March 2018

3keel

The effectiveness of standards in driving adoption of sustainability practices:

A State of Knowledge Review

Lead authors:
Dr Gillian Petrokofsky (University of Oxford) and Dr Steve Jennings (3Keel LLP)

iseal alliance

Executive summary

Why this study?

Voluntary sustainability standards are an important means of providing assurance that products and materials traded in complex, global supply chains have been produced in an ethical and environmentally benign way. Understanding what these standards deliver on the ground is an important part of the debate about their ongoing relevance.

We address this question through reviewing the evidence that standards systems drive the adoption of sustainable practices. The focus is on individuals or organisations changing their practices – such as the use of agricultural inputs, retention of wildlife habitat, and providing better conditions for hired workers – as a necessary step towards the final impacts of standards systems. We focus on changes in practice rather than the potential outcomes, such as increased biodiversity or improved wellbeing. This focus is to increase understanding of how standards bring about change and because recent reviews have focused on outcomes.

The research questions we focus on are:

RQ 1: What is the effectiveness of standards and certification in driving the adoption of more sustainable practices by certified entities over a period of time?

RQ 2: How do standards and certification tools operate to achieve practice adoption? To what extent does adopting practices lead to continuous improvement in entities over time in identified thematic areas?

RQ 3: What lessons can standards systems learn for this area of their work?

Four complementary types of evidence are used. Systematic mapping provides information on the amount and quality of evidence available to address the research questions. It also provides a filtered and manageable set of research from a large and chaotic literature and presents papers of direct relevance that can then be the source for detailed narrative analysis. The narrative analysis of these papers provides insight into what is happening and why. However, there are questions, commodities and geographies within this body of evidence that are rarely researched. We add additional insight through analysis of standards systems' monitoring and compliance data, and from interviews with expert informants.

What is the evidence base to understand practice adoption linked to sustainability standards?

One hundred and sixteen studies which reported relevant outcomes from entities certified to a sustainability standard, and which included a counterfactual, were filtered from an original body of over thirteen thousand studies from the peer-reviewed and 'grey' literature. The resulting body of evidence was characterised by the following:

- There is a concentration of research into coffee and forestry, and on Rainforest Alliance, Organic, Fairtrade and FSC standards. How representative the findings are for other schemes and sectors is not well understood.
- There is a concentration on producers, specifically smallholders and producer cooperatives. Little research focuses on other parts of the supply chain.
- There is a focus on research in developing countries in the tropics and sub-tropics, with comparatively little on developed and temperate countries. Some of the potential inter-relations between context and certification, therefore, have a limited body of robust literature.

- There are almost no studies that enumerate changes in practice over time. In particular, little is known about what changes to practice take place before certification.
- To put our findings in context, a recent systematic review on a related topic (Oya et al., 2016), following an identical filtering process, found a total of 158 publications containing relevant studies, from an initial search result of 10,753 publications, i.e. 1.5% of the initial set were analysed. This compares with 0.86% in our study – 116 final papers from a set of 13,515 papers – a similar outcome.

Do sustainability standards drive adoption of better sustainability practices?

The evidence for practice adoption was assessed in six thematic areas which covered environmental, social and economic practices. These were:

1. Conservation and biodiversity;
2. Input use (fertilisers, herbicides, pesticides, etc.);
3. Community benefits and development;
4. Occupational health and safety;
5. Good production practices; and
6. Management systems.

There is evidence in each of these thematic areas that certification and standards can contribute to the adoption of improved practices. This is typically expressed as a difference in practices between certified and non-certified entities.

The evidence for practice adoption may be more robust for some sustainability themes than for others. For example, all seven of the papers enumerating changes in Health and Safety practices showed a positive impact of certification, as did all of the twelve enumerating community benefits and development. This does not 'prove' that practice adoption is a universal effect of certification for these themes, but there are logics from other areas of research that suggest why practice adoption may differ between themes and contexts. Evidence that standards do not result in practice adoption seems to be more frequently reported for:

- Situations where existing levels of practice are high and may not have to change to meet the requirements of a standard (or the requirements of the standard are low), or conversely where the gap between existing practice and the standard is too great;
- Hard to detect practices (e.g., freedom of association, or discriminatory hiring) or ones carried out infrequently (e.g., fertiliser application);
- Changes in practice after certification. There is circumstantial evidence – but few empirical studies – that the biggest changes to practice may occur before certification.

The same factors that motivate individuals and organisation to seek certification are reported to drive the adoption of improved sustainability practices. These include market access, price premium, gaining an advantage over competitors, managing reputational risks, and responding to demands from customers. In addition to these, the most commonly cited driver of practice change was external technical, institutional or financial support. In addition, the process of assessing compliance with the standard on a regular basis (audits) appears to drive some practice adoption. Audits also play a role in ensuring that levels of performance are maintained. The role of standards systems in preventing slippage of good practice may be under-appreciated, but is nonetheless critical to ensuring sustainable outcomes.

Broader considerations

Three other considerations arise from this study:

• There is little formal evidence about whether entities that choose certification are different from those that do not. It is therefore unclear how much certification is reaching all producers, or whether it is preferentially used by those that are already performing well relative to others within the same context;

• Changing practice is a type of behaviour change. For standards schemes to be more effective at driving practice adoption may therefore be aided by increased understanding and application of disciplines such as behavioural psychology; and

• Standards may influence norms, policies and national standards, which in turn affect practices on the ground. Standards may, therefore, be having a greater effect on practice than is reported through studies on certified entities alone.

Recommendations

For research

- The amount of longitudinal research needs to increase so that evidence can be provided to answer key questions, such as whether entities that seek certification are different from those that do not, and the extent to which practices are adopted and maintained after certification.
- Impact evaluators should insist on study designs that include a relevant comparison population, or before and after comparison.
- More primary research on certification and standards located further up the supply chain, in under-represented sectors (e.g., mining, cotton, fish); and under-represented geographies (especially in Europe and North America), would significantly aid understanding of how certification systems work.

For practice

- Many impact evaluations were filtered out of the final body of evidence due to lack of counterfactuals and/or clearly defined measurable outcomes. Recognising the steps made to increase the rigour of impact evaluations, and the cost of rigorous studies, we recommend that impact evaluations are used only to answer key strategic questions. The resources thereby saved to be invested in rapid, repeat longitudinal surveys, gathering stories of change, and in experiential learning.
- Further invest in understanding and implementing behavioural and organisational change.
- Consider developing innovative capacity for detecting short-term, hidden changes in complex and rapidly changing supply chains, to overcome the difficulties in changing (and auditing) ‘hard to detect’ practices.
- Find ways to capture evidence of pre-certification change either by asking certification and assurance bodies to share that data or to provide analysis of it.
- Think of how a continuous improvement model of standard-setting and auditing can address some of the challenges that the study raises about when practice adoption may be less likely.

For policy

- As a significant proportion of existing impact studies lack the counterfactual to be able to detect the impact, we suggest that only rigorous impact evaluations are funded,

which may mean only funding them for key strategic issues. The corollary is to increase investment in longitudinal research.

- Providing the funds and policy environment to ensure that actors anywhere in the supply chain receive adequate technical, financial and institutional support to adopt improved practices is a key multiplier to investment in sustainability standards.

For businesses and governments:

- Be clear about what sustainability standards are designed to do. If the aim is to raise the bar of the lowest performing producers in a given sector, then standards alone may not be the solution. Support may be needed to bring producers whose starting point is extremely low up to a point where they can then enter certification and maintain good practices.
- Be aware of the role standards play in helping adopt and maintain good practices.

Contents

1 Introduction

1.1 Background to the study

Voluntary sustainability standards have grown rapidly in number and importance in global commodity markets over recent years. There are also challenges as businesses, in particular, evolve their public policy commitments and seek to ensure that standards are providing the benefits they require in a cost-effective way. A key strand to these debates is understanding what voluntary sustainability standards deliver on the ground.

The standards community has made progress in researching its own impacts in recent years, investing in increasingly robust data collection, monitoring and evaluation systems. In addition, the impact of sustainability standards has recently been the subject of a number of formal, academic reviews. Most notably, Oya et al. (2017) and DeFries et al. (2017) performed systematic reviews on the impacts of standards on sustainability outcomes. Both reported mixed signals – certification having positive impacts on some metrics but no evidence of impacts on others – and the findings of both were limited by a paucity of high-quality impact evaluations.

If the findings of formal reviews on the impact of standards and certification systems are not straightforward, then perhaps we need to take a step back. For standards to have an effect on sustainability impacts, they must first affect the practices of the individuals and organisations that are certified. After all, if a standard did not require actions to be taken to conserve biodiversity, for example, then how could an increase in bird species be attributed to that standard? If the standard did not require actions to make working conditions safe, how could improved health and safety be a direct result of the standard?

Practice adoption is thus the necessary step before impact. By being a step earlier in the process of change it may be more straightforward to measure as it may be less influenced by complex changes to the context.

Focusing on practice adoption aligns with the way that standards systems operate. The approach commonly adopted by standards systems (although there are variations) is to develop a set of practices or performance benchmarks, certify or license entities

Box 1.

Examples of the approaches taken by voluntary sustainability standards schemes to achieve change.

The following provides examples of how standards schemes try to change practices in order to bring about more sustainable production:

- Defining sustainable practices and training producer groups, trainers and auditors in implementing them (e.g., UTZ, 2017).
- Engagement and consensus on the issues, defining best practice in standards, and an assurance system to ensure that producers apply them (e.g., Forest Stewardship Council, 2015).
- Renewed certification being dependent upon performance improving beyond a minimum acceptable level (e.g., Marine Stewardship Council, 2011).
- Creating the demand for improved practices via sourcing initiatives; and the capacity to deliver them through training, standards, and auditing (e.g., Responsible Jewellery Council, undated).

A scheme might use several of these approaches. Alongside practice adoption, additional changes, such as changes in policy, consumer awareness, increased leadership and/or collaboration in the supply chain are also viewed as necessary to reach sustainable impacts.

that adopt these practices, build a market for the demand for certified products, check compliance with these practices through assurance models, and evaluate over time if this delivers the stated sustainability goals and performance. Some of the ways that standards systems seek to support practice adoption are illustrated in Box 1.

1.2 Purpose of the study

The aim of this research is to understand the effectiveness of sustainability standards and certification tools in driving the adoption of more sustainable practices in certified entities, thereby contributing to the achievement of key sustainability outcomes.

The objectives of the study were to:

- Generate robust research that comprehensively analyses and reviews available data and empirical evidence to help understand the role that standards and certification tools play in driving practice adoption in certified entities over time;
- Contribute to our theoretical and conceptual understanding of how standards systems operate to drive practice adoption in various sectors and regions, and how this results in achievement of sustainability outcomes;
- Synthesise learnings on how standards systems can improve this line of work and specific recommendations emerging from the research to inform future practice.

1.3 What do we mean by practice adoption?

For the purpose of this study, we define practice adoption as when an individual or organisation changes their way of doing particular activities. These include activities that would be considered closer to 'outputs' in a Theory of Change or logframe (e.g., a change in management system) and other activities that would be more akin to outcomes (e.g., reduced pesticide use). We do not include detailed consideration of final impacts (e.g., increased biodiversity or increased household income) as these have been the subject of recent systematic reviews.

An individual's or organisation's practice can, of course, change for numerous reasons, not only as a result of joining a sustainability standards scheme. Changes to public policy, industry norms or access to finance, will all support or impede practice adoption. New technology may drive practice adoption even in the absence of changes to the above factors. A farmer may adopt new practices that are seen on a neighbouring farm that is certified, without undergoing certification themselves.

The ability to adopt practice will be further enabled or hindered by the development context within which the individual or organisation operates, along with the entity's own capacities and assets. Practices that require a cultural change – such as gender equality – may be difficult to embed in the absence of an understanding of why the cultural shift is needed.

Standards and certification systems are another potential driver of what practices are adopted and which are not. Standards systems define what constitute good practice and sustainable outcomes, and, as market-based mechanisms, provide incentives for their adoption. These incentives include customer specifications, access to markets, premium prices and risk management (e.g., reputational risk).

The reality that sustainability standards are only one of many factors promoting changes in practice within the production system places subsequent requirements on the standard of evidence needed to answer the research questions. A change in practice

could have occurred for many reasons, so we can only attribute that change to standards and certification with any confidence if there is a clear counterfactual, such as an otherwise similar organisation subject to the same internal and external changes except that it has not been certified.

1.4 Research approaches

Three main research questions are addressed by this review:

RQ 1: What is the effectiveness of standards and certification in driving the adoption of more sustainable practices by certified entities over a period of time?

RQ 2: How do standards and certification tools operate to achieve practice adoption? To what extent does adopting practices lead to continuous improvement in entities over time in identified thematic areas?

RQ 3: What lessons can standards systems learn for this area of their work?

As discussed above, recent systematic reviews that have researched questions on sustainability standards have found that there are relatively few studies that have robust methods, including counterfactuals. With this constraint in mind, we use four approaches to address the research questions. These are:

1. Systematic mapping;
2. A narrative review of the literature identified by the mapping;
3. Analysis of standards system data that was provided by some of ISEAL's members; and,
4. Key informant interviews.

These approaches provide different and complementary types of evidence (Figure 1). Systematic mapping (see below) provides core information on the amount and quality of evidence available to address the research questions. It also provides a filtered and manageable set of research from a large and chaotic literature of vastly varying quality and presents papers of direct relevance to the research questions that can then be the source for detailed narrative analysis.

The narrative analysis of the papers identified by the systematic mapping provides insight into what is happening and why. However, there are inevitably strengths and weaknesses in this body of evidence: questions, commodities and geographies that are rarely researched. We fill these gaps through analysis of sustainability standards system data and through interviews with expert informants. (see Figure 1)

1.4.1 Systematic mapping

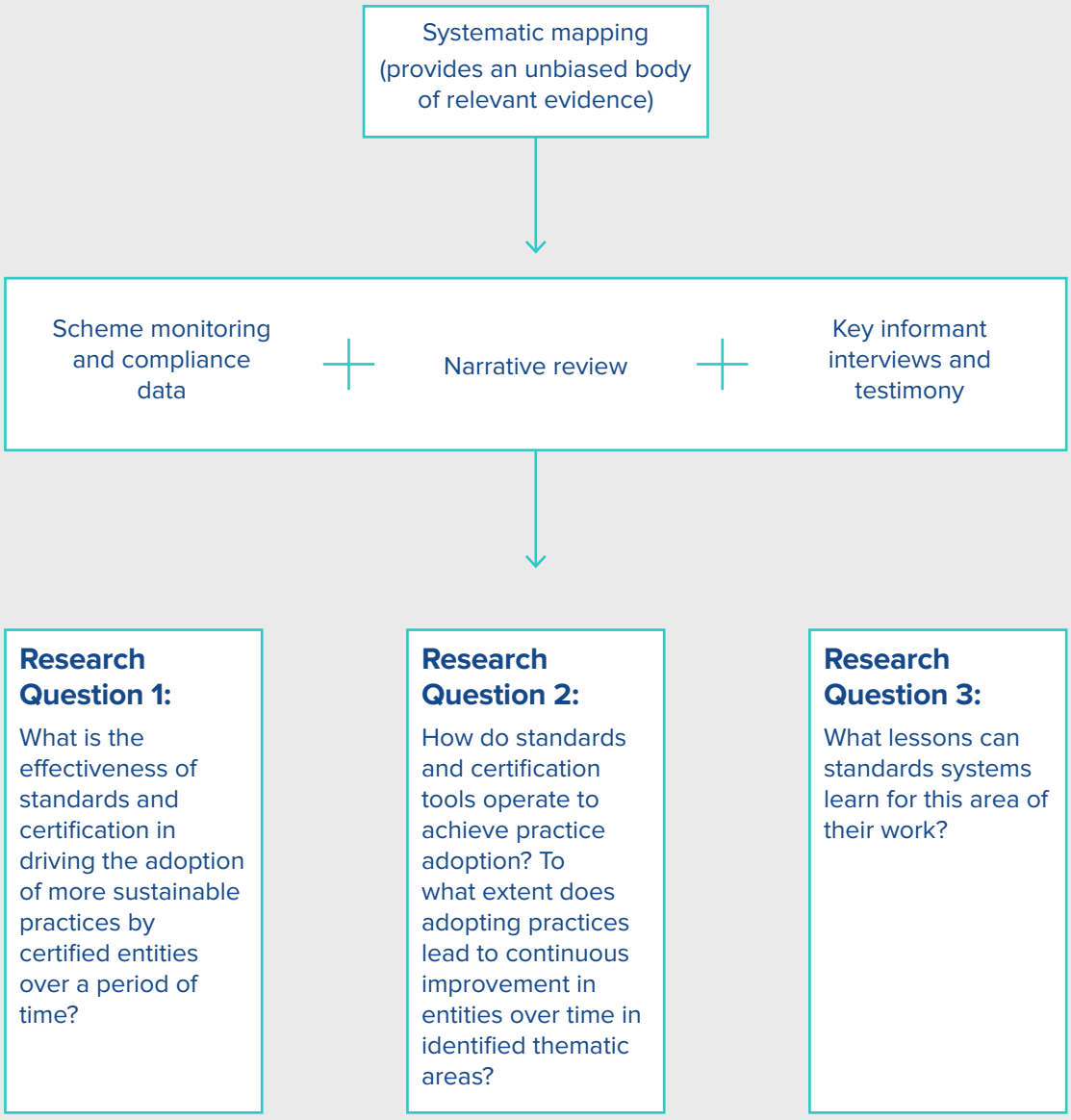
The particular set of questions posed in the current evaluation lends itself best to the systematic map approach. Systematic mapping is particularly valuable for broad, multi-faceted questions that can include multiple interventions, populations or outcomes (see Table 1 for definitions of these terms). They can cover the breadth of evidence needed for policy- or practice-relevant questions of the type under consideration here.

Systematic mapping follows the same rigorous processes as systematic reviews to evaluate relevant evidence and minimise the potential biases and lack of transparency of traditional literature reviews (James et al., 2016). The main differences between the two systematic approaches are summarised in Appendix A1.5.

In systematic mapping, the evidence is presented in a searchable database, with clearly

Figure 1:

Schematic representation of the research approach



defined elements that are coded similarly across the body of evidence collected through an extensive search of multiple sources (including academic journals and other sources of information, such as organisational databases, collections of theses, unpublished reports, and publications suggested by stakeholders interested in the review question). As systematic maps may include multiple populations, interventions or exposures, the database usually enables cross-tabulations of the data to be carried out to explore the evidence base thoroughly. It then becomes possible to identify trends and knowledge gaps and clusters. In further contrast with systematic reviews, systematic maps are unlikely to include detailed extraction of study results or statistical synthesis of results. The mapping process involves rigorous (i) searching for evidence, (ii) selection of relevant evidence ('filtering'), and (iii) presentation of key elements of individual studies in the evidence set ('coding'). These are described in detail in Appendix 1.

Table 1:

Key terms of a systematic review or map defined

Question element	Definition
Population (of subjects)	Statistical samples or populations of subject(s) (e.g., ecosystem, species, etc.), to which the interventions will be applied, or exposed to described conditions.
Intervention/Exposure	Policy, action or environmental variable impacting the populations or to which the subject populations are exposed.
Comparator	What the exposure or intervention are compared to. Either a control with no intervention/exposure or an alternative intervention or a counterfactual scenario.
Outcome	Consequences of the intervention or exposure. All relevant variables that can be reliably measured.

Following good practice for systematic evaluation, the method for the systematic mapping was guided by an Advisory Group (Appendix A1.1), who represented a broad range of expertise in sustainability standards including research, policy, and practice. It was agreed that the review would include any activity certified to a recognised sustainability standard, and not be limited to the standards schemes of ISEAL members. The thematic areas were agriculture (including tea, coffee, cocoa, horticulture, livestock, and dairy), fishing, aquaculture, forestry (embracing forestry timber and non-timber products including forest fruits, edible fungi, bushmeat, etc.) and forest resources (forests as carbon sinks or bioreserves), textiles, mining, and handicrafts. Publications (from 1990 to October 2017) related to sustainability standards in these areas were searched in online bibliographic databases (covering academic journals, conferences, theses, books, and other reports) and on the websites of certification standards organisations. These were assessed in a series of systematic filtering stages by a team of reviewers using the methods described in full in the Protocol, developed in collaboration with the Advisory Group (Appendix 1), to arrive at a final set of publications which comprise the systematic map of evidence relevant to the research questions.

1.4.2 Narrative review of literature identified by the systematic mapping

The body of research identified by the systematic map provides an unbiased sample of published and unpublished research identified as being most relevant to the research question by a rigorous, transparent systematic mapping process. We use this body of literature to investigate the research questions for each of the six thematic areas of focus. This differs from a conventional literature review in the sense that the body of research is not selected on the basis of individual knowledge or from a literature search tailored to what is easily available.

This body of literature reports a range of outcomes, including evidence of practice adoption as well as no difference in practice adoption between certified and non-certified entities. In the absence of statistical meta-analysis, it is not possible to infer which of these reported outcomes is likely to be correct. It is common in conventional literature

reviews for weight to be given simply to the number of studies reporting any given result, but this 'vote counting' approach is discredited and does not form part of this systematic analysis.

1.4.3 Analysis of standards' monitoring and compliance data

Many standards schemes have spent substantive resources putting in place systems that capture data on various types of change. This data is used internally for scheme monitoring and evaluation purposes, but also represents an important body of evidence on questions around practice adoption. However, some schemes lack the capacity to use this data for public-facing research, partly due to confidentiality issues and also due to limited resources available to analyse the data. This may be beginning to change, as a recent analysis of Marine Stewardship Council non-conformity data shows (Gorham et al., submitted).

This study provided an opportunity to complement findings from the systematic mapping with insights from analysis of scheme data. All ISEAL member schemes were approached to contribute data to this study, with six ultimately being able to do so. The data were independently analysed by the lead researchers under strict confidentiality norms (Appendix 2 provides a full description of the methods used).

1.4.4 Key informant interviews

The final research approach, key informant interviews, were undertaken to add insight into sectors (e.g., mining) and questions (e.g., motivations for practice adoption) that are not fully covered by the preceding research methods. The nine informants were selected for their experience in otherwise under-represented areas and were interviewed using a semi-structured interview under the Chatham House rule.

Interviews were conducted in December 2017 and January 2018, with each interview lasting 45-60 minutes. This is an innovative strategy to extend the collaborative stakeholder approach always taken in systematic reviews - framing review questions and developing the precise review method – to include a qualitative element of interviewing stakeholders to fill in some of the knowledge gaps revealed by the systematic mapping.

1.5 Structure of the report

The main part of this report has three sections:

- **The state of knowledge review.** This section reports the outcomes of the systematic mapping and describes the quality, quantity and characteristics of the research that is identified as being most relevant to the research questions.
- **Thematic assessments.** This section focuses on six sustainability themes (in broad terms: two social, two environmental and two economic), describes the output of the systematic mapping for each area, provides a narrative review of the research identified by the systematic mapping, and adds information and insight from the analysis of certification system data and from expert informant interviews.
- **Conclusions and recommendations.** This section draws out the main findings from the preceding sections and provides recommendations for different stakeholders.

2 The State of Knowledge: results of a systematic mapping process

This section presents a narrative summary of the systematic mapping process undertaken following the method outlined in the Protocol (Appendix 1). The results are also presented visually on an interactive global map at: **https://oxlel.github.io/evidencemaps/certification/**

2.1 Method

2.1.1 Search

The comprehensive search strategy in three major bibliographic databases:

- Web of Science published by Thomson Reuter's (formally ISI) Web of Science, New York, USA **http://apps.webofknowledge.com/**
- SCOPUS published by Elsevier **http://www.elsevier.com/online-tools/scopus**
- CAB Abstracts published by CAB International, Wallingford, UK **http://www.cabdirect.org/**

To ensure coverage of sources not captured by the above databases the meta-search engine Google Scholar (http://scholar.google.com) was searched using Harzing's Publish or Perish open-source software (https://harzing.com/resources/publish-or-perish). The first 2000 records were exported.

Searches were carried out according to the Protocol (Appendix A1.3 of the Protocol shows all key words and the logic used to combine them into search strings) following good practice guidance (Livoreil et al., 2017). The comprehensiveness of the search was tested against a reference set of 15 papers (referred to in figures and tables as 'Golden') known to be of direct relevance to the review questions (Appendix A4 of the Protocol).

Searches in the bibliographic databases and Google Scholar yielded 23,628 articles. These were screened for relevance to the review questions in a sequence: title-only, title-plus-abstract, full-text. The rule in systematic review is to retain any article that is not wholly and clearly out of scope, which means retaining a large number of papers that may be rejected at the next stage of screening, or not rejected until a thorough examination of the full text. The process is time-consuming and labour-intensive. We therefore deployed machine-learning at the title-only stage using a commonly-used tool for systematic review, Abstrackr, which 'learns' from decisions made by review screeners over several hundred records. It then begins to sort records into 'most likely' to be useful, while suggesting how many of the remaining records are likely to be not useful.

Combining machine-learning with human screening is becoming more widely recognised as an important contribution to enabling systematic reviews to be completed in shorter periods of time. It is also more reflective of resources available for this type of work from funders and policy-makers (Thomas et al., 2017).

Searches of organisational websites were screened for relevance before downloading, accounting for the relatively low number from these sources shown in Figure 1. References from the recent systematic review by Oya et al. (2017) were assessed for relevance and downloaded if relevant and not already captured in the database searches. Figure 1 shows the progress of all articles screened during the review process.

A flyer describing the project and seeking relevant non-journal literature was prepared and distributed at ISEAL's Members' Week 2017 (Appendix A5).

2.1.2 Screening for inclusion

The criteria used to determine whether an article was included are documented in the Protocol, and are repeated here for clarity of interpretation of the results presented.

It is important to distinguish here between 'studies' and 'articles'. In systematic review terminology, the word 'studies' refers to independent investigations which may be reported in a single 'article' (any document, not confined to journal articles) or multiple articles (where results taken over different time periods are reported separately). There were few of these cases in the current review, but they occur frequently in the environmental and forestry literature.

Relevant Population

Any entity that is the direct receiver of certification by a sustainability standard was treated as the relevant population. This encompassed parts of the supply chain beyond producers (e.g., trader, buyer etc.). Groups involved in supply chain custody but not directly certified were not included in the systematic mapping.

Relevant Intervention

Studies reporting on entities that are certified under a sustainable certification scheme were included. These were not limited to those represented in ISEAL Members.

Relevant Counterfactual

Studies were included which compared certified entities with similar uncertified entities, or compared before-and-after certification in a single entity or groups of entities. The aim was to include only studies that were explicit in attributing changes in behaviour, outcomes or impacts, directly to certification; not those where changes were simply a result of developing better practice for other reasons. The review has attempted to differentiate between these two possible pathways to changed practice by the requirement for a study design that has a counterfactual. Studies were not included if they compared two or more certified entities without considering an uncertified entity (or a before-after comparison of the multiple entities), whether different certification standards or the same.

Relevant Outcomes

Although the review question deals with practice adoption and change, the application of the concept of 'Outcomes' from standard systematic review methodology was applied to the current review because there had to be evidence of measured change. This change of practice fits well with the notion of impact as measured by outcomes.

Through discussion with the Advisory Group a very simple classification of practices adopted in response to standards was devised. Studies were included if they reported data on at least one of these pieces of information. The documented outcomes were grouped into three broad categories (Social, Economic, and Environmental) for ease of document handling, however, the interplay between the three categories is recognised.

General inclusion criteria

Studies were **included** if they reported:

- A sustainability standard/certification (not restricted to ISEAL member standards);
- The study design included a counterfactual – a clear comparison(s) between certified and non-certified entities or in the same entity before and after certification (where the situation before is considered the counterfactual to the situation after certification);
- At least one of the agreed sustainability outcomes of focus in this study.

It is recognised that there is a very large amount of conceptual and theoretical literature on sustainability standards and practices, these studies were excluded unless they were

reporting controlled studies with measured outcomes. This exclusion criterion should not be interpreted as making a statement on the inherent quality or degree of interest in such studies. They serve a different purpose from that undertaken in the current review. Likewise, there is likely to be a large literature that discusses the influence of standards on consumer behaviour and on policy development, but because these arenas were outside the area of interest of the review (practice change related to key sustainability impacts), these studies were also excluded.

Studies were **excluded** if they reported only:

- Issues such as consumer, supply chain, chain of custody certification and so on;
- Policy based/regulations/co-ops, unless there is an explicit standard mentioned;
- Appellation of origin/protected designated area, etc.;
- Food safety/quality;
- Farmer field schools or Co-ops, unless they also reported on certification;
- Studies that reported empirical results of perception surveys amongst the population, but did not actually measure actual change were likewise rejected;
- Kosher or Halal (not sustainability standards per se);
- Land ownership certification.

Studies in languages other than English, French, Spanish and Portuguese were excluded. The original aim had been to include narrative qualitative evidence (for example interviews or focus group meetings that were reported as dialogue with no numeric analysis), but this was not pursued in the systematic map arm of the project, because it was given prominence in one of the other two strands of work (key informant interviews).

Screening was carried out by two independent research assistants and, following best practice guidance (Frampton et al., 2017), Kappa tests were performed at the start of each screening stage on a random subset of 100 studies to confirm alignment of criteria interpretation between the reviewers (Cohen, 1960).

2.1.3 Data extraction/coding

A total of 24 variables were coded by seven research assistants. They recorded only the presence of data but not the results presented in the studies because the systematic mapping does not include statistical analysis of reported data.

The variables were: Education, Poverty, Health and Safety, Nutrition, Gender, Wellbeing, Community Benefits and Development, Knowledge Exchange, Transparency, Management Systems, Price Premium, Profitability, Costs, Market Advantage, Yield, Good Production Practices (including good forest management), Financial Security, Deforestation, Conservation and Biodiversity, Pollution, Soil Management, Water Management, Input Use, and Post-Harvest Practices.

In the initial stages of coding, studies were assessed by the lead coder and a second coder sitting together and discussing interpretation of the outcomes to check consistency of coding. The lead coder changed near the end of the project, but the same process was undertaken with other coders who started working on the project. This method of working collaboratively ensured a high degree of consistency.

In addition to the 24 outcome variables, details of the studies (publication details and geo-location) were also recorded. Where geolocation data was not presented in studies, they were determined from Google maps using the names of study locations in the paper. For studies on large geographical areas, a point near the nearest town, capital city, or mid-ocean point was recorded.

2.1.4 Mapping and visualising the studies

All included studies were recorded in a consistent way on a spreadsheet in Microsoft Excel (one of the project outcomes and lodged with the ISEAL Alliance) and presented on an interactive global map following Thorn et al. (2015) and Martin et al. (2017) on a platform hosted by University of Oxford's Long-term Ecology lab (**https://oxlel.github.io/evidencemaps/certification**).

The latitude and longitude of the study population was coded. Where multiple locations were recorded in the coding sheet the study appears as a circle on the world map in each location, colour-coded for commodity. Filters can be selected to enable users to interrogate the map for combinations of different variables. The map has been designed to be useable on most commonly-used devices (PC, tablet, mobile phone). Any data added to the systematic map over time can theoretically be drawn into the online global map, therefore, any updates and amendments will be possible in response to stakeholder feedback and popular use.

2.2 Results of the systematic mapping

Figure 2 shows the number of studies considered at each stage of the screening process. When duplicates had been removed, over 13,000 studies were identified from the search. Further screening reduced this to 371 articles for which the full text was reviewed, and of these, 116 were included in the final map. The most important reason for exclusion was the lack of a counterfactual or comparator, noting that this is likely to be an underestimate for the reasons outlined above.

Figure 2:

Stages of the selection process showing numbers of articles assessed at each stage. NB: 'Not processed' includes 20 articles coded as language-exclude, which should have been passed to a coder with French, Spanish or Portuguese knowledge and 13 marked as 'Query'.

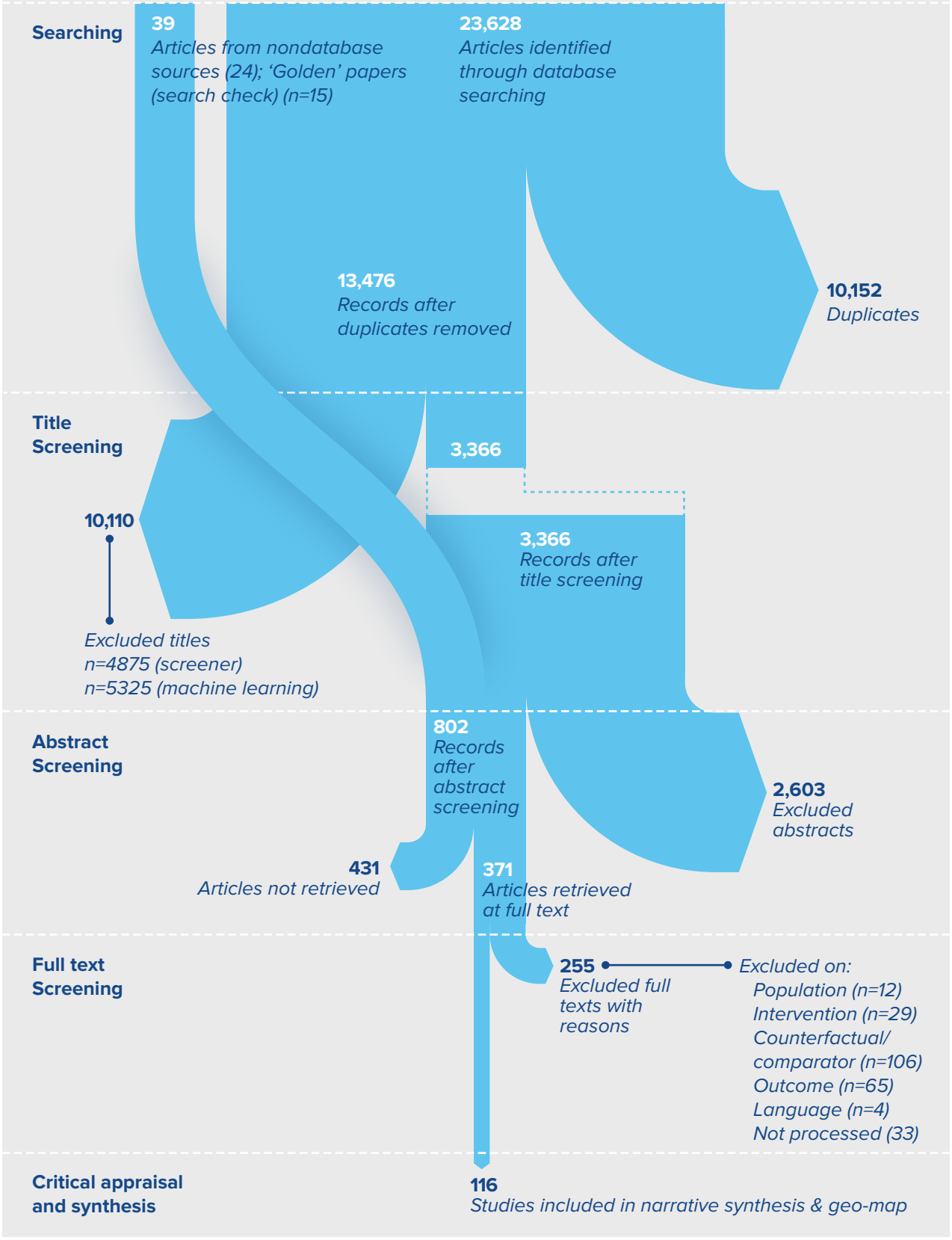


Figure 3:

Searchable map of the final set of studies. (Size of markers indicates the number of studies for the location)

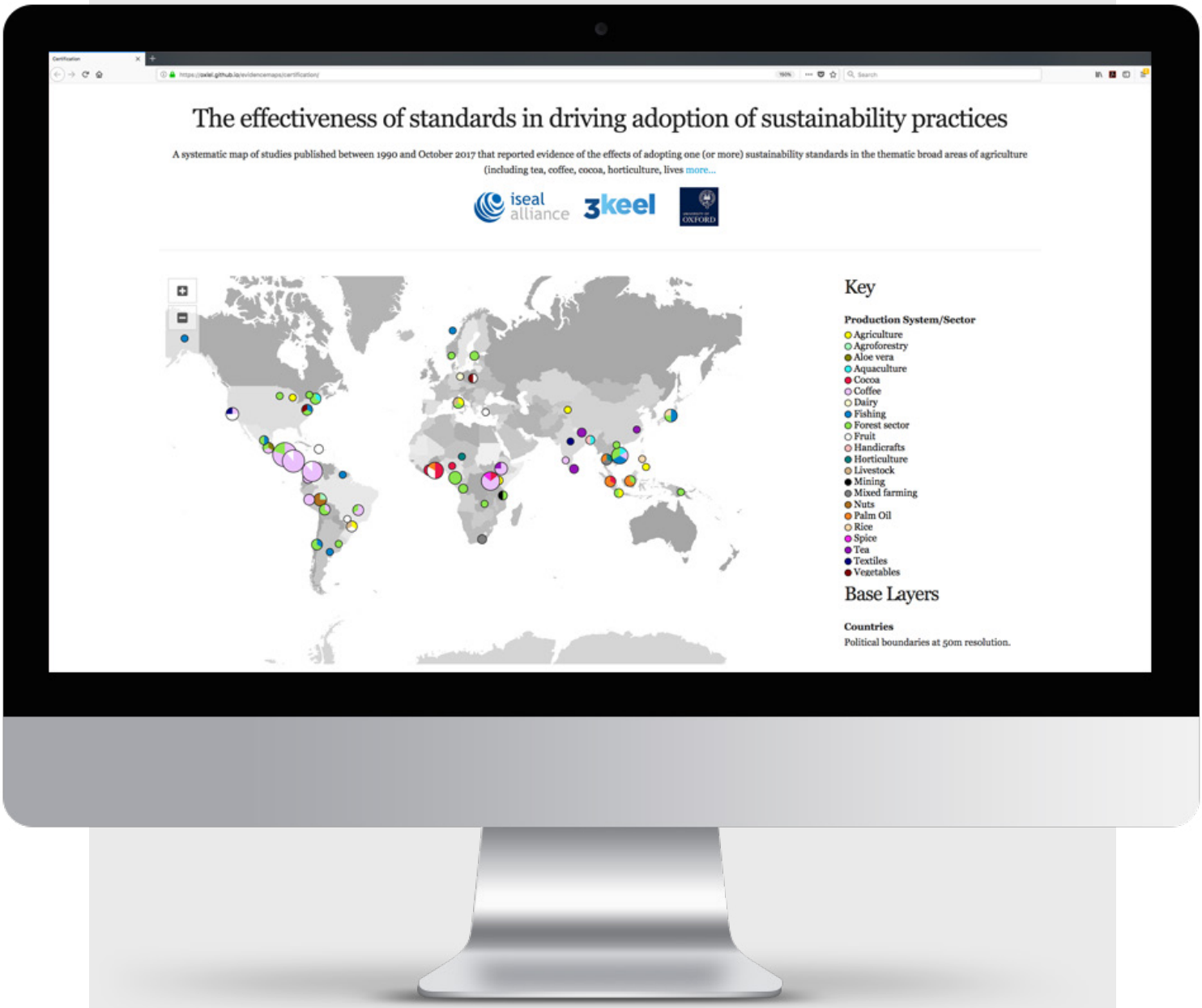


Figure 3 shows the commodities reported in the filtered studies by geographical location(s). This interactive map can display any of the information extracted from the filtered papers according to the location(s) of the study. The map is available at **https://oxlel.github.io/evidencemaps/certification/**

Figure 4:

Number of articles by publication year

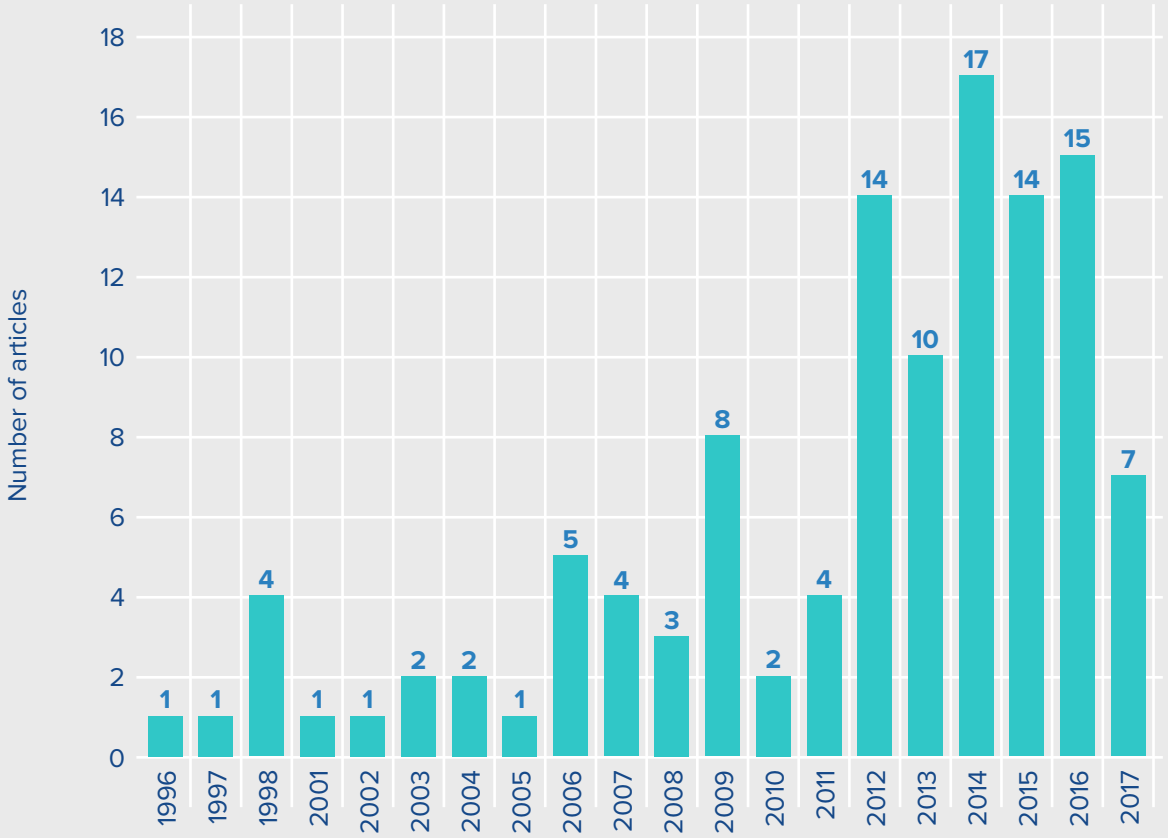


Figure 4 shows the number of articles by publication year. The majority of studies are from Journals published within the last five years (72% since 2010), which is typical of many systematic reviews and maps, owing to the exponential increase in publications over time. Journal articles predominate, providing 83% of the papers identified by the systematic mapping, and with contributions from reports, conference papers, and other types of publications (Figure 5).

Figure 5:

Publication type of articles included in the map (n=116)

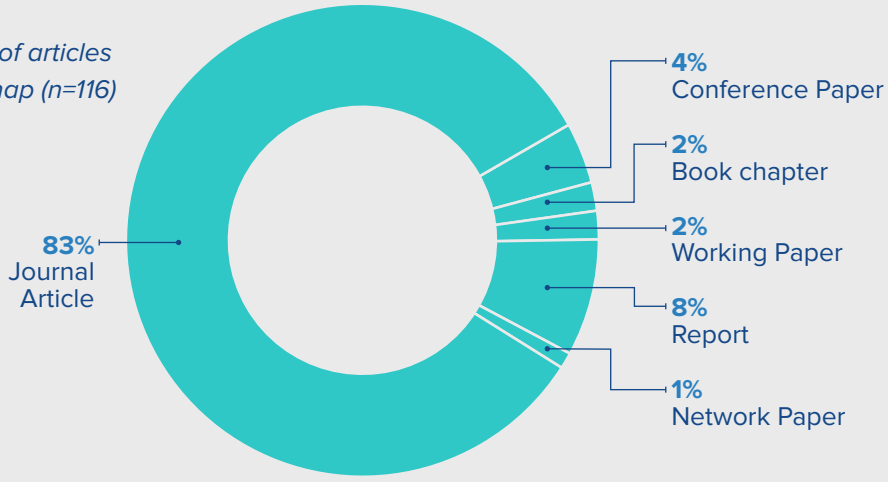


Figure 6:

Number of studies by certified commodity. Many studies reported more than one commodity.

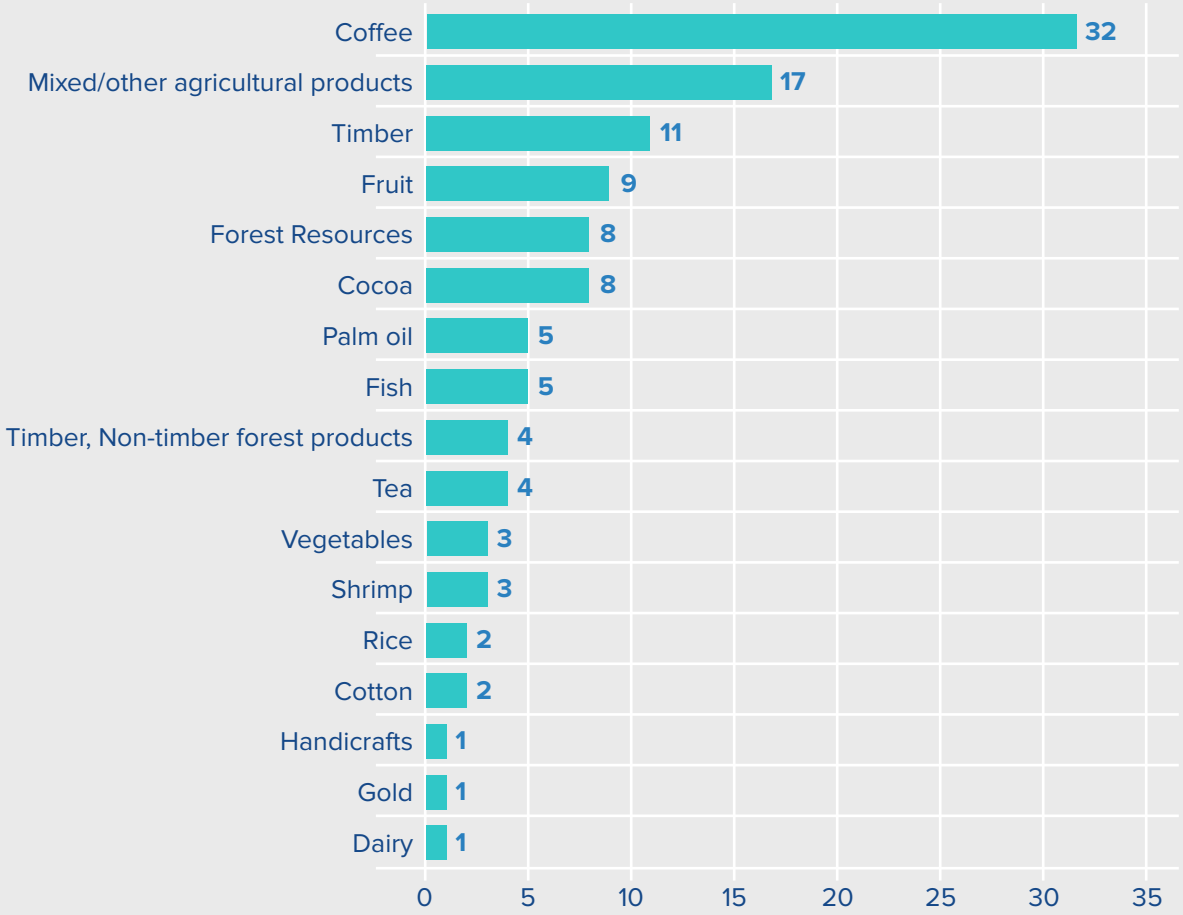


Figure 6 shows the number of studies by certified commodity. Where more than one commodity were reported in a study, all commodities were coded, and 'mixed products' refers to certification of multiple products that are not differentiated by the study. Studies on coffee predominated, with relevant outcomes reported in 32 of the studies identified by the systematic map. A significant proportion of studies are from the forestry sector, either timber, forest resources and/or non-timber forest products studies (23 studies combined). Few studies reported on mining, cotton or livestock.

Table 2 shows the number of studies reporting outcomes for specific sustainability standards. Multiple standards were reported in many studies, with The Rainforest Alliance, Organic, Fairtrade and Forest Stewardship Council predominating. These four standards account for 70% of the total number of studies reviewed in this research. There are also a large number of other standards that are reported in two studies or fewer.

Table 2:

Sustainability standard focus in reported studies (multiple standards were reported in many papers)

Certification scheme	Number of studies
Rainforest Alliance	37
Organic	37
Fairtrade	24
FSC	21
UTZ	9
Global G.A.P.	7
Marine Stewardship Council	6
ISO14001	3
Others (reported in two studies or fewer)	25

Figure 7:

Number of studies reporting specific variables. Many studies included multiple variables

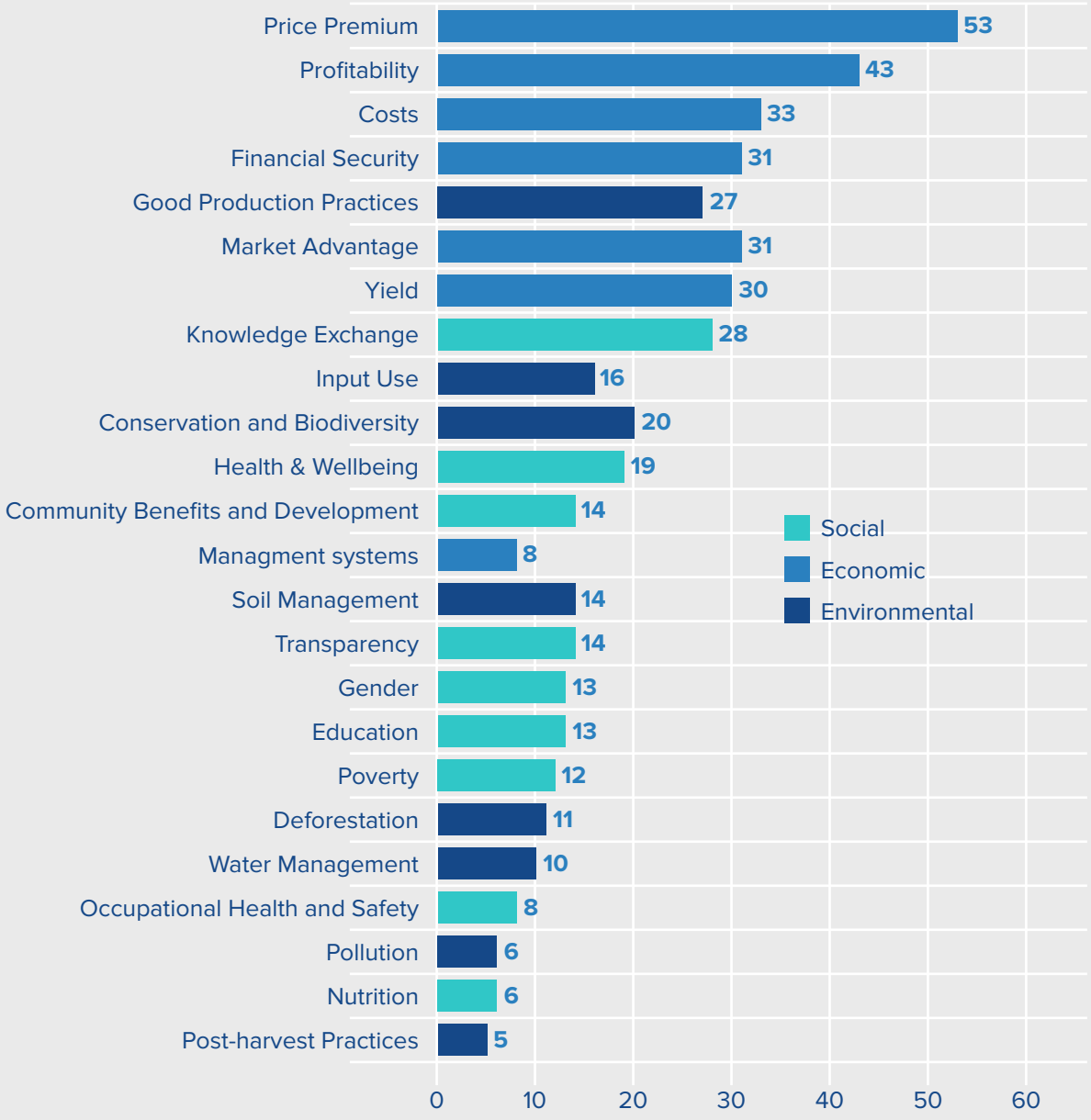


Figure 7 shows the number of studies for certified commodities and Table 3 shows how the reported outcome variables in the three categories Social, Economic, Environmental varied between sectors by specific variables. Economic variables dominate, with environmental variables the least well represented. Studies typically report several specific outcomes. The most common ones reported are realization of a price premium, profitability and costs.

The outcomes that were selected for inclusion include a mixture of final impacts and practices (see Section 1.3). The reason for including papers with this range of outcomes was because it is impossible to know what evidence is available at the beginning of the research process. Specifying a broad range of outcomes therefore increased the chances of identifying reports that would be or use to the research question from the mapping exercise. Examples of reported outcomes that are essentially impacts of certification include price premium and financial security. In addition to reporting impacts, these papers sometimes also report the practices that have led to the impacts reported. Frequently reported examples of reported outcomes that are practices include Good Production Practices, Input Use and Conservation & Biodiversity. These papers are almost always directly reporting data on whether practices have changed or not, or by how much they have changed. For more detailed narrative analysis, it is therefore this second group of reports, that are primarily focused on practices, that is of most direct relevance to this research. The very few studies on textiles and mining is striking and a clear knowledge gap.

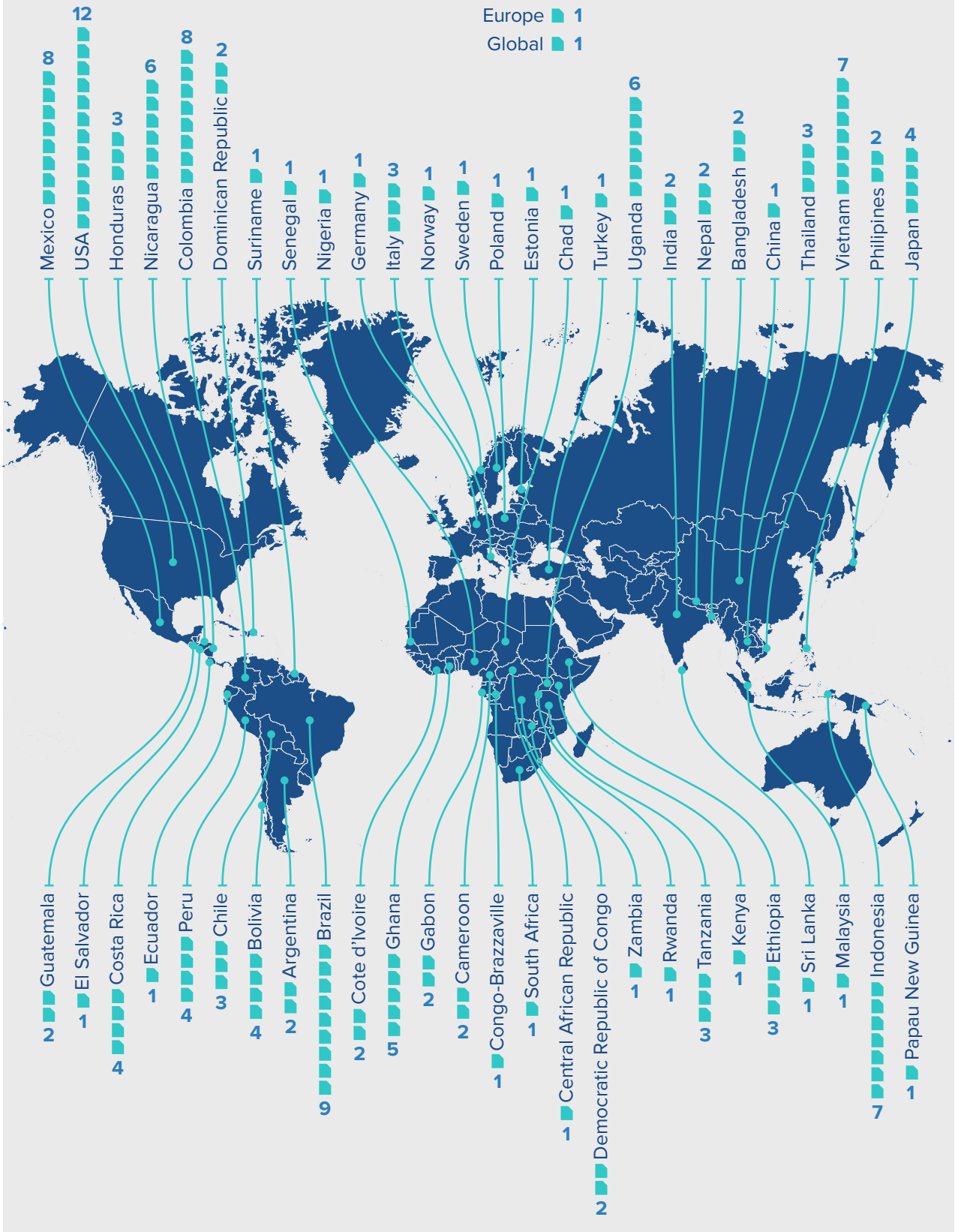
Table 3:

Outcome variables by production sector. Some studies report multiple variables and multiple sectors, all of which are recorded.

Production sector	Social Variables	Economic Variables	Environmental Variables
Coffee	14	17	6
Timber/Forest Resources	9	11	8
Agriculture/Mixed Farming/Rice	5	6	6
Cocoa	6	5	5
Fruit	3	5	5
Fishing	1	4	1
Tea	4	3	2
Dairy/Livestock	1	3	1
Horticulture/Vegetables	2	3	2
Palm Oil	2	2	3
Textiles		1	
Mining		1	

Figure 8:

Frequency with which countries are reported from. Where studies report from multiple countries all are recorded



The distribution of countries reported from is shown in Figure 7. A total of 54 countries are represented. Latin America and the Caribbean is the best represented region, with multiple studies from Brazil, Colombia, Mexico, Nicaragua. Costa Rica and Peru all being amongst the most frequently reported countries. Asia and Africa are reasonably well represented, but there are few studies from Europe or North America (Table 4).

Table 4:

Regional representation of studies

North America	Latin America and the Caribbean	Africa	Europe	Asia
12	58	9	34	34
8%	39%	6%	23%	23%

2.3 Reflections on the systematic mapping

Mapping the evidence base for factors that may contribute to adoption of sustainability standards using the robust methodology developed for systematic reviews, shows clearly how unevenly evidence is distributed across sectors, commodities and countries. This finding is entirely consistent with those of other systematic reviews and systematic maps of similar topics. The research effort is not always driven by need for knowledge; it is dictated by trends in funding, which itself can follow the changing landscape of policy priorities and/or public interest and these trends only become apparent when knowledge is collated over long time periods.

The systematic filtering process has shown how relatively few studies that set out to assess impact have a research design that includes a counterfactual. This absence of controlled experimentation (whether in social or ecological studies) accounts for the huge drop-off from the many thousand that initially appear to have relevant information from broad searches, to the very small set of studies that make it through the selection process. This is a matter of some concern and something which can be addressed in future research requirements by funders, donors and policy makers.

As other systematic reviews have shown, the majority of research in this area concentrates on agriculture and forestry, with a specific focus on coffee and timber. There is very little evidence from fisheries and almost none on mining, sectors where sustainability standards are no less needed. It is likely that the principal reason for the difference between these sectors and commodities is age of the standards, but there are undoubtedly other factors that have led to a dearth of research in these key areas. The knowledge gaps shown in the systematic map could help stimulate a research agenda better attuned to evidence need.

It is also notable that the focus of much of the empirical research is on ‘ultimate outcomes’ rather than the more fundamental question of ‘practice adoption’. This to some extent limits the insights into questions of how standards lead to change, under what circumstances and for whom. Again, this is an important area for future well-designed primary research projects.

3 Key sustainability themes in focus

3.1 Introduction

The proceeding chapter provides an overview of the extent of the evidence that standards and certification systems drive changes in sustainability practices. This chapter looks in more detail at the evidence for practice adoption in some specific thematic areas. For each of these themes, we investigate the evidence for the three research questions (see the Introduction to the report).

Each sustainability theme is presented in the same manner: an introduction that briefly situates the theme within sustainability standards, a review of the evidence for practice adoption, assessment of changes over time, and the proposed mechanisms that drive (or hinder) adoption of that practice.

3.2 Materials and methods

3.2.1 Themes covered

We focus on six areas of practice, which together provide coverage of environmental, social and economic aspects of sustainability. These themes are: conservation and biodiversity; input use; community benefits and development; health and safety; good production practices; and management systems.

These six areas of practice were selected because they had the strongest combined evidence from the number of papers identified in the systematic mapping and from the data obtained from the standards systems that contributed to this study (See Appendix 2). The exception to this is the final one, management systems, which has a relatively limited literature but which was suggested as an important area for investigation by a number of key informants.

A broader range of thematic areas are covered within Section 3.3, below.

3.2.2 Evidence used

As described in the Introduction, four types of evidence were used to understand the research questions:

Information from the systematic mapping on the papers that reported outcomes in each thematic area. As in the preceding Chapter, this provides information on the state of knowledge for each thematic area, based on a transparent and rigorous method.

Literature review. The papers identified by the systematic map are reviewed to provide further information on the precise nature of the changes reported, how these might vary between sectors and contexts, the reasons for practice adoption and non-adoption, differences between certified and non-certified entities (or pre- and post-certification), and the mechanisms that drive practice adoption.

Analysis of standards' monitoring and compliance data. As described in the Section 1.4.3, data from the monitoring and compliance systems of six ISEAL Alliance members was made available.

The data was analysed in two complementary ways to generate insights into practice adoption. Firstly, non-conformity data was analysed to show the thematic and temporal pattern of non-conformities. Non-conformities recorded by standards schemes when independent auditors find that the performance of a certified entity does not meet the

requirement of the standard. The underlying rationale is that a decrease in the number of non-compliances over time indicates that the practice of certified entities is becoming increasingly aligned to the standard. Note that only criteria applicable for the whole period of certification were included, as in some standards schemes the requirements increase over time. However, this data is essentially about practices not in compliance with a standard rather than about practice adoption per se, and so does not, for example, take into account instances when a certified entity 'flip-flops' between compliance and non-compliance on a particular issue (which arguably indicates that practice adoption has been superficial).

The second approach used was to develop a metric to measure practice adoption from a variety of different types of data (including non-compliance data and performance data). This metric coded changes as either practice adoption (where a positive change was indicated); maintenance of practice (if there was either no evidence of change or where change was followed by reversion to previous practice), or 'unresolved' if a decrease in practice had occurred at the end of the time period.

Both metrics are presented as an average rate per certified entity, because the number of certified entities providing data varies both over time (fewer entities have long time series, see Table A2.1 in and Figure A2.3 in Appendix 2), and with each practice area (e.g., non-compliances are not evenly distributed between criteria). The metrics are presented for broad sustainability themes, as the precise requirements of standards vary for any given area of practice. The results are therefore an aggregate, and not necessarily fully applicable to any of the individual schemes that provided data. A full description of the data and methods used is given in Appendix 2.

Key informant interviews. Semi-structured interviews were conducted with eight informants in areas and topics not well covered by the literature, including the role of supply chain actors, sectors such as mining, and the motivations for practice adoption.

3.3 Conservation and Biodiversity

3.3.1 Introduction

There are a wide range of practices that are intended to improve conservation and maintain or increase biodiversity, and the requirements of standards on conservation and biodiversity are formulated differently in standards working in different production sectors. Practices essentially include measures designed to protect natural habitat and species, and in some cases the wider landscape (Table 5). These practices are, of course, not the only sustainability practices that can yield positive impacts on biodiversity: other environmental practices (e.g., pollution control) and production practices (e.g., soil management, use of inputs) all contribute to conservation and biodiversity. However, in this section we limit the discussion to evidence on activities that are explicitly framed around conservation and biodiversity.

There is a growing literature evaluating the impacts of certification in terms of habitat protection and species diversity. This literature often assesses the practices that have brought about positive impacts on biodiversity as well as the biodiversity itself, and therefore enumerates some important changes in practice. It should though be noted that it has been argued that some production systems, such as large-scale plantations, have such inherently low biodiversity value that adopting conservation practices within them is unlikely to have a significant impact (Azhara et al.,2015). This suggests that in some cases there may be a disjunction between the practice that standards require certified entities to adopt and the likely conservation impacts of those practices.

3.3.2 Evidence of practice adoption

The systematic mapping identified twenty papers reporting outcomes relating to conservation and biodiversity. Eight of the papers identified by the systematic mapping are from the forestry sector, three from coffee and the remaining studies deal with crops such as bananas, cocoa, tea and fisheries. Fourteen of the papers are from the tropics and sub-tropics, with six from temperate zones. Half of the papers report improved conservation outcomes in certified entities, with the remaining studies showing either a mix of outcomes depending on the metric or the context, or no change between certified and non-certified.

This body of literature provides some evidence of sustainability standards leading to the adoption of conservation and biodiversity practices in certified entities. FSC certified forests in Vermont, USA reportedly implemented measures to retain more dead wood – an important habitat for many forest invertebrates – than non-certified forests (Azhara et al., 2015). The majority of surveyed managers of certified forests in Chile and Argentina indicated that they had increased protection for threatened species, planning for biological diversity, and old growth reserves in order to receive certification, as well as making a number of procedural changes relating to conservation and biodiversity (Cubbage et al., 2010). Forest stands in Norway that had been harvested after the introduction of certification showed an increasing retention of trees and greater width of buffer-strips left along rivers, bogs and lakes, when compared to stands harvested before

Table 5:

Illustrative requirements within sustainability standards that are focused on conservation and biodiversity

Certification Scheme	Scale	Practice required by the standard
Marine Stewardship Council	Species	PI 2.3.1 The UoA meets national and international requirements for protection of ETP species
Roundtable on Sustainable Palm Oil	Species	Indicator 5.2.2: Where rare, threatened or endangered (RTE) species, or HCVs, are present or are affected by plantation or mill operations, appropriate measures that are expected to maintain and/or enhance them shall be implemented through a management plan.
Better Cotton Initiative	Species	Criterion 4.1: Practices are adopted that enhance biodiversity on and surrounding the farm
LEAF Marque	Habitat	Traditional field boundaries, environmental/landscape features and other natural habitats are retained.
Rainforest Alliance	Habitat	Existing native vegetation outside natural ecosystems is maintained, including: a) Existing agroforestry shade tree cover; b) Existing vegetated zones adjacent to aquatic ecosystems; c) Large native trees, except when these pose hazards to people or infrastructure.
Responsible Jewellery Council	Wider landscape	Criterion 3.6.2: Members in the Mining Sector shall respect legally designated protected areas by ensuring that: a. Members have a process to identify nearby legally designated protected areas. b. Members comply with any regulations, covenants or commitments attributed to these areas. c. Decisions to proceed with exploration, development, operation and closure activities take into account the presence of, and impact on, legally designated protected areas
Forest Stewardship Council	Wider landscape	Criterion 6.8: The Organization shall manage the landscape* in the Management Unit to maintain and/or restore a varying mosaic of species, sizes, ages, spatial scales and regeneration cycles appropriate for the landscape values in that region, and for enhancing environmental and economic resilience

the introduction of certification (Sverdrup-Thygesen et al., 2015). Certified companies were the only ones in a survey in the Congo Basin that implemented measures such as protection of ecologically sensitive areas, control of hunting, along with a suite of good production practices that have positive environmental impacts, such as reduced impact logging and improved road planning (Nasi, Billand & Vanvliet, 2012).

In the coffee sector, Rainforest Alliance certified farmers reportedly retain more forest than non-certified producers in Colombia (Rueda, Thomas & Lambin, 2015) and Ethiopia (Takahashia & Todo, 2014). The latter study quantified this as an increase in the probability of forest retention by 19.3%. Mexican Bird Friendly® certified coffee farmers retained more shade trees, with a greater canopy cover, than non-certified shade grown farmers (Caudill & Rice, 2016).

Some of the key practices around conservation and biodiversity in the marine fisheries sector include marine mammal bycatch, finfish discard rates, and gear impacts. Selden et al. (2017) compared the incidence of these between MSC-certified fisheries and non-certified fisheries in the USA. Discards of non-target finfish and bycatch rates of marine mammals were indistinguishable between certified and non-certified U.S. fisheries. MSC-certified fisheries did have lower average gear destructiveness scores than non-certified stocks when weighted by landings, but not when weighted by the number of fisheries.

Analysis of standards schemes' monitoring and compliance data illustrates further the role of standards and certification in driving the adoption of conservation and biodiversity practices (Figure 11). This analysis is based on the metric of practice adoption that was derived from data on nearly 2000 certified entities. The metric coded changes as either practice adoption (where a positive change was indicated); maintenance of practice (if there was either no evidence of change or where change was followed by reversion to previous practice), or 'unresolved' if a decrease in practice had occurred at the end of the time period. It is limited to substantive changes in practice 'on the ground' with procedural changes (e.g., changes in documentation) excluded. Data is presented per certified entity, so that, for example, a score of 0.1 would imply that one certified entity out of every one hundred has adopted a substantive practice within this thematic area. Full details of the data and analytical methods are given in Appendix 2.

Figure 9 shows that the estimated rate of adoption of substantive conservation and biodiversity practices is 0.16 per certified entity. This was the third highest rate of practice adoption of any thematic group, which suggests that adoption of conservation and biodiversity practice may be a reasonably frequent occurrence for certified entities. In addition, practice maintenance was recorded at a rate of 0.3 per certified entity. This is when either practice has remained stable, or when an entity has recorded a negative change in practice, followed by a return to a previous level (i.e., no overall change). This suggests that certification may play a key role in maintaining conservation and biodiversity practice over time, preventing slippage into inferior practices.

However, not all research into the role of standards in driving practices related to conservation and biodiversity outcomes provides evidence of practice adoption: a study comparing FSC and PEFC certified forests with non-certified forests in Sweden concluded that *"Our data analyses, using several different sources, still cannot show the cause–effect relationship between forest certification and environmental protection"* (Johansson & Lidestav, 2011). The authors cited different harvesting practices, inherent differences between certified and non-certified forest managers, and the impact of a damaging storm as potential confounding factors in their study.

3.3.3 Changes in conservation and biodiversity practice over time

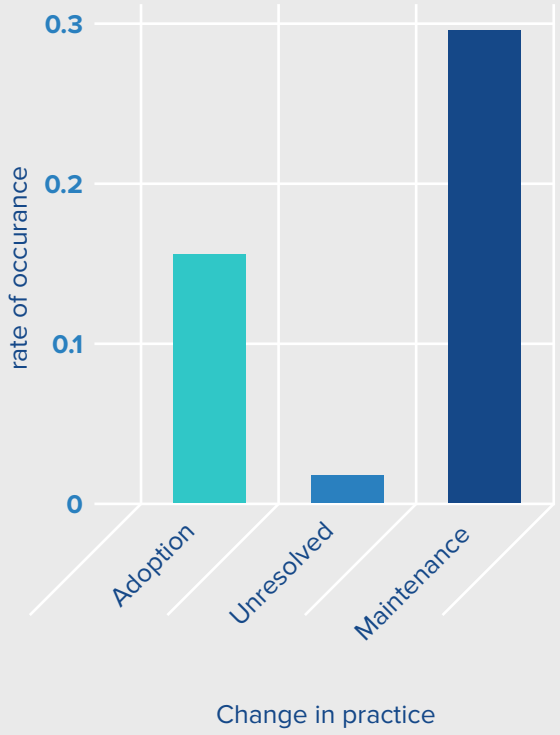
The published literature is largely based on relatively short time periods and single surveys, with few studies evaluating changes in conservation and biodiversity practices amongst certified entities over time. However, one of the studies identified by the systematic mapping did find evidence that certified Colombian coffee farmers increased a range of conservation practices after becoming certified. These practices included

maintaining a diversity of shade trees, watershed protection, and infrastructure for wastewater management (Rueda & Lambin, 2013).

Figure 9 provides a suggestion of how certification might influence conservation and biodiversity practices in certified entities over time. As reported in the previous section, there are an estimated 0.3 incidences of maintenance of substantive conservation and biodiversity practice per certified entity. This suggests that certification plays a role in the maintenance of conservation and biodiversity practice over time, preventing slippage into inferior practices. This interpretation is consistent with the low rate (two per one hundred entities) of certified entities that have cases of unresolved conservation and biodiversity practice (e.g., where a non-conformity has been raised but either the data on its closure is not available or that it has not been closed). In addition, analysis of standards schemes' non-conformity data (see Appendix 2 for details) shows a distinctive pattern of conservation and biodiversity non-conformities over time. In year one, 0.58 non-compliances are raised for each certified entity Figure 10. This is followed by few non-conformities being raised on substantive conservation and biodiversity practices until year six, which shows a spike of 0.78 non-conformities per certified entity in year six, with few in between.

Figure 9:

The average number of incidences per certified entity of adopting, maintaining or having unresolved conservation and biodiversity practices



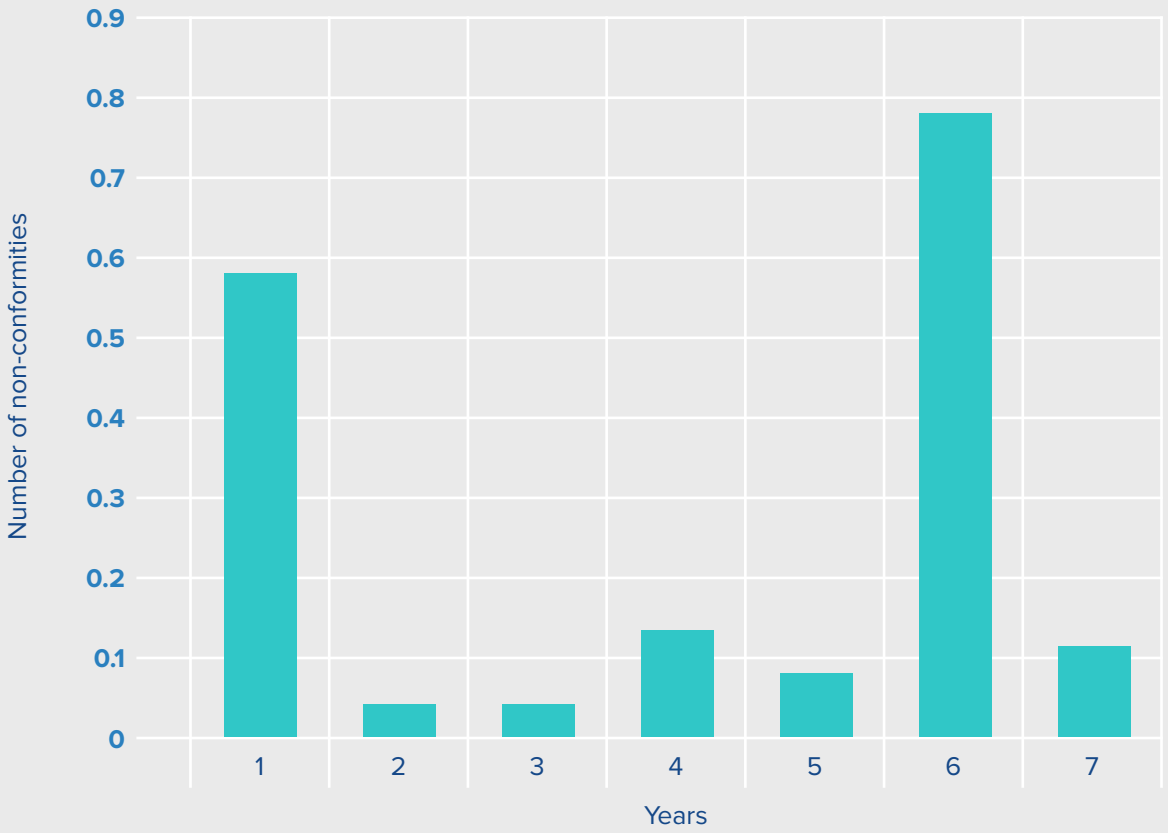
3.3.4 Reasons for practice adoption

The general motivations behind certification, such as access to markets, premium pricing, managing reputational risk, amongst others, are also likely to encourage certified entities to adopt conservation and biodiversity practices. For example, Rueda, Thomas & Lambin (2015) suggested that the mechanism by which certification caused certified coffee farmers to retain more tree cover than non-certified farmers was the 15-20% premium paid to certified producers for their coffee, a conclusion perhaps borne out by the indication that poorer producers with access to less land conserved proportionately more forest than wealthier producers. An almost identical conclusion was reached by Takahashia & Todo (2014) for Rainforest Alliance certified coffee producers in Ethiopia.

In addition to these general motivations for certification, the literature suggests some other mechanisms relating to conservation and biodiversity practice adoption. The first of these is information. Dumont et al. (2014) found that a higher proportion of certified cocoa farmers in Côte d'Ivoire (76 %) had received information about shade trees than non-certified farmers (15 %), however this had not translated into a greater diversity of trees in certified cocoa farms when compared with non-certified ones. Similarly, the entry of certification into the Norwegian forest sector was thought to bring greater awareness and knowledge of conservation practices (Sverdrup-Thygeson, Borg & Bergsaker, 2008).

Figure 10:

Average number of substantive non-conformities relating to conservation per certified entity



A second set of mechanisms have been suggested, that concern the process of certification itself. Sverdrup-Thygeson, Borg & Bergsaker (2008) suggest that Norwegian forest managers may adopt practices such as retention of riparian buffer zones as a result of non-conformities raised by independent, third party auditors. The focus of certification audits on performance in the field, as opposed to guidelines and policies that are somewhat removed from the practicalities of managing individual forests, were deemed to be an important aspect of the role that certification can play in driving practice adoption.

3.3.5 Summary

Much of the evidence for adoption of conservation and biodiversity practices come from studies on the impact of the forest certification, with most of the remaining research from tropical agriculture, including coffee. A large proportion of this literature indicates that entities certified in these sectors often improve their conservation with other studies showing mixed impacts or no change between certified and non-certified entities. Analysis of certification system data is consistent with the suggestion the adoption of substantive conservation and biodiversity practices may be relatively frequent amongst certified entities. In addition, verification audits may play a role in maintaining improved conservation practices.

Some authors suggest that certification and standards play a role in raising awareness of environmental issues, and in increasing knowledge of what actions can be taken, thereby supporting practice adoption.

3.4 Input use

3.4.1 Introduction

We include within this theme the requirements on the use of inputs that are found in most of the sustainability standards that cover agricultural production, and in others including forestry and aquaculture standards (e.g., FSC and ASC). The inputs include fertilisers, pesticides, herbicides, fungicides, antibiotics and growth enhancers,

The requirements of sustainability standards include outright prohibition of synthetic inputs (e.g., organic standards); enforcement of legal prohibitions with limits on the quantity and type of legal inputs that can be used; or more general invocations of 'wise use'. These restrictions are often twinned with encouragement within the standards for complementary or alternative approaches to the use synthetic inputs, such as increased use of organic fertilisers (and noting that the very closely related area of Integrated Pest Management is treated within the section on Good Production Practices). Finally, the importance of context is broadly acknowledged: whereas a European farmer might need to find ways to make input use more efficient, an African farmer may not be using enough

Table 6

Illustrative examples of requirements concerning input use from sustainability standards

Scope	Certification Scheme	Practice required by the standard
Legal	Rainforest Alliance	Criterion 3.4: No use of Rainforest Alliance prohibited pesticides and application of only legally registered pesticides
Legal	Better Cotton Initiative	Criterion 1.2: Only pesticides that are i. registered nationally for the crop being treated and ii. correctly labelled in the national language are used
Limitations on use	Bonsucro	Indicator 4.1.5: Agrochemicals applied per hectare per year Verifier: kg active ingredient/ha/yr. Standard: <5
Limitations on use	LEAF Marque	Control point 3.7 Steps are taken to minimise damage to beneficial and non-target species [when using plant protection products]
Encouraging additional/alternative approaches	UTZ	GB.47 Organic fertilizers and by-products available at farm level are used first and supplemented by inorganic fertilizer if nutrients are still lacking

of certain input types. Some illustrative examples of the requirements of sustainability standards are given in Table 6.

The motivation for standards to include consideration of input use seems to be predominantly environmental: criteria regarding input use are typically located within principles on the environment. However irresponsible use of inputs can also result in human health issues (e.g., pesticide residues in food).

3.4.2 Evidence for practice adoption

The systematic map identified sixteen papers that report outcomes on input use. Seven of these papers reported outcomes from Latin America, five from Asia, with the remainder from Africa, Europe and North America. The main crops systems studies were coffee (four), cocoa (two), timber (two) and a variety of horticultural crops (six). A significant proportion of this literature is evaluations of whether certified entities improve their use of synthetic fertilisers or pesticides compared with non-certified entities. This evidence is therefore directly concerned with practice adoption, and less focused on final impacts than the literature for some other thematic areas.

The findings of these studies are split between those that show that standards have resulted in either improvements in the use of synthetic inputs or an increase in use of alternative inputs (nine) and those which show no difference between certified and non-certified farms (two), a combination of difference in some metrics and no change in others (four), and an interaction between forest certification and input use in agriculture (one).

It is axiomatic that a conversion to organic agriculture reduces the use of synthetic inputs, but perhaps more pertinent are the practices that organic farmers adopt as a result. For example, organic tea farmers in China and Sri Lanka (Qiao et al., 2015), rice farmers in the Philippines (Mendoza, 2004) and vegetable growers (Fernandez-Cornejo et al., 1998) and soy farmers (Hartman et al., 2016) in the US essentially adopt a number of labour intensive practices to replace synthetic inputs as they converted to organic production. Costa Rican coffee farmers' decrease use of chemical pesticides, fertilizers, and herbicides and increased their use of organic fertilizer (Blackman, & Naranjo, 2012). Coffee farmers that were certified Fairtrade and organic in Peru used more organic fertilizer than farmers that were only certified organic (Ruben & Fort, 2011). The same study found no difference in organic fertiliser use between conventional Fairtrade certified and non-Fairtrade certified farmers on their coffee crop, but greater use of organic fertilizer by the certified farmers on their other crops.

Beyond organic certification, results are more varied. Several studies do demonstrate that certified farmers adopt practices around input use, including adopting alternative measures to compensate for reduced input use. For example, UTZ certified coffee farms in Colombia had higher levels of fertilization, and a higher share of producers applied fertilizers based on technical recommendations (García et al., 2014). Eight of the ten surveyed managers of certified forests in Chile and Argentina claimed to have adopted improved practices on the control and use of chemicals (Cubbage et al., 2010), and EUREPGAP certified fruit growers in São Paulo state, Brazil, used fertilisers more efficiently than non-certified growers (Lourenzani et al., 2006).

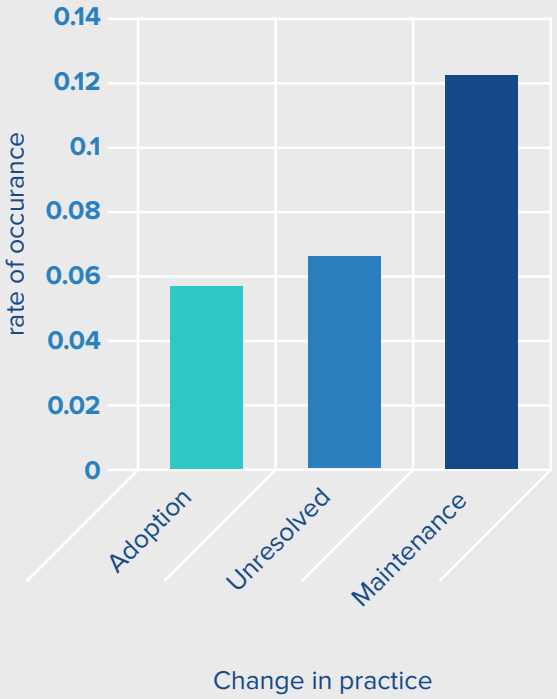
However, there are also reports of inconsistent or no differences in input use between certified and non-certified entities. RSPO certified palm oil plantations in Thailand used less glyphosate herbicide for weed control than non-certified plantations, but there was little consistent difference on fertiliser use (Saswattecha et al., 2015). The use of the

main insecticide was not significantly different between the Rainforest Alliance certified and non-certified coffee farmers in Colombia (Rueda & Lambin, 2013). Both certified and noncertified farmers applied organic and synthetic fertilizers and used synthetic herbicides, with no significant differences between the two groups with regard to these practices. Similarly, there was little difference in the application of pesticides between Rainforest Alliance certified banana farms and non-certified farms in Costa Rica (Bellamy et al., 2016). Pesticide management practices did not differ between GAP and non-GAP horticulturalists in Thailand (Montano, Nawata, & Panichsakpatana, 2016). In a context where fertiliser use amongst coffee farmers was low, no significant difference in the proportion of UTZ certified and non-certified farmers using synthetic fertilisers was found (Giovannucci et al., 2013).

An interesting interaction between FSC certification and agriculture was noted by Kalonga & Kulinwda (2017), who observed that smallholders in Tanzania who had both certified forests and agricultural plots invested part of their increased incomes from forestry in buying fertilisers for their agriculture.

Figure 11:

The average number of incidences per certified entity of adopting, maintaining or having unresolved input use practices



Analysis of standards schemes' monitoring and compliance data on input use showed a relatively low rate of change, with only around 0.06 occurrences of practice adoption per certified entity (Figure 11, and see Appendix 2 for details of the analysis). This was one of the lowest rates of practice adoption for the thematic areas addressed, and would seem to be consistent with the mixed evidence within the literature of practice adoption on input use other than in organic certification.

3.4.3 Changes in input use over time

Few studies measure input-use by certified entities over time, and therefore the evidence of continuous improvement in practice is limited. That being said, Bennett et al., 2013 explicitly looked at changes in input use over time between UTZ certified and non-certified coca farms in Ghana. The UTZ certified farms used more fertiliser than non-certified farms (which in the context of farmers who had generally been unable to purchase fertiliser in the past and whose yields were low, can be considered a positive finding). Use rose through the three-year study period, and the difference between certified and non-certified farms became greater over time (a difference of 19% rising to 48%).

The analysis of practice adoption from standards schemes' monitoring and compliance data provides further insight into changes in practice over time. The rate of maintenance of practices on input use was 0.12 per certified (Figure 13) was 0.12 per certified entity. This is when either practice has remained stable, or when an entity has recorded a negative change in practice, followed by a return to a previous level (i.e., no overall

change). This implies that that, in common with other thematic areas, the major role of certification processes in embedding practice adoption on input use may be to reverse slippages in performance. In addition, input use has the highest rate of unresolved changes in practice over time of any of the thematic groups (0.07). This indicates that a decline in performance has occurred without any later return to good practice, either because the subsequent good practice occurred after the data was gathered or because the entity has not renewed their certification.

Analysis of non-compliances shows that surprisingly few are raised at the beginning of the certification process, with numbers gradually building (Figure 12). None, however were recorded in certified entities after the fifth year of certification. These data could indicate that new practices on input use do not quickly become embedded as a norm by certified entities, and/or insufficient practices are not readily detected or prioritised by auditors in the early years of certification.

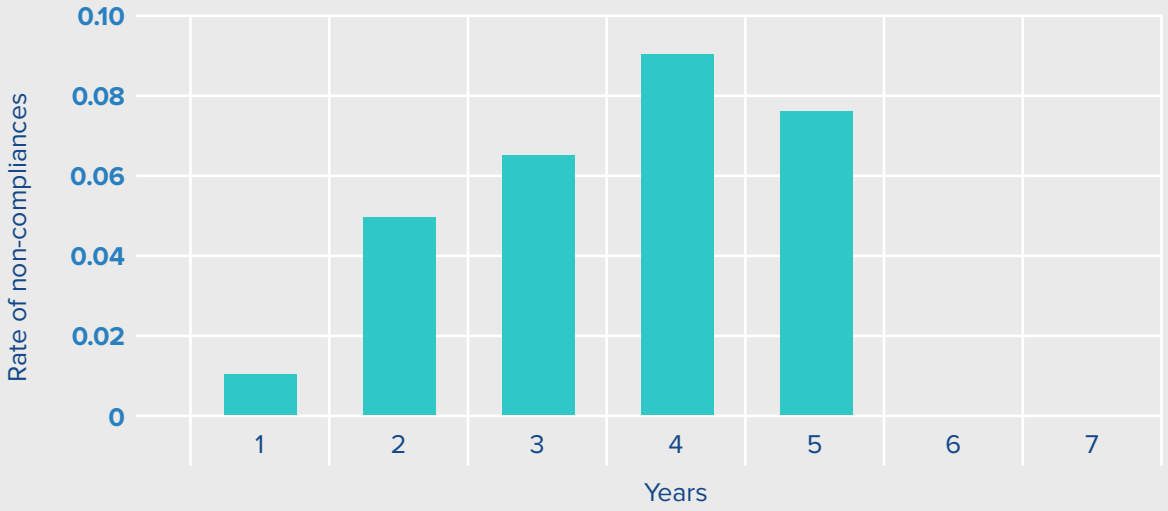
3.4.4 Reasons for practice adoption

It would be anticipated that a general motivation for changing practices on input use would be to increase profit, either through decreasing expenditure or increasing yields, or a combination of both. This motivation is not always linked to certification, and can be ambiguous. For example, 81% of Thai horticulture farmers surveyed perceived that adopting Global GAP would reduce the cost of chemical inputs, which could act as a motivation to gain certification and/or change practices on input use (Kersting & Wollni, 2012). However, the authors of that study caution that since Global GAP only permits use of chemicals which are registered for use for the target crop, and these tend to be more expensive than the inputs customarily used by Thai farmers, that the perceived gain may not materialise. This could presumably then result in abandoning the improved practices.

Suggested drivers of changes in input use practices that are more closely related to sustainability standards include information and technical assistance. Rueda, Thomas & Lambin (2015) suggest that certification promotes a more responsive institutional network of technical assistants and Colombian coffee farmers that circulate knowledge. This view is supported and supplemented by Ruben & Fort (2011) who attributed the

Figure 12:

Average number of substantive non-conformities relating to conservation per certified entity



increased use of organic fertilisers amongst non-organic Fairtrade coffee cooperatives in Peru to the enhanced technical assistance and access to agricultural finance provided by these cooperatives to their members. Similarly, UTZ was credited with having contributed significantly to more effective fertiliser-use and increased use of organic fertilisers amongst Indonesian cocoa farmers. This was attributed to the training and follow up provided to UTZ certified farmers but not available to non-certified farmers (Aidenvironment, 2016). Indeed, the absence of a difference in pesticide use between GAP certified and non-certified farmers in Thailand was attributed to the lack of government resources to provide technical support (Montano, Nawata, & Panichsakpatana, 2016).

3.4.5 Summary

There seems to be a strong suggestion that the technical support that certification systems can bring to farmers may be critical in supporting them to adopt more sustainable practices concerning input use. Whilst organic certification and reduced use of synthetic inputs are unsurprisingly usually closely linked, the evidence from other schemes that certification alone changes practices on input use appears to be more mixed.

3.5 Community benefits and development

3.5.1 Introduction

Within this thematic group we include a range of issues that relate to the contribution that a certified entity can make to the development of the local communities within or alongside which it operates. These include fundamental issues of respecting and upholding the rights of others, including the customary rights of indigenous peoples to use forest, land and fishing resources. It also includes keeping good relations with communities, including active consultation with them on issues that they are affected by. Several standards also include explicit requirements for a certified entity to contribute actively to local development, through fair local purchasing of products and services, by supporting community initiatives, or through a community investment fund. Some examples of this range of community benefits and development requirements within sustainability standards are given in Table 7.

3.5.2 Evidence for practice adoption

The systematic mapping identified fourteen papers with outcomes relating to community benefits and development. There is a strong focus on Latin America within this literature: eight studies reported results from one or more countries from Latin America and the Caribbean, a further four from Africa, one from Asia, and none from Europe or North America. Several papers studied more than one certification scheme, with the most common being Rainforest Alliance (ten studies), Fairtrade (eight), and FSC (four). Six of the studies focused on coffee and four on forestry. Ten of the identified studies reported that certified entities had changed practices with regard to community development, with a further four papers reporting no change or change for only some of the metrics used.

The identified literature is often concerned with the impacts of certification on community development rather than on the practices that lead to these impacts per se. The impacts reported include improved infrastructure, health (Jari et al., 2013, Cerutti et al., 2014; Miteva et al., 2015), or greater availability of labour for local women (Qiao et al., 2015). They also include examples where communities perceive few benefits, either because the 'rules' of certification are not well understood (Loconto & Sumbua, 2010), or because the producers had only recently become certified and so had a limited budget for community projects (Utting-Chamorro, 2005).

However, alongside evaluating impacts on community development, some of the identified studies also consider the practices that lie behind these impacts. A case in point is the research into Fairtrade premiums, which are ideally channelled to community development projects proposed by a village and approved by the producer organisation. The allocation process requires farmers and community members to adopt of a number of practices around organising, management and democratic management. Given that the administration of a community development fund is likely to be a new activity for most, and probably all, producers that become Fairtrade certified, it is very likely that certification drives these democratic practices. Adoption of democratic and organising practices is recorded for dual organic and Fairtrade certified tea farmers in China and

Table 7:

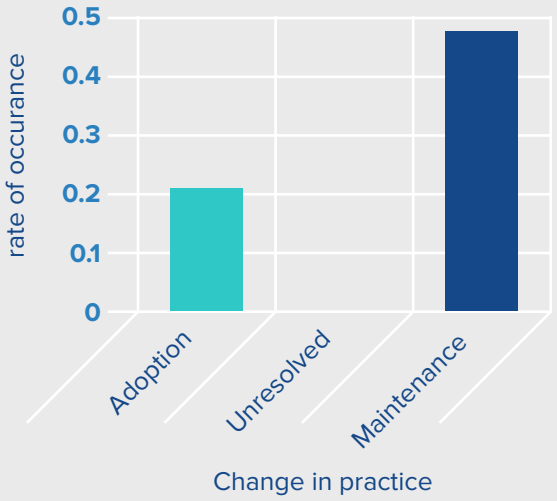
Illustrative examples of requirements concerning community benefits from sustainability standards

Type of community benefit	Certification Scheme	Practice required by the standard
Use rights	Rainforest Alliance	Criterion 4.19 Legitimate right to use the land is demonstrated by ownership, leasehold, or other legal documents or by documentation of traditional or community use rights. Right to use the land is not legitimately disputed by current or former local residents or communities, including in relation to past dispossession or forced abandonment. In the event of land conflict, legitimate right may be demonstrated if a conflict resolution process has been implemented, documented, and accepted by the affected parties
Use rights	MSC	PI 3.1.1 The management system exists within an appropriate and effective legal and/or customary framework which ensures that it: - Is capable of delivering sustainability in the UoA(s) - Observes the legal rights created explicitly or established by custom of people dependent on fishing for food or livelihood; - Incorporates an appropriate dispute resolution framework
Indigenous rights	FSC	Criterion 3.4 The Organization* shall recognize and uphold* the rights, customs and culture of Indigenous Peoples* as defined in the United Nations Declaration on the Rights of Indigenous Peoples (2007) and ILO Convention 169 (1989).
Community relations	Aquaculture Stewardship Council	Indicator 7.1.1 Evidence of regular and meaningful consultation and engagement with community representatives and organizations
Community relations	LEAF Marque	Control Point 9.1 There is regular communication and participation with local community initiatives to communicate a balanced and positive approach to farming.
Local purchasing	Roundtable on Sustainable Palm Oil	Growers and millers deal fairly and transparently with smallholders and other local businesses.
Community development	Roundtable on Sustainable Biomaterials	Criterion 5a. In regions of poverty, the socioeconomic status of local stakeholders impacted by the operations shall be improved.
Community development	Responsible Jewellery Council	Criterion 8.1 Members shall seek to support the development of the communities in which they operate through the support of community initiatives.
Community development	Fairtrade	Criterion 2.1.19 The Fairtrade Premium benefits workers, their families and their communities through Fairtrade Premium projects that address their needs and preferences as decided and adequately justified by workers.

Sri Lanka (Qiao et al., 2015), and coffee farmers in Peru (Ruben & Fort, 2011). The interpretation that requirements to manage funds inclusively are adopted due to certification is supported by evidence that Fairtrade and Rainforest Alliance coffee farmers in Colombia had a greater identification and satisfaction with cooperative organization compared to non-certified producers and producers certified to a private standard, with Rainforest Alliance certified farmers showing higher involvement of women in production and household decision-making than Fairtrade ones (Ruben & Zuniga, 2011). Perhaps inevitably. practices of democratic decision making on the use of the premium may not always be adopted, with the community in general, and women in particular, barely being included in decisions in one reported case from the Dominican Republic (Shreck, 2002).

Figure 13

The average number of incidences per certified entity of adopting, maintaining or having unresolved community development practices



Examples of producers adopting practice for community benefits from outside Fairtrade certification include a number of assessments of certified forests. Comparisons of certified with non-certified forest management units in Cameroon, Gabon and The Republic of Congo reported that FSC certified forest management companies invested more in community development (providing roads, electricity and schools) than their non-certified counterparts (Cerutti et al., 2014). FSC certified forests in Tanzania provided training for villagers on agricultural intensification to improve their livelihoods, which was not the case in villages not associated with non-certified forest management units (Kalonga & Kulindwa, 2017). The same study reported the improved social performance in terms of community relations, and respecting rights was not only greater in certified forest management units than non-certified ones, but the fact that these companies were not reputed for their social and legal performances in the past means that it is plausible that certification has brought about these changes. Markopoulis (1998) cites the case of FSC certification being used explicitly by an indigenous group in Bolivia as a strategy for securing their customary land rights. However, forest managers in Argentina and Chile reported that they did not have to change practice in order to respect the use rights of others (Cubbage et al., 2010).

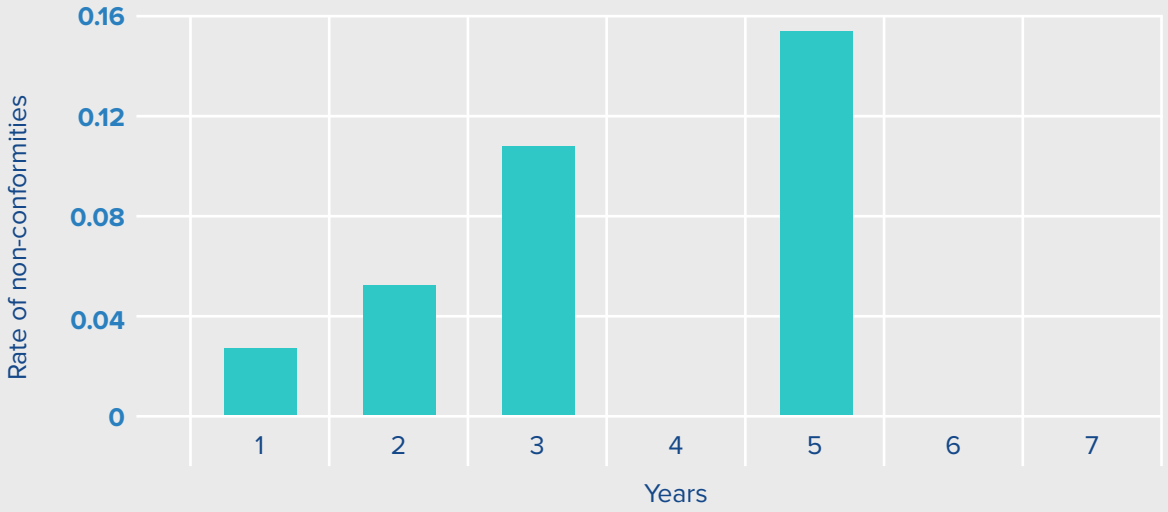
Analysis of standards schemes' monitoring and compliance data showed that community benefits and development had the highest rate of adoption of all of the thematic groups assessed. The rate of adoption was over 0.2 per certified entity (i.e., one incidence of substantive practice adoption for every five certified entities, Figure 13). This can perhaps be understood as a result of the types of practice required by standards to provide benefits to communities being outside the traditional norms of farming and forestry, and hence an area for improvement for many certified entities as they become certified.

3.5.3 Changes in community development practices over time

Few studies measure community development practices over time, and therefore the evidence of continuous improvement in practice is limited. One of the few that does, illustrated that the practice of disbursing Fairtrade premiums to community projects is likely to become cumulative over time (Ruben & Fort, 2011). The improved operations

Figure 14:

The average number of non-conformities on community benefits per certified entity over time



of Fairtrade cooperatives over time were considered likely to be in part funded by the Fairtrade Premium.

Analysis of standards schemes' monitoring and compliance data showed that the rate of maintenance of community benefit practices is amongst the highest of any thematic group, with an average of 0.5 episodes of practice maintenance per certified entity (Figure 13). This indicates that practice is either unchanging when established and/or slipping before returning to the previous level, and suggests that certification may be important in maintaining practice in this area. Analysis of the non-compliances raised on practices concerning community benefits increases from years one to three, thereafter there is little clear pattern in the number of non-conformities raised over time (Figure 14).

3.5.4 Reasons for practice adoption

In addition to the general motivations for certification described in Section 3.3.4, the literature suggests another mechanism for adoption of practices on community benefits and development. The process of auditors issuing non-conformities caused Bolivian forest managers to increase their community engagement and consultation in decisions (Markopoulos, 1998). As the author puts it "the conditions imposed on the project by Smart Wood [the Certification Body] appear to be having a positive, if gradual, effect on social relations."

3.5.5 Summary

There seems to be a strong suggestion that certification often results in improved democratic organisation and decision making, and/or greater engagement with local communities. This holds even for certification systems that do not have requirements on the disbursement of a premium. Certification is a plausible drive of these changes, not least because providing benefits to communities is often outside the conventional norms of farming and forestry. However, the limited number of studies outside Rainforest Alliance, Fairtrade and FSC certification, and from regions other than Latin America and the Caribbean, make it difficult to assess how widespread these effects are.

3.6 Occupational health and safety

3.6.1 Introduction

Occupational health and safety practices aim to reduce the risks of work-related accidents and ill health, and respond appropriately to problems that do occur. There are ethical reasons for ensuring good health and safety practice, as well as legal requirements on employers. In addition, poor health and safety costs businesses money through lost productivity. Health and safety is therefore a key element of ensuring appropriate working conditions.

For these reasons, sustainability standards almost invariably contain requirements on health and safety (Table 8). In some standards, these requirements cover a range of practices, from planning and risk analysis, to training the workforce on health and safety issues, through to implementation. Implementation requirements include displaying safety information, use of personal protective equipment (PPE), and the availability of first aid kits and medical response should an incident occur.

As well as being an important part of working conditions, health and safety has another characteristic that makes it interesting within an assessment of the effectiveness sustainability standards. It is one of the most visible issues that is typically covered by standards: it is obvious whether someone is wearing appropriate PPE or not, and whether first aid kits are available, in a way that gender discrimination, harassment or Freedom of Association are not. As one informant put it, with particular reference to Africa “I see uptake of health and safety practices but much less on other labour issues, like discrimination”. We might therefore expect certified entities to quickly adopt health and safety practices in order to gain certification and to avoid non-compliances. Thus, as well as being a critical area in its own right, understanding the adoption of health and safety practice can potentially reveal specific insights into the relation between certification systems and practice adoption.

3.6.2 Evidence for practice adoption

The systematic mapping identifies eight papers that compare health and safety outcomes in certified entities compared with non-certified entities. All of these publications enumerate practice adoption directly, and in some cases the impacts of those practices (e.g., the impact of certification on accident rates). The studies lean heavily towards forestry (five reports), the remainder covering crops such as tea, coffee, bananas and various fruit. There is a strong focus on Latin America and the Caribbean (five papers), with two reporting outcomes from Europe and two from Africa (note that two papers report from more than one region). All of the studies identified in the systematic map suggest that certified entities have better occupational health and safety practices than non-certified entities.

Managers of FSC certified forests interviewed in Argentina and Chile (Cubbage et al., 2010) and in Italy (Galati et al., 2017) stated that they had had taken actions to reduce health and safety risk. In the latter case this included conducting health and safety

Table 8

Illustrative examples of requirements concerning health and safety from sustainability standards

Type of H&S activity	Certification Scheme	Practice required by the standard
Planning	Rainforest Alliance	Criterion 4.14: The farm management and group administrator develop and implement an Occupational health and safety (OHS) plan [additional text omitted]
Training	Better Cotton Initiative	Criterion 6.1.1 Workers receive regular health and safety training appropriate to the work they perform
Training	Aquaculture Stewardship Council	ASC Salmon Standard. Version 1.1 (2017) Criterion 6.5.1: Percentage of workers trained in health and safety practices, procedures and policies on a yearly basis [performance requirement is 100%]
Implementation	Roundtable on Sustainable Biomaterials	Criterion 4f: Conditions of occupational safety and health for workers shall follow internationally recognised standards
Implementation	Fairtrade	Criterion 3.3.34: When you carry out hazardous work, you and the members of your organization must display all information, safety instructions, re-entry intervals and hygiene recommendations clearly and visibly in the workplace in the local language(s) and with pictogram
Implementation	Goodweave	Criterion 9.3.a [part of] Where necessary the producer should provide the workers with PPE, supply educational materials on its importance and require its usage (e.g., for those handling waste water, toxic dyes and/or chemicals, or operating dangerous machinery)
Implementation	Responsible Jewellery Council	Criterion 21.6: Members shall ensure that appropriate Personal Protective Equipment (PPE) is provided free of charge and verify that it is current, worn and used correctly.

training of staff and contractors. The perception of stakeholders to Estonian State Forest Management Centre was that the FSC certification has increased safety and health care of forest workers. The use of safety equipment by the Centre's contractors was one of the biggest changes in practice that these stakeholders identified as a result of certification (Hain & Ahas, 2007).

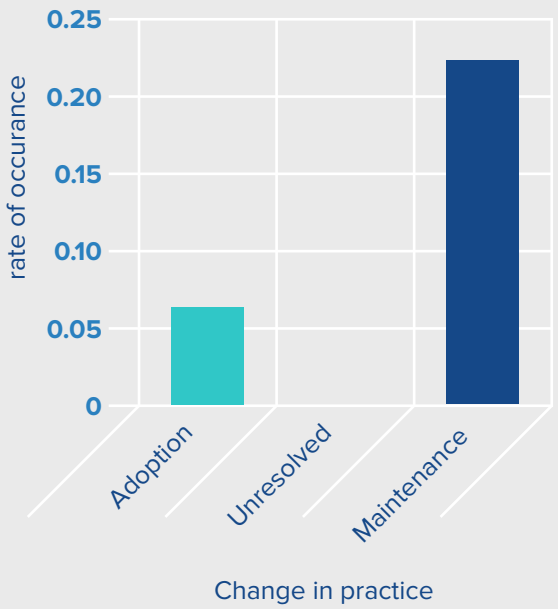
Studies that compare certified and non-certified entities or use before and after certification time series reach similar conclusions. FSC certified forest managements in Cameroon, Gabon and The Republic of Congo were more likely to have functional medical facilities; provide health and life insurance; have more and better-quality equipment to respond to emergencies (dedicated vehicles, on-site medicines, mobile or satellite phones); provide safety gear; provide well planned training; and have procedures to control and verify the use of safety equipment (Cerutti et al., 2014). In Chile, FSC certification of fast growing timber plantations resulted in changes to

health and safety management practices such as enforcing occupational health and safety regulations, penalising high accident rates or non-reported accidents, better supervisors' control of operations, training follow-up, certification of workers competencies, and improvements in contractor selection. These changes in practice resulted in significant trends of reduced accident rate, risk rate, and average time lost per accident. FCS certified plantations showed a greater decrease in these measures than plantations certified under ISO 14001 or OHSAS 18001 (Ackerknecht et al., 2005).

Other certification systems also appear to be associated with the adoption of improved health and safety practices. Workers in Fairtrade banana plantations in the Dominican Republic reportedly used more safety measures than workers in non-certified plantations, and those in Colombia also reported improvements in safety measures. These practices included safety measures for chemical use (e.g., use of protective equipment, awareness of the chemicals used), and health and safety training. There was no significant difference between workers in Fairtrade certified and non-certified plantations in Ghana, but this was likely to have been the result of a very small sample size (Rijn et al., 2016). A significantly greater proportion of employees in UTZ certified coffee farms in Colombia received training in job safety in the first two years of certification, and the proportion of farms with available first aid kits and that provided protective gear was also significantly higher than in non-certified farms (García et al., 2014).

Figure 15:

The average number of incidences per certified entity of adopting, maintaining or having unresolved health and safety practices



Analysis of standards schemes' monitoring and compliance data illustrates further the role of standards and certification in driving the adoption of occupational health and safety practices. The analysis presented in Figure 15 is based on the metric of practice adoption which coded changes as either practice adoption (where a positive change was indicated); maintenance of practice (if there was either no evidence of change or where change was followed by reversion to previous practice), or 'unresolved' if a decrease in practice had occurred at the end of the time period. It is limited to substantive changes in practice 'on the ground' and purely procedural changes (e.g., changes in documentation) are excluded. Appendix 2 gives full details of the methods. Figure 15 shows that the estimated rate of adoption of substantive health and safety practices is 0.08 practice adopted per certified entity. This is one of the lowest rates of practice adoption of any thematic group. This might at first appear counter to the evidence of published studies, but due to the visibility of health and safety issues, it is possible that changes in practice are frequently made prior to certification.

3.6.3 Changes in health and safety practices over time

Four studies explicitly look at changes in health and safety practices in certified entities over time. Ackerknecht et al. (2005) chart accident statistics over time for four Chilean forest companies, two of which gained FSC certification during the study period.

The occurrence of health and safety incidents amongst the contractors of these two companies is all-but eliminated after certification, whilst incidences within the companies themselves generally decline but with significant noise in the data.

A study on UTZ certified coffee farms in Colombia noted a drop off in the provision of health and safety training after certification. Training in job safety and agrochemical handling was carried out by 23.8% of the farms during the second year and had decreased to 2.7% by the fourth (Garcia et al., 2014). Workers in Fairtrade banana plantations in the Ghana reported that all safety measures for handling chemicals were better in a plantation that had been certified for many years than in one that had only recently gained certification (Rijn et al., 2016).

One of the few studies to consider changes that occur between entities deciding to attempt certification and becoming certified reported that awareness of health and safety practices compliant with the EUREPGAP standard changed from 50% to 80% as a result of pre-certification training (Lourenzani et al., 2006).

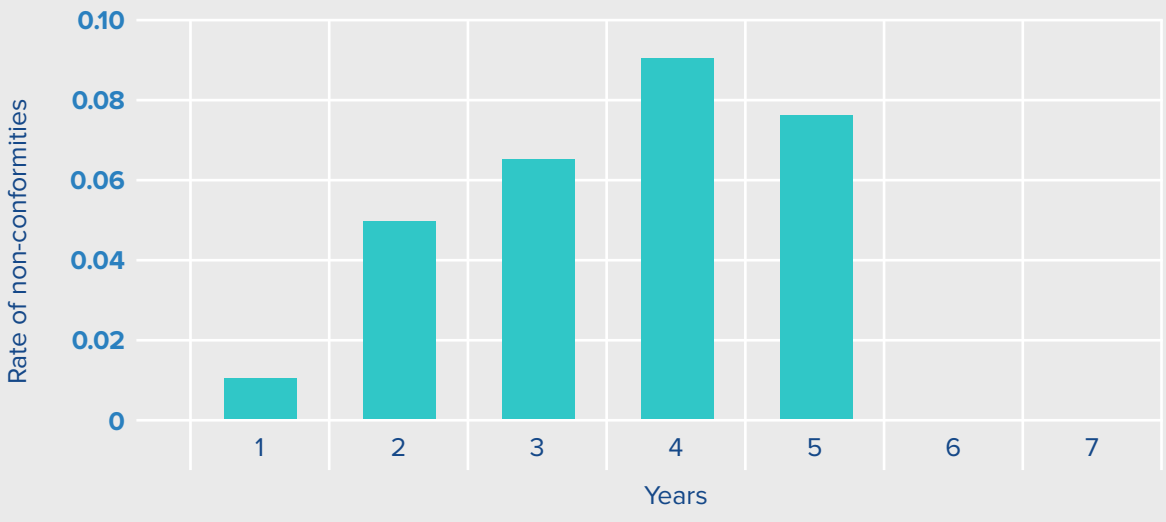
Analysis of standards schemes' monitoring and compliance data provides additional insight into the potential role of standards in driving the adoption of health and safety practices over time. Approximately 0.2 cases of practice maintenance per certified entity were recorded, indicating that certification may be helping to maintain good practice (Figure 15). Analysis of scheme's non-conformity data (see Appendix 2 for details) shows an overall low rate of non-compliances, with a peak in occurrence in Years 3-4 after certification (Figure 16). However, even at the peak, there are fewer than 0.1 non-compliances per certified entity per year. Note that this data does not include any from forest certification, where the majority of the literature on the role of certification in health and safety practices is focused. Given that the risk of serious accidents is perhaps more likely in forestry than in many other sectors, the pattern of practice adoption and non-conformities may well be different in that sector.

3.6.4 Reasons for practice adoption

A number of studies highlight training as an important part of raising awareness and knowledge of health and safety, which can then lead to adoption of improved practices

Figure 16

The average number of non-conformities on health and safety per certified entity over time



(e.g., (Rijn et al., 2016; Lourenzani et al., 2006).

In addition, there is also a suggestion from a survey of managers of FSC certified forests in Argentina and Chile that the auditing process and non-conformities play an important role: more than half of the non-conformities raised by auditors focused on workers' training and safety (Cubbage et al., 2010). This may serve to counterbalance the perception in some countries that health and safety requirements in forest certification can be excessive, which is a barrier that has to be overcome to adopt better practice (Hain & Ahas, 2007).

3.6.5 Summary

The literature identified by the systematic map consistently reports that certification results in the adoption of improved health and safety practices. Whilst the majority of this evidence is from the forestry sector, and FSC certification in particular, there is at least some evidence that the same holds true in agricultural contexts. Sectors such as fisheries remain under-researched.

Expert opinion suggests that part of the reason for this finding is to do with the detectability of health and safety practice to auditors: non-conformities in this area are easier for auditors to detect than many other thematic areas. Therefore, any entity wishing to gain a certificate is likely to have to be in compliance with health and safety requirements.

3.7 Management systems

3.7.1 Introduction

Many people who have been involved with certification over a significant period assert that one of the most consistent changes they see is improvement in – or perhaps more precisely, formalisation of – the management systems of certified entities. Commenting on a plantation sector in a particular country, one informant put it "Before certification, their management was in the stone age. And behind closed doors, they admit it".

Having good management systems are, of course, not sufficient to guarantee sustainability: many companies the world over, have strong management systems yet poor environmental and social performance. However, practitioners observe that entities

Table 9

Illustrative examples of requirements within sustainability standards concerning management systems

System element	Certification Scheme	Practice required by the standard
Management document	Roundtable on Sustainable Palm Oil	Criterion 1.1.2 Records of requests for information and responses shall be maintained.
Management document	Rainforest Alliance	Criterion 1.11 The farm management and group administrator document all training according to training topic, name, organization and title of the trainer, focal crop, number of women and men trained, and an attestation of each worker or group member that s/he participated in the training
Implemented system	Responsible Jewellery Council	9.1 a Members shall establish policy/ies that prohibit bribery in all business practices and transactions carried out by the Member and by agents acting on behalf of the Member.
Implemented system	Roundtable for Sustainable Biomaterials	Principle 2: Planning, Monitoring and Continuous Improvement - Sustainable operations are planned, implemented, and continuously improved through an open, transparent, and consultative impact assessment and management process and an economic viability analysis
Implemented system	Marine Stewardship Council	PI 3.1.2 The management system has effective consultation processes that are open to interested and affected parties [additional text omitted for brevity]
Management outcome	Bonsucro	1.1.1 National laws complied with

with weak management systems tend to be the ones that struggle to gain and maintain sustainability certification. If an entity has a process of deciding what should be achieved, defining the activities to be implemented in order to achieve them, measuring progress and revising targets it can apply to all facets of sustainability, driving both performance and ongoing improvements.

A number of sustainability standards have specific requirements about management practices required of certified entities. These include the purely procedural requirement to have a particular type of document, such as a management plan or a particular set of records. Often though, the requirement is for documents that contain specific elements, or that are implemented and measured. A final approach is to simply specify a management outcome, rather than the content of documents (see Table 9).

Management systems are often codified in documents, and verification of conformity with a standard often relies heavily on documentation in this practice area (Jennings, 2017). For this reason, the general focus of this section – as with all other thematic areas – is on the practical implementation of management systems as it tangibly affects sustainability, rather than record keeping or the existence of documents *per se*.

3.7.2 Evidence for practice adoption

The literature identified in the systematic map includes eight reports that evaluate management systems in certified entities, although this is rarely a primary focus of the research. This perhaps reflects a bias of researchers towards environmental and social outcomes, rather than organisational processes. Nonetheless, there are indications that the need to develop functional cooperatives in order to participate in Fairtrade and Rainforest Alliance certification has increased management capacity amongst smallholder farmers, compared to non-certified farmers (e.g., Qiao et al., 2015; Ruben & Fort, 2011; Ruben & Zuniga, 2011). A similar finding is reported from coffee growers in Nicaragua, although the attribution to a specific certification scheme was complicated by multiple certificates that most farmers possessed (Utting-Chamorro, 2005).

By contrast, in the highly developed context of the USA, half of the surveyed farmers that enrolled into the Michigan Agriculture Environmental Assurance Program stated that they needed to make only minor changes to their management systems in order to achieve certification (Stuart, Benveniste & Harris, 2014).

One of the more comprehensive analyses of management practices under certification studied MSC certified fisheries in Latin America and the Caribbean. The fisheries gained higher scores on MSC Principle 3 (effective management), than they achieved on Principle 1 (sustainable fish stocks) or Principle 2 (minimising environmental impact; Pérez-Ramírez et al., 2016).

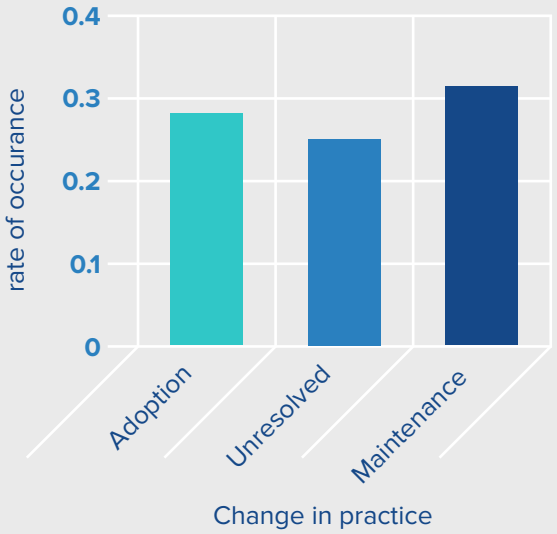
Three studies of FSC certification also provide evidence of adoption of improved management system practices. FSC certified forests in Tanzania had significantly improved implementation of forest bylaws than non-certified forests, which was interpreted as being an indicator of good organisational governance (Kalonga & Kulindwa, 2017). Similarly, FSC certified forest management units in Cameroon, Gabon and The Republic of Congo adopted management practices that resulted in more consistent legal compliance, staff training and other outcomes, which were rarely present in non-certified forests (Cerutti et al, 2014). The majority of surveyed forest managers in Argentina and Chile reported that they had adopted some new elements of management systems (e.g., legal planning), but not others (Cubbage et al., 2010).

Analysis of standards schemes' monitoring and compliance data illustrates further the

role of standards and certification in driving the adoption of improved management systems. Figure 17 shows that the rate of adoption of management system practices at nearly 0.3 per certified entity. This is one of the highest rates of any thematic areas assessed. This analysis includes only substantive practices, with the procedural requirements for management systems that predominate within several standards not represented. This suggests that the adoption of improved management practices is relatively common within certified entities.

Figure 17:

The average incidence of adopting, maintaining or having unresolved management systems practices per certified entity



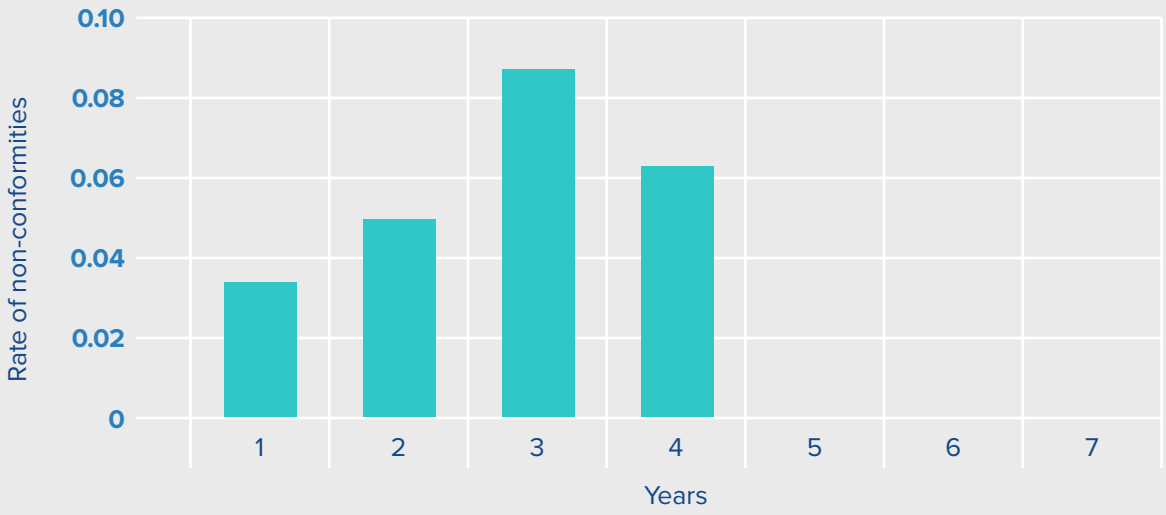
3.7.3 Changes in practice over time

Few of the studies identified by the systematic mapping evaluate continual improvement of management systems over time. However, Pérez-Ramírez et al. (2016) noted that although many of the MSC certified fisheries in Latin America and the Caribbean were regarded as having achieved certification in part because of existing management practices and governance, but in some cases (e.g., the seabob shrimp fishery in Suriname) new management arrangements and practices emerged after certification.

Analysis of standards schemes' non-conformity data illustrates the temporal pattern of non-conformities relating to substantive management systems practices (Figure 18). Non-conformities increase over the first three years of certification, but essentially absent from the fifth year onwards. This implies that certification may play a role in embedding management practices, further supported by the high rate of maintenance of practice (over 0.3 per certified entity) show in Figure 17.

Figure 18

The average number of non-conformities on management systems per certified entity over time



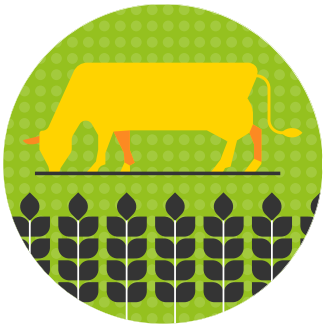
3.7.4 Reasons for practice adoption

In addition to general motivations for certification, such as market access and price premiums, there is also a suggestion from a study of FSC certified forests in the Congo Basin that the requirement within the FSC Principles and Criteria to comply with applicable laws and regulations motivated forest managers to improve their management systems to control illegal activities (Cerutti et al, 2014). This motivation was presumably heightened by the fact that compliance with FSC Principles and Criteria is audited.

Small-scale vegetable farmers in Thailand who participated in Global GAP had the Quality Management System run either by the donor, by the exporter, or by farmers themselves. Support by donors, exporters and public–private partnerships was found to be vital to enable small-scale farmers to adopt the standard (Kersting & Wollni, 2013).

3.7.5 Summary

The adoption of improved management systems by entities seeking certification is seen as a critical step towards sustainability by many people who work with sustainability standards. This is an under-researched field. There is, however, some indication that certification may drive the adoption of improved and/or more formal management systems, and a suggestion that these may become embedded over a few years of certification. The extent to which these continue to adapt and improve over time remains poorly understood.



3.8 Good Production Practices

3.8.1 Introduction

Good production practices refer to the suite of practices that aim to ensure a high-quality production for the long term and by using resources efficiently. Each sector has its own specific practices and terminology: good agricultural practices, sustainable forest management, and sustainable harvest (or catch) in fisheries. Within each of these sectors there are numerous competing definitions of what good production practice is, but even though there is significant variation in what constitutes good production practices, the overarching aim of all of them is that production is sustained for the long term.

Table 11:

Illustrative examples of requirements concerning good production practices from sustainability standards

Scope	Certification Scheme	Practice required by the standard
Specific practices	Aquaculture Stewardship Council	Indicator: 5.2.3 Percentage of medication events that are prescribed by a veterinarian Requirement: 100%
Specific practices	UTZ	Control Point G.B.39: New plantings follow a suitable crop pattern to ensure a well-established cropping system
Specific practices	Roundtable on Sustainable Palm Oil	Criterion 4.5 Pests, diseases, weeds and invasive introduced species are effectively managed using appropriate Integrated Pest Management techniques
Specific practices	Fairtrade	Criterion 3.2.32 You and your members must not intentionally use genetically engineered seed or planting stock for Fairtrade crop(s). You must implement practices to avoid GM contamination in seed stocks.
Efficiency	Better Cotton Initiative	Criterion 5.2 Seed cotton is harvested, managed and stored to minimise trash, contamination and damage
Long term harvest	Forest Stewardship Council	Criterion 5.2 The Organization shall normally harvest products and services from the Management Unit at or below a level which can be permanently sustained.
Long term harvest	Marine Stewardship Council	PI 1.1.2 Where the stock is reduced, there is evidence of stock rebuilding within a specified timeframe.

Sustainability standards vary in their approaches to good production practices, some requiring specific practices such as Integrated Pest Management, or prohibiting others such as the use of Genetically Modified Organisms (Table 11). Efficient use of resources is included in some standards (noting that specific requirements on efficient fertiliser, pesticide and water use is treated in Section 3.5, above), whilst some standards specify the long-term outcome of sustained harvest.

The good production practices here concern agricultural, forest and fisheries sectors, but even within these sectors we have kept some closely related issues including soil management, water management and input use separate as there is sufficient information on these aspects to justify differentiation.

3.8.2 Evidence of practice adoption

The systematic map identified twenty-seven studies that reported outcomes relating to good production practices. Several of the studies were conducted in multiple countries, with fourteen including data from Latin America and the Caribbean, six from Asia, four from Europe, and three from North America. Only two studies reported outcomes from African countries. Nine of the papers focused on the forestry sector, seven on coffee and four on fish. Other sectors and commodities, such as horticulture and tea, are also represented. Several papers included multiple certification schemes, with the most common ones studied including the FSC (nine studies), Rainforest Alliance (six), organic certification (four), Fairtrade (four), UTZ (three), and Marine Stewardship Council (three).

The majority of these papers (sixteen) provide some evidence that good production practices are adopted as a result of certification, with a further two showing a mix of positive change and no change. Nine showed no change between certified and non-certified entities, or before and after certification. Examples of changed production practice are found across a range sectors, and there the circumstances where change appears to be limited appear to be dependent on the context more than the sector. Contexts where little change is reported include those where the entities already have high levels of technical production practices and so are in effect already meeting the requirements of the standard. At the opposite end of the spectrum, there examples of entities with limited technical and financial resources may be unable to change production practices.

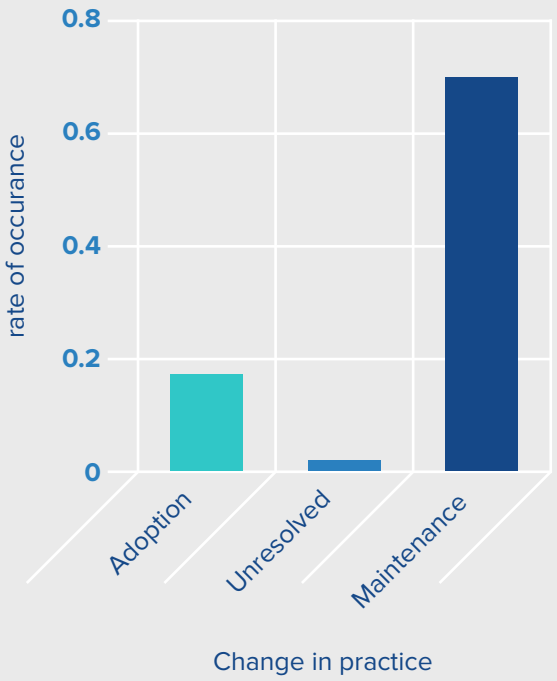
Amongst the research which showed evidence that certification had driven the adoption of some good production practices are a number of studies on coffee. A significantly greater number of Rainforest Alliance certified farmers in Colombia adopted integrated management strategies to control the two most prevalent phytosanitary problems than did non-certified farmers. All but one of the certified farmers had partially or fully adopted the rust-resistant varieties of coffee, whereas susceptible varieties were still common on non-certified farms. However, the use of biological controls was not significantly different between certified and non-certified coffee, neither was the use of the main insecticide (Rueda & Lambin, 2013). Similar results were found for UTZ certified coffee farmers in the same country (García et al., 2014). Good Agricultural Practices were more commonly applied by Nicaraguan coffee producers certified under the Rainforest Alliance and Café Practices than those certified as Fairtrade or non-certified producers (Ruben & Zuniga, 2011). Costa Rican organic-certified coffee growers had significantly better management practices than non-certified farmers (Blackman & Naranjo, 2012).

The forestry sector is also well represented by studies that compare the forest management practices of certified and non-certified entities. Management of FSC certified forest management units in the Congo Basin was characterised by Nasi, Billand

& Vanvliet (2012) as 'optimization for a single good (timber) but with consideration given to the preservation of other goods and services for long term sustainability' in contrast to non-certified forest management units that 'can safely be considered as having a single use management system, with timber as the sole commodity for an immediate profit, and without long term sustainability concerns'.

Figure 19:

The average incidence of adopting, maintaining or having unresolved good production practices per certified entity



The degree to which an entity has to change production practices in order to meet the requirements of a standard will depend on a number of factors. Organisations in countries that have a longer history of technical forest management might be generally expected to require less profound changes in order to achieve and retain a certificate. Thus, in contrast to the countries of the Congo Basin, forest managers in Argentina and Chile reported that some changes had been required to their production practices in order to achieve certification (e.g., management of invasive species, improved management plans), whilst practices were already in place (e.g. sustainable yield constraints, forest health protection measures; Cubbage et al., 2010). Similarly, Johansson & Lidestav (2011) found little difference in environmental management between certified and non-certified forest management units in Sweden, a country with a long tradition of forest technical management, although there was more harvesting on PEFC certified forests than on non-certified ones. Camino & Alfaro (1998) concluded that "Those Forest Management Units (FMUs) in Latin America (and probably in the rest of the world), which have achieved certification, undoubtedly have above-average management practices both in natural forests and plantations."

Research from other sectors has also found no significant difference in the adoption of good production practices between certified and non-certified entities when practices were already at a high level. Examples include coffee farmers in India certified by UTZ (Marie-Vivien et al., 2014) and Global GAP certified broiler chickens in Brazil, where all studied farms adhered to minimum welfare standards, regardless of certification (Souza et al., 2015). Twenty-eight per cent of surveyed maize farmers reported not having had to make any changes to receive certification, as they had already been implementing all of the practices required to gain Michigan Agriculture Environmental Assurance Program certification, with 50% reporting having had to make only made minor changes (Stuart, Benveniste & Harris, 2014). In the marine fisheries sector, Selden et al. (2016) found no difference in the rates of discard of non-target finfish, bycatch rates of marine mammals and observer coverage in MSC certified and non-certified fisheries in the U.S.. The MSC-certified fisheries did have lower average gear destructiveness scores than non-certified stocks when weighted by landings, but not when weighted by the number of fisheries.

The opposite effect – where producers do not possess the skills, experience or capital to change production practices – has also been reported. In the forestry sector, Bass et al. (2001) argued that the changes required often presents a disproportionate challenge

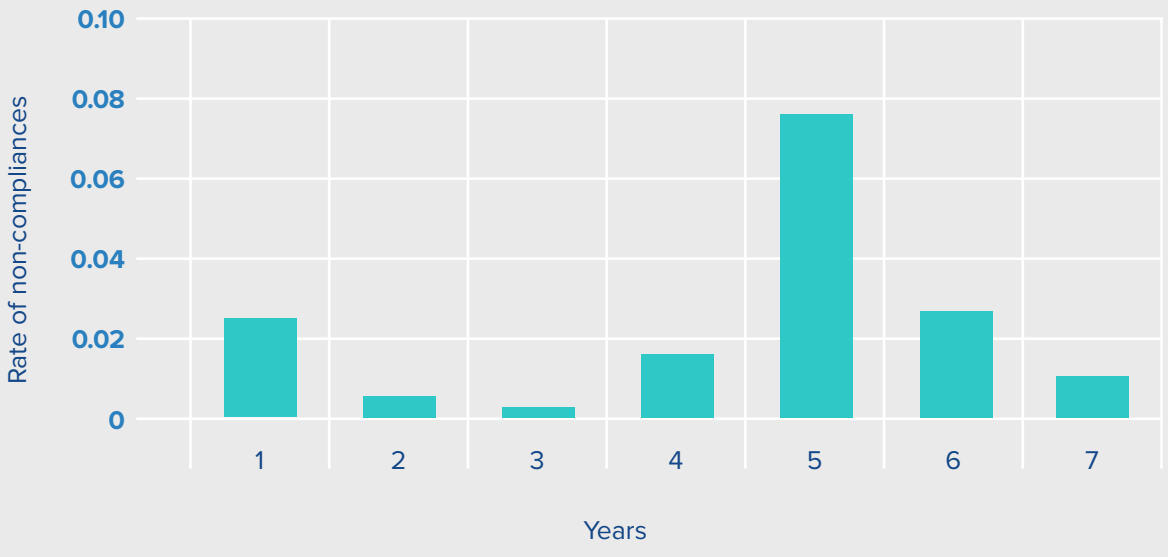
to small-scale and community forest enterprises. Amongst surveyed forest managers in Italy, over half of whom were small enterprises, the ability to bear the cost of training staff in new forest management techniques – or pay consultants to do so – is seen as a challenge to adopting management practices compliant with FSC Principles and Criteria (Galati et al., 2017). Further examples include Vietnamese smallholder farmers who had gained FSC certification for their planted Acacia plots, but who, as farmers, lacked knowledge and experience of forestry (Hoang, Hoshino & Hashimoto, 2015), and a forest management unit in Bolivia that did not change their silvicultural regime even though this had been the subject of non-conformities (Markopoulis, 1998).

The requirements that different certification schemes place on good production practices also influences the degree of change that organisations are required to make to their production practices in order to become certified. Email surveys of organizations that had received forest management certification under the FSC in the United States and Sustainable Forestry Initiative (SFI) in the United States and Canada showed that firms implemented on average 13-14 changes in practice. Of these, FSC certified companies were required to make more changes to forest management and environmental practices than SFI certified companies (Moore, Cubbage & Eicheldinger, 2012).

Analysis of standards schemes' monitoring and compliance data provides additional insight on the role of standards and certification in driving the adoption of improved production practices (see Appendix 2 for details of the analysis). Figure 19 shows that a rate of adoption of good production practices of 0.17 per certified entity. This is one of the highest rates of any thematic areas assessed. This is consistent with the indications in the literature that certification can, in some circumstances at least, drive the adoption of good production practices. The rate of maintenance of good practice was also the highest amongst all of the thematic areas assessed, at 0.7 per certified entity. This is when either practice has remained stable, or when an entity has recorded a negative change in practice, followed by a return to a previous level (i.e., no overall change). This suggests that certification may play a key role in maintaining good production practices over time, and preventing slippage into inferior practices.

Figure 20:

The average number of non-conformities on good production practices per certified entity over time



3.8.3 Changes in good production practices over time

Little of the identified literature presents a time series on good production practices, and it is therefore difficult to make any firm conclusions on continual improvement on this theme. Garcia et al. (undated) noted that the values of an aggregated index of Colombian coffee farm sustainability increases over time for certified farms, but the difference between certified and control groups tended to decrease. This provides a tantalising suggestion that certified entities are adopting sustainable production practices, and these practices gradually become the norm.

However, there is at least one example in the literature where no change in production practices was found over time. Research into marine fisheries in the Northeast Atlantic concluded that the majority of fisheries were overexploiting the fish stock, that this had been occurring before and continued after certification (Optiz et al., 2016).

Analysis of standards schemes' monitoring and compliance data illustrated the potentially important role of standards systems in maintain good production practises over time (Figure 19). Further analysis of non-conformities showed a distinct pattern is of an increase in frequency 4-6 years after certification (Figure 20). This suggests that either performance in good production practice dips over time, or perhaps that auditors pay less attention to these elements of the standards initially.

3.8.4 Reasons for change

The general motivations behind certification, such as access to markets and maintaining a competitive advantage (Utting-Chamorro, 2005), premium pricing (Ruben & Zuniga, 2015), have also been cited as reasons for certified entities to adopt improved production practices.

In addition to these, a number of other mechanisms have been proposed. These include external support, increased awareness of and training on sustainability issues, and certification processes. For example, support from donors, exporters and public private partnerships was considered vital for Thai vegetable farmers to achieve Global GAP certification (Kersting & Wollni, 2017). Similarly, Nepalese tea farmers required significant support from factories, local NGO experts, and overseas organic certifiers to successfully convert to organic certification (Mohan, 2016). Even though the performance of MSC certified fisheries in Latin America and the Caribbean was considered to be high, the technical and financial support from governmental agencies was considered necessary to reach the standard of resource management required for certification (Pérez-Ramírez et al., 2016). An important lesson in support is provided by the example of Vietnamese shrimp farmers, who were certified to an organic standard but who apparently showed little difference in production practices to non-certified farmers. This was ascribed to their certification being part of a project for which certification, rather than sustainability, seemed to have become the objective (Baumgartner & Nguyen, 2017).

As good production practices require technical knowledge, it might be expected that increased awareness of the issues and training on practices would support practice adoption. The emergence of certification was credited with raising awareness of sustainability issues amongst forest managers in Norway (Sverdrup-Thygeson, Borg & Bergsaker, 2008). The greater uptake of Good Agricultural Practices amongst UTZ certified coffee farmers in Colombia was at least in part believed to have been the result of the additional training they received compared to non-certified farmers (García et al., 2014). Similarly, training at the pre-certification stage was found important in changing production practices for fruit farmers (Lourenzani et al., 2006) and agroforestry

smallholders (Pinto et al., 2013) in Brazil.

Finally, the process of raising non-conformities and the focus of certification audits on performance in the field was thought to have contributed to Norwegian forest managers adopting improved management practices (Sverdrup-Thygeson, Borg & Bergsaker, 2008).

3.8.5 Summary

There are indications from the literature of improved production practices in certified entities when compared with non-certified entities, and before and after certification. Good production practices are a technical area, and where no change in practice is seen, it often seems to be either because certification occurred in a context where existing practice was already meeting the requirements of the standard, or conversely, because normative practice was so poor and resources to change it too limited, resulting in the barrier to certification being too high. As also might be expected in a technical area, practice adoption often appears to be aided by external technical support, including training.

3.9 Summary of thematic findings

This section of the report has summarized the evidence for standards and certification systems driving practice adoption on selected themes. It used evidence derived from a narrative review of literature identified by systematic mapping, from analysis of standards systems' monitoring and compliance data, and through key informant interviews.

Practice adoption is reported for each of the thematic areas assessed. The evidence for practice adoption, though, may be more robust for some sustainability themes than for others. For example, all eight of the papers enumerating changes in health and safety practices showed a positive impact of certification, as did all of the fourteen reporting community benefits and development. This does not 'prove' that practice adoption is a universal effect of certification in these themes, but there are logics from other areas of research that suggest why practice adoption may differ between themes and contexts.

There seem to be three instances where practice adoption by certified entities may be less frequent. The first of these is when the requirements of the standard are already being met by existing practice. This may explain why several studies indicate no difference, or a difference in some but not all parameters, between certified and non-certified entities from developed countries.

The second circumstance is when existing practice is far short of that required by a standard, and the technical and financial resources required for change are limited. This underlines the importance of technical and financial support as a complement to certification, particularly for many small-scale producers in developing countries.

Finally, practice adoption may happen less frequently with practices that are hard to detect. Some practices may be less detectible than others, either because they are only implemented occasionally (e.g., application of pesticides) and/or because they are difficult to investigate (e.g., discriminatory hiring, harassment). This may help to explain why the literature on some sustainability themes, such as use of agricultural inputs, show a mixture of positive effects and no change. In general, though, as certification is by no means the only factor affecting practice adoption, we should not expect unequivocal findings.

Most of the studies identified by the systematic mapping use comparisons of certified

entities with non-certified ones, sometimes using rigorous propensity matching to ensure that the two samples are alike in most characteristics other than the certificate they hold. As long as the certified and non-certified groups are indeed identical, then it can perhaps be assumed that the differences in practice are due to certification. Nonetheless, there is a dearth of longitudinal research that would explicitly show change in practices and outcomes in the pre-certification stage, and continuous improvement thereafter. In particular, there is almost no formal evidence of changes made by entities between the time that they decide to pursue certification and obtaining a certificate. Expert informants consistently stated their experience that this was when the largest changes in practice occurred.

There are a small number of papers that provide data with counterfactuals on ongoing changes to practice after certification. Overall, they give mixed results, with some ongoing improvements to practice shown, but probably the most common outcome is maintenance of practice. This is consistent with maintenance metric developed for this research, which showed that maintenance of practice – either through no recorded change or due to a slippage in performance that was subsequently rectified – was more common than either practice adoption or unresolved slippages in practice. This suggests that many certified entities are at least maintaining practice over time, and also that the audit process actively contributes to this process of practice maintenance.

The same general factors that tend to motivate entities to become certified, including price premiums, market access, market share, and managing reputational risk are also cited as motivating entities to adopt improved practice. A small but not irrelevant proportion of entities are thought to improve practices because it aligns with their existing values. In addition to these motivations, external financial and/or technical support seems to be a common enabler of practice adoption in many of the certified entities studied in the literature.

4 Conclusions and recommendations

4.1 The evidence base

This study has applied a systematic mapping method as a primary tool to assess the quality, quantity and type of evidence available to understand the effectiveness of standards in driving practice adoption amongst certified entities. From a body of over thirteen thousand reports from the peer-reviewed and 'grey' literature that contained at least some information relevant to the research questions, only 371 (2.9%) appeared to be the subject of an appropriate population, intervention, have a counterfactual, and be reporting on an appropriate outcome. Of these, just 116 (less than 1% of the original sample) proved to be appropriate on closer examination. This study does not claim to have assessed all possible research, but nonetheless has examined a far larger base of evidence than is common in conventional literature reviews.

The predominant differentiator for research which was carried through all stages of filtering and those that were excluded was the presence of a counterfactual. The majority of studies of the impacts and outcomes of certification simply do not have a control population, nor measurement of before and after certification. Without a counterfactual, it is impossible to evaluate whether standards systems have had an effect on practice. This is the first major finding of the research: a huge amount of effort and expense is put into evaluations that are formally incapable of answering the question they pose themselves.

The second major finding derives from the 116 studies that were comprehensively coded as appropriate to answer the research question. Amongst this evidence base, there are some notable concentrations of effort, and some notable gaps:

- There is a concentration of research into certain sectors, particularly agriculture and forestry, and within these sectors, coffee and timber production. Important commodities from a sustainability perspective, including cotton, aquaculture, sugar, tea, bananas and mining are barely represented or not at all.
- There is a concentration of research into four certification schemes, Rainforest Alliance, Fairtrade, FSC and organic. These are amongst the longest-established and largest certification schemes, and so it is perhaps not surprising that they should have received a greater attention from research. How representative the findings from these schemes are for certification as a whole is not well understood.
- There is a concentration on producers, and specifically on smallholder producers and producer cooperatives. This focus means that a large proportion of the supply chain of any commodity is under-represented.
- There is a focus on research in developing countries in the tropics and sub-tropics, with comparatively little on developed and temperate countries. Some of the potential inter-relations between context and certification have a limited base of robust literature.
- Positive results bias – the reporting of positive impacts being more common than reporting no or negative impacts – is a well-documented problem in the peer-reviewed journal literature (Leimu & Koricheva 2004, Lortie et al. 2007). Research that reports new, positive results are considered by publishers to be inherently more interesting than research that finds nothing significant and nothing new. Although non-journal articles were included in the present research, they were undoubtedly less comprehensively covered by our search strategy than the studies found more readily through the three large bibliographic databases of academic journals.

It should be noted that not all of the studies identified by the systematic mapping report practice. Many report impacts, and these are beyond the scope of this study. For example, increases in household income, educational attainment within producer families are impacts that will almost always be the result of multiple changes in practice, but often it is only the impact that is reported, not the practices that lie behind them.

A final important finding of the systematic mapping is that even amongst the studies that do have a counterfactual, this is almost invariably a control population of non-certified producers. There are very few studies that follow producers through time, and present a 'before and after' counterfactual. This means that there is a relatively limited formal evidence base on the similarities and differences that exist between entities that seek certification and those that don't; changes that occur before certification; and continual improvement after it. The importance of this is discussed below.

4.2 Do sustainability standards drive adoption of better practices?

The systematic mapping evaluates the strength of evidence available to answer the research questions posed by this study. Critically, it also provides a filtered and manageable set of documents from a large and chaotic literature of vastly varying quality, and this was used as the basis for understanding what the literature concludes about the research questions. We reviewed in detail the literature identified by the systematic mapping that reported outcomes for six sustainability themes. Given some of the structural gaps highlighted above, we added additional evidence from analysis of data provided by some of ISEAL's members, and from key informant interviews.

There is evidence in each of the six thematic areas that certification and standards contribute to the adoption of improved practices by certified entities. This is typically expressed as a difference in practices between certified and non-certified entities. The studies that report greater adoption of sustainable practices amongst certified entities come from a wide range of geographies, sectors and standards systems.

As might be expected, there is also evidence that certification does not always result in the adoption of better practices. There seem to be at least three situations where practice adoption becomes less common. Firstly, practice adoption seems to be reported less consistently when existing levels of practice are high (or the requirements of the standard modest). In this regard, it is interesting to note that practice adoption was consistently reported for practices that are aimed at supporting community development, perhaps because these practices are rarely the norm within non-certified entities. At the other end of the spectrum, entities with low levels performance may simply not be able to make the changes in practice required by the standard.

Secondly, some practices are more visible than others. Expert informants suggest that these are the most likely to change in part because these are the practices that are most readily auditable. For example, whether a riparian zone has been maintained or workers are wearing PPE when using hazardous chemicals is easily and immediately apparent. Whether the same company is allowing freedom of association for its workers, using discriminatory hiring practices or harassing staff is much harder to detect, as is understood in the academic literature on auditing: "An audit-only approach is unlikely to be effective in tackling a number of problems also intimately linked to the working conditions of workers, particularly in discovering violations that are intangible such as anti-union policies or forms of discrimination and harassment"(Clean Clothes Campaign, 2015). It was notable that all of the studies identified by the systematic mapping reported positive changes in some of the most intrinsically detectable areas of practice, such as health and safety. Practices that are less detectable, either because they are intangible

or because they are carried out infrequently, may be less prone to practice adoption due to there is less chance of independent auditors finding evidence of practice below the requirement of the standard. Such practices may be also less detectable to researchers, and hence change may be under-reported in the literature.

As described above, few studies attempt to evaluate practice adoption over time. This has a number of consequences for our understanding of practice adoption. Firstly, analysis of standard systems' monitoring and compliance data suggests that maintaining practice – either because no change had occurred or due to the process of responding to non-conformities – was far more common than the adoption of new practices. The corollary of this, as argued by several of the expert informants to this study, is that the main changes to practice may happen before certification, in the period between an entity deciding to apply for certification and the first full certification audit. Unfortunately, there is insufficient evidence within the literature and within schemes' monitoring systems to prove or disprove this, but it remains a possibility.

The literature identified by the systematic mapping also provides insight into the reasons for practice adoption. These are often framed as motivations for certification, and in the context where an entity needs to adopt practices to meet the requirements of a standard the motivations can be synonymous. These commonly cited drivers include the possibility of accessing new (more lucrative) markets, gaining a price premium, gaining an advantage over competitors, managing reputational risks, and responding to demands from customers.

In addition to these, some specific enablers of practice adoption were repeatedly cited. The most commonly of these was the provision of external financial, technical and institutional support. Examples in the literature included support from donors, the government, NGOs, actors higher up the supply chain, and capacity building packages offered by certification systems themselves.

On one level, the frequency with which external support is cited as an important enabler of practice adoption can be explained by the strong focus on developing countries and small-scale producers within the identified literature. However, as discussed above, it is precisely these producers that are frequently some distance from meeting the requirements of international sustainability standards. For this group at least, external support in addition to certification processes appears to be important in driving practice adoption. This is significant, because it implies that although certification is viewed as a market mechanism, it can be most effective when structural market failures in capacity are addressed.

An allied suggestion is that the capacity of certified smallholder producer groups can be greater than those of non-certified producers, which is both a form of management practice adoption and a mechanism for supporting 'on farm' practice adoption. Several informants emphasised the importance of support, and that it needed to be a sustained effort. Firstly, it often takes repeated training to embed practices that are carried out infrequently. Secondly, some individuals and organisations need to be accompanied on a journey of awareness through to recognising their responsibility and agency for sustainable outcomes. The conclusion to be drawn here is that external support and group strengthening appear to be important enablers of practice adoption, and particularly amongst small-scale producers.

Finally, there is confirmation that the process of assessing compliance with the standard on a regular basis (audits) appears to be both driving some practice adoption as well as ensuring that performance is maintained. This is indicated in the literature within the forestry (Cubbage et al., 2010) and marine fisheries sectors (Gorham et al., submitted).

4.3 The Matthew Effect?

The 'Matthew Effect' is the phenomenon whereby one person (or organisation's) initial advantage accumulates (Merton, 1968). In the context of certification, we use it to refer to the possibility that those entities that gain certification were already the best performing, with the attendant logic that they will therefore have had to make the least changes to practice to meet the requirements of a standard. This is a critical consideration when assessing the efficacy of standards in driving practice adoption (and, for that matter, impact), because if it holds true then it limits the 'global' effect of standards on practice adoption.

The most robust way of testing for the Matthew effect would be to have a time series in which the changes in practice of farmers that became certified were compared with those of farmers who remained un-certified. However, as mentioned earlier, the literature is dominated by studies that use a single survey to evaluate the differences between already certified entities and non-certified entities, with very few that have longitudinal data (but see, for example, CRECE, 2014). We consider the paucity of time series data a major gap in understanding the processes and effects of certification.

4.4 Behavioural change

If standards schemes seek to drive practice adoption, and through that, achieve sustainability impacts, then they are really seeking to bring about changes in the behaviour of individuals and of organisations. It is beyond the scope of this research to review all of the potential approaches to behavioural change from disciplines such as behavioural psychology, business studies, development studies (to name just a few relevant disciplines). However, we suggest that it is worth brief consideration as a complement to the 'day-to-day' technical business of managing standards systems. After all, the ultimate goal of sustainability standards is to embed behaviours that result in sustainable outcomes within their chosen area of focus.

There is also a large body of knowledge on the barriers, triggers and motivators to behavioural change for relevant sectors. For example, researchers such as Gershon Feder and colleagues have over decades sought to understand why and how farmers adopt new production technologies (e.g., Feder, Just & Zilberman, 1985; Feder & Umali, 1993). As the research cited in this report demonstrates, much less is known about how producers adopt social and environmental aspects of sustainability, but investments in providing technical, financial and institutional support to producers are attempts to support behavioural change that many sustainability standards schemes employ.

Understanding the barriers, motivations and triggers to behavioural change is one part of the process: making behavioural change happen is another. There are numerous established approaches to bringing about behavioural change, such as Unilever's Five Levers for Change described in Box 2. Whilst few if any standards organisations are likely to have the scale of resources that Unilever has, it seems reasonable to suggest that investment in approaches to behavioural change such as Unilever's Five Levers for Change might be a significant complement to the compliance-based approaches that are at the core of many sustainability standards operations. It would also build upon and inform the existing work of standards schemes on training and in-country partnerships.

4.5 Beyond the certified entity

This study has focused on the adoption of improved practice by certified entities. There

is another sense in which standards can drive improved sustainability standards: by influencing sectoral norms and national policies. For example, Auld, Gulbrandsen & McDermott (2008) argue that FSC has stimulated the development of national standards and codes of conduct. Expert informants proffered anecdotal evidence that sustainability standards had driven changes in forest policies in Germany and Canada, and was being used to influence land use zoning within an African nation. Another interviewee described how the changes legal compliance with labour laws, including health and safety, made by certified companies had begun to change the practices of all companies within the sector locally. The extent to which further standards, codes of conduct and policies change practices on the ground can of course be questioned, but this route to certification impact may have the potential scale of impact to warrant further primary research.

4.6 Recommendations

This section suggests some recommendations for different stakeholders, based on the findings of the research.

Box 2:

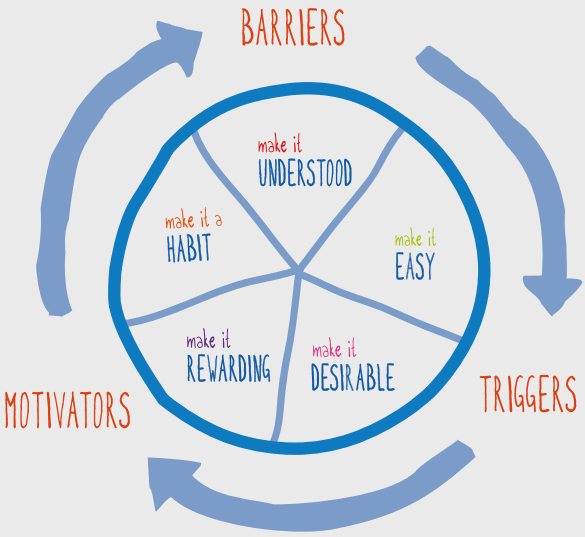
How to change behaviour: Unilever's five levers for change

Unilever are aware that changing certain behaviours such as handwashing and brushing teeth can have enormous health benefits in many developing contexts. Improving these behaviours also potentially expands the market for products like soap and toothpaste, for which Unilever can compete for market share.

Drawing on work and knowledge from a wide range of disciplines, Unilever developed a way of bringing about behavioural change in individuals that they systematised as the 'five levers for change' (Figure 21). Underlying making behavioural change is a sound understanding of the barriers to, triggers of and motivations for change. These are used to making the reasons why change is necessary easy to understand, to make the desired change easy to do, make it desirable and rewarding, and engrain it as a habit. Unilever have used this system to start influencing behaviour on a grand scale. For example, the Lifebuoy Soap handwashing programme alone has reached 379 million people across 29 countries between 2010-2016, with evaluations showing clear improved handwashing behaviour amongst target populations.

Figure 21:

The Five Levers for Change (Unilever, undated.)



4.6.1 For research

- Research with well-designed counterfactuals – either matched control populations of before and after evaluations – is essential to demonstrate change in practice.
- The majority of research that was identified as relevant to the question of the effectiveness of standards on practice adoption was based on single surveys. The amount of longitudinal research needs to increase so that evidence can be provided to key questions of the Matthew Effect, and continual improvement.
- More primary research on certification and standards located further up the supply chain, and in a broader suite of sectors (e.g., mining, tea, sugar cane, textiles and fish) and countries (particularly in Europe and North America), would significantly aid understanding of how certification systems work and the interactions between context and outcomes.
- The impacts of sustainability standards on policies, practices, codes of conduct and national standards is an under-researched area. By influencing these types of policy mechanism, standards may be having a significant impact on broader practises that may be an important corollary to 'on the ground' certification impacts.

4.6.2 For practice

- We recognise the steps made in the standards community to increase the rigour of impact evaluations, and we also recognise that rigorous impact evaluation with counterfactual is prohibitively expensive to use routinely. Our suggestion is to insist on impact evaluations that have rigorous counterfactuals, but do so only when necessary to answer key strategic questions. The resources saved could be invested in rapid, repeat longitudinal surveys, gathering stories of change, and investing in experiential (rather than information based) learning.
- Further invest in understanding and implementing ways of incentivising changes in behavioural and organisational change: expertise on technical aspects of standards can always be hired in, but the fundamental aim of all sustainability standards is to bring about change.
- It is well understood that visible outcomes are more likely to be logged by auditors than intangible and hidden outcomes, and there is a suggestion in the literature and from expert opinion that it is the visible, detectible outcomes where certification may be having the greatest impact. Do we need to develop innovative mechanisms for detecting short-term, hidden changes in complex and rapidly changing supply chains as a complement to auditing?
- Find ways to capture evidence of pre-certification change either by asking certification and assurance bodies to share that data or asking them for analysis of it.
- Think of how a continuous improvement model of standard-setting and auditing can address some of the challenges that the study raises about when practice adoption may be less likely.

4.6.3 For policy

- Most disciplines suffer from short-term funding cycles, and research on standards and certification is no different. As a significant proportion of existing impact studies lack the counterfactual to be able to detect impact, we suggest that only rigorous impact evaluations are funded, and that may mean only funding them for key strategic issues moments. The corollary is to increase investment in longitudinal research.

- Standards are market mechanisms, but that does not mean that they will always work in a free market. Providing the funds and policy environment to ensure that actors anywhere in the supply chain are actively supported to adopt improved practices is a key multiplier to investment in sustainability standards.

4.6.4 For businesses and governments

- Be clear about what sustainability standards are designed to do. If the aim is to raise the bar of the lowest performing producers in a given sector, then standards alone may not be the solution. Support may be needed to bring producers whose starting point is extremely low up to a point where they can then enter certification and maintain good practices.
- Be aware of the role standards play in helping adopt and maintain good practices.

Acknowledgements

The authors would like to acknowledge the following whose generosity with their time and knowledge contributed significantly to the research:

The following researchers were fundamental completing the systematic mapping: Ellen Baker, Andrew Martin, Oliver Hooper, Nelson Mhlanga (all University of Oxford) and Rebecca Guerriero (3Keel LLP).

The Advisory Group to the systematic mapping: Emma Keller (WWF); Paul Jepson, (University of Oxford, Centre for the Environment), Vidya Rangan (ISEAL Alliance), Catherine McCosker (3Keel LLP), Don Seville (Sustainable Food Lab); Sini Savilaakso (Metsäteho Oy); Henk Gilhuis (UTZ); and Katie Longo (Marine Stewardship Council); Ruerd Ruben (Wageningen University & Research).

The following generously contributed certification system data: Aurea Nardelli (Roundtable on Sustainable Biomaterials); Harveen Kour and Daniel Castro (Fairtrade); Daniel Teng and Laszlo Mathe (Accreditation Services International); Soo Chin Oi & Elisa Chong (Roundtable on Sustainable Palm Oil); Katie Longo and Ashleigh Arton (Marine Stewardship Council; Jennifer Clark and Kathryn Green (LEAF Marque); Kendra Pasztor and Shannon Avison (Better Cotton Initiative).

Key informants: Jan-Kees Vis (Unilever), Cristian Soto (Colombian Coffee Growers), George Watene (Global Coffee Platform), KInjal Shah (Responsible Jewellery Council), Sonke Fischer (Accreditation Services International), Julia Young (WWF), Sally Smith (independent researcher) and Indira Moreno Echeverri (UTZ)

This work was funded by the ISEAL Alliance. Particular mention must go to Vidya Rangan and Sophie Mendes at ISEAL for their generous and thoughtful support.

Bibliography

ACKERKNECHT, C., BASSABER, C., REYES, M. & MIRANDA, H. (2005). Environmental certification systems and impacts of their implementation on occupational health and safety in Chilean forest companies. New Zealand Journal of Forestry Science, 35(2/3),153–165

AIDENVIRONMENT (2016). Evaluation of UTZ in the Indoneisan cocoa sector. Aidenvironment, Amsterdam, The Netherlands.

AQUACULTURE STEWARDSHIP COUNCIL (2017). Salmon Standard. Version 1.1.

AQUACULTURE STEWARDSHIP COUNCIL (undated). Theory of change. Available at https://www.asc-aqua.org/what-we-do/how-we-make-a-difference/theory-of-change/ Last accessed 08 February 2018.

AULD, G., GULBRANDSEN, L.H., & MCDERMOTT, C.L. (2008). Certification Schemes and the Impacts on Forests and Forestry. Annual Revue of Environmental Resources, 33, pp187–211

AZHARA, A. SAADUN, N., PUANA, C.L., KAMARUDIN, K., Aziz, N., NURHIDAYU, S. & Fischer, J. (2015). Promoting landscape heterogeneity to improve the biodiversity benefits of certified palm oil production: Evidence from Peninsular Malaysia. Global Ecology and Conservation 3, 553–561

BASS, S., THORNBER, K., MARKOPOULOS, M., ROBERTS S. & GRIEG-GRAN , M.(2001). Certification's impacts on forests, stakeholders and supply chains. Instruments for sustainable private sector forestry series. International Institute for Environment and Development, London.

BELLAMY, A.S., SVENSSON, O., VAN DEN BRINK, P.J., & TEDENGREN, M. (2016). What is in a label? Rainforest-Alliance certified banana production versus non-certified conventional banana production. Global Ecology and Conservation 7, 39–48

BENNETT, M., GIOVANNUCCI, D., Rue, C., AYERAKWA, H., HOLMES, A.A. (2013). Cocoa Farms in Ghana: An Evaluation of the Impact of UTZ Certification on the Sustainability of Smallholders supported by the Solidaridad Cocoa Programme (2010-2012). Committee on Sustainability Assessment (COSA) and Institute of Statistical Social and Economic Research (ISSER) University of Ghana

BLACKMAN, A. & NARANJO, M.A. (2012). Does eco-certification have environmental benefits? Organic coffee in Costa Rica. Ecological Economics 83, 58-66

BOIRAL, O. & GENDRON, Y. (2011). Sustainable Development and Certification Practices: Lessons Learned and Prospects. Business Strategy and the Environment. Volume 20, Issue 5, pages 331–347.

CAMINO, R. de, & ALFARO, M. (1998). Certification in Latin America: Experience To Date. Rural Development Forestry Network. Network paper 23c. Overseas Development Institute, London, UK

CAMPBELL COLLABORATION. 2015. Campbell collaboration systematic reviews: policies and guidelines version 1.2. Campbell Policies and Guidelines Series No. 1 ed. Oslo, Norway

CAUDILL, S.A. & A. RICE, R.A. (2016). Do Bird Friendly® Coffee Criteria Benefit Mammals? Assessment of Mammal Diversity in Chiapas, Mexico. PLOS ONE | DOI:10.1371/journal.pone.0165662

CERUTTI P.O, LESCUYER, G., TSANGA, R., KASSA, S.N., MAPANGOU, P.R., MENDOULA, E.E, MISSAMBA-LOLA, A.P, NASI, R., ECKEBIL, P.P.T. & YEMBE, R.Y. (2014). Social impacts of the Forest Stewardship Council certification: An assessment in the Congo basin. Occasional Paper 103. CIFOR, Bogor, Indonesia.

CLEAN CLOTHES CAMPAIGN 2005. Looking for a quick fix: How weak social auditing is keeping workers in sweatshops, Amsterdam. Available at https://www.evb.ch/fileadmin/files/documents/Konsum/Quickfix05.pdf

COHEN, J. 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37-46

COLLABORATION FOR ENVIRONMENTAL EVIDENCE (CEE), 2013. Guidelines for systematic review and evidence synthesis in environmental management. Version 4.2.

CUBBAGE, F., DIAZ, D., YAPURA, P. & DUBE, F. 2010. Impacts of forest management certification in Argentina and Chile. Forest Policy and Economics 12: 497–504

DEFRIES, R.S, FANZO, J., MONDAL, P., REMANS, R. & WOOD, S.A. (2017), Is voluntary certification of tropical agricultural commodities achieving sustainability goals for small-scale producers? A review of the evidence. Environmental Research Letters. 12 033001

FERNANDEZ-CORNEJO, J., GREENE, C., PENN, R., & NEWTON, D. (1998). Organic vegetable production in the U.S.: Certified growers and their practices. American Journal of Alternative Agriculture, Volume 13 (2), 69-77

FEDER, G., & UMALI, D.L. (1993). The adoption of agricultural innovations: a review. Technological Forecasting and Social change 43 (3-4), 215-239

FEDER, G. JUST, R.E. & ZILBERMAN, D. (1985). Adoption of agricultural innovations in

developing countries: A survey. Economic Development and Cultural Change 33 (2), 255-298.

FOREST STEWARDSHIP COUNCIL (2015). FSC's theory of change. Available at https://ic.fsc.org/en/fsc-theory-of-change Last accessed 08 February 2018.

FOSTER, B.C, WANG, D., KEETON, W.S. (2008). An Exploratory, Post-Harvest Comparison of Ecological and Economic Characteristics of Forest Stewardship Council Certified and Uncertified Northern Hardwood Stands. Journal of Sustainable Forestry 26(3), 171-191

FRAMPTON, G. K., LIVOREIL, B. & PETROKOFSKY, G. 2017. Eligibility screening in evidence synthesis of environmental management topics. Environmental Evidence, 6, 27.

GALATI, A. GIANGUZZI, G., TINERVIA, S., CRESCIMANNO M. & LA MELA VECA, D.S. (2017). Motivations, adoption and impact of voluntary environmental certification in the Italian Forest based industry: The case of the FSC standard. Forest Policy and Economics, 83,169-176

GARCÍA, C., GARCÍA, J., OCHOA, G., MORA, J. C. & CASTELLANOS, J. F. (2014). Impact Evaluation of UTZ Certified Coffee Program in Colombia. (2008-2012). CRECE. Manizales, Colombia

GARCÍA, G. OCHOA, J. GARCÍA, J. MORA, J. & CASTELLANOS, J.F. (undated). Use of Polychoric Indexes to Measure the Impact of Seven Sustainability Programs on Coffee Growers' Livelihood in Colombia. Centre for Regional Coffee and Entrepreneurial Research, Manizales, Colombia

GARSIDE, R. 2014. Should we appraise the quality of qualitative research reports for systematic reviews, and if so, how? Innovat Eur J Soc Sci Res, 27.

GORHAM, T.M., GOOD, S.G., LÉFÈBURE, R, & Longo, C.S. (submitted). Evidence for effects of seafood ecolabeling on managing fisheries' ecosystem impacts.

HAIN, H. & R. AHAS (2007). Can Forest Certification Improve Forest Management? Case Study of the FSC Certified Estonian State Forest Management Centre. International Forestry Review, 9(3), 759-770

HARTMAN, G. L.; PAWLOWSKI, M. L.; HERMAN, T. K.; EASTBURN, D. (2016). Organically Grown Soybean Production in the USA: Constraints and Management of Pathogens and Insect Pests. Agronomy, 6, 16; doi:10.3390/agronomy6010016

HIGGINS, J. P. & GREEN, S. 2011. Cochrane handbook for systematic reviews of interventions, Chichester, Wiley.

HOANG, H.T.N., HOSHINO, S. & HASHIMOTO, S. (2015). Forest stewardship council certificate for a group of planters in Vietnam: SWOT analysis and implications. Journal of Forest Research 20, 35–42

JAMES, K. L., RANDALL, N. P. & HADDAWAY, N. R. 2016. A methodology for systematic mapping in environmental sciences. Environmental Evidence, 5, 7.

JENNINGS, S. (2017). Expecting too much, getting too little? A think piece on sustainability certification auditing in the oil palm sector. WWF Switzerland.

JOHANSSON, J. & Gun LIDESTAV (2011). Can voluntary standards regulate forestry? Assessing the environmental impacts of forest certification in Sweden. Forest Policy and Economics 13, 191–198

KALONGA, S.K. & KULINDWA, K.A. (2017). Does forest certification enhance livelihood conditions? Empirical evidence from forest management in Kilwa District, Tanzania.

Forest Policy and Economics 74, 49–61

KERSTING, S. & WOLLNI, M. (2012). New institutional arrangements and standard adoption: Evidence from small-scale fruit and vegetable farmers in Thailand. Food Policy 37, 452–462

LIVOREIL, B., GLANVILLE, J., HADDAWAY, N. R., BAYLISS, H., BETHEL, A., DE LACHAPELLE, F. F., ROBALINO, S., SAVILAAKSO, S., ZHOU, W., PETROKOFSKY, G. & FRAMPTON, G. 2017. Systematic searching for environmental evidence using multiple tools and sources. Environmental Evidence, 6, 23.

LOCONTO A. M. & SIMBUA, E.F. (2010). Organizing Smallholder Production Sustainability: Lessons Learned From Fairtrade Certification In The Tanzanian Tea Industry. ISDA, Montpellier, France. CiradInra-SupAgro, 16 pp.

LOURENZANI, W.L, LOURENZANI, A.E.B.S, PIGATTO, G. & SIMON, E.J. (2006). O papel da certificação no programa de desenvolvimento da fruticultura na região da nova alta Paulista. Informações Econômicas, SP, 36 (2), 30-37

MARIE-VIVIEN, D., GARCIA, C.A. KUSHALAPPA, C.G. & VAAST, P. (2014). Trademarks, Geographical Indications and Environmental Labelling to Promote Biodiversity: The Case of Agroforestry Coffee in India Development Policy Review, 32 (4), 379–398.

MARINE STEWARDSHIP COUNCIL (2011). MSC Theory of Change. Available at https://www.msc.org/documents/msc-brochures/msc-theory-of-change/view Last accessed 08 February 2018.

MARKOPOULOS, M.D. (1998). The Impacts of Certification on Community Forest Enterprises: A Case Study of the Lomerío Community Forest Management Project, Bolivia Oxford Forestry Institute (OFI), University of Oxford, UK.

MARTIN, A.C., JEFFERS, E.S., PETROKOFSKY, G., MYERS-SMITH, I. & MACIAS-FAURIA, M. 2017. Shrub growth and expansion in the Arctic tundra: an assessment of controlling factors using an evidence-based approach. Environmental Research Letters, 12 (8). DOI: 10.1088/1748-9326/aa7989

MENDOZA, T.C. (2004). Evaluating the Benefits of Organic Farming in Rice Agroecosystems in the Philippines. Journal of Sustainable Agriculture, 24 (2), 93-115

MERTON, R.K. (1068). The Matthew Effect in Science. Science, 159 (3810), 56-63

MITEVA, D.A., LOUCKS, C.J. & PATTANAYA, S.K. (2015). Social and Environmental Impacts of Forest Management Certification in Indonesia. PLOS ONE DOI:10.1371/journal.pone.012967

MOHAN, S. (2016). Institutional Change in Value Chains: Evidence from Tea in Nepal. World Development, 78, 52–65

MONTANO, J., NAWATA, E. & PANICHSAKPATANA, S. (2016). Do GAP farmers do better than non-GAP farmers? Pesticide management practices of horticultural farmers in Damnoen Saduak, Thailand. Tropical Agriculture and Development 60 (1), 1-9

MOORE, S.E., CUBBAGE, F. & EICHELDINGER, C. (2012). Impacts of Forest Stewardship Council (FSC) and Sustainable Forestry Initiative (SFI) Forest Certification in North America. Journal of Forestry, 110 (2), 79-88

NASI, R., BILLAND A. & VANVLIET, N. (2012). Managing for timber and biodiversity in the Congo Basin. Forest Ecology and Management 268, 103–111

O'ROURKE, D. 2000. Monitoring the monitors: A critique of PricewaterhouseCoopers labor monitoring. Available at http://www.bollettinoadapt.it/old/files/document/18107ROURKE_2000.pdf

OPITZ, S., HOFFMANN, J., QUAAS, M., MATZ-LÜCK, N., BINOHLAN, C. & FROESE, R. (2016). Assessment of MSC-certified fish stocks in the Northeast Atlantic. Marine Policy 71, 10–14

OYA, C, SCHAEFER, F, SKALIDOU, D, MCCOSKER, C & LANGER, L, 2017. Effects of certification schemes for agricultural production on socio-economic outcomes in low- and middle-income countries: A systematic review. 3ie Systematic Review 34. London: International Initiative for Impact Evaluation (3ie)

PÉREZ-RAMÍREZ, M., CASTREJÓN, M., GUTIÉRREZ N.L., & DEFEO, O. (2016). The Marine Stewardship Council certification in Latin America and the Caribbean: A review of experiences, potentials and pitfalls. Fisheries Research 182, 50–58

PINTO, L.F.G., GONÇALVES, E.T., BERKEMBROCK, D., YAMADA, M & SORICE, L. (2013). Características, potencial e limitações de diferentes modalidades de certificação socioambiental para sistemas agroflorestais. Cadernos de Ciência & Tecnologia, Brasília, 30 (1/3), 11-32

QIAO, Y., HALBERG, N., VAHEESAN, S., & SCOTT, S. (2015). 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 Agriculture and Food Systems: 31(3); 246–25.

RESPONSIBLE JEWELLERY COUNCIL (undated). Theory of Change. Available at https://www.responsiblejewellery.com/files/Theory-of-Change1.jpg Last accessed 08 February 2018.

RIJN, F.C., JUDGE, L.O. , FORT, R., KOSTER, T., WAARTS Y.R. & RUBEN, R. 2016. Fairtrade certification in the hired labour banana sector. Wageningen, LEI Wageningen UR (University & Research centre), LEI Report 2015-056

ROE, D., DAY, M., BOOKER, F., ZHOU, W., ALLEBONE-WEBB, S., KÜMPEL, N., HILL, N. A. O., WRIGHT, J., RUST, N., SUNDERLAND, T. C., REDFORD, K. & PETROKOFSKY, G. 2014. Are alternative livelihood projects effective at reducing local threats to specified elements of biodiversity and/or improving or maintaining the conservation status of those elements?: a systematic review protocol. Environmental Evidence, 3, 6.

RUBEN, R. & ZUNIGA, G. (2011). How standards compete: comparative impact of coffee certification schemes in Northern Nicaragua. Supply Chain Management: An International Journal 16 (2), 98–109

RUBEN, R. & FORT, R. (2011). The Impact of Fair Trade Certification for Coffee Farmers in Peru. World Development, 40 (3), 570–582

RUEDA, X. & LAMBIN, E.F. (2013). Responding to Globalization: Impacts of Certification on Colombian Small-Scale Coffee Growers. Ecology and Society18 (3): 21

RUEDA, X. & Lambin, E.F. (2013). Responding to Globalization: Impacts of Certification on COLOMBIAN Small-Scale Coffee Growers. Ecology and Society18 (3): 21.

RUEDA, X, THOMAS, N.E., & LAMBIN, E.F. (2015). Eco-certification and coffee cultivation enhance tree cover and forest connectivity in the Colombian coffee landscapes. Regional Environmental Change 15, 25–33

RUTTEN-SÜLZ, M. (2014). A snapshot from the field: five countries over five years. The 4C Association, Germany.

SASWATTECHA, K., KROEZE, C., JAWJIT, W. & HEIN, L. (2015). Assessing the environmental impact of palm oil produced in Thailand. Journal of Cleaner Production 100, 150-169

SELDEN, R.L., VALENCIA, S.R., LARSEN, A.E., CORNEJO-DONOSO, J. & WASSERMAN, A.A. (2016). Evaluating seafood eco-labeling as a mechanism to reduce collateral impacts of fisheries in an ecosystem-based fisheries management context. Marine Policy 64, 102–115

SHRECK, A. (2002). Just Bananas? Fair Trade Banana Production in The Dominican Republic. International Journal of Sociology of Agriculture and Food 10, 2: 13-23

SILVA-CASTAÑEDA, L. (2012). A forest of evidence: third-party certification and multiple forms of proof—a case study of oil palm plantations in Indonesia. Agriculture and Human Values, Volume 29, Issue 3, pp 361-370

SMITH DUMONT, E., GNAHOUA, G. M., OHOUO, L., SINCLAIR, F. L. , VAAST, P. (2014). Farmers in Côte d'Ivoire value integrating tree diversity in cocoa for the provision of ecosystem services. Agroforestrty Systems 88:1047–1066

SOUZA, A.P.O., OLIVEIRA SANS, E.C. de, MÜLLER B.R. & MOLENTO C.F.M. (2015). Broiler chicken welfare assessment in GLOBALGAP® certified and noncertified farms in Brazil. Animal Welfare, 24: 45-54

STUART, D., BENVENISTE, E. & HARRIS, L.M. (2014). Evaluating the use of an environmental assurance program to address pollution from United States cropland. Land Use Policy 39, 34–43

SVERDRUP-THYGESON, A., BORG, P. & BERGSAKER, E. (2008) A comparison of biodiversity values in boreal forest regeneration areas before and after forest certification. Scandinavian Journal of Forest Research, 23 (3), 236-243

TAKAHASHIA R. & TODO, Y. (2014). The impact of a shade coffee certification program on forest conservation using remote sensing and household data. Environmental Impact Assessment 44, 76-81

THE VIA INITIATIVE (2017). M-2b Guidelines for the Study of Corrective Action Requests (CARS). Available from the ISEAL Alliance.

THOMAS, J., NOEL-STORR, A. N., MARSHALL, I., WALLACE, B., MCDONALD, S., MAVERGAMES, C., GLASZIOU, P., SHEMILT, I., ANNELIESE SYNNOT, A., TURNER, T. ELLIOTT, J. 2017. Living systematic reviews: 2. Combining human and machine effort. Journal of Clinical Epidemiology, 91: 31-37

THORN, J., SNADDON, J., WALDRON, A., KOK, K., ZHOU, W., BHAGWAT, S., WILLIS, K. & PETROKOFSKY, G. 2015. How effective are on-farm conservation land management strategies for preserving ecosystem services in developing countries? A systematic map protocol. Environmental Evidence 2015, 4:11.

UNILEVER (undated). Inspiring Sustainable Living: expert insights into consumer behaviour and Unilever's Five Levers for Change. Unilever PLC, London

UNILEVER (undated). Changing Handwashing Habits for better health. https://www.unilever.com/sustainable-living/improving-health-and-well-being/health-and-hygiene/changing-handwashing-habits-for-better-health/. Last accessed 25 January 2018.

UTTING-CHAMORRO, K. (2005) Does fair trade make a difference? The case of small coffee producers in Nicaragua, Development in Practice, 15:3-4, 584-599.

UTZ (2017). Theory of Change. Available at https://utz.org/wp-content/uploads/2017/07/UTZ-Theory-of-Change_2017.pdf Last accessed 08 February 2018.

Appendix 1: Effectiveness of sustainability standards in driving practice adoption: results and learnings. A systematic map protocol

Authors

Gillian Petrokofsky[1], Ellen Baker[1], Andrew Martin[1], Oliver Hooper[1], Nelson Mhlanga[1], Rebecca Guerriero[2], Steve Jennings[2]

1. University of Oxford

2. 3Keel LLP

Keywords

Certification, sustainability standards, practice adoption, systematic mapping, visualisation

Background

The present study aims to determine what evidence exists in the published and grey literature on what drives the adoption and maintenance sustainability standards by undertaking a systematic mapping approach, following as closely as possible good practice of systematic review (Collaboration for Environmental Evidence (2013), Campbell Collaboration (2015), Cochrane Collaboration (Higgins & Green 2011)) and guidance for systematic maps (James et al 2016) to synthesise existing evidence on the business benefits of using credible sustainability standards.

The particular set of questions posed in the current evaluation lends itself best to the systematic map approach. Systematic mapping is particularly valuable for broad, multi-faceted questions that can include multiple interventions, populations or outcomes; they can cover the breadth of evidence needed for policy- or practice-relevant questions of the type under consideration here. Systematic mapping follows the same rigorous processes as systematic reviews to evaluate relevant evidence and minimise the potential biases and lack of transparency of traditional literature reviews. In systematic mapping, the evidence is presented in a searchable database, with clearly defined elements that are coded similarly across the body of evidence collected through an extensive search of multiple sources (including academic journals and other sources of information, such as organisational databases, collections of theses, unpublished reports, and publications suggested by stakeholders interested in the review question). As systematic maps may include multiple populations, interventions or exposures, the database usually enables cross-tabulations of the data to be carried out to explore the evidence base thoroughly. It then becomes possible to identify trends and knowledge gaps and clusters. In further contrast with systematic reviews, systematic maps are unlikely to include detailed extraction of study results or statistical synthesis of results. The mapping process involves rigorous (i) searching for evidence, (ii) selection of relevant evidence ('filtering'), and (iii) presentation of key elements of individual studies in the evidence set ('coding'). These are described in detail below.

Table A1.1 shows sectors of interest to the review, with information on their linkage with ISEAL. The review will not be limited to ISEAL Members' coverage or information, but the review will draw heavily on existing information available through ISEAL.

Table A1.1

Sectors and sustainability standards

Sector of interest	Broad sustainability theme or issue of interest	Specific outcomes of interest linked to standards' practices or performance benchmarks	Relevant ISEAL member standards
Agriculture – coffee, cocoa, tea, palm oil, bananas, horticulture	Farming practices in smallholder farming and larger commercial agriculture set-ups	• Crop productivity • Quality of crop • Use of pesticides and fertilizers • Soil conservation • Water management	BCI, Bonsucro, Fairtrade, GCP, RSPO, RTRS, Rainforest Alliance / SAN, LEAF, UTZ.
Fishing – open sea fishing and aquaculture	Fishing and farming practices	• Quality of product • Reduction in bycatch (open sea fishing) • Use of antibiotics and other toxics during aquaculture farming	ASC and MSC
Forestry	Forestry management practices (noting that the certified entity could be a large private or publicly owned forest land or collectively owned forest land)	• Forestry and logging practices • Practices related to forest management	FSC
Agriculture, forestry, fishing, textiles and mining	Labour practices and occupational health and safety	• Practices related to gender variables, issues and outcomes • Occupational health and safety • Workers' rights and empowerment	All relevant members
Agriculture, forestry, fishing, textiles and mining	Water usage and management practices	• Water consumption • Wastewater management • Water preservation including preservation of natural water bodies	All relevant members
Agriculture, forestry, fishing, textiles and mining	Biodiversity and habitat preservation	• Deforestation • Species preservation	All relevant members

Advisory Group

An Advisory Group (AG) was assembled to represent as wide a range of standards as possible, and included academics with an interest in sustainable standards and/or the application of systematic evaluation techniques (see Appendix A1.1 for list of AG members). The AG met in Oxford on October 4th, 2017. Inputs from the AG were incorporated into the draft protocol (including conceptual framework, definitions and boundaries of the key elements for the systematic map (the 'PICOs' - see below), keywords, and websites for grey literature. A shared Dropbox was set up to invite further inputs from the AG. Comments received have informed the sections below. A second meeting will be convened in early 2018 to discuss data analysis.

Method

Figure 1 shows the conceptual framework devised to examine the review within the context of drivers and barriers, and other confounding factors, that affect adoption, maintenance and evaluation of success.

Primary review questions:

RQ1. What is the effectiveness of sustainability standards and certification in driving practice adoption?

RQ2. To what extent does adopting practices lead to continuous improvement in entities over time in identified thematic areas?

Secondary review questions:

1. Are there differences in the sustainability practices of organizations that get certified and those that don't? This question helps us understand the extent to which compliance with a standard is the result of practice adoption as opposed to being the verification of pre-existing practice within an individual company.

2. Are there differences in the sustainability practices of organizations that commit to certification (i.e., between scoping and the first certification audit)? This research question helps us to understand the degree to which practice adoption happens before certification.

3. (i) How do sustainability practices change after certification? This question helps us understand how the cycle of a certified entity's responses to repeated audit findings drives changes in practice.

(ii)How do standards and certification tools operate to achieve practice adoption?

Objectives of the systematic map

1. Bring together formally documented literature that examines changes in practice adoption, both intended (+) and not (-), caused by application of sustainability standards.

2. Highlight areas, both geographically and by sector, which represent a gap in current study and knowledge.

3. Qualitative discussion of how standards and certification tools operate to achieve practice adoption.

4. Produce an online interactive map, searchable by topic.

Search strategy

Following good practice guidance (Livoreil et al 2017) we will review studies published in peer-reviewed journals and the 'grey' literature (conference papers, book chapters, reports from NGO Organizations, and unpublished reports available on Organizational websites) by following an agreed search strategy, to be finalised at the first Advisory Group meeting. We are aware of the possible limitations of confining the work to published work only, and in particular to studies published in peer-reviewed journals, and we will therefore work with the Advisory Group to disseminate a wide call for existing evidence from Organizations working in this field in order to obtain non-journal studies of relevance to the review. Subject to AG approval, we will limit studies to be included to those published from 1990 to the present.

Language

Literature published in English, French, Spanish and Portuguese will be read to full text level. Those in other languages will be recorded in the systematic map, where possible, but not assessed comprehensively for practical reasons. After consultation with the Advisory Group it was decided that this should give good coverage of the current body of literature, leaving only Russia and China as areas omitted which may have significant collections of relevant studies. Any future assessments or evaluations built on this map may then include these languages to remove this potential source of publication bias.

Key search terms

An evidence synthesis process starts with a question that is usually structured into "building blocks" (concepts or elements), some of which are then used to develop the search strategy (Livoreil et al 2017). The current search strategy (and later data extraction uses the "PICO" elements which are commonly used in CEE evidence synthesis. Appendix A1.2 reproduces a table describing the PICO elements in general (Livoreil et al 2017).

Following these best practice guidelines, the review will be structured using the PICO framework (see below) and key concepts and keywords associated with each category will be developed and tested for comprehensiveness and specificity:

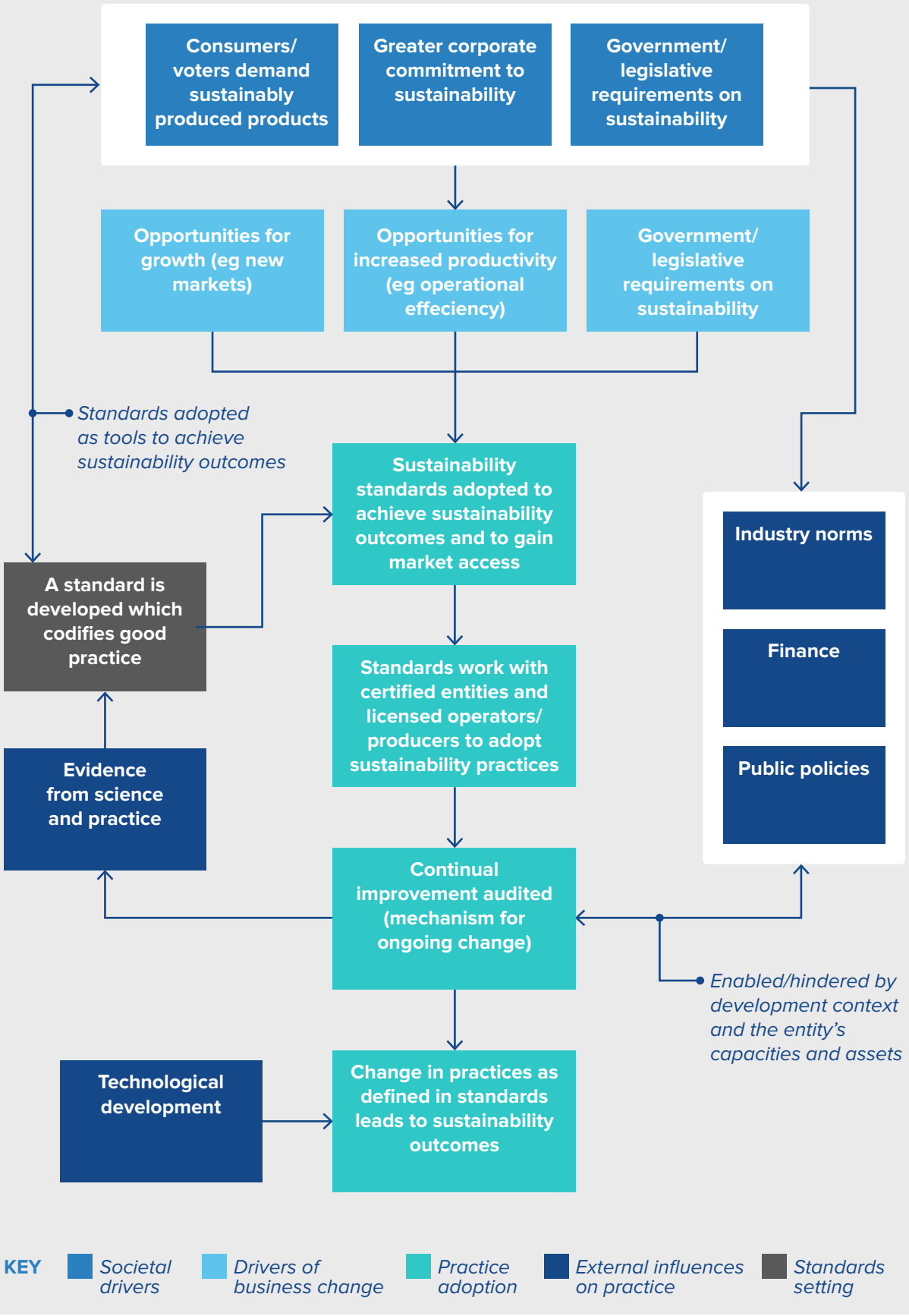
- **Population (P)** –Organizations and growers producing or marketing commodities for which certification schemes have been developed
- **Intervention (I)** – Certification scheme adopted by organization/grower (e.g., FSC, Fairtrade, etc.)
- **Counterfactual/Control (C)** - Non-adoption (in time or space, e.g., before/after certification; certified vs non-certified compared in neighbouring or similar farms, etc.). The systematic map will include only studies that have matched or pre/post comparisons with adoption of a certification scheme/sustainability standard.
- **Outcomes (O)** – measures of sustainability (social, economic and environmental measures).

We amplify these four PICO elements below in section on Study screening, inclusion and exclusion criteria.

The search logic for the three principle databases searched are included in Appendix A1.3 These were developed from key terms contributed by the review team and from members of the Advisory Group, following the first Advisory Group meeting.

Figure A1.1:

Conceptual framework



Sources of publications
Bibliographic databases

- Web of Science published by Thomson Reuter's (formally ISI) Web of Science, New York, USA **http://apps.webofknowledge.com/**
- SCOPUS published by Elsevier **http://www.elsevier.com/online-tools/scopus**
- CAB Abstracts published by CAB International, Wallingford, UK **http://www.cabdirect.org/**

Search engine searching

To ensure coverage of sources not captured by the above databases the met-search engine Google Scholar (**http://scholar.google.com**) will be searched using Harzing's Publish or Perish open-source software (**https://harzing.com/resources/publish-or-perish**). The first 2000 records will be exported.

Grey literature and other sources not captured above

Based on consultation with the advisory committee a selection of journals, Organizational archives, online databases and papers will be searched. Again, any arising duplicates compared to bibliographic searches will be removed before inclusion in the reference file. We will distribute a purpose-designed flyer requesting grey literature for input we hope to capture any key research papers, reports or case studies relevant to the research questions. In addition to peer-reviewed work other formally published grey literature will be collated from the websites of the major standards Organizations and from ISEAL. Key terms used to search grey literature will be amended from strings used for peer-reviewed bibliographic databases to take account of the limited search capability of Organizational websites.

Comprehensiveness of search

Following good practice, the search logic was developed and tested iteratively against a set of studies which were assessed to be of central interest to the review question. The studies are not guaranteed to progress through all stages of inclusion into the final systematic map, but they are sufficiently germane to the review question to be used to test search logic and breadth of keywords in the search strings. The list of these papers is included in Appendix A1.4.

Study screening, inclusion and exclusion criteria

All results from the literature searches will be collated in referencing software using EndNote and Mendeley. Inbuilt software in these programs will be used to remove duplicates prior to screening. Screening will be carried out by two research assistants with Kappa tests performed at the start of each screening stage on a random subset of 100 studies to confirm alignment of criteria interpretation between reviewers (Cohen 1960). The online tool Abstrackr will be used to record reviewer decisions, with a conservative approach taken to study rejection. Screening will begin at title level followed by abstract and full text, studies in French, Spanish or Portuguese will be screened by either or both of the project leads due to language constraints. An online calculator for free-marginal kappa will be used, available at http://justusrandolph.net/kappa/. If the kappa coefficient is below 0.6, then the kappa analysis will be repeated on additional sets of randomly selected studies until agreement reaches at least 0.6. To supplement the kappa analysis of agreement, random samples of rejected studies will

be examined and any disagreements between the reviewers discussed and resolved through regular meetings following good practice guidance (Frampton et al 2017).

General inclusion criteria

In order to be **included**, studies must:

- Report a sustainability standard/certification
- Report comparison between certified non-certified
- Report outcomes – these will be economic, social or environmental

Studies will be **excluded** if they report only:

- Environmental management (unless also includes relevant terms like a certain
- Consumer, supply chain, chain of certification
- Policy based/regulations/co-ops unless there is an explicit standard mentioned
- Appellation of origin/protected designated area, etc.
- Food safety/quality
- Farmer field schools or Co-ops, unless they also mention certification
- Personal reasons and/or motivations for adoption (e.g., if the study reports "I wanted to become certified because I thought I would get a price premium" reject. Compare with "Since being certified I have evidence of a price premium" would be accepted.
- Kosher or Halal (not sustainability standards per se)
- Food safety
- Land ownership/certification

More detailed inclusion criteria are listed according to PICO definitions below:

P - Population

Any entity that is the direct receiver of certification by a sustainability standard will be treated as the relevant population. This will encompass parts of the supply chain beyond producers. i.e. trader, buyer etc. helping to remove biases and allow results to be more widely applicable. Groups involved in supply chain custody but not directly certified will therefore not be included in the systematic map but may form part of the discussion.

I/C- Intervention (and counterfactual)

Studies which compare practice adoption pre- and post- certification in a single entity and those which compare between equivalent certified and non-certified entities will be considered to contain the relevant intervention. Examination of a set pre-certification period may show if there is something else driving practice adoption other than certification e.g., collaboration with NGOs. Method sections in selected studies should be explicit enough to identify whether activities are associated with certification or simply developing better practice. Studies will not be included if they compare two or more certified entities without considering an uncertified entity (or a before-after comparison of the multiple entities), whether different certification standards or the same.

O- Outcomes

Through discussion with the Advisory Group a very simple classification of practices adopted in response to certification was devised (positive, neutral, negative outcomes). Studies must report one of these as an outcome to be included in the systematic map. This will include both intended/unintended consequences of certification and will largely be drawn from the requirements of known sustainability standards. The outcomes fall into three broad categories (Social, Economic, Environmental) for ease of document handling; the interplay between the three categories is recognized. The 24 outcomes are:

Education, Poverty, health and safety, Nutrition, Gender, Wellbeing, Community Benefits and Development, Knowledge Exchange, Transparency, Management systems, Price Premium, Profitability , Costs, Market Advantage, Yield, Good Production Practices (including good forest management), Financial Security, Deforestation, Conservation and Biodiversity, Pollution, Soil Management, Water Management, Input Use, Post-harvest Practices.

Studies will not be included if the outcomes are based only on modelling. Also excluded will be studies which report only narrative perceptions of outcomes or impact, without quantitative data. This was not because this type of study is not considered equally important, but because it requires a different type of systematic review with a strong focus on coding qualitative results from social research papers.

Relevant types of study design

Studies will be included in the analysis if they report measured sustainability criteria under the three broad headings 'Social', Environmental' 'Economic', and they compare certified and non-certified entities of the same entities or entities before-and-after certification. Within the broad sustainability criteria, we will record the most frequently-occurring sub-categories iteratively as we proceed with data extraction (see below). We will note other important information, such as funding sources which may lead to possible conflicts of interest, and social/environmental factors that may have influenced results (extreme weather, political or social upheaval, etc.).

Study quality assessment

Critical appraisal (often known as 'quality control') is not mandatory for systematic maps prepared for the Collaboration for Environmental Evidence, but was considered highly desirable by the commissioners and, initially, the Advisory Group and it was envisaged that critical appraisal would closely follow Roe et al (2014), which had been based on previous CEE systematic reviews and had been subsequently adapted by other reviews. Since starting the project, however, the new CEE draft guidelines (in press) have cast doubt on the robustness of the type of approach we had thought might be suitable (Frampton, pers com).

The aim of the current project is not to advance theory in this area, and the Advisory Group agreed that the systematic mapping approach provided sufficient robust filtering and would not benefit from an approach that was not generally now thought to be suitable. Using Garside (2014) for assessing the quality of qualitative research was not undertaken for similar reasons.

Coding variables

Data to enable analysis of the impacts of certification on sustainability will be determined in collaboration with the Advisory Group at the first meeting. Descriptive data extracted for the systematic evidence map will include, but not be limited to: year of publication, type of publication, location of study, type of intervention (certification scheme), length of time under certification, reported drivers of adoption, barriers to adoption, reported sustainability outcomes (positive, neutral, or negative). Qualitative data for inclusion in the report will be discussed with Advisory Group members at the first meeting and an analysis method agreed.

Table A1.2 shows the outcome/impact variables that were developed with input from the Advisory Group iteratively.

Table A1.2

Coding variables

Coder ID	Info
Study ID	
Report ID	Source ID/endnote
Rejected	
Author	First only
Publication date	
Publication type	Book, paper etc.
Journal	If applicable
Study	Can be multiple
Study context comments	
Country	
Latitude-Longitude	
Production system/sector	e.g., fishing, agriculture
Commodity	e.g., coffee
Certification scheme/standard	
Population	Certified entity: Individual, co-op, company etc.
Sampling Unit	
Position in supply chain	Producer or processor
Study design	Matched or pre-post
Sample size (only matched design)	Certified
	Non-certified
Treatment arms	e.g., control, FT, RA
Social	Education
	Poverty
	health and safety
	Nutrition
	Gender
	Wellbeing
	Community benefits/development
	Knowledge exchange
Economic	Transparency
	Management systems
	Price premium
	Profitability
	Costs
	Yield
	Market advantage
	Financial security
	Good Production Practices (including good forest management)
Environmental	Deforestation
	Conservation & Biodiversity
	Pollution
	Soil management
	Water management
	Input use [fertilizer, insecticide, fungicide, etc.]
	Post-harvest practices
Notes Aqua	
Notes	

Visualizing the results of the systematic map

Included studies will be presented on an interactive map (following Thorn et al (2015) and Martin et al (2017) on a platform initially hosted by University of Oxford's Long-term Ecology lab. The studies are selected on the basis of coding for latitude-longitude. Where multiple locations are recorded in the coding sheet, the study will appear as a circle on the world map in each location, colour-coded for commodity. Filters will be selected to enable users to interrogate the map for combinations of different variables. The map will be useable on most commonly-used devices (pc, tablet, mobile phone). Any data added to the systematic map over time can be drawn into the online global map, so that updates and amendments will be possible in response to stakeholder feedback and popular use.

References

COLLABORATION FOR ENVIRONMENTAL EVIDENCE (CEE), 2013. Guidelines for systematic review and evidence synthesis in environmental management. Version 4.2.

CAMPBELL COLLABORATION. 2015. Campbell collaboration systematic reviews: policies and guidelines version 1.2. Campbell Policies and Guidelines Series No. 1 ed. Oslo, Norway

COHEN, J. 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37-46

HIGGINS, J. P. & GREEN, S. 2011. Cochrane handbook for systematic reviews of interventions, Chichester, Wiley.

JAMES, K. L., RANDALL, N. P. & HADDAWAY, N. R. 2016. A methodology for systematic mapping in environmental sciences. Environmental Evidence, 5, 7.

LIVOREIL, B., GLANVILLE, J., HADDAWAY, N. R., BAYLISS, H., BETHEL, A., DE LACHAPELLE, F. F., ROBALINO, S., SAVILAAKSO, S., ZHOU, W., PETROKOFSKY, G. & FRAMPTON, G. 2017. Systematic searching for environmental evidence using multiple tools and sources. Environmental Evidence, 6, 23.

FRAMPTON, G. K., LIVOREIL, B. & PETROKOFSKY, G. 2017. Eligibility screening in evidence synthesis of environmental management topics. Environmental Evidence, 6, 27.

GARSIDE, R. 2014. Should we appraise the quality of qualitative research reports for systematic reviews, and if so, how? Innovat Eur J Soc Sci Res, 27.

MARTIN, A.C., JEFFERS, E.S., PETROKOFSKY, G., MYERS-SMITH, I. & MACIAS-FAURIA, M. 2017. Shrub growth and expansion in the Arctic tundra: an assessment of controlling factors using an evidence-based approach. Environmental Research Letters, 12 (8). DOI: 10.1088/1748-9326/aa7989

ROE, D., DAY, M., BOOKER, F., ZHOU, W., ALLEBONE-WEBB, S., KÜMPEL, N., HILL, N. A. O., WRIGHT, J., RUST, N., SUNDERLAND, T. C., REDFORD, K. & PETROKOFSKY, G. 2014. Are alternative livelihood projects effective at reducing local threats to specified elements of biodiversity and/or improving or maintaining the conservation status of those elements?: a systematic review protocol. Environmental Evidence, 3, 6.

THORN, J., SNADDON, J., WALDRON, A., KOK, K., ZHOU, W., BHAGWAT, S., WILLIS, K. & PETROKOFSKY, G. 2015. How effective are on-farm conservation land management strategies for preserving ecosystem services in developing countries? A systematic map protocol. Environmental Evidence 2015, 4:11.

Appendix A1.1. Advisory Group members

Emma Keller, WWF
Paul Jepson, University of Oxford, Centre for the Environment
Vidya Rangan, ISEAL Alliance
Catherine McCosker, 3Keel
Don Seville, Sustainable Food Lab
Sini Savilaakso, University of Helsinki
Henk Gilhuis, UTZ
Katie Longo, MSC
Ruerd Ruben, Wageningen University & Research

Appendix A1.2

Explanation of the "PICO" elements of systematic reviews & maps

Elements of a reviewable PICO/PECO question, often structured as "does intervention (I) or exposure (E) applied to populations (P) produces outcome (O) [compared to comparator (C)]?" (Livoreil et al 2017)

Question element	Definition
Population (of subjects)	Statistical samples or populations of subject(s) (e.g., ecosystem, species, etc.), to which the interventions will be applied, or exposed to described conditions
Intervention/Exposure	Policy, action or environmental variable impacting the populations or to which the subject populations are exposed
Comparator	What the exposure or intervention are compared to. Either a control with no intervention/exposure or an alternative intervention or a counterfactual scenario
Outcome	Consequences of the intervention or exposure. All relevant variables that can be reliably measured

Appendix A1.3 Search strings for major bibliographic databases

CAB abstracts# 6309 (20/10/17)

(certif* OR member*) adj3 (aquaculture OR forest* OR forest management OR silvicultur* OR agricultur* OR mining OR farm* OR cultivat* OR fish* OR wood* OR timber* OR fruit* OR banana* OR textile* OR commodit* OR material* OR product* OR mineral* OR jewel* OR agroforestry* OR crop* OR coffee OR oilseed OR cocoa OR tea OR palm oil OR oil palm OR oilpalm OR palmoil OR oilseed OR oil seed OR dairy OR livestock OR eco-agriculture).ab,de,ti.

OR

(Marine Stewardship Council OR Forestry Stewardship Council OR ISEAL OR Fairtrade OR "fair trade" OR Bonsucro OR UTZ OR "Linking Environment And Farming" OR RSPO OR Goodweave OR Alliance for water stewardship OR Aquaculture Stewardship Council OR Equitable Origin OR Responsible Jewellery Council OR Roundtable on Sustainable Biomaterials OR Union for Ethical Biotrade OR Golf Environment Organization OR Rainforest Alliance OR Roundtable on sustainable palm oil OR Good Agricultural Practice* OR Sustainability Impact Assessment OR European Sustainability Criteria OR Life Cycle Assessment OR Programme for Endorsement of Forest Certification Scheme OR PEFC

OR Indonesian Sustainable Palm Oil certification system OR ISPO OR “Coffee and Farmer Equity” OR Sustainable Agriculture Network OR Dairyman sustainability index).ab,de,ti

)AND(

((improv* OR impact* OR restor* OR develop* OR enhance*) AND (practice adoption OR continu* improvement OR legal compliance OR compliance with law OR transparency OR complaints OR grievance procedure or riparian buffer zone* OR labour OR freedom of association OR workers rights OR income OR discriminat* OR gender OR land rights OR indigenous OR set-aside OR integrated pest management OR low impact logging OR sustainable forest management OR fisheries improvement project or “customary use” OR “customary right” OR “free prior and informed consent” OR market* OR non-monetary OR smallholder* OR small-holder*)).ab,de,ti.

OR

((improv* OR enhance* OR increas* OR reduc* OR decreas*) adj4 (revenue OR livelihood* OR living standard* OR habitat OR (water adj(management OR stewardship OR conservation OR quality)) OR conservation value OR carbon stock OR biodiversity OR soil management OR soil conservation OR “employment” OR “social development*” OR “economic development” OR “social impact” OR price premium OR bycatch OR poverty OR greenhouse gas emission* OR deforest* OR fertilizer OR pesticide OR herbicide OR impact logging OR well-being OR environmental impact or rare species or threatened species or “rare, threatened and endangered” or ecosystem or “health and safety” OR personal protective equipment OR satisf* OR price)).ab,de,ti.

().ab,de,ti. = abstract, descriptor index and title

Scopus #9463 (20/10/17)

(“certif*” OR “member*”) W/3 (“aquaculture” OR “forest*” OR {forest management} OR “silvicultur*” OR “agricultur*” OR “mining” OR “farm*” OR “cultivat*” OR “fish*” OR “wood*” OR “timber*” OR “fruit*” OR “banana*” OR “textile*” OR “commodit*” OR “material*” OR “product*” OR “mineral*” OR “jewel*” OR “agroforestry*” OR “crop*” OR “coffee” OR “oilseed” OR “cocoa” OR “tea” OR {palm oil} OR {oil palm} OR “oilpalm” OR “palmoil” OR “oilseed” OR {oil seed} OR “dairy” OR “livestock” OR {eco-agriculture})

OR

({Marine Stewardship Council} OR {Forestry Stewardship Council} OR ISEAL OR Fairtrade OR {fair trade} OR Bonsucro OR UTZ OR {Linking Environment And Farming} OR RSPO OR Goodweave OR {Alliance for water stewardship} OR {Aquaculture Stewardship Council} OR {Equitable Origin} OR {Responsible Jewellery Council} OR {Roundtable on Sustainable Biomaterials} OR {Union for Ethical Biotrade} OR {Golf Environment Organization} OR {Rainforest Alliance} OR {Roundtable on sustainable palm oil} OR {Good Agricultural Practice*} OR {Sustainability Impact Assessment} OR {European Sustainability Criteria} OR {Life Cycle Assessment} OR {Programme for Endorsement of Forest Certification Scheme} OR PEFC OR {Indonesian Sustainable Palm Oil certification system} OR ISPO OR {Coffee and Farmer Equity} OR {Sustainable Agriculture Network} OR {Dairyman sustainability index})

)AND(

((improv* OR impact* OR restor* OR develop* OR enhance*) AND (“practice adoption” OR {continu* improvement} OR {legal compliance} OR {compliance with law} OR transparency OR complaints OR {grievance procedure} or {riparian buffer zone*} OR labour OR {freedom of association} OR {workers rights} OR income OR discriminat* OR gender OR {land rights} OR indigenous OR {set-aside} OR {integrated pest management} OR {low impact logging} OR {sustainable forest management} OR {fisheries improvement project} or {customary use} OR {customary right} OR {free prior and informed consent} OR market*

OR {non-monetary} OR smallholder* OR {small-holder*}))

OR

((improv* OR enhance* OR increas* OR reduc* OR decreas*) W/4 (revenue OR livelihood* OR {living standard*} OR habitat OR (water W/1(management OR stewardship OR conservation OR quality)) OR {conservation value} OR {carbon stock} OR biodiversity OR {soil management} OR {soil conservation} OR employment OR {social development*} OR {economic development} OR {social impact} OR {price premium} OR bycatch OR poverty OR {greenhouse gas emission*} OR deforest* OR fertilizer OR pesticide OR herbicide OR {impact logging} OR {well-being} OR {environmental impact} or {rare species} or {threatened species} or {rare, threatened and endangered} or ecosystem or {health and safety} OR {personal protective equipment} OR satisf* OR price))

Web of Science #6856 (20/10/17)

TS=((certif* OR member*) NEAR/3 (aquaculture OR forest* OR "forest management" OR silvicultur* OR agricultur* OR mining OR farm* OR cultivat* OR fish* OR wood* OR timber* OR fruit* OR banana* OR textile* OR commodit* OR material* OR product* OR mineral* OR jewel* OR agroforestry* OR crop* OR coffee OR oilseed OR cocoa OR tea OR "palm oil" OR "oil palm" OR oilpalm OR palmoil OR oilseed OR "oil seed" OR dairy OR livestock OR eco-agriculture))

OR

TS=("Marine Stewardship Council" OR "Forestry Stewardship Council" OR ISEAL OR Fairtrade OR "fair trade" OR Bonsucro OR UTZ OR "Linking Environment And Farming" OR RSPO OR Goodweave OR "Alliance for water stewardship" OR "Aquaculture Stewardship Council" OR "Equitable Origin" OR "Responsible Jewellery Council" OR "Roundtable on Sustainable Biomaterials" OR "Union for Ethical Biotrade" OR "Golf Environment Organization" OR "Rainforest Alliance" OR "Roundtable on sustainable palm oil" OR "Good Agricultural Practice*" OR "Sustainability Impact Assessment" OR "European Sustainability Criteria" OR "Life Cycle Assessment" OR "Programme for Endorsement of Forest Certification Scheme" OR PEFC OR "Indonesian Sustainable Palm Oil certification System" OR ISPO OR "Coffee and Farmer Equity" OR "Sustainable Agriculture Network" OR "Dairyman sustainability index")

)AND(

TS=((improv* OR impact* OR restor* OR develop* OR enhance*) AND ("practice adoption" OR "continu* improvement" OR "legal compliance" OR "compliance with law" OR transparency OR complaints OR "grievance procedure" OR "riparian buffer zone*" OR labour OR "freedom of association" OR "workers rights" OR income OR discriminat* OR gender OR "land rights" OR indigenous OR "set-aside" OR "integrated pest management" OR "low impact logging" OR "sustainable forest management" OR "fisheries improvement project" OR "customary use" OR "customary right" OR "free prior and informed consent" OR market* OR "non-monetary" OR smallholder* OR small-holder*))

OR

TS=((improv* OR enhance* OR increas* OR reduc* OR decreas*) NEAR/3 (revenue OR livelihood OR "living standard*" OR habitat OR "water management" OR "water stewardship" OR "water conservation" OR "water quality" OR "conservation value" OR "carbon stock" OR biodiversity OR "soil management" OR "soil conservation" OR "employment" OR "social development*" OR "economic development" OR "social impact" OR "price premium" OR bycatch OR poverty OR "greenhouse gas emission*" OR deforest* OR fertilizer OR pesticide OR herbicide OR "impact logging" OR well-being OR "environmental impact" OR "rare species" OR "threatened species" OR "rare, threatened and endangered" OR ecosystem OR "health and safety" OR "personal protective equipment" OR satisf* OR price))

Timespan=1990-2017, Search language=Auto

Google scholar (23/10/17) #1000

Any of the words: practice OR adoption OR change OR revenue OR change OR increase OR decrease

Must include the phrase: Sustainability standard

1990-present, search carried out using Harzing's publish or perish 6.15.5992.6503

Appendix A1.4 Reference papers for search comprehensiveness

BECCHETTI, L. & COSTANTINO, M. 2008. The effects of fair trade on affiliated producers: an impact analysis on Kenyan farmers. World Development (Oxford), 36, 823-842.

BOLWIG, S., GIBBON, P. & JONES, S. 2009. The economics of smallholder organic contract farming in tropical Africa. World Development (Oxford), 37, 1094-1104.

CERUTTI, P. O., LESCUYER, G., TSANGA, R., KASSA, S. N., MAPANGOU, P. R., MENDOULA, E. E., MISSAMBA-LOLA, A. P., NASI, R., ECKEBIL, P. P. T. & YEMBE, R. Y. 2014. Social impacts of the Forest Stewardship Council certification: an assessment in the Congo basin. CIFOR Occasional Paper, 58-pp.

CHIPUTWA, B., SPIELMAN, D. J. & QAIM, M. 2015. Food standards, certification, and poverty among coffee farmers in Uganda. World Development (Oxford), 66, 400-412.

JENA, P. R., CHICHAIBELU, B. B., STELLMACHER, T. & GROTE, U. 2012. The impact of coffee certification on small-scale producers' livelihoods: a case study from the Jimma Zone, Ethiopia. Agricultural Economics, 43, 429-440.

KALONGA, S. K. & KULINDWA, K. A. 2017. Does forest certification enhance livelihood conditions? Empirical evidence from forest management in Kilwa District, Tanzania. Forest Policy and Economics, 74, 49-61.

NORMAN, S. & WARMARK, C. 2003. Farm sustainability in the semi-arid tropics - a case study of small farms in Nicaragua using sustainability indicators. Minor Field Studies - International Office, Swedish University of Agricultural Sciences, 93-pp.

QIAO, Y., HALBERG, N., SAMINATHAN, V. & 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 Agriculture and Food Systems, 31, 246-257.

RIJSBERGEN, B. V., ELBERS, W., RUBEN, R. & NJUGUNA, S. N. 2016. The ambivalent impact of coffee certification on farmers' welfare: a matched panel approach for cooperatives in Central Kenya. World Development (Oxford), 77, 277-292.

RUBEN, R. & FORT, R. 2012. The impact of Fair Trade certification for coffee farmers in Peru. World Development (Oxford), 40, 570-582.

RUBEN, R. & ZUNIGA, G. 2011. How standards compete: comparative impact of coffee certification schemes in Northern Nicaragua. Supply Chain Management: An International Journal, 16, 98-109.

RUEDA, X. & LAMBIN, E. F. 2013. Responding to globalization: impacts of certification on Colombian small-scale coffee growers. Ecology and Society, 18, Art-21.

RUEDA, X., THOMAS, N. E. & LAMBIN, E. F. 2015. Eco-certification and coffee cultivation enhance tree cover and forest connectivity in the Colombian coffee landscapes. Regional Environmental Change, 15, 25-33.

STEMLE, A., UCHIDA, H. & ROHEIM, C. A. 2016. Have dockside prices improved after MSC certification? Analysis of multiple fisheries. Special Issue: Fisheries certification and eco-labeling: benefits, challenges and solutions., 182, 116-123.

TAKAHASHI, R. & TODO, Y. 2013. The impact of a shade coffee certification program on forest conservation: a case study from a wild coffee forest in Ethiopia. Journal of

Environmental Management, 130, 48-54.

UTTING-CHAMORRO, K. 2005. Does fair trade make a difference? The case of small coffee producers in Nicaragua. Special issue: Do more good, do less harm: development and the private sector., 15, 584-599.

WAKAMATSU, H. 2014. The impact of MSC certification on a Japanese certified fishery. Marine Resource Economics, 29, 55-67.

Appendix A1.5

Main differences between systematic reviews and systematic maps, adapted from James et al. (2016)

Stage in 'evidence synthesis'	Systematic map	Systematic review
Objective	Describes the state of knowledge for a question or topic	Aims to answer questions with a quantitative or qualitative answer
Question formulation	Question can be open-framed or closed-framed. Topic can be broad or narrow	Question is usually closed-framed
Search strategy	No limitation on research evidence that can be included (e.g. primary and secondary research)	Evidence is limited to primary qualitative or quantitative research. For example, comparative, prevalence or occurrence type studies
Article screening	Articles not obtainable at full text (where the full document is not available) or studies with limited data may be included	Evidence is limited to primary qualitative or quantitative research. For example, comparative, prevalence or occurrence type studies
Data coding and data extraction	Information describing the study and its methods are coded. Study results may not be extracted	Information describing the study and its methods and studies' qualitative and or quantitative results coded
Critical appraisal	Critical appraisal optional	All included studies critically appraised for study internal and external validity
Synthesis	Trends in the literature, knowledge gaps and clusters identified but no 'synthesis of study results' carried out	Qualitative or quantitative synthesis of study results where possible using appropriate methodology (e.g. meta-analysis). Knowledge gaps identified
Report	Describes and catalogues available evidence relating to a topic of interest, identifying knowledge gaps and knowledge clusters. Implications for policy, practice and research made	Narrative and qualitative or quantitative synthesis study results (e.g. meta-analysis) to answer the question where feasible. Implications for policy and practice, and identification of knowledge gaps for future research

Appendix 2: Methods for analysing data from sustainability standard systems

Background and Purpose

The aim of this research study is to understand the effectiveness of standards and certification tools in driving the adoption of more sustainable practices in certified entities over time. One of the inputs used to answer this research question is internal data that some ISEAL members generously made available for the research.

The main research question, can only be answered through research that compares practice adoption of entities that are (or are about to become) certified with those that are not (won't become) certified. For this reason, the main focus of the report is on the systematic mapping, which provides a rigorous assessment of existing research. This body of knowledge should tell us what we know about the extent to which sustainable practices are adopted by entities that become certified, and the degree to which this can be attributed to the interaction between certified entities and certification systems. However, there will inevitably be gaps in this knowledge: sectors with little or no rigorous research, technical areas and geographies that have not been well researched. It is also likely that other important questions remain incompletely answered by the formal literature, such as why and in what circumstances certification systems drive practice adoption? For these reasons, it is important to complement the systematic map with other forms of information and analysis.

To this end, we consider that ISEAL Members' data can play an important role in complementing the systematic map. The data can inform questions around what types of practices are adopted, how commonly they might be adopted, and in which areas of sustainability. It can potentially provide a longer time horizon than the typical 'snapshot' of most published research. What analysis of this data cannot do is provide attribution that changes in practice are a result of certification, because the internal 'scheme' data is not, for the most part, collected within rigorous, controlled experimental or quasi-experimental designs. For example, if data from audits is used to assess changes in practice, a putative change in practice could be the result of the certified entity indeed doing things differently, or a change in the approach use by auditors, or a change in sectoral practices (e.g., a result of changes in the law) that are nothing to do with the certification system. Analysis of this data is illustrative, and should not be interpreted as implying attribution. However, it does put an upper boundary on the extent of practice adoption: i.e., if auditing and sectoral practices did not change, this is the maximum impact that certification has on driving practice adoption.

Types of data available and relevance

There is no standard process of data collection and data management amongst certification systems. The following types of data seem to be the most common amongst twelve ISEAL Members contacted:

1. **Reach.** All schemes have data – typically publicly available – that describes core attributes of the scheme: how many entities are certified, in how many countries, the number of hectares certified, the market share of certified products, etc.
2. **Case studies.** Many schemes have invested in collecting case studies of how certified entities have changed. These are essentially anecdotal in nature, but nonetheless

important for communications, and can often illustrate the motivations behind, and barriers to changing practice.

3. **Self-declarations.** Some schemes have a process by which certified entities declare their progress on the standard, in addition to compliance with the standard being assessed by independent auditors.
4. **Non-conformities.** Non-conformities are decided by auditors (assessors) when they deem that there is insufficient evidence to demonstrate that the certified entity (or entity seeking certification) is not in compliance with part of the standard. Each non-compliance is explicitly linked to part of the standard (a criterion, an indicator, or equivalent) and so is traceable to a particular area of practice. In some schemes, non-conformities are ranked 'major' or 'minor'. Major non-conformities have to be rectified before the organisation can gain a certificate, whereas certified entities typically have a set time period to rectify with minor non-conformities.
5. **Performance data.** Some schemes collect data on outcomes as part of the compliance process and/or to support learning. Note that other schemes commission independent researchers to conduct impact research on specific topics, geographies, or types of certified entity, and this body of evidence is included within the systematic map as appropriate.

Table A2.1:

Schemes providing data to the study

Scheme	Sector	Type of certified entity	Type of data	Period/area covered
Fairtrade	General agricultural	Developing country and smallholder focus	Non-conformity	2015-16
RSB	Crops used as biomass/biofuel feedstock	Producers, processors and traders	Non-conformity	2012-17
RSPO	Palm oil	Mostly large plantations	Non-conformity	2008-17
LEAF	General agricultural	Farms	Self-declared	2009-16
MSC	Marine fish	Fisheries	Non-conformity	2000-16
BCI	Cotton	Mostly smallholders, developing country focus	Performance with comparison	2014-16

The same type of data was not available from all of the schemes contacted, and in addition the different types of data are not all as useful for understanding practice adoption. Therefore, a decision was made to work with data from the final three categories: self-declarations, non-conformities and outcome data. This allowed a broad range of schemes, sectors and types of practice to be included. Six schemes agreed to make their data available: Fairtrade, the Roundtable on Sustainable Biomaterials, the Roundtable on Sustainable Palm Oil, LEAF Marque, the Marine Stewardship Council, and the Better Cotton Initiative (Table A2.1).

Each of these types of data has advantages and disadvantages concerning analysis of practice adoptions. Self-declared data can in theory show a pattern of change over time. It is generated by those with the deepest knowledge of sustainability in the specific context in which it is being practiced, and because it is relatively quick and cheap to collect it can also be comprehensive. The disadvantage is that it is likely to suffer from bias, with some certified entities being over-optimistic, whilst others are overly pessimistic about their achievements. The data may be of lower quality as there is no independent check on it.

Non-conformity data is potentially useful information for understanding practice adoption because there is typically a set time for a certified entity to change its practice to become consistent with the standard. Therefore, if a non-conformity is raised and then closed it can provide information that a practice that was not previously compliant with a standard now is. However, it is well established that different auditors (and certification bodies) do not raise non-conformities in an identical way. Some auditors are stricter than others, and not all audit teams have deep knowledge of all of the areas covered by the standard.

Some performance (or outcome) data may not be relevant to questions of practice adoption. For example, data on income is information on an outcome, not a practice. However, other measured outcomes are strongly aligned to practice adoption. For example, changes in the quantity and types of pesticides and fertilisers used over time indicates a change in practice, which may have numerous impacts (changes in yield, biodiversity, etc). Where relevant, performance data can be amongst the most rigorous available, as it often includes collecting data from analogous but not certified entities (i.e. controls).

Finally, the fact that schemes provided data sets of different lengths and for varying numbers of certified entities, means that not all schemes contributed equally to evidence. The findings are an aggregate, and not therefore necessarily fully applicable to any of the individual schemes that provided data.

Overall analytical approach

The data provided had at least the following characteristics in common: a unique identifier for each certified entity; a time series of change data (non-conformity, self-declaration of progress or performance) for each entity; and a descriptor of the change (e.g., a criterion in the standard against which the data had been recorded).

The first step was to make the data comparable. This required two actions. Firstly, the precise change being described under each scheme was different, because the wording and intent of the standards that underlie each scheme vary to a lesser or greater extent. To make data comparable therefore required aggregating data into meaningful thematic groups. For example, one scheme may require that the habitat of rare species is maintained or enhanced whereas another scheme might ask for an implemented nature conservation plan. These specifics can be meaningfully aggregated to ‘conservation’ practices even though the precise wording of the standard, and the context of the change – marine or terrestrial, developing country or developed – is different. The Thematic Groups were selected to be coherent (where possible) with the outcomes recorded in the systematic map, and were only included if at least two schemes provided data to the Thematic Group. The Groups (Table A2.2) included almost all of the available data. A small number of changes were excluded when a thematic area was specific to one scheme (e.g., the RSPO has a number of criteria concerning the use of fire in land preparation that do not have analogous criteria in the other standards). This was used in the analysis of non-conformity data (see next section).

Table A2.2:

Thematic groups used to aggregate data, with examples

Scheme	Schemes included	Examples
Labour	LEAF, Fairtrade, RSB, RSPO, BCI	• RSPO Criterion 6.5: Pay and conditions for employees and for contract workers always meet at least legal or industry minimum standards and are sufficient to provide decent living wages.
Health and safety	LEAF, Fairtrade, RSB, RSPO	• RSB Criterion 4f: Conditions of occupational safety and health for workers shall follow internationally recognized standards • LEAF Q 3.7.1: We ensure that steps are taken to protect operator, human, wildlife and pet safety
Community benefits & development	LEAF, Fairtrade, RSB, RSPO	• RSB Criterion 4f: Conditions of occupational safety and health for workers shall follow internationally recognized standards • LEAF Q 3.7.1: We ensure that steps are taken to protect operator, human, wildlife and pet safety
Conservation	MSC, LEAF, Fairtrade, RSB, RSPO	• MSC PI 2.3.1 The UoA meets national and international requirements for protection of ETP species • RSPO Indicator 5.2.2: Where rare, threatened or endangered (RTE) species, or HCVs, are present or are affected by plantation or mill operations, appropriate measures that are expected to maintain and/or enhance them shall be implemented through a management plan.
Input use	LEAF, Fairtrade, RSB, RSPO, BCI	• BCI: Toxic load indicator score per hectare
Energy use & GHG emissions	LEAF, Fairtrade, RSB, RSPO	• RSPO Criterion 5.6: Plans to reduce pollution and emissions, including greenhouse gases, are developed, implemented and monitored
Waste and Recycling	LEAF, Fairtrade, RSB, RSPO	• LEAF Q 4.3.3 We recycle or dispose of our plastics in compliance with the law and best practice
Soil Management	LEAF, Fairtrade, RSB, RSPO	• LEAF Q2.2.2: We adopt a general policy to conserve and build up soil organic matter
Water Use	LEAF, Fairtrade, RSB, RSPO	• RSPO Criterion 4.4 Practices maintain the quality and availability of surface and ground water
Good Production Practices	MCS, LEAF, RSB, Fairtrade, RSPO	• MSC PI 1.1.1 Stock status • Crop rotations are an essential part of our cropping decisions. (Q 3.2.1, LEAF)
Management Systems	MCS, LEAF, RSB, Fairtrade, RSPO	• RSPO Criterion 2.1 There is compliance with all applicable local, national and ratified international laws and regulations.
Chemical Handling and Storage	LEAF, Fairtrade, RSB, RSPO	• LEAF Q2.6.9: Our fertiliser store is safely sited and properly maintained to current best practice • RSB Criterion 11 d 5. The handling, storage and disposal of pesticides shall comply with the FAO's Guidelines on Good Practices for Ground and Aerial Applications of Pesticides

Secondly, because the data included records of non-conformities, self-declared scales of progress and performance metrics, it had to be transformed into a common scale that allows comparison. This entailed coding the change on a three-point scale: practice adoption, no change or unresolved (i.e., the situation where it is not clear whether the practice of the certified entity has regressed or will return to a previous state). A description of the coding protocol is given in Table A2.3). This data was used to analyse practise adoption more explicitly (see next section).

Table A2.3:

Protocol for coding practice adoption from certification system data

Change	Scheme	Coding protocol
Practice adoption	MSC, RSPO, RSB, Fairtrade	A non-conformity is raised on the initial audit with subsequent confirmation that it has been addressed
	LEAF Marque	Progress against each criterion is self-declared on a four-point scale from 'not started' to ''fully achieved', with practice adoption recorded as a change up that scale between reporting years
	BCI	A decrease in the metric or smaller increase in total pesticide use per hectare, TLI score per hectare, water use per hectare and total synthetic fertiliser use per hectare than comparison (non-certified) farmers.
Maintenance of performance	MSC, RSPO, RSB, Fairtrade	A non-conformity is raised on a subsequent audit and closed
	LEAF Marque	No change on the scale of 'not started' to ''fully achieved' between reporting years
	BCI	Not scored
Unresolved	MSC, RSPO, RSB, Fairtrade	A non-conformity is raised at any time but not closed
	LEAF Marque	A downward movement of the scale 'not started' to 'fully achieved' between reporting years
	BCI	An increase or smaller decrease in total pesticide use per hectare, TLI score per hectare, water use per hectare and total synthetic fertiliser use per hectare than comparison (non-certified) farmers.

1 The data kindly provided by the LEAF Marque scheme was answers provided by farmers to annual online questionnaires. The questions changed in significantly after 2014, and so to create a time series only those questions that dealt with the same underlying issue were used in the analysis (61 questions in total).

Data on non-conformities and self-declared answers to questionnaires can vary widely in the degree of 'on the ground' change or action they imply. Some describe substantive activities that clearly demonstrate that things are being done differently (e.g., reducing GHG emissions, change in use of agricultural inputs, biodiversity conservation or implementation of labour rights). Others address procedural changes that may or may not lead to on the ground change (e.g., 'adequate calculation of GHG emissions' or 'keeping financial records'). As this study is about practice adoption, it is the former that is of interest. Each coded change was then categorised as 'substantive' or 'procedural', which

is recommended good practice for using non-conformity data in particular. This reduced the overall number of questions within each of the thematic groups listed above, but did not eliminate any thematic group.

Finally, the data was arrayed in time series, with the year of the first declaration, audit or measurement coded as year zero, the next as year 1 and so on. This provides a way of comparing entities from different schemes with relation to what is likely to be one of the main trends in the data: the influence of certification over time (i.e., on average the largest changes are likely to be at the beginning, and once an entity has 'learned' how to achieve certification there are likely to be fewer changes in its practice with respect to the standard. Certified entities with just a single data point were omitted, as they do not allow change to be estimated.

Analysis

The data is analysed in two ways. Firstly, the non-conformity data was analysed to show the thematic and temporal pattern of non-conformities within each thematic group. This is a straightforward metric, and one which has been used in a number of studies. The underlying rationale is that a decrease in the number of non-compliances over time indicates that the practice of certified entities is becoming increasingly aligned to the standard. However, this data is essentially about practices not in compliance with a standard rather than about practice adoption per se, it does take into account non-conformities that are not acted upon, or instances when a certified entity 'flip-flops' between compliance and non-compliance on a particular thematic area (which arguably indicates that practice adoption has been superficial).

In addition, the auditing procedures that raise non-conformities have inherent biases, and it is well established that some non-conformities are far easier to detect than others. For example, non-conformities in areas such as workplace discrimination, harassment, and freedom of association are far harder to detect than non-conformities on health and safety or the maintenance of riparian zones. That means that some areas of sustainability practice are likely to be systematically under-represented in non-conformity data. Finally, changes to practice brought about as a result of non-conformities being raised have been shown in some systems to be a small proportion of all the changes in practice achieved.

The second approach is described in Table A2.3, above, and provides an explicit metric for practice adoption, maintenance of performance, and unresolved situations. With this metric, only positive changes – closing a non-conformity, improving performance or self-declared progress – with respect to the starting conditions are considered as practice adoption. This analytical definition partly addresses the issue of changing auditor or reporter behavior in generating apparent changes in practice, because only non-conformities raised on the initial audit count towards practice adoption.

Under this definition, a negative change reported after the initial certification is considered a regression in practice, and therefore improvement is return to previous practice – maintenance of performance - not practice adoption. This information is however important, as it potentially sheds light on the role that certification plays in supporting entities to maintain sustainable performance over time, as the context changes around them.

Where there is no evidence that practice has returned after a negative change, either because the entity lost its certificate or because the negative change was in the final year and hence there is no data to show that performance has been re-established, then this is considered 'unresolved'. This final category therefore includes incidences of maintenance of performance that are still in progress as well as genuine regressions in practice. All of the data types are used in this analysis, which is reported per thematic group.

Overview of final data

The total number of certified entities included in the analysis was 1,982. These came from a total of 77 countries. The UK had the highest number of certified entities, followed by India, Malaysia, Indonesia and Pakistan, with these five countries accounting for 55% of the total. The included certified entities had between one year (i.e., two data points) and 14 years of data.

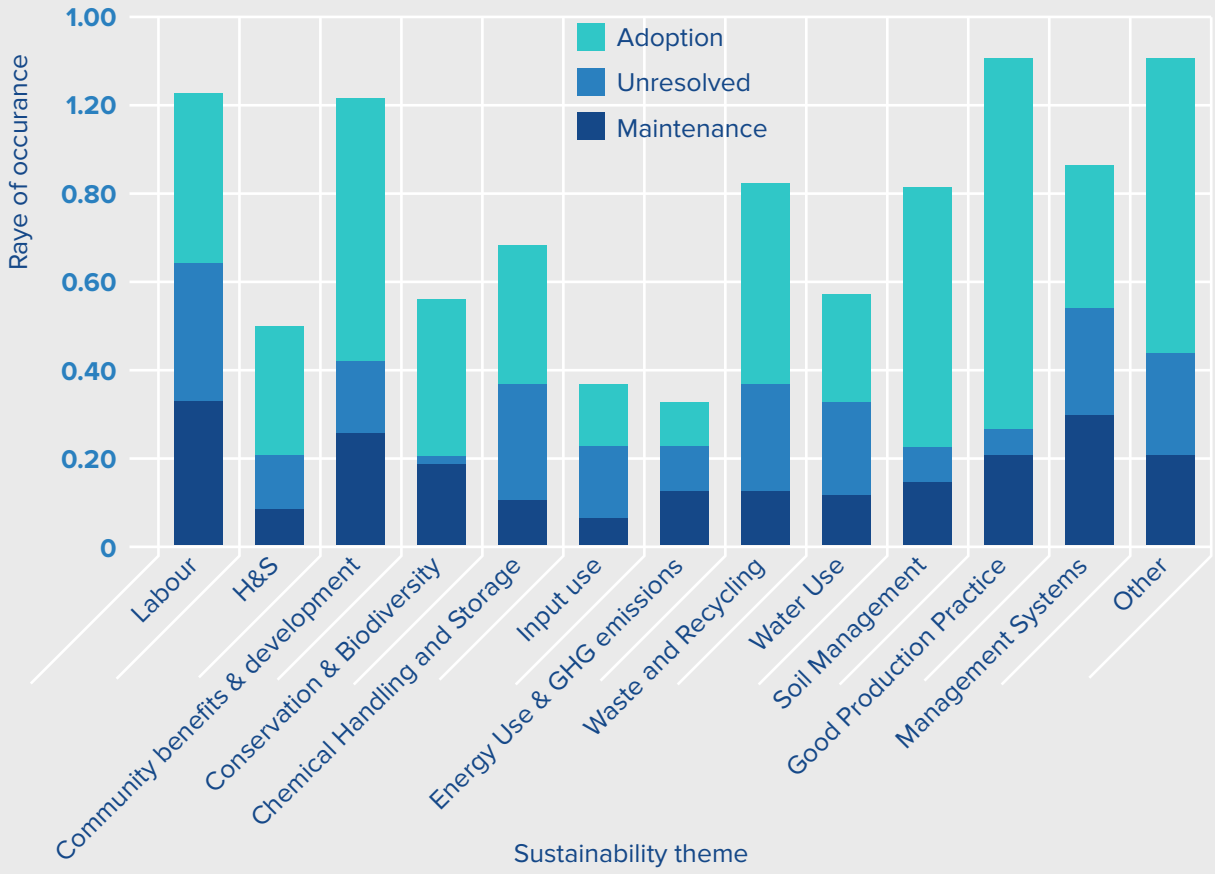
Figure A2.1 shows the rates of practice adoption, practice maintenance and unresolved practice for all of the sustainability themes assessed.

Maintenance is by far the commonest outcome, accounting for 58% occurrences of practice change. It is more common than adoption for all themes excepting energy use and greenhouse gas emissions. The highest rates were found in Good Production Practices, Community Benefits and Development, and Soil Management, each of which record around 60-80 incidences per one hundred certified entities. As this category includes incidences of recorded slippage in practices followed by recovery, and when practice has been not changed from the starting state, this suggests that practices are relatively stable, and that one of the chief roles of certification may be to ensure that practice does not slip over time.

The highest rates of practice adoption are found in Labour, Management Systems and Community Benefits and Development, each of which have over 25 incidences per one

Figure A2.1:

Comparison of the rates of practice adoption, practice maintenance and unresolved practice amongst sustainability themes



hundred certified entities. Adoption is more common than unresolved practice change in seven of the thirteen themes. This final category includes slippages in practice that have not yet been restored, and some that may never be if the entity leaves certification.

Considering non-conformities alone, there is a logarithmic decline in the average number of non-conformities raised over time (Figure A2.2). However, the number of certified entities in the data set also declines rapidly over time (i.e., very few have been certified for more than three or four years), and so when the average number of non-compliances per certified entity is plotted, there is no apparent trend over time (Figure A2.3).

Figure A2.2:

The number of substantive environmental, social and economic non-conformities over time

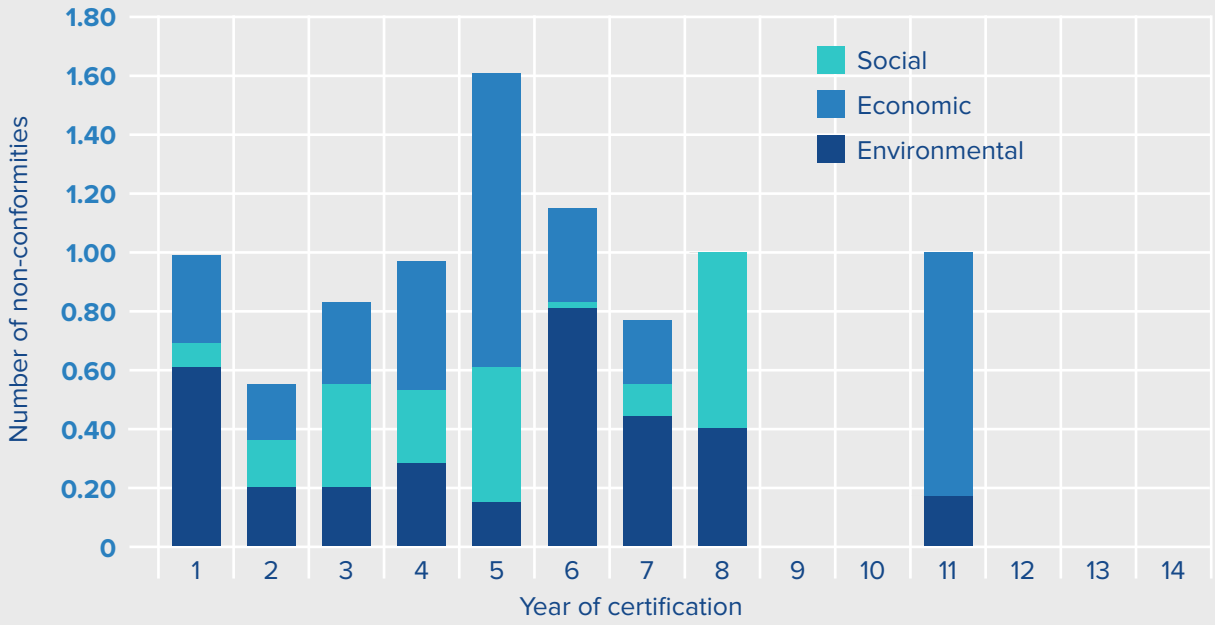
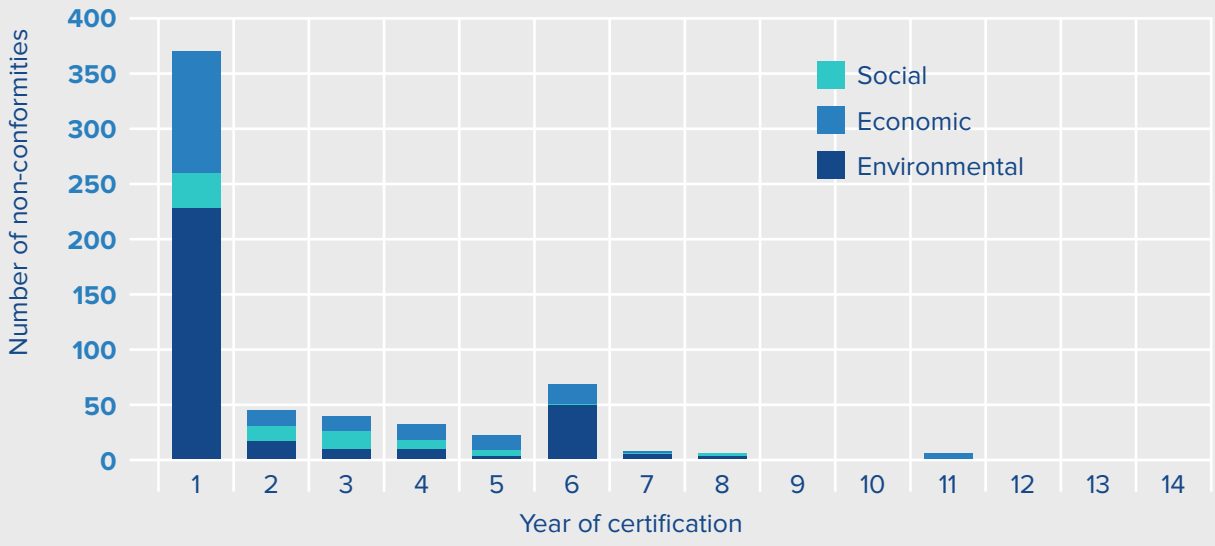


Figure A2.3:

The average number of substantive non-conformities per certified entity over time



About 3Keel

3Keel is an Oxford-based firm of sustainability advisors specialised in working with food systems, supply chains and landscapes.

The company was formed in 2013 by Tom Curtis, Simon Miller and Richard Sheane. The name 3Keel comes from the keel of a boat – giving it stability in rough seas – and this is how we see our work helping to give organisations resilience in changing times. It may also be revealing of one of the partners' love of nautical pursuits!

In 2014 3Keel changed from a limited company to become a partnership with the arrival of Steve Jennings and Will Schreiber, boosting our expertise in international development and corporate sustainability. Our complementary skills and experience allow for the cross-fertilisation of ideas across the company, making us truly integrated thinkers who enjoy advising on complex issues.

1 Standingford House
26 Cave Street
Oxford
OX4 1BA
United Kingdom

+44 (0)1865 236500
office@3keel.com

Funded with UK aid from the UK government.