trend (→), based on the Mann-Kendall test (see Methods for details).

<table>
<thead>
<tr>
<th>Fish groups</th>
<th>URI sites</th>
<th>Reef Check sites</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Big</td>
<td>Cra</td>
</tr>
<tr>
<td>Butterfly Fish</td>
<td>↑</td>
<td>→</td>
</tr>
<tr>
<td>Grunts</td>
<td>$\downarrow$</td>
<td>→</td>
</tr>
<tr>
<td>Parrotfish</td>
<td>↑</td>
<td>→</td>
</tr>
<tr>
<td>Snappers</td>
<td>→</td>
<td>→</td>
</tr>
<tr>
<td>Groupers</td>
<td>→</td>
<td>→</td>
</tr>
<tr>
<td>Nassau Grouper</td>
<td>→</td>
<td>→</td>
</tr>
</tbody>
</table>
Table 5. Results of repeated measures mixed effects linear models to test the null hypothesis that the measurements collected by the two programs were equivalent (see section 2.4 for details). Presented are p-values for each term in the model, and significant terms are in bold. Models contained terms for the effect of program (URI versus Reef Check), year (change over time) and the program x year interaction. A program by year interaction indicates that the trend over time (slope) differs between programs.

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Program</th>
<th>Model term</th>
<th>Program x year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hard Coral (HC)</td>
<td>0.875</td>
<td><strong>0.002</strong></td>
<td>0.869</td>
</tr>
<tr>
<td>Soft corals (SC)</td>
<td>0.213</td>
<td>0.237</td>
<td>0.236</td>
</tr>
<tr>
<td>Macroalgae (NIA)</td>
<td>-(a)</td>
<td>0.123**(a)</td>
<td>-(a)</td>
</tr>
<tr>
<td>Sponges (SP)</td>
<td>0.139</td>
<td>0.126</td>
<td>0.126</td>
</tr>
<tr>
<td>Rubble (RB)</td>
<td>0.398</td>
<td><strong>0.028</strong></td>
<td>0.396</td>
</tr>
<tr>
<td>Sand (SD)</td>
<td><strong>0.001</strong></td>
<td><strong>0.003</strong></td>
<td>0.854</td>
</tr>
<tr>
<td>Sea fans</td>
<td>-(b)</td>
<td>-(b)</td>
<td>-(b)</td>
</tr>
<tr>
<td>Butterflyfish</td>
<td>0.438</td>
<td>0.634</td>
<td>0.43</td>
</tr>
<tr>
<td>Grunts</td>
<td>0.092</td>
<td>0.053</td>
<td>0.093</td>
</tr>
<tr>
<td>Parrotfish</td>
<td><strong>0.001</strong></td>
<td><strong>0.014</strong></td>
<td><strong>0.001</strong></td>
</tr>
<tr>
<td>Snappers</td>
<td>-(c)</td>
<td>-(c)</td>
<td>-(c)</td>
</tr>
<tr>
<td>Nassau groupers</td>
<td>-(c)</td>
<td>-(c)</td>
<td>-(c)</td>
</tr>
<tr>
<td>Groupers</td>
<td>-(c)</td>
<td>-(c)</td>
<td>-(c)</td>
</tr>
</tbody>
</table>

(a) For Reef Check, all macroalgal data were zeros, so no comparison between programs is possible and the test for a year effect is for the URI program only; (b) for sea fans, the two programs recorded data in different units so a test for a year effect was performed separately for each program (URI, p=0.606; Reef Check; p=0.002); (c) for snappers, Nassau groupers, and other groupers there were too many zero counts to reliably fit a model.
Figure 1. Paired comparisons of benthic indicators measured at the same site at the same time (n = 7). Plotted are marginal means (± 95% CI) plus results of paired t-tests (NS = p≥0.05; * p≤ 0.05; *** = p<0.002; na = no test performed because all Reef Check counts were zero). Full names of the benthic indicators are Hard Coral (HC), Soft corals (SC), Macroalgae (NIA), Sponges (SP), Rubble (RB), and Sand (SD).
Figure 2. Paired comparisons of fish counts performed at the same site at the same time (n=7). Plotted are marginal means (± 95% CI) plus results of paired t-tests (NS = p≥0.05; * = 0.05; na = no test performed because all URI counts were zero). No Nassau groupers were counted by either group during the paired surveys, so no plot is presented for this indicator.
Figure 3. Hard coral % cover estimates for each site.
Figure 4. Soft coral % cover estimates for each site. See Figure 3 for colour codes identifying the sites.
Figure 5. Parrot fish counts for each site. See Figure 3 for colour codes identifying the sites.
Figure 6. Snapper counts for each site. Note that the y-axis is scaled differently for the two programs. See Figure 3 for colour codes identifying the sites.
Supplementary online material

Available online are a map of the study sites (Fig. S1) and plots for site-specific changes over time in sponges (Fig. S2), macroalgae (Fig. S3), rubble (Fig. S4), sand (Fig. S5), sea fans (Fig. S6), butterflyfish (Fig. S7), grunts (Fig. S8) and groupers (Fig. S9).
Fig. S1. Map of the study sites. URI sites are indicated with white circles and Reef Check sites with yellow circles.
Fig. S2. Changes in sponge % cover at each site.
Fig. S3. Changes in macroalgal % cover at each site.
Fig. S4. Changes in sand % cover at each site.
Fig. S5. Changes in rubble % cover at each site.
Fig. S6. Changes in sea fan abundance at each site. For URI sites, abundance is estimated as % cover, whereas for Reef Check sites units are colonies per m².
Fig. S7. Changes in butterfly fish density at each site.
Fig. S8. Changes in grunt density at each site. Note that grouper counts were much lower during the URI surveys than during Reef Check surveys.
Fig. S9. Changes in grouper density at each site. Note that densities reported by the URI observer were much lower than those recorded by Reef Check observers.