Target Setting and Burden Sharing in Sustainability Assessment beyond the Farm Level

Jay Whitehead[1], Yuan Lu[2], Holly Still[2], Jonathan Wallis[2], Hannah Gentle[3], Henrik Moller[2,3]

[1] Agribusiness & Economic Research Unit, Lincoln University, Lincoln 7647, New Zealand
[2] Centre for Sustainability, University of Otago, PO Box 56, Dunedin 9054, New Zealand
[3] Ecosystems Consultants Ltd., 30 Warden St., Dunedin 9010, New Zealand

Keywords: Target setting and benchmarking, Fairness, Responsibility, TempAg

Abstract

While great progress has been made towards monitoring agricultural sustainability through the use of indicators, setting sustainability indicator targets that motivate the transformation of farming systems for sustainability and resilience is often overlooked. This paper examines the role of target setting and benchmarking comparisons in sustainability assessment. A review of 186 indicator metrics and their targets from 12 sustainability assessment frameworks showed a preponderance of practice-based rather than performance-based measures. Many targets were implicit and embedded within the way ratings or standards were measured rather than explicitly derived from external information or processes. Ratio scales were rarely used for indicator measurement. Given these limitations, most assessment frameworks are weak tools for the comparison of agricultural sustainability between sectors, regions or nations. We then considered the equity implications of sustainability burden and benefit sharing and drew lessons from recent international climate change negotiations to recommend guidelines when erecting production level sustainability targets and benchmark comparisons between farms, regions, sectors and countries in the way being considered by the TempAg network. Equitable participation by multiple stakeholders in the process of erecting targets is important to achieve fair outcomes that underpin lasting commitment to sustainability. Scrupulous application of equity and fairness is more likely to change values of the farming families, food processors and distributors and consumers for collective action. Adjusting targets to match local social, economic and ecological constraints on farming performance may be fairer, but this local tuning also challenges the design of and use of targets and benchmarks that have been upscaled to regional and national levels for informing sustainability policies across temperate agriculture as a whole. So will TempAg targets and benchmarking help or hinder transformation for sustainability and resilience?

1. Introduction

Food production is required to grow substantially if it is to meet global demand of nine billion people by 2050 (OECD-FAO, 2010). Achieving this increase in food production to supply global markets, while meeting consumer and citizen expectations, and maintaining biodiversity and ecosystem services is an important challenge. Agricultural systems will need to improve in all parts of the world in order to ensure the long term sustainability of food, fibre and biofuel production (Pretty et al., 2008). The search for practical solutions to enhancing sustainability requires both a farm-level focus complemented with a view towards developing appropriate social and economic policies at regional, national and international levels. The 'Collaborative Research Network on Sustainable Temperate Agriculture' ('TempAg'), is one recent and potentially effective initiative to help co-ordinate sustainability interventions at multiple spatial and temporal scales (Gregory 2016, this symposium). It is a coalition of agricultural researchers and policy makers from across temperate and high altitude production areas of the OECD that amongst much else, aims to *identify and critically evaluate the available tools and guidelines to assess sustainability and transform agricultural production systems across multiple scales.* The TempAg team therefore proposed elements of workshops 2.2 (farm level) and 2.4 (beyond-farm level) sustainability assessment tools as a valuable part of this IFSA 2016 symposium.

Agricultural performance improvements will be accelerated by erecting sustainability ‘reference values’ such as performance targets, critical thresholds, minimum standards and benchmarks. Without targets or benchmarks, measures of sustainability indicators provide little opportunity for risk management by decision makers like farmers, processors and distributors, marketers, policy analysts and government. Indeed, without reference values, sustainability assessment is in danger of being seen as measurement for measurement’s own sake, and farmers are less likely to see the exercise as relevant and an opportunity for themselves, rather than a cost and threat imposed by ‘outsiders’. This paper is the first of a series from the New Zealand Sustainability Dashboard project (Benge et al. 2016, this symposium) to focus on how reference values are currently designed and used. We first briefly present an overview of reference value definitions and structure as deployed across 12 sustainability frameworks that currently operate in very different contexts around the world. We then go on to consider how reference values may actually be used for encouraging change and apportioning responsibilities at two levels: first we present a review of the international climate change mitigation agreements as a potential example of the way the tools being tested by TempAg might be used by an international organisation such as OECD or FAO to meet a collective target for agricultural sustainability; and second, a hypothetical example how those principles from climate change mitigation might be applied at a local collective industry for both benchmarking and target setting for transforming production.

2. Reference Setting and Benchmarking in Sustainability Assessment

2.1 Types of reference values

We reviewed the wider sustainability assessment literature and then selected a stratified random selection of up to 20 indicators across four pillars (economic, social, environment, governance) for each of 12 sustainability frameworks currently in operation around the world: Sustainability Assessment of Food and Agriculture systems (SAFA); The Sustainability Consortium (TSC); GLOBAL.G.A.P.; LEAF Marque Standard (Linking Environment and Farming); International Federation of Organic Agriculture Movements (IFOAM) Standard; BioGro Organic Standards; Response-Inducing Sustainability Evaluation (RISE) 3.0; BioBio (Farmland biodiversity indicators); OECD Agri-environmental indicators; Mauri Model; The Sustainable Agriculture Network - Sustainable Agriculture Standard (SAS); Sustainability Monitoring and Assessment RouTine (SMART). Altogether we reviewed 186 indicators, scored them for the presence or absence of reference values, and devised a typology of how indicators and reference values were constructed.

Our extensive literature search identified only a few papers that attempted a critical overview of agricultural sustainability reference setting structure ((van der Heide et al., 2007; Van Cauwenbergh et al., 2007; Bastian et al., 2007; Acosta-Alba and van der Werf, 2011). Considerable confusion arises from conflicting definitions of targets and benchmarking, so our first plea is for sustainability assessors to converge on a standardised set of definitions.

Out of the 186 indicators randomly selected from the 12 sustainability frameworks, the majority (96%) had a ‘target’, either in the form of a standard (a minimum standard required; 47%), or rating (a measure or evaluation of performance; 49%). The proportion of indicators with targets (96%) was higher than previously anticipated because ‘embedded’ targets, or implicit expectations, were categorised as a ‘target’ rather than the absence of a reference value. For example, a rating of farm management on a 5-point/category ordinal scale ranging from ‘poor to ‘excellent’ implicitly suggests that an ‘excellent’ rating could be achieved by all farmers. The ‘excellent’ rating is the implicit or embedded target. Most targets (93%) were implicit, while only 7% of targets were independent of the way the rating scale was measured and instead derived from external information on acceptable or optimal performance. Our second plea is for sustainability assessors to motivate improvement by making targets more explicit and direct.

The majority (65%) of the targets were “practice-based” (assessing adherence to specific best farming practices), whereas 31% were “performance-based” (monitoring farming outputs). The remaining 4% of targets were a mixture of practice and performance-based. We found that most practise-based ratings were loosely defined and deployed statistically weak metrics for trend analysis. More fundamentally, they make an overarching assumption that improved sustainability (of some unstated amount) will emerge if a given practice is in place (e.g. soil health is monitored in some way). Our third plea is that more performance-based rating systems are deployed for improved assessment and learning.

The majority (58%) of targets were simple binaries (usually the presence or absence of a desired practice). Some (24%) used semi-quantitative ordinal scales, and only 18% deployed a measurable target using a ratio scale for measurement. Many of the latter were “secondarily derived” i.e. aggregations at an industry or product level to calculate the percentage of producers or suppliers that achieved some binary performance or practice criterion at the individual farm level. Binary and ordinal scales have several well-recognised limitations of scale depression; low sensitivity for measuring change and limitations of how they can be combined for upscaling and aggregation of indicators and targets. There were several examples where the subsequent manipulations and interpretations of binary and ordinal scales violated fundamental properties of measurement scales and statistics. Our fourth plea is that true ratio scales of measurement are used for indicators and targets at the farm level and not just in secondary aggregations of the data beyond the farm scale.

2.2 Targets are not always needed: Internal benchmarking for encouraging improvement

Many of the applications of sustainability measures appear to be designed for internal comparison of relative performance between farms now (spatial comparison), or changes in their own performance with past years (temporal comparisons). Provided that the metric has been scored in a relatively consistent manner, continuous improvement can result by the comparison a farmer sees with their neighbours or at least other producers facing the same or similar constraints. In this way, those signalled to be in the bottom quartile of performers may be motivated to improve and climb past their colleagues next season or as they develop their systems. This in turn will potentially trigger renewed efforts of the previous leaders. The underlying model is one of an “improvement escalator” where the overall average performance will climb when farmers compete with each other and become aware that it is indeed possible to improve. In this model, benchmarking is a type of passive incentivisation tool that requires no particular target or plan. It has the advantage of local relevance and naturally fits with the way farmers often monitor their own performance by comparing with their neighbours. Benchmarking oneself against earlier performance is a temporal version of the same internal benchmarking tool.

It would be possible to set targets for rates of improvement, but we found these to be relatively rare.

3. Equity when setting targets: Lessons from Climate Change Negotiations

Specific, quantitative, time-bound targets can be linked to indicators so that performance can be interpreted clearly on a 'distance-to-target' basis (Moldan et al., 2012). Target setting for sustainability assessment requires two distinct, yet closely interconnected steps. The first is to define the target either quantitatively or qualitatively, while the second is to assign responsibility for meeting the target. Figge (2005) argues that society defines in political processes the 'goalposts' of sustainable development, and suggests that it is these targets that parties need to meet. Voluntary sustainability assessment initiatives are becoming an increasingly common way to address sustainability concerns. Regardless of what targets are set, participants to a voluntary sustainability initiative are unlikely to willingly adopt a sustainability performance target unless they perceive it to be fair. A balance is required between the overall target that is expected to be met, and the fairness of each party's obligations for meeting the target. Equity concerns play an important role in the establishment of sustainability performance targets because acceptability of collective responsibility will be enhanced if the target is perceived as fair amongst those participants expected to enact it.

International climate change negotiations and the associated literature have devoted significant attention to both the setting of a performance target, and the assigning of responsibilities for achieving that target amongst nations. A common determination that global warming should be limited to 2°C is a clear performance target that has been developed largely through scientific research. The division of obligations amongst multiple disparate nations for meeting this target however, involves ongoing ethical and political debates around equity and distributive justice ideals, as each nation expresses its own vision of fairness (Lange et al., 2010). The concepts of equity, justice, or fairness, here used interchangeably, have been central to discussions on sustainability since its inception (Pearce, 1987). The normative foundations for equity concerns are based on philosophical and moral theories of distributive justice (Pearce, 1987). Distributive justice is concerned with fairness in outcomes.

Providing equal opportunities and need satisfaction for people is a central feature of the concept of sustainable development (Langhelle, 2000). Multiple international treaties and agreements intended to confront complex and interconnected issues, like those presented by sustainability, address equity through the concept of 'Common but Differentiated Responsibility' and 'Respective Capabilities' (CBDR & RC) (UN, 2015).

The Rio Declaration is provides one of the clearest enunciations of CBDR & RC in an international agreement:

> In view of the different contributions to global environmental degradation, States have common but differentiated responsibilities. The developed countries acknowledge the responsibility that they bear in the international pursuit of sustainable development in view of the pressures their societies place on the global environment and of the technologies and financial resources they command (UN, 1992 Principle 7).

Interconnected ecological networks and economic interdependence between countries mean that each country can be subject to the environmental and consumption choices of others. Yet, each country alone does not have the capability to address these issues which require co-operation thereby, promoting the idea of 'common' responsibility (Rajamani, 2000). At the same time as global issues require global co-operation, it has been recognised that the differences in country's capabilities, technology, historic responsibility, and needs (amongst other factors) mean that all countries do not have an equal opportunity to address global issues, and therefore, their responsibilities to act should be 'differentiated'. Both the United Nations Framework Convention on Climate Change, and the Kyoto Protocol were explicitly based on the concept of CBDR, which continues to play a key role in the post-2012 climate change negotiations.

Perceptions of fairness in the allocation of the burdens associated with sustainable development can influence the viability of a sustainability proposal both at an international level, and at an individual level (Carlsson et al., 2011). At the centre of climate change negotiations is a debate on the equity and fairness implications of the burdens imposed by emissions mitigation (Dannenberg et al., 2010). While climate change mitigation is a direct concern of agricultural sustainability, the broader framework of international climate change negotiations can also provide guidance on the likely challenges that could occur in setting agricultural sustainability performance targets for a wide and diverse group of participants.

Lange (2010) demonstrates that equity considerations underpin many of the differences between country's interpretations of which path to climate change mitigation is optimal. Countries are likely to only accept treaties with international obligations if they are perceived as fair (Stalley, 2013). This same fairness requirement underlies the acceptance of other, lower level, sustainability initiatives. What is deemed to be fair rests upon the weight each entity puts on different distributive justice principles.

4 Constructing Burden Sharing Criteria

Distributive justice principles that provide grounds for a departure from absolute equality have been discussed extensively in both philosophy and welfare economics (Yaari & Bar-Hillel, 1984). The search for appropriate equity principles on which departures from equality can be justified has experienced a resurgence of interest in recent years due to the global challenge of climate change, and the associated international negotiations (Carlsson et al., 2011; Lange et al., 2010). Despite a vast array of fairness principles being described by the literature, there is a considerable convergence on three basic principles of distributive justice (Underdal & Wei, 2015). They are; *need*, which refers to a minimum required threshold for goods or benefits, *capacity*, which refers to the ability to contribute to problem solving, and *responsibility*, which refers to culpability for contributing to an issue.

4.1 Need Principle

The principle of need provides an absolute standard that must be achieved through any distribution. Multiple studies have found evidence of support for meeting basic needs as a central requirement of distributive fairness (Carlsson et al., 2011). The principle of need implies a threshold below which an entity would not be obliged to accept any burden for addressing an issue. This is clearly evident in the Kyoto Protocol, where Non-Annex 1 countries were excluded from emission mitigation targets (UN, 1998). The selection of a need threshold should be undertaken within the context of the issue being addressed. In the research literature, one of the most common approaches has been to create a threshold, and grant exemptions, at the point where average income falls below an official poverty line (Baer, 2013; Underdal & Wei, 2015). For example, the United Nations classifies countries into three broad categories, based primarily on their Gross National Income (GNI). In 2013, countries with less than $1,035 GNI were classified as low income countries (UN, 2014), under the Kyoto Protocol these countries generally fell into the Non-Annex 1 category and were excluded from any requirements to reduce GHG emissions due to their more urgent development needs (UN, 1998).

4.2 Responsibility

Distributive justice theory distinguishes between an agent's role in causing damage and that agent's moral responsibility for the damage it has caused (Underdal & Wei, 2015). Konow (2001) emphasises that responsibility should only be considered in respect to variables which can be influenced by an agent. However the 'Brazilian Proposal' in climate negotiations argues that countries should be considered culpable for historic emissions, despite present day governments having no control over the actions of past governments, and past governments having had no understanding of the adverse effects of GHG emissions (Klinsky & Dowlatabadi, 2009). Climate change negotiations coalesced around the year 1990, in which the first IPCC report was published to develop a mechanism to assign responsibility for historical damage to the world's atmosphere. It has been argued in climate change research that, as the responsibilities of individuals within a country vary widely due to power and income inequalities, the best level of analysis for determining responsibility should be individuals, or small entities rather than countries (Newell et al., 2015). However, Underdal & Wei (2015, p. 38) developed a responsibility assignment mechanism that assigns proportional responsibility for CO_2 emissions amongst whole countries as shown in Box 1.

> Countries with per capita CO_2 emissions above the world average … have proportional responsibility for all their own emissions. Countries emitting between 50% and 100% of the world average … are proportionally responsible for emissions within that interval only. Countries emitting <50% of the world average … are granted full exemption.

Box 1. Responsibility assignment mechanism for CO_2 emissions.

This approach eliminates responsibility for those who have contributed little to the issue, it gives partial responsibility to those who have contributed at a level below average, and full responsibility to those above average. In doing so, it protects the development of low emitters, and assigns them no moral responsibility for the issue. The latest round of climate negotiations, COP 21 in Paris however, moved away from the Kyoto Protocol approach of completely exempting developing countries from responsibility, towards a more bottom-up approach, whereby countries now determine their own emission targets (UN, 2015). Under the new 'Paris agreement', while developing countries are still expected to make smaller mitigation commitments than developed countries, they can no longer be said to be 'exempt'. Assigning these countries a 'low responsibility' for their emissions is a more accurate interpretation of the latest climate change agreement.

4.3 Capability Principle

Capability refers to an entities capacity to contribute, and can only be properly assessed with reference to a specific task or function. Capability in climate change negotiations, has related largely to a countries material wealth often measured by GDP (Füssel, 2010) which is seen to be a determinant of its ability to contribute to climate change mitigation and adaption. A wide range of interpretations of capability have been put forward by researchers and policymakers. At one end of the spectrum is the capability approach, which focuses upon people's capability to achieve outcomes that they "value and have reason to value" (Sen, 1999, p. 18). Due to the substantial cultural and socio-economic variation at all levels of societies on what ‘a life that people value and have reason to value’ means, the capabilities approach presents a significant challenge to apply in an international, or even a nationwide context. The Human Development Index (HDI), which is based on some of the key concepts of the capabilities approach, and developed in part by Sen himself, does however provide a measure of capability (Winkler et al., 2013) which has been applied to multiple countries. The HDI comprises a composite measure of health, education, and standard of living (gross national income per capita), which combined are considered to measure human development, but can be reframed as a measure of capability (Winkler et al., 2013).

Confronted with complexity, and data limitations however, capability in international climate change has largely been limited to simply ‘capacity to pay’ for mitigation, measured by GDP per capita (Underdal & Wei, 2015). In a similar vein, a relatively prominent approach to equitable burden sharing in climate change negotiations known as the Greenhouse Development Rights (GDR) framework, defines capability as “income above a threshold, below which individuals are presumed to have ‘development’ as their appropriate priority and thus be exempted from climate-policy burdens” (Baer et al., 2009, p. 270). What makes the GDH framework unique amongst other methods for determining capability in climate change negotiations is that its site of focus is at the individual household level, rather than a national level, making it particularly relevant for agricultural sustainability assessment.

5. Burden Sharing of Targets and Equitable Benchmarking: a hypothetical example for New Zealand agriculture

The three principles of CBDR & RC and criteria for measuring and categorising entities against them might be used at a much more local scale to erect fair targets and to determine which other entities to benchmark their performance against. In order to identify potential problems and opportunities, we have conducted a thought experiment in which we apply them to three measures of environmental sustainability measured on New Zealand orchards, vineyards and farms by the ARGOS and NZ Sustainability Dashboard project (Merfield et al. 2015): efficient use of energy; appropriate application of artificial fertilisers; and minimal yet sufficient application of chemical sprays to achieve Integrated Production goals.

Following the United Nations approach to basing a *needs* threshold on a monetary criterion (UN, 2014), a *need* threshold within an agricultural sustainability assessment context could also be set based on a financial measure. In accordance with the long-term requirements of sustainability, solvency, understood as a ratio between liabilities and equity, can provide an indicator of a farms ability to meet its basic needs and continue operations (see Table 1).

While there is no formal ranking of the importance of the three primary distributive justice principles by the UNFCCC, it appears to be widely accepted that insofar as needs refer to basic goods or fundamental human rights, the needs principle is the most important, and provides a gateway test for entry into assessment against the other principles (Underdal & Wei, 2015). We therefore suggest a need threshold could be set at a solvency ratio of zero, under which entities are no longer able to meet their debt obligations (Fig. 1). Any farm that is insolvent is unlikely to be in a position to take on significant additional sustainability burdens in a voluntary sustainability initiative and therefore be exempt from additional sustainability burdens

Table 1: A hypothetical application of three primary principles of distributive justice applied to target setting and benchmarking for individual orchards, vineyards and farms.

Equity Principle	Example Criteria	Categorisation
Need	Solvency e.g. • Solvency Ratio = $$ potentially earned - $$ owed for production and land	• **Low Need** – Solvency ratio greater than zero • **High Need** – Solvency ratio Less than zero
Responsibility	Historic performance against a sustainability issue relative to the group average, measured as distance below or above some optimum level (or band of levels) e.g. • Difference in fruit/spray/ha from IPM target. • Decreased fruit/J energy invested/ha from the maximum predicted from yield curve.	• **Low Responsibility** – Performance above the group average • **Medium Responsibility** – Performance between 50 percent and 100 percent of the group average • **High Responsibility** – Performance below 50 percent of the group average
Capability	Solvency beyond the need threshold e.g. • Solvency Ratio = $$ potentially earned - $$ owed for production and land (only applicable where solvency>0)	• **Low Capability** – Solvency ratio below 50 percent of the group average • **Medium Capability** – Solvency ratio between 50 percent and 100 percent of the group average • **High Capability** – Solvency ratio below above the group average

The approach to categorising *responsibility* provided by Underdal & Wei (2015) shown in Box 1 can also be adapted to apply in an agricultural sustainability assessment scenario. For the purpose of this paper, the point is not to quantify a required standard of performance improvement, such as a certain percent reduction in emissions, but rather to categorise farms based on their level of responsibility into groups of high, medium, or low for issues like application of fertiliser, chemical sprays or energy consumption (see Table 1 and Fig. 1). However several complications must be considered, rather than simply assuming risk from sprays or fertiliser is linearly related to application rates. Sprays are applied to New Zealand kiwifruit and vineyards within strict Integrated Pest Management guidelines based on pest insect counts and trigger thresholds to protect market access requirements – applying too few sprays can build resistance amongst the pests, and would fail to protect the crop and exports; applying too many sprays creates unnecessary toxic risks for the wider environment and human health risks amongst consumers. If we assumed the consequences of over-spraying were the same as those from under-spraying, the absolute (+ or -) deviation from optimum could measure responsibility for change as well as 'distance to target'. Similar approaches could apply to fertiliser use where the optimum application is set at sufficient inputs to maintain soil health and production. On the other hand, responsibility measured by energy use is more likely to be linearly and directly related to environmental harm, especially where the energy subsidies for food production are based on fossil fuel inputs. However, even in the energy case, ethical consideration of providing food security could be included by scaling all energy inputs against the amount and quality of food produced per hectare: adding successively more energy subsidies into production eventually will

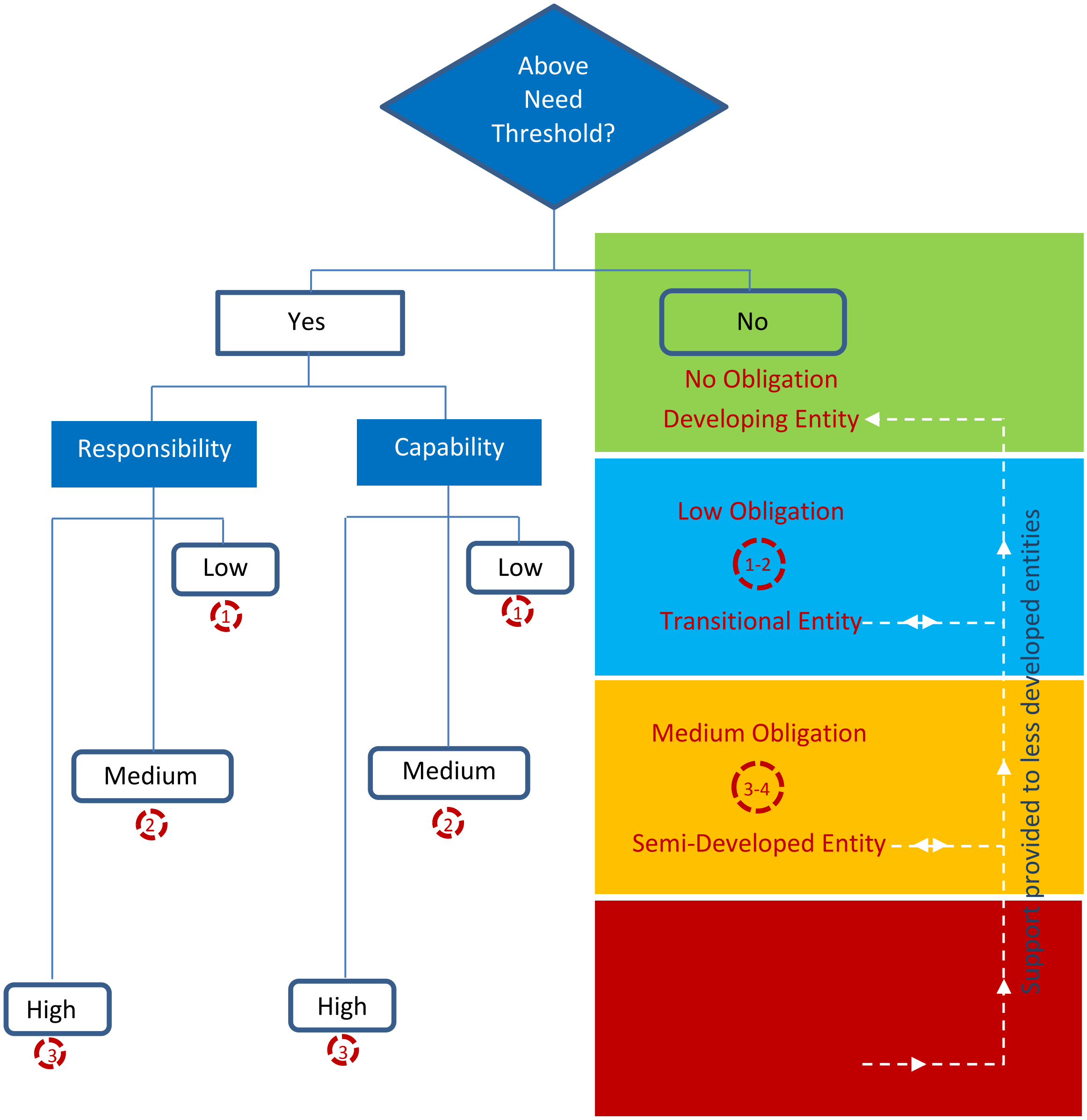

Figure 1. *A potential burden sharing framework for sustainability targets for individual farms, orchards, vineyards or forests within an industry-wide sustainability programme. The underlying principles are based on the climate change mitigation negotiations between nations. Although applied to burden sharing, the same principles can be applied to fair benchmarking, in which the farm's current performance is only scaled against other farms that face the same (low, medium, high) levels of capability*.

produce little extra fruit, so we expect (fruit/J/ha; Table 1) to level and then inflect downward when too much energy is used. Responsibility for adjusting energy inputs might therefore be best measured as the relative inputs above the point on a yield curve where fruit/J/ha first begins to flatten as energy inputs climb.

There are many objections that can be raised to adopting a simple GDP per capita, or other monetary approach to defining *capability* alone, particularly from the more rigorous perspective provided by the capabilities approach (Sen, 1999). Levels of development or capability cannot be entirely understood by an increase in individual consumption or GDP. However, like the GDH framework (Baer, 2013) and other approaches (see Winkler et al., 2013), a first approximation might be to adopt a simple monetary measure of *capability* based on the previous definition of *need* to assign high, medium, or low capability for in the same manner as proposed for the responsibility categorisation (Box 1,Table 1, Fig. 1).

Elements of this same approach might also form the basis of equitable benchmarking when comparing performance against other farms i.e. once the need threshold has been met, is it more fair and acceptable to compare current performance of only those farms that share the same *capability* to do something about the problem?. If a proxy measure was available for *responsibility* (such as historical discharge of an accumulating pollutant), then it might be feasible to define benchmark panels based on some combination of both *responsibility* and *capability*.

6. General Discussion & Conclusions

The framework presented by this paper is intended to provide a starting point for discussions around burden-sharing when setting targets for agricultural sustainability assessments. Our review of 12 sustainability frameworks showed that most included targets, but in general they lacked an explicit rationale for how they are derived. This lack of transparency is likely to undermine their usefulness for encouraging collective action amongst all the participants in food and fibre production, distribution, marketing and consumption. Most of the indicators were practice-based i.e. a measure of the presence or absence of a best practice (farming input), and simply assume that they will lead to sustainability (a farm output). Performance-based indicators could test this fundamental and widespread assumption that we can adequately steer sustainability by monitoring farm inputs, but the necessary performance-based scoring systems were relatively uncommon. Frequently, the quantitative measures presented are secondary calculations on aggregated scores above the farm level (e.g. what proportion of farms in a given agricultural sector or product line followed best practice). We expect such general primary scores of inputs at the farm level to be relatively crude tools for learning and incentivising change, because the scale of measurement is binary or ordinal, and the definitions are necessarily generalised, making them hard to evaluate and potentially not trusted by decision makers working further along the food supply chain.

Outsiders beyond the farm will potentially be interested in a much “bigger picture” formed by large-grained and aggregated metrics, whereas producers must make decisions on fine-grained and locally tuned information to guide their own investments and land care. “No one size fits all” when designing sustainability metrics or indicator sets (de Olde et al. 2016) and the TempAg research team has rightly identified that the scale at which comparisons are to be attempted has a crucial influence on what is measured, how, by whom and for what purpose. We found it difficult to propose limits for how sustainability performance could be compared by TempAg or OECD between nations which farm temperate agroecosystems, because it is not yet clear how such comparisons would ever be used in a policy context. Is the intent to create league tables like those commonly used by OECD to compare social and economic wellbeing across its member states? We presumed yes, or at least that such broad scale comparisons would be wanted if they could be robust enough. Our review of existing sustainability frameworks suggested that the measures at farm scales are far too crude to yet allow this aggregated comparison of absolute measures. We conclude that individual farms, local communities or industries, and even national agricultural sectors, should instead best use the existing sustainability assessments for learning and improvement by 'internal *benchmarking'*. In this use, sustainability metrics are only used as relative and proxy measures of change and improvement in their own local contexts rather than looking across to compare performance in very different ecologies and socio-economic constraints and opportunities. This may be a slower way of incentivising change, but it is practical and the local comparison ensures equity and local relevance in ways that farmers looking over at their immediate neighbours can quickly accept as valid and fairer.

When setting out to review the design criteria for setting targets and making benchmark comparisons, we quickly encountered the more general question of how targets and benchmarking comparisons can be made fair and thereby enduring and collaborative in effect. Although our hypothetical application of CBDR & RC proxy measures suggests that some proxy measures can conceivably measure responsibility and capability to shift local farm inputs and management, two further overarching complications are likely to arise when applying the CBDR & RC framework to target setting and equitable benchmarking: (a) are all dimensions of sustainability performance to be considered equally important; and (b) are responsibility and capability to be treated as equally important after first meeting need thresholds? It seems inevitable that a farm may fall into the low performance bracket for say energy, yet medium for fertiliser, and also fertiliser is a very important component of energy use on the farm. So overall targets for whole farming systems adjustment must weight these different components in some agreed way. Similarly, a solvent farm may score low on responsibility yet high on capability for the same environmental input issue. Figure 1 has an embedded assumption that CBDR has the same driving importance as RC, in which case a target response or benchmark comparison would be scaled as the average of the two fairness criteria. A valid moral argument can be made that entities with a low responsibility should have low obligations (Konow, 2001). However, collective challenges like sustainability rely heavily of group participation in a social contract (Rawls, 1971), so some will argue that all those with the capability to act have a moral obligation to do so, even if they played a relatively little role in creating the problem, or are currently performing better than many of their counterparts to minimise future impacts.

Clearly our thought experiment and Table 1 & Figure 1 work best as heuristic devices to illustrate several potential complications in applying the CBDR & RC approach at a farm, national or international levels. Much work is left to be done in devising fair bases for comparisons of agricultural sustainability performance by TempAg. Nevertheless, the international experience in dealing with a shared common problem like climate change mitigation illustrates that failure to take proper account of need, responsibility and capability when setting targets or making benchmark comparisons would undermine collective action across temperate agriculture systems. Feeding a growing human population using global markets and distribution systems, avoiding land degradation and loss of biodiversity, global biosecurity are just some of the ways that finding shared solutions to agricultural sustainability is no less complicated or urgent as combatting climate change. Extreme care is needed in ensuring and negotiating fairness and ethical concerns when erecting targets and comparing farming performance across multiple scales and jurisdictions within temperate agricultural systems.

Acknowledgements

The New Zealand Sustainability Dashboard Project is funded by NZ's Ministry of Business, Innovation, and Employment and several primary industry partners. For more information visit http://www.nzdashboard.org.nz/.

References

Acosta-Alba, I., & van der Werf, H. (2011). The use of reference values in indicator-based methods for the environmental assessment of agriculural systems. *Sustainability, 3,* 424-442.

Baer, P. (2013). The greenhouse development rights framework for global burden sharing: reflection on principles and prospects. *Wiley Interdisciplinary Reviews: Climate Change, 4*(1), 61-71.

Baer, P., Athanasiou, T., Kartha, S., & Kemp-Benedict, E. (2009). Greenhouse development rights: A proposal for a fair global climate treaty. *Ethics Place and Environment, 12*(3), 267-281.

Bastian, O., Corti, C., & Lebboroni, M. (2007). Determining environmental minimum requirements for functions provided by agro-ecosystems.. *Agron. Sustain. Dev.*, *27*, 279-291.

Benge, J., Barber, A., Le Quellec, I.,. & Manhire, J. (2016). Implementation of sustainability assessment and reporting in horticulture: a case study of New Zealand's wine and kiwifruit sectors. This symposium, Workshop 2.2.

Carlsson, F., Kataria, M., Lampi, E., Löfgren, Å., & Sterner, T. (2011). Is fairness blind?—The effect of framing on preferences for effort-sharing rules. *Ecological Economics, 70*(8), 1529-1535. doi:10.1016/j.ecolecon.2011.03.015

van Cauwenbergh, N., Biala, K., Bielders, C., et al. (2007). SAFE—A hierarchical framework for assessing the sustainability of agricultural systems. *Agr. Ecosyst. Environ, 120*, 229-242.

Dannenberg, A., Sturm, B., & Vogt, C. (2010). Do Equity Preferences Matter for Climate Negotiators? An Experimental Investigation. *Environmental and Resource Economics, 47*(1), 91-109. doi:10.1007/s10640-010-9366-5

Figge, F., & Hahn, T. (2005). The cost of sustainability capital and the creation of sustainable value by companies. *Journal of industrial ecology, 9*(4), 47-58.

Füssel, H.-M. (2010). How inequitable is the global distribution of responsibility, capability, and vulnerability to climate change: A comprehensive indicator-based assessment. *Global Environmental Change, 20*(4), 597-611. doi:10.1016/j.gloenvcha.2010.07.009

Gregory, P.J. (2016). TempAg; An international Research Consortium for Sustainable Agriculture in Temperate Regions This symposium, Workshop 2.4.

van der Heide M., Brouwer, F.M., Bellon, S., et al. (2007). Review of approaches to establish reference levels to interpret indicators. SEAMLESS Report No. 30, SEAMLESS integrated project, EU 6th Framework Programme, contract no. 010036-2, www.SEAMLESS-IP.org, 44 pp, ISBN no. 90-8585-118-1 and 978-90-8585-118-9.

Klinsky, S., & Dowlatabadi, H. (2009). Conceptualizations of justice in climate policy. *Climate Policy, 9*(1), 88-108.

Konow, J. (2001). Fair and square: the four sides of distributive justice. *Journal of Economic Behavior & Organization, 46*(2), 137-164.

Lange, A., Löschel, A., Vogt, C., & Ziegler, A. (2010). On the self-interested use of equity in international climate negotiations. *European Economic Review, 54*(3), 359-375.

Langhelle, O. (2000). Sustainable development and social justice: expanding the Rawlsian framework of global justice. *Environmental Values, 9*(3), 295-323.

Merfield, C., Manhire, J., Moller, H., et al. (2015). Are organic standards sufficient to ensure sustainable agriculture? Lessons from New Zealand's ARGOS and Sustainability Dashboard projects *Sustainable Agriculture Research* 4: 158-172.

Moldan, B., Janoušková, S., & Hák, T. (2012). How to understand and measure environmental sustainability: Indicators and targets. *Ecological Indicators, 17*, 4-13.

Newell, P., Bulkeley, H., Turner, K., Shaw, C., Caney, S., Shove, E., & Pidgeon, N. (2015). Governance traps in climate change politics: re-framing the debate in terms of responsibilities and rights. *Wiley Interdisciplinary Reviews: Climate Change, 6*(6), 535-540. doi:10.1002/wcc.356

OECD-FAO. (2010). *Agricultural outlook 2010–2019*. (Paris: Organisation for Economic Cooperation and Development.

de Olde, E., Moller, H., Marchand, F. et al. (In press, 2016). When experts disagree: the need to rethink indicator selection for assessing sustainability of agriculture. *Environment, Development and Sustainability*

Pearce, D. (1987). Foundations of an ecological economics. *Ecological Modelling, 38*(1–2), 9-18. doi:http://dx.doi.org/10.1016/0304-3800(87)90042-1

Pretty, J., Smith, G., Goulding, K., Groves, S., Henderson, I., Hine, R., . . . Vis, J. (2008). Multi-year assessment of Unilever's progress towards agricultural sustainability I: indicators, methodology and pilot farm results. *International journal of agricultural sustainability, 6*(1), 37-62.

Rajamani, L. (2000). The Principle of Common but Differentiated Responsibility and the Balance of Commitments under the Climate Regime. *Review of European Community & International Environmental Law, 9*(2), 120-131. doi:10.1111/1467-9388.00243

Rawls, J. (1971). *A Theory of Justice*. Oxford: Oxford University Press.

Sen, A. (1999). *Development as freedom*: Oxford Paperbacks.

Stalley, P. (2013). Forum: Principled Strategy: The Role of Equity Norms in China's Climate Change Diplomacy. *Global Environmental Politics, 13*(1), 1-8.

UN. (1992). *Rio Declaration on Environment and Development*: UN General Assembly.

UN. (1998). *Kyoto Protocol to the United Nations Framework Convention on Climate Change*. Kyoto: The United Nations.

UN. (2014). *World Economic Situation and Prospects 2014*. New York: United Nations.

UN. (2015). *Adoption of the Paris Agreement*. Paris: United Nations.

Underdal, A., & Wei, T. (2015). Distributive fairness: A mutual recognition approach. *Environmental Science & Policy, 51*, 35-44.

Winkler, H., Letete, T., & Marquard, A. (2013). Equitable access to sustainable development: operationalizing key criteria. *Climate Policy, 13*(4), 411-432.

Yaari, M. E., & Bar-Hillel, M. (1984). On dividing justly. *Social choice and welfare, 1*(1), 1-24.