Review

Producer-Level Benefits of Sustainability Certification

ALLEN BLACKMAN*† AND JORGE RIVERA†‡

†Environment for Development Center for Central America, Turrialba, Costa Rica

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, 9 examined the economic effects and 2 the environmental effects of certification. The results of 4 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 credible data 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 estas iniciativas tienen tales beneficios. Identificamos estudios revisados por pares, ex post y a nivel de productores 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 cómo 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. Los métodos utilizados en 11 estudios generaron resultados creíbles. De estos 11 estudios, 9 examinaron los efectos económicos y 2 los efectos ambientales de la certificación. Los resultados de 4 de los 11 estudios, todos examinando efectos económicos, mostraron que la certificación tiene beneficios a nivel de productores. Por lo tanto, la evidencia para soportar la hipótesis de que la certificación beneficia al ambiente o a los productores es limitada. Se deben generar más datos creíbles mediante la incorporación de evaluaciones independientes más rigurosas en el diseño e implementación de proyectos que promueven la certificación.

Palabras Clave: ecocertificación, eco-etiquetas, paisaje dominado por humanos
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 2 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 5 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 4 criteria. First, they examined certification in 1 of 5 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 4 criteria into 3 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 the 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 with relatively less soil erosion disproportionately decided to obtain organic certification. This could happen if a disproportionate number of 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 3 other main approaches (all quasi-experimental) to constructing a credible counterfactual 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 producers’ decision 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 the probability of certification, but do not directly affect the social, economic, or environmental outcome. 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 through 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 2 studies, both moderately 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).

<table>
<thead>
<tr>
<th>Economic Sector</th>
<th>Rigorous: quantitative, credible counterfactual</th>
<th>Moderately rigorous: quantitative, no credible counterfactual</th>
<th>Qualitative: no credible counterfactual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bananas</td>
<td>Fort &amp; Ruben 2008a</td>
<td>Melo &amp; Wolf 2007</td>
<td>Shreck 2002</td>
</tr>
<tr>
<td></td>
<td>Ruben &amp; van Schendel 2008</td>
<td>Ruben et al. 2008</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zúñiga-Arias &amp; Sáenz Segura 2008</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fort &amp; Ruben 2008b</td>
<td>Jaffee 2008</td>
<td>Raynolds et al. 2004</td>
</tr>
<tr>
<td></td>
<td>Lyngbaek et al. 2001</td>
<td>Kilian et al. 2004</td>
<td>Sick 2008</td>
</tr>
<tr>
<td>Fish products</td>
<td>None</td>
<td>Méndez et al. 2010</td>
<td>Utting-Chamorro 2005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Philpott et al. 2007</td>
<td>Valkila &amp; Nygren 2010</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Valkila 2009</td>
<td></td>
</tr>
<tr>
<td>Forest products</td>
<td>None</td>
<td>Hicks &amp; Schnier 2008</td>
<td>Biao et al. 2005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ward 2008</td>
<td>Phillips et al. 2003</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Cubbage et al. 2003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Guibrandsen 2005</td>
<td>Ebeling &amp; Yasue 2009</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hartsfield &amp; Ostermeier 2003</td>
<td>Markopoulos 2003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Humphries &amp; Kainer 2006</td>
<td>McDaniel 2003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kukkonen et al. 2008</td>
<td>Quevedo 2007</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nebel et al. 2005</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rickenbach &amp; Overdevest 2006</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tourism</td>
<td>Rivera 2002</td>
<td></td>
<td>Ayuso 2007</td>
</tr>
<tr>
<td></td>
<td>Rivera &amp; de Leon 2004</td>
<td></td>
<td>Goodman 2000</td>
</tr>
<tr>
<td></td>
<td>Rivera et al. 2006</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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\(^b\).

<table>
<thead>
<tr>
<th>Economic sector</th>
<th>Rigorous: quantitative, credible counterfactual</th>
<th>Moderate rigorous: quantitative, no credible counterfactual</th>
<th>Qualitative: no credible counterfactual</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Type</td>
<td>Environmental focus</td>
<td>All</td>
</tr>
<tr>
<td>Bananas</td>
<td>3</td>
<td>FT (3)</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Organic (1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Eurepap (1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RA (1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Utz (1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coffee</td>
<td>5</td>
<td>FT (3)</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Organic (5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fish products</td>
<td>0</td>
<td>n/a</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forest products</td>
<td>0</td>
<td>n/a</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tourism</td>
<td>3</td>
<td>CST (1)</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SSP (2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>2</td>
<td>18</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)In many studies, producers with multiple certifications were evaluated: 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.

\(^b\)Includes studies in which both environmental effects and social and economic effects were analyzed.

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, 6 on bananas, 5 on tourism, and 4 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 population 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 2 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, 2 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 4 of the 11 provided evidence of such benefits (Table 3). Of these 4, one tested for environmental effects and 3 tested for economic effects. In 2 of the 3 studies of economic effects, 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 7 rigorous studies (2 studies of environmental effects and 5 of social and economic effects) did not show that certification benefits producers.

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

<table>
<thead>
<tr>
<th>Economic sector</th>
<th>Positive social and economic effect</th>
<th>Positive environmental effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bananas</td>
<td>3</td>
<td>n/a</td>
</tr>
<tr>
<td>Coffee</td>
<td>5</td>
<td>n/a</td>
</tr>
<tr>
<td>Fish products</td>
<td>0</td>
<td>n/a</td>
</tr>
<tr>
<td>Forest products</td>
<td>0</td>
<td>n/a</td>
</tr>
<tr>
<td>Tourism</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>4</td>
</tr>
</tbody>
</table>
Bananas

We found 6 published articles on the certification of bananas as sustainable that met our criteria for inclusion in the evidence base (Table 1). All 6 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 6 articles are in Ruben (2008), an edited volume on fair-trade certification. We classified 3 of the studies as rigorous (quantitative analyses that constructed a credible counterfactual), 2 as moderately rigorous (quantitative studies that did not construct a credible counterfactual), and 1 as a qualitative study (did not construct a credible counterfactual).

These 6 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 (2008a) 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 2 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 5 study sectors, the greatest number of published studies of certification effects, 20 in all, examined coffee (Table 1). We classified 5 of these studies as rigorous, 8 as moderately rigorous, and 7 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 (Lyngebaek et al. 2001; Fort & Ruben 2008b; 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 6 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 2 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 7 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 3 studies are more mixed.

Fish Products

We found 4 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, 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 2 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 4 aquaculture operations.

**Forest Products**

We found 11 published studies of the effects of forest-product certification (0, rigorous; 6, moderately rigorous; 5, 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 the 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 5 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, 10 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 the 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 1 or 2 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. Chacon, J. Ferris, D. McLaughlin, 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


Conservation Biology
Volume **, No. *, 2011
Benefits of Sustainability Certification

