

Consideration of species community composition in statistical analyses of coral disease risk

Susan Harrell Yee^{1,*}, Deborah L. Santavy¹, John W. Kern², Sarah Hession²

¹United States Environmental Protection Agency, Gulf Ecology Division,
National Health and Environmental Effects Research Laboratory, Gulf Breeze, Florida 32561, USA

²Kern Statistical Services, Inc., Sauk Rapids, Minnesota 56379, USA

ABSTRACT: Diseases are increasing in marine ecosystems, and these increases have been attributed to a number of environmental factors including climate change, pollution, and overfishing. However, many studies pool disease prevalence into taxonomic groups, disregarding host-species composition when comparing sites or assessing environmental impacts on patterns of disease presence. We used simulated data under a known environmental effect to assess the ability of standard statistical methods (binomial and linear regression, ANOVA) to detect a significant environmental effect on pooled disease prevalence with varying species abundance distributions and relative susceptibilities to disease. When one species was more susceptible to a disease and both species only partially overlapped in their distributions, models tended to produce a greater number of false positives (Type I error). Differences in disease risk between regions or along an environmental gradient tended to be underestimated, or, even in the wrong direction, when highly susceptible taxa had reduced abundances in impacted sites, a situation likely to be common in nature. Including relative abundance as an additional variable in regressions improved model accuracy, but tended to be conservative, producing more false negatives (Type II error) when species abundance was strongly correlated with the environmental effect. An analysis of field data on bleaching from the Florida Keys, FL, USA, affirmed general results of model simulations. Investigators should be cautious of underlying assumptions about species similarity in susceptibility and species composition when interpreting pooled data on disease risk.

KEY WORDS: Coral · Disease · Bleaching · Community composition · Species distribution · Regression models

—Resale or republication not permitted without written consent of the publisher—

INTRODUCTION

Disease incidence for many marine species has increased in frequency over the past few decades, and many factors including pollution, overfishing, and climate change have been hypothesized to contribute to such proliferations (Harvell et al. 2002, Lafferty et al. 2004, Aeby et al. in press). While many studies assess disease patterns for a single species (Bruckner & Bruckner 2006, Gochfeld et al. 2006), species-specific information is often challenging to analyze due to insufficient sample sizes, and disease prevalence is pooled to assess patterns for a larger taxonomic group (e.g. Porter et al. 2001, Willis et al. 2004, Lips et al. 2006, Weil & Cróquer 2009). Studies that disregard

changes or differences in species composition, density and disease susceptibility of host communities when reporting disease prevalence can lead to erroneous conclusions regarding patterns of disease risk (Williams et al. 2010, Aeby et al. 2011). This can be especially critical when the composition of susceptible host populations has changed in response to past disease infections or bleaching mortality when base populations grossly differ between sites or sampling units.

Coral diseases (Santavy et al. 2001, 2006, Sutherland et al. 2004, Willis et al. 2004) have increased in frequency over the past 4 decades, and are thought to be a major cause of dramatic declines in coral abundance (Aronson & Precht 2001, Patterson et al. 2002, Weil et al. 2006, Bruckner & Hill 2009). Coral disease out-

*Email: yee.susan@epa.gov

breaks are spatially and temporally variable, and likely influenced by a variety of abiotic and biotic factors, including elevated sea temperatures (Gil-Agudelo & Garzón-Ferreira 2001, Jones et al. 2004, Selig et al. 2006, Bruno et al. 2007), sedimentation (Voss & Richardson 2006), pollution (Kaczmarek et al. 2005), and nutrients (Bruno et al. 2003). Mass bleaching events, in which coral lose their symbiotic algae, have also caused extensive coral mortality in recent decades (Hoegh-Guldberg 1999, Hughes et al. 2003, Wilkinson 2008). Mass bleaching events can be predicted by sustained periods of elevated sea temperatures (Liu et al. 2005) and may be exacerbated by elevated solar radiation (Gleason & Wellington 1993, McField 1999) or diminished by storm events (Manzello et al. 2007). Both coral disease and bleaching patterns are influenced by different species susceptibility and host density (Bruno et al. 2007, Aeby et al. 2011), even though disease and bleaching outbreaks have been observed to coincide (Jones et al. 2004, Miller et al. 2006, Muller et al. 2007, Brandt & McManus 2009).

A number of studies have examined associations between environmental factors and temporal or spatial patterns of disease or bleaching prevalence (e.g. Gil-Agudelo & Garzón-Ferreira 2001, Kuta & Richardson 2002, Borger & Steiner 2005, Page & Willis 2006, Voss & Richardson 2006, Aeby et al. in press). Studies to look at patterns of disease or bleaching across broad spatial or temporal scales often focus on pooled observations across all susceptible coral species (Goreau et al. 2000, Porter et al. 2001, Weil & Cróquer 2009, Zvuloni et al. 2009). Observations of coral disease or bleaching prevalence are often reported in broad taxonomic groupings, based on coral cover (Bruno et al. 2007) or abundance by pooling species to genus or family levels (Edmunds 1991, McClanahan et al. 2001, Willis et al. 2004, Brandt & McManus 2009). In the Indo-Pacific, where species diversity is high, it may not be practical to classify corals to species level. However, it may not be correct to assume that all species within a single genus are equally susceptible to diseases (Aeby et al. 2011). Similarly, species vary in their susceptibility to bleaching (McField 1999, Marshall & Baird 2000, Floros et al. 2004, Yee et al. 2008), yet models to predict bleaching events under elevated temperatures are typically derived from pooled observations of coral bleaching across taxa (Berkelmans et al. 2004, Liu et al. 2005, Maina et al. 2008). Combining data across multiple host taxa may obscure independent disease trends (Lafferty et al. 2004, Williams et al. 2010).

Statistical analyses of disease or bleaching patterns relative to environmental gradients for pooled data have inherent assumptions that the composition of susceptible species is consistent between compared study areas, an assumption that is not often tested

(Kuta & Richardson 1996, Bruno et al. 2007, Brandt & McManus 2009). A key issue is whether sources of environmental variation contributing to patterns of disease prevalence across regions or along a gradient are confounded with patterns in species composition. For example, the variance associated with differences in relative species abundances among different locations are difficult to separate from the factors tested for association with temporal or spatial differences in bleaching or disease prevalence (Williams et al. 2010, Aeby et al. 2011). Unaccounted for variations in species composition that are associated with environmental gradients may mask true patterns of disease or bleaching. Some studies have acknowledged that results from the analysis of pooled disease data could be influenced by differences in the spatial distribution among coral species (Borger 2003, Page et al. 2009). Recognizing these issues, other studies have conducted statistical analyses of bleaching prevalence while considering weightings of species relative abundances (McClanahan et al. 2007, Yee et al. 2008), species richness, and evenness (Ward et al. 2006), percent coral cover (Page & Willis 2006, Bruno et al. 2007), or reef type (Berkelmans et al. 2004, Wooldrige & Done 2004). If adjustments are not made for variation in species composition, actual effects due to environmental gradients may be either over- or understated, depending on the nature of species abundance patterns and susceptibility.

In the present study, simulated communities were generated using statistical models to investigate the ability to detect significant environmental effects on disease prevalence when species are pooled. Because simulated communities were generated based on known environmental effects, statistical model predictions can be compared to these specified true effects. Species were allowed to vary in their susceptibility to disease (low vs. high) and their distribution (widespread vs. narrow), either between 2 distinct regions or along an environmental gradient. We generated simulated data on pooled disease prevalence under various species distribution and susceptibility scenarios, then fit models to the simulated data to test for a significant environmental effect on disease prevalence. We compared models that only included an environmental effect (ANOVA or simple regression) to those that also included species abundance as a covariate (ANCOVA or multiple regression). For each model and scenario, we investigated the propensity for Type I or Type II errors, i.e. false positives or false negatives, and the accuracy of the predicted environmental effect relative to the known true effect. We assessed the relevancy of simulation results toward detecting regional differences in coral bleaching by analyzing field data from the Florida Keys, FL, USA.

METHODS

Simulated data. Discrete regions: We simulated data on disease prevalence for sites within 2 discrete regions, reflecting, for example, a control region and an impacted region. Sites are considered the sampling unit, such as a transect or quadrat, with N replicate sites per region. We simulated a community of 2 species, which varied in both their susceptibility to disease (Fig. 1a,c,d) and their distribution between the 2 regions (Fig. 1b,d,f). Each species was assumed to

have some baseline or minimum disease susceptibility in the control region ($p_{i,0}$), defined as the mean probability of species i being infected with disease in the control region. In general, we assume one species is more susceptible to disease infection ($p_{1,0} > p_{2,0}$). We examined 3 possible abundance distributions for each species i , determined by their mean abundance per site ($A_{i,j}$) in each region j (control: $j = 0$; impacted: $j = 1$): uniformly abundant across both regions, corresponding to a widespread distribution ($A_{i,0} = A_{i,1}$); more abundant in the control region, corresponding to a narrow

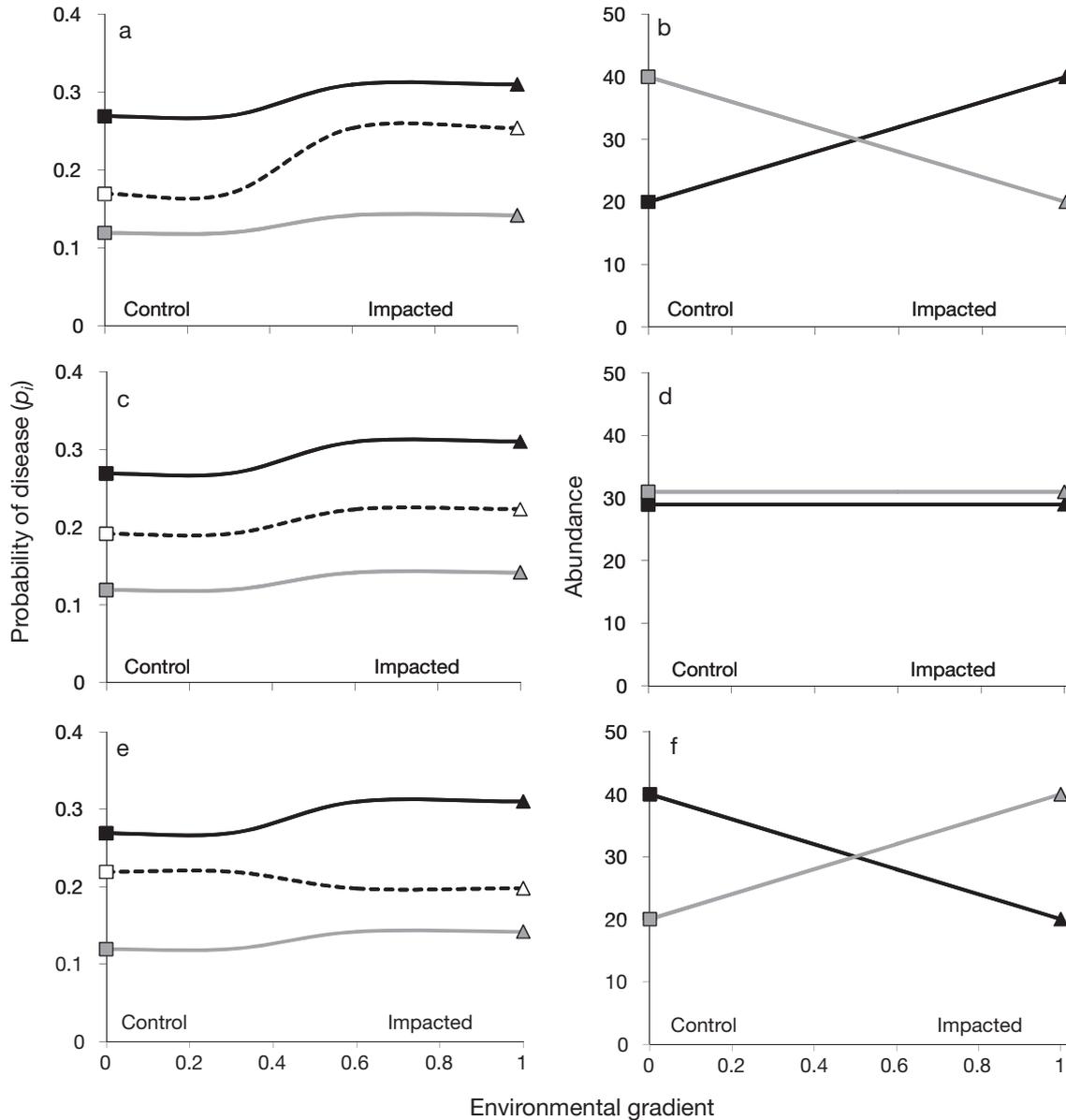


Fig. 1. Probabilities of disease along an environmental gradient for a more susceptible species (p_1 ; black lines) and a less susceptible species (p_2 ; grey lines), and the observed pooled probabilities of disease across species (p_{pooled} ; dashed lines) for alternative abundance distributions in which: (a,b) the most susceptible species is more abundant at highly impacted sites, (c,d) the 2 species are similarly abundant along the gradient, and (e,f) the most susceptible species is more abundant at non-impacted sites. Symbols indicate probability of disease and abundance for the case of 2 discrete regions: control (squares) and impacted (triangles)

distribution ($A_{i,0} > A_{i,1}$); or more abundant in the impacted region, also corresponding to a narrow distribution ($A_{i,0} < A_{i,1}$). Varying the distribution of each species gives several different scenarios for species composition between regions. For example, both species may be widespread (Fig. 1d) or the more susceptible species may be more abundant in either the control (Fig. 1f) or impacted region (Fig. 1b), while the distribution of the less susceptible species is the opposite.

In the impacted region, individuals of each species are assumed to have an additional probability of disease infection above their baseline susceptibility in the control region. Because probabilities are constrained between 0 and 1, we used a sigmoidal logistic model to simulate the probability of disease incidence (Fig. 1a). This nonlinear model can be written as a linear equation by rewriting the probability in terms of the odds that an event occurs, which is the probability an event occurs relative to its converse, and taking the natural log (Quinn & Keough 2002). We can then simulate an increase in the log-odds of disease infection between the control and impacted regions as:

$$\ln\left(\frac{P_{i,\text{Region}}}{1-P_{i,\text{Region}}}\right) = \ln\left(\frac{P_{i,0}}{1-P_{i,0}}\right) + m \times \text{Region} \quad (1)$$

where $p_{i,j}$ is the probability of disease infection for each species i ($i = 1, 2$) in either the control (region = 0) or impacted regions (region = 1), $p_{i,0}$ is the baseline susceptibility of each species in the control region, and m is the known increase in the log-odds of disease for an individual present in the impacted region. Eq. (1) defines the mean log-odds of disease infection in the control or impacted regions, with the variability in disease incidence at replicate sites within each region distributed around this mean.

To generate random communities, we drew the total number of individuals of each species i at each site within each region j from a Poisson distribution, with mean abundance given by $A_{i,j}$. The number of infected individuals was then drawn from a binomial distribution, with mean probability of infection given by $p_{i,j}$ in each region (Eq. 1). We kept track of the number of individuals of each species and the total number of diseased individuals at each site. Disease prevalence was pooled across species to obtain the fraction of diseased individuals at each site. Observed disease prevalence was then fit by models which included region (0, 1), as well as the relative abundance of the species at each site as variables:

$$\ln\left(\frac{P_{\text{pooled,Region}}}{1-P_{\text{pooled,Region}}}\right) = \beta_0 + \beta_1 \times \text{Region} + \beta_2 \frac{A_{1,\text{Region}}}{A_{1,\text{Region}} + A_{2,\text{Region}}} \quad (2)$$

where β are the estimated model coefficients giving the baseline disease risk (β_0), the additional risk due to the impacted region (β_1), and the effect of relative abundance as a covariate (β_2).

We assessed our ability to detect significant differences in disease prevalence between regions when species susceptibilities were different by generating simulated data under 2 cases: species with similar ($p_{1,0} = p_{2,0}$) or different ($p_{1,0} > p_{2,0}$) baseline susceptibilities to disease in the control region. We also assessed the effects of species compositional differences between regions by varying the distribution of the 2 species. We used 7 distribution scenarios for the 2 species, in which each species had a mean abundance of $A_{i,j} = 40$ in its preferred region and $A_{i,j} = 20$ in its less preferred region (Table 1). In 3 of the scenarios, species completely overlapped: both widespread (WW), both more abundant in control region (CC), or both more abundant in impacted region (II). In 2 scenarios, species were partially overlapped with the less susceptible species widespread and the most susceptible species more abundant in either the control (CW) or impacted region (IW). There were also 2 scenarios where species were minimally overlapped with the most susceptible species more abundant in either the control (CI) or impacted region (IC), and the less susceptible species having the opposite distribution. In a separate set of simulations, we further narrowed the distribution of the most susceptible species (IW, CW) by reducing the abundance in the non-preferred region from $A_{1,j} = 20$ to $A_{1,j} = 1$. For all simulations, each of the 2 regions consisted of 20 sites. For each combination of parameters, we ran 500 simulations and fitted models to each set of simulated data.

Table 1. Abundance distributions ($A_{i,j}$) used to generate the 7 species-distribution scenarios illustrated in Fig. 3

Scenario designation	Species 1			Species 2		
	Preferred distribution	Control ($A_{1,0}$)	Impacted ($A_{1,1}$)	Preferred distribution	Control ($A_{2,0}$)	Impacted ($A_{2,1}$)
IC	Impacted	20	40	Control	40	20
IW	Impacted	20	40	Widespread	40	40
II	Impacted	20	40	Impacted	20	40
WW	Widespread	40	40	Widespread	40	40
CC	Control	40	20	Control	40	20
CW	Control	40	20	Widespread	40	40
CI	Control	40	20	Impacted	20	40

Environmental gradient: In addition to 2 discrete regions (Fig. 1, symbols), we also allowed susceptibility and abundance to vary along an environmental gradient (Fig. 1, lines). The environmental gradient could represent a specific environmental variable, such as temperature or chemical concentration. For consistency with the case of discrete regions, we assume sites occur along a continuous linear environmental gradient, standardized to range from control ($X = 0$) to highly impacted conditions ($X = 1$). Similar to the discrete case (Eq. 1), log-odds of disease for each species increase linearly along the environmental gradient as:

$$\ln\left(\frac{p_i(X)}{1-p_i(X)}\right) = \ln\left(\frac{p_{i,0}}{1-p_{i,0}}\right) + mX \quad (3)$$

where baseline probabilities of disease risk under control conditions ($X = 0$) are given by $p_{i,0}$ for each species i . In both the discrete region case and the environmental gradient, m represents the size of the effect, the only difference being whether the predictor variable is discrete (region) or continuous (X).

In contrast to the discrete region case, where species mean abundance was fixed within each region, in the gradient case, we allowed species abundances to increase or decrease linearly along the gradient. Each species i was assumed to have a linear distribution in abundance along the environmental gradient $A_i(X) = A_{i,0} + \alpha_i X$, with an abundance of $A_{i,0}$ at one end of the gradient and a slope ranging from $-\alpha_i$ to $+\alpha_i$ defining whether the species abundance is negatively, positively, or not correlated with the environmental variable X .

Random communities were generated by drawing the total number of individuals of each species i at each of 40 sites along gradient X from a Poisson distribution, with mean abundance given by $A_i(X)$. The number of infected individuals was drawn from a binomial distribution, with mean probability of infection given by $p_i(X)$ in Eq. (3). We examined how the ability to detect a significant environmental effect depends on the difference in baseline susceptibility ($p_{1,0} - p_{2,0}$) and the disparity in species overlap ($\alpha_1 - \alpha_2$), where α_i is the slope of each species abundance along the environmental gradient. For each species, we varied α_i from -40 to 40 and $A_{i,0}$ from 0 to 40 , in combinations that maintained a maximum abundance of 40 individuals and did not allow negative abundances. The disparity in distribution then ranged from -80 (Species 1 decreases in abundance along the gradient, $\alpha_1 = -40$; Species 2 increases, $\alpha_2 = 40$) to 80 (Species 1 increases in abundance along the gradient, $\alpha_1 = 40$; Species 2 decreases, $\alpha_2 = -40$). We examined 3 cases of differential susceptibility, none ($p_{1,0} = p_{2,0} = 0.12$), medium ($p_{1,0} = 0.27$, $p_{2,0} = 0.12$),

and high ($p_{1,0} = 0.5$, $p_{2,0} = 0.12$). For each combination of parameters, we ran 500 simulations and fitted models to each set of simulated data.

Statistical methods. Regression models: Environmental effects between regions or along a gradient in simulated data were estimated using linear regression, logistic regression, and beta-binomial regression models (McCullagh & Nelder 1989). Unlike linear regression, logistic and beta-binomial regression have assumptions that the underlying probability distribution is binomial (presence/absence of disease) rather than normal. Furthermore, beta-binomial regression, unlike standard logistic regression, accounts for colonies being clustered within independent sites. Colonies within a site cannot be considered independent because they experience the same local site conditions, and can be expected to respond similarly. The fitted beta-binomial model has 1 additional parameter, which measures the degree of correlation within a cluster, or overdispersion, and provides a correction to the amount of variability explained by the model. Linear and logistic regressions were implemented using the package 'glm', and beta-binomial regressions were implemented using the package 'aod' in the statistical package R (www.r-project.org).

For each simulation, using beta-binomial regression or logistic regression models, we evaluated whether the log-odds of pooled disease risk were significantly different between regions or along the environmental gradient, and estimated the predicted size of the effect (m) by the slope of the regression. Similarly, linear regression was used to determine whether the fraction of diseased individuals was significantly related to the environmental variable, which, in the case of discrete regions, is equivalent to analysis of variance (ANOVA).

When observations of disease prevalence are pooled across species, the observed response depends on the relative abundances of species within a site (Fig. 1). We, therefore, compared models with only an environmental effect to models in which relative species abundance was included as an additional covariate in models (Eq. 2). We also evaluated 2 other metrics of species composition as model covariates (total species abundance and species presence/absence), but, overall, models with relative species abundance performed better (not shown).

The beta-binomial model is expected to perform better when mean disease susceptibilities ($p_{i,\text{region}}$) vary between sites within a region, such that susceptibilities within a site are more similar than those between sites (i.e. within-site correlation would be expected), a situation which is likely to occur in nature. Therefore, to demonstrate the improved performance of the beta-binomial model, we compared model accuracy using the beta-binomial regression model, a logistic regres-

sion model, and a linear regression model in 2 cases: (1) using a between-site variability in disease susceptibility of $\sigma = 0$, in which the mean baseline disease susceptibility ($p_{i,0}$) was the same across all sites within a region, and (2) incorporating between site variability by drawing the mean baseline susceptibility ($p_{i,0,k}$) for each site k from a normal distribution with mean $p_{i,0}$ and a standard deviation of $\sigma = 0.5$. It is expected that the logistic model, which treats individual colonies as independent experimental units, would overstate the level of significance of effects resulting in inflated Type I error rates. Normal theory linear regression models are expected to perform better than logistic regression, because each sampling station is treated as an experimental unit, but may be less efficient than beta-binomial regression because the model is not informed by the number of colonies sampled within experimental units.

Model accuracy: Model accuracy was assessed through comparison of power curves. We varied the true effect (m) of either the impacted region (Eq. 1) or the environmental variable (Eq. 3) from 0 to 3, and examined the fraction of simulated data sets for which the effect of region was significant (p-value < 0.05). For $m = 0$, there is no additional disease risk in the impacted site or along the environmental gradient, such that one expects a low detection rate of a significant region effect, i.e. a Type I error rate of approximately 5%. For sufficiently high m , there is an additional disease risk in the impacted region or disease risk increases along the environmental gradient, such that one expects a high detection rate of a significant region effect, i.e. with power to detect a region effect of approximately 95% and a Type II error rate of approximately 5%. We also evaluated model accuracy in binomial regressions by comparing the true region effect or slope along the environmental gradient used to generate simulated data (m) to the predicted region effect given by the estimated model coefficient (β_1 ; Eq. 2).

Field data example. We evaluated the applicability of model simulation results to field data by using coral bleaching data for the Florida Keys to evaluate whether the probability of coral bleaching differed between 2 regions subjected to differing levels of anthropogenic stress. Local anthropogenic stressors may reduce coral resistance to bleaching (West & Salm 2003). Reefs in the upper Florida Keys region are in close proximity to land, and are subjected to potential anthropogenic stress from pollutant runoff. Reefs in the Dry Tortugas and New Grounds, in contrast, are open oceanic and distant from land-based sources of pollution.

Field surveys were conducted at permanent sampling stations along the Florida Keys reef tract from the

Upper Keys (25° N, 80° W) to the Dry Tortugas (24° N, 83° W). Site locations, characteristics, and survey methods are detailed elsewhere (Santavy et al. 2001, 2006). Survey sites were randomly selected from areas known to contain hard coral bottom, approximately 5 to 10 km offshore. Sites ranged in depth from 1 to 23 m, and contained at least 5% live coral cover. Coral colonies were counted within the 2 m circular band located from 8 to 10 m from the center of each station, with a total survey area per station of 113 m². Surveys were conducted from 7 August to 12 September 1998, during a period of elevated water temperatures when a mass bleaching event was occurring.

We pooled bleaching observations across species to determine the total number of bleached colonies per survey station. We compared the probability of coral bleaching between 2 regions, the Upper Keys (9 stations) and Dry Tortugas/New Grounds (18 stations), using beta-binomial regression. We compared model results with region as the only independent variable, to models in which species relative abundances were also included. We used stepwise AIC (Akaike information criterion) to determine which species to include in models (Burnham & Anderson 2002). To better understand results for pooled species bleaching, we also examined the probability of bleaching for individual coral species in separate beta-binomial regressions conducted for each species.

RESULTS

Regression models

In general, linear, logistic, and beta-binomial models produced similar power curves, where the probability of detecting significant differences in disease risk between regions increased with strength of the known effect (m) used to generate simulated data (Fig. 2a). When between-site variability in susceptibility was low within each region ($\sigma = 0$), power curves for each of the 3 types of regression models were similar, with detection of significant region effects slightly lower in the linear regression model than either logistic or beta-binomial regression (Fig. 2a). However, the beta-binomial regression model performed best when mean susceptibility was variable among sites within a region ($\sigma = 0.5$; Fig. 2b). The logistic regression model had a high rate of false positives (Type I error), detecting significant effects of region at rates >20%, compared with the nominal 5% stated false positive rate, when effects were absent ($m = 0$). We focus the rest of our results on beta-binomial regressions, noting that results were similar or worse for other types of regression models.

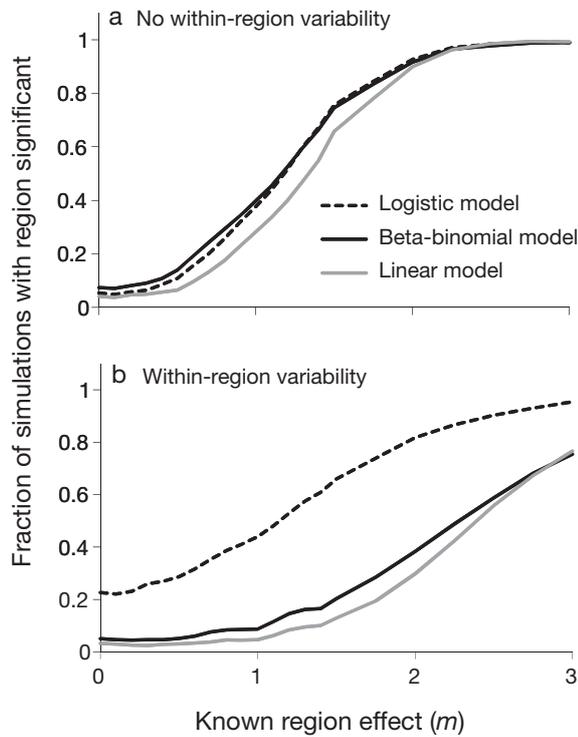


Fig. 2. Fraction of simulations in which significant differences between impacted and control regions were detected for logistic, beta-binomial, or linear models, including both region and relative abundance as model variables. Baseline susceptibility among sites was either: (a) fixed mean within a region ($\sigma = 0$) or (b) variable among sites within a region ($\sigma = 0.5$). Simulation parameters were such that Species 1 was more susceptible to disease ($p_{1,0} = 0.27, p_{2,0} = 0.12$) and more abundant in the control region ($A_{1,0} = 40, A_{1,1} = 1$), and Species 2 was uniformly distributed across regions ($A_{2,0} = A_{2,1} = 40$)

Species composition between regions

When species had similar baseline disease susceptibilities ($p_{1,0} = p_{2,0}$), the distribution of each species across the 2 regions only had minor impacts on our ability to detect significant differences between regions using beta-binomial regression (Fig. 3a; Table 1). The ability to detect significant differences between regions was highest when both species were widely distributed with similar abundances in both impacted and control regions (WW). Detection of significant differences between regions was slightly lower when one or both species were limited in their distribution, with reduced abundance in either control (CC) or impacted regions (II).

When the 2 species differed in their baseline disease susceptibilities ($p_{1,0} > p_{2,0}$), species distribution greatly affected the detection of significant differences between regions (Fig. 3b; Table 1). In cases where the 2 species overlapped in distribution as either widespread (WW), both more abundant in control region

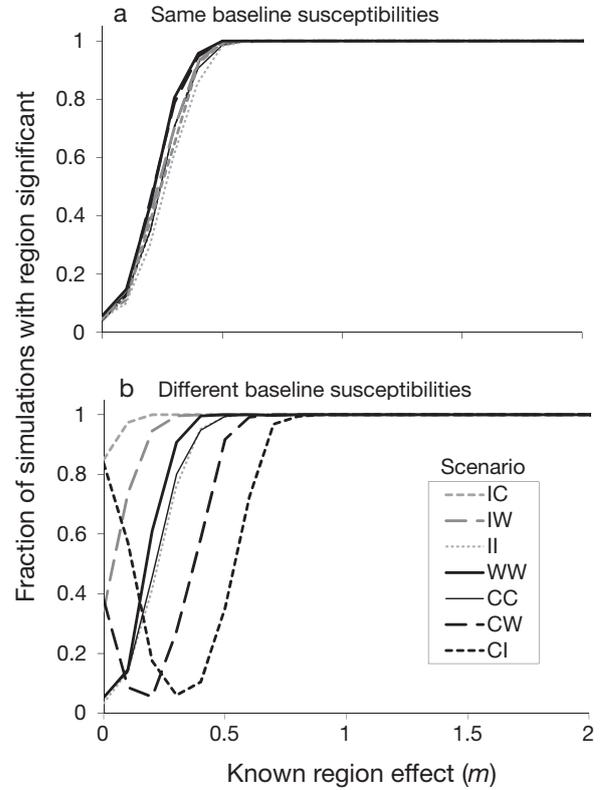


Fig. 3. Fraction of simulations in which region was significant for simulations when: (a) Species 1 and 2 have equal baseline disease susceptibility ($p_{1,0} = p_{2,0} = 0.12$) or (b) Species 1 is more susceptible than Species 2 ($p_{1,0} = 0.27, p_{2,0} = 0.12$) for 7 combinations of species distributions (see Table 1 for scenarios)

(CC), or both more abundant in impacted region (II), power curves where the 2 species differed in baseline susceptibility (Fig. 3b) were only slightly different from the ideal scenario in which both species were widespread with equal baseline susceptibility (Fig. 3a). However, cases where the 2 species only partially overlapped in their distribution produced large Type I and Type II error rates (Fig. 3b).

When the more susceptible species was more abundant in impacted sites (IC, IW), there was a high rate of false positives (Type I error) in which significant differences between regions were erroneously detected (Fig. 3b). The error rate increased when the distribution of Species 2 was changed from widespread (IW) to the opposite distribution of Species 1 (IC). The error rate also increased as the more susceptible species became more highly limited in its distribution (IW: Fig. 3b, $A_{1,0} = 20$ when rare; Fig. 4a; $A_{1,0} = 1$ when rare). Indeed, the predicted difference between regions in such cases was much greater than the known value (m) used to generate simulated data (Fig. 4b). In the control region, the pooled disease prevalence is dominated by the lower susceptibility of Species 2, whereas the higher susceptibility of Spe-

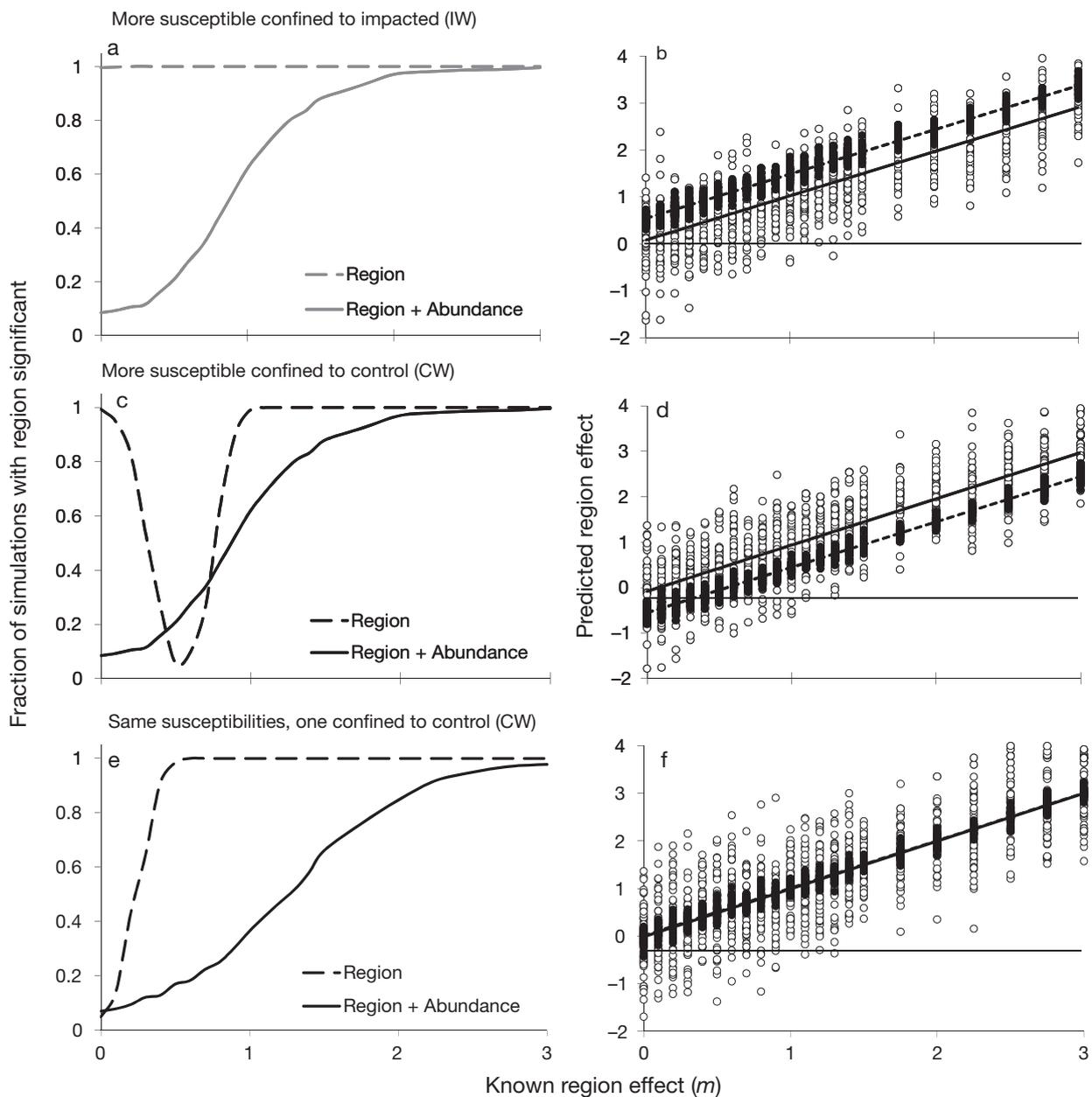


Fig. 4. Model accuracy for models without (a,c,e: dashed lines; b,d,f: filled circles) and with abundance (a,c,e: solid lines; b,d,f: open circles) for simulations in which Species 1 is more susceptible than Species 2 ($p_{1,0} = 0.27$, $p_{2,0} = 0.12$) and highly limited in distribution to either: (a,b) impacted regions (IW; $A_{1,0} = 1$, $A_{1,1} = 40$) or (c,d) control regions (CW; $A_{1,0} = 40$, $A_{1,1} = 1$), or in which (e,f) species have the same baseline susceptibilities ($p_{1,0} = 0.12$, $p_{2,0} = 0.12$), but Species 1 is limited to control regions (CW). Species 2 is equally abundant in both regions ($A_{2,0} = A_{2,1} = 40$). Power curves (a,c) give the fraction of simulations in which differences between regions were significant, and fitted lines (b,d) give the relationship between the known (m) and predicted size (β_1) of the effect

cies 1 contributes more strongly in the impacted region, thus overestimating the true difference in disease risk between regions.

Error rates were also high when the more susceptible species, Species 1, was more abundant in the control region and Species 2 was either widespread (CW) or had the opposite distribution to Species 1 (CI: Fig. 3b). The error rate was further exaggerated as the

more susceptible species became more highly limited in its distribution (CW: Fig. 3b; $A_{1,0} = 20$ when rare; Fig. 4c; $A_{1,0} = 1$ when rare). The difference in disease risk between regions was significant for a high fraction of simulated datasets, even when the known true difference was very low ($m \sim 0$). In fact, the predicted difference between regions was actually the opposite of the true value, such that disease prevalence was pre-

dicted to be higher in the control region (Fig. 4d). In the control region, the pooled disease prevalence is dominated by the higher susceptibility of Species 1, whereas the lower susceptibility of Species 2 contributes more strongly in the impacted region. This creates an underestimate of the known region effect, and the estimated effect might even be negative depending on the size of m .

Relative abundance as a covariate

In general, incorporating relative abundances as an additional explanatory variable in beta-binomial regressions greatly improved the Type I error rate, reducing the fraction of false positives (Fig. 4a,c). On average, predicted differences in disease risk between the 2 regions for models with abundance as a covariate more closely matched known values of m (Fig. 4b,d). There was greater variability, however, in the estimated region effect between simulations in models with both region and abundance. In cases where susceptibilities were similar between species, there was a greater rate of false negatives in models with abundance when species were narrowly distributed (Fig. 4e,f). In such circumstances, models with or without abundance gave reasonable estimates of m on average, but significant differences between regions were detected as much as 60% less often, when the true difference between regions was moderate ($m = 1$). In cases where a species distribution was limited to a particular region, then region and abundance were highly correlated, and region provided little additional explanatory power when abundance was also in the model.

Distribution along a gradient

To generalize results from discrete regions, we allowed species disease susceptibilities and abundances to vary linearly along an environmental gradient. For regressions which only included the environmental effect as a variable, model accuracy depended strongly on 2 factors: the difference in baseline susceptibilities between the 2 species and the disparity in species distributions (Fig. 5a). Disparity in distribution, quantified as the difference between the slopes of the 2 species abundances along the environmental gradient ($\alpha_1 - \alpha_2$), had little effect on model accuracy when species had similar sensitivities. However, as the differences in species susceptibilities increased ($p_{1,0} > p_{2,0}$), regressions with only the environmental effect tended to over- or underestimate the known environmental effect (m). The error was increasingly strong as

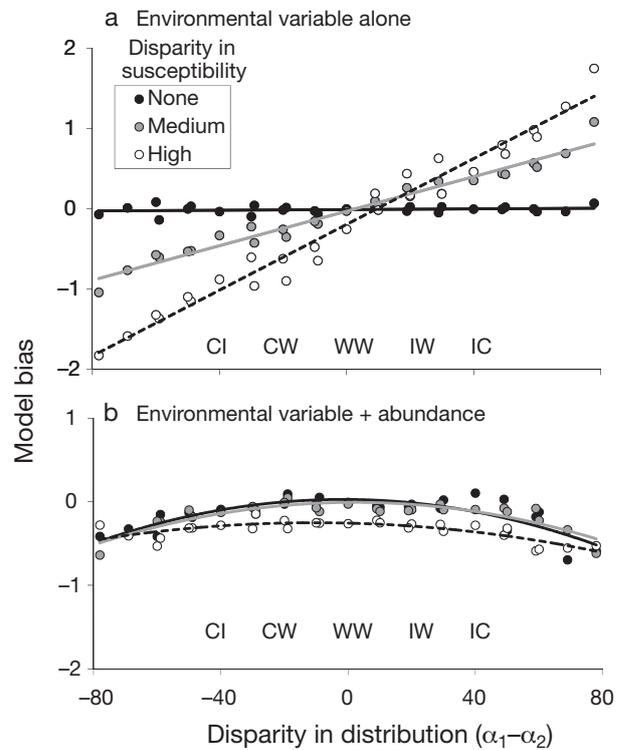


Fig. 5. Model accuracy (predicted effect β_1 – known effect m) for simulations in which species abundances varied linearly along a gradient from control to impacted sites fitted with models of: (a) the environmental variable alone or (b) the environmental variable + abundance. Slopes α_i ranged from -40 to 40 for each species (see Fig. 1), with the disparity in distribution given by the difference between the slopes ($\alpha_1 - \alpha_2$). Disparity values for the discrete cases from Table 1 are labeled (CI, CW, WW, IW, IC, see Table 1 for scenarios). The difference in baseline susceptibilities between species was none ($p_{1,0} = p_{2,0} = 0.12$), medium ($p_{1,0} = 0.27, p_{2,0} = 0.12$), or high ($p_{1,0} = 0.5, p_{2,0} = 0.12$). Points for each combination of slopes are means across 500 simulations

the disparity in species distributions increased. When the more susceptible species abundance decreased linearly along the gradient from control to impacted ($-\alpha_1$), and Species 2 had the opposite distribution ($+\alpha_2$), regression models tended to underestimate the environmental effect. In contrast, models overestimated the environmental effect when the more susceptible species increased along the gradient with higher abundance in impacted sites ($+\alpha_1$), with the error exaggerated as the distribution of Species 2 became increasingly narrow in the opposite direction ($-\alpha_2$).

Including relative abundance as an additional variable in regressions improved model accuracy (Fig. 5b). The estimated environmental effect was similar to the known effect, regardless of species distribution or differences in susceptibilities. However, when species abundances were strongly correlated with the environmental gradient (large α_i ; large $|\alpha_1 - \alpha_2|$), significant

environmental effects were harder to detect and estimated as $m = 0$, thus underestimating the known effect.

Results for the environmental gradient were comparable to those of the case of discrete regions (Figs. 3 & 4), and lead to a number of generalizations (Table 2). If susceptibilities are likely to be similar across species, or species composition is likely similar between regions or along the environmental gradient, then simple regression or ANOVA to detect differences should be sufficient. However, if species differ in susceptible and composition is variable, then multiple regression or ANCOVA including relative abundance as an additional factor should minimize potential under- or over-estimates of the true environmental impact, as long as there is not high collinearity between factors.

Field data example

The range of bleaching susceptibilities and species abundance distributions used to parameterize model simulations were similar to those observed in Florida Keys field surveys. The percent of observed bleached colonies per station ranged from 0 to 47% across 14 species (Fig. 6a). Species mean abundances ranged from <1 colony to 21 colonies per station (Fig. 6b), compared to ranges of from 0 to 40 used in model simulations.

We used beta-binomial regression to assess differences in the probability of coral colony bleaching between the Upper Keys and Dry Tortugas/New Grounds. When only regional differences were inclu-

ded in the model, we found significant differences between the 2 regions, with a predicted higher rate of bleaching in the Upper Keys (Fig. 6c; Table 3). However, species composition varied substantially between the 2 regions (Fig. 6b). In particular, communities in the Upper Keys tended to be dominated by a species that was frequently observed bleaching, *Siderastrea siderea*, whereas Dry Tortugas/New Grounds stations had a high relative abundance of a species that was infrequently observed bleaching, *Acropora cervicornis* (Fig. 6a). These differences in relative species abundances contributed to an observed low rate of bleaching in the Dry Tortugas and a high rate of bleaching in the Upper Keys when species were pooled (Fig. 6c).

Including relative species abundances into models greatly reduced the predicted difference between regions, such that the difference between regions was no longer significant (Table 3). Differences in the relative species abundance alone explained the difference in pooled observations of bleaching. Four species (*Acropora cervicornis*, *Colpophyllia natans*, *Montastraea annularis*, and *M. cavernosa*) were included in the best fit model through stepwise AIC. A fifth species (*Siderastrea siderea*) was not included in the regression because of high collinearity with the region effect (Pearson correlation coefficient, $r = 0.729$). Furthermore, we analyzed individual species bleaching rates and found that only 2 of 14 species showed significant differences in bleaching between regions (*Dichocoenia stokesii*, $m = 22.2$, $p = 0.007$; *M. cavernosa*, $m = 3.4$, p -value < 0.0001), affirming our assessment that overall bleaching rates did not differ between regions.

Table 2. Summary of results for analyzing pooled data under different assumptions of species susceptibilities and species composition along an increasing gradient of a potential environmental impact

Do species differ in susceptibility?	Does relative species composition vary?	Are species relative abundances strongly correlated with gradient (collinearity)?	Correlation between abundance of most susceptible species and gradient?	Recommended analysis	Potential bias
No	No	No or Yes	Negative or Positive	Simple regression or ANOVA	Minimal
No	Yes	No or Yes	Negative or Positive	Simple regression or ANOVA	Minimal
Yes	No	No or Yes	Negative or Positive	Simple regression or ANOVA	Minimal
Yes	Yes	No	Negative or Positive	Multiple regression or ANCOVA	Minimal
Yes	Yes	Yes	Positive	Simple regression or ANOVA	Overestimate of environmental impact
Yes	Yes	Yes	Negative	Simple regression or ANOVA	Underestimate or opposite estimate of environmental impact

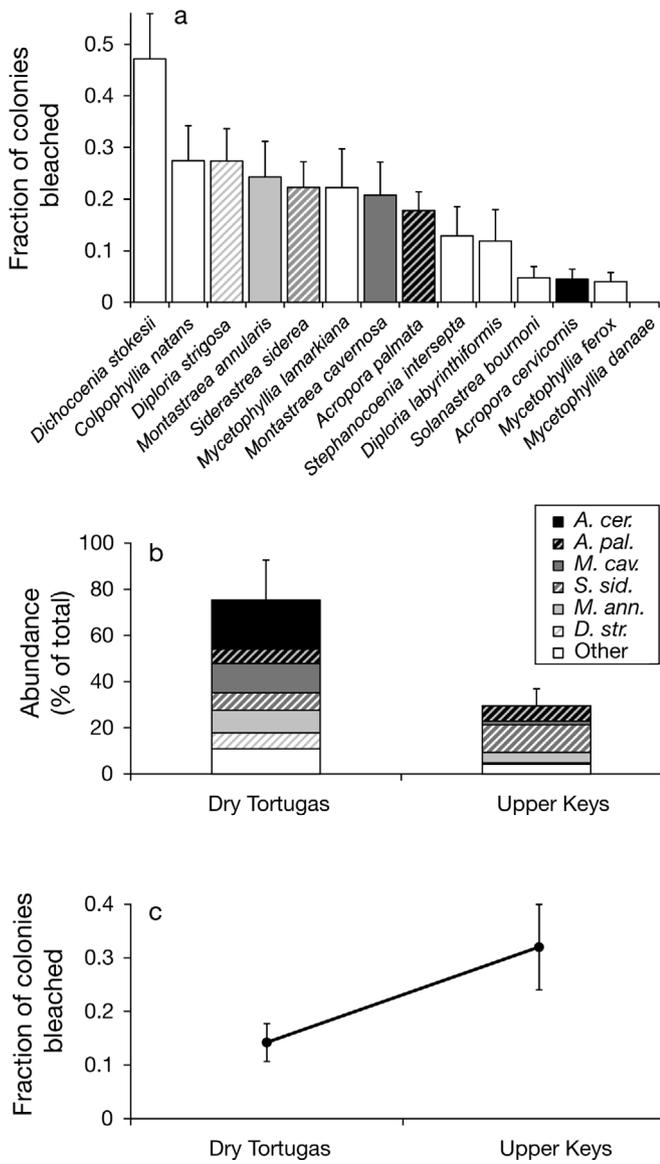


Fig. 6. Observed bleaching and coral community composition during 1998 Florida Keys, USA, station surveys including: (a) mean fraction of bleached colonies per station for each species, with the 6 most abundant species shaded, (b) relative contribution of the 6 most abundant species (full names in a) to mean colony abundance per station in each region, and (c) the mean fraction of bleached colonies per station in each region, pooled across species

DISCUSSION

Our results emphasize the need to account for species composition when assessing differences in disease prevalence along an environmental gradient. Unless species have similar distributions or similar susceptibilities between the regions being compared, using standard regression or ANOVA methods to detect differences will likely over- or underestimate the true effect, with a tendency toward false positives. Regressions may even predict opposite patterns to the true effect, particularly when more susceptible species are limited in their distribution to non-impacted or high-quality regions, a situation that may be common in nature (Gaston 1994, Nijboer & Schmidt-Kloiber 2004). Furthermore, because pooled disease prevalence is inherently linked to relative species abundances, increased sampling effort, either through higher abundances or additional sampling sites, will not improve the ability to detect significant differences between regions or along a gradient.

Our results indicate that incorporating species relative abundances as a covariate greatly improved model accuracy. However, situations where species abundance is strongly correlated with the environmental effect tend to produce large numbers of false negatives, not surprising given known issues with collinearity among predictor variables in regression (Quinn & Keough 2002). Collinearity also likely contributed to high variability in estimates of model coefficients in models with both environmental effects and abundance.

We used simplified 2-species communities in our simulations to demonstrate that, even under a relatively simple scenario, assessments of pooled disease prevalence can lead to misinterpretations of true effects. Errors will be likely further exaggerated in species-rich communities, because of the increasing likelihood of variability in species susceptibilities and shifts in community composition. These issues are a concern for coral reef studies because species composition can vary widely between reefs or over time (Loya et al. 2001, Somerfield et al. 2008). Additionally, species have been observed to be variable in their susceptibilities to either disease (Gil-Agudelo & Garzón-

Table 3. Beta-binomial regressions on Florida Keys, USA, bleaching data, in which bleaching observations were pooled across species (*Colpophyllia natans*, *Montastraea cavernosa*, *Acropora cervicornis*, *Montastraea annularis*)

Model	df	Predicted region effect (m)	Region (p-value)	Model (p-value)
Region	3	1.13	0.012	<0.001
Region + <i>C. natans</i> + <i>M. cavernosa</i> + <i>A. cervicornis</i> + <i>M. annularis</i>	7	0.28	0.432	<0.001
<i>C. natans</i> + <i>M. cavernosa</i> + <i>A. cervicornis</i> + <i>M. annularis</i>	6	0	-	<0.001

Ferreira 2001, Santavy et al. 2006) or bleaching (McField 1999, Marshall & Baird 2000, Floros et al. 2004, Yee et al. 2008).

Results from our analysis of field data, which represented a community of 14 species of variable susceptibility to bleaching, matched the general conclusions from our simplified 2-species model simulations. Our field data demonstrated a real scenario of data in which species had different susceptibilities to bleaching and species had different distributions between regions. Because abundances of highly susceptible species were positively correlated with the region effect, we would anticipate from our simulations that a regression model with region alone would tend to overestimate the region effect, which was confirmed through statistical analysis of our field data. In our case, field data included several other species without strong collinearity, which could be included in multiple regressions, providing a potentially less biased estimate of the region effect. Without historical data, however, it is impossible to ascertain whether the currently observed lack of differences between regions reflects a past event of disease or bleaching mortality that removed highly susceptible species from the impacted region.

Many field studies may have limited or no information on species coral bleaching. In such cases, researchers may need to rely on personal observations or subsampling to ascertain potential biases in their data analysis or to acknowledge that species compositional differences may be contributing to perceived differences. Even when species-specific information on disease or bleaching is available, low sample sizes may limit the ability to conduct species-specific analyses. When pooled data are necessary, it is advisable to only pool data across species with similar distributions and susceptibilities. Furthermore, susceptibilities can vary within a species, such as across depths (Gleason & Wellington 1993) or size classes (Nugues 2002), and a high density of coral may exacerbate disease outbreaks (Borger & Steiner 2005, Bruno et al. 2007). Also, bleached coral may be more susceptible to infection, such that disease outbreaks follow bleaching events (Harvell et al. 2001, Muller et al. 2007). Multiple regression methods provide one way to detect significant differences in susceptibilities between regions or along an environmental gradient while controlling for additional environmental variables, e.g. reef depth, size distribution, species total abundances, or historical disease/bleaching incidence, which may be confounded with the variable of interest.

Several studies have attempted to adjust for differences in species distributions in statistical analyses of thermal stress on disease prevalence by including coral cover as an additional explanatory variable (Page

& Willis 2006, Bruno et al. 2007). It is unclear, however, whether variations in coral cover were coincident with variation in species composition, which may have influenced interpretations of disease patterns between reefs. An alternative approach is to quantify risk through an index, which calculates a weighted average across species by assigning weightings to species depending on their relative susceptibilities (McClanahan et al. 2007). The multiple regression approaches we propose here are functionally similar to a weighted average, but allow coefficients on species relative abundances to be estimated by fitting models to sample data, rather than assigning *a priori* weightings on species susceptibilities.

Because the presence or absence of disease is a binary response variable, we chose to model the probability of disease using binomial regression. However, because individuals co-occur at a given site, they will be subjected to similar local conditions and cannot be treated as independent data points, an assumption of standard logistic regression. Beta-binomial models were selected to account for both types of variation that occur in reality: variation between individuals at the same site (binomial variation) and variation between sites within the same region or treatment group (extra-binomial variation). The standard regression model accurately represents the extra-binomial variation, but averages away the within-site binomial variation. Methods that ignore extra-binomial variation may underestimate the true standard errors of estimated treatment differences, leading to false positives (McCullagh & Nelder 1989). This would be a form of pseudoreplication (Hurlbert 1984).

Although we have focused on coral disease and bleaching prevalence, we emphasize that statistical detection of patterns for pooled effects across species can apply to other systems as well. For example, estimates of indices of ecological quality are susceptible to the inclusion of taxa with small distribution ranges (Nijboer & Schmidt-Kloiber 2004). More sophisticated methods for estimating ecological condition incorporate estimates of species distributions and tolerance to stress (Howe et al. 2007). Furthermore, it is well known that patterns in ecosystem functioning along a gradient may be driven by changes in species composition (Tilman 1999). Pooling of species may obscure strong responses of individual functional groups or species (Micheli et al. 2004). Even predictions of presence/absence of single species might be misleading if the prevalence of a target organism is not taken into account (Manel et al. 2002).

Our understanding of coral diseases is relatively young but growing (Harvell et al. 2007). In many cases, our ability to assess patterns is limited by the availability of data, particularly when looking for broad-scale

patterns across species groups (Goreau et al. 2000, Porter et al. 2001). Options for managing disease are limited, and it is important to compare contrasting environments to gain insights into factors influencing disease prevalence (Page et al. 2009). However, researchers should be cautious of potential species compositional differences that may confound our ability to assess the impacts of environmental factors on patterns of disease.

Acknowledgements. We thank 3 anonymous reviewers for helpful comments on prior versions. Field data were collected under research permits to D.L.S, E. Mueller, and L. MacLaughlin by Florida Keys National Marine Sanctuary, Dry Tortugas National Park, and Biscayne National Park from 1998 to 2005. Ship support was provided by the US Environmental Protection Agency (US EPA) (OSV 'Anderson') and the NOAA (RV 'Nancy Foster'). Field support was provided by: J. Campbell, L. MacLaughlin, E. Mueller, J. Patrick, M. Parsons, R. Quarles, and others. This paper has been reviewed in accordance with the US EPA's peer and administrative review policies and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use. This is Contribution Number 1410 from the Gulf Ecology Division.

LITERATURE CITED

- Aeby GS, Williams GJ, Franklin EC, Haapkyla J and others (2011) Growth anomalies on the coral genera *Acropora* and *Porites* are strongly associated with host density and human population size across the Indo-Pacific. *PLoS ONE* 6:e16887
- Aeby GS, Williams GJ, Franklin EC, Kenyon J, Cox EF, Coles S, Work TM (in press) Patterns of coral disease across the Hawaiian archipelago: relating disease to environment. *PlosONE*
- Aronson RB, Precht WF (2001) White-band disease and the changing face of Caribbean coral reefs. *Hydrobiologia* 460:25–38
- Berkelmans R, De'ath G, Kininmonth S, Skirving WJ (2004) A comparison of the 1998 and 2002 coral bleaching events on the Great Barrier Reef, spatial correlation, patterns, and predictions. *Coral Reefs* 23:74–83
- Borger JL (2003) Three scleractinian coral diseases in Dominica, West Indies: distribution, infection patterns, and contribution to coral tissue mortality. *Rev Biol Trop* 51:25–38
- Borger JL, Steiner SCC (2005) The spatial and temporal dynamics of coral diseases in Dominica, West Indies. *Bull Mar Sci* 77:137–154
- Brandt ME, McManus JW (2009) Disease incidence is related to bleaching extent in reef-building corals. *Ecology* 90: 2859–2867
- Bruckner AW, Bruckner RJ (2006) Consequences of yellow band disease (YBD) on *Montastraea annularis* (species complex) populations on remote reefs off Mona Island, Puerto Rico. *Dis Aquat Org* 69:67–73
- Bruckner AW, Hill RL (2009) Ten years of change to coral communities off Mona and Desecheo Islands, Puerto Rico, from disease and bleaching. *Dis Aquat Org* 87:19–31
- Bruno JF, Petes LE, Harvell CD, Hettinger A (2003) Nutrient enrichment can increase the severity of coral diseases. *Ecol Lett* 6:1056–1061
- Bruno JF, Selig ER, Casey KS, Page CA and others (2007) Thermal stress and coral cover as drivers of coral disease outbreaks. *PLoS Biol* 5:e124
- Burnham KP, Anderson D (2002) Model selection and multi-model inference. Springer-Verlag, New York, NY
- Edmunds PJ (1991) Extent and effect of black band disease on a Caribbean reef. *Coral Reefs* 10:161–165
- Floros CD, Samways MJ, Armstrong B (2004) Taxonomic patterns of bleaching within a South African coral assemblage. *Biodivers Conserv* 13:1175–1194
- Gaston KJ (1994) Rarity. Population and community biology, Ser 13. Chapman & Hall, London
- Gil-Agudelo DL, Garzón-Ferreira J (2001) Spatial and seasonal variation of dark spots disease in coral communities of the Santa Marta area (Columbian Caribbean). *Bull Mar Sci* 69:619–629
- Gleason DF, Wellington GM (1993) Ultraviolet radiation and coral bleaching. *Nature* 365:836–838
- Gochfeld DJ, Olson JB, Slattery M (2006) Colony versus population variation in susceptibility and resistance to dark spot syndrome in the Caribbean coral *Siderastrea siderea*. *Dis Aquat Org* 69:53–65
- Goreau T, McClanahan T, Hayes R, Strong AE (2000) Conservation of coral reefs after the 1998 global bleaching event. *Conserv Biol* 14:5–15
- Harvell D, Kim K, Quirolo C, Weir J, Smith G (2001) Coral bleaching and disease: contributors to 1998 mass mortality in *Briareum asbestinum* (Octocorallia, Gorgonacea). *Hydrobiologia* 460:97–104
- Harvell CD, Mitchell CE, Ward JR, Altizer S, Dobson AP, Ostfeld RS, Samuel MD (2002) Climate warming and disease risks for terrestrial and marine biota. *Science* 296: 2158–2162
- Harvell D, Jordan-Dahlgren E, Merkel S, Rosenberg E and others (2007) Coral disease, environmental drivers, and the balance between coral and microbial associates. *Oceanography (Wash DC)* 20:173–195
- Hoegh-Guldberg O (1999) Climate change, coral bleaching and the future of the world's coral reefs. *Mar Freshw Res* 50:839–866
- Howe RW, Regal RR, Niemi GJ, Danz NP, Hanowski JM (2007) A probability-based indicator of ecological condition. *Ecol Indic* 7:793–806
- Hughes TP, Baird AH, Bellwood DR, Card M and others (2003) Climate change, human impacts, and the resilience of coral reefs. *Science* 301:929–933
- Hurlbert SH (1984) Pseudoreplication and the design of ecological field experiments. *Ecol Monogr* 54:187–211
- Jones RJ, Bowyer J, Hoegh-Guldberg O, Blackall LL (2004) Dynamics of a temperature-related coral disease outbreak. *Mar Ecol Prog Ser* 281:63–77
- Kaczmarek LT, Draud M, Williams EH (2005) Is there a relationship between proximity to sewage effluent and the prevalence of coral disease? *Caribb J Sci* 41:124–137
- Kuta KG, Richardson LL (1996) Abundance and distribution of black band disease on coral reefs in the northern Florida Keys. *Coral Reefs* 15:219–223
- Kuta KG, Richardson LL (2002) Ecological aspects of black band disease of corals: relationships between disease incidence and environmental factors. *Coral Reefs* 21:393–398
- Lafferty KD, Porter JW, Ford SE (2004) Are diseases increasing in the ocean? *Annu Rev Ecol Syst* 35:31–54
- Lips KR, Brem F, Brenes R, Reeve JD and others (2006) Emerging infectious disease and the loss of biodiversity in a neotropical amphibian community. *Proc Natl Acad Sci USA* 103:3165–3170
- Liu G, Strong AE, Skirving W, Arzayus LF (2005) Overview of NOAA coral reef watch program's near-real time satellite

- global coral bleaching monitoring activities. Proc 10th Int Coral Reef Symp 1:1783–1793
- Loya Y, Sakai K, Yamazato K, Nakano Y, Sambali H, Van Woesik R (2001) Coral bleaching: the winners and losers. *Ecol Lett* 4:122–131
- Maina J, Venus V, McClanahan TR, Ateweberhan M (2008) Modelling susceptibility of coral reefs to environmental stress using remote sensing data and GIS models. *Ecol Modell* 212:180–199
- Manel S, Williams HC, Ormerod SJ (2002) Evaluating presence–absence models in ecology: the need to account for prevalence. *J Appl Ecol* 38:921–931
- Manzello DP, Brandt M, Smith TB, Lirman D, Hendee JC, Nemeth RS (2007) Hurricanes benefit bleached corals. *Proc Natl Acad Sci USA* 104:12035–12039
- Marshall PA, Baird AH (2000) Bleaching of corals on the Great Barrier Reef: differential susceptibilities among taxa. *Coral Reefs* 19:155–163
- McClanahan TR, Muthiga NA, Mangi S (2001) Coral and algal changes after the 1998 coral bleaching: interaction with reef management and herbivores on Kenyan reefs. *Coral Reefs* 19:380–391
- McClanahan TR, Ateweberhan M, Graham NAJ, Wilson SK, Ruiz Sebastián C, Guillaume MMM, Bruggemann JH (2007) Western Indian Ocean coral communities: bleaching responses and susceptibility to extinction. *Mar Ecol Prog Ser* 337:1–13
- McCullagh P, Nelder JA (1989) Generalized linear models, 2nd edn. Chapman & Hall, London
- McField MD (1999) Coral response during and after mass bleaching in Belize. *Bull Mar Sci* 64:155–172
- Micheli F, Halpern BS, Botsford LW, Warner RR (2004) Trajectories and correlates of community change in no-take marine reserves. *Ecol Appl* 14:1709–1723
- Miller J, Waara R, Muller E, Rogers C (2006) Coral bleaching and disease combine to cause extensive mortality on reefs in US Virgin Islands. *Coral Reefs* 25:418
- Muller EM, Rogers CS, Spitzack AS, van Woesik R (2007) Bleaching increases likelihood of disease on *Acropora palmata* (Lamarck) in Hawksnest Bay, St. John, US Virgin Islands. *Coral Reefs* 27:191–195
- Nijboer RC, Schmidt-Kloiber A (2004) The effect of excluding taxa with low abundances or taxa with small distribution ranges on ecological assessment. *Hydrobiologia* 516:347–363
- Nugues MM (2002) Impact of a coral disease outbreak on coral communities in St. Lucia: What and how much has been lost? *Mar Ecol Prog Ser* 229:61–71
- Page C, Willis B (2006) Distribution, host range and large-scale spatial variability in black band disease prevalence on the Great Barrier Reef, Australia. *Dis Aquat Org* 69:41–51
- Page CA, Baker DM, Harvell CD, Golbuu Y and others (2009) Influence of marine reserves on coral disease prevalence. *Dis Aquat Org* 87:135–150
- Patterson KL, Porter JW, Ritchie KB, Polson SW and others (2002) The etiology of white patch, a lethal disease of the Caribbean elkhorn coral *Acropora palmata*. *Proc Natl Acad Sci USA* 99:8725–8730
- Porter JW, Dustan P, Jaap WC, Patterson KL and others (2001) Patterns of spread of coral disease in the Florida Keys. *Hydrobiologia* 460:1–24
- Quinn GP, Keough MJ (2002) Experimental design and data analysis for biologists. Cambridge University Press, New York, NY
- Santavy DL, Mueller E, Peters EC, MacLaughlin L, Porter JW, Patterson KL, Campbell J (2001) Quantitative assessment of coral diseases in the Florida Keys: strategy and methodology. *Hydrobiologia* 460:39–52
- Santavy DL, Campbell J, Quarles RL, Patrick JM and others (2006) The epizootiology of coral diseases in South Florida. EPA/600/R-05/146, US Environmental Protection Agency, Gulf Ecology Division, Gulf Breeze, FL
- Selig ER, Harvell CD, Bruno JF, Willis BL, Page CA, Casey KS, Sweatman H (2006) Analyzing the relationship between ocean temperature anomalies and coral disease outbreaks at broad spatial scales. In: Phinney JT, Strong A, Skirving W, Kleypas J, Hoegh-Guldberg O (eds) Coral reefs and climate change: science and management. AGU Monograph Series, Coastal and Estuarine Studies, Vol 61, Washington, DC, p 111–128
- Somerfield PJ, Jaap WC, Clarke KR, Callahan M and others (2008) Changes in coral reef communities among the Florida Keys, 1996–2003. *Coral Reefs* 27:951–965
- Sutherland KP, Porter JW, Torres C (2004) Disease and immunity in Caribbean and Indo-Pacific zooxanthellate corals. *Mar Ecol Prog Ser* 266:273–302
- Tilman D (1999) The ecological consequences of changes in biodiversity: a search for general principles. *Ecology* 80:1455–1474
- Voss JD, Richardson LL (2006) Coral diseases near Lee Stocking Island, Bahamas: patterns and potential drivers. *Dis Aquat Org* 69:33–40
- Ward JR, Rypien KL, Bruno JF, Harvell CD and others (2006) Coral diversity and disease in Mexico. *Dis Aquat Org* 69:23–31
- Weil E, Cróquer A (2009) Spatial variability in distribution and prevalence of Caribbean scleractinian coral and octocoral diseases. I. Community-level analysis. *Dis Aquat Org* 83:195–208
- Weil E, Smith GW, Gil-Agudelo DL (2006) Status and progress in coral reef disease research. *Dis Aquat Org* 69:1–7
- West JM, Salm RV (2003) Resistance and resilience to coral bleaching: implications for coral reef conservation and management. *Conserv Biol* 17:956–967
- Wilkinson C (2008) Status of coral reefs of the world. Australian Institute of Marine Science, Townsville
- Williams GJ, Aeby GS, Cowie RO, Davy SK (2010) Predictive modeling of coral disease distribution within a reef system. *PLoS ONE* 5:e9264
- Willis BL, Page CA, Dinsdale EA (2004) Coral disease in the Great Barrier Reef. In: Rosenberg E, Loya Y (eds) Coral health and disease. Springer-Verlag, Berlin, p 69–104
- Wooldridge S, Done T (2004) Learning to predict large-scale coral bleaching from past events: a Bayesian approach using remotely sensed data, *in-situ* data, and environmental proxies. *Coral Reefs* 23:96–108
- Yee SH, Santavy DL, Barron MG (2008) Comparing environmental influences on coral bleaching across and within species using clustered binomial regression. *Ecol Model* 218:162–174
- Zvuloni A, Artzy-Randrup Y, Stone L, Kramarsky-Winter E, Barkan R, Loya Y (2009) Spatio-temporal transmission patterns of black-band disease in a coral community. *PLoS ONE* 4:e4993