Conservation Biology

Review

Producer-Level Benefits of Sustainability Certification

ALLEN BLACKMAN*† AND JORGE RIVERA†‡

*Resources for the Future, Washington, D.C. 20008, U.S.A., email blackman@rff.org
†Environment for Development Center for Central America, Turrialba, Costa Rica
‡George Washington University, Washington, D.C. 20052, U.S.A.

Abstract: *Initiatives certifying that producers of goods and services adhere to defined environmental and social-welfare production standards are increasingly popular. According to proponents, these initiatives create financial incentives for producers to improve their environmental, social, and economic performance. We reviewed the evidence on whether these initiatives have such benefits. We identified peer-reviewed, ex post, producer-level studies in economic sectors in which certification is particularly prevalent (bananas, coffee, fish products, forest products, and tourism operations), classified these studies on the basis of whether their design and methods likely generated credible results, summarized findings from the studies with credible results, and considered how these findings might guide future research. We found 46 relevant studies, most of which focused on coffee and forest products and examined fair-trade and Forest Stewardship Council certification. The methods used in 11 studies likely generated credible results. Of these 11 studies, nine examined the economic effects and two the environmental effects of certification. The results of four of the 11 studies, all of which examined economic effects, showed that certification has producer-level benefits. Hence, the evidence to support the hypothesis that certification benefits the environment or producers is limited. More evidence could be generated by incorporating rigorous, independent evaluation into the design and implementation of projects promoting certification.*

Keywords: ecocertification, ecolabel, human-dominated landscape

Beneficios de la Certificación de la Sustentabilidad a Nivel de Productores

Resumen: *Las iniciativas de certificación de productores de bienes y servicios que se apegan a estándares ambientales y de producción de bienestar social son cada vez más populares. De acuerdo con los proponentes, estas iniciativas crean incentivos financieros para que los productores mejoren su desempeño ambiental, social y económico. Revisamos la evidencia para ver si esas iniciativas tienen tales beneficios. Identificamos estudios ex post y a nivel de productores sometidos a revisión cientifica external, en sectores económicos en los que la certificación es particularmente prevalente (plátano, café, productos pesqueros, productos forestales y operaciones turísticas), los clasificamos considerando si su diseño y métodos generaron resultados creíbles, sintetizamos los hallazgos de los estudios con resultados creíbles y consideramos como estos hallazgos pueden dirigir investigaciones en el futuro. Encontramos 46 estudios relevantes, muchos de ellos enfocados en café y productos comerciales y con certificación de comercio justo y Forest Stewardship Council. Los métodos utilizados en 11 estudios generaron resultados creíbles. De estos 11 estudios, nueve examinaron los efectos económicos y dos los efectos ambientales de la certificación. Los resultados de cuatro de los 11 estudios, todos examinando efectos económicos, mostraron que la certificación tiene beneficios a nivel de productores. Por lo tanto, la evidencia para apoyar la hipótesis de que la certificación beneficia al ambiente o a los productores es limitada. Se deben generar mas evidencia mediante la incorporación de evaluaciones independientes más rigorosas en el diseño e implementación de proyectos que promueven la certificación.*

Palabras Clave: ecocertificación, eco-etiquetas, paisaje dominado por humanos

Paper submitted August 26, 2010; revised manuscript accepted October 7, 2010.

DOI: 10.1111/j.1523-1739.2011.01774.x

Introduction

Initiatives certifying that producers of goods and services adhere to defined environmental and social-welfare production standards are increasingly popular. Today, 10% of the timber, 7% of the coffee, and 12% of the wild fish products traded in international markets are certified as being sustainably produced by organizations such as the Forest Stewardship Council, Rainforest Alliance, and the Marine Stewardship Council (Eilperin 2010). According to proponents, such certification spurs producers to improve their environmental, social, and economic performance (Rice & Ward 1996; Giovannucci & Ponte 2005). In theory, it does so by enabling the consumer to differentiate among goods and services on the basis of their environmental and social attributes and effects. This ability to differentiate facilitates price premiums and expands market access for certified products. Price premiums and market access, in turn, create financial incentives for producers to meet certification standards.

Yet certification programs that aim to improve commodity producers' environmental, social, and economic performance face substantial challenges. They must set sufficiently stringent standards and ensure that monitoring and enforcement are strict enough to exclude poorly performing producers. In addition, they must offer high enough price premiums or new customers plentiful enough to offset the costs of certification and attract a considerable number of applicants. Even if these two challenges are met, however, certification schemes still can be undermined by selection effects. Commodity producers already meeting certification standards have strong incentives to join certification programs; they need not make additional investments to qualify and can obtain price premiums and other benefits. But certification programs that mainly attract such producers will have a limited additional effect on producer behavior and few environmental, social, or economic benefits.

Although a fast-growing academic literature examines sustainability certification, little is known about whether certification actually affects producers' environmental, social, and economic performance. Relatively few studies evaluate the effects of certification, and most of these few do not correct for selection effects or have other methodological limitations.

We sought to assess the evidence base on the environmental, social, and economic effects of sustainability certification in five economic sectors in which such certification is particularly prominent. We did this by identifying studies of certification effects that meet certain criteria (defined below), classifying them on the basis of whether they use methods likely to generate credible results, summarizing their findings, and considering how these findings might guide future research.

Methods

To identify studies of sustainability certification, we searched library catalogs, lists of references in related articles, and digital databases, including Econlit, Google, Science Direct, Scirus, and Scopus. In constructing electronic searches, we identified as many studies as possible by using a variety of search terms—including *certification*, *ecolabel*, and *label*—alone and in combinations with the names of the economic sectors on which we focused.

Criteria for Inclusion

Studies included in the evidence base met four criteria. First, they examined certification in one of five economic sectors in which certification is particularly prominent: bananas, coffee, fish products, forest products, and tourism. Hence, we excluded studies of manufacturing and staple crops such as corn and soybeans. Second, they were published in peer-reviewed journals or books published by third parties (i.e., not by the institution with which the authors were affiliated). This criterion was meant to enforce some measure of quality control and to make our literature search manageable. Third, they focused specifically on identifying environmental, social, and economic effects of certification at the level of the primary producer of goods and services. Hence, we excluded studies of buyers of certified products, the drivers of certification, and design recommendations for certification programs. And finally, they presented an original ex post analysis of an actual experience with certification. Hence, we excluded general discussions of certification, studies that were based on secondary sources, and models simulating, rather than measuring, certification effects.

Method Categories

We grouped studies that met our four criteria into three categories on the basis of the methods used. The first category, which we term *rigorous*, was quantitative studies that constructed a credible counterfactual outcome and could therefore be considered a test of the causal effect of certification. A counterfactual outcome is an estimate of the certified producers' outcomes had they not been certified. It is the baseline against which the environmental, social, and economic performance of certified producers is measured. The second category, which we term *moderately rigorous*, was quantitative studies that did not construct a credible counterfactual. The third category was qualitative studies that did not construct a credible counterfactual.

Constructing a Counterfactual

To generate evidence on the causal effects of certification on environmental, social, and economic outcomes—rather than just statistical correlations between certification and such outcomes—an evaluation must construct a credible counterfactual. The effect of certification is defined as the difference between the actual outcome and the counterfactual outcome.

Most researchers evaluating the effects of certification programs use simplistic counterfactual outcomes that likely bias their results. One common approach, applicable to panel (i.e., time series) data on certified producers' outcomes, is to use certified producers' precertification outcomes as the counterfactual outcome. The implicit identifying assumption (i.e., the key assumption needed for an unbiased estimate of the effect of certification) is that if certified producers had not been certified, their outcomes, on average, would have stayed the same. This assumption is violated when outcomes change during the study period because of contemporaneous confounders (i.e., factors unrelated to certification that affect outcomes). For example, if the counterfactual outcome in a study of the economic effects of fair-trade coffee is certified producers' precertification household income, then effect is measured as the difference between precertification and postcertification average household income. If this difference is positive, statistically significant, and large, the evaluator may conclude certification increased average household income. But estimates of certification effect derived in this manner are biased upward, and the evaluator's finding of a causal effect is misleading if producers' household incomes rose after certification for reasons unrelated to certification. These reasons might include increases in international prices for coffee, advantageous weather, or advances in processing and marketing.

A second common approach, applicable to cross-sectional (i.e., snapshot) data on certified and uncertified producers outcomes, is to use all uncertified producers' outcomes to estimate the counterfactual outcome. In other words, all uncertified producers serve as a control group. The implicit identifying assumption is that if certified producers had not been certified, their outcomes would have been the same, on average, as those of uncertified producers. This assumption is violated when producers decide to obtain certification or are selected by certifiers on the basis of characteristics that affect the measured outcomes. Unless corrected for, such selection effects can bias estimation of the counterfactual and, therefore, estimation of certification effects; this problem is known as selection bias. For example, if the counterfactual outcome in a study of the effect of organic coffee certification on environmental outcomes is soil erosion on all uncertified producers' farms, then the effect is measured as the difference between average soil erosion for all certified and all uncertified farms. If this difference is negative, significant, and large, the evaluator may conclude certification drove reductions in soil erosion. Estimates of the effect of certification derived in this manner are biased upward, and the evaluator's finding of a causal effect is misleading if producers with relatively less soil erosion disproportionately decided to obtain organic certification. This could happen if producers who had already adopted soil-conservation measures sought organic certification because they knew they would not have to invest in additional conservation measures to meet certification standards.

Four principal approaches to constructing a credible counterfactual have been used (Frondel & Schmidt 2005; Ferraro 2009; Greenstone & Gayer 2009). One requires randomized or experimental design of certification projects, also known as field experiments (Burtless 1995; Duflo & Kremer 2005). For certification projects, this amounts to randomly selecting producers to receive certification from among a group of qualified and interested candidates. The outcome for the uncertified control group is then used as the counterfactual outcome for a randomly constituted sample of certified producers. This approach requires building evaluation into the design of the certification initiative.

The three other main approaches to constructing a credible counterfactual (all quasi-experimental) entail use of statistical techniques to correct for potential selection bias in data from a certification initiative in which certification has not been assigned randomly. The first such method, which is applicable to either cross-sectional or panel data, is to use matching, wherein certified producers are paired with uncertified producers that have very similar, if not identical, observable characteristics that plausibly affect outcomes (Morgan & Harding 2006; Caliendo & Kopeining 2008). Outcomes for this matched control sample serve as the counterfactual outcome. For example, in a study of the soil-erosion effects of organic coffee certification, certified producers would be matched with uncertified producers of similar size, education, and previous history of adopting conservation practices. Soil erosion for this matched control group would be used as the counterfactual. This method depends on the dual identifying assumptions that no unobservable characteristics of the producers in question (e.g., management skill) affect both producers' decisions to participate in the certification program and measured outcomes and that all uncertified producers in the matched control sample have characteristics that make them suitable for certification. Various techniques are available for matching producers when the number of observable characteristics is large.

The second quasi-experimental method, instrumental variables, is applicable to either cross-sectional or panel data. The method takes advantage of known correlations between certification and instruments, which are

characteristics of certified producers that plausibly affect both probability of certification, but do not directly affect the social, economic, or environmental outcomes. In other words, instruments do not affect outcomes except through the probability of certification (Angrist & Krueger 2001; Morgan & Winship 2007). These instruments can be used to control for selection bias. For example, in a study of the environmental effects of organic coffee certification on farm income, one might use distance from the farm to a certifying agency's headquarters as an instrument. One drawback of this approach is that credible instruments are not easy to identify. A second limitation is that this method measures only the effect of certification on the subset of producers whose decision to certify is affected by the instrument. As a result, this method estimates an effect of certification (technically, a local average treatment effect) that is conceptually different from estimates derived with other methods (Morgan & Winship 2007).

The third quasi-experimental method, difference in differences, requires panel data (Card & Krueger 1994; Bertrand et al. 2004). It entails testing whether changes in certified producers' outcomes measured before and after certification are significantly different from changes in the same outcomes during the same period for similar uncertified producers. This method is analogous to the before-after-control-impact approach used in natural science. The identifying assumption of this method is that without certification, changes in outcomes occurring during the study period, which starts prior to certification and ends after it, would be the same for certified and uncertified producers.

The studies in the evidence base that construct a counterfactual rely almost exclusively on quasi-experimental matching. Only Bolwig et al. (2009) used instrumental variables. In no studies was an experimental-design or a difference-in-difference approach used.

A final issue about methods concerns the evaluation of long- versus short-term effects. Most researchers evaluating certification effects either explicitly or implicitly tested for effects that occur within a few years of certification. For example, a before–after analysis explicitly tests for such effects, whereas cross-sectional analyses implicitly test for them. In cases like these, evaluations may not take full account of benefits that occur over a longer period. Such benefits may be significant, particularly in the case of ecological systems that change slowly. For example, the full effect of some timber-harvesting practices on forest regeneration may become apparent only after many years. Evaluators of certification effects should be explicit about such limitations, although some are not.

Results

Forty-six published studies met our criteria for inclusion in the evidence base (Table 1). Of these, 11 were rigorous, 18 were moderately rigorous, and 17 were qualitative (Table 2). Of the rigorous studies, all attempted to identify certification effects by using either propensity score matching or regression analyses to analyze cross-sectional data. In only two studies, both moderately

Table 1. Studies of the effects of sustainability certification included in our evidence base by economic sector and study category.

	Rigorous: quantitative, credible counterfactual	*Moderately rigorous: quantitative, no credible counterfactual*	*Qualitative: no credible counterfactual*
Bananas	Fort & Ruben 2008*a* Ruben & van Schendel 2008 Zúñiga-Arias & Sáenz Segura 2008	Melo & Wolf 2007 Ruben et al. 2008	Shreck 2002
Coffee	Arnould et al. 2009 Bolwig et al. 2009 Fort & Ruben 2008*b* Lyngbaek et al. 2001 Sáenz Segura & Zúñiga-Arias 2008	Bacon 2005 Bacon et al. 2008 Jaffee 2008 Kilian et al. 2004 Martínez-Torres 2008 Méndez et al. 2010 Philpott et al. 2007 Valkila 2009	Bray et al. 2002 Parrish et al. 2005 Raynolds et al. 2004 Sick 2008 Utting 2009 Utting-Chamorro 2005 Valkila & Nygren 2010
Fish products	none	Hicks & Schnier 2008 Ward 2008	Biao et al. 2005 Phillips et al. 2003
Forest products	none	Gulbrandsen 2005 Hartsfield & Ostermeier 2003 Humphries & Kainer 2006 Kukkonen et al. 2008 Nebel et al. 2005 Rickenbach & Overdevest 2006	Cubbage et al. 2003 Ebeling & Yasue 2009 Markopoulos 2003 McDaniel 2003 Quevedo 2007
Tourism	Rivera 2002 Rivera & de Leon 2004 Rivera et al. 2006	none	Ayuso 2007 Goodman 2000

Table 2. Studies on the effects of sustainability certification included in our evidence base by economic sector, method category, type of certification[a], and environmental focus[b].

	Rigorous: quantitative, credible counterfactual			*Moderately rigorous: quantitative, no credible counterfactual*			*Qualitative: no credible counterfactual*			*All*	
	all	*type*	*environmental focus*	*all*	*type*	*environmental focus*	*all*	*type*	*environmental focus*	*total*	*environmental focus*
Bananas	3	FT (3) Organic (1)	0	2	FT (2) Organic (1) Eurepgap (1) RA (1)	1	1	FT (1)	1	6	2
Coffee	5	FT (3) Organic (3)	0	8	FT (7) Organic (6) RA (1) Utz (1)	3	7	FT (6) Organic (1)	3	20	6
Fish products	0	n/a	0	2	MSC (1) DS (2)	2	2	Organic (1) MSC (1)	2	4	4
Forest products	0	n/a	0	6	FSC (6) RA (1)	4	5	FSC (3) SFI (1) RA (1)	3	11	7
Tourism	3	CST (1) SSP (2)	2	0	n/a	0	2	Multiple	2	5	4
Total	11		2	18		10	17		11	46	23

[a] *In many studies, producers with multiple certifications were evaluated.*
[b] *Includes studies in which both environmental effects and social and economic effects were analyzed.*
CST, certification for sustainable tourism; DS, dolphin safe; FSC, Forest Stewardship Council; FT, fair trade; MSC, Marine Stewardship Council; RA, Rainforest Alliance; SFI, Sustainable Forest Initiative; SSP, Sustainable Slopes Program.

rigorous, was an attempt made to identify certification effects with a before–after comparison (Hartsfield & Ostermeier 2003; Hicks & Schnier 2008). No studies compared certified and noncertified producers both before and after certification (i.e., difference in differences).

The economic sectors most heavily represented in the evidence base were coffee and forest products. Of the 46 studies in the evidence base, 20 focused on coffee, 11 on forest products, six on bananas, five on tourism, and four on fish products (Table 2).

Half the studies examined environmental outcomes, often in combination with social and economic outcomes. The nature of the environmental outcomes varied depending on the economic sector and certification. In general, however, authors used both direct and indirect measures of environmental "health." For example, authors of studies of coffee used direct measures such as species richness, rates of soil erosion rates, and humus depth (Philpott et al. 2007; Martinez-Torres 2008), and indirect measures such as soil-conservation practices (Bray et al. 2002; Jaffee 2008). Authors of fisheries studies used direct measures such as size of fish stocks and indirect measures such as dolphin bycatch (Ward 2008). And authors of studies of forest products used direct measures such as species richness, the rate of forest regeneration, and the rate of deforestation (Nebel et al. 2005; Kukkonen et al. 2008) and indirect measures such as the adoption of sustainable management practices (Hartsfield & Ostermeier 2003). Although environmental effects featured prominently in the 46 studies in the evidence base, they were the focus of only two of the 11 rigorous studies, both of tourism (Table 2).

Counting the studies that focused on specific types of certification is problematic because in many studies more than one type of certification was analyzed. For example, many of the coffee studies examined producers with both fair-trade and organic certifications. That said, two certification schemes were particularly well represented: fair-trade certification and Forest Stewardship Council certification (Table 2).

The 11 rigorous studies provided very weak evidence for the hypothesis that sustainable certification has positive environmental, social, or economic effects at the producer level. Only four of the 11 provided evidence of such benefits (Table 3). Of these four, all tested for economic effects. In two of these four studies, both on coffee, the authors remark that the benefits are either idiosyncratic or inconsistent (Arnould et al. 2009; Bolwig et al. 2009). The results of the remaining seven rigorous studies (two studies of environmental effects and five of

Table 3. Rigorous studies (quantitative and construct a credible counterfactual) in our evidence base that show a positive socioeconomic or environmental effect.

Economic sector	*Number*	*Positive social and economic effect*	*Positive environmental effect*
Bananas	3	1	n/a
Coffee	5	2	n/a
Fish products	0	n/a	n/a
Forest products	0	n/a	n/a
Tourism	3	1	0
Total	11	4	0

social and economic effects) did not show that certification benefits producers.

Bananas

We found six published articles on the certification of bananas as sustainable that met our criteria for inclusion in the evidence base (Table 1). All six focused on fair-trade certification and therefore on the effect of certification on producers' social and economic status (the main concern of such certification). Except for Melo and Wolf (2007) and Shreck (2002), these six articles are in Ruben (2008), an edited volume on fair-trade certification. We classified three of the studies as rigorous (quantitative analyses that constructed a credible counterfactual), two as moderately rigorous (quantitative studies that did not construct a credible counterfactual), and one as a qualitative study (did not construct a credible counterfactual).

These six studies provided mixed evidence on the effect of fair-trade certification on producers' social and economic status. Of the rigorous studies, only Fort and Ruben (2008*a*) found that certification provides substantial benefits. Specifically, they found that certification improves producers' productivity in Peru by a substantial amount, presumably by generating on-farm investment. In the other rigorous studies (Ruben & van Schendel 2008; Zúñiga-Arias & Sáenz Segura 2008), the authors concluded that certification has little or no effect on producers' social and economic status in Ghana and Costa Rica. The results of the two moderately rigorous studies (Melo & Wolf 2007; Ruben et al. 2008) showed that in Ecuador, certified producers have higher social and economic status and that their farms have better environmental performance than uncertified farms. Recall, however, that failure to control for selection bias can lead to an overly optimistic assessment of certification benefits. Finally, in the Dominican Republic, Shreck (2002), who conducted the one qualitative study, found mixed evidence of certification benefits.

Coffee

Among our five study sectors, the greatest number of published studies of certification effects, 20 in all, examined coffee (Table 1). We classified five of these studies as rigorous, eight as moderately rigorous, and seven as qualitative. These 20 studies do not provide compelling evidence that coffee certification has significant social, economic, or environmental benefits. Of the rigorous studies, only Arnould et al. (2009) and Bolwig et al. (2009) showed that certification has significant social or economic benefits. Both studies include substantial caveats. Arnould et al. (2009) found that even though certification generates statistically and economically significant increases in price and volume of coffee sold, it does not affect education or health, whereas Bolwig et al. (2009) cautioned that the social and economic benefits they identify are mainly due to an idiosyncratic design feature of the certification program they studied. The results of the other rigorous studies (Lyngbaek et al. 2001; Fort & Ruben 2008*b*; Sáenz Segura & Zúñiga-Arias 2008) showed that certification has either minimal social and economic benefits or net costs.

Even though one would expect a failure to control for selection bias to spur unduly positive assessments of certification benefits, results of all the moderately rigorous studies (Kilian et al. 2004; Bacon 2005; Philpott et al. 2007; Bacon et al. 2008; Jaffee 2008; Martínez-Torres 2008; Valkila 2009; Méndez et al. 2010) showed no strong correlation between certification and social, economic, and environmental benefits. In six of these studies, all except Kilian et al. (2004) and Philpott et al. (2007), fair-trade and organic certification in Latin America were examined, and the results of all showed that even though certification may boost farm-level prices, reduce price variability, and have other benefits, it does not have significant positive effects on important social and economic measures of household welfare. Of the two remaining moderately rigorous studies, Kilian et al. (2004) concluded that product quality, not certification, drives price premiums for certified coffee, and Philpott et al. (2007) found that ant and bird species richness was no higher on certified farms than on uncertified ones.

It is only among the seven qualitative studies of coffee, all of which examined fair-trade and organic certification, that we found broad support for the proposition that certification has significant benefits. Bray et al. (2002), Parrish et al. (2005), Raynolds et al. (2004), and Utting (2009) found that certification has an overall positive effect, whereas results of the other three studies are more mixed.

Fish Products

We found four published studies of the social, economic, or environmental effects of fish-product certification (Table 1). We classified Hicks and Schnier (2008) and Ward (2008) as moderately rigorous and Biao et al. (2005) and Phillips et al. (2003) as qualitative. The results of these studies regarding certification benefits are mixed. Ward (2008), which examines dolphin-safe and Marine Stewardship Council (MSC) certification globally, found that certification does not increase fish stocks. In their analysis of the first several years of implementation of MSC certification, Phillips et al. (2003) found there are financial benefits to producers, but also found a lack of evidence of ecological benefits. Hicks and Schnier (2008), which focuses on dolphin-safe certification, and Biao et al. (2005), which focuses on organic aquaculture certification, document mostly positive results of certification. The methods of the last two studies have significant limitations, however. Hicks and Schnier (2008) purport to show that dolphin-safe certification increases

the environmental performance of U.S.-flagged ships in the eastern tropical Pacific, but they used what amounts to a before–after comparison, did not control for self-selection into and out of the U.S. fleet, and examined a mandatory certification scheme. Biao et al. 2005, who examined organic shrimp production in China, used purely qualitative data from just four aquaculture operations.

Forest Products

We found 11 published studies of the effects of forest-product certification (zero, rigorous; six, moderately rigorous; five, qualitative) (Table 1). Nine of these studies were of Forest Stewardship Council (FSC) certification. As a group, the results suggest certification has few direct economic benefits, such as higher producer prices or improved market access, and few direct environmental benefits, such as reduced deforestation or improved regeneration.

Of the moderately rigorous studies, Hartsfield and Ostermeier (2003), Humphries and Kainer (2006), and Rickenbach and Overdevest (2006) focus mainly on economic, not environmental, benefits. Hartsfield and Ostermeier (2003) and Rickenbach and Overdevest (2006) found that certification does not improve producer prices or expand market access, but that it does provide less-tangible benefits, such as building the human capital and reputation of certified enterprises. Similarly, Gulbrandsen (2005), Kukkonen et al. (2008), and Nebel et al. (2005) focused on environmental benefits and found that although certification does not reduce deforestation, speed forest regeneration, or improve forest status, it may improve forest management.

The same conclusions emerge from the qualitative studies. Cubbage et al. (2003), Markopoulos (2003), McDaniel (2003), and Quevedo (2007) focused on economic effects and found that certification's direct economic benefits are negligible, although less-tangible benefits may be significant. In their analysis of environmental effects, Ebeling and Yasue (2009) conclude that certification is unlikely to stem deforestation in countries such as Bolivia with limited capacity for forest governance.

Tourism

We found five studies that focused on the environmental or social and economic effects of certification on tourism (Table 1). We categorized Rivera (2002), Rivera and de Leon (2004), and Rivera et al. (2006) as rigorous, and Ayuso (2007) and Goodman (2000) as qualitative. Rivera (2002) found that hotel certification in Costa Rica generates significant price premiums and therefore presumably has an economic benefit. However, Rivera and de Leon (2004) and Rivera et al. (2006) demonstrate that certification of ski slopes in the United States has not improved environmental performance and may even have generated environmental costs. Goodman (2000), which is based on a single in-depth case study, found mostly positive results of certification, whereas Ayuso's (2007) results are equivocal. Ayuso (2007) concludes that compared with other voluntary instruments, certification is relatively effective, but because producers already meeting certification standards disproportionately choose to participate, certification typically does not change behavior.

Discussion

Our literature review suggests that more studies of the causal effects of certification (i.e., studies that construct a credible counterfactual) are needed. Several more specific gaps in the literature are also apparent. First, certain economic sectors have not been examined. The 11 studies of causal effects focus on bananas, coffee, and tourism. Studies on fish and forest products are lacking. Second, certain types of certification have not been examined. Eight of the 11 studies of causal effects examined fair-trade and organic certification. Studies on other types of certification, such as Sustainable Agriculture Network and Euregap, are lacking. Finally, nine of the 11 studies of causal effects of certification focus on social and economic rather than environmental effects.

How can certification projects be designed to generate more credible evidence of certification effects? An array of institutions, including nongovernmental organizations, national governments, and multilateral and bilateral international organizations, fund projects that purport to either expand participation in existing certification systems or develop new systems. However, few include rigorous evaluations. We recommend the following to help fill this gap.

First, require projects to clearly articulate the general and specific objectives of certification and to spell out measurable indicators of success, such as the percent increase in species richness of native birds and insects in an area where a certified crop is grown. Explicit objectives and indicators of success will facilitate project evaluation and strengthen incentives to design and implement certification projects in a manner that generates these effects.

Second, require that projects include a detailed plan for project evaluation and a budget sufficient to implement it.

Third, require that all phases of project evaluations—including design, implementation, and dissemination—be conducted by an independent third party. Allowing certification programs to evaluate their own efforts may create conflicts of interest.

Fourth, design evaluations to maximize opportunities for knowledge creation. Several recent articles discuss design principles for evaluation of environmental projects (Frondel & Schmidt 2005; Ferraro 2009; Greenstone & Gayer 2009). Here, we list recommendations as they apply to the design of certification programs. Evaluations

should not be solely ex post exercises. Rather, they should be planned at the same time as the certification project itself and built into project design. Collect outcome data for certified producers (the treatment group) and noncertified producers (the control group), ideally before and after certification. Collecting ex post data from certified producers is generally straightforward and of low cost. More difficult—but critically important—is collecting data from uncertified producers and baseline data from both groups.

Furthermore, when practical, incorporate a randomized design that generates a control sample of noncertified producers that is very similar to certified producers. This can be accomplished by, for example, compiling a sampling frame of matched locations (e.g., towns, villages) targeted for certification and then randomly selecting a subsample where certification is actually promoted. Another means of implementing randomized design is to delay the award of certification by one or two years for a random sample of producers that have successfully applied for certification. After a control group is constructed, introduce a second layer of randomization when practical to create additional knowledge about certification drivers and effects. This entails creating several distinct treatment groups by, for example, randomly assigning different types of certification (e.g., Rainforest Alliance and Bird Friendly for coffee producers) across applicants to gauge their relative effects; by randomly varying the amount and type of certification subsidies (financial and technical) provided to producers to gauge their effectiveness; or by allowing for slight changes in certification requirements across randomly selected applicants.

Fifth, train project personnel in the principles of project evaluation to facilitate cooperation with third-party evaluators.

Finally, promote transparency in the evaluation process and plan and budget for dissemination of the evaluation results. Transparency helps minimize opportunities for gaming and builds consumers' trust in the institution of third-party certification. Widespread dissemination via websites and academic publication maximizes the benefit of evaluation.

Acknowledgments

This research was funded by the United Nations Environment Program in cooperation with the Science and Technical Advisory Panel (STAP) of the Global Environment Facility and by the Swedish Research Council Formas. We are grateful to D. Cunningham and P. Ferraro of STAP and to seminar participants at the University of Michigan for helpful comments; A. Chacón, J. Ferris, D. McLaghlin, and A. Stern for research assistance; and the editors and reviewers of this journal for helpful comments and suggestions.

Supporting Information

A bibliography of 213 published studies considered for inclusion in the evidence base is available online (Appendix S1). The authors are solely responsible for the content and functionality of these materials. Queries (other than absence of the material) should be directed to the corresponding author.

Literature Cited

Angrist, J., and A. Krueger. 2001. Instrumental variables and the search for identification: from supply and demand to natural experiments. Journal of Economic Perspectives **15:**69–87.

Arnould, E., A. Plastina, and D. Ball. 2009. Does fair trade deliver on its core value proposition? Effects on income, educational attainment, and health in three countries. Journal of Public Policy and Marketing **28:**186–201.

Ayuso, S. 2007. Comparing voluntary policy instruments for sustainable tourism: the experience of the Spanish hotel sector. Journal of Sustainable Tourism **15:**144–159.

Bacon, C. 2005. Confronting the coffee crisis: can fair trade, organic and specialty coffees reduce small-scale farmer vulnerability in Northern Nicaragua? World Development **33:**497–511.

Bacon, C. M., V. Méndez, M. Gómez, D. Stuart, and S. Flores. 2008. Are sustainable coffee certifications enough to secure farmer livelihoods? The millennium development goals and Nicaragua's fair trade cooperatives. Globalizations **5:**259–274.

Bertrand, M., E. Duflo, and S. Mullainathan. 2004. How much should we trust differences-in-differences estimates? Quarterly Journal of Economics **119:**249–275.

Biao, X., J. Li, and X. Wang. 2005. Report on sustainable aquaculture. Asia-Pacific Marine Finfish Aquaculture Network. Aquaculture Asia **10:**11–17.

Bolwig, S., P. Gibson, and S. Jones. 2009. The economics of smallholder organic contract farming in tropical Africa. World Development **37:**1094–1104.

Bray, D., J. Sanchez, and E. Murphy. 2002. Social dimensions of organic coffee production in Mexico: lessons for eco-labeling initiatives. Society and Natural Resources **15:**429–446.

Burtless, G. 1995. The case for randomized field trials in economic and policy research. Journal of Economic Perspectives **9:**63–84.

Caliendo, M., and S. Kopeining. 2008. Some practical guidance for the implementation of propensity score matching. Journal of Economic Surveys **32:**31–72.

Card, D., and A. Krueger. 1994. Minimum wages and employment: a case study of the fast food industry in New Jersey and Pennsylvania. American Economic Review **84:**772–784.

Cubbage, F., S. Moore, J. Cox, L. Jervis, J. Edeburn, D. Richter, W. Boyette, M. Thompson, and M. Chesnutt. 2003. Forest certification of state and university lands in North Carolina: a comparison. Journal of Forestry **101:**26–31.

Duflo, E., and M. Kremer. 2005. Use of randomization in the evaluation of development effectiveness. Pages 205–232 in G. Pitman, O. Feinstein, and G. Ingram, editors. Evaluating development effectiveness. Transaction Publishers, New Brunswick, New Jersey.

Ebeling, J., and M. Yasue. 2009. The effectiveness of market-based conservation in the tropics: forest certification in Ecuador and Bolivia. Journal of Environmental Management **90:**1145–1153.

Eilperin, J. 2010. Environmental certification becoming increasingly crowded and contested field. Washington Post 3 May:A4.

Ferraro, P. 2009. Counterfactual thinking and impact evaluation in environmental policy. New Directions for Evaluation **122:** 75–84.

Fort, R., and R. Ruben. 2008*a*. The impact of fair trade on banana producers in northern Peru. Pages 49–73 in R. Ruben, editor. The impact of fair trade. Wageningen Academic Publishers, Wageningen, The Netherlands.

Fort, R., and R. Ruben. 2008*b*. The impact of fair trade on coffee producers in Peru. Pages 75–98 in R. Ruben, editor. The impact of fair trade. Wageningen Academic Publishers, Wageningen, The Netherlands.

Frondel, M., and C. Schmidt. 2005. Evaluating environmental programs: the perspective of modern evaluation research. Ecological Economics **55**:515–26.

Giovannucci, D., and S. Ponte. 2005. Standards as a new form of social contract? Sustainability initiatives in the coffee industry. Food Policy **30:**284–301.

Goodman, A. 2000. Implementing sustainability in service operations at Scandic hotels. Interfaces **30:**202–214.

Greenstone, M., and T. Gayer. 2009. Quasi-experimental and experimental approaches to environmental economics. Journal of Environmental Economics and Management **57**:21–44.

Gulbrandsen, L. H. 2005. The effectiveness of non-state governance schemes: a comparative study of forest certification in Norway and Sweden. International Environmental Agreements **5:** 125–149.

Hartsfield, A., and D. Ostermeier. 2003. The view from FSC-certified land mangers. Journal of Forestry **101:**32–36.

Hicks, R. L., and K. Schnier. 2008. Eco-labeling and dolphin avoidance: a dynamic model of tuna fishing in the eastern Tropical Pacific. Journal of Environmental Economics and Management **56:**103–116.

Humphries, S., and K. Kainer. 2006. Local perceptions of forest certification for community-based enterprises. Forest Ecology and Management **235:**30–43.

Jaffee, D. 2008. Better, but not great: the social and environmental benefits and limitations of Fair Trade for indigenous coffee producers in Oaxaca, Mexico. Pages 196–222 in R. Ruben, editor. The impact of fair trade. Wageningen Academic Publishers, Wageningen, The Netherlands.

Kilian, B., C. Jones, L. Pratt, and A. Villalobos. 2004. Can the private sector be competitive and contribute to development through sustainable agricultural business? A case study of coffee in Latin America. International Food and Agribusiness Management Review **7:**21–45.

Kukkonen, M., H. Rita, S. Hohnwald, and A. Nygren. 2008. Treefall gaps of certified, conventionally managed and natural forests as regeneration sites for Neotropical timber trees in northern Honduras. Forest Ecology and Management **255:**2163–2176.

Lyngbaek, A., R. Muschler, and F. Sinclair. 2001. Productivity and profitability of multistrata organic versus conventional coffee farms in Costa Rica. Agroforestry Systems **53:**205–213.

Markopoulos, M. 2003. The role of certification in community based forest enterprise. Pages 105–131 in E. Medinger, C. Elliot, and G. Oesten, editors. Social and political dimensions of forest certification. Fortsbuch, Rebagen Oberwinter, Germany.

Martínez-Torres, M. E. 2008. The benefits and sustainability of organic farming by peasant coffee farmers in Chiapas, Mexico. Pages 99–126 in C. Bacon, V. Méndez, S. Gliessman, D. Goodman, and J. Fox, editors. Confronting the coffee crisis: fair trade, sustainable livelihoods, and ecosystems in Mexico and Central America. MIT Press, Cambridge, Massachusetts.

McDaniel, J. 2003. Community-based forestry and timber certification in southeast Bolivia. Small-scale Forest Economics, Management and Policy **2:**37–341.

Melo, C. J., and S. A. Wolf. 2007. Ecocertification of Ecuadorian bananas: prospects for progressive north–south linkages. Studies in Comparative International Development **42:**256–278.

Méndez, V. E., C. Bacon, M. Olson, S. Petchers, D. Herrador, C. Carranza, L. Trujillo, C. Guadarrama-Zugasti, A. Cordón, and A. Mendoza. 2010. Effects of fair trade and organic certifications on small-scale coffee farmer households in Central America and Mexico. Renewable Agriculture and Food Systems **25:**236–251.

Morgan, S., and D. Harding. 2006. Matching estimators of causal effects: prospects and pitfalls in theory and practice. Sociological Methods and Research **35:**3–60.

Morgan, S., and C. Winship. 2007. Counterfactual and causal inference: methods and principles for social research. Cambridge University Press, New York.

Nebel, G., L. Quevedo, J. Bredahl Jacobsen, and F. Helles. 2005. Development and economic significance of forest certification: the case of FSC in Bolivia. Forest Policy and Economics **7:**175–186.

Parrish, B., V. Luzadis, and W. Bentley. 2005. What Tanzania's coffee farmers can teach the world: a performance-based look at the fair trade-free trade debate. Sustainable Development Journal **13:**177–189.

Phillips, B., T. Ward, and C. Chaffee. 2003. Eco-labeling in fisheries: What is it all about? Blackwell Science, Oxford, United Kingdom.

Philpott, S., P. Bichier, R. Rice, and R. Greenberg. 2007. Field testing ecological and economic benefits of coffee certification programs. Conservation Biology **21:**975–985.

Quevedo, L. 2007. Forest certification in Bolivia. Pages 303–336 in B. Cashore, F. Gale, E. Meidinger, and D. Newsom, editors. Confronting sustainability: forest certification in developing and transitioning countries. Yale School of Forestry and Environmental Studies Press, New Haven, Connecticut.

Raynolds, I., D. Murray, and P. Taylor. 2004. Fair trade coffee: building producer capacity via global networks. Journal of International Development **16:**1109–1121.

Rice, R., and J. Ward. 1996. Coffee, conservation and commerce in the western hemisphere: How individuals and institutions can promote ecologically sound farming and forest management in northern Latin America. Smithsonian Migratory Bird Center, Washington, D.C.

Rickenbach, M., and C. Overdevest. 2006. More than markets: assessing Forest Stewardship Council (FSC) certification as a policy tool. Journal of Forestry **104:**143–147.

Rivera, J. 2002. Assessing a voluntary environmental initiative in the developing world: the Costa Rican certification for sustainable tourism. Policy Sciences **35:**333–360.

Rivera, J., and P. de Leon. 2004. Is greener whiter? The Sustainable Slopes Program and the voluntary environmental performance of western ski areas. Policy Studies Journal **32:**417–437.

Rivera, J., P. de Leon, and C. Koerber. 2006. Is greener whiter yet? The Sustainable Slopes Program after five years. Policy Studies Journal **34:**195–224.

Ruben, R., editor. 2008. The impact of fair trade. Wageningen Academic Publishers, Wageningen, The Netherlands.

Ruben, R., L. Clercx, D. Cepeda, and T. de Hopp. 2008. Fair trade impact of banana production in El Guabo Association, Ecuador: a production function analysis. Pages 155–167 in R. Ruben, editor. The impact of fair trade. Wageningen Academic Publishers, Wageningen, The Netherlands.

Ruben, R., and L. van Schendel. 2008. The impact of fair trade in banana plantations in Ghana: income, ownership and livelihoods of banana workers. Pages 137–153 in R. Ruben, editor. The impact of fair trade. Wageningen Academic Publishers, Wageningen, The Netherlands.

Sáenz Segura, F., and G. Zúñiga-Arias. 2008. Assessment of the effect of fair trade on smallholder producers in Costa Rica: a comparative study in the coffee sector. Pages 117–135 in R. Ruben, editor. The impact of fair trade. Wageningen Academic Publishers, Wageningen, The Netherlands.

Shreck, A. 2002. Just bananas? Fair trade banana production in the Dominican Republic. International Journal of Sociology of Agriculture and Food **10:**13–23.

Sick, D. 2008. Coffee, farming families, and fair trade in Costa Rica: New markets, same old problems? Latin American Research Review **43:**193–208.

Utting, K. 2009. Assessing the impact of fair trade coffee: towards an integrative framework. Journal of Business Ethics **86:**127–149.

Utting-Chamorro, K. 2005. Does fair trade make a difference? The case of small coffee producers in Nicaragua. Development in Practice **15:**584–599.

Valkila, J. 2009. Fair trade organic coffee production in Nicaragua: Sustainable development or a poverty trap? Ecological Economics **68:**3018–3025.

Valkila, J., and A. Nygren. 2010. Effects of fair trade certification on coffee farmers, cooperatives, and laborers in Nicaragua. Agriculture and Human Values **27:**321–333.

Ward, T. 2008. Barriers to biodiversity conservation in marine fishery certification. Fish and Fisheries **9:**169–177.

Zúñiga-Arias, G., and F. Sáenz Segura. 2008. The impact of fair trade in banana production of Costa Rica. Pages 99–116 in R. Ruben, editor. The impact of fair trade. Wageningen Academic Publishers, Wageningen, The Netherlands.