THRESHOLD EFFECTS IN META-ANALYSES WITH APPLICATION TO BENEFIT TRANSFER FOR CORAL REEF VALUATION

LUKE FITZPATRICK, CHRISTOPHER F. PARMETER, AND JUAN AGAR

Abstract. Policymakers and advocates often use benefit transfers to estimate the economic value of environmental amenities when primary valuation studies are infeasible. Benefit transfers based on meta-analyses, which synthesize site and methodological characteristics from valuation studies of similar underlying amenities, generally outperform traditional site-to-site transfers. We build on earlier meta-analyses of willingness-to-pay for tropical coral reef recreation by introducing a meta-regression model with threshold effects, with a goal of increasing transfer reliability. We estimate a threshold in coral reef quality and find that increases in live coral cover have a large impact on individuals' WTP for recreation at degraded coral reefs. Relaxing the assumption of users' constant valuation across the distribution of this characteristic improves the performance of coral reef benefit transfers in some instances: tests of convergent validity reveal that including the threshold effect reduces the mean transfer error and the interquartile range of transfer errors in 5 out of 8 tests.

1. Introduction

A common problem facing government agencies, policy makers, and advocates is how to properly value an amenity that benefits the public, but is not traded in markets. Examples include improvements in air quality, decreases in violent crime rates, and reductions in health

---

Key words and phrases. meta-regression model, convergent validity, threshold, contingent valuation.
Luke Fitzpatrick, Department of Economics, Ohio University; e-mail: fitzpatl@ohio.edu.
Christopher F. Parmeter, Corresponding Author, Department of Economics, University of Miami; e-mail: cparmeter@bus.miami.edu.
Juan Agar, NOAA, National Marine Fisheries Service; e-mail: juan.agar@noaa.gov.
We would like to thank two anonymous referees, Peter Edwards and Robert Johnston for useful comments and suggestions. We also gratefully acknowledge Robert Johnston for providing the metadata. The support of NOAA's Office of Science and Technology is gratefully acknowledged. The views and opinions expressed or implied in this article are those of the authors and do not necessarily reflect the position of the National Marine Fisheries Service, NOAA.

risks. In some instances, cost and/or time constraints prevent analysts from conducting a primary valuation study. When a site-specific study is infeasible, those who conduct valuation work turn to *benefit transfers*. A benefit transfer is defined as the use of results from extant primary research to predict welfare estimates for amenities when primary valuation estimates are infeasible (Johnston et al. 2015). In practice, analysts would transfer the value associated with an amenity from an already-studied site (the study site) to estimate the unknown value of the same amenity at the unstudied site (the policy site).

Frequently, there is no sufficiently comparable study site which analysts can use to estimate the value at the policy site (Parsons & Thur 2008); this is one cause for concern in benefit transfer (Johnston & Rosenberger 2010). In these situations, researchers can use meta-analysis for benefit transfer. As a tool for benefit transfer, meta-analysis integrates findings from multiple primary studies of a common amenity, and helps the analyst understand how values depend on site characteristics and valuation methodology. For example, researchers recently combined a meta-analysis of willingness-to-pay for coral reef recreation with a visitation model to estimate the value of foregone recreation that would occur under different ecological scenarios (Brander et al. 2015).[1] An attractive feature of meta-analysis is the ability to control for factual and methodological heterogeneity that might plague traditional benefit transfers (Nelson & Kennedy 2009, Rolfe, Brouwer & Johnston 2015).

Many meta-analyses of environmental valuation make use of meta-regression (Smith & Huang 1995, Shrestha & Loomis 2001, Van Houtven, Powers & Pattanayak 2007, Braden, Feng & Won 2011), in which the dependent variable is an estimate of welfare, typically willingness-to-pay (WTP), drawn from primary studies that analyze similar underlying amenities. In a metadataset, a single study could provide multiple observations of WTP. The independent variables in a meta-regression model (MRM) characterize the site attributes and methodological approaches that could explain variation in WTP for the underlying amenity

[1]Recent reviews on the use of meta-analyses in the benefit transfer and valuation literatures include Johnston & Rosenberger (2010), Boyle, Kaul & Parmeter (2015), and Johnston et al. (2015); more critical reviews are Smith & Pattanayak (2002) and Nelson & Kennedy (2009).

(Nelson & Kennedy 2009). The researcher then combines the coefficient estimates from the MRM with the policy site characteristics to estimate welfare associated with the policy site.[2]

The standard approach in meta-analysis is to employ a linear MRM (Nelson & Kennedy 2009), and a linear-in- parameters relationship between site characteristics and WTP. However, the recent work of Kaul et al. (2013) discusses the possibility of nonlinear effects in a MRM. One way to model non-linear impacts of site characteristics on WTP is to include the logged values of continuous right-hand side variables (Johnston et al. 2005, Brander et al. 2007, Londoño & Johnston 2012).

We depart from earlier meta-analyses by modeling a discontinuous relationship between WTP and (some) site characteristics within a threshold model. In a threshold model, consumers' marginal WTP (MWTP) for improvements in a particular characteristic could vary depending on the level of that specific characteristic. For example, in a MRM analyzing estimates of WTP for improved air quality, populations may demonstrate large MWTP for improvements when concentrations of pollutants are above some level, and MWTP of zero when pollutants are below that level.

Of concern when using benefit transfer is the reliability of the transfer approach (Boyle et al. 2009, Londoño & Johnston 2012). To quantify reliability, researchers compute benefit transfer errors: the difference between a benefit transfer's estimate of WTP and the actual WTP (Rosenberger & Stanley 2006). The percentage transfer error between the transfer estimate (V_T) and known estimate (V_P), calculated as:

$$PTE = \frac{V_T - V_P}{V_P} \times 100 \tag{1}$$

is a common measure of transfer reliability; note that transfer errors are typically measured in terms of absolute value percentages.[3] Small transfer errors imply reliable benefit transfer,

[2]When using meta-analysis for benefit transfer, the researcher must choose values for methodological variables when estimating welfare for the study site. The most common approach is to use the sample means for each methodological variable within the metadata (Moeltner, Boyle & Paterson 2007).

[3]Here we employ the notation of Rosenberger (2015).

and benefit transfers generated from meta-analyses tend to be more reliable than those from alternative methods (Johnston & Rosenberger 2010).

Unfortunately, in most policy applications, primary study estimates are unavailable, and so reliability is unknown (i.e., benefit transfer is required only when high quality, site specific primary studies are unavailable). However, researchers can use test cases where primary study estimates are available to gain insights into the sort of errors that might be expected in actual transfers (Johnston & Rosenberger 2010, Rosenberger & Stanley 2006). In such cases, reliability is quantified using convergent validity tests that compare a transferred welfare estimate to an available primary study estimate for a particular site.

Among other things, reliability hinges on a properly specified MRM. To see why, suppose a researcher use a MRM to estimate the unknown value of an unstudied policy site, V_P. To construct the estimate V_T, the researcher uses the set of coefficients for site characteristics and methodological approaches from the MRM, β_S, and the observed policy site variables X_P: $V_T\left(\beta_S, X_P\right)$. If the coefficient estimates from the MRM are biased, then the transfer estimate will also be biased, compromising reliability.

Our aim here is to determine if accounting for threshold effects in biophysical characteristics in a MRM of coral reef values can improve benefit transfer reliability. We find that threshold effects can improve the reliability of meta-analysis based benefit transfers, measured either by mean transfer error or the interquartile range from the full distribution of transfer errors. Moreover, even if there does not exist a threshold, the ability to test for a threshold allows one to consider the homogeneity of a sample of valuations across disparate coral reefs, which is one criteria that Boyle et al. (2009) listed for performing valid benefit transfers. We point out here that threshold estimation is not as simple as sample splitting; sample splitting implies that the value of the threshold is known to the analyst, which is typically not true in applied valuation.

This work builds on the suggestion of Kaul et al. (2013) to model nonlinearity in MRMs. Additionally, we follow the seminal work of Brander et al. (2007), Londoño & Johnston

(2012) and Brander et al. (2015) and use meta-analysis to study coral reef valuation. We estimate a MRM with threshold effects on metadata comprised of primary studies that estimate the recreational value of coral reefs. To capture discontinuous effects of biophysical characteristics on WTP for reef recreation, we propose a threshold effect model that incorporates a discontinuity in the relationship between percentage of live coral cover and WTP for recreation. To our knowledge, this approach is new for both estimating a MRM, and for conducting a benefit transfer. We follow the work of Londoño & Johnston (2012), who used a more refined valuation dataset than that of Brander et al. (2007), and enhance it with additional valuation studies. We then perform convergent validity tests (leave-one-observation-out and leave-one-study-out cross validity tests) across both methods to gain insights into potential improvements in transfer reliability. We find that accounting for the threshold leads to smaller mean transfer error and error variance when considering the pooled data from our expanded sample. Another interesting result stemming from the use of threshold MRMs is that the interquartile range of transfer errors is reduced in the majority of our convergent validity tests. Further results are described in section 5.

The remainder of the paper is structured as follows. Section 2 details common meta regression methodology, and describes our inclusion of a threshold in the MRM. Section 3 describes the growing interest in the economic value of coral reef ecosystems. Section 4 provides a brief description of the coral reef valuation dataset of Londoño & Johnston (2012), and of our additions to this dataset. Section 5 details our estimates of the MRMs and presents the convergent validity tests. Section 6 concludes.

2. Methodology

Our approach builds on the standard multi-level MRM, which is recommended for meta-analysis of this type of metadata (Nelson & Kennedy 2009). Multi-level models allow for within-study correlation across the different observations of estimated of welfare (Bateman & Jones 2003, Johnston et al. 2005). The benchmark multi-level MRM which estimates the impact of methodological ($\bar{x}_{js}$) and site ($\bar{z}_{js}$) characteristics on reported measures of welfare $\bar{y}_{js}$ for valuation estimate s from study j is:

$$\bar{y}_{js} = \bar{x}_{js}\beta + \bar{z}_{js}\delta + u_s + \varepsilon_{js}. \tag{2}$$

Here, u_s captures systemic study level effects and ε_{js} is a standard *iid* observation specific error with constant variance. Clustering by study is standard in the meta regression literature, though alternative clustering strategies may be deployed (clustering by author or region, for example). β and δ are vectors of parameters to be estimated to discern the impact that research methods and environmental characteristics have on estimated welfare. We assume that $E(u_s) = 0$ and $Var(u_s) = \sigma_u^2$.

While the standard multi-level MRM is capable of explaining empirical variation in estimated welfare, if threshold effects existed in any of the site characteristics, then this estimation approach would be inconsistent. For example, in the context of our application, suppose that WTP per visitor day for recreation at a coral reef depended on the size of the reef. It is likely that for small reefs the perceived value by individuals is different than for large reefs. Compare the size of the Great Barrier Reef in Australia to the Biscayne National Park in the Florida Keys. They differ in size by almost 500 to 1. In these instances, it is conceivable that the meta regression function itself actually differs by the size of the reef.

One way to account for a nonlinear relationship between WTP and site characteristics is to use their logged values in the MRM. For example, in Londoño & Johnston (2012), their MRM includes some reef characteristics in logarithmic form. Other approaches include nonparametric estimation, as in Kaul et al. (2013), or including quadratic terms. Here, we propose an approach new to meta-analysis: threshold models.

Assume that within the MRM a threshold exists in a single variable $\bar{z}$, and that the effects of the other variables could differ across the threshold. In this case, the threshold MRM is:

$$\bar{y}_{js} = \bar{x}_{js}\beta_1 + \delta_1 \bar{z}_{js} + u_s + \varepsilon_{js}, \quad \bar{z}_{js} \leq \gamma \tag{3}$$

$$\bar{y}_{js} = \bar{x}_{js}\beta_2 + \delta_2 \bar{z}_{js} + u_s + \varepsilon_{js}, \quad \bar{z}_{js} > \gamma. \tag{4}$$

Hansen (2000) uses this specification to study cross-country growth rates. This approach allows for different non-threshold variable coefficients on each side of the threshold.[4] An alternative approach builds the discontinuity into the effect that the threshold variables has, while holding the effect of other variables constant. In this case the threshold MRM is:

$$\bar{y}_{js} = \bar{x}_{js}\beta + \delta_1 \bar{z}_{js} + \delta_2 \bar{z}_{js} 1\{\bar{z}_{js} > \gamma\} + u_s + \varepsilon_{js}. \tag{5}$$

Here $\bar{z}_{js}$ has a constant effect on estimated welfare, measured through δ_1, and if a threshold exists, the additional threshold effect is determined through δ_2. Further, the effect of the remaining moderator variables is captured through β, and is constant regardless of γ.

Threshold models are a common way to capture the presence of a discontinuity in applied economic and econometric research. A typical strategy used to estimate a threshold is profiling (see Hansen 2000), where a range of threshold levels are selected *a priori* and the regression model is estimated by splitting the sample based on the selected threshold level, as in the case of (3) and (4), or, for (5), constructing a variable $1\{\bar{z}_{js} > \gamma\}$ for each level of the threshold and estimating a single multi-level model. The estimated threshold level leads to the optimal value of some criterion, most commonly the minimum residual sum of squared errors. For the multi-level model, rather than using residual sum of squares, we can instead rely on the total likelihood. See Appendix A for more details on our implementation of the profiling strategy used to estimate the threshold MRM.

[4] Van Houtven et al. (2007) use a similar idea to test for metadata consistency by splitting their sample metadata based on CV and TC studies. In this case there is no unknown threshold.

We estimate and present results for both threshold models, but prefer the second approach for three reasons. First, it is not obvious whether or how research methods' effect on valuation estimates should change based on a threshold in another variable. Intuitively, it seems unlikely that $\beta_1 \neq \beta_2$ based on a biophysical characteristic. Second, although the sample splitting threshold model may fit well in sample, out of sample prediction could be substantially degraded. This is especially salient since one of the goals of a using a MRM for benefit transfer is to enhance out of sample prediction. Finally, given the sample sizes common in the meta-analysis literature, many degrees of freedom would be lost by specifying the sample splitting MRM in 4 and 5.

3. Tropical Coral Reefs

Among the most biologically diverse ecosystems in existence, coral reefs supply a large number of ecosystem services: they act as marine recreation sites, foster recreational and commercial fishing opportunities, and buffer coasts during storms (Brander et al. 2007, Londoño & Johnston 2012). In recent decades, 33-50% of all reefs have perished (Hoegh-Guldberg et al. 2015). This loss of habitat places a quarter of marine species at risk. Locally, unsustainable fishing practices, coastal development, and water pollution can degrade entire coral ecosystems. Globally, warming waters and ocean acidification threaten entire species. Widespread coral bleaching - their biological response to environmental stress - highlights the growing amount of strain on reef ecosystems. In advance of the 2015 U.N. Climate Change Conference, the International Society for Reef Studies advocated a long-term atmospheric carbon dioxide concentration of 350 parts per million to prevent further coral losses (Hoegh-Guldberg et al. 2015).

Quantifying the economic benefits that reefs provide is imperative for designing efficient conservation policy (Brander et al. 2007). Although there is growing quantitative research into these benefits, estimates for specific sites are often unavailable. This stems from logistical constraints of conducting research in areas where coral reefs are located (Brander et al. 2007).

As a result, estimating the value of coral reefs is a useful application of benefit transfer. Reliably transferring coral reef valuations is non-trivial. As noted in Brander et al. (2007) and Londoño & Johnston (2012), there exist few high-quality studies of coral reef valuation, and studies diverge in methodology and welfare measures considered.

In light of these challenges, it is unsurprising that there exist only three published meta-analyses of the value of coral reefs. Brander et al. (2007) use 33 valuation studies to form a meta-dataset with 73 observations. Their welfare measure is WTP per visit at coral reefs, and the mean transfer error in their leave-one-observation convergent validity test is 186%.[5] Brander et al. (2007, p. 215) acknowledge that their convergent validity tests reveal that the results of their meta-analysis are "unlikely to be acceptable in most policy-making scenarios."

Such large errors suggest room for improvement in reliably transferring benefit estimates of coral reefs with meta-analysis. Londoño & Johnston (2012) show that increasing commodity consistency and controlling for site characteristics lead to substantial improvements in reliability. They only include studies that are explicitly linked to recreation, and exclude studies that estimate values per unit area. These enhancements lead to significant improvements in reliability: the mean transfer error from leave-one-observation-out convergent validity tests are 98.8% for their full sample, and 93.8% for their sample consisting only of contingent valuation studies. More recently, Brander et al. (2015) estimate a MRM using an expanded version of the metadata used in Brander et al. (2007), and combine it with a reef visitation model to estimate the foregone value of reef recreation resulting from declining reefs in Southeast Asia.[6]

[5]Woodward & Wui (2001) perform a meta-analysis of another heterogeneous commodity (wetlands) and report similarly large errors.

[6]Although they expand upon the metadata of Brander et al. (2007), Brander et al. (2015) also pare down their data so that their studies are welfare consistent, resulting in a smaller meta dataset than that used in Brander et al. (2007).

4. Data

The majority of our data come directly from Londoño & Johnston (2012). They use metadata drawn from studies that estimate individual users' WTP per visitor per day of coral reef recreation. Coral reef recreation includes snorkeling, scuba diving, and glass bottom boat tours.

Londoño & Johnston (2012) construct a refined and expanded meta-dataset compared to Brander et al. (2007) by excluding studies that did not comply with conditions for welfare and commodity consistency: they use only observations linked explicitly to WTP per visitor day for reef recreation. This amounts to removing observations with WTP linked to other reef-related services, WTP for conservation, or values linked to visiting an area close to a reef without recreating there.

To ensure meta data consistency with Londoño & Johnston (2012), we deflate WTP values from our new studies to year 2000 international dollars. Variables fall into two categories: methodological characteristics ($\bar{x}_{js}$ in (2)) and site-specific attributes ($\bar{z}_{js}$ in (2)). Methodological characteristics for each observation include: method of elicitation, type of payment vehicle, study sample size, publication status, and sampling method. Site attributes include: size of the reef, percentage of reef covered by live coral (reef quality), type of reef (artificial vs. man made), dummy variables for marine protected area (MPA) status, and whether or not the reef is used for snorkeling and scuba diving. Additionally, a dummy variable for studies conducted on East African reefs is also included. This captures the fact that, in the nascent reef tourism economy of Africa, the average visitor tends to be less wealthy than for other reefs worldwide, and the most common type of reef tourism package is at the low end, leading to a different type of experience.

Table 1 describes the independent variables included in our MRMs and presents variable means. These variables are identical to those used in Londoño & Johnston (2012) and represent attributes with the potential to influence WTP. For a full description of the variables used, we refer the reader to Londoño & Johnston (2012).

Table 1. Description of Variables Used for Meta-Analysis.

Variable	Description	Mean
$\ln(WTP)$	Willingness-to-pay per person per day for tropical coral reef recreation. All values are deflated to year 2000 purchasing power parity U.S. dollars.	37.5
Discrete choice	Binary variable equal to one for observations from contingent valuation (CV) surveys using choice-based methods. These include choice experiments and dichotomous choice. The default includes open-ended or iterative bidding CV methods.	0.14
Payment card	Binary variable equal to one for observations from CV surveys using payment card methods. The default includes open-ended or iterative bidding CV methods.	0.35
Travel cost	Binary variable equal to one for observations from travel cost studies. The default includes open-ended or iterative bidding contingent valuation methods.	0.16
Trip cost	Binary variable equal to one for observations from studies that use additional trip cost for a payment vehicle. The default includes any payment vehicle not described as trip cost or donation based.	0.49
Donation	Binary variable equal to one for observations from studies that use a donation for a payment vehicle. The default includes any payment vehicle not described as trip cost or donation based.	0.12
ln (sample size)	Logged value of study sample size.	343.1
Onsite study	Binary variable equal to one for observations using onsite surveys to collect data. The default includes any study not administered on-site.	0.65
Publication	Binary variable equal to one for observations taken from studies published in peer-reviewed journals.	0.47
East Africa	Binary variable equal to one for observation taken from studies performed for reefs in East Africa.	0.13
ln (reef area)	Logged value of reef area (in square kilometers.)	21,267
MPA status	Binary variable equal to one for observations from studies of reefs in designated or proposed marine protected areas.	0.87
Snorkel/dive	Binary variable equal to one for observations taken from studies of specifically snorkeling/diving related reef values.	0.31
Reef quality	Percentage of live coral cover at the reef.	0.32
Natural reef	Binary variable equal to one for observations from natural (as opposed to man-made) coral reefs.	0.90

Means of variables included in logarithmic form are presented in levels here.

To update the metadata of Londoño & Johnston (2012), we searched (on July 7, 2015) the academic website Research Papers in Economics for papers written since 2007 (the terminal year of all studies included in Londoño & Johnston (2012)) featuring the word "coral" for observations that could potentially be added to the metadata.[7] Only studies that estimate use value for tropical reefs were taken for potential inclusion. Three studies met the conditions for commodity and welfare consistency laid out by Londoño & Johnston (2012). Following the guidelines of Stanley et al. (2013), Fitzpatrick and Parmeter separately coded welfare measures, site characteristics, and methodological variables before comparing findings, clarifying any differences, and then finally adding them to the meta data.

This produced an additional six observations: four from Ransom & Mangi (2010) and one each from Kragt, Roebeling & Ruijs (2009) and Phewmau (2013). Five of the six new observations were from contingent valuation studies and one observation was from a travel cost study. This resulted in a total of 91 observations, with 15 observations coming from travel cost valuation methods and 76 coming from CV methods. Further, Ransom & Mangi's (2010) study was conducted in the Mombasa Marine National Park and Reserve, which gives us an additional study, with four observations, from East Africa, nearly doubling the total number that appeared in Londoño & Johnston (2012). We rely on several sources for our new estimates of percent live coral cover. For the observations from Kenya, we use Wilkinson (2008), which was also used by Londoño & Johnston (2012). The estimates of coral cover for the Similan Islands and the Cairns Recreational Area come from Brown et al. (2011) and Death et al. (2012), respectively.

5. Results

5.1. **MRM Estimates.**

5.1.1. *Benchmark MRM.* We estimate four different baseline MRMs, which are summarized in Table 2. Columns (1) and (2) closely reproduce the estimates of Londoño & Johnston

[7]The most recent meta-analysis of coral reef valuation, Brander et al. (2015), only included studies up to 2012, and does not include the three studies that we use here.

(2012) across the full sample and the subsample of CV estimates.[8] Columns (3) and (4) estimate the same MRM, but use our expanded metadata. There are a few points to mention from Table 2. Adding six observations to the original metadata induced changes in some estimates' magnitude and statistical significance for both pooled and CV specific models. Note that each coefficient estimate represents the change in logged WTP for coral reef recreation following a marginal (for logged variables) or one-unit (for binary variables) increase in each characteristic.

Our results largely mirror earlier meta-analyses of recreational values of coral reefs. With respect to methodological variables, our results are similar to Londoño & Johnston (2012). We find that among contingent valuation studies, dichotomous choice methods and payment card elicitation mechanisms are associated with smaller estimates of WTP, relative to open-ended questionnaires. This result is robust for both the pooled and CV-only samples. We also find that, similar to Johnston, Besedin & Ranson (2006), donation based methods are associated with lower WTP, and that revealed preference methods generate larger estimates of WTP (Rosenberger & Loomis 2000, Moeltner et al. 2007). We find some evidence for publication bias: for the combined sample, studies published in peer-reviewed journals feature WTP nearly 60% higher than unpublished studies, *ceteris paribus*. With respect to site characteristics, we find that reefs in MPAs and reefs used for snorkeling and diving are associated with higher WTP. We also find that reefs with higher percentage of live coral cover generate higher WTP, as do natural (as opposed to man-made) reefs. In one divergence from the findings of Londoño & Johnston (2012), we find that larger reefs are associated with higher WTP.

[8]Despite estimating the same model on the same dataset, our estimates do not exactly match those of Londoño & Johnston (2012). We believe this results from the fact that the model has no closed form solution, and different software packages (we use R and Londoño & Johnston (2012) use SAS) that use different solution algorithms and employ different convergence criteria could lead to slightly divergent results. Importantly, we reproduce the *qualitative* results of Londoño & Johnston (2012).

Table 2. Meta-Regression Estimates: Combined and Contingent Valuation Studies. Values in parentheses are standard errors.

	Londoño & Johnston (2012) data		Full data	
	(1)	(2)	(3)	(4)
Discrete choice	-0.893^{***}	-0.577^{*}	-0.841^{***}	-0.432
	(0.302)	(0.301)	(0.323)	(0.353)
Payment card	-1.115^{***}	-0.893^{***}	-0.978^{***}	-0.647^{*}
	(0.261)	(0.275)	(0.281)	(0.340)
Travel cost	0.002		0.357	
	(0.354)		(0.361)	
Trip cost	1.005^{***}	0.629^{**}	0.578^{**}	0.425
	(0.277)	(0.306)	(0.275)	(0.377)
Donation	-0.970^{**}	-0.684^{*}	-1.050^{**}	-0.504
	(0.414)	(0.415)	(0.440)	(0.458)
ln(sample size)	0.169	0.247^{**}	0.259^{**}	0.252^{**}
	(0.110)	(0.100)	(0.113)	(0.101)
Onsite study	-0.278	-0.434	-0.123	-0.142
	(0.282)	(0.273)	(0.297)	(0.305)
Publication	0.616^{***}	0.313	0.566^{**}	0.312
	(0.208)	(0.241)	(0.225)	(0.300)
East Africa	-0.867^{**}	0.004	-1.018^{***}	-0.293
	(0.366)	(0.452)	(0.370)	(0.493)
ln(reef area)	-0.003	-0.009	0.037	0.071^{*}
	(0.037)	(0.043)	(0.037)	(0.043)
MPA status	0.600^{*}	0.819^{**}	0.518	0.694
	(0.342)	(0.366)	(0.370)	(0.426)
Snorkel/dive	0.203	0.309	0.480^{**}	0.505^{*}
	(0.244)	(0.276)	(0.239)	(0.297)
Reef quality	2.031^{***}	1.867^{***}	2.277^{***}	1.883^{***}
	(0.470)	(0.567)	(0.491)	(0.602)
Natural reef	0.643^{*}	0.592^{**}	0.539	0.599^{**}
	(0.331)	(0.263)	(0.357)	(0.264)
Constant	1.280^{***}	1.287^{***}	1.347^{***}	1.296^{***}
	(0.333)	(0.309)	(0.361)	(0.336)
Observations	85	71	91	76
Log Likelihood	-94.566	-67.736	-108.799	-78.356
AIC	223.132	167.473	251.599	188.713
BIC	264.657	203.676	294.283	226.004

Note: *p<0.1; **p<0.05; ***p<0.01

5.1.2. *Single Variable Threshold in Reef Quality.* Next, we estimate the single variable threshold MRM with a threshold in reef quality, which is the proportion of reef surface area covered by live stony coral. We might expect a threshold in reef quality to exist given that declines in coral cover have not only been shown to decrease the abundance of reef-associated fish species, but also to threaten the structural integrity of reefs, making them more susceptible to erosion and decomposition (Bell & Galvin 1984, Coker, Wilson & Pratchett 2014).

Table 3 presents results from a model that contains a threshold in reef quality as presented in (5) for the same subsamples as in Table 2. As hypothesized, an increase in reef quality at a site with low live coral cover generates a larger increase in WTP than at sites with more live coral cover. In each of the four models estimated, the coefficient on reef quality interacted with low live coral cover is over 10 times greater than the coefficient on non-interacted reef quality. All four of the models estimate a threshold in reef quality of 0.12, which implies 28 observations from low-quality reefs in both datasets. Given the noisiness associated with measuring reef quality, we urge caution in reading directly into this amount of coral cover as the sample sizes here are quite small. In terms of model fit, the threshold model unambiguously improves upon the baseline model, based on both the Akaike and Bayesian information criteria.

One result that appears to be more robust across the results from the threshold models is the type of reef. In Table 2, with the full set of 91 observations (see column (3)), the estimated effect on the type of reef was positive but statistically insignificant, a departure from the other three models. However, here we see that the effect is remarkably consistent across all four models, always having a positive and statistically significant effect. This same type of phenomenon occurs with the snorkel/dive dummy variable. Accounting for a threshold in reef quality leads to an estimate on snorkel/dive that is statistically significant

Table 3. Single Variable Threshold (Reef Quality) Meta-Regression Model Estimates: All Studies.

	Londoño & Johnston (2012) Data		Full Data	
	(1)	(2)	(3)	(4)
Discrete Choice	−0.384	−0.219	−0.307	−0.097
	(0.306)	(0.287)	(0.328)	(0.356)
Payment Card	−0.510*	−0.491*	−0.351	−0.290
	(0.285)	(0.271)	(0.305)	(0.343)
Travel Cost	0.472		0.786**	
	(0.347)		(0.351)	
Trip Cost	0.721***	0.383	0.338	0.200
	(0.265)	(0.249)	(0.261)	(0.328)
Donation	−1.280***	−0.964**	−1.310***	−0.717*
	(0.389)	(0.383)	(0.412)	(0.430)
ln(Sample Size)	0.202**	0.224**	0.272***	0.250**
	(0.102)	(0.097)	(0.105)	(0.102)
Onsite Study	−0.090	−0.370	0.098	−0.086
	(0.263)	(0.246)	(0.280)	(0.286)
Publication	0.469**	0.245	0.429**	0.226
	(0.195)	(0.192)	(0.210)	(0.256)
East Africa	−0.443	0.466	−0.462	0.142
	(0.354)	(0.407)	(0.370)	(0.471)
ln(Reef Area)	−0.009	−0.010	0.034	0.062
	(0.034)	(0.036)	(0.034)	(0.039)
MPA Status	0.174	0.543*	0.051	0.492
	(0.333)	(0.329)	(0.362)	(0.400)
Snorkeling/Dive	0.379*	0.513**	0.703***	0.701**
	(0.229)	(0.241)	(0.228)	(0.274)
Reef Quality	3.375***	2.865***	3.600***	2.662***
	(0.551)	(0.645)	(0.564)	(0.663)
Natural Reef	0.754**	0.679**	0.666**	0.639**
	(0.306)	(0.268)	(0.332)	(0.275)
Reef Quality∗I(Reef Quality $\leq \gamma$)	12.252***	9.864***	13.068***	9.864**
	(3.114)	(3.255)	(3.315)	(4.210)
Constant	0.430	0.576	0.490	0.726*
	(0.375)	(0.372)	(0.398)	(0.402)
Observations	85	71	91	76
$\widehat{\gamma}$	0.12	0.12	0.12	0.12
Below/Above	28/57	28/43	28/63	28/48
Log Likelihood	−87.454	−63.478	−101.627	−76.276
Akaike Inf. Crit.	210.908	160.955	239.253	186.553
Bayesian Inf. Crit.	254.876	199.421	284.449	226.175

Note: *p<0.1; **p<0.05; ***p<0.01

and of roughly equal magnitude across all four models, whereas this was not the case when we omitted the threshold.[9]

The large estimate effects for the interaction term need to be interpreted carefully. Given that our estimated threshold is 12% live coral cover, the average WTP to improve coral cover from 0% to 12%, using the estimates in column (1) (the original Londoño & Johnston (2012) data), is $exp((3.375 + 12.252) * 0.12) - 1 = 5.522$, which amounts to 552%, in line with the mean WTP estimates reported in Londoño & Johnston (2012) to improve reef quality from 0% to 100%. The threshold estimates here suggest that low quality reefs have quite high valuations placed on them for small improvements in coral reef cover.

5.1.3. *Sample Splitting Threshold in Reef Quality.* For comparison we estimate a sample splitting threshold MRM for reef quality. Table 4 presents estimates from this model.

Comparing Table 4 and Table 2 reveals that there are several differences between the sample splitting model and the single variable threshold model. Sample splitting produces more favorable AIC and BIC scores, so that version of the MRM model is more favorable in terms of model selection, despite the increase in the number of parameters to be estimated. This suggests that the in-sample fit of the sample splitting MRM is better than the individual threshold MRM. However, some parameter estimates from this model are troubling. The estimated elasticity of WTP for reef quality is an order of magnitude larger than in the single variable threshold model. The value of 53.237 implies that to improve reef quality from 0% up to the threshold, which is 17%, average mean WTP is 852,000%, while to improve reef quality from the threshold level to 100%, the average WTP is 1,262%. Additionally, it is not clear why the methodological characteristics would vary depending upon reef quality, which the sample splitting model imposes.

[9]Appendix B presents estimates for a threshold in reef area, as opposed to reef quality. These models always produced AIC and BIC scores which favored modeling the MRM with a threshold in reef quality and for brevity we do not discuss these results directly in the text.

Table 4. Threshold Sample Splitting Meta-Regression Model (Reef Quality) Estimates: All Studies.

	Londoño & Johnston (2012) Data		Full Data	
	$\leq \widehat{\gamma}$	$> \widehat{\gamma}$	$\leq \widehat{\gamma}$	$> \widehat{\gamma}$
Discrete Choice	-3.722^{***}	-0.217	-3.722^{***}	-0.161
	(0.670)	(0.299)	(0.670)	(0.348)
Payment Card	0.486	-0.677^{**}	0.486	-0.349
	(0.970)	(0.307)	(0.970)	(0.353)
Travel Cost	1.889^{***}	-0.004	1.889^{***}	0.583
	(0.466)	(0.369)	(0.466)	(0.403)
Trip Cost	-0.988^{**}	0.902^{***}	-0.988^{**}	0.295
	(0.401)	(0.308)	(0.401)	(0.327)
Donation	0.301	-2.055^{***}	0.301	-2.056^{***}
	(0.428)	(0.446)	(0.428)	(0.519)
ln(Sample Size)	0.850^{***}	0.194^{*}	0.850^{***}	0.256^{**}
	(0.142)	(0.110)	(0.142)	(0.125)
Onsite Study	0.414	-0.273	0.414	0.251
	(0.353)	(0.330)	(0.353)	(0.360)
Publication	0.369	0.638^{***}	0.369	0.524^{*}
	(0.296)	(0.239)	(0.296)	(0.274)
East Africa		-0.669^{*}		-0.531
		(0.346)		(0.405)
ln(Reef Area)	0.156^{**}	-0.055	0.156^{**}	0.038
	(0.064)	(0.042)	(0.064)	(0.042)
MPA Status	3.395^{***}	-0.309	3.395^{***}	-0.728
	(1.172)	(0.474)	(1.172)	(0.550)
Snorkeling/Dive	-0.379	-0.091	-0.379	0.707^{**}
	(0.286)	(0.308)	(0.286)	(0.300)
Reef Quality	53.237^{***}	3.147^{***}	53.237^{***}	3.710^{***}
	(12.666)	(0.581)	(12.666)	(0.644)
Natural Reef	0.726^{***}	1.404^{*}	0.726^{***}	1.314
	(0.175)	(0.819)	(0.175)	(0.969)
Constant	-3.869^{***}	-0.029	-3.907^{***}	-0.193
	(1.255)	(0.832)	(1.259)	(0.993)
Observations	30	55	30	61
$\widehat{\gamma}$	0.17		0.17	
Log Likelihood	-11.041	-49.046	-11.041	-65.613
Akaike Inf. Crit.	54.081	132.091	54.081	165.226
Bayesian Inf. Crit.	76.500	166.216	76.500	201.110
Note:			*p<0.1; **p<0.05; ***p<0.01	

Further, when we switch attention to convergent validity and transfer errors we will see that the sample splitting threshold MRM for reef quality is inferior to both the regular MRM as well as the single variable threshold MRM. This is not unexpected, however. The increase in the number of parameters generated through the sample splitting, with the relatively low number of observations should lead to poor out of sample performance, which is what convergent validity is based off of.

5.2. **Benefit Transfer Reliability.** Next, we consider whether and how a threshold MRM can improve benefit transfer reliability. One approach to reveal the economic value of modeling a threshold is to use convergent validity tests. Londoño & Johnston (2012) use convergent validity tests to demonstrate that more rigid adherence to commodity and welfare consistency and the inclusion of more methodological variables resulted in substantially lower transfer errors. We find that accounting for thresholds in site characteristics can further improve transfer reliability.

We test for convergent validity using leave-one-observation-out (LOOO) and leave-one-study-out (LOSO) tests (Brander et al. 2007, Londoño & Johnston 2012, Boyle et al. 2013). In the former, we estimate equation (5) excluding observation js, then use our estimated coefficients to predict WTP for observation js, using that observation's site and methodological covariates. In the latter, we estimate equation (5) excluding study s, and use the estimated coefficients to predict each observation from study s. A transfer error is the difference between actual and predicted WTP (see equation (1).) To facilitate comparison between the leave-one-out approaches and across model types, we present transfer errors in absolute value percentage form: the absolute value of the ratio of transfer error to actual WTP.[10]

5.2.1. *Benchmark MRM.* Table 5 presents the results from convergent validity tests of the benchmark model applied to the original Londoño & Johnston (2012) data, as well as our

[10]We first exponentiate our estimates of WTP to account for the logarithmic form of the dependent variable in our MRM.

Table 5. Out of sample percent transfer errors: Absolute values, full meta regression model.

Meta-regression Model	Q_1	Q_2	Q_3	Mean	Std. Dev.
Combined Transfers, Londoño & Johnston (2012) data					
Omit Observation	28.73	56.28	87.87	98.82	195.71
Omit Study	22.99	58.91	94.87	104.11	192.51
Contingent Valuation Transfers, Londoño & Johnston (2012) data					
Omit Observation	24.80	40.87	76.92	96.83	227.01
Omit Study	24.97	54.73	84.62	154.07	282.90
Combined Transfers, Full Dataset					
Omit Observation	29.53	62.99	89.61	109.83	213.39
Omit Study	33.72	68.02	98.70	120.45	209.36
Contingent Valuation Transfers, Full Dataset					
Omit Observation	23.15	45.67	88.10	97.46	200.46
Omit Study	28.90	63.21	96.31	172.62	293.26

expanded dataset. In addition to mean transfer error, we present the quartiles and standard deviation of all transfer errors.[11]

Several points are worth noting in Table 5. First, as in Londoño & Johnston (2012), transfer errors are in general larger in the LOSO settings. Second, with the additional six observations there is a strict degradation in reliability for the pooled MRM, and degradation in all but one quartile (the first quartile in the LOOO distribution) of errors for the contingent valuation sample. Third, Table 5 indicates that strict adherence to welfare consistency does not improve reliability in all instances. With our expanded metadata, transfer errors are larger for the contingent valuation sample than for the pooled sample. This result, previously noted in Londoño & Johnston (2012), suggests that relaxing requirements for commodity consistency may actually improve transfer reliability based on the LOSO analysis; for the LOOO analysis, it is always the case that transfer errors are improved under strict commodity consistency. Fourth, more than 75% of the transfer errors are less than 100% (for both observation and study approaches) suggesting that the large mean error and standard

[11]We include quartiles because the distribution of transfer errors from each experiment is heavily skewed, making the means misleading measures of central tendency.

deviations are driven by a few, large transfer errors. This is true for both the original metadata and our expanded metadata.[12]

5.2.2. *Single Variable Threshold in Reef Quality MRM.* Table 6 contains convergent validity results using the single variable threshold model. For the LOOO approach, this threshold model improves reliability (reduces mean transfer error) in three out of four settings. Only in the expanded contingent valuation sample does the threshold MRM decrease reliability. For the LOSO experiments, the threshold improves reliability in two instances (Londoño & Johnston's (2012) contingent valuation sample and our expanded full sample). In the other two, reliability is either slightly degraded or largely degraded. Note that to calculate both the leave-one-observation and leave-one-study out analyses here we need to re-estimate the threshold each time an observation or study is omitted.

Table 6. Out of sample percent transfer errors: Absolute values, full meta regression model with threshold in reef quality.

Meta-regression Model	Q_1	Q_2	Q_3	Mean	Std. Dev.
Combined Transfers, Londoño & Johnston (2012) data					
Omit Observation	37.37	55.10	82.78	96.18	197.85
Omit Study	38.59	60.23	86.83	105.13	178.42
Contingent Valuation Transfers, Londoño & Johnston (2012) data					
Omit Observation	32.21	43.86	76.21	93.45	203.17
Omit Study	32.84	58.34	79.43	126.12	225.54
Combined Transfers, Full Dataset					
Omit Observation	30.01	63.09	91.72	101.40	198.34
Omit Study	28.10	60.69	91.84	104.20	177.16
Contingent Valuation Transfers, Full Dataset					
Omit Observation	25.65	56.86	90.59	116.50	266.59
Omit Study	47.93	76.03	117.76	224.80	390.23

For the pooled MRM using our expanded metadata, adding a threshold for reef quality reduces the mean error and variability in transfer errors for both the LOOO and LOSO analyses. Further, the threshold leads to better performance throughout the transfer error

[12]For comparison, Appendix C includes an identical table to Table 5, except that reef quality enters the MRM in logarithmic form. The results there suggest that for the full Londoño & Johnston (2012) dataset, the LOOO and LOSO prediction errors are worse than when reef quality enters linearly.

distribution for the LOSO analysis of full sample. We emphasize here that the threshold MRM does not always out perform the baseline MRM and this is important practically because it suggests that modeling a threshold can add more noise than signal to the benefit transfer process. This is particularly true for our expanded contingent valuation sample, which sees larger mean errors from both the LOOO and LOSO analyses. Still, we interpret lower transfer errors for the combined sample as evidence that accounting for thresholds in the impacts of biophysical properties on WTP can improve benefit transfer reliability.[13]

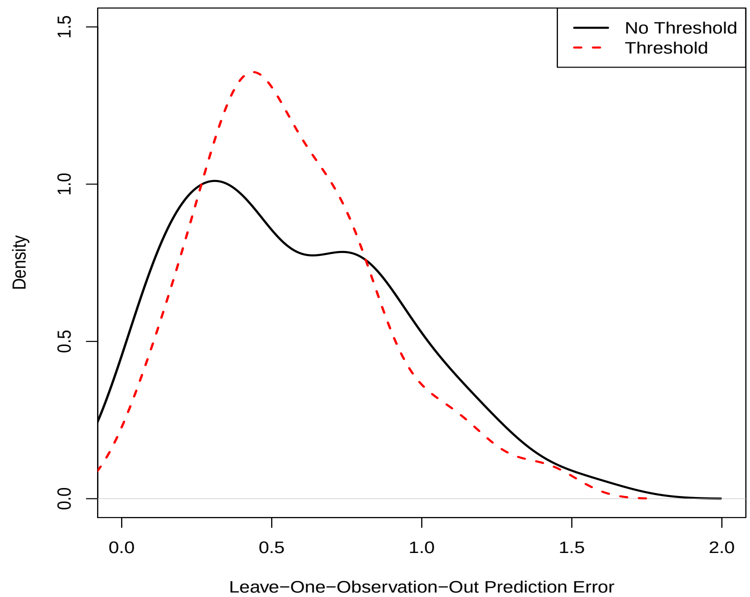
To visually demonstrate the potential value added of including a threshold in a MRM for benefit transfer, Figures 1 and 2 present kernel density estimates of the predicted transfer error for both the LOOO and LOSO studies using the original Londoño & Johnston (2012) metadata. When constructing the density we omit outlier transfer errors from both the simple and threshold MRM, defined as any transfer error larger than 1.5 times the interquartile range beyond the upper quartile.[14] This is done to focus on the visual aspects of the densities, since adjusting the range of the figure to present the whole density would yield less information than simply not accounting for these outliers.

While the lower end of the transfer error distribution favors the simple MRM, the upper tail for both the LOOO and LOSO predicted transfer errors is smaller for the threshold MRM and the distribution has collapsed around the median. Certainly, using either the mean or median, there are instances when the baseline MRM is preferred to the threshold MRM, especially using the LOSO results. However, using a robust measure of spread (the interquartile range (IQR)) the threshold MRM yields gains beyond the baseline MRM. In five of the eight settings the IQR is lower for the threshold MRM, and is equal in a sixth.

[13]A table similar to Table 6 appears in Appendix B. As noted earlier, we do not discuss results from convergent validity tests for a threshold in reef area for the MRM given that this model fit the metadata worse, according to both AIC and BIC scores, than a threshold in reef area.

[14]This lead to the removal of seven predicted transfer errors from the baseline MRM and nine from the threshold MRM. Six of these were from the same observations across both models, suggesting broad comparability.

FIGURE 1. Estimated density of predicted transfer error (LOOO) for the simple and single variable threshold MRMs using Londoño and Johnston's (2012) metadata.



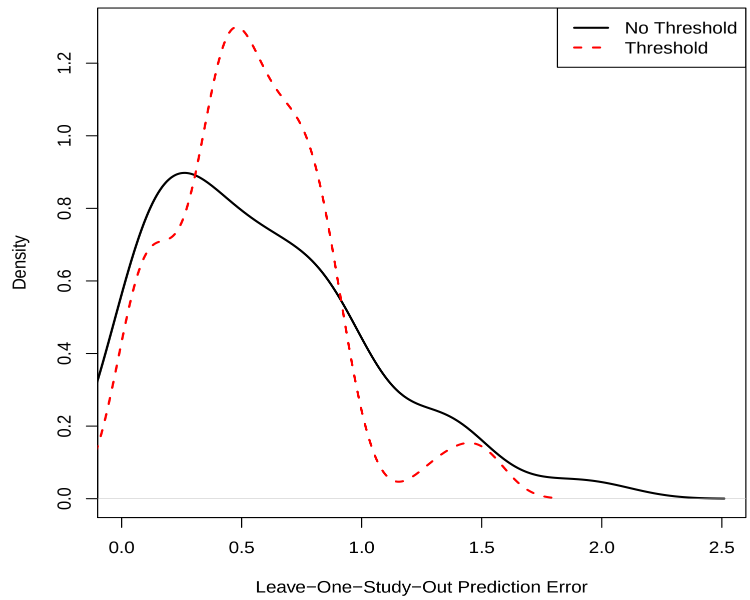
We thus view these two models as complements, rather than competitors when it comes to meta-analysis benefit transfer.[15]

Another way to view the convergent validity results from the benchmark and single variable threshold MRMs is that the threshold MRM is less likely to generate either extremely high or low transfer errors. This can be taken as evidence that the threshold model has less risk relative to the benchmark MRM model when it comes to benefit transfer.

5.2.3. *Sample Splitting Threshold in Reef Quality.* Studying convergent validity based on splitting the entire metadata sample for a threshold in reef quality leads to a different conclusion that what we saw in Table 10. Our convergent validity results in Table 7 are largely disappointing relative to both the benchmark Londoño & Johnston (2012) findings and our

[15]Given the small sample size, it is not surprising that the LOSO tests favor the baseline MRM more than the LOOO tests. For certain tests, a large number of observations are left out, making it difficult to identify the threshold.

Figure 2. Estimated density of predicted transfer error (LOSO) for the simple and single variable threshold MRMs using Londoño and Johnston's (2012) metadata.



single variable threshold results in Table 6. Outside of the LOOO analyses producing lower Q_1 transfer errors in three of the four experiments relative to Table 5 and in all four of the experiments relative to Table 6, in general the sample splitting threshold MRM does not predict average WTP as accurately as the other MRMs. In every setting average transfer error, and the standard deviation of the transfer error is larger than the other two approaches.

More troubling are the LOSO results. Across all four experiments, there always existed at least two transfer errors that were orders of magnitude larger than any transfer error from the other two approaches. Given the small sample size of the meta dataset, removal of a study with more than one or two observations could have a distinct effect on the estimate of the threshold in the leave-one-out exercise, leading to inferior prediction.

Figures 3 and 4 presents kernel density estimates of the predicted transfer error for both the LOOO and LOSO studies using the original Londoño & Johnston (2012) metadata

TABLE 7. Out of sample percent transfer errors: Absolute values, sample splitting meta regression model with threshold in reef quality. A tabular entry of ∞ implicates the the corresponding quantity was in excess of $10e + 56$.

Meta-regression Model	Q_1	Q_2	Q_3	Mean	Std. Dev.
Combined Transfers, Londoño & Johnston (2012) data					
Omit Observation	24.87	44.60	87.21	149.38	530.53
Omit Study	36.37	71.90	121.07	∞	∞
Contingent Valuation Transfers, Londoño & Johnston (2012) data					
Omit Observation	20.37	45.78	81.71	145.68	338.78
Omit Study	42.16	84.33	127.56	∞	∞
Combined Transfers, Full Dataset					
Omit Observation	23.17	50.02	100.00	159.88	516.75
Omit Study	43.00	81.12	156.88	∞	∞
Contingent Valuation Transfers, Full Dataset					
Omit Observation	36.02	61.01	88.92	148.73	291.04
Omit Study	54.04	87.70	222.27	∞	∞

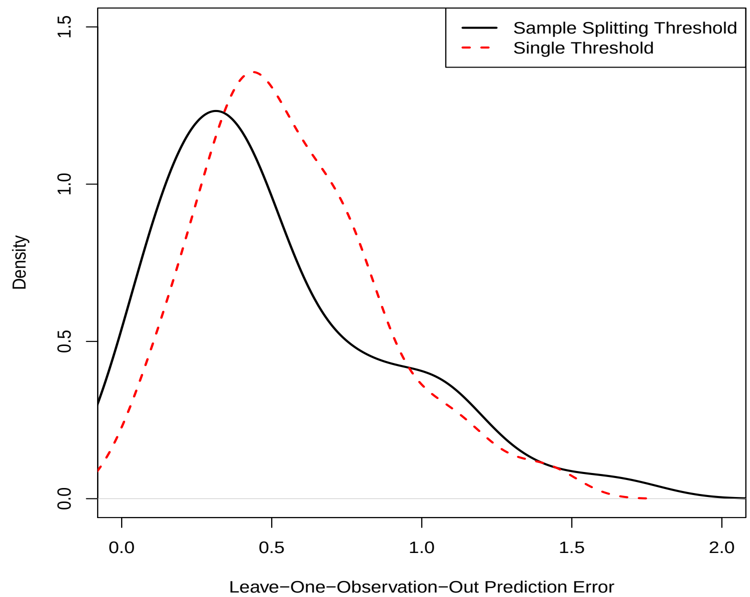
comparing the single variable and sample splitting threshold MRMs. As before, when we construct the density we omit outlier transfer errors from both threshold MRMs.[16]

The lower end of the transfer error distribution favors the sample splitting threshold MRM for both LOOO and LOSO predicted transfer errors. The upper tail of the LOSO predicted transfer error distribution is larger for the sample splitting threshold MRM as opposed to the single variable threshold MRM; with the small sample size, it is not surprising that the single variable threshold MRM is preferred to the sample splitting threshold MRM with respect to avoiding large errors. Also, these results really only capture the central portion of these densities as we have removed outliers prior to presenting these results. As Table 7 made clear, the largest transfer errors for the sample splitting MRM are orders of magnitude larger than the baseline and single variable threshold MRMs.

This is not to discredit the sample splitting threshold MRM or discourage its use. Here our sample size is probably too small to unequivocally favor this model over the single variable threshold MRM even though the model selection criterion strongly support such a claim. A

[16]For the sample splitting threshold MRM, 9 observations were omitted for the LOOO experiment and 14 observations were omitted for the LOSO experiment.

FIGURE 3. Estimated density of predicted transfer error (LOOO) for the single variable and sample splitting threshold MRMs using Londoño and Johnston's (2012) metadata.

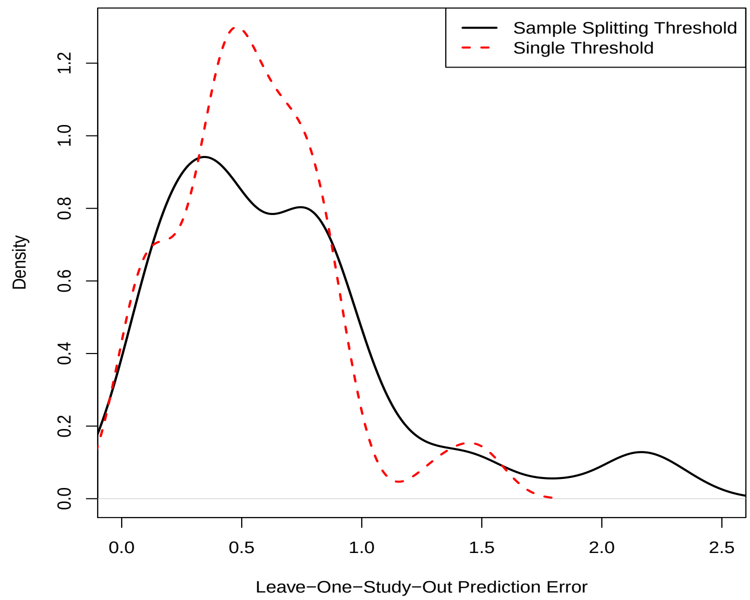


larger dataset for a different benefit transfer may yield more convincing evidence. We end by noting that both models, in various ways, offer improvement over the benchmark, linear in parameters MRM.

6. Conclusions

Londoño & Johnston (2012) demonstrate that focusing on metadata consistency and controlling for biophysical characteristics at valuation sites can improve benefit transfer reliability. Their meta-analysis of coral reef valuation studies and subsequent tests of convergent validity reveal transfer errors comparable in magnitude to other recreational amenities (Rosenberger & Phipps 2007), which is an improvement upon an earlier coral reef meta-analysis (Brander et al. 2007). To explore the potential for further reliability improvements, we focus our attention on the functional relationship between continuous site characteristics

Figure 4. Estimated density of predicted transfer error (LOSO) for the single variable and sample splitting threshold MRMs using Londoño and Johnston's (2012) metadata.



and WTP. While previous meta-analyses capture non-linearities with logarithmic or quadratic functional forms or non-parametric estimation (Kaul et al. 2013, Boyle et al. 2015), we propose MRMs that capture non-linearities using thresholds.

Our threshold MRMs allow for discontinuous changes in the impact of one or more methodological or site characteristics across the distribution of a given characteristic. If these impacts differ across the threshold, then a meta-analysis that assumes a uniform impact might generate larger transfer errors. Comparing tests of convergent validity between the benchmark and threshold MRMs reveals that the threshold models can improve benefit transfer reliability.

Compared to Londoño & Johnston (2012), our threshold approach leads to mixed performance in terms of mean transfer error. When comparing the full distribution of transfer errors, however, our preferred threshold model generates fewer extremely large transfer errors.

Therefore practitioners may find the threshold MRM to be less likely to generate unacceptably large transfer errors, compared to a MRM that assumes uniform impacts of biophsyical impacts. Although we find some evidence that threshold MRMs can lead to improvements in transfer reliability, the magnitude of errors suggests meta-analysis of coral reefs is better suited for broad welfare guidance, rather than targeted policy recommendations (Londoño & Johnston 2012).

Situations that require precise point estimates of values, such as cost-benefit analyses and applications in litigation, are still better served by primary studies of individual sites (Bergstrom & De Civita 1999). Still, that we can reduce variation in the transfer error distribution reveals the importance of accounting for threshold effects in meta-analysis of environmental amenities, and the potential for the use of threshold models in future meta-analyses.

References

Bateman, I. & Jones, A. (2003), 'Contrasting conventional with multi-level modeling approaches to meta-analysis: Expectation consistency in UK woodland recreation values', *Land Economics* **79**(2), 235–258.

Bell, J. & Galvin, R. (1984), 'Influence of live coral cover on coral reef fish communities', *Marine Ecology Progress Series* **15**(3), 265–274.

Bergstrom, J. & De Civita, P. (1999), 'Status of benefits transfer in the United States and Canada: A review', *Canadian Journal of Agricultural Economics* **47**(1), 79–87.

Boyle, K. J., Kaul, S. & Parmeter, C. F. (2015), Meta-analysis: Econometric advances and new perspectives toward data synthesis and robustness, *in* R. Johnston, R. Rolfe, R. Rosenberger & R. Brouwer, eds, 'Benefit Transfer of Environmental and Resource Values: A Handbook for Researchers and Practitioners', Springer, chapter 17.

Boyle, K. J., Kuminoff, N. V., Parmeter, C. F. & Pope, J. C. (2009), 'Necessary conditions for valid benefit transfers', *American Journal of Agricultural Economics* **91**(5), 1328–1334.

Boyle, K. J., Parmeter, C. F., Boehlert, B. & Paterson, R. (2013), 'Due diligence in meta-analyses to support benefit transfer', *Environmental and Resource Economics* **55**(3), 357–386.

Braden, J., Feng, X. & Won, D. (2011), 'Waste sites and property values: A meta-analysis', *Environmental and Resource Economics* **50**(1), 175–201.

Brander, L., Eppink, F., Schagner, P., van Beukering, P. & Wagtendonk, A. (2015), GIS-based mapping of ecosystem services: The case of coral reefs, *in* R. Johnston, R. Rolfe, R. Rosenberger & R. Brouwer, eds, 'Benefit Transfer of Environmental and Resource Values: A Handbook for Researchers and Practitioners', Springer, chapter 20.

Brander, L., van Beukering, P. & Cesar, H. (2007), 'The recreational value of coral reefs: A meta-analysis', *Ecological Economics* **63**(1), 209–218.

Brown, B., Dunne, P., Phuongsuwan, N. & Somerfield, P. (2011), 'Increased sea level promotes coral cover on shallow reef flats in the Andaman Sea, eastern Indian Ocean', *Coral Reefs* **30**(1), 867–878.

Coker, D. J., Wilson, S. K. & Pratchett, M. S. (2014), 'Importance of live coral habitat for reef fishes', *Reviews in Fish Biology and Fisheries* **24**(1), 89–126.

Death, G., Fabricius, K., Sweatman, H. & Puotinen, M. (2012), 'The 27-year decline of coral cover on the Great Barrier Reef and its causes', *Proceedings of the National Academy of Sciences* **109**(44), 17995–17999.

Hansen, B. E. (2000), 'Sample splitting and threshold estimation', *Econometrica* **68**(3), 575–603.

Hoegh-Guldberg, O., Eakin, C., Hodgson, G., Sale, P. & Veron, J. (2015), Climate change threatens the survival of coral reefs, Technical report, International Society for Reef Studies, Hawaii.

Johnston, R. J., Besedin, E., Iovanna, R., Miller, C., Wardwell, R. & Ranson, M. (2005), 'Systematic variation in willingness to pay for aquatic resource improvements and implications for benefit transfer: A meta-analysis', *Canadian Journal of Agricultural Economics* **53**(3), 221–248.

Johnston, R. J., Besedin, E. & Ranson, M. (2006), 'Characterizing the effects of valuation methodology in function-based benefits transfer', *Ecological Economics* **60**(1), 407–419.

Johnston, R. J. & Rosenberger, R. S. (2010), 'Methods, trends and controversies in contemporary benefit transfer', *Journal of Economic Surveys* **24**(3), 479–510.

Johnston, R., Rolfe, J., Rosenberger, R. & Brouwer, R. (2015), *Benefit Transfer of Environmental and Resource Values: A Guide for Researchers and Practitioners*, Springer, New York, New York.

Kaul, S., Boyle, K. J., Kuminoff, N. V., Parmeter, C. F. & Pope, J. C. (2013), 'What can we learn from benefit transfer errors? Evidence from 20 years of research on convergent validity', *Journal of Environmental Economics and Management* **66**(1), 90–104.

Kragt, M. E., Roebeling, P. C. & Ruijs, A. (2009), 'Effects of Great Barrier Reef degradation of recreational reef-trip demand: A contingent behavior approach', *The Australian Journal of Agricultural and Resource Economics* **53**(2), 213–229.

Londoño, L. M. & Johnston, R. J. (2012), 'Enhancing the reliability of benefit transfer over heterogeneous sites: A meta-analysis of international coral reef values', *Ecological Economics* **78**(1), 80–89.

Moeltner, K., Boyle, K. & Paterson, R. (2007), 'Meta-analysis and benefit transfer for resource valuation - addressing classical challenges with Bayesian modeling', *Journal of Environmental Economics and Management* **53**(1), 250–269.

Nelson, J. P. & Kennedy, P. E. (2009), 'The use (and abuse) of meta-analysis in environmental and resource economics: An assessment', *Environmental and Resource Economics* **42**(3), 345–377.

Parsons, G. R. & Thur, S. M. (2008), 'Valuing changes in quality of coral reef ecosystems: a stated preference study of SCUBA diving in Bonaire National Marine Park', *Environmental and Resource Economics* **40**(4), 593–608.

Phewmau, K. (2013), 'Recreational valuation of the coral diving activities at Similan island national park in Thailand', *The Empirical Econometrics and Quantitative Economics Letters* **2**(1), 79–86.

Ransom, K. P. & Mangi, S. C. (2010), 'Valuing recreation benefits of coral reefs: The case of Mombasa marine national park and reserve, Kenya', *Environmental Management* **45**(2), 145–154.

Rolfe, J., Brouwer, R. & Johnston, R. (2015), Meta-analysis: Rationale, issues, and applications, *in* R. Johnston, R. Rolfe, R. Rosenberger & R. Brouwer, eds, 'Benefit Transfer of Environmental and Resource Values: A Handbook for Researchers and Practitioners', Springer, chapter 16.

Rosenberger, R. (2015), Benefit transfer validity and reliability, *in* R. Johnston, R. Rolfe, R. Rosenberger & R. Brouwer, eds, 'Benefit Transfer of Environmental and Resource Values: A Handbook for Researchers and Practitioners', Springer, chapter 14.

Rosenberger, R. & Loomis, J. (2000), 'Using meta-analysis for benefit transfer: In-sample convergent validity tests of an outdoor recreation database', *Water Resources Research* **36**(4), 1097–1107.

Rosenberger, R. & Phipps, T. (2007), Correspondence and convergence in benefit transfer accuracy: Meta-analytic review of the literature, *in* S. Navrud & R. Ready, eds, 'Environmental Value Transfer: Issues and Methods', Springer, chapter 3.

Rosenberger, R. S. & Stanley, T. D. (2006), 'Measurement, generalization, and publication: Sources of error in benefits transfers and their management', *Ecological Economics* **60**(2), 372–378.

Shrestha, R. & Loomis, J. (2001), 'Testing a meta-analysis model for benefit transfer in international outdoor recreation', *Ecological Economics* **39**(1), 67–83.

Smith, V. & Huang, J. (1995), 'Can markets value air quality? A meta-analysis of hedonic property value models', *Journal of Political Economy* **103**(1), 209–227.

Smith, V. K. & Pattanayak, S. K. (2002), 'Is meta-analysis a Noah's ark for non-market valuation?', *Environmental and Resource Economics* **22**(1-2), 271–296.

Stanley, T. D., Doucouliagos, H., Giles, M., Heckemeyer, J. H., Johnston, R. J., Laroche, P., Nelson, J., Paldam, M., Poot, J. & Pugh, G. (2013), 'Meta-analysis of economics research reporting guidelines', *Journal of Economic Surveys* **27**(2), 390–394.

Van Houtven, G., Powers, J. & Pattanayak, S. (2007), 'Valuing water quality improvements in the united states using meta-analysis: Is the glass half-full or half-empty for national policy analysis?', *Resource and Energy Economics* **29**(1), 206–228.

Wilkinson, C. (2008), *Status of the Coral Reefs of the World: 2008*, Australian Institute of Marine Science, Townsville, Australia.

Woodward, R. T. & Wui, Y. (2001), 'The economic value of wetland services: A meta-analysis', *Ecological Economics* **37**(2), 257–270.

Yu, P. (2012), 'Likelihood estimation and inference in threshold regression', *Journal of Econometrics* **167**(2), 274–294.

Appendix A. Implementation of the Threshold Model

Consider either the sample splitting threshold model in (3) and (4) or the single variable threshold model in (5). If the threshold, γ, were known then estimation of (3) and (4) is straightforward, requiring the researcher to split the sample and estimate the multi-level MRM using only observations for which $\bar{z}_{js}$ are less than or equal to γ, and then using only observations for which $\bar{z}_{js}$ greater than γ. To estimate (5), for the known value of γ, the indicator function $1\{\bar{z}_{js} > \gamma\}$ is first created and then a single, multi-level model is estimated. However, rarely in practice do we know γ, and so it must be estimated.

To develop an estimator of γ in either threshold model, (3) and (4) or (5), we use the idea of profiling, proposed in Hansen (2000). This can be done as follows. First, we create a grid of quantiles of $\bar{z}_{js}$. The reason for using a grid over the quantiles of $\bar{z}_{js}$ is that the use of any other values is redundant since any value of γ between $\bar{z}_{(K)}$ and $\bar{z}_{(K+1)}$ (where $\bar{z}_{(G)}$ is the Gth quantile of $\bar{z}_{js}$) will produce the same partition of the data as the quantiles. Then, we set γ equal to each value in the grid and evaluate the corresponding multi-level model. The value of γ which produces the highest aggregate maximum likelihood is the estimator for the threshold value. We say aggregate maximum likelihood because in the sample splitting threshold model we have to account for the likelihood of the estimated multilevel model in (3) and (4). For the indicator threshold variable, (5), we have a single maximum likelihood value.

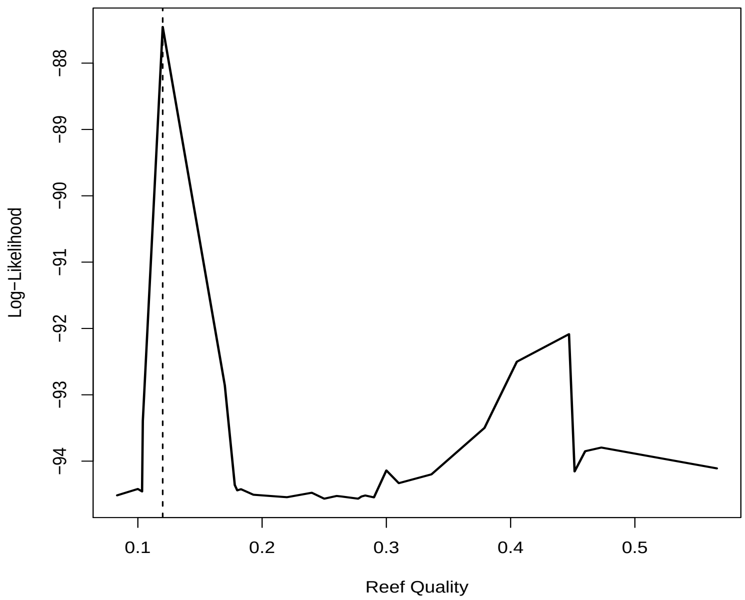
Two important user issues exist here, the number of quantiles in the grid to evaluate over and the number of observations to leave on the ends of the grid. It is necessary to leave enough room on both sides of the threshold to allow the model to be well identified. In practice it is common to leave out roughly 1-10% of the observations on each side of the threshold (depending upon the sample size), but there currently does not exist a well defined cut-off. Naturally, as the sample size increases the proportion of the data on one side of the threshold can be reduced. For example, with 400 observations, if we were to leave 3% of the data on each side of the threshold then we would search between $\bar{z}_3$ and $\bar{z}_{(97)}$. Remember, since we do not know where the threshold hold is, we have to leave enough data on both ends of the range of $\bar{z}$ to ensure the threshold can be reliably estimated. The number of points to use in the grid also will have an impact on the estimator of γ; if too few points are chosen the estimator of γ could be quite noisy.

Several alternative estimators exist for γ based on the grid of quantiles. For example, Yu (2012) uses the center value between the two quantiles, rather than the lower quantile value. More specifically, when searching along the grid suppose the value of γ that leads to the highest aggregate likelihood is $\bar{z}_{(G)}$, Yu's (2012) estimator of γ is $\hat{\gamma}_M = \left(\bar{z}_{(G)} + \bar{z}_{(G+1)}\right)/2$, whereas the lower quantile estimator would be $\hat{\gamma}_L = \bar{z}_{(G)}$. Simulations in Yu (2012, Table 1) reveal marginal improvements in use of $\hat{\gamma}_M$ over $\hat{\gamma}_L$.

For clarity we present plots of our likelihood function for the threshold MRM of reef quality using the metadata of Londoño & Johnston (2012). For the grid we have 36 unique values of reef quality across the 85 observations. Looking at the interdecile range, i.e. the 10th to 90th quantile, this leaves us with only 29 quantile values in the data to estimate the threshold MRM.[17] As is clear from Figure 5 we have a well defined maximum of the likelihood function (identified with a vertical dashed line) even though the profiling does not produce a convex relationship. This is not concerning in practice as, with a single variable threshold, this amounts to univariate optimization which can be done visually over the fixed range of quantiles.

[17]Given that we are omitting ≈20% of the 36 unique values, this leaves us with $36 - 7 = 29$ points.

FIGURE 5. Estimated log likelihood for threshold MRM in reef quality using Londoño and Johnston's (2012) metadata. The dashed line represents the value of reef quality where the profiled log likelihood is maximized.



Appendix B. A Threshold in Reef Area

B.1. **A single variable threshold.** Aside from investigating the presence of a threshold in reef quality for the MRM, we also consider a threshold for the size of the reef. Table 8 presents estimates allowing for a threshold only for the area of the reef under study for the full MRM (this is the model in (5)).

We see similar results to those modeling a threshold in reef quality. Many of the same qualitative conclusions from Londoño & Johnston's (2012) research appear. However, regarding fit of the threshold MRM, we see that for all four of the different variants of the model that both the AIC and BIC criterion suggest that a threshold in reef quality fits the metadata better than a threshold in reef area. One quantitative difference is that when we switch from using all of the observations to just those that model WTP using contingent valuation methods, the threshold in reef area becomes much smaller. This differs from the threshold in reef quality, that was invariant across the valuation method.

Aside from studying the estimates of the MRM with a threshold in reef size, we can see if modeling the presence of a size threshold produces improvements in prediction errors. Table 9 presents LOOO and LOSO statistics using the threshold indicator model (5) for area of the reef (the same which appears in Table 8 but with a threshold in reef quality).

We observe similar behavior, a reduction in mean transfer error for several of our hold out sample experiments, smaller upper quartiles, and a condensing of the interquartile range.

B.2. **A Sample Splitting Threshold in Reef Area.** For comparison we also estimated the sample splitting threshold MRM for reef area. Table 10 presents estimates from the meta regression with a potential threshold in the area of the reef under study for the full MRM (that is (3) and (4)).

Table 8. Single Variable Threshold (Reef Area) Meta-Regression Model Estimates: All Studies.

	Londoño & Johnston (2012) Data		Full Data	
	(1)	(2)	(3)	(4)
Discrete Choice	-1.092^{***}	-0.563^{*}	-0.948^{***}	-0.287
	(0.305)	(0.304)	(0.325)	(0.356)
Payment Card	-1.142^{***}	-0.873^{***}	-0.987^{***}	-0.551
	(0.254)	(0.283)	(0.278)	(0.335)
Travel Cost	-0.108		0.251	
	(0.347)		(0.362)	
Trip Cost	1.138^{***}	0.778^{**}	0.676^{**}	0.290
	(0.274)	(0.326)	(0.278)	(0.372)
Donation	-0.916^{**}	-0.619	-1.014^{**}	-0.546
	(0.402)	(0.413)	(0.435)	(0.449)
ln(Sample Size)	0.146	0.247^{**}	0.241^{**}	0.263^{***}
	(0.108)	(0.097)	(0.112)	(0.100)
Onsite Study	-0.271	-0.370	-0.082	-0.208
	(0.273)	(0.274)	(0.294)	(0.300)
Publication	0.731^{***}	0.228	0.641^{***}	0.279
	(0.208)	(0.254)	(0.226)	(0.291)
East Africa	-1.119^{***}	-0.092	-1.222^{***}	0.068
	(0.371)	(0.457)	(0.386)	(0.529)
ln(Reef Area)	-0.102^{*}	-0.040	-0.030	0.135^{**}
	(0.056)	(0.048)	(0.055)	(0.059)
MPA Status	0.759^{**}	0.851^{**}	0.603	0.533
	(0.339)	(0.370)	(0.369)	(0.429)
Snorkeling/Dive	0.280	0.335	0.534^{**}	0.490^{*}
	(0.239)	(0.280)	(0.238)	(0.290)
Reef Quality	1.735^{***}	1.795^{***}	2.048^{***}	1.953^{***}
	(0.473)	(0.563)	(0.505)	(0.594)
Natural Reef	0.636^{**}	0.601^{**}	0.546	0.614^{**}
	(0.321)	(0.253)	(0.352)	(0.263)
ln(Reef Area) $*$I(ln(Reef Area) $\leq \gamma$)	0.188^{**}	0.092	0.137	-0.136
	(0.081)	(0.057)	(0.085)	(0.086)
Constant	1.611^{***}	1.357^{***}	1.585^{***}	1.122^{***}
	(0.353)	(0.309)	(0.386)	(0.348)
Observations	85	71	91	76
$\widehat{\gamma}$	-1.05	-0.73	-6.45	-5.60
Log Likelihood	-91.950	-66.489	-107.520	-77.134
Akaike Inf. Crit.	219.900	166.978	251.041	188.268
Bayesian Inf. Crit.	263.868	205.443	296.236	227.891

Note: *p<0.1; **p<0.05; ***p<0.01

Table 9. Out of sample percent transfer errors: Absolute values, single variable threshold (Reef Area) meta regression model.

Meta-regression Model	Q_1	Q_2	Q_3	Mean	Std. Dev.
Combined Transfers, Londoño & Johnston (2012) data					
Omit Observation	24.30	53.70	82.16	96.94	167.90
Omit Study	26.30	49.18	81.10	111.08	183.50
Contingent Valuation Transfers, Londoño & Johnston (2012) data					
Omit Observation	20.97	38.25	74.01	93.44	242.80
Omit Study	21.36	55.48	89.66	156.37	294.24
Combined Transfers, Full Dataset					
Omit Observation	27.95	56.74	90.65	107.06	190.14
Omit Study	36.37	65.92	89.67	130.44	215.38
Contingent Valuation Transfers, Full Dataset					
Omit Observation	24.06	49.58	85.30	98.30	181.51
Omit Study	38.09	71.02	108.23	175.26	274.62

Table 10. Threshold Sample Splitting Meta-Regression Model (Reef Area) Estimates: All Studies.

	(1)	(2)	(3)	(4)
Discrete Choice	0.486	−0.994***		−0.930***
	(0.407)	(0.335)		(0.357)
Payment Card	−0.212	−2.051***	3.575***	−1.072***
	(0.227)	(0.438)	(0.413)	(0.405)
Travel Cost	4.542***	−0.629		0.384
	(0.616)	(0.392)		(0.393)
Trip Cost	0.567***	1.391***	−3.640***	0.369
	(0.208)	(0.368)	(0.446)	(0.330)
Donation	0.341	−1.614***	0.301***	−1.575***
	(0.222)	(0.494)	(0.080)	(0.533)
ln(Sample Size)	0.076	0.302**	0.093**	0.305**
	(0.070)	(0.148)	(0.042)	(0.126)
Onsite Study	−1.343***	−0.313		−0.233
	(0.267)	(0.339)		(0.326)
Publication	0.007	0.777***	−3.803***	0.524*
	(0.151)	(0.271)	(0.412)	(0.271)
East Africa	0.624**	−1.688***		−0.962**
	(0.253)	(0.428)		(0.421)
ln(Reef Area)	−0.147**	−0.088	1.992***	0.034
	(0.067)	(0.056)	(0.185)	(0.050)
MPA Status	3.112***	0.281		0.388
	(0.304)	(0.442)		(0.421)
Snorkeling/Dive	−0.895***	−0.126		0.508*
	(0.281)	(0.312)		(0.299)
Reef Quality	5.104***	1.696***	−19.973***	2.552***
	(0.674)	(0.576)	(2.175)	(0.584)
Natural Reef		0.662**		0.505
		(0.323)		(0.378)
Constant	1.548***	1.407***	18.172***	1.282***
	(0.301)	(0.369)	(1.498)	(0.411)
Observations	21	64	11	80
$\widehat{\gamma}$	-1.56		-5.83	
Log Likelihood	5.870	−67.203	14.365	−98.139
Akaike Inf. Crit.	20.260	168.406	−8.730	230.278
Bayesian Inf. Crit.	36.972	205.107	−4.751	270.772

Note: *p<0.1; **p<0.05; ***p<0.01

Appendix C. Convergent Validity Tests for Logarithm of Reef Quality

As an alternative way to modeling nonlinearities in a MRM, we estimated the MRM using both datasets but included reef quality in logarithmic form instead of with a threshold. While different than a threshold, it still allows a degree of nonlinearity into the MRM. Table 11 presents leave-one-observation and leave-one-study out tests of convergent validity for the baseline MRM using both the original Londoño & Johnston (2012) data as well as the updated dataset where Reef Quality is included in logarithmic form as opposed to level form. We present the quartiles of all transfer errors as well as the mean transfer error and the standard deviation of all transfer errors.

Table 11. Out of sample percent transfer errors: Absolute values, full meta regression model, using logarithm of Reef Quality.

Meta-regression Model	Q_1	Q_2	Q_3	Mean	Std. Dev.
Combined Transfers, Londoño & Johnston (2012) data					
Omit Observation	28.82	57.38	91.40	106.41	217.04
Omit Study	29.73	61.78	87.07	118.41	237.40
Contingent Valuation Transfers, Londoño & Johnston (2012) data					
Omit Observation	19.70	45.76	75.69	96.71	238.80
Omit Study	16.36	53.29	84.92	129.82	262.04
Combined Transfers, Full Dataset					
Omit Observation	29.69	59.64	91.45	120.63	249.54
Omit Study	40.67	65.95	91.48	140.56	265.86
Contingent Valuation Transfers, Full Dataset					
Omit Observation	23.62	44.60	81.65	97.55	207.74
Omit Study	21.98	60.02	95.62	152.38	261.09

In comparison to the convergent validity findings of Londoño & Johnston (2012), accounting for nonlinearity between WTP and reef area through the use of a logarithm in reef quality produced almost uniformly worse relative prediction errors using the original metadata of Londoño & Johnston (2012). However, focusing attention on just contingent valuation studies produce almost the opposite result, with a logarithm of reef quality having lower transfer errors throughout the distribution. This same pattern emerges using the updated metadata. For the full set of studies, both contingent valuation and travel cost, placing reef quality linearly into the MRM results in lower transfer errors than when reef quality appeared logarithmically. However, focusing on just the studies using contingent valuation, this result flips, with the logarithm of reef quality outperforming reef quality entering in a linear fashion. Clearly the results here suggest that some type of nonlinear relationship between reef quality and WTP exists in the metadata.