

TOPICAL REVIEW • OPEN ACCESS

Is voluntary certification of tropical agricultural commodities achieving sustainability goals for small-scale producers? A review of the evidence

To cite this article: Ruth S DeFries *et al* 2017 *Environ. Res. Lett.* **12** 033001

View the [article online](#) for updates and enhancements.

Related content

- [Assessing the potential additionality of certification by the Round table on Responsible Soybeans and the Roundtable on Sustainable Palm Oil](#)
Rachael D Garrett, Kimberly M Carlson, Ximena Rueda *et al.*
- [Effectiveness of Roundtable on Sustainable Palm Oil \(RSPO\) for reducing fires on oil palm concessions in Indonesia from 2012 to 2015](#)
Megan E Cattau, Miriam E Marlier and Ruth DeFries
- [Evaluating the effectiveness of palm oil certification in delivering multiple sustainability objectives](#)
Courtney L Morgans, Erik Meijaard, Truly Santika *et al.*

Recent citations

- [Improving smallholder livelihoods and ecosystems through direct trade relations: Highquality cocoa producers in Ecuador](#)
Romaïke S. Middendorp *et al*
- [Socio-environmental conflicts: An underestimated threat to biodiversity conservation in Chile](#)
Daniela M. Carranza *et al*
- [Do smallholder farmers benefit from sustainability standards? A systematic review and meta-analysis](#)
Eva-Marie Meemken

Environmental Research Letters



TOPICAL REVIEW

OPEN ACCESS

RECEIVED
29 November 2015

REVISED
20 February 2017

ACCEPTED FOR PUBLICATION
23 February 2017

PUBLISHED
14 March 2017

Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Is voluntary certification of tropical agricultural commodities achieving sustainability goals for small-scale producers? A review of the evidence

Ruth S DeFries^{1,6}, Jessica Fanzo², Pinki Mondal³, Roseline Remans⁴ and Stephen A Wood⁵

¹ Department of Ecology, Evolution and Environmental Biology, Columbia University, New York, NY, United States of America

² Berman Institute of Bioethics, Johns Hopkins University, Baltimore, MD, United States of America

³ Center for International Earth Science Information Network, Columbia University, Palisades, NY, United States of America

⁴ Bioversity International, 3001 Heverlee, Belgium

⁵ School of Forestry and Environmental Studies, Yale University, New Haven, CT, United States of America

⁶ Author to whom any correspondence should be addressed.

E-mail: rd2402@columbia.edu

Keywords: certification, tropical agriculture, coffee, cacao, fairtrade, sustainability, Rainforest Alliance

Supplementary material for this article is available [online](#)

Abstract

Over the last several decades, voluntary certification programs have become a key approach to promote sustainable supply chains for agricultural commodities. These programs provide premiums and other benefits to producers for adhering to environmental and labor practices established by the certifying entities. Following the principles of Cochrane Reviews used in health sciences, we assess evidence to evaluate whether voluntary certification of tropical agricultural commodities (bananas, cocoa, coffee, oil palm, and tea) has achieved environmental benefits and improved economic and social outcomes for small-scale producers at the level of the farm household. We reviewed over 2600 papers in the peer-review literature and identified 24 cases of unique combinations of study area, certification program, and commodity in 16 papers that rigorously analyzed differences between treatment (certified households) and control groups (uncertified households) for a wide range of response variables. Based on analysis of 347 response variables reported in these papers, we conclude that certification is associated on average with positive outcomes for 34% of response variables, no significant difference for 58% of variables, and negative outcomes for 8% of variables. No significant differences were observed for different categories of responses (environmental, economic and social) or for different commodities (banana, coffee and tea), except negative outcomes were significantly less for environmental than other outcome categories ($p = 0.01$). Most cases (20 out of 24) investigated coffee certification and response variables were inconsistent across cases, indicating the paucity of studies to conduct a conclusive meta-analysis. The somewhat positive results indicate that voluntary certification programs can sometimes play a role in meeting sustainable development goals and do not support the view that such programs are merely greenwashing. However, results also indicate that certification is not a panacea to improve social outcomes or overall incomes of smallholder farmers. Rigorous analysis, standardized criteria, and independent evaluation are needed to assess effectiveness of certification programs in the future.

1. Introduction

Civil society and the private sector are increasingly relying on voluntary certification labels, such as Fairtrade and organic, to pursue social and environ-

mental sustainability in supply chains for agricultural products. The basic rationale behind certification programs is straightforward. Producers receive premiums and other benefits for following sustainability criteria defined by a certifying entity. Consumers

(either retail or buyers in the supply chain) pay a premium in exchange for assurance that producers in reality met these criteria.

Coffee, cocoa and other export-oriented, tropical crops are a particular focus for certification programs. The focus on these commodities arises for several reasons. State-level governance in the tropics in many instances does not prioritize sustainable production or is unable to implement effective environmental and labor regulations, leaving open the possibility for non-governmental organizations and the private sector to develop markets for certified products. Second, despite recent reductions in numbers of people living below the poverty line, most of the world's poor are smallholder farmers and laborers in rural areas in the tropics (Cruz *et al* 2015). Improving livelihoods of smallholder farmers and ensuring adequate working conditions for agricultural laborers are consequently critical aspects of sustainable supply chains. Third, habitat conservation is a major factor in sustainable supply chains from an environmental viewpoint and most of the world's biodiversity is located in the humid tropics (Fisher and Christopher 2007). Finally, several tropical crops are widely traded commodities and certification programs could have a large effect on international supply chains. In 2015, the total value of exports for coffee, palm oil, bananas, tea and cocoa—the commodities we consider for this review—were approximately 21.2, 12.4, 8.4, 5.1 and 2.7 trillion US dollars respectively (United Nations 2015b). These factors combine to highlight certification as a potentially influential mechanism to address several sustainable development goals, including 'no poverty' (goal 1), 'responsible consumption' (goal 12), and 'life on land' (goal 15) (United Nations 2015b), as well as private sector commitments to sustainable supply chains.

Ultimately, whether voluntary certification programs fulfill their potential rests on whether benefits are in fact realized on the ground. Producers will only adopt the practices if they perceive that they will be better off than they would otherwise be following other options, including non-certified production or other livelihood strategies. Consumers will only be willing to pay premiums if they trust the certifying entity to achieve positive impacts. Empirical and rigorous data on the effectiveness of voluntary certification programs is fundamental to their ultimate success.

Previous reviews of certification programs highlight a dearth of studies that rigorously assess whether certification schemes are meeting their goals for socially and environmentally sustainable production (Blackman and Rivera 2011, Lernoud *et al* 2016, Milder *et al* 2015, Potts *et al* 2014). A particular challenge is comparison of impacts of certified and non-certified production that accounts for the counterfactual, in other words what the impacts would have been if the same certified producers were

not certified (see methods section below on methods to account for selection bias and establish a counterfactual control group). The growing literature also highlights several other concerns with voluntary certification programs. Smallholders who cannot afford the transaction costs associated with certification or do not have access to information can be further marginalized. Authors argue that certification programs need to strengthen mechanisms to ensure that poorer farmers have the opportunity to benefit and that the market is not further concentrated in the hands of large plantations and wealthier farmers (Brandi *et al* 2015). An additional critique is that certification schemes have to date been driven predominantly by non-governmental organizations and private sector entities from the North, usurping the role of governments in the South to promote their own priorities and establish capacity to regulate production (McDermott 2013, Schleifer 2016).

Certification programs cover a large number of products, including forestry products, fisheries, handicrafts and many crops (Lernoud *et al* 2016, Potts *et al* 2014). For a manageable subset, we focus this review on five major tropical, land-based food commodities. The review assesses case studies in the peer-review literature and summarizes existing results on the impacts of certification at the household (farm) level for smallholder producers who participate in certification programs. We follow the procedures of Cochrane Reviews (Higgins and Green 2011) using statistically rigorous analysis (defined below) to address the question: are the impacts of certified production on economically, socially and environmentally sustainable production significantly different than non-certified production and, if so, what are the impacts? In other words, we rigorously assess the currently-available evidence to test whether voluntary certification programs are achieving results on the ground.

2. Background on voluntary certification programs

Early certification programs were based on eco-labeling to distinguish products that excelled in terms of the environmental sustainability of their production (Potts *et al* 2014). The number of certification programs has multiplied over the last several decades and broadened to include criteria related to economic, social and environmental sustainability. Certification programs generally use a multi-stakeholder process to establish general principles, supported by detailed guidelines and checklists of practices to achieve these principles (these checklists are available in the individual websites for certification programs). To be certified, producers are required to follow these guidelines and undergo a verification process carried out by auditors.

Table 1. Major certification programs for commodities considered in this review. Objectives and principles were obtained from programs' respective websites. Date is from (Potts *et al* 2014). SAN = Sustainable Agriculture Network, IFOAM = International Foundation for Organic Agriculture. Commodities certified are those of the 5 considered in this review and not all commodities certified by the programs.

Program ^a	Date founded	Key objectives and principles for certification standards	Commodities certified
Fair Trade International	1997	Ensure that producers receive prices that cover their average costs of sustainable production; Provide an additional Fairtrade Premium which can be invested in projects that enhance social, economic and environmental development; Enable pre-financing for producers who require it; Facilitate long-term trading partnerships and enable greater producer control over the trading process; Set clear core and development criteria to ensure that the conditions of production and trade of all Fairtrade certified products are socially, economically fair and environmentally responsible	Bananas, Coffee, Cocoa, Tea
GlobalGAP (Good Agricultural Practices)	1997	Food safety and traceability; Environment (including biodiversity); Workers' health, safety and welfare; Animal welfare; Includes Integrated Crop Management; Integrated Pest Management; Quality Management System; and Hazard Analysis and Critical Control Points	Bananas
Organic	1972	IFOAM is umbrella organization with affiliates in more than 100 (IFOAM) countries. Organic standards are typically set at national and regional level.	Bananas, Coffee, Cocoa, Tea, Palm oil
Rainforest Alliance/SAN	1987	Social and environmental management system; Ecosystem conservation; Wildlife protection; Water conservation; Fair treatment and good working conditions; Occupational health and safety; Community relations	Bananas, Coffee, Cocoa, Tea, Palm oil
UTZ Certified	2002	Standards cover management, farming practices, working conditions and environment	Coffee, Cocoa, Tea
Roundtable for Sustainable Palm Oil	2004	Environmental responsibility and conservation of natural resources and biodiversity; Commitment to transparency; Responsible consideration of employees, and of individuals and communities affected by growers and mills; Compliance with applicable laws and regulations; Responsible development of new plantings; Commitment to long-term economic and financial viability; Use of appropriate best practices by growers and millers; Commitment to continuous improvement in key areas of activity	Palm oil

^a Other certification programs for coffee include Starbucks CAFÉ, 4 C Association, and Bird Friendly.

The guiding principles and objectives of most programs for the commodities considered in this review address the three pillars of sustainable development—economic, social and environmental—to varying degrees (table 1). For example, Fairtrade International focuses on equity and producer control, but also includes some guidelines related to environmental concerns. Several programs (e.g. Fairtrade, organic, Rainforest Alliance) certify multiple consumer-facing products such as coffee and cocoa produced by smallholder farmers. More recently, and particularly in the case of commodities such as palm oil that are present in many products but are not directly evident to retail consumers, producers and other stake-holders have formed roundtables to develop certification guidelines (Garrett *et al* 2016). Roundtables have developed for commodities produced by large-scale producers on plantations.

The share of standard-compliant production has grown considerably in recent years. Compliant production grew from 15% to 40%, 3% to 22%, 2% to 15%, 6% to 12%, and 2% to 3% between 2008 and 2012 for coffee, cocoa, palm oil, tea and bananas respectively (Potts *et al* 2014). However, supply is greater than demand and a considerable portion of compliant production is sold as non-compliant,

meaning that producers incur the cost of adhering to standards but do not receive the benefit through premiums for that share of production (figure 1(a)).

Of the five commodities, oil palm is the largest share in terms of area harvested (figure 1(b)) and production (figure 1(c)), as well as the most recent addition to the suite of certified tropical commodities. Coffee has the largest share of certified production (40% of all production standard-compliant and 12% sold as standard-compliant) with a relatively long history of certification following the market crash in the 1990s with the collapse of the International Coffee Agreement (Auld 2010). Shade-grown coffee, which harbors greater biodiversity than higher-yielding, hybrid full sun coffee (De Beenhouwer *et al* 2013, Perfecto *et al* 1996), has been a particular target for certification programs such as Rainforest Alliance for several decades. These varying histories of certification of different commodities are reflected in the number and nature of case studies available for this review.

3. Methods

For this review, we follow the approach defined for Cochrane Systematic Reviews of interventions

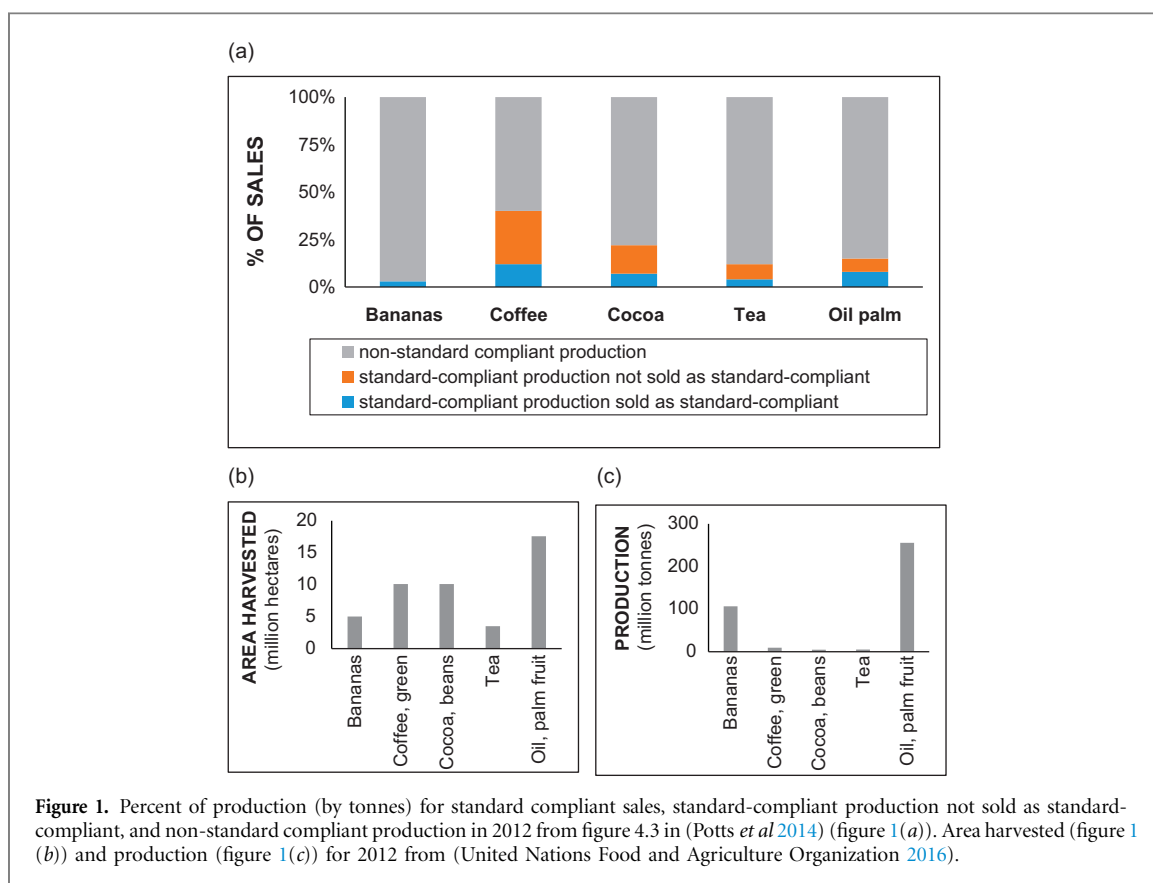


Figure 1. Percent of production (by tonnes) for standard compliant sales, standard-compliant production not sold as standard-compliant, and non-standard compliant production in 2012 from figure 4.3 in (Potts *et al* 2014) (figure 1(a)). Area harvested (figure 1(b)) and production (figure 1(c)) for 2012 from (United Nations Food and Agriculture Organization 2016).

(Higgins and Green 2011). Cochrane Reviews, coordinated by an independent network of researchers and practitioners, are used in the healthcare field to gather and summarize reliable evidence. Key characteristics of a systematic review are (Green *et al* 2008):

- A clearly stated set of objectives with pre-defined eligibility criteria for studies;
- An explicit, reproducible methodology;
- A systematic search that attempts to identify all studies that would meet the eligibility criteria;
- An assessment of the validity of the findings of the included studies, for example through the assessment of risk of bias; and
- A systematic presentation, and synthesis, of the characteristics and findings of the included studies.

Ideally, Cochrane Reviews are based on results of randomized control trials. In non-experimental studies assessing the effectiveness of certification programs, randomized control trials are less feasible than in the healthcare field due to the impracticability of establishing treatment and control groups of households participating in certification programs. In lieu of randomized control trials, rigorous studies that assess effectiveness of certification programs compare certified and non-certified households using statistical approaches to guard against selection bias (Blackman and Rivera 2011). Selection bias can occur

if treatment and control groups are not comparable in terms of biophysical (e.g. soil, topography), socioeconomic (e.g. preferential bias to participate in certification programs, access to markets, education), or other characteristics.

Several methods address the risk of bias, which all involve the construction of a credible counterfactual control group. One method is to pair certified producers with non-certified producers in the study design to ensure that each pair is similar in terms of co-variables that could feasibly affect the result. The treatment and control groups are constructed from the matched pairs (Stuart 2010). A second method is propensity score matching which identifies regions of ‘common-support’ between treatment and control groups. The propensity score predicts probability of certification based on confounding factors. Observations in the control group with a propensity score lower than the minimum in the treatment group, and observations in the treatment group with a propensity score higher than the maximum in the control group, are eliminated from the sample in order to include only comparable observations in the two groups. Matching is carried out through one-to-one pairing of observations with the same propensity score, nearest neighbor, or kernel matching which weights observations in the control group according to propensity-score distance from the observation in the treatment group (Austin 2011, Ruben and Fort 2012). A third method is difference-in-difference, which is common in econometrics. The method uses balanced panel data

to compare changes in outcomes before and after certification in treatment and control groups, implying that the difference in outcomes between the two groups is attributable to certification (Donald and Lang 2007). A difference-in-difference method is only possible if observations are available before and after the time certification began for both control and treatment groups.

The process undertaken for this review is: (1) establish criteria for which studies to include in the review, (2) identify potential studies, (3) assess each study with particular attention to risk of bias, (4) categorize response variables from studies included in the review, and (5) evaluate results.

3.1. Criteria for inclusion in review

We identified the following criteria to determine which papers to include in the review:

- *The study must include primary, quantitative, and empirical data that compares economic, social, and/or environmental outcomes between certified and non-certified households.* A large body of literature contains useful and relevant qualitative analysis assessing impacts and perceptions about certification programs at all levels of the supply chain, e.g. (Arce, 2009, Byerlee and Rueda, 2015). We recognize the value of this literature, but restrict this review to studies that quantitatively compare treatment (certified) and control (non-certified) groups. There is also a large body of literature that evaluates differences in biodiversity for various levels of shade and agroforestry management for coffee and cocoa (see De Beenhouwer *et al* (2013) for a meta-analysis of biodiversity benefits from agroforestry). Again, we recognize the value of these studies in setting certification standards. However, because shade-grown coffee and cocoa can occur with or without certification, we do not include these studies in the review unless they explicitly assess differences in biodiversity of certified and non-certified farms or whether certification accounts for differences in management.
- *The study must be published in the peer-review literature.* We did not include studies in the grey literature or studies commissioned by certifying entities, e.g. (Anon, 2016, Milder and Newsom 2015). We restrict the search to peer review literature to increase confidence that methods and statistical analyses meet standards of the scientific community. In the case of book chapters, we contacted the book's editor to identify whether the chapters went through a peer review process.
- *The study explicitly accounts for risk of bias in comparing certified and non-certified groups.* Studies that compare these two groups without accounting for confounding factors with methods such as

those described above are identified as high risk. Studies that do not report statistical significance of differences between groups are not included.

We only include English-language papers.

3.2. Identification and assessment of potential studies

We used the following search terms in Web of Science to identify potential studies to include in the review: certification, certified, or sustainable as primary search terms and coffee, cocoa (or cacao), banana, tea and oil palm as secondary terms. Each combination of primary and secondary terms were a separate search. Date of the search was February 25, 2016. We downloaded the citation information and abstracts for the papers that met the search criteria. We eliminated duplicates with EndNote bibliography software.

For the studies that met the search criteria (see section 4.1), we skimmed all the titles and abstracts to determine potential relevance for the review. This step eliminated studies that are technical, agronomic studies about the five crops or in other ways not related to certification. Following this initial screening, we downloaded the full papers for the remaining studies. We then assessed whether each paper met the criteria.

For each study that we identified as meeting the criteria, we recorded the date of study, location of the study at the country level, certification program(s), number of households surveyed, response variables assessed in the study, and whether the difference for each response variable between certified and non-certified households was positive, negative or not significant along with level of significance (p -value). We recorded only the sign of the difference, rather than the value, because the literature does not consistently report impact in absolute or percentage terms and does not generally provide the underlying data that would allow comparisons of magnitude of impact beyond positive, negative or not significant.

Studies range considerably in the number of variables collected through surveys and assessed for statistical significance in difference between treatment and control groups. To avoid artifacts from the difference in number of response variables in different studies, we use the average fraction in each study of impacts variables that are positive, negative or not significantly different (rather than taking the fraction of all response variables from all studies) as the key metric to evaluate results.

Some studies include comparisons of impacts among certification programs rather than or in addition to comparison between certified and non-certified households, e.g. (Woubie *et al* 2015). We include only those comparisons with non-certified households. In addition, some response variables, e.g. input costs, indicate a disadvantage for producers if positive difference between the treatment and control groups. In those few cases, we reversed the sign so that

Table 2. Categories for response variables used in this review (see table S2 for detailed variables for individual cases within each of these categories).

Category	Aggregated Category
Yield of respective crop	ECONOMIC
Revenue from respective crop	
Household income	
Non-income well-being	SOCIAL
Impacts on habitat conservation	ENVIRONMENTAL
Other environmental impacts	
Unrelated to sustainability	OTHER

a positive sign for a response variable indicates an advantage and a negative sign indicates a disadvantage associated with certification. In addition, if the paper reports certification as the control and non-certification as the treatment, we reversed the sign.

For studies that reported results from multiple study areas, treatment groups with different certification programs, or commodities, we identified each unique combination of study area/certification and program/commodity as a separate case. If multiple studies reported on the same study area with the same certification program and commodity, we combined results into a single case.

3.4. Comparison of response variables

In the absence of standard evaluation metrics for assessing certification programs, researchers typically define response variables based on their interests and experiences. Some studies emphasize economic impacts such as income while some studies emphasize environmental impacts such as whether farmers planted trees outside coffee plots. The response variables vary greatly across the studies. A seemingly simple variable such as yield, for example, could be reported as yield per hectare, yield per tree, or yield per farm. Non-income well-being variables include a range of impacts, for example from satisfaction with technical assistance to whether the household has piped water.

Consequently, we first categorized the response variables identified within 7 categories and aggregated these categories into three that represent the pillars of sustainability: Economic, Social, and Environmental (table 2). We also included a category for ‘other’ for variables unrelated to sustainability.

We assigned each of the response variables in the studies according to these categories.

3.5. Evaluation of results

We evaluated the results for each case according to the average fraction of response variables that are positive, negative, or not significant between treatment (certified) and control (non-certified). We calculated these fractions for each of the 24 cases. We analyzed the average fractions across cases (in effect average of averages) in three different groupings by: commodity; category of response variables (environment, economic,

social); and certification program. We used a Kruskal-Wallis to test whether differences in fractions are significant across these groupings. Separate tests were applied for positive, not significant, and negative fractions for each of the groups. These tests allowed us to address the question, for example, of whether positive outcomes are significantly different for environmental, economic, or social response variables.

Overall assessment is based on average fractions across cases of positive, not significant, and negative differences in response variables combining all commodities, certification programs and categories of response variables.

3.6. Cross-checks on reproducibility of methods

As recommended for a Cochrane Review (Green and Higgins 2008), more than one person on the authorship team participated in decisions to guard against errors and bias. Three decisions in the methods are subject to potential error: (1) the search for papers to include in the first-stage evaluation, (2) selection of papers to include in the analysis, and (3) assignment of response variables to categories. The first author carried out the search, selection, and category assignments. Co-authors independently searched for papers in other search engines and with other search terms to test whether relevant papers were excluded; applied the criteria for a subset of papers that met the search terms to test reproducibility of the selection process; and reviewed categories assigned to response variables to check for discrepancies.

4. Results

4.1. Identification of studies

In total, 2658 papers met the search criteria, with the highest number for coffee—31% (table 3). The second step of skimming the titles and abstracts to identify relevant papers drastically reduced the total number to 185. The third step to identify papers that meet the criteria in section 3.1 further eliminated most studies, leaving sixteen papers. Fourteen papers in addition to these sixteen only met the first two search criteria (primary, quantitative, and empirical data; and peer-review literature) but did not meet the third criterion (explicitly accounting for risk of bias). Table 4 includes the papers with low risk of bias and table S1 (available at stacks.iop.org/ERL/12/033001/mmedia) includes papers that met the first two criteria but were considered to have higher risk of bias (Bacon 2005, Bacon *et al* 2008, Bacon *et al* 2014, Elder *et al* 2012, Elder *et al* 2013, Fort and Ruben 2009b, Gockowski *et al* 2013, Jacobi *et al* 2015, Jaffee 2009, Lyngbaek *et al* 2001, Melo and Wolf 2005, Mendez *et al* 2010, Ochieng *et al* 2013, Philpott *et al* 2007, Saswattecha *et al* 2015, Dumont *et al* 2014). We evaluate results only for the 16 papers with low risk of bias, which resulted in 24 cases (each case is a unique combination

Table 3. Numbers of papers assessed for the review.

	Coffee	Cocoa	Bananas	Tea	Oil palm	Total
Number of papers meeting search criteria (after removing duplicates)	811	230	340	618	659	2658
Number of papers considered relevant after initial screening	101	30	19	7	28	185
Number of papers with low risk of bias (included in analysis)	13	0	2	1	0	16
Number of papers with risk of bias (not included in analysis)	8	3	1	1	1	14
Number of cases included in analysis	20 ^a	0	2	2 ^b	0	24

^a Twenty cases are contained in the 13 papers for coffee. Three certification programs are included in (Chiputwa *et al* 2015) and two in (Van Rijsbergen *et al* 2016). Three study regions for coffee are included in each (Arnould *et al* 2009) and (Elbers *et al* 2015). Two study regions are included in (Schoonhoven-Speijer and Ruben 2015). The same study region is reported in (Rueda and Lambin 2013) and (Rueda *et al* 2015).

^b One paper (Qiao *et al* 2016) includes two study areas for tea.

Table 4. Sources and characteristics of 24 cases used for the review. See table S2 for number of samples and response variables included in these cases. PSM = propensity score matching; diff-in-diff = difference in difference method.

Reference	Location	Commodity	Certification	Matching method
(Fort and Ruben 2009b) ^a	Peru	Banana	Fairtrade	PSM
(Zinuga-Aria and Saenz-Segura 2009), (Ruben <i>et al</i> 2009)	Costa Rica	Banana	Fairtrade	PSM
(Saenz-Segura and Zuniga-Arias 2009), (Ruben <i>et al</i> 2009)	Costa Rica	Coffee	Fairtrade	PSM
(Rueda and Lambin 2013); (Rueda <i>et al</i> 2015)	Colombia	Coffee	Rainforest Alliance	pair matching
(Bolwig <i>et al</i> 2009)	Uganda	Coffee	Organic	full information maximum likelihood
(Arnould <i>et al</i> 2009)	Nicaragua	Coffee	Fairtrade	test confounding factors
	Peru	Coffee	Fairtrade	test confounding factors
	Guatemala	Coffee	Fairtrade	test confounding factors
(Ruben and Fort 2012) ^b	Peru	Coffee	Fairtrade	PSM
(Van Rijsbergen <i>et al</i> 2016) ^c	Kenya	Coffee	Fairtrade	Diff-in-diff
	Kenya	Coffee	Fairtrade/UTZ	Diff-in-diff
	Uganda	Coffee	Fairtrade	PSM
(Chiputwa <i>et al</i> 2015)	Uganda	Coffee	Rainforest Alliance	PSM
	Uganda	Coffee	UTZ	PSM
(Takahashi and Todo 2013)	Ethiopia	Coffee	Rainforest Alliance	PSM
(Jena <i>et al</i> 2012)	Ethiopia	Coffee	Fairtrade/org	PSM
(Ruben and Zuniga 2011)	Nicaragua	Coffee	Fairtrade	PSM
	Uganda1 ^d	Coffee	UTZ	Diff-in-diff
(Elbers <i>et al</i> 2015)	Uganda2 ^e	Coffee	UTZ	Diff-in-diff
	Uganda3 ^f	Coffee	UTZ	Diff-in-diff
(Schoonhoven-Speijer and Ruben 2015)	Kenya1 ^g	Coffee	UTZ	test prob of certification
	Kenya2 ^h	Coffee	UTZ	test prob of certification
(Qiao <i>et al</i> 2016)	China	Tea	Organic	PSM
	Sri Lanka	Tea	Organic	PSM

^a This case study is also summarized in (Ruben *et al* 2009).

^b This case study is also reported in (Fort and Ruben 2009a) and summarized in (Ruben *et al* 2009).

^c This case study is also reported in (van Rijsbergen *et al* 2015).

^d Study region is Old Ankole community, Ibanda district.

^e Study region is New Ankole community, Ibanda district.

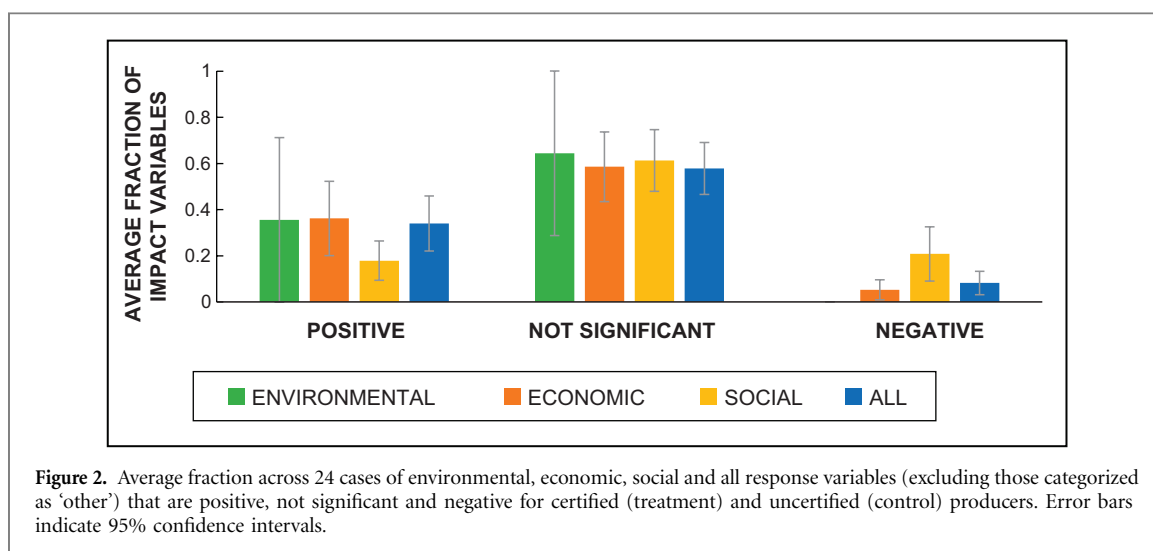
^f Study region is Kulika district.

^g Study region is Embu district.

^h Study region is Mathiyoia district.

of study region, commodity, and certification program) because several papers included results from multiple study regions or certification programs in a single paper (see footnote to table 4). The 24 cases contained a total of 347 response variables, not including 125 categorized as 'other' (table S2).

We found no papers that met the criteria for cocoa or oil palm and only one for tea (with 2 cases) and two papers for bananas (table 4). The predominance of papers assessed impacts of coffee and Fair Trade. Two papers assessed impacts of multiple certification (one each for Fair Trade/UTZ and Fair Trade/organic),



which farmers undertake to maintain options for different markets.

Papers on coffee spanned across Latin America and Africa, but papers on banana were only from Latin America and papers on tea only from Asia. The imbalance in the commodities, certification programs, and geographical locations rigorously evaluated in the literature indicates the need for additional, targeted studies and the inability to form robust conclusions that compare across commodities and certification programs.

The cross-checks on reproducibility of results indicated no major discrepancies but identified one additional edited book containing peer-review cases which we subsequently added to the analysis. A Google Scholar search also identified a relevant paper outside the time window for this review (Bellamy *et al* 2016), which we did not include. In the second step to select papers based on the criteria, independent selection by co-authors for subsets of 30 papers on oil palm and 25 papers on cocoa agreed with the first author's decisions. Co-authors' assignment of categories for response variables also agreed with a few minor discrepancies.

4.2. Evaluation of studies

Comparison across the 24 cases indicates that on average 34% of response variables were significantly positive, 58% not significant and 8% significantly negative (figure 2). We interpret this result as a moderately positive impact from certification programs across the pillars of sustainability at the household level.

When aggregating results by category of response variable (environmental, economic, and social), each of these categories covers a large range of impacts that fit loosely into these categories (table S2). Environmental and economic response variables had approximately the same average fraction of positive outcomes (36% for both) compared with social response variables (18%). Environmental impacts variables had the largest average fraction of not significant response variables across cases (64%) and the lowest fraction of negative outcomes (0%) (table 5).

However, Kruskal-Wallis test indicates that the differences in means among the three categories are not statistically significant except for negative impacts ($p = 0.01$).

Results show a pattern within categories (figure S1). For economic response variables categorized as 'revenue from commodity,' 56% had positive outcomes compared with 24% for 'household income.' This difference could represent the relative success of certification in providing premiums but less success with improving smallholders' overall economic situation. Within environmental response variables, positive impacts on habitat conservation are unanimous (although only 2 cases). Average fraction of positive impacts on other environmental practices (e.g. use of organic fertilizers) is only 22 percent.

Aggregation by commodity indicates a predominance of studies on coffee (20 cases compared with 2 each for tea and bananas). Coffee had the highest average fraction of positive outcomes (36%) although the difference with banana (24%) and tea (25%) was not statistically significant.

Comparison across certification programs is hampered by lack of cases that cover the range of commodities and outcome categories required for a valid comparison (table S3). For example, cases with organic certification only included economic response variables and no cases covered fair trade tea. Based on the available cases, Rainforest Alliance had a marginally non-significantly higher average percentage of positive response variables (77%, $p = 0.06$) and Utz had the highest average percentage of negative response variables (22%, $p = 0.06$) compared among Fairtrade, organic, Rainforest Alliance, and Utz. Many more studies are needed to confidently assess whether these different programs have led to varying outcomes.

5. Conclusions

Based on studies that rigorously compare certified and non-certified producers, we conclude that voluntary

Table 5. Average fraction across 24 cases of impacts variables (excluding those categorized as ‘other’) that are positive, not significant and negative for certified (treatment) and uncertified (control) producers grouped by category of response variable, commodity, and certification program. P-values are results of Kruskal-Wallis test to test difference in means.

	No. of cases ^a	Mean no. (range) of response variables per case	Average fraction (st dev) across cases of response variables by sign between treatment (certified) and control (non-certified)		
			Positive	Not significant	Negative
By category of response variable					
Environmental	6	4.00 (1–15)	0.36 (0.45)	0.64 (0.45)	0 (0)
Economic	20	11.05 (2–26)	0.36 (0.39)	0.59 (0.37)	0.05 (0.11)
Social	16	6.13 (2–13)	0.18 (0.17)	0.62 (0.28)	0.21 (0.24)
<i>p</i> -value			n.s.	n.s.	0.01
By commodity					
Banana	2	24.00 (14–34)	0.24 (0.15)	0.66 (0.01)	0.10 (0.14)
Coffee	20	14.00 (1–39)	0.36 (0.33)	0.55 (0.31)	0.09 (0.14)
Tea	2	10.00 (10–10)	0.25 (0.35)	0.75 (0.35)	0 (0)
<i>p</i> -value			n.s.	n.s.	n.s.
By certification program^b					
FairTrade	10	17.60 (3–39)	0.39 (0.27)	0.55 (0.23)	0.06 (0.09)
Organic	4	6.25 (2–10)	0.38 (0.48)	0.63 (0.48)	0 (0)
Rainforest Alliance	2	12.50 (1–24)	0.77 (0.32)	0.23 (0.32)	0 (0)
Utz	6	14.67 (1–19)	0.13 (0.18)	0.65 (0.29)	0.22 (0.19)
<i>p</i> -value			0.06	n.s.	0.06
All	24	14.54 (1–39)	0.34 (0.32)	0.58 (0.30)	0.08 (0.13)

^a Number of cases sums to more than 24 for category and certification program because cases can include more than one category or program.

^b Comparison by certification program does not include multiple certification (Fairtrade/organic and Fairtrade/UTZ) because only one case was available for these programs.

certification programs for the tropical commodities considered in this review generally lead to some positive impacts (on average 34% of response variables) with most impacts (on average 58%) showing no difference between certified and non-certified producers. This result is not surprising considering that variables quantified in the studies covered a broad set of impacts, many of which are not explicit goals of the certification programs (table S2). Negative impacts, which could occur, for example, if a household loses income by producing a certified product without recouping the costs of compliance through premiums or other benefits, are less common (on average 8% of response variables).

Based on the small sample of 24 cases and particularly the small number of cases for commodities other than coffee, we have only moderate confidence in these results. We cannot rule out selection bias of study region that would inflate (if researchers choose study regions where certification programs are more impactful than randomly selected study regions) or deflate (if researchers choose study regions where certification programs are less impactful) results. A randomized selection of study regions would be required to eliminate this potential bias.

From a consumer’s point of view, the results indicate that premiums for certified products do have a generally positive impact on the ground. These positive impacts are most pronounced for conserving habitat and increasing revenue from the commodity for the producer compared with more diffuse impacts

on environment and overall household income. We conclude that certification programs can play a role in advancing sustainable development goals, although consumers should be aware that these programs are not a panacea especially for the considerable social hardships facing smallholder producers. However, the imbalance that creates more supply than demand (figure 1), as well as the usurpation of governance responsibilities by actors from the North, are serious obstacles to long-term contributions of voluntary certification programs to sustainable development goals.

From a researcher’s point of view, the results confirm the arguments of Milder *et al* (2015) that shared evaluation criteria and procedures are urgently needed for the future success of certification programs. Lack of consistency in response variables and non-standard reporting were challenging for this review. In particular, studies with rigorous matching procedures and construction of a counterfactual control group are needed to add to the body of evidence. Even within the 24 cases with low bias, studies were inconsistent in how they reported results, e.g. percentages or absolute differences in response variables between treatment and control groups. Studies also used different definitions of response variables. Consequently, the possibility to conduct a rigorous meta-analysis with pooled data from different studies was limited. Moreover, the imbalance in the number of cases assessing the impacts of coffee certification compared with the other commodities (20 out of 24) prohibits

robust conclusions comparing outcomes for certification for different commodities. A further limitation is the inability to assess the magnitude of the differences in response variables beyond significantly positive, negative or no difference.

We support an independent process, such as advocated by the Committee on Sustainability Assessment (COSA 2013), to identify evaluation criteria, study design, and analysis to evaluate certification programs in the future. With current discussion about certification at the jurisdictional landscape level (Tscharntke *et al* 2015), an approach to evaluate effectiveness of certification will need to consider construction of control groups at the landscape-scale, preferably from the outset of such programs.

Finally, we note the challenges and benefits of following the procedures identified in the healthcare field for Cochrane Reviews. Randomized control trials are not generally feasible in many topics related to sustainability, putting a large onus on researchers to develop statistically valid counterfactual control groups. The strict procedures for Cochrane Reviews could benefit many aspects of sustainability studies by promoting rigorous analysis of primary data. In addition, such reviews can provide a basis for evidence-based decisions by the public, governments, and non-governmental organizations.

Acknowledgments

This review is based on the principles of Cochrane Reviews but is not affiliated with the Cochrane Group nor has it been reviewed or endorsed by the Cochrane Group. No authors received funding from certification entities.

References

- Anon 2016 UTZ Impact report 2016 UTZ, Amsterdam, The Netherlands, p 38
- Arce A 2009 Living in times of solidarity: fair trade and the fractured life worlds of Guatemalan coffee farmers *J. Int. Dev.* **21** 1031–41
- Arnould E J, Plastina A and Ball D 2009 Does fair trade deliver on its core value proposition? Effects on income, educational attainment, and health in three countries *J. Public Policy Mark.* **28** 186–201
- Auld G 2010 Assessing certification as governance: effects and broader consequences for coffee *J. Env. Dev.* **19** 215–41
- Austin P C 2011 An introduction to propensity score methods for reducing the effects of confounding in observational studies *Multivar. Behav. Res.* **46** 399–424
- Bacon C 2005 Confronting the coffee crisis: can fair trade, organic, and specialty coffees reduce small-scale farmer vulnerability in northern Nicaragua *World Dev.* **33** 497–511
- Bacon C M, Mendez V E, Gomez Flores M E, Stuart D and Diaz Flores S R 2008 Are sustainable coffee certifications enough to secure farmer livelihoods? The millennium development goals and Nicaragua's fair trade cooperatives *Globalizations* **5** 259–74
- Bacon C M, Sundstrom W A, Gomez M E F, Mendez V E, Santos R, Goldoftas B and Dougherty I 2014 Explaining the 'hungry farmer paradox': smallholders and fair trade cooperatives navigate seasonality and change in Nicaragua's corn and coffee markets *Glob. Environ. Change* **25** 133–49
- Bellamy A S, Svensson O, van den Brink P J and Tedengren M 2016 What is in a label? Rainforest-Alliance certified banana production versus non-certified conventional banana production *Glob. Ecol. Conserv.* **7** 39–48
- Blackman A and Rivera J 2011 Producer-level benefits of sustainability certification *Conserv. Biol.* **25** 1176–85
- Bolwig S, Gibbon P and Jones S 2009 The economics of smallholder organic contract farming in tropical africa *World Dev.* **37** 1094–104
- Brandi C, Cabani T, Hosang C, Schirmbeck S, Westermann L and Wiese H 2015 Sustainability standards for palm oil challenges for smallholder certification under the RSPO *J. Env. Dev.* **24** 292–314
- Byerlee D and Rueda X 2015 From public to private standards for tropical commodities: a century of global discourse on land governance on the forest frontier *Forests* **6** 1301–24
- Chiputwa B, Spielman D J and Qaim M 2015 Food standards, certification, and poverty among coffee farmers in Uganda *World Dev.* **66** 400–12
- COSA 2013 The COSA measuring sustainability report: coffee and cocoa in 12 countries. The Committee on Sustainability Assessment, Philadelphia, PA
- Cruz M, Foster J, Quillin B and Schellekens P 2015 Ending extreme poverty and sharing prosperity: progress and policies *Policy Research Note PRN/15/03*, World Bank Group (<http://pubdocs.worldbank.org/pubdocs/publicdoc/2015/10/109701443800596288/PRN03-Oct2015-TwinGoals.pdf>)
- De Beenhouwer M, Aerts R and Honnay O 2013 A global meta-analysis of the biodiversity and ecosystem service benefits of coffee and cacao agroforestry *Agric. Ecosyst. Environ.* **175** 1–7
- Donald S G and Lang K 2007 Inference with difference-in-differences and other panel data *Rev. Econ. Stat.* **89** 221–33
- Elbers W, van Rijsbergen B, Bagamba F and Hoebink P 2015 *The Impact of Utz Certification on Smallholder Farmers in Uganda* ed R Ruben and P Hoebink (The Netherlands: Wageningen Academic) pp 53–81
- Elder S D, Zerriffi H and Le Billon P 2012 Effects of fair trade certification on social capital: the case of rwandan coffee producers *World Dev.* **40** 2355–67
- Elder S D, Zerriffi H and Le Billon P 2013 Is Fairtrade certification greening agricultural practices? An analysis of Fairtrade environmental standards in Rwanda *J. Rural Stud.* **32** 264–74
- Fisher B and Christopher T 2007 Poverty and biodiversity: measuring the overlap of human poverty and the biodiversity hotspots *Ecol. Econ.* **62** 93–101
- Fort R and Ruben R 2009a The impact of fair trade certification on coffee producers in Peru *The Impact of Fair Trade* ed R Ruben (The Netherlands: Wageningen Academic) pp 75–98
- Fort R and Ruben R 2009b The impact of fair trade on banana producers in Northern Peru *The Impact of Fair Trade* ed R Ruben (The Netherlands: Wageningen Academic Publishers) pp 49–74
- Garrett R D, Carlson K M, Rueda X and Noojipady P 2016 Assessing the potential additionality of certification by the round table on responsible soybeans and the roundtable on sustainable palm oil *Environ. Res. Lett.* **11** 045003
- Gockowski J, Afari-Sefa V, Sarpong D B, Osei-Asare Y B and Agyeman N F 2013 Improving the productivity and income of Ghanaian cocoa farmers while maintaining environmental services: what role for certification *Int. J. Agr. Sustain.* **11** 331–46
- Green S and Higgins J 2008 Chap 2 preparing a cochrane review *Cochrane Handbook for Systematic Reviews of Interventions* ed J Higgins and S Green (London, UK: Wiley-Blackwell) pp 11–30

- Green S, Higgins J, Alderson P, Clarke M, Mulrow C and Oxman A 2008 Chapter 1. Introduction *Cochrane Handbook of Systematic Reviews of Interventions* ed J Higgins and S Green (London, UK: Wiley-Blackwell) pp 3–10
- Higgins J and Green S 2011 *Cochrane Handbook for Systematic Review of Interventions Version 5.1.0 The Cochrane Collection* (www.handbook.cochrane.org)
- Jacobi J, Schneider M, Pillco Mariscal M, Huber S, Weidmann S, Bottazzi P and Rist S 2015 Farm resilience in organic and nonorganic cocoa farming systems in alto beni, Bolivia *Agroecol. Sust. Food* **39** 798–823
- Jaffee D 2009 ‘Better but not great’: the social and environmental benefits and limitations of fair trade for indigenous coffee producers in Oaxaca, Mexico *The Impact of Fair Trade* ed R Ruben (The Netherlands: Wageningen Academic Publishers) pp 195–222
- Jena P R, Chichaibelu B B, Stellmacher T and Grote U 2012 The impact of coffee certification on small-scale producers’ livelihoods: a case study from the jimma zone, Ethiopia *Agr. Econ.* **43** 429–40
- Lernoud J, Potts J, Sampson G, Voora V, Willer H and Wozniak J 2016 *The State of Sustainable Markets-Statistics and Emerging Trends 2015*
- Lynghaek A E, Muschler R G and Sinclair F L 2001 Productivity and profitability of multistrata organic versus conventional coffee farms in Costa Rica *Agroforest. Syst.* **53** 205–13
- McDermott C L 2013 Certification and equity: applying an ‘equity framework’ to compare certification schemes across product sectors and scales *Environ. Sci. Policy* **33** 428–37
- Melo C J and Wolf S A 2005 Empirical assessment of eco-certification—the case of Ecuadorian bananas *Organ. Environ.* **18** 287–317
- Mendez V E, Bacon C M, Olson M, Petchers S, Herrador D, Carranza C, Trujillo L, Guadarrama-Zugasti C, Cordon A and Mendoza A 2010 Effects of fair trade and organic certifications on small-scale coffee farmer households in Central America and Mexico *Renewable Agric. Food Syst.* **25** 236–51
- Milder J and Newsom D 2015 *SAN/Rainforest Alliance Impacts Report*. Rainforest Alliance/Sustainable Agriculture Network, New York, p 122
- Milder J C, Arbutnot M, Blackman A, Brooks S E, Giovannucci D, Gross L, Kennedy E T, Komives K, Lambin E F and Lee A 2015 An agenda for assessing and improving conservation impacts of sustainability standards in tropical agriculture *Conserv. Biol.* **29** 309–20
- Ochieng B O, Hughey K F D and Bigsby H 2013 Rainforest alliance certification of kenyan tea farms: a contribution to sustainability or tokenism *J. Cleaner Prod.* **39** 285–93
- Perfecto I, Rice R A, Greenberg R and Van der Voort M E 1996 Shade coffee: a disappearing refuge for biodiversity *BioScience* **46** 598–608
- Philpott S M, Bichier P, Rice R and Greenberg R 2007 Field-testing ecological and economic benefits of coffee certification programs *Conserv. Biol.* **21** 975–85
- Potts J, Lynch M, Wilkings A, Huppé G, Cunningham M and Voora V 2014 The state of sustainability initiatives review 2014: Standards and the green economy. International Institute for Sustainable Development (IISD) and the International Institute for Environment and Development (IIED) 332
- Qiao Y, Hallberg N, Vaheesan S and Scott S 2016 Assessing the social and economic benefits of organic and fair trade tea production for small-scale farmers in Asia: a comparative case study of China and Sri Lanka *Renewable Agric. Food Syst.* **31** 246–57
- Ruben R and Fort R 2012 The impact of fair trade certification for coffee farmers in Peru *World Dev.* **40** 570–82
- Ruben R, Fort R and Zúñiga-Arias G 2009 Measuring the impact of fair trade on development *Dev. Pract.* **19** 777–88
- Ruben R and Zuniga G 2011 How standards compete: comparative impact of coffee certification schemes in Northern Nicaragua *Supply Chain Manag.* **16** 98–109
- Rueda X and Lambin E F 2013 Responding to globalization: impacts of certification on colombian small-scale coffee growers *Ecol. Soc.* **18** 21
- Rueda X, Thomas N E and Lambin E F 2015 Eco-certification and coffee cultivation enhance tree cover and forest connectivity in the Colombian coffee landscapes *Reg. Environ. Change* **15** 25–33
- Saenz-Segura F and Zuniga-Arias G 2009 Assessment of the effect of fair trade on smallholder producers in Costa Rica: a comparative study in the coffee sector *The Impact of Fair Trade* ed R Ruben (The Netherlands: Wageningen Academic) pp 117–35
- Saswattecha K, Kroeze C, Jawjit W and Hein L 2015 Assessing the environmental impact of palm oil produced in Thailand *J. Cleaner Prod.* **100** 150–69
- Schleifer P 2016 Private governance undermined: India the roundtable on sustainable palm oil. Global environmental politics
- Schoonhoven-Speijer M and Ruben R 2015 *Maintaining Sustainable Livelihoods: Effects of Utz Certification on Market Access, Risk Reduction and Livelihood Strategies of Kanya Coffee Farmers* ed R Ruben and P Hoebink (The Netherlands: Wageningen Academic) pp 149–73
- Smith Dumont E, Gnahoua G M, Ohouo L, Sinclair F and Vaast P 2014 Replication data for: farmers in Cote d’Ivoire value integrating tree diversity in cocoa for the provision of ecosystem services World Agroforestry Centre
- Stuart E A 2010 Matching methods for causal inference: a review and a look forward *Stat. Sci.* **25** 1
- Takahashi R and Todo Y 2013 The impact of a shade coffee certification program on forest conservation: A case study from a wild coffee forest in Ethiopia *J. Environ. Manage.* **130** 48–54
- Tscharntke T, Milder J C, Schroth G, Clough Y, DeClerck F, Waldron A, Rice R and Ghazoul J 2015 Conserving biodiversity through certification of tropical agroforestry crops at local and landscape scales *Conserv. Lett.* **8** 14–23
- United Nations 2015a Transforming our World: the 2030 agenda for sustainable development: A/RES/70/1. United Nations New York
- United Nations 2015b UN Comtrade Database, Department of Social Economic Affairs Statistics Division (Ed.)
- United Nations Food and Agriculture Organization 2016 FAOSTAT, Rome
- Van Rijsbergen B, Elbers W, Ruben R and Njuguna S N 2016 The ambivalent impact of coffee certification on farmers’ welfare: a matched panel approach for cooperatives in central kenya *World Dev.* **77** 277–92
- van Rijsbergen B, Elbers W, van Kempen L, Ruben R and Hoebink P 2015 The effects of coffee certification in Kenya *Coffee Certification in East Africa* ed R Ruben and P Hoebink (The Netherlands: Wageningen Academic) pp 83–125
- Woubie A, Muradian R and Ruben R 2015 Impact of multiple certification on smallholder coffee farmers’ livelihoods: evidence from southern Ethiopia *Coffee Certification in East Africa* ed R Ruben and P Hoebink (The Netherlands: Wageningen Academic) pp 127–47
- Zinuga-Aria G and Saenz-Segura F 2009 The impact of Fair Trade in banana production of Costa Rica *The Impact of Fair Trade* ed R Ruben (The Netherlands: Wageningen Academic) pp 99–116