DOCTOR OF PHILOSOPHY

The use of Randomised Control Trials in evaluating conservation interventions/ the case of Watershared in the Bolivian Andes

Pynegar, Edwin Lindsay

Award date:
2018

Awarding institution:
Bangor University

The use of Randomised Control Trials in evaluating conservation interventions: the case of *Watershared* in the Bolivian Andes

Edwin Lindsay Pynegar

A dissertation submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy (PhD)

PRIFYSGOL
BANGOR
UNIVERSITY

March 2018

Contents

i. List of tables 3

ii. List of figures 4

iii. Acknowledgments 5

iv. Thesis Abstract 7

1 Introduction 9

1.1 Evidence Use in Conservation 9

1.2 Payments for Ecosystem Services 9

1.3 *Watershared* 12

1.4 The *Área Natural de Manejo Integrado Río Grande – Valles Cruceños* 15

1.5 *Watershared* in the *ANMI Río Grande*, and its evaluation using a Randomised Control Trial: Case Study 16

1.6 Research objectives, questions, methods, and structure of the thesis 22

1.7 The use of existing data and links to other research projects 25

2 What role should Randomised Control Trials play in providing the evidence base underpinning conservation? 27

2.1 Abstract 27

2.2 Introduction 27

2.3 Under what circumstances might an RCT evaluation be useful? 32

2.4 What affects the feasibility of RCT evaluation? 34

2.5 What factors affect the quality – the 'internal validity' – of an RCT evaluation? 36

2.6 Conclusions 39

3 Impact of Payments for Watershed Services on water quality: an evaluation using a Randomised Control Trial 42

3.1 Abstract 42

3.2 Introduction 42

3.3 Methods 45

3.4 Results 51

3.5 Discussion 55

3.6 Conclusions 58

4 Can Payments for Watershed Services improve health in communities where it is implemented? 60

4.1 Abstract 60

4.2 Introduction 60

4.3 Case Study: Socio-Ecological Context 63

4.4 Methods 65

4.5 Results 70
4.6 Discussion 73
4.7 Conclusions 78
5 The devil in the detail: experiences from the implementation of a large-scale socio-ecological Randomised Control Trial 79
5.1 Abstract 79
5.2 Introduction 79
5.3 Methods 81
5.4 Results 85
5.5 Discussion 93
5.6 Conclusions 100
6 Land use and conservation of catchments supplying water to communities: two case studies 102
6.1 Abstract 102
6.2 Introduction 102
6.3 Health records, water quality, and water supply systems: further evidence of the influence of land use on water quality and health 103
6.4 How much of a catchment needs to be conserved to achieve improvements in water quality? Evidence from the *Reserva de Patrimonio Natural* (*REPANA*) in La Aguada 107
6.5 Conclusions 109
7 Discussion 111
7.1 Key findings 111
7.2 What does the experience of evaluating *Watershared* tell us about the value of RCTs in conservation impact evaluation? 112
7.3 How can *Watershared*-like programs be implemented in a way more likely to achieve their intended goals? 117
7.4 Implementation of *Watershared* as an RCT in the Bolivian Chaco 122
7.5 Conclusions 124
Reference List 126
Appendix A. List of communities randomly allocated to treatment and control groups. 150
Appendix B. Example of the text of a *Watershared* agreement between a landowner and *Natura* . 152
Appendix C. List of incentives available to those signing *Watershared* agreements 155
Appendix D. Chapter 3 Supplementary Information 159
Appendix E. The use of Coliscan Easygel to conduct water monitoring in low-resource contexts. ... 164
Appendix F. Chapter 4 Supplementary Information 169
Appendix G. Chapter 6 Supplementary Information 180

i. List of tables

Table 1.1. Characteristics of *Watershared* agreements in the *ANMI Río Grande – Valles Cruceños*... 21
Table 2.1. Consequences of behavioural effects when compared with results obtained in a hypothetical double-blind RCT.......... 41
Table 3.1. Variables hypothesized to be important in predicting *E. coli* concentration in 2015.......... 51
Table 3.2. Number and proportions of intakes visited in 2015 in treatment and control communities with and without cattle access. 54
Table 5.1. The selected randomisation unit – community – presented problems for many of the intended outcomes of the *Watershared* intervention. 86
Table 5.2. Number of intakes feeding school tap which remained in the same location between each possible pair of rounds of water monitoring undertaken in the 129 communities included in the RCT 89
Table 6.1. Cases of diarrhoeal disease attended at community health centres in 2014 and 2015 in the three largest communities of the *ANMI Río Grande* area.. 104

ii. List of figures

Figure 1.1. Visual representation of a reciprocal agreement. 14
Figure 1.2. a) Locations of the 65 treatment and 64 control communities included in the RCT. b) Location of the *ANMI Río Grande – Valles Cruceños* protected area within Bolivia. 18
Figure 1.3. The *Watershared* theory of change relevant to the intervention as implemented in the *ANMI Río Grande*, and potential threats to its validity 20
Figure 2.1. Summary of our suggested decision-making process for evaluators relating to RCT feasibility and quality, and alternative evaluation options if RCTs are inappropriate 31
Figure 2.2. The Bolivian NGO *Fundación Natura Bolivia* conducted an RCT of their PES-like conservation program, *Watershared*, in the Bolivian Andes between 2011 and 2016. 32
Figure 3.1. Schematic of an example community with 2 water intakes and 6 households, showing locations of sampling sites. 47
Figure 3.2. Relative *E. coli* levels in treatment and control sites in 2010 and 2015. 52
Figure 3.3. The effects of 2010 *E. coli* concentration and whether a site is in a treatment or control community on 2015 *E. coli* concentration.. 53
Figure 3.4. The effect of site features which are predictors of 2015 E. coli concentration in the most likely GLMM 54
Figure 4.1. Examples of typical water supply systems in communities of the ANMI Río Grande. 63
Figure 4.2. Effects of deforestation and extensive cattle grazing on water supplies, as presented by *Natura* when offering the *Watershared* agreements to community members. 64
Figure 4.3. Conditional average of models testing predictors of *E. coli* levels in water supply systems 71
Figure 4.4. Coefficient plot (with 95% CI values) showing the effect of age and measured *E. coli* concentration in household water supplies on diarrhoeal disease levels in communities over the previous 12 months. 72
Figure 4.5. Coefficient plot (with 95% CI values) showing the effect of age and measured *E. coli* concentration in household water supplies on incidence of diarrhoeal disease over the previous 12 months in communities. 72
Figure 4.6. Coefficient plot (with 95% CI values) showing conditional average of models relating to water supply infrastructure and institutions and conservation actions on self-reported levels of diarrhoeal disease over the previous 12 months among children in communities 73
Figure 4.7. Improved water source supplying the community of Paredones. 77
Figure 5.1. Watershared mapped in the community of Piraimirí. 87
Figure 5.2. Randomisation units do not necessarily align for different outcomes of interest. 88
Figure 5.3. Multiple water intakes and water intake movement presented problems for monitoring in the community of Huertas. 90
Figure 5.4. The voluntary nature of the intervention has consequences for evaluating the effectiveness using an RCT. 92
Figure 5.5. Conceptual model of the relationship between conservation actions and outcomes of interest.. 98
Figure 6.1. Locations of current and former water intakes supplying Pucará, catchments, and sites measured in 2016. 105
Figure 6.2. Cases of diarrhoea attended at the health centre in Pucará, by month, in 2014 and 2015. Supply was switched to the new intake in August of 2015. 106
Figure 6.3. Boundary between *REPANA* area (forested land, background) and non-*REPANA* area (non-forested degraded land in foreground). 108
Figure 6.4. Map of La Aguada, the *REPANA* area, and the catchment within which it is located. 109

iii. Acknowledgments

I would like to thank from the bottom of my heart everyone who has helped and supported me in researching and writing this thesis over the past four years. It would not have happened without you.

In particular, I would like to thank my supervisors, Professor Julia Jones, Dr. James Gibbons and Dr. Nigel Asquith, for their support, patience, and extremely hard work – and also for continuing to believe in me even when I may not have believed in myself. I owe many thanks to Dr. Emma Wiik and Dr. Rémi d'Annunzio for their generosity in sharing data. I am also deeply indebted to Professor Jörn Scharlemann, Daniel Ingram, Martin Jung, Richard Hazell and Dr. Claudia Gray for being willing to 'adopt' me for a time at the University of Sussex, and also would like to thank Professor Morag McDonald for being so kind in facilitating this.

I can hardly overstate how grateful I am to have had such excellent colleagues at *Fundación Natura Bolivia*, and I look forward to more collaboration in the future. I am also extremely grateful for the opportunities *Natura* has provided me with, especially in working on the design of the new randomised control trial currently being implemented in the Chaco and representing them at the 3ie Development Priorities Window post-award workshop in New Delhi. In particular, I'd like to thank María Teresa Vargas, Tito Vidaurre, Lucindo Gonzales, Natalia Araujo, Huáscar Azurduy, Dr. Veronica Chávez, Dr. Andrea Markos, Máximo García, Hugo Vallejos, Robert Rueda, and many others for being willing to share their time, resources, knowledge and experience with me. I owe particular and heartfelt thanks, however, to Dionicio Toledo for being such a careful and skilled driver on the challenging roads of the ANMI Río Grande, for introducing me to the local communities, and for sharing his profound knowledge of local society, as well as for being prepared to work what amounted, often, to antisocial hours.

I'd also like to mention the excellent friends I made in Santa Cruz, who not only ensured that I enjoyed my time there but also helped me to understand Bolivia better. Particular thanks go to Dr. Patrick Bottazzi and David Crespo for their immense friendliness and hospitality, Manuel Menacho for his support and knowledge in hacking through the thickets of Bolivian bureaucracy at the Oficina de Migración, Dr. Ximena Vélez-Liendo and her family for making me so welcome, sharing her profound knowledge of Bolivia's ecology and in particular that of the Andean spectacled bear, as well as introducing me to Bolivian wine and gin, as well as Yara Rojas, Sharon Pereyra, Sara Franco, Dr. Steffen Reichle, María Angelica Steinbach, Rein Meirte, Bart Crezee, Lotta Chan, the Suárez Sosa family, and many more.

Finally I would like to dedicate this thesis to my family and friends, as a small token of my appreciation for your love and support, in the good times and the bad. You know who you are.

iv. Thesis Abstract

There is growing awareness among conservation practitioners, funders and policymakers of the importance of evaluating the impact of an intervention on outcomes of interest. Historically conservation has been weak at undertaking high-quality counterfactual evaluations, which has resulted in calls for more Randomised Control Trials (RCTs) of conservation interventions. This thesis studies one of the few RCT evaluations of a large-scale conservation intervention, a payments for ecosystem services-like program in the Bolivian Andes known as *Watershared*. *Watershared* provides in-kind incentives (barbed wire, fruit trees, cement, and irrigation equipment) to upstream farmers in exchange for them agreeing to keep cattle out of riparian forests and avoiding clearing forested land. By doing so it aims to conserve forest carbon stocks and biodiversity as well as improving water quality, health outcomes and wellbeing in the communities in which it is implemented.

I review the evaluation literature to identify the circumstances under which RCTs may be an appropriate approach for conservation impact evaluations. I identify seven key issues relating to RCT feasibility, utility and quality and demonstrate how these issues were dealt with in the RCT of *Watershared*. I use the RCT to evaluate the impact of the *Watershared* intervention on water quality, as measured by *Escherichia coli* concentration in community water supplies in 129 communities at baseline (2010) and endline (2015). The before-after RCT-based evaluation shows no effect of the intervention on water quality at the landscape scale. However detailed analysis from the endline data shows that presence of cattle faeces is a predictor of lower quality water. I conclude that the intervention (excluding cattle from riparian forest) has the potential to improve water quality but that the intervention as implemented did not result in an increase in cattle exclusion from the areas most likely to influence water quality (forest close to water intakes). One of the ultimate aims of *Watershared* is to reduce waterborne disease in the communities; I evaluate this using both a household survey conducted in all communities and a further set of *Escherichia coli*-based water monitoring data taken in 2016. While factors such as age predict disease incidence, none of the actions directly related to *Watershared* (e.g. fencing off water intakes) were found to do so. *E. coli* levels did however significantly predict disease incidence, and a number of monitored environmental factors affected *E. coli* levels including land use in catchments supplying water intakes. Therefore the environmental context is important for water quality, including aspects of it related to *Watershared* actions, but we did not find that the intervention resulted in measurable impacts on disease levels.

Analysing data from the *Watershared* RCT revealed challenges which are relevant to the design of RCT evaluations of other interventions. Many of these challenges stem from attempting to evaluate multiple outcomes at once; the appropriate randomisation unit for each outcome may not be the

same and impacts may spill over between intervention and control units. The voluntary nature of interventions such as *Watershared* (and other PES-like programs) also make interpretation of RCT results challenging. While RCTs can be successfully used to evaluate multi-outcome socio-ecological conservation interventions, this is not straightforward and substantial planning is needed to ensure they provide robust evidence. As a result, evaluators should not expect to be able to conduct a high-quality RCT cheaply. I therefore argue that if RCTs are to be more widely used in conservation, this will require greater collaborations between researchers and practitioners. I also conclude that design of successful payments for watershed services-type programs will be dependent upon a good understanding of the functioning of action-outcome relationships, and that some socio-ecological contexts will facilitate implementation of such programs while others may hinder it. Thus program designers must consider both ecological processes and social contexts, and the interactions between the two, if programs are to have the potential to achieve their intended goals.

1 Introduction

1.1 Evidence Use in Conservation

It is widely agreed that decision-making in conservation practice and policy should be based on scientific evidence (Pullin & Knight 2001; Sutherland *et al.* 2004; Segan *et al.* 2011; Pullin *et al.* 2013). Despite this, historically, evidence of intervention effectiveness has not been well used to inform decision-making processes in conservation (Sutherland *et al.* 2004; Pullin *et al.* 2004). Worse still, much of the evidence that does exist is known to be generally of poor quality due to the failure of conservation researchers and practitioners to undertake *counterfactual* evaluations (Baylis *et al.* 2016). As a result, there has been little systematic high-quality evidence produced of the effectiveness of many kinds of conservation intervention, nor of their relative effectiveness when compared with other possible interventions. This includes interventions that have become very widely implemented in recent years, such as *payments for ecosystem services* (Ferraro & Pattanayak 2006; Pattanayak, Wunder & Ferraro 2010; Miteva, Pattanayak & Ferraro 2012). While this is now slowly beginning to change (e.g. Samii *et al.* 2014; Börner *et al.* 2016; Jayachandran *et al.* 2017), it is clear that conservation remains far behind other fields of applied science – such as medicine, development microeconomics, and education – in the use of counterfactual-type evaluations to measure effectiveness. This thesis represents one of the first uses of a randomised control trial (RCT) to evaluate a payments for ecosystem services-like program, and so represents a contribution towards changing this.

1.2 Payments for Ecosystem Services

Ecosystem services are considered to be the benefits that ecosystems of all kinds provide to human societies (e.g. Daily 1997). These services may be categorised as provisioning, regulating, supporting, or cultural services (Millennium Ecosystem Assessment 2005), and their value is of the order of trillions of dollars annually (Costanza *et al.* 2014). Many or most cases of environmental destruction or degradation can therefore be conceptualized as a reduction in the supply or availability of ecosystem services (Costanza *et al.* 2014). In many such cases the party disadvantaged by this will not be the same as that whose actions caused the reduction in supply of the service initially (Pearce & Turner 1990; Daly & Farley 2010). Such changes are therefore negative externalities of human activities; ecosystem services are often public or otherwise common goods (Fisher, Turner & Morling 2009) and their losses are a form of market failure (Balmford *et al.* 2002; Jack, Kousky & Sims 2008; also see Gómez-Baggethun *et al.* 2010). As a consequence, much environmental policymaking is implicitly or explicitly focused on maintaining or increasing the supply of ecosystem services. This may be done via legislation and its enforcement (including but not restricted to protected area establishment), voluntary standards, capacity building, and initiatives to influence behaviour by changing economic and other incentives (see Salafsky *et al.* 2008). This last case includes using market mechanisms to influence behaviour and/or providing direct or indirect payments for conservation. Both of these may be considered under the umbrella of 'Payments for Ecosystem Services' (PES).

Payments for Ecosystem Services interventions are a conservation tool predicated upon the understanding that individuals respond to incentives and that incentivising activities (normally land uses) that produce more ecosystem services will result in their greater availability (Ferraro & Kiss 2002; Gómez-Baggethun *et al.* 2010). Ecosystem service users pay or otherwise compensate ecosystem service providers for conducting some action understood to maintain or increase supply of the ecosystem service in question; generally this is a relatively unprofitable but environmentally more benign and thus socially desirable resource use. Thus PES programs are intended to change incentive structures in order to internalise the cost of environmentally and socially harmful activities (Coase 1960; Engel, Pagiola & Wunder 2008).

The past two decades have seen an increasing amount of interest in the use of PES programs in environmental and conservation policymaking and they have been implemented throughout the world (e.g. Ezzine-De-Blas *et al.* 2016; Börner *et al.* 2017). In the industrialised and newly industrialised world, PES has frequently been implemented through large government-run programs although often not under that name. For example, the Conservation Reserve Program in the United States has since 1985 incentivised farmers to protect soils through ceasing farming in environmentally sensitive areas (e.g. Goodwin & Smith 2003); the EU Common Agricultural Policy since 2001 has paid farmers to conserve habitats and reduce agrochemical inputs on their land (e.g. Baylis *et al.* 2008); and in China, the Sloping Land Conversion Program pays participants to reforest marginal agricultural lands (Xu *et al.* 2004). However PES has also become increasingly widely used throughout the developing world. The interest in PES comes from belief that the market-like aspects of PES have the potential to increase efficiency in the use of scarce conservation funding (Ferraro & Simpson 2002). Also as many ecosystem service providers are poor, payments can act as a development mechanism (e.g. Folke 2006; Wunder 2008). This is often seen as positive, as command-and-control approaches to conservation can strongly negatively affect already marginalised people and prohibit traditional resource uses (e.g. Roe & Elliott 2004). For example, Costa Rica (e.g. Sánchez-Azofeifa *et al.* 2007; Porras *et al.* 2013) and Mexico (e.g. Muñoz-Piña *et al.* 2008; Alix-Garcia, Shapiro & Sims 2012) have national payment programs for forest carbon and hydrological service provision; in Colombia sugarcane producers attempt to maintain their water supplies through payments to upstream forest owners (Rodríguez-de-Francisco & Budds 2015); and in Tanzania payments for biodiversity conservation are made to communities (Nelson 2008).

While PES is often conceptualised as a market-based solution to environmental degradation, with ES users buying 'units' of ES provision as required from ES suppliers who then change land uses in response to these purchases (e.g. Gómez-Baggethun *et al.* 2010), almost no real PES programs work in this way (Tacconi 2012; Santos de Lima, Krueger & García-Marquez 2017). (An exception may be Nestlé compensating farmers for reducing agricultural intensity to prevent nitrates from contaminating the aquifer from which Vittel mineral water is drawn [Perrot-Maître 2006].) This is because the relationship between actions and ES provision is often extremely complex and uncertain with time lag between actions and responses also often being excessively long (Santos de Lima, Krueger & García-Marquez 2017), monitoring, reporting and verification is costly and again technically difficult, production of carbon credits (for example) relies on a hypothetical baseline scenario which is open to dispute (Angelsen 2017), land tenure systems in which landowners and land users are not the same open up questions of just compensation (e.g. Naughton-Treves & Wendland 2014), and political considerations and ideology can lead to rejection outright on the grounds of PES representing 'neoliberal conservation' (e.g. Arsel & Büscher 2012).

More often, government agencies compensate landowners at a particular rate or rates for doing or not doing certain actions or maintaining land uses which are *believed* to provide ESs on behalf of their citizenry as a whole. In other cases, non-governmental organizations may set up PES programs to compensate individuals for conserving biodiversity or conducting other actions that the NGO or its funders feel worthwhile. In yet others, water users may directly contribute towards compensating landowners in conserving land believed important for provision of that water and the maintenance of its quality and quantity. This typology covers most conservation PES interventions currently in existence (Ezzine-de-Blas *et al.* 2016), and it shows that rather than the theoretical Coasean model of ecosystem service users and providers brought together to buy and sell ESs in a market or pseudo-market, PES programs instead consist of a Pigouvian model of subsidies for actions believed to be socially beneficial (Tacconi 2012; Martin-Ortega, Ojea & Roux 2013).

Perhaps because of these contradictions, 'PES' is a highly contested term, with many differing definitions (Tacconi 2012). PES is generally considered a subcategory of conservation with incentives, as contrasted with a 'fences and fines' command-and-control approach to conservation (Jack, Kousky & Sims 2008). Many implementers of such interventions consider their program to be PES, as PES has become a fashionable kind of intervention based upon its theoretical ability to achieve multiple conservation goals simultaneously with poverty alleviation (c.f. Redford, Padoch & Sunderland 2013). However it has become clear that PES conservation programs may generally be defined by their (presumed) *conditionality* – that actors receiving money or other incentives from a program are expected in return to conduct the incentivised action, and that such incentives can be stopped or removed if the recipients do not comply (Sommerville, Jones & Milner-Gulland 2009; Wunder 2015; Ezzine-de-Blas *et al.* 2016). PES is therefore a form of conditional transfer, an intervention widely implemented throughout the developing world in the past two decades (e.g. Rawlings 2005).

There has been a great deal of theoretical discussion in the academic and policy world about PES. It has been widely critiqued, particularly for being overly complex for realistic implementation in actually existing human-environment systems (Norgaard 2010; Muradian *et al.* 2013; Santos de Lima, Krueger & García-Marquez 2017) and for framing nature in a particular way which may be imposed upon, and clash with, local understandings or pre-existing norms of natural resource management (e.g. Rode, Gómez-Baggethun & Krause 2015). It is also controversial for potentially opening up ecosystem service provision much more widely to the private sector and thus allowing for greater 'privatisation' or commodification of nature (e.g. Arsel & Büscher 2012; Silvertown 2015). Despite this, effectiveness or otherwise of such interventions is fundamentally an empirical question. Good-quality impact evaluations of PES effectiveness in achieving its goals – change in provision of services due to the program – have been relatively rare as rarely have interventions been established with impact evaluation in mind (e.g. Pattanayak, Wunder & Ferraro 2010; Börner *et al.* 2017); many PES programs do not even measure baselines (Naeem *et al.* 2015). However, a number of recent reviews have shown that effects of PES on ecological outcomes have been mixed, with some having essentially zero effect and some large, lasting effects (Samii *et al.* 2014; Börner *et al.* 2016, 2017); the effect sizes are similar to those of other kinds of forest conservation interventions (Börner *et al.* 2016). There currently exists little evidence to suggest that PES can consistently achieve improvements in both ecological and human development/welfare indicators, although more research is required (Samii *et al.* 2014; also see Grieg-Gran, Porras & Wunder 2005; Pagiola, Arcenas & Platais 2005). The number of PES evaluations has recently become large enough to identify features of contexts or interventions which make those interventions significantly more likely to be effective (Ezzine-de-Blas *et al.* 2016; Börner *et al.* 2017).

1.3 *Watershared*

This thesis is based on work undertaken together with the non-governmental organization *Fundación Natura Bolivia* (*Natura*). *Natura* is an environmental non-governmental organisation founded in 2003 and based in the city of Santa Cruz de la Sierra in eastern Bolivia. It considers its mission to be to support communities to conserve their water sources through conservation of their forests (see *naturabolivia.org*), and its main program is known as *Watershared* in English, or *Acuerdos Recíprocos por Agua* (*Reciprocal Water Agreements*) in Spanish. This intervention is now being implemented in around 40 municipalities in 5 of the 9 departments of Bolivia, but for this thesis we focus on

Watershared as it is implemented in the region of the Santa Cruz Valleys (the easternmost extent of the Andean mountain range near the city of Santa Cruz de la Sierra).

As the name suggests, *Watershared* is based on reciprocity between downstream water users and upstream 'water providers' (Asquith, Vargas & Wunder 2008; Asquith 2016). Water providers are landowners who own forested land upstream in catchments, which is locally considered to be responsible for maintaining the supply of water in adequate quality and quantity to downstream communities. In the Santa Cruz Valleys, dry-season water flows have fallen by 50% in the past 20 years. This has been damaging to livelihoods in much of the area, as they are based on high-intensity irrigated agriculture and horticulture on fertile land in valley bottoms. Many local people blame deforestation by colonists in the catchment headwaters for this change and consider these forests to be 'water factories', a perception encountered in much of the tropical Andes (Murtinho *et al.* 2013; Rodríguez-de-Francisco & Budds 2015). This is despite the fact that scientific evidence on the forest-water relationship is unclear, difficult to study (Santos de Lima, Krueger & García-Marquez 2017), highly location-dependent (Biggs, Carpenter & Brock 2009), and almost certainly not that (as is widely believed) greater forest cover results in greater total water quantity (Bruijnzeel 2004; Le Tellier, Carrasco & Asquith 2009; Lele 2009; Ponette-González *et al.* 2015).

Watershared, as traditionally conducted, therefore involves downstream water users each financially contributing a modest monthly amount (generally less than 1 USD) to a water fund, which also receives contributions from the municipal government in the location in question and the NGO *Natura*. This money is then used to deliver incentives in the form of goods, such as beehives, cement, barbed wire, or fruit tree saplings, to forest owners in the upper catchments in return for an agreement not to deforest that land during the following three years. These incentives are relatively small (around 1 USD equivalent per hectare of forest placed in conservation per year) and are delivered annually upon compliance with the signed agreement.

Natura has historically been somewhat ambiguous as to whether or not they consider the *Watershed* intervention to be PES (Asquith & Wunder 2008; Butler 2012; Asquith 2016). It involves no markets, trading of credits, or buying of ES 'units'. The incentives that are provided are relatively small (Wunder, Engel & Pagiola 2008), and *Natura* makes no attempt to quantify and then compensate opportunity costs of forest conservation for landowners. Nor do they attempt to calculate and subsequently provide upstream landowners with the marginal value of the ecosystem services that they provide to those downstream. Thus the agreements (as the Spanish name suggests) are based upon reciprocity, 'nudge' or behavioural economics, and to a certain extent altruism, rather than the transactional model implied by much PES theory (Asquith 2016). Sociologists have analysed how the program's framing places it at the intersection of market-based, reciprocity-based, and redistribution-based policy (Bétrisey & Mager 2015), and it 'piggy-backs' on top of, rather than attempts to replace, pre-existing norms of reciprocity relating to environmental management (figure 1.1; Bétrisey & Mager 2015; Grillos 2017; c.f. Rode, Gómez-Baggethun & Krause 2015). This is not however unique to the case of *Watershared* (Kosoy *et al.* 2007). The Bolivian national government has also been overtly hostile to PES in general and to REDD+ in particular, claiming it to be a neoliberal and privatizing model of conservation, and does not allow such programs to exist within Bolivia. *Watershared*, has been cited in the Bolivian government's Conjoined Mechanism for Adaptation and Mitigation to climate change as an example of successful non-market-based conservation taking place within the country (T. Vidaurre, pers. comm.).

Figure 1.1. Visual representation of a reciprocal agreement: slide taken from the presentation used to present the *Watershared* agreements to community members. It describes them as a version of an *'aine'* or *'ayni'* (also see Grillos 2017) – a locally used term meaning an agreement between two or more people to help each other; thus, a reciprocal agreement.

Despite this, *Watershared* does largely meet the definition of a PES program according to Wunder (Wunder 2015), as it is a *conditional* transaction based upon rewards for a particular land use believed to provide benefits for parties other than those actually implementing the intervention. Also, *Watershared* has been considered a payments for watershed services program by many third parties in the academic and policy literature (Martin-Ortega, Ojea & Roux 2013; Bétrisey & Mager 2015; Engel 2016). Because of this I do consider it to be a form of PES (or at the very least a PES-like program; Robertson & Wunder 2005), while noting that it is a very long way from a market-framed PES program predicated on neoclassical economic theory.

1.4 The *Área Natural de Manejo Integrado Río Grande – Valles Cruceños*

The *Área Natural de Manejo Integrado* (*ANMI*) Río Grande – Valles Cruceños is a protected area of 734,000 hectares located in the Andean region of Santa Cruz Department (Dirección de Áreas Protegidas del Gobierno Autónomo de Santa Cruz *et al.* 2009). It was created in 2007 and its protection was ratified by the Autonomous Departmental Government of Santa Cruz in July of 2012 (Gobierno Autónomo Departamental de Santa Cruz 2012). The area is located between two National Parks, Amboró to the north and Iñao to the south, thus forming part of the Amboró-Tariquía conservation corridor. It is extremely biodiverse, counting as flagship species the red-fronted macaw (*Ara rubrogenys*) endemic to the dry valleys of Bolivia's Andes, and Sunkha palm (*Parajubaea sunkha*) as well as the Andean spectacled bear (*Tremarctos ornatus*) and Andean condor (*Vultur gryphus*), but it is believed to contain a total of over 2000 plant species and 362 bird species (Dirección de Áreas Protegidas del Gobierno Autónomo de Santa Cruz *et al.* 2009) in five very different ecoregions (Ibisch *et al.* 2003). This ecological and biological diversity is a consequence of the variation in altitude throughout the region, ranging between 700 and 3000 metres above mean sea level. As a consequence, temperature and precipitation also varies greatly across the area (Antúnez *et al.* 2009).

As well as this exceptional ecological diversity, the area contains a number of sites of global historical and cultural interest. These include the pre-Inca religious complex and World Heritage site known as *El Fuerte de Samaipata*, and the *Ruta del Che*, a number of locations associated with the final 1966-67 campaign of the Argentine-Cuban revolutionary leader Ernesto "Che" Guevara, including where he was captured and executed by the Bolivian army (Gobierno Autónomo Departamental de Santa Cruz 2017).

The area's legal protection is intended to conserve this biodiversity, but also to provide the ecosystem service of hydrological regulation and reduced erosion and sedimentation through conservation of the forest cover within the area (Gobierno Autónomo Departamental de Santa Cruz 2012). Despite this statute explicitly prohibiting any activity threatening or modifying the flow of the Río Grande, the national government of Bolivia is planning to construct a 600-megawatt hydroelectric dam (the *Represa Rositas*) across the Río Grande. This would flood a number of communities within the protected area, including some of the *Ruta del Che* sites, and would irreversibly alter the ecology of the area (e.g. Jemio 2017).

The area contains a number of towns and villages and was estimated to have a total population of 19,499 in 2008 (Antúnez *et al.* 2009). Most people living in the area are farmers, with a mixed farming system consisting of agriculture (maize, potato and wheat production), horticulture, and livestock rearing. Most landowners own cattle which are managed through an extensive grazing system, with the cattle living in the forests for at least part of each year. While cattle have both economic and cultural importance in the area, they also have major negative externalities associated with their presence. They degrade the forest and prevent its regeneration through both eating it and trampling it, destroy access roads to communities in the wet season preventing vehicles from entering or exiting, and defecate in water bodies including those which supply communities with drinking water, reducing water quality and increasing prevalence of water-borne diseases within the communities within the area (Paredes & Isurza 2012).

Local people are well aware of this issue. Therefore, cattle exclusion from water bodies, particularly those supplying households or communities with water for drinking, cooking and sanitation, has been practised for decades in the region (e.g. Asociación Zabalketa de Cooperación y Desarrollo 2008). This may be conducted and/or paid for by individuals, community-based water committees, or by NGOs, such as by the *Instituto de la Capacitación del Oriente*, another organization which works in the area (http://www.ico-bo.org/).

1.5 *Watershared* in the *ANMI Río Grande*, and its evaluation using a Randomised Control Trial: Case Study

Watershared in the ANMI Río Grande was implemented as a randomised control trial (RCT). One hundred and twenty-nine eligible communities within five of the municipalities in the area (Vallegrande, Pucará, Postrervalle, Moro Moro and Samaipata) were stratified by municipality, community size, and cattle density. Sixty-five of these communities were then randomly allocated to a treatment group, which from August 2011 was offered a version of the *Watershared* program as well as an environmental education presentation relating to the importance of forest conservation and cattle exclusion, and the other sixty-four to the control group, which initially was offered the environmental education presentation only. *Natura* set *Watershared* up as an RCT in the ANMI Río Grande as a pioneering study of impact evaluation in conservation, but also they did not have adequate funding to implement the program in all communities straight away in 2011 and wished to do so without favouritism or any possibility of undue influence determining which communities received the program. Subsequently, after endline data collection, *Watershared* was offered to members of all communities from September 2016. Figure 1.2 shows a map of the community locations within the *ANMI Río Grande* and their allocations to treatment or control. Appendix A contains a list of communities allocated to treatment and control groups.

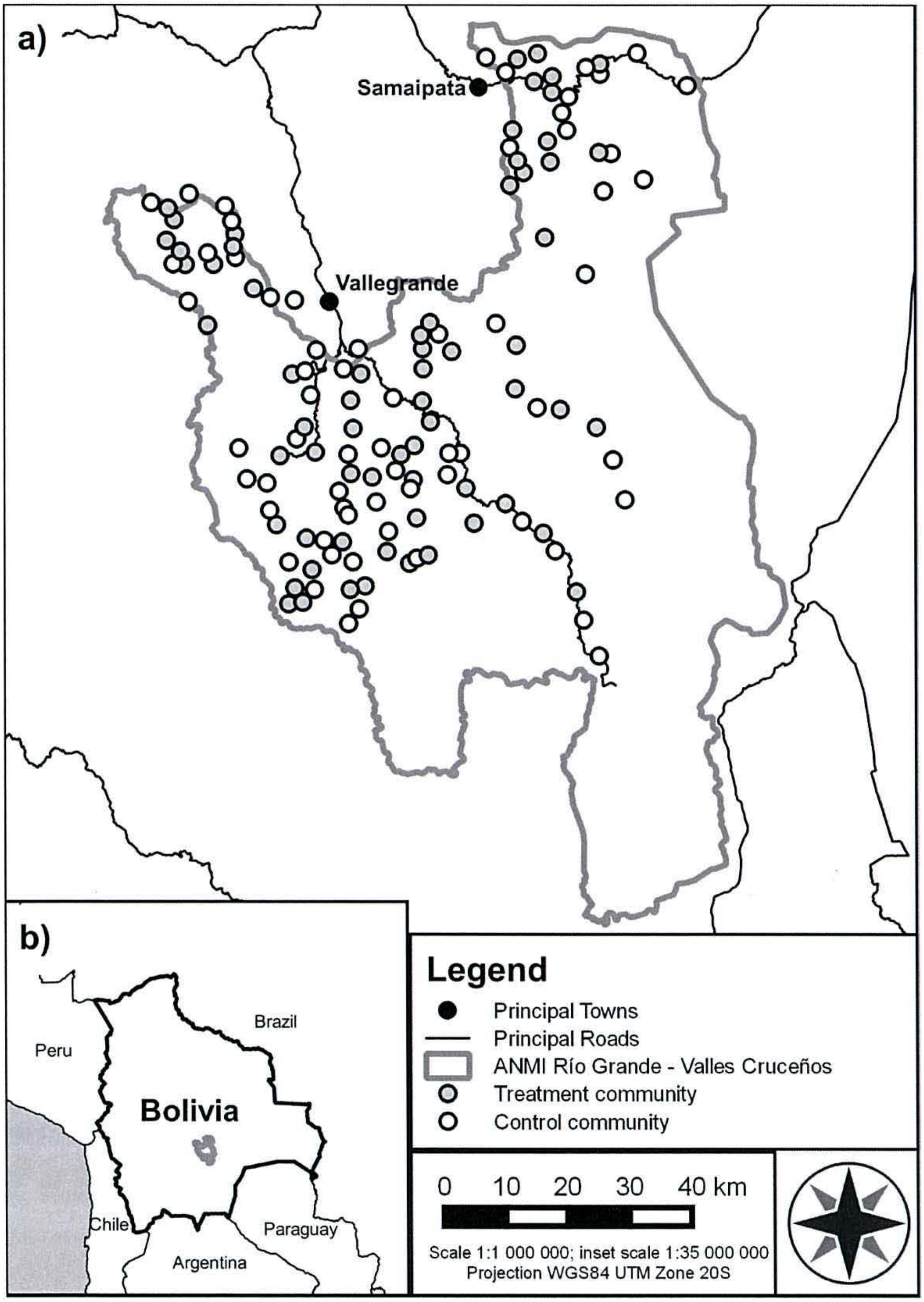


Figure 1.2. a) Locations of the 65 treatment and 64 control communities included in the RCT. b) Location of the *ANMI Río Grande – Valles Cruceños* protected area within Bolivia.

The way in which *Watershared* was implemented in the *ANMI Río Grande – Valles Cruceños*, and hence the intervention that we study in this thesis, was slightly modified from the basic model described in section 1.3 above. The *Watershed* program as described above was offered to landowners throughout the area. However a second version of the program was also offered from 2011 to members of the 65 treatment communities within the area, in which both forest conservation and exclusion of cattle and other livestock from watercourses were incentivised. This was known within *Natura* as the 'experimental ARA', in contrast to the 'traditional ARA' described in section 1.3 which has been implemented since 2003 throughout many parts of Bolivia where *Natura* works. We will refer to this 'experimental' program as *Watershared* from now onwards in the thesis unless otherwise specified.

This version of *Watershared* involved offering conservation agreements to landowners at three different 'levels' within these 65 communities. Increases in land use restrictions were compensated by increased value of incentives offered. Level 1 agreements (the highest level), for which landowners had to exclude cattle from riparian forests, aimed not only to store forest carbon, regulate water availability, and conserve biodiversity as in previous versions of *Watershared*, but also to improve water quality and to reduce stream bank erosion and levels of waterborne disease in communities. A complete theory of change, as we conceptualise it, is shown in figure 1.3. Similar programs with related goals have been implemented in various parts of Latin America, and in Mexico conserved riparian forest was estimated to provide a value of 90 USD per hectare per year in reduced disease burden and other ecosystem services (Mokondoko, Manson & Pérez-Maqueo 2016).

Agreements had a duration of three years, after which landowners could agree to extend them if they wished to do so (in the case of level 1 or level 3 agreements). The text of a level 1 agreement is shown in appendix B. The details of the agreement levels, their obligations, and the value of incentives provided, are shown in table 1.1. Agreements were offered twice annually. Upon signing the agreement, landowners were compensated with their selected goods (the list available is shown in appendix C). Funding for incentives was not provided through a water fund to which community members contributed due to the lack of water cooperatives and other institutions within the communities, as well as the fact that such payments might not have been accepted. Instead, funding for incentives came directly from the NGO and the municipal governments. Thus the reciprocal link between ecosystem service users and providers was somewhat weaker than previously. Two further rounds of incentives were then delivered upon compliance with the agreement's conditions, after rounds of compliance monitoring. In practice, non-compliance was only sanctioned with non-delivery of incentives in cases of major non-compliance (essentially deforestation and/or heavy cattle presence in the case of level 1 agreements); in cases where the landowner was persistently non-compliant but

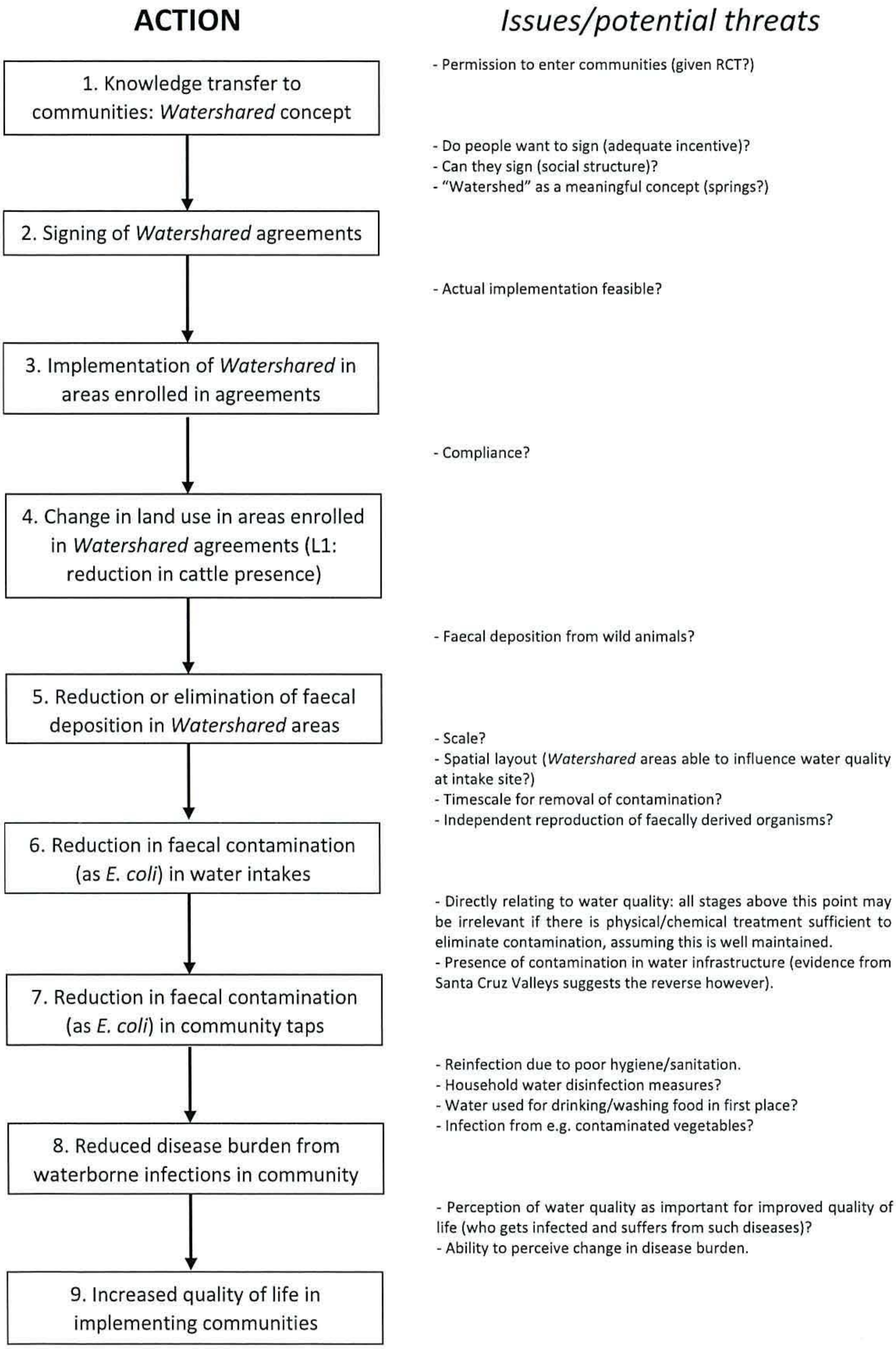


Figure 1.3. The *Watershared* theory of change relevant to the intervention as implemented in the *ANMI Río Grande*, and potential threats to its validity.

the incentives were consumable, they had to repay the equivalent value in cash. Returned goods from sanctions were returned to the community to which the landowner belonged rather than the NGO. By the end of 2014, members of 498 households had each signed at least one *Watershared* conservation agreement; non-compliance serious enough to result in the removal of incentives only occurred in eight cases.

Table 1.1. Characteristics of *Watershared* agreements in the *ANMI Río Grande – Valles Cruceños*.

Level of agreement	**1**	**2**	**3**
Principal activities required	No deforestation; no mineral extraction; no dumping of waste; absence of livestock from area (particularly cattle).	No deforestation; no mineral extraction; no dumping of waste; commitment to reduce livestock presence in area.	No deforestation; no mineral extraction; no dumping of waste.
Eligible land	Forested land within 100m of a watercourse /water body	Forested land within 100m of a watercourse /water body	All non-agricultural and undeveloped land
Timescale of agreement	3 years	3 years	3 years
Incentive value (per year)	10 USD/hectare, plus 100 USD initially†	3 USD/hectare	1 USD/hectare
Compliance monitoring	Transects within area under agreement	Transects within area under agreement	Forest change analysis from satellite imagery
Agreement renewable	Yes	No‡	Yes

†The 100 USD equivalent incentive is a flat one-off payment, regardless of the size of the area placed in conservation, and is not available on renewal after three years. It is intended to allow the landowner to construct troughs away from the conserved watercourses in order that their cattle or other livestock may continue to drink.

‡The fact that the level 2 agreement is not renewable implies that by the end of the three-year agreement period, the landowner must either be prepared to remove all livestock (in order to sign a level 1 agreement) or convert the agreement to the lower level of level 3 if this is not feasible (or to take the land out of conservation altogether).

The areas of greatest interest for *Natura* and for evaluators were the level 1 areas, as these were expected to provide all the ecosystem services which were the objectives of the program. The level 1 areas were intended to also be additional, as *Natura* expected landowners to have to construct fencing in order to exclude livestock from *Watershared* areas (which otherwise it would have access to). Achieving additionality is a major difficulty with almost all PES or PES-like programs. 'Rational' landowners would be expected to take advantage of informational rent to enrol areas which comply with land-use requirements, but they *would have complied anyway* even without the program, thereby not providing any additional ecosystem services (although the extent to which such behaviour may be considered problematic or undesirable can vary by program: the Costa Rican national PES program does not consider achieving additionality an objective of the program design and is not mentioned in the law framing the program itself [Porras *et al.* 2013]).

It is becoming clear that a predictor of program effectiveness is low levels of compliance with the action before program implementation (Börner *et al.* 2017). In the case of *Watershared*, in practice many landowners enrolled areas of land which cattle could not access anyway, therefore not being required to do anything to comply and so having no additionality. Subsequent analyses showed that areas of each of the 3 levels had a modest but significant level of additionality associated with them (Bottazzi *et al.* 2018).

1.6 Research objectives, questions, methods, and structure of the thesis

There are two principal objectives of the research in this thesis, with a number of research questions associated with each. We describe these here:

1) To examine the use of randomised control trials in evaluating large-scale socio-ecological conservation interventions.

- With reference to a review of the literature from a number of fields in which RCTs are more widely used, which factors may make such randomised evaluations in conservation more or less appropriate, feasible, and of high quality?

- How, and to what extent, were these issues dealt with in the case of the RCT of the *Watershared* intervention?

- Using RCT and PES theory as well as subsequent geospatial analysis of the *Watershared* program's implementation, what lessons were learnt from difficulties with conducting analyses of the program? How can implementers learn from this in designing future RCTs of PES-like evaluations?

2) To evaluate the effectiveness of the *Watershared* incentive-based conservation program in improving outcomes of interest related to water quality (including health and disease), taking advantage of the randomised evaluation setup where appropriate.

- Using the RCT setup and microbial water quality monitoring data from baseline and endline, did the *Watershared* program improve water quality at the community scale?

- Using the more comprehensive and reliable endline water quality data only, what predictors significantly determine water quality as measured?

- Using both household survey response data and water monitoring data, what predicts waterborne disease levels in communities amongst children? To what extent do factors related to *Watershared* and its associated actions achieve this change?

Concluding, we discuss what we learned regarding the effectiveness of the design of *Watershared*, the extent to which the RCT setup contributed to that learning, and how careful planning may make RCTs more likely to achieve their intended objectives, with reference to a new RCT of *Watershared* being conducted in the Chaco of Tarija in the south-east of Bolivia.

The thesis is structured as follows:

The Introduction above provides background to the themes which we investigate and analyse in this thesis.

The second chapter, "What role should Randomised Control Trials play in providing the evidence base underpinning conservation?", reviews the existing literature on the use of Randomised Control Trials in other fields where they have been widely used and investigates the extent to which they have potential for use in conservation and environmental management. This literature review was conceptualised by Edwin Pynegar and was then written by Edwin Pynegar, James Gibbons, Nigel Asquith and Julia P. G. Jones. It examines which factors may affect their utility, feasibility and quality, and illustrates this throughout with a discussion of the situation in the RCT of *Watershared* in the ANMI Río Grande. This is shortly to be resubmitted to a journal under the same title.

The third chapter, "Impact of Payments for Watershed Services on water quality: an evaluation using a Randomised Control Trial", consists of an analysis, using the randomised control trial setup, of the impact of *Watershared* on water quality in community water supplies, one of the key outcomes of interest for the RCT. It was based on the RCT designed originally by Professor B. Kelsey Jack (Tufts University) and Nigel Asquith, and the monitoring protocol used designed by Nigel Asquith, José Luis Isurza, and Sandy Rojas Banegas. The research design and monitoring protocol were then modified in 2014 and 2015 by Edwin Pynegar, James Gibbons and Julia P. G. Jones. The chapter therefore uses a 2010 baseline and 2015 endline to study relative rates of change in *Escherichia coli* concentration (our monitored metric of contamination) between treatment and control community sites. We also use the 2015 endline data, which contains a more comprehensive set of environmental characteristics, to predict which features of water intake sites are associated with elevated *E. coli* concentration. This enabled us to establish a more complete understanding of the effects of the intervention. We go on to discuss implications of the findings for the design of *Watershared* and possible modifications to it that may be worth exploring. This chapter is shortly to be resubmitted to a journal under the title "An evaluation of the impact of Payments for Ecosystem Services using a Randomised Control Trial", authored by Edwin Pynegar, Nigel Asquith, James Gibbons and Julia P. G. Jones.

The fourth chapter, "Does Payments for Watershed Services improve health in communities where it is implemented?", extends the work on water quality to investigate the links between *Watershared* and health outcomes in the communities. This chapter was conceptualised by Edwin Pynegar with help from James Gibbons, Nigel Asquith and Julia P. G. Jones, and then written by these four authors. The research partially used survey data collected by Patrick Bottazzi and David Crespo Rocha and their team in an endline household survey taken between October 2015 and June 2016 relating to water supply and management and reported waterborne disease prevalence. We then also took *E. coli* and other coliform concentration measurements from selected communities' water systems. We again find that a number of features of environmental context predict *E. coli* levels in community water supplies, and that *E. coli* levels at taps predict diarrhoeal disease levels amongst children. However we did not find that actions directly related to the *Watershared* intervention resulted in lowered probability of disease. This chapter is currently in preparation for publication.

The fifth chapter, "The devil in the detail: experiences from the implementation of a large-scale socio-ecological Randomised Control Trial" examines challenges which we encountered during the process of evaluating the intervention using the randomised control trial setup and analyses them in an empirical, quantitative way. It was conceptualised by Edwin Pynegar and written by Edwin Pynegar, James Gibbons, Nigel Asquith and Julia P. G. Jones. Specifically we investigate how selection of randomisation unit is likely to be highly problematic in any socio-ecological RCT evaluating multiple outcomes of interest, how monitoring and measurement has implications for potential for the intervention to detect changes as well as risking attrition, and how the nature of the intervention – not just voluntary, but voluntary in the extent and location to which it was implemented – results in meaningful evaluation via RCT being difficult. This chapter is currently in preparation for publication. While the chapter answers questions related to the first research objective, the thesis is structured with it subsequent to the chapters (3 and 4) related to the second research objective. We feel this to be more appropriate, as a reflection on the RCT's quality and the challenges that we encountered after evaluating the program's effects on outcomes of interest.

The sixth chapter, "Land use and conservation of catchments supplying water to communities: two case studies", is a short additional chapter included to present two additional observations (with accompanying data) which we believe support the conclusions both of the preceding chapters and of the thesis as a whole. This chapter was written by Edwin Pynegar. Records from community health centres show a substantial difference in attended cases of gastrointestinal disease between communities, which is associated with presence of *E. coli* in water systems. We also show how a change in water supply source in one of these communities led to a spike in gastrointestinal disease, and local knowledge suggested the land use in the new catchment above the intake was responsible for this. We also conducted measurements of water quality in a catchment where long-term forest conservation has been conducted, and found that despite this, the level of contamination remained high.

The seventh chapter, the Discussion, examines the ways in which the design of *Watershared* as implemented in the *ANMI Río Grande* influenced its outcome in achieving its intended goals and the implications of this for PES design more generally, the extent to which the RCT setup *per se* was necessary and sufficient for learning from *Watershared*, and how this learning has then been used in planning and undertaking a new RCT of a modified version of *Watershared* in the Chaco of Tarija department.

As the same or similar methods were used in a number of the chapters, we repeat some material describing the case study and the methods in certain parts of the thesis. This is because the thesis chapters have been presented as stand-alone papers.

1.7 The use of existing data and links to other research projects

We became involved with the evaluation of *Watershared* in the ANMI Río Grande in the autumn of 2013. By this time, the RCT described above had been implemented, and the *Watershared* agreements had been offered to landowners in treatment communities since August of 2011. Therefore, as discussed in section 1.6 above, we were not involved in designing and setting up the RCT evaluation initially. A number of datasets had been collected by *Natura* and collaborators, including a baseline household socio-economic survey, a baseline biodiversity survey, and three rounds of water quality monitoring (taken in 2010, 2012 and 2014). We also had access to *Natura*'s GIS databases showing the locations of areas placed under agreements.

In collaboration with *Natura* we collected the water quality data in 2015 and 2016 which was used in the third and fourth chapters of this thesis. We modified and elaborated the protocol, principally by incubating the samples at 35-37^{0}C as specified by the manufacturer (Micrology Labs 2016). In 2016 we also collected additional information relating to the ecological status of the catchments upstream of water monitoring sites. We also mapped centres of communities using a number of datasets.

The work described in this thesis was conducted in parallel with other research led by Dr. Patrick Bottazzi (Bangor University). An endline household socio-economic survey was conducted as part of this work, into which I added questions relating to the potential for 'copying' spillover effects (see chapter 5, *Methods*). We also used the community boundaries drawn by David Crespo Rocha from data produced by the Bolivian government agency *Instituto Nacional de la Reforma Agraria* (*INRA*), and a map of forest cover change in the ANMI Río Grande from 2011 to 2016 produced from Sentinel-2 satellite imagery by Dr. Rémi d'Annunzio (FAO).

We use the pronouns 'I' and 'we' differently throughout this thesis. Chapters 2 through to 5 were conceptualised and written by a number of authors (Edwin Pynegar, Nigel Asquith, James Gibbons and Julia P. G. Jones, with support from Tito Vidaurre, María Teresa Vargas, and others) and thus the pronoun used is 'we'. The Introduction, Discussion and chapter 6 were conceptualised and written principally by Edwin Pynegar only and so the pronoun used here is 'I' unless otherwise appropriate.

2 What role should Randomised Control Trials play in providing the evidence base underpinning conservation?

2.1 Abstract

There is general agreement that conservation decision-making should be evidence-informed, but many evaluations of intervention effectiveness do not attempt to account for confounding variables and so provide weak evidence. Randomised Control Trials (RCTs), in which experimental units are randomly allocated to treatment or control groups, offer an intuitive means of calculating the effect of an intervention through establishing a reliable counterfactual and avoid pitfalls of alternative quasi-experimental approaches. However, RCTs may not be the most appropriate way to answer some kinds of evaluation question, are not feasible in all circumstances, and factors such as spillover and behavioural effects risk prejudicing their quality. Some of these challenges may be greater in situations where the intervention aims to influence ecological outcomes through changing human behaviour (socio-ecological interventions). The external validity of RCT impact evaluation has also been questioned. We offer guidance and a series of criteria for deciding when RCTs may be a useful approach for evaluating the impact of conservation interventions, and what must be considered to ensure an RCT is of high quality. We illustrate this with examples from one of the few RCTs of a socio-ecological intervention – an incentive-based conservation program in the Bolivian Andes. Those who care about evidence-informed environmental management should aim to avoid a re-run of the polarized debate surrounding RCTs' use in fields such as development economics and take a pragmatic approach to impact evaluation, while also actively integrating learning from these fields. If this can be achieved, they will have a useful role to play in robust impact evaluation.

2.2 Introduction

Land managers, policymakers and other stakeholders make decisions about how ecosystems should be managed. There are increasing calls that such decisions should be firmly rooted in robust evidence (Sutherland *et al.* 2004; Segan *et al.* 2011; Baylis *et al.* 2016). Reasons why current decisions may not be evidence-based include decision makers' lack of access to evidence (Pullin *et al.* 2004) and inertia to changing established practices (Sutherland *et al.* 2004). However there are also clear limitations in the available evidence on the likely impacts of potential conservation interventions in a given situation (Ferraro & Pattanayak 2006; Pattanayak, Wunder & Ferraro 2010).

Impact evaluation (described by the World Bank as assessment of changes in outcomes of interest attributable to specific interventions; Independent Evaluation Group 2012) requires a counterfactual: an understanding of what would have occurred without that intervention (Margoluis *et al.* 2009;

Miteva, Pattanayak & Ferraro 2012; Ferraro & Hanauer 2014; Baylis *et al.* 2016). It is well recognized that simple before-and-after comparison of units exposed to the intervention is flawed, as some factor other than the intervention may have caused the change in the outcome of interest (Ferraro & Hanauer 2014; Baylis *et al.* 2016). Comparing groups exposed and not exposed to the intervention is also flawed as the groups may differ in other, potentially unobserved, ways that affect the outcome.

One solution is to replace simple post-project monitoring with more robust quasi-experiments, in which a variety of approaches may be used to construct a counterfactual scenario statistically. *Statistical matching*, including *propensity score matching*, involves comparing outcomes in units where an intervention is implemented with outcomes in similar (statistically selected) units lacking the intervention. This is increasingly used for conservation impact evaluations such as determining the effectiveness of a sustainable agriculture program (Margoluis *et al.* 2001) and in investigating the impact of national park establishment (Andam *et al.* 2008) or Community Forest Management (Rasolofoson *et al.* 2015) on deforestation. Other quasi-experimental approaches include *instrumental variables* (where easily observable variables correlated with the intervention but not the outcome are used as a proxy for the treatment), the *regression-discontinuity* approach (which compares outcomes of interest in units just above and below an initial eligibility criterion for implementation of the intervention:; as the criterion is arbitrary, units on either side will be essentially identical other than in implementation of the intervention), and *difference-in-differences* (which compares changes in outcomes in units exposed to an intervention with changes in a comparison group which was not exposed). Butsic *et al.* (2017) provide much more information on quasi-experiments' use in a conservation context.

Quasi-experiments should, and increasingly do, have a major role to play in conservation impact evaluation, and in some situations will be the only robust option available to evaluators. Their use has become substantially more common in recent years, which should be greatly welcomed, and meta-analyses of the effectiveness of certain interventions have recently begun to be published based upon quasi-experimental analyses (Samii *et al.* 2014; also Börner *et al.* 2016). However, because the intervention is not allocated at random, unknown differences between experimental and control groups may bias quasi-experiments' results (e.g. Michalopoulos, Bloom & Hill 2004). This problem, known as unobserved heterogeneity, historically led many in development economics to question their usefulness (e.g. Leamer 1983; also Levitt & List 2009; Angrist & Pischke 2010).

Randomised Control Trials ('RCTs'; also referred to as Randomised Controlled Trials) offer an outwardly straightforward solution to the limitations of other approaches to impact evaluation. By randomly allocating from the population of interest those units (individuals, areas or communities) which will receive a particular intervention (the 'treatment group'), and those which will not (the 'control group'), there should be no substantial differences in the types of unit that are in the treatment group when compared with the control group (e.g. White 2013). Evaluators can therefore assume that in the absence of the intervention, the outcomes of interest would have changed in the same way in the two groups making the control group a valid counterfactual for measuring the effect of the intervention can be calculated. Complete balance in all characteristics between treatment and control groups can only be guaranteed with extremely large sample sizes (e.g. Bloom 2008). However baseline data collection, stratification, and checking for balance between treatment and control groups can greatly reduce the probability of unbalanced groups (Glennerster & Takavarasha 2013) and if differences remain this can be resolved through its inclusion as a covariate in subsequent analyses (Senn 2013). In any program, there may be a difference between the units which were potentially exposed to the intervention (all units in the treatment group) and those actually exposed (a sub-set of the intervention group). This arises because many interventions are voluntary and take-up will not be 100%, or units may fail to comply or drop out for many reasons. Evaluators therefore often calculate both the mean effect on units in the intervention group as a whole (the 'intention to treat') and the effect of the actual intervention on a treated unit (the 'treatment on the treated', e.g. Glennerster & Takavarasha 2013).

The relative simplicity and intuitiveness of RCTs may make them particularly appealing to policymakers, especially when compared with the statistical 'black box' of quasi-experiments, and this may make them more persuasive than other impact evaluation methods to sceptical audiences (Banerjee, Chassang & Snowberg 2016). While the different kinds of quasi-experiment have associated with each of them a large number of assumptions in order for the counterfactual to be valid, and indeed the validity of the effect size estimate for any such quasi-experiment may be dependent upon the extent to which those assumptions are met, experimental evaluations such as RCTs avoid many of these problems and thus in some ways are conceptually simpler than quasi-experiments (Glennerster & Takavarasha 2013). RCTs are also substantially less dependent on theoretical understanding of *how* the intervention might or might not work. Interpretation of results and separate issues associated with experimental evaluations, many of which we go on to discuss in this thesis, may remain challenging, and there may be some cases in which such issues affect RCTs more than quasi-experiments. We discuss these issues here, in chapter 5, and in the Discussion.

RCTs are central to the paradigm of evidence-based medicine: since the 1940s tens of thousands of RCTs have been conducted and they are often considered the 'gold standard' for testing treatments' efficacy (Barton 2000). They are also widely used in agriculture, education, social policy (Bloom 2008), labour economics (List & Rasul 2011), and, increasingly over the last two decades, in development economics (Banerjee & Duflo 2011; Glennerster & Takavarasha 2013). The governments of both the United Kingdom and the United States have strongly supported the use of RCTs in evaluating policy effectiveness (Haynes *et al.* 2012; Council of Economic Advisers 2014). The United States Agency for International Development explicitly states that experimental impact evaluation provides the strongest evidence, and alternative methods should be used only when random assignment is not feasible (USAID 2016). However there are both philosophical (e.g. Cartwright 2010) and practical (Deaton 2010; Deaton & Cartwright 2016) critiques of RCTs' use, and their recent spread in development economics has led to a polarized debate (e.g. Ravallion 2009; Picciotto 2012). This debate notwithstanding, some development RCTs have acted as a catalyst for the widespread implementation of interventions. A now classic RCT testing treatment of parasitic worm infection on health and educational outcomes in Kenyan schoolchildren (Miguel & Kremer 2004) has led to the creation of initiatives such as Deworm the World (http://www.evidenceaction.org/dewormtheworld/) and the consequent treatment of over 95 million children.

Calls for the use of RCTs in evaluating environmental interventions have been increasing (Greenstone & Gayer 2009; Pattanayak 2009; Miteva, Pattanayak & Ferraro 2012; Samii *et al.* 2014; Ferraro & Hanauer 2014; Baylis *et al.* 2016; Curzon & Kontoleon 2016; Börner *et al.* 2016). Many kinds of conservation interventions aim to deliver ecological outcomes through changing human behaviour through incentive structures or rules (e.g. agri-environment schemes, provision of alternative livelihoods, protected areas, payments for ecosystem services, and certification schemes). We term these *socio-ecological interventions*. There are clear lessons to be learnt from RCTs in development economics, which also aim to achieve development outcomes through changing human behaviour and therefore face similar issues. A few pioneering RCTs of such large-scale socio-ecological interventions have recently been concluded, evaluating: an incentive-based conservation program in Bolivia (described in this chapter; also see Grillos [2017]); a payment program for forest carbon in Uganda (Jayachandran *et al.* 2017); and unconditional cash transfers in support of conservation in Sierra Leone (Kontoleon *et al.* 2016). We expect that RCT evaluation in conservation will become more widespread in the coming years.

We examine the potential of RCTs in developing the evidence base supporting (or otherwise) use of conservation interventions and thereby supporting evidence-informed decision making. We discuss the factors influencing the usefulness, feasibility, and quality of RCT evaluation of conservation and aim to provide insights for researchers and practitioners interested in conducting high-quality evaluations. The structure of the chapter is mirrored by a checklist (figure 2.1) which can be used to assess the feasibility of an RCT in a given context. We also illustrate these points throughout the chapter with the implementation of the recent RCT of the incentive-based conservation program *Watershared* by the NGO *Fundación Natura Bolivia* (*Natura*) in Bolivia (figures 1.2 and 2.2).

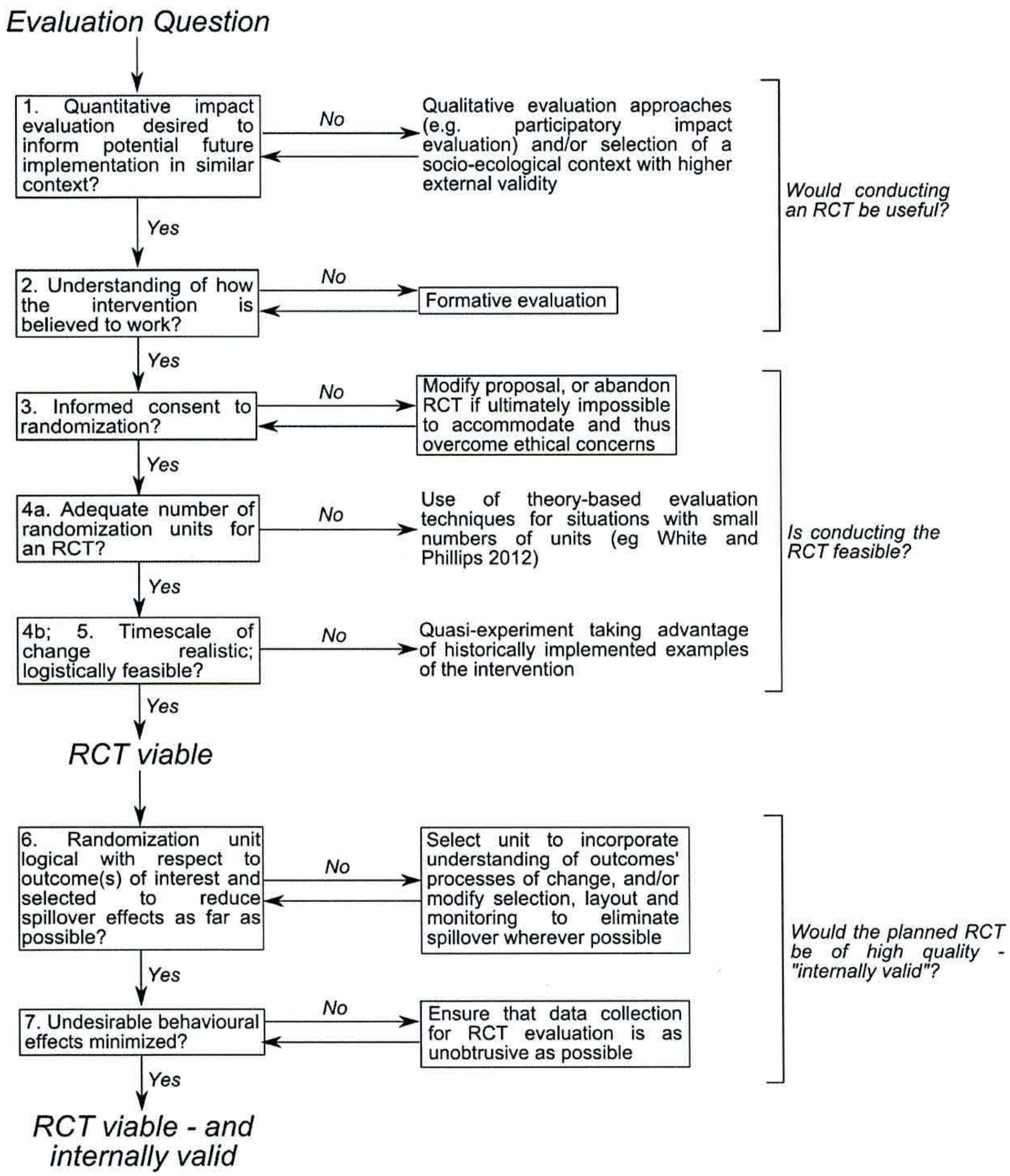
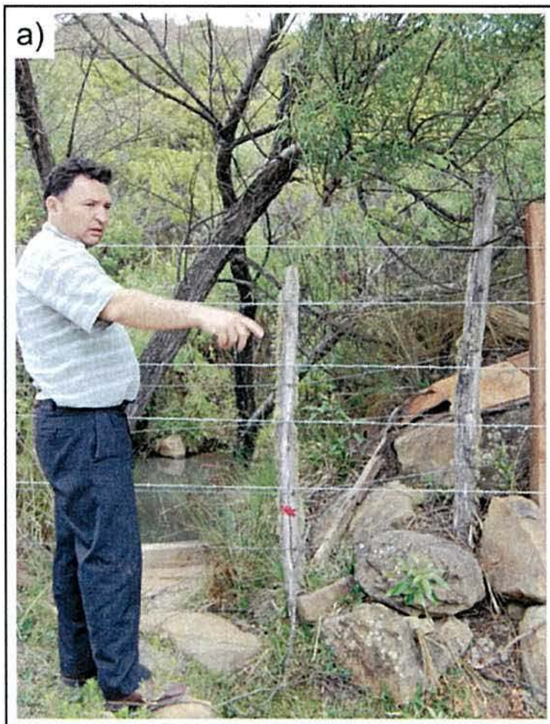


Figure 2.1. Summary of our suggested decision-making process for evaluators relating to RCT feasibility and quality, and alternative evaluation options if RCTs are inappropriate. Decisions or actions for evaluators to take during the process of RCT design are in boxes. Pattanayak (2009), Stern *et al.* (2012) and White & Phillips (2012) are good introductions to the alternative evaluation methods mentioned.



The non-governmental organisation *Fundación Natura Bolivia* has conducted a large-scale RCT in the Andean forests of the *Área Natural de Manejo Integrado Río Grande - Valles Cruceños* protected area in the Santa Cruz Department, Bolivia, evaluating the impacts of the *Watershared* incentive-based conservation intervention on environmental, health, and socio-economic indicators. Landowners were offered in-kind incentives for not deforesting and for preventing cattle from gaining access to and contaminating watercourses or springs. 65 of the 129 included communities in the study area were offered incentives plus an environmental education program (the treatment group), while the other 64 (control) communities received the environmental education program only.
Photographs: *Fundación Natura Bolivia*.

Figure 2.2. The Bolivian NGO *Fundación Natura Bolivia* conducted an RCT of their PES-like conservation program, *Watershared*, in the Bolivian Andes between 2011 and 2016. a) Water source fenced off to prevent cattle access. b) Community meeting in which incentive-based conservation is explained to local landowners.

2.3 Under what circumstances might an RCT evaluation be useful?

2.3.1 RCTs quantitatively evaluate an intervention's impact in a particular context

Many different approaches can be used to evaluate an intervention's impact. We focus on quantitative approaches, which allow the magnitude of the effect of an intervention on outcomes of interest to be estimated, as is often required by policy makers. However, evaluators should bear in mind that more qualitative approaches such as participatory or theory-based impact evaluation methods (e.g. Stern *et al.* 2012) might be more suitable in cases where the intervention was implemented in very few units (White & Phillips 2012) or when evaluators seek a detailed understanding of the pathways of change from intervention through to outcome (Cartwright 2010). RCT results indicate *whether* an intervention works and to what extent, but policymakers may also wish to know *why* it works, to allow prediction of project success in other contexts.

This issue of *external validity* – the extent to which knowledge obtained can be generalized to other contexts – is a major focus of the debate surrounding RCT use in development economics (e.g. Deaton 2010; Cartwright 2010). Advocates for RCTs accept such critiques as partially valid (White 2013), but note that RCTs provide complementary, not contradictory knowledge to other approaches to impact evaluation. Additionally the question of whether learning obtained in one location or context can be applicable to another is an epistemological question common to much applied research and is not limited to RCTs (Glennerster & Takavarasha 2013).

Solutions to the external validity problem include conducting qualitative studies alongside an RCT (researchers will inevitably develop an understanding of the causal processes involved anyway), or using covariates to explore which factors influence outcome. The most obvious solution, however, is to conduct RCTs of the same kind of intervention in different socio-ecological contexts (White 2013). While this is challenging due to the spatial and temporal scale of RCTs evaluating socio-ecological interventions, a number of groups of researchers have recently undertaken RCTs of incentive-based conservation programs (Kontoleon *et al.* 2016; Jayachandran *et al.* 2017; as well as the RCT described in this thesis). A study consisting of six separate RCTs on three continents, with over 10,000 participants in total, which evaluated a multifaceted development approach targeted at extremely poor households (Banerjee *et al.* 2015), has shown that multiple simultaneous RCTs of an intervention can be conducted (and in this case the pattern of lasting positive effects on income and assets was found across all countries).

In Bolivia, the NGO *Natura* wished to evaluate quantitatively the effects of the *Watershared* intervention (an incentive-based Payment for Ecosystem Services-like program) on water quality, biodiversity indicator species, deforestation rates, and human wellbeing. Similar socio-ecological systems exist throughout Latin America and incentive-based forest conservation projects have been widely implemented in montane forested regions. *Natura* is currently undertaking a complementary RCT of the intervention in the drier Bolivian Chaco (where land is held communally by indigenous people) and is in the process of designing a third, in a different part of the Chaco, which will evaluate, amongst other questions, the relative effectiveness of framing the intervention as a Payments for Ecosystem Services program or as a reciprocal agreement on its eventual outcomes. Additionally, in follow-up surveys at the end of the evaluation period, researchers have also extensively used qualitative methods to understand more profoundly processes of change within treatment communities.

2.3.2 **RCTs are likely most usefully conducted when the intervention is well developed**

Impact evaluation is a form of summative evaluation (Scriven 1967), meaning that it involves measuring outcomes. This can be contrasted with formative evaluation, which develops and improves the design of an intervention. Many evaluation theorists recommend a cycle of formative and summative evaluation, by which interventions may progressively be understood, refined, and evaluated (Rossi, Lipsey & Freeman 2004). This is similar to the thinking behind adaptive management

(Lindenmayer & Likens 2009). Summative evaluation alone is somewhat inflexible as once started, aspects of the intervention cannot be changed. The substantial investment of time and resources in an RCT is therefore likely to be most appropriate when implementers are confident that they have an intervention whose functioning is reasonably well developed and understood (Pattanayak 2009; Cartwright 2010). Again, outputs from formative and summative evaluation represent complementary and not contradictory knowledge.

In Bolivia, *Natura* has been undertaking incentive-based forest conservation in the Bolivian Andes since 2003, and cattle exclusion from water sources had been conducted in the region for decades by another NGO and by local communities. Lessons learnt from these experiences were integrated into the design of the *Watershared* intervention as evaluated by the RCT which began in 2010.

2.4 What affects the feasibility of RCT evaluation?

2.4.1 Ethical challenges

Randomisation involves withholding the intervention from the control group so the decision to randomize is not a morally neutral one. A central ethical principle in medical RCTs is that to justify a randomised experiment, there must be significant uncertainty surrounding whether the treatment is in fact better than the control (a principle known as equipoise). The mechanisms through which an environmental intervention is intended to result in changes are often complex and poorly understood, meaning that in environmental RCTs there may indeed be uncertainty about whether the treatment is better than the control. Additionally, it is unclear whether obtaining equipoise should even always be an obligation for evaluators (e.g. Brody 2012), as how well – not just whether – an intervention works, and how cost-effective it is, are also important results for policymakers. It may be argued that lack of availability of high-quality evidence leading to resources being wasted on ineffective or only modestly effective interventions is also unethical (List & Rasul 2011). Decisions such as these are not solely for researchers to make and must be handled with sensitivity (White 2013).

Another central principle of research ethics states that no-one should be a participant in research without giving their free, prior and informed consent. Depending on the scale at which the intervention under evaluation is implemented, it may not be possible to obtain consent from every individual in an area. This can be overcome by randomising by community or administrative unit (not by individual) and then giving individuals the opportunity of opting into or out of the offered intervention. This may result in challenges for interpretation as the level at which the intervention is implemented (the individual) is different from the level at which the randomisation is conducted.

In Bolivia, the complex nature of the socio-ecological system, and the lack of initial understanding of the ways in which the intervention might affect or not affect it, meant there was real uncertainty about the effectiveness of *Watershared* on outcomes of interest. However, had monitoring shown immediate significant improvements in water quality in the experimental communities, *Natura* would have stopped the RCT and immediately implemented the intervention in all communities. Consent was granted by community leaders for the randomisation and individual households could choose to join the program or not.

2.4.2 Spatial and temporal scale

Larger numbers of randomisation units in an RCT allow reliable detection of smaller effect sizes (Bloom 2008). This is easily achievable in small-scale experiments, such as those studying the effects of nest boxes on bird abundance or of wildflower verges on farmland invertebrate biodiversity; such trials have been a mainstay of applied ecology for decades (c.f. Fisher 1935). However, increases in scale of the intervention will make RCT implementation more challenging. A large randomisation unit (such as a protected area) will mean few available randomisation units, increasing the effect size required for a result to be statistically significant and decreasing the experiment's power (Bloom 2008; Glennerster & Takavarasha 2013). Large randomisation units are also likely to increase costs and logistical difficulties. However we emphasise that this does not make such evaluations impossible; two recent RCTs of a purely ecological intervention – impact of use of neonicotinoid-free seed on bee populations – were conducted across a number of sites throughout northern and central Europe (Rundlöf *et al.* 2015; Woodcock *et al.* 2017). When the number of units available is extremely small, RCTs will clearly not be possible and evaluation methods based upon expected theories of change may be more appropriate (White & Phillips 2012).

For some interventions, measurable changes in outcomes may take years or even decades, due to long life cycles of relevant species and the slow and stochastic nature of many ecosystem changes. It is unlikely to be realistic for researchers or practitioners to set up and monitor RCTs over such timescales. In these cases RCTs are likely to be an inappropriate means of impact evaluation, and the best option for evaluators would likely consist of a well-designed quasi-experiment taking advantage of a historically implemented example of the intervention.

In the Bolivian case, an RCT of the *Watershared* intervention was feasible as the intervention units are relatively small (communities of 2 to 185 households) and baseline data allowed stratified random allocation of 129 communities to control or treatment. The RCT was run over 5 years (2011-2016). Effects on water quality should be observable over this timescale as cattle exclusion may result in decreases in waterborne bacterial concentration in under 1 year (Meals, Dressing & Davenport 2010). However impacts on biodiversity may be expected to take substantially longer.

2.4.3 Available resources

RCTs require substantial human, financial and organizational resources for their design, implementation, monitoring, and subsequent evaluation. These resources are over and above the additional cost of monitoring in control units, because RCT design, planning, and the subsequent analysis and interpretation require substantial effort. USAID advises that a minimum of 3% of a project or program's budget be allocated to external evaluation (USAID 2016), while the World Health Organization recommends 3-5% (WHO 2013). The UN's Evaluation Group has noted that the sums allocated within the UN in the past cannot achieve robust impact evaluations without major uncounted external contributions (UNEG Impact Evaluation Task Force 2013). Conducting a high-quality RCT is certainly not cheap; many conservation practitioners are already well aware of this (Curzon & Kontoleon 2016).

Collaborations between researchers (with independent research funding) and practitioners (with a part of their program budget allocated to evaluation) can be an effective way for high quality impact evaluation to be conducted. This was the case with the evaluation of *Watershared* in Bolivia*:* the NGO had funding for implementation of the intervention from development and conservation organizations while the additional costs of the RCT came from research grants and collaborations with universities. Additionally, there are a number of organizations whose goals include conducting and funding high-quality impact evaluations (including RCTs), such as Innovations for Poverty Action (www.poverty-action.org), the Abdul Latif Jameel Poverty Action Lab (J-PAL; www.povertyactionlab.org), and the International Initiative for Impact Evaluation (3ie; www.3ieimpact.org).

2.5 What factors affect the quality – the 'internal validity' – of an RCT evaluation?

2.5.1 Potential for 'spillover', and how selection of randomisation unit may affect this

Evaluators must decide upon the unit at which allocation of the intervention is to occur. In medicine the unit is normally the individual, although some interventions may be allocated to groups. In development economics units may be individuals, households, schools, communities, or other groups while in conservation units could also potentially include fields, farms, habitat patches, protected areas, or others. Units selected should, however, logically correspond to the process of change by which the intervention is understood to lead to the desired outcome (Glennerster & Takavarasha 2013).

In conservation RCTs, surrounding context will often be critical to interventions' functioning. This is also true of some RCTs in medicine or development economics, and hence evaluators can learn from these fields. Spatial context means that evaluators need to consider the potential for outcomes to 'spill over' between units – with positive effects from the intervention in treatment units affecting control units, or vice versa (Glennerster & Takavarasha 2013; Baylis *et al.* 2016). It is easy to imagine species of interest moving from one unit to another because of habitat connectivity or water flowing down from a treatment area to a control one. These kinds of spillover, which we refer to as *biophysical* as they relate to ecological processes, thus cause changes achieved in treatment areas to affect outcomes of interest in control areas and thus reduce an intervention's apparent effect size. If an intervention were to be implemented in all areas rather than solely treatment areas (presumably the ultimate goal for practitioners), such effects would not occur. Spillover is particularly likely to occur if the randomisation unit and the natural unit of the intended ecological process of change do not align, meaning in practice the intervention would be implemented in areas which would affect outcomes at control sites, and vice versa.

Spillover effects are thus a property of the trial itself, and are recognized as important in some situations in development economics. For example, the influential RCT investigating treatment of worm infection in Kenyan schoolchildren used schools as the randomisation unit as children in the same school are likely to interact and re-infect each other more frequently than with children at other schools. It was explicitly designed to allow measurement of spillover (Miguel & Kremer 2004); and showed (notwithstanding the re-analysis by Davey *et al.* [2015]) that deworming in treatment schools resulted in decreased worm burden in children attending nearby non-treatment schools. Such spillover also affected one of the very few attempts to conduct a large-scale environmental management RCT: the UK Government's RCT of badger culling in south-western England (Donnelly *et al.* 2005).

Preliminary consideration of spatial relationships between units, and the relationship between randomisation units and the process of change for the indicators, is critical for reducing or eliminating spillover and thus successfully undertaking internally valid conservation RCTs. Spillover may also be reduced by selecting indicators and/or sites to monitor which would still be relevant and meaningful but would be unlikely to suffer from spillover (such as by choosing a species to monitor with a small range size, or ensuring that a control area's monitoring site would not be directly downstream of a treatment area's in an RCT of a payments for watershed services program).

In the evaluation of *Watershared,* it proved difficult to select a randomisation unit that was politically feasible and worked for all outcomes of interest. *Natura* used the community as the randomisation unit as it would have been extremely difficult to have offered *Watershared* agreements to some members of communities and not to others. Community boundaries thus had to be drawn (these did not previously exist) and these did not always align well with area of land in the catchment of the communities' water sources. Thus while *Natura* did all it could to ensure that no community water quality monitoring site was directly downstream of another, land under conservation agreements in one community would sometimes be located in the catchment upstream of the monitoring site of another, risking biophysical spillover. We examine empirically the extent to which this spillover took place, and its consequences, in chapter 5.

2.5.2 Consequences of human behavioural effects on evaluation of socio-ecological interventions

There is a key difference between *ecological* interventions that aim to have a direct impact on an ecosystem and *socio-ecological* interventions which seek to deliver ecosystem changes by changing human behaviour. Medical RCTs are generally double-blinded so neither the researcher nor the participants know who has been assigned to the treatment or control group. Double-blinding is possible for some ecological interventions such as pesticide impacts on non-target invertebrate diversity in an agroecosystem: implementers do not have to know whether they are applying the pesticide or a control. This was partially achieved in the large-scale study of neonicotinoids cited above (Rundlöf *et al.* 2015). However, it is harder to carry out double-blind trials of the effects of socio-ecological interventions, as the intervention's consequences can be observed by the researchers, and participants will know whether they are being offered the intervention or not.

Lack of blinding creates potential problems. Participants in control communities may observe activities in nearby treatment communities and implement aspects of them on their own, reducing the measured impact of the intervention. They may, however, also feel resentful at being excluded from a supposedly beneficial intervention and therefore reduce pre-existing pro-conservation behaviours (Alpízar *et al.* 2017). It may be possible to reduce or eliminate such phenomena through selecting units whose individuals infrequently interact with each other. Evaluators of the *Watershared* program in Bolivia were concerned that members of control communities might decide to protect watercourses themselves after seeing successful results elsewhere (which would be encouraging, suggesting local support for the intervention, but which would interfere with the evaluation by reducing the effect size of the intervention detected). They therefore included questions in their follow-up socio-economic surveys to identify this effect; these revealed only one case in over 1500 household surveys.

The second issue with lack of blinding is that RCT design is intended to achieve that treatment and control groups are not systematically different immediately after randomisation. However those allocated to control or treatment may have different expectations or show different behaviour or effort simply as a consequence of the awareness of being allocated to a control or treatment group, meaning that a systematic difference between the two groups would have been introduced (Chassang, Padró i Miquel & Snowberg 2012). Hence the outcome observed may not depend solely on the efficacy of the intervention; some authors have claimed that these effects may be large (Bulte *et al.* 2014).

Overlapping terms have been introduced into the literature to describe the ways in which actions of participants in experiments vary due to differences in effort between treatment and control groups (summarised in table 2.1). The 'Hawthorne effect' describes the phenomenon that participants in an experiment may behave differently because they know that they are being studied (e.g. Levitt & List 2011). The 'Pygmalion' and 'golem' effects, in which participants may adjust effort to meet experimenter expectations, are a form of this (Babad, Inbar & Rosenthal 1982). Similarly, treatment-group interviewees may give answers that they believe evaluators wish to hear, known as experimenter demand. The related 'John Henry effect' may arise when individuals in control groups increase effort to compete with the treatment group (Saretsky 1972). In addition, it is rational for subjects to increase effort expended on implementing an intervention if they believe the intervention to be effective (Chassang, Padró i Miquel & Snowberg 2012). The consequence of these 'rational effort' effects can be that performance increases when people believe in the intervention (Babad, Inbar & Rosenthal 1982). Therefore, if an intervention appears to achieve a large change in an outcome of interest, that may be because true efficacy of the intervention was large, or because participants *believed* it to be large and thus expended large amounts of effort on implementing it.

We do not believe that potential behavioural effects invalidate RCT evaluation as some have claimed (Scriven 2008), as part of an intervention's impact in subsequent implementation will also be due to implementers' expended effort (Chassang, Padró i Miquel & Snowberg 2012). It remains unclear whether behavioural effects are large enough to result in incorrect inference, or even exist at all (Bausell 2015). In the case of the evaluation of *Watershared*, compliance monitoring is an integral part of incentive-based or conditional conservation, so any behavioural effect driven by increased monitoring should be thought of as an effect of the intervention itself rather than a confounding influence on outcome. Any such effects may be reduced through low-impact monitoring (Glennerster & Takavarasha 2013). In Bolivia, water quality measurement was unobtrusive (few community members were aware of *Natura* technicians being present) and infrequent (either annual or biennial); deforestation monitoring was even less obtrusive as it was based upon satellite imagery; and socio-economic surveys were undertaken equally in treatment and control communities.

2.6 Conclusions

Scientific evidence supporting an intervention's use does not necessarily lead to the uptake of that intervention. Policy is at best *evidence-informed* rather than *evidence-based* (Adams & Sandbrook 2013) because cost and political acceptability inevitably influence decisions, and frameworks to integrate evidence into decision-making are often lacking (Segan *et al.* 2011). However, improving available knowledge of intervention effectiveness is still important. For example, managers are more likely to report an intention to change their management strategies when presented with high-quality evidence of intervention effectiveness (Walsh, Dicks & Sutherland 2015). The potential for evidence to have influence is higher when it is driven by the needs of practitioners: links between researchers and policymakers or practitioners throughout the design and implementation of impact evaluation studies are therefore valuable (Cook *et al.* 2013).

RCTs can be used to establish a reliable counterfactual allowing robust estimation of intervention effectiveness, and hence cost-effectiveness, and interest in their use is increasing within the conservation community. Like any evaluation method, they are clearly not suitable in all circumstances, and there exist significant practical challenges with their implementation. Even when feasible, evaluators must design RCTs with great care to avoid spillover and behavioural effects and thus maintain internal validity. We would argue that it still remains unclear whether, to what extent, and in which contexts, RCTs are likely to provide estimates of treatment effects more accurate than quasi-experiments (c.f. Michalopoulos, Bloom & Hill 2004; Bulte *et al.* 2014), due to confounding experimental effects. This research question deserves a great deal more attention. There also will inevitably remain some level of subjectivity whether a location or context for subsequent implementation of an intervention is similar enough to one where an RCT was carried out to allow the learning to be confidently applied. We hope that those interested in evaluating the impact of conservation interventions can avoid the polarization and controversy surrounding their use in development economics while learning from their implementation in other fields. RCTs may then make a substantial contribution towards building a more robust evidence base to underpin conservation decisions.

Table 2.1. Consequences of behavioural effects when compared with results obtained in a hypothetical double-blind RCT. Hawthorne '1', '2' and '3' refer to the three kinds of effect discussed in Levitt & List (2011). References: [a] - (Jakovljevic 2014). [b] - (Rosenthal & Jacobson 1968). [c] - (Babad, Inbar & Rosenthal 1982). [d] - (Levitt & List 2011). [e] - (Orne 1962).

Effect name	Description/Explanation	Other names	Effect on outcome in treatment units	Effect on outcome in control units	Effect on estimated effect size of intervention
'Hawthorne 1'	Act of observation increases effort	-	Increases	Increases	Unknown
'Hawthorne 2'	Changes in intervention increase effort	Halo effect of uncontrolled novelty[a]	None / Increases	None	None / Increases
'Hawthorne 3'	Experimental subjects tend to meet what they believe to be experimenters' expectations	Pygmalion effect[b]; golem effect[c]; Rosenthal effect[a]; experimenter demand[d]; demand characteristics[e]	Increases	None / Decreases	Increases
Rational effort	Experimental subjects base effort on their own expectations of the intervention's effectiveness	Galatea effect[c]	Increases	None / Decreases	Increases
'John Henry'	Individuals in control group increase effort in an attempt to compete with the intervention group	-	None	None / Increases	None / Decreases

3 Impact of Payments for Watershed Services on water quality: an evaluation using a Randomised Control Trial

3.1 Abstract

Payments for Ecosystem Services (PES) aim to incentivize land users to manage their land in ways which benefit society. However, as with many complex socio-ecological interventions, robust evaluation of PES is challenging and rare. We evaluate whether a conservation program in the Bolivian Andes, which incentivizes landowners to avoid deforestation and exclude cattle from riparian forests, delivers improvements in microbial water quality (as measured by *Escherichia coli* contamination), using a Randomised Control Trial (RCT). One hundred and twenty-nine communities were randomly allocated to a treatment or control group following baseline data collection in 2010. Endline data were collected in 2015. Although *E. coli* contamination was higher in control communities in 2015, this difference pre-existed the intervention. Presence of cattle faeces adversely affected water quality, showing the effectiveness of excluding cattle, but the intervention did not have a demonstrable effect at the landscape scale. This is likely due to landowners often not enrolling the most important land from a water quality perspective, so linkages between the incentivized intervention and the desired ecosystem service are weak. Program effectiveness is fundamentally an empirical question, and this pioneering RCT shows their potential in robustly evaluating large-scale conservation interventions and contributing to the evidence base available to decision-makers.

3.2 Introduction

In the past two decades there has been growing interest in the potential of Payments for Ecosystem Services (or Payments for Environmental Services – the terms are largely interchangeable [Wunder 2015]) as an approach to improving the management of ecosystems in order to increase supply of valued services. PES programs aim to change the economic incentives that land managers face in supplying off-site environmental benefits from their land to make environmentally sound land uses more economically favourable (Engel, Pagiola & Wunder 2008). The focus of many PES programs is the maintenance or increased availability of good quality water in adequate quantity (e.g. Martin-Ortega, Ojea & Roux 2013), and there are plentiful examples of water-focused PES in Latin America (Martin-Ortega, Ojea & Roux 2013; Grima *et al.* 2016), and to a lesser extent in Asia and Africa (e.g. Calvet-Mir *et al.* 2015). However, despite the interest in the approach and the increasing number of real-world examples, there are very few robust evaluations of the extent to which PES programs deliver the services (such as clean water in adequate quantity) they seek to supply (Pattanayak, Wunder & Ferraro 2010; Miteva, Pattanayak & Ferraro 2012; Grima *et al.* 2016).

The importance of provision of clean water to global health and development is highlighted by its inclusion as goal 6 in the UN Sustainable Development Goals (United Nations 2015). However at least 1.8 billion people still rely on drinking water sources contaminated with faecal matter (Bain *et al.* 2014a). Many of these sources lack adequate physical or chemical treatment, and so the quality of the drinking water depends to a great extent on land use and ecosystem management around and upstream of those water sources. Hence provision of clean water can be considered as an ecosystem service or as a precursor to multiple ecosystem services benefiting society (Keeler *et al.* 2012). Faecal contamination of drinking water may cause a whole host of diseases to be transmitted, with pathogens of faecal origin including viruses, bacteria, protozoa and helminths. Gastrointestinal illnesses caused by consumption of such contaminated water are a major cause of mortality and morbidity in the developing world (Prüss-Ustün *et al.* 2014).

Escherichia coli is a bacterium that lives only in the guts of warm-blooded animals, and is thus widely used as a faecal indicator (Leclerc *et al.* 2001). While some strains of *E. coli*, such as O157:H7, are pathogenic (enterohaemorrhagic or enterotoxic *E. coli*), the majority are not but are associated with the presence of other harmful organisms found in faecal matter which are more difficult to culture and more hazardous for experimenters to handle (Ashbolt, Grabow & Snozzi 2001). Sources of such faecal contamination may include faulty sewerage systems and leaking septic tanks (Richards *et al.* 2016), open defecation by people (Spears, Ghosh & Cumming 2013), or the presence of wildlife (Ahmed *et al.* 2012). However, a major source of contamination is the presence of domestic livestock, particularly free-roaming cattle (Crane *et al.* 1983). Therefore, cattle exclusion has been practiced as a means of reducing faecal contamination of watercourses. In the UK, for example, the Good Agricultural and Environmental Conditions standard 1 requires farmers in receipt of certain subsidies to maintain buffer strips and refrain from spreading manure within areas close to water bodies (UK Government 2016). There is evidence of such actions being effective at significantly reducing *E. coli* concentration and other faecal contamination of water supplies (Sunohara *et al.* 2012) with associated timescales under 1 year (Meals, Dressing & Davenport 2010). However, many uncertainties remain about the extent to which these interventions, incentivized via a PES program, can deliver consistent benefits in water quality and consequent improvements in human health outcomes at the landscape scale.

Evaluating the effectiveness of environmental interventions is challenging and complex, and as a consequence there is little robust evidence of relative effectiveness or cost-effectiveness of many interventions (Pattanayak, Wunder & Ferraro 2010; Miteva, Pattanayak & Ferraro 2012). Randomised Control Trials (RCTs), in which experimental units are randomly allocated to treatment or control groups, allow the creation of robust counterfactuals from which to infer what would have happened in the absence of the intervention. Therefore, a subsequent comparison of outcomes of interest between the treatment and control units can be used to calculate the intervention's mean effect size. RCTs are highly promoted in many areas of public policy including medicine, labour economics, education, and recently in development economics (Glennerster & Takavarasha 2013; Council of Economic Advisers 2014). Although small-scale Randomised Control Trials have been a mainstay of applied ecological experiments for decades (Fisher 1935), there are very few examples of RCTs of large-scale socio-ecological interventions, i.e. those in which desired biophysical changes are mediated by human behavioural change, such as PES, and there have been calls for their increased use (Greenstone & Gayer 2009; Baylis *et al.* 2016). We know of only a single, very recently published Randomised Control Trial of a PES, evaluating the impact of a program in Uganda on deforestation (Jayachandran *et al.* 2017).

The Bolivian non-governmental organization *Fundación Natura Bolivia* (*Natura*) began using incentives (in kind rather than in cash) to encourage upstream landowners to conserve forests and water in the Andean region of Bolivia in 2003, and now has 210,000 hectares under conservation agreements with 4500 families (Asquith 2016). The intervention is known locally as *Acuerdos Recíprocos por Agua* (Reciprocal Water Agreements) and internationally as *Watershared* (Asquith 2016). There exist widespread perceptions locally that upstream deforestation has reduced dry-season flows in local rivers, making agriculture more marginal and climate-vulnerable, and that defecation from cattle causes contamination of water sources and thus affects the quality of drinking water and ultimately the prevalence of gastro-intestinal diseases (Paredes & Isurza 2012; Rojas Banegas 2012). Thus *Watershared* aims to protect the quantity and quality of water flowing to downstream communities through incentivizing landowners to cease deforestation and to prevent livestock from accessing watercourses. This is intended to contribute additionally to conserving the exceptional biodiversity of the region and the carbon stored in these forests. Although implementers do not use the term PES when describing the program (Asquith 2016), it meets the recent definition of PES published by Wunder (Wunder 2015): "voluntary transactions between service users and service providers that are conditional on agreed rules of natural resource management for generating offsite services".

Given the growing interest in evaluating effectiveness of different conservation approaches, *Natura* established *Watershared* in 129 communities of a newly established protected area as a Randomised Control Trial (RCT) to allow robust evaluation. We use this unique setup, described in more detail in *Methods* below, to investigate the effectiveness of the intervention at delivering improvements in microbial water quality. We address two interconnected questions: firstly, did the implementation of *Watershared* in treatment communities result in an improvement in water quality relative to control communities; and secondly, do the features of *Watershared* agreements (e.g. cattle exclusion, absence of faeces) have a measurable impact on water quality at a site, accounting for other predictors?

3.3 Methods

3.3.1 **Context and RCT design**

This article focuses on the *Watershared* intervention of *Natura* in the *Área Natural de Manejo Integrado (ANMI) Río Grande – Valles Cruceños*, a protected area of 7340km² in the Andean region of the Santa Cruz Department in eastern Bolivia (figure 1.2). Forests in this area are perceived locally as 'water factories' providing high-quality water for human and animal consumption and irrigation, despite the mixed scientific evidence on this topic (Bruijnzeel 2004; Ponette-González *et al.* 2015). Gastrointestinal illnesses are endemic; for example, in 2015 the health centre of Moro Moro, a community with a population of 743 in 2008 (Antúnez *et al.* 2009), treated 236 cases of diarrhoea (information from *Servicio Nacional Integral de Salud, Centro de Salud Moro Moro*, obtained 4th April 2016). Faecal contamination from cattle is widely considered an important contributor to the high prevalence of these diseases as the traditional farming system involves cattle grazing freely within the forests from where most communities take their water. While some communities have rudimentary sedimentation and filtration systems, these are of limited effectiveness and often become clogged with sediment after each rainfall event; chlorination or other chemical treatment is rare (R. Rueda, pers. comm.). Forest conservation and cattle exclusion have been conducted in the area since the 1980s (Robertson & Wunder 2005) with the aim of providing the ecosystem service of clean water.

In 2010, 129 communities within the *ANMI Río Grande – Valles Cruceños* were selected for inclusion in a Randomised Control Trial (figure 1.2), these being all the communities falling within the five main municipalities in the ANMI area (Vallegrande, Samaipata, Pucará, Postrervalle and Moro Moro). Consent to randomisation was granted by community leaders on the understanding that should the program be found to be effective, it would subsequently be implemented in all communities. Communities were randomly allocated to control (64 of these communities in which conservation agreements were not offered) or treatment (65 communities in which agreements were offered) groups following stratification based on municipality, community size, and estimated cattle density. The RCT was not blinded as participants inevitably knew whether they belonged to a treatment or control community. However, in order to avoid observer bias effects during data collection, those conducting water quality monitoring did not know which communities belonged to the treatment or control group.

Individuals belonging to treatment communities were offered the chance to conserve land belonging to them under *Watershared* agreements as well as receiving an education program on the importance of cattle exclusion and forest conservation for the maintenance of water quality and quantity. Individuals belonging to control communities were offered the same environmental education program only. The implementing NGO *Natura* offered landowners in treatment communities three-year conservation agreements to conserve upstream forest and exclude cattle in return for in-kind incentives such as fruit trees, beehives, barbed wire, or cement. Participants could enrol their land in one of three kinds of agreements (for details see table 3.S1). In this paper we only consider level 1 agreements, in which landowners are offered $10/hectare/year in-kind equivalent (plus the equivalent of $100 regardless of the size of the area enrolled) in return for conserving forested land within 100m of a watercourse and excluding cattle from these areas. Landowners were offered the opportunity to enrol their land twice per year, beginning in August 2011. Compliance monitoring and distribution of the in-kind compensations was conducted yearly.

Our analyses are based upon two rounds of monitoring of the quality of water intended for human consumption. A baseline was taken between February and July of 2010 by the NGO *Natura* before the sites were allocated to control or treatment groups. However the allocation was not stratified by measured water quality at the sites, and the baseline data was not otherwise used until our team (who started work on this project in 2013) examined it to explore whether there was evidence of pre-existing differences between sites. A more detailed endline monitoring round was undertaken between March and May of 2015, in the first wet season subsequent to completion of the first signed agreements, with more stringent protocols for handling water samples introduced by our team. In 2015 we also measured a number of other potential indicators of water quality.

The communities within the RCT are small (between 2 and 185 families in 2010) with diverse water supplies. Some have a single water intake, others multiple intakes and in a few cases no functional intake at all (community members take water directly from streams or other water bodies). Resource and logistical constraints meant that not all intakes and taps could be sampled and so the tap supplying the community's school, along with the intake supplying that tap, were taken as sampling sites based on the assumption that these would have the greatest importance for health (figure 3.1). In cases where the community had no school, we monitored at the intake which supplied the greatest number of households and a representative tap fed by that intake. In those cases in which the community had no functional water system at all, we took a sample in the water body where the greatest number of households took their water from. Thus most communities had two site measurements (intake and tap), whereas a few (those lacking an intake) only had one.

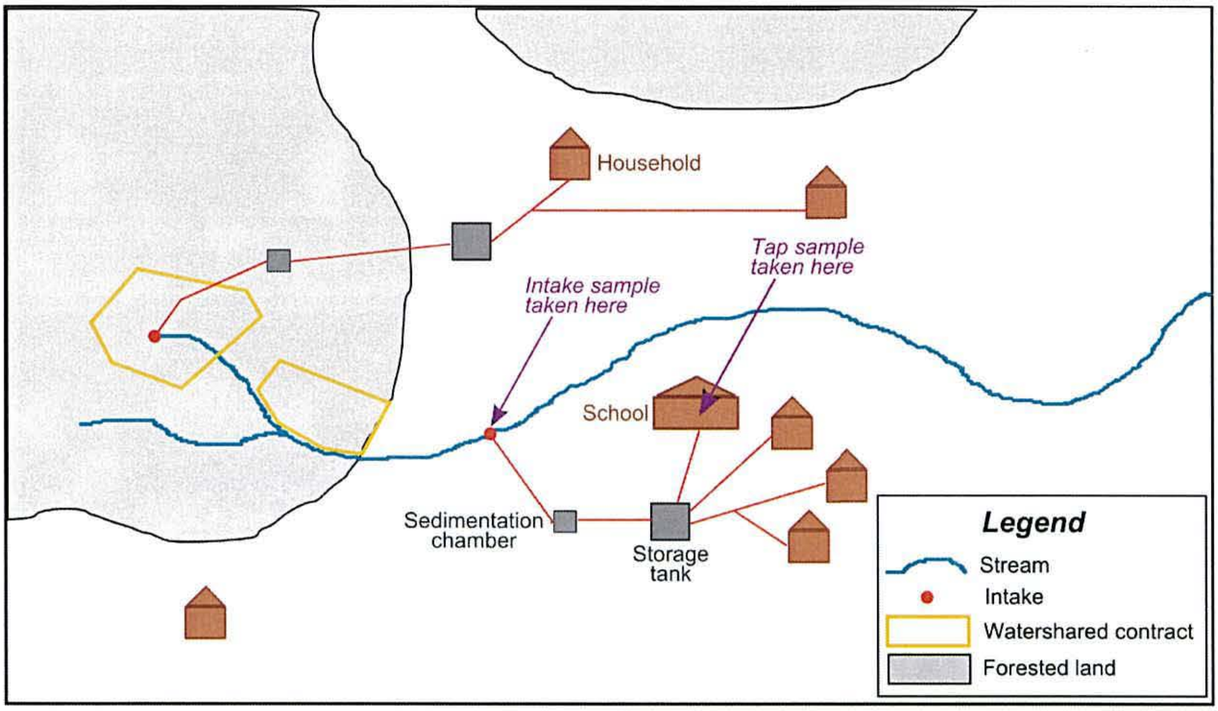


Figure 3.1. Schematic of an example community with 2 water intakes and 6 households, showing locations of sampling sites.

3.3.2 **Water quality monitoring**

In 2010 technicians from *Natura* monitored water quality in 241 sites in 125 communities. Independently, the randomisation process assigned 129 communities to control or treatment groups, 123 of which had been monitored at baseline. In 2015 monitoring was carried out by a team from Bangor University and *Natura*; we monitored 249 sites in 127 communities (123 of which were part of the RCT and therefore can be categorized as control or treatment); logistical problems meant the remaining baseline sites could not be visited[1]. At each site we monitored the presence of faecal contamination. In 2015 we also recorded other biophysical characteristics of the site which may predict water quality (including both physico-chemical properties of the water and disturbances in the surrounding environment).

[1]Between visits by the Natura technicians in 2010 and our team in 2015, the location of both water outtakes and the main tap serving the community had changed in a number of sites. In fact only 83 sites in 47 communities had remained at the same location between 2010 and 2015 and thus were directly comparable (table 3.S2).

The principal metric of contamination used was the concentration of *Escherichia coli* colony forming units (CFUs) in water samples. *E. coli* concentration, along with that of other non-*E. coli* bacteria belonging to the coliform group, was quantified using the Coliscan Easygel method (Micrology Labs, Goshen, IN, USA). Coliscan Easygel allows enumeration of coliforms as after incubation *E. coli* colonies appear purple, blue-purple or dark blue due to metabolism of both beta-galactosidase and beta-glucuronidase. Other non-*E. coli* coliforms are pink based upon metabolism of beta-galactosidase only. Colonies of a blue-green or sky blue colour (metabolism of beta-glucuronidase only) and white colonies were not counted (Micrology Labs 2016). The Easygel method (which uses only 5ml of water per sample) does not comply with the World Health Organization's 100ml standard for coliform monitoring but studies have shown that it is reasonably robust and not susceptible to false negatives (Chuang, Trottier & Murcott 2011). This method had been selected by *Natura* due to the logistical challenges with using alternative methods such as membrane filtration in the remote and low-resource context of the study area (many sites are reachable only difficult drives and long walks). When we modified the protocol in 2015 (to overcome some of the limitations of the 2010 protocol), we elected to retain the method for the same reason.

In 2010, one sample was placed into sterile Coliscan Easygel sampling flasks (35ml) taking care to avoid any external contamination. Up to two days later (but normally on the same day) the *Natura* team then inoculated Easygel Petri dishes using 5ml of the water from each flask. After solidification the Petri dishes were sealed and incubated at ambient temperature for 48 hours, after which numbers of *E. coli* and other non-*E. coli* coliform CFUs were counted. Petri dishes and Coliscan bottles were sterilized with bleach before being disposed of.

In 2015, four separate samples were taken using sterile Coliscan Easygel sampling flasks (35ml each) and placed on ice within 1 hour of sampling. Within 6 hours of sampling (although generally within 4) we produced Easygel Petri dishes using 5ml of water from each flask as inoculum. After solidification we sealed the Petri dishes and incubated them for 24 hours at 35-37°C in a portable incubator (NQ28 model, Darwin Chambers, St Louis, MO, USA). In locations where no mains electricity was available we maintained a constant incubation temperature through use of a 12V vehicle power supply or supply from a car battery. After incubation we counted *E. coli* and other non-*E. coli* coliform CFUs. Petri dishes and Coliscan bottles were subsequently sterilized by boiling for a minimum of 1 hour and then disposed of. We discuss the reasons for the selection and modification of this protocol (given the low-resource context in which the water quality monitoring was conducted) in appendix E, as a reference for researchers and practitioners interested in undertaking similar monitoring.

In 2015 we also measured in each site a number of physico-chemical parameters of water: temperature, dissolved oxygen in mg/l and per cent of saturation value, pH, salinity and conductivity in each site with an HQ40d portable multi-parameter meter and IntelliCAL LDO101, PHC101 and CDC401 rugged probes respectively (HACH Company, Loveland, CO, USA). We measured turbidity in formazin attenuation units through the use of a DR/850 colorimeter (HACH Company, Loveland, CO, USA). Additionally, at the intake sites, we recorded other variables that may predict *E. coli* concentration, including the presence or absence of cattle (judged based upon presence of faeces, hoof prints, or cattle paths recently used) and the presence or absence of cattle faeces in the riparian forest, in the water, or on banks. Some were recorded at the intake itself and others along a 10m transect upstream (uphill in the case of intakes in springs) of the intake. Details of all monitored variables are available in table 3.S3.

We used *Natura's* community database to determine which intakes supplied treatment or control communities (we did not have this information when conducting field sampling to avoid any observer bias effects). We used GIS software (ArcGIS 10.2, ESRI, Redlands, CA, USA) and *Natura*'s shapefiles to calculate which monitored intakes fell within land enrolled in *Watershared* agreements, and then used data held by *Natura* on which of these areas of interest were compliant with the requirements in the agreements. In a number of the earliest sites monitored during 2015 we disturbed the sediment in the water intake while taking samples; sites in which this happened were recorded as such.

3.3.3 **Statistical analysis**

We used generalized linear mixed models (GLMMs) to calculate the effect of the intervention on delivering improvements in water quality. Firstly we explored whether the implementation of *Watershared* resulted in a measurable improvement in water quality relative to control communities. To do this we calculated the difference in changes from 2010 to 2015 between treatment and control community sites, using RCT treatment status of the community to which the site belonged as a dummy variable. Secondly we used the much richer 2015 data to explore whether the features of *Watershared* (specifically cattle exclusion, absence of faeces) have a measurable impact on water quality at a site, accounting for other predictors.

We used the *glmmADMB* package in R (Fournier *et al.* 2012; R Development Core Team 2014; Skaug *et al.* 2016) to produce GLMMs predicting *E. coli* concentrations, specifying a negative binomial error structure and log-link. We included the water system identifier throughout as a random effect, as measurement at an intake and then a tap supplied by that intake represents repeated measures of the same water system. We used model selection based upon comparisons of the Akaike's Information

Criterion (AIC) and compared relative goodness of models through Akaike weighting. We determined 95% confidence intervals for predictors in the principal model of interest.

In the first set of models (those evaluating the difference in change from 2010 and 2015 between treatment and control sites), we only included sites where the intakes and taps have remained at the same location between 2010 and 2015 and the intake is unambiguously associated with a treatment or control community (site N=83, water system N=47). We estimated a GLMM with *E. coli* concentration in 2015 as the response variable and site treatment status (whether a site is in a control or treatment community) and 2010 *E. coli* concentration as potential predictors. We also included an interaction term between 2010 *E. coli* concentration and site treatment status (if this interaction were a significant predictor, this would represent a significant effect of the intervention on water quality). Given the different volume of water sampled in 2010 (5ml) and 2015 (20ml), we included an offset term of $\log_e(4)$ in each of the models to ensure equivalence between 2010 and 2015 *E. coli* CFU counts. To aid in interpretation of these results, we then ran further GLMs comparing relative levels *E. coli* concentration in 2010 with treatment or control community site status as a predictor. We did this for both consistent intake sites between 2010 and 2015 (site N=47) and for all intake sites unambiguously associated with a treatment or control community (site N=123).

In the second set of models (analysing the predictors of *E. coli* concentration at sites in 2015), we used all sites monitored which had a complete set of predictors, with the exception of a single site in the community of Torrecillas which is in the Río Mizque, a river with an associated catchment size of 9768km^2, meaning that site was qualitatively different from all other sites and thus was removed (final site N=219, water system N=124). For predictor variable selection we first removed closely correlated predictors and then classified variables we considered likely to be important in predicting *E. coli* concentration (table 3.1). We associated predictor data relating to intake features with both intakes and their respective associated taps. We then produced GLMMs for all purely biophysical traits of sites, while also including all 2-way and 3-way interactions between temperature, pH, and salinity. To determine intervention effectiveness, we then added features that related directly to the intervention (cattle access, whether the site was in a level 1 *Watershared* agreement, and faeces presence) and again conducted model selection based on AIC minimization. We subsequently compared this model with one including RCT status as a predictor (N=211, water system N=119; some sites were removed due to intakes not being unambiguously attributable to treatment or control communities).

Table 3.1. Variables hypothesized to be important in predicting *E. coli* concentration in 2015. Codes are used in subsequent model selection tables (tables 3.S3a, 3.S3b and 3.S4).

Variable	**Code**	**Classification**	**Base level**
Site type	ST	Intake; Tap	Intake
Intake category	IC	Stream; Spring	Stream
Sediment disturbance	SD	Undisturbed; Disturbed	Undisturbed
Intake substrate	IS	Rock only; with sand; with mud	Rock only
Cattle presence	C	Absent; Present	Absent
Agriculture presence	A	Absent; Present	Absent
Turbidity	Tu	Continuous; FAU/100	-
Temperature	T	Continuous	-
Salinity	S	Continuous	-
pH	pH	Continuous	-
Cattle access	CA	Yes; No	Yes
Faeces presence	F	Absent; Present in forest; Present in water or on stream banks	Absent
Compliant level 1 *Watershared* area	ARA	None; Intake entirely within conserved area	None
Water System RCT status	RCT	Control; Treatment; Unclassifiable	Control

To establish whether in practice land use differs between treatment and control communities, we also determined (for all intakes monitored in 2015 with data on cattle access and classifiable as treatment or control) whether relative proportions of intake sites protected from cattle differed between the treatment and control communities. We tested for a significant difference using a chi-squared test.

3.4 Results

3.4.1 The intervention had no significant effect on *E. coli* concentration once baseline levels were accounted for

E. coli concentration in 2015 appeared higher at sites associated with control communities than at sites associated with treatment communities. However, this pattern was also evident in the 2010 data, before communities were allocated to control or treatment groups (figure 3.2). *E. coli* concentrations measured in 2010 are consistently lower than in 2015. This pre-existing tendency towards lower *E. coli* concentration in treatment community sites was found if all sites are compared, rather than just those which remained in the same location between 2010 and 2015 and thus that are directly comparable (figure 3.S1). Comparisons of 2010 *E. coli* concentrations between treatment and control community intake sites show significantly lower levels in treatment sites for both consistent sites (N=47, p=0.012, generalized linear model [GLM]) and all sites (N=123, p=0.020, GLM with negative binomial error structure).

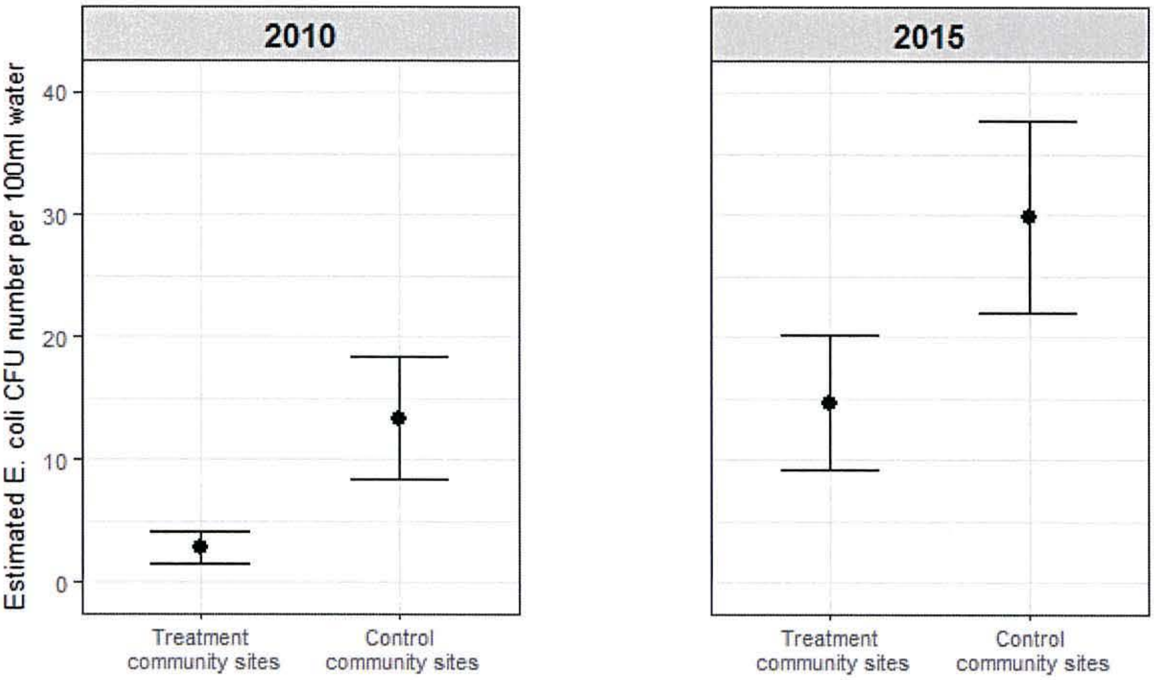


Figure 3.2. Relative *E. coli* levels in treatment and control sites in 2010 and 2015. *E. coli* CFU numbers shown per 100ml sample equivalent.

Once the pre-existing difference between control and treatment community sites is taken into account, there is no significant effect of a site being in a control or treatment community on *E. coli* concentration in 2015 (figure 3.3, table 3.S4a). There is also no significant interaction between RCT status and *E. coli* concentration in 2010, showing that rates of change in *E. coli* concentration between 2010 and 2015 are not significantly different in sites associated with treatment or control community (figure 3.4, table 3.S4a; site N=83, water system N=47; generalized linear mixed model [GLMM] with negative binomial error structure). Model selection shows that AIC decreases when this term is removed (table 3.S4b). This shows that, using the robust RCT design and both the baseline and endline datasets, there is no significant effect of the intervention on *E. coli* concentration at the landscape scale.

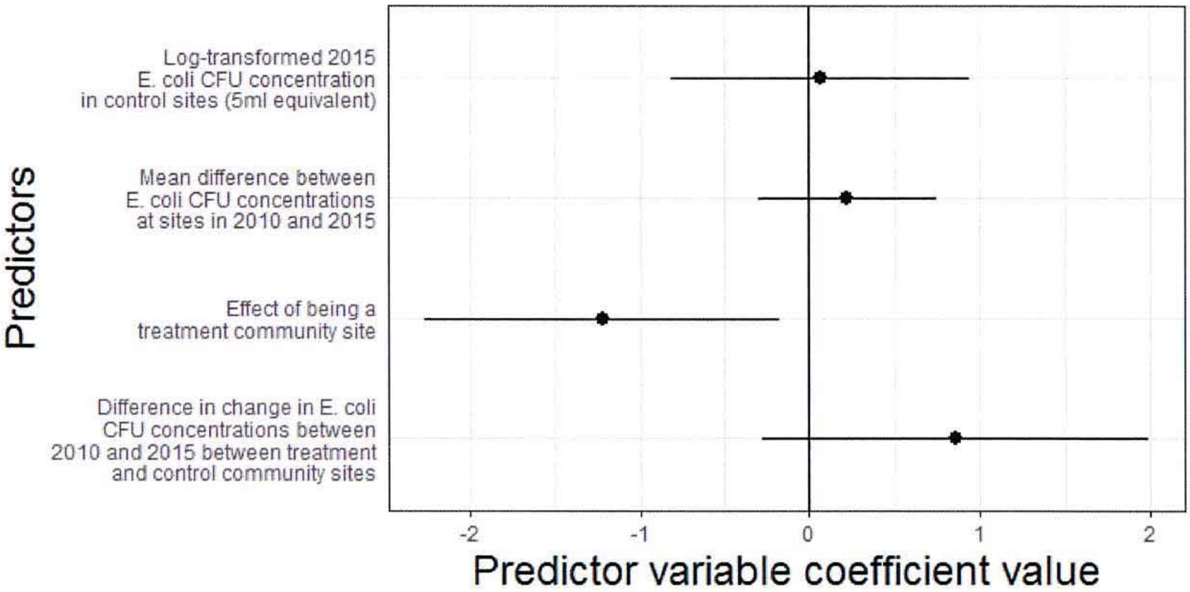


Figure 3.3. The effects of 2010 *E. coli* concentration and whether a site is in a treatment or control community on 2015 *E. coli* concentration. Error bars show 95% confidence intervals.

3.4.2 Cattle faeces in water is one of the significant predictors of *E. coli* concentration

E. coli concentration in 2015 is significantly predicted by a number of variables (figure 3.4; site N=219, water system N=124, GLMM with negative binomial error structure). The details of model selection can be seen in tables 3.S5a (for purely biophysical model selection) and table 3.S5b for model selection including parameters relating directly to the intervention. Intakes are significantly more contaminated than taps; sites associated with stream intakes are significantly more contaminated than sites associated with spring intakes; turbidity, and disturbance of the sediment by the research team during sampling are both also associated with higher recorded contamination. In terms of variables directly connected to the intervention, the presence of cattle faeces in or close to the water is a significant predictor of contamination, although faeces presence in the wider forest, while showing a positive trend, is not significant at 95% CI. Details of the model can be found in table 3.S5c.

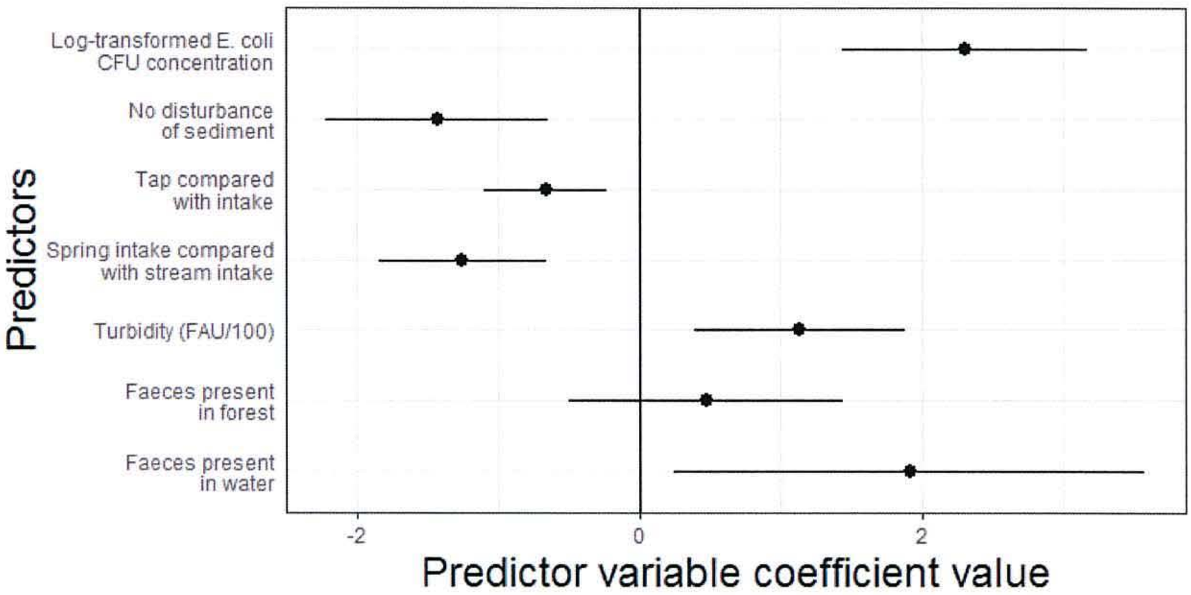


Figure 3.4. The effect of site features which are predictors of 2015 E. coli concentration in the most likely GLMM (model 16; see tables 3.S5a, b and c). Error bars show 95% confidence intervals.

We also found that including RCT status as a predictor shows that treatment community sites are significantly less contaminated than control community sites in 2015, and that inclusion of this in the model increases model explanatory power and reduces model AIC (table 3.S6; figure 3.S2).

3.4.3 Treatment and control communities do not differ with respect to cattle access to water intakes

Not all intakes in treatment communities are within sites that are protected under *Watershared* agreements and many intakes in control communities are protected from cattle despite the absence of such agreements (table 3.2). There is therefore no significant difference between the number of intakes with cattle access in control and treatment community sites (N=129; p=0.97; chi-squared test).

Table 3.2. Number and proportions of intakes visited in 2015 in treatment and control communities with and without cattle access.

	Treatment community intake	**Control community intake**
N	68	61
Compliant level 1 *Watershared* conservation agreement (%)	16 (24%)	0 (0%)
Sites with cattle access (%)	28 (41%)	24 (39%)
Sites with no cattle access (%)	40 (59%)	37 (61%)

3.5 Discussion

We found no evidence that the PES-like *Watershared* intervention, which aimed to improve water quality by excluding cattle from riparian forests, has had a significant impact on water quality at the landscape scale. While a simple comparison of *E. coli* concentration in 2015 (endline only) between treatment and control communities does show a lower level of *E. coli* concentration in treatment sites, a likely explanation for this is the pre-existing difference in *E. coli* level between control and treatment communities. Once this is taken into account, the difference disappears. We cannot explain why the randomisation conducted in 2010 to allocate communities to control or treatment did not achieve balance in terms of original levels of contamination. However, as demonstrated in this study, where baseline data is available from control and treatment groups, perfect balance between treatment and control groups is not necessary to achieve valid inference (Senn 2013).

The protocol used in monitoring in 2010 was relatively weak, which is the reason for the modifications when our team designed the 2015 endline protocol. However these baseline data, despite their weaknesses, provide some evidence that the difference in contamination levels between control and treatment sites pre-existed the intervention. While the difference between 2010 and 2015 *E. coli* concentrations was likely due to mistreatment of the samples in 2010, this was done equally to samples from both (future) treatment and control community sites, and so no bias between treatment and control communities would thus be expected to be introduced. Also, not using the 2010 data would have led us to a substantially different conclusion, superficially attractive in that it would suggest an endline difference between the treatment and control community sites in the treatment sites' favour. However, our other analyses and further data present no feasible nor realistic theory of change to explain it. Treatment community sites do not systematically differ in any trait associated with lower contamination (faeces presence, stream/spring intake site, etc.), there is no systematic difference in levels of protection between intervention and control intakes, and while there is a highly significant difference in levels of conserved areas between catchments of treatment and control community intakes, the absolute levels are still very low. Combined with the fact that the systematic difference did appear to pre-date the intervention, we would have felt it incomplete and insufficiently rigorous not to include these data, and we would certainly have not felt confident in claiming a significant difference between treatment and control directly attributable to the program. These data are therefore very important to the interpretation of the endline data. The problem with the baseline data collection highlights the importance of sufficient resources (including the need for interdisciplinary expertise and technical know-how as well as financial resources) being provided if robust evaluation is to be carried out (see section 7.2).

The more reliable 2015 data allowed us to show that presence of cattle faeces in water or on the stream banks does result in higher *E. coli* contamination at individual sites. This suggests that excluding cattle from water sources can indeed contribute to improving water quality. This should perhaps not be surprising given that fresh cattle faeces can contain more than 10^8 colony forming units per kilogram (Weaver, Entry & Graves 2005). However, the presence of cattle faeces is only one predictor of water quality. Intakes fed by streams were much more contaminated than those fed by springs. This is implied in the literature (Howell, Coyne & Cornelius 1995; WaterAid 2013) and in fact the Millennium Development Goal definition of an improved water source allows some springs to be considered improved without further chemical treatment while no stream or river intakes can (Bain *et al.* 2014b). We also found that intakes were more heavily contaminated than taps. This suggests that although the sedimentation and filtration chambers in many of the water systems may not always be effective, they have at least some positive effect on water quality. Turbidity was also an unsurprising important predictor; turbidity is well-known in the literature as a predictor of *E. coli* contamination (LeChevallier, Evans & Seidler 1981).

There is therefore an apparent paradox in that the program has not had an impact at the landscape scale, but that the intervention it seeks to incentivise (excluding cattle) does apparently improve water quality. We suggest a number of possible explanations.

First, field observations showed that the majority of land in catchments upstream of intakes was not enrolled in the *Watershared* program and hence was not under compliant conservation agreements. Moreover, the areas actually enrolled in conservation agreements at treatment sites were almost all small (1-10 hectares). Livestock-derived *E. coli* can enter water intakes through a number of routes including overland flow and movement of groundwater (Oliver *et al.* 2010), and not solely through direct deposition (one of the forms of point-source pollution) which is what the *Watershared* agreements attempt to prevent. The small areas conserved at or above the intakes may well have reduced faeces presence and so reduced *E. coli* concentration at these sites. However, upstream or uphill of these intakes contamination may have continued to enter water bodies through multiple routes. Freshwater sediments can act as an *E. coli* reservoir, implying that even if it were the case that flow routes were shut down, reductions in *E. coli* concentration should not be expected in the short term (Pachepsky & Shelton 2011). Instructively, evidence from a 25-year-old conservation area in the community of La Aguada, near to the *ANMI Río Grande* protected area, shows that despite the 20% of the catchment nearest to the intake being under conservation, the water remains contaminated (see chapter 6).

Second, the way in which the *Watershared* intervention is designed means there is no obligation, or even extra incentive, for landowners to conserve land surrounding or even in the same catchment as monitored intakes. Farmers are free to enrol any land which meets the criteria (forest within 100m of a stream or spring). Therefore the intervention is not spatially targeted towards critical areas, and in fact only 16 of the 68 water intakes in treatment communities are located inside enrolled and compliant parcels of level 1 land. Also, many communities have excluded cattle from water intakes independently of the *Watershared* program, meaning there is no significant difference in the proportion of intakes protected from cattle between treatment and control communities.

Third, it is possible that there has not been sufficient time between implementation of the intervention and subsequent evaluation for differences to become apparent. Some of the *Watershared* areas included in the analysis had been enrolled in the latter part of 2014, hence in some cases only a few months before the endline monitoring was undertaken. It is well known that *E. coli* may be able to live in sediment at the bottom of water bodies for long periods of time (Pachepsky & Shelton 2011; Cho *et al.* 2016) which will mean that any impact of the intervention could not yet be detectable.

Such mismatches in spatial and temporal scale between intervention implementation and biophysical processes leading to desired changes are examples of two issues frequently encountered in incentive-based conservation programs including PES (Jack, Kousky & Sims 2008). First, the link between the land use incentivized (the proxy) and the ecosystem service desired is weak and poorly understood. In the *Watershared* intervention, it is unclear how much land in the catchments of interest would have to be protected, where, and over what timescale, to obtain a significant improvement in water quality at the landscape or even the catchment scale. Second, the marginal benefits from service provision (or in this case the land use proxy for service provision) are also not spatially constant. Land enrolled in areas around or directly upstream of intakes will probably have an effect on monitored water quality while areas under conservation elsewhere (for example below the water intake) obviously cannot. In such cases, spatial targeting and differentiated payments could increase program efficiency.

For the *Watershared* intervention to have a significant effect on water quality, its design may have to be modified so that areas most valuable in terms of their potential ecosystem service provision are enrolled (ideally whole catchments upstream of water intakes). However such targeting is challenging because of the high informational costs in terms of identifying the land eligible for enrolment (Jack, Kousky & Sims 2008) and because of potential issues in terms of social acceptability and perceived fairness locally. The fact that several landowners may own land in one catchment further complicates such an approach (we discuss the implications of this further in chapters 5 and 7).

There is also likely to be a limit to the impact that livestock exclusion can achieve and this will depend on the extent to which faecal contamination derives from other sources such as wildlife, inadequate sanitation infrastructure, spreading of manure on agricultural land, or from open defecation. While there is little human habitation or infrastructure in most of the catchments (the communities are downhill of the gravity-fed intakes), contamination from wildlife may indeed reduce the efficacy of the intervention. Those involved in promoting similar interventions should check the extent to which cattle contamination is indeed the driver of microbial water quality issues in the region, perhaps using genetic testing of *E. coli* (Carson *et al.* 2001).

There are clear challenges to designing conservation interventions to deliver improvements in microbial water quality, including the likely limitations due to non-livestock-derived contamination. Context-appropriate engineering solutions, such as protection of springs used for drinking water (Kremer *et al.* 2011), use of springs rather than streams as drinking water sources, construction of filtration systems, or introduction of household-level interventions (Clasen *et al.* 2007a), may be more effective at improving water quality. Such solutions however do not provide the desired co-benefits of the intervention, such as forest carbon storage, biodiversity conservation, and increases in local incomes. Future work may aim to combine both conservation and engineering solutions, and involve more direct conservation actions such as purchase or rent of particularly sensitive or important catchments.

3.6 Conclusions

Global interest in PES exists because it is seen as an efficient way to provide multiple benefits; for example to improve water quality, maintain water quantity, store forest carbon and conserve biodiversity while also providing socio-economic benefits. While there have been weighty critiques of this optimistic view from a theoretical perspective (Kosoy & Corbera 2010; Silvertown 2015), the effectiveness of PES in achieving its intended goals is fundamentally an empirical question. The evidence base concerning the delivery of benefits is mixed (Calvet-Mir et al. 2015; Grima et al. 2016) but is generally of poor quality (Pattanayak, Wunder & Ferraro 2010; Miteva, Pattanayak & Ferraro 2012). Although we focus on a single outcome (the supply of clean water), our analysis represents one of the first attempts at a robust, quantitative, counterfactual evaluation of a PES-like program. We found that E. coli contamination was lower in treatment communities post-treatment; however this difference pre-existed the implementation of the intervention, highlighting the importance of a randomised design and baseline data for impact evaluation. We conclude that this particular program would require much more intensive targeting (which would increase substantially the transaction costs and design complexity of the intervention) to have a significant impact on water quality.

Although Randomised Control Trials are not practical in all situations, they certainly have an important role to play in building the evidence base for understanding the impact of controversial, but rapidly spreading, environmental management approaches such as Payments for Ecosystem Services.

4 Can Payments for Watershed Services improve health in communities where it is implemented?

4.1 Abstract

One of the ultimate goals of many Payments for Ecosystem Services programs and similar interventions is the improvement of health outcomes in communities, as a consequence of water supplies being of improved quality and more consistent quantity. Uptake of the *Watershared* program, implemented by the Bolivian NGO *Fundación Natura Bolivia*, is explicitly advocated to communities as having the potential to improve these outcomes. Using both data on levels of diarrhoeal disease among children over the previous year reported in a household survey and direct measures of microbial contamination of water sources for a sub-set of water systems, we examine whether there is evidence that *Watershared* affected health (and related outcomes) in local communities. We find that a number of environmental characteristics, including those relating to land uses in catchments upstream of water intakes, affect microbial water quality in community water systems as measured by *E. coli* concentration. *E. coli* concentration in household water supplies in turn was found to be significantly positively associated with both the number of diarrhoeal episodes per child and incidence of diarrhoea over the past year, although this association was not found in the case of school water supplies. Using data from the full household survey (i.e. not restricted to households for whom microbial water quality data was available) we found that diarrhoeal disease levels were predicted by factors such as a child's age and household treatment of water but not by factors directly related to the *Watershared* intervention. We also did not find any evidence that diarrhoeal disease levels were lower in treatment communities, nor that they were lower in households which sourced their water from *Watershared* areas. We therefore conclude that land use may affect water quality which in turn does significantly influence health and disease levels in communities, but that the *Watershared* program itself did not have a measurable effect on the health of the communities in which it was implemented.

4.2 Introduction

Access to clean water in adequate quantity is a key requirement for human health and wellbeing. Goal 6 of the UN Sustainable Development Goals states that, by 2030, there should be universal and equitable access to clean water (United Nations 2015). However, as of 2014, 1.8 billion people were estimated to consume drinking water contaminated with faecal matter (Bain *et al.* 2014a). Diarrhoeal diseases, due to lack of access to clean drinking water as well as poor sanitation and hygiene, are particularly harmful to children and are estimated to claim the lives of 361,000 children under the age of 5 annually (Liu *et al.* 2012; Prüss-Ustün *et al.* 2014).

Engineering can offer solutions to the lack of access to clear water; for example through construction of water treatment plants to purify water via processes of sedimentation, filtration and chlorination (e.g. Crittenden *et al.* 2012). However, such treatment systems are highly capital- and knowledge-intensive and focused on large systems, which precludes their use in much of the developing world, especially in rural or semi-rural contexts with low population densities (Shannon *et al.* 2008). Rapidly growing peri-urban areas are also often poorly served by such infrastructure (Mintz *et al.* 2001). This is borne out empirically by the large rural-urban disparities in access to improved (principally piped) water supplies in the developing world (McDonald *et al.* 2014). Recognition of this difficulty has led to the spread of household-level treatment solutions, such as ceramic pot filters, solar disinfection, and chlorine or iodine tablets (Mintz *et al.* 2001; Clasen *et al.* 2007b; Hunter 2009). These may be effective when implemented correctly, but both take-up and compliance will never be complete throughout a community (Enger *et al.* 2013). They also may be dependent on external logistics, have an associated cost to be borne by users, and can require a certain level of technical knowledge on the part of those users to be effective (e.g. Roberts *et al.* 2001; Montgomery & Elimelech 2007). Consequently such interventions may not be sustainable solutions over the medium to long term, particularly for more marginalised members of communities (e.g. Clasen 2009; Freeman *et al.* 2012).

These well-known difficulties with achieving adequate treatment of water in rural areas in the developing world have led to a focus on the fact that the quality of available water will be closely linked to, and dependent on, environmental conditions in the catchment from which the water is originally taken (McDonald *et al.* 2014; Herrera *et al.* 2017). Even where such engineering solutions exist, poor quality water arriving increases both infrastructure and operational costs as well as elevating the quantity of effluent produced from such processes (e.g. McDonald *et al.* 2016). In this way, good-quality water can be conceptualised as an ecosystem service, or as a precursor to multiple ecosystem services (e.g. Keeler *et al.* 2012). Forested land in catchments can filter pathogens and pollutants (Pattanayak & Wendland 2007), controls erosion and sediment loading (Hamilton 2008), and provides hydrological regulation preventing contamination through flooding (Bruijnzeel 2004). Forest conservation also leads to relative improvements of water quality through displacement of other more damaging land uses (e.g. Herrera *et al.* 2017). Animal agriculture results in microbial contamination from animal faeces as well as elevated nutrient levels and eutrophication (Ongley 1996); crop production results in sedimentation from runoff, eutrophication and contamination with agrochemicals (Ongley 1996); settlement and infrastructure construction in such areas also change capacity for hydrological regulation and risks point-source faecal contamination (Hamilton 2008); and industrial activities can risk chemical contamination (Hamilton 2008).

Conservation interventions increasingly mention the maintenance or improvement of water quality and community health through conservation of watershed forest ecosystems among their objectives (Abell 2017). This may be achieved by command-and-control interventions (c.f. Lubell *et al.* 2002), locally agreed natural resource management policies (Blanchard, Vira & Briefer 2015), incentive-based conservation such as agri-environment schemes or other kinds of payments for ecosystem services (e.g. Porras *et al.* 2008; Martin-Ortega, Ojea & Roux 2013), or a combination of all of these. However the relationships between ecosystem condition, ecosystem services, and health, are often poorly understood and are systematically understudied (Alexander *et al.* 2013; Myers *et al.* 2013; Bauch *et al.* 2015; Whitmee *et al.* 2015).

Watershared is a Payments for Ecosystem Services-like forest conservation intervention, undertaken by the NGO *Fundación Natura Bolivia (Natura)* in Bolivia since 2003. Through incentivising conservation of forested areas and excluding cattle from watercourses and other water bodies, it aims to achieve provision of water, in greater quantity and of improved quality, to participating communities. In the whole of Bolivia, *Watershared* agreements have been implemented across around 40 municipalities, with 4500 households benefiting from incentives and 210,000 hectares under conservation (Asquith 2016). Improved health outcomes are one of the goals of the intervention (see Abell 2017), based on the fact that drinking water quality is one of the chief predictors of gastrointestinal disease in vulnerable populations (Ashbolt 2004; Bain *et al.* 2014a; b). We focus on *Watershared* as implemented in the *Área Natural de Manejo Integrado* (*ANMI*) *Río Grande – Valles Cruceños*, a 7340km^2 protected area in the Andean region of the Santa Cruz Department, eastern Bolivia.

We combine self-reported levels of diarrhoeal disease (from a household survey) with direct sampling of microbial water quality to explore the factors influencing both (both those relating to land use, which *Watershared* seeks to influence, and those relating to management of water supplies) on levels of *Escherichia coli* in those water supplies and health outcomes in communities. We explore the extent to which *E. coli* concentration in water supplies to households and schools may predict levels of diarrhoeal disease. We also use the wider survey data to quantify the extent to which levels of diarrhoeal disease are affected by *Watershared* actions and other factors of interest. We therefore examine to what extent the hypothesis that *Watershared*, and similar programs, can improve community health outcomes is supported.

4.3 Case Study: Socio-Ecological Context

In 2011 *Natura* offered *Watershared* to approximately half of the communities within the within the *ANMI Río Grande – Valles Cruceños* protected area, as the intervention was set up as a randomised control trial (RCT). More details of the intervention and the experiment are available in chapters 1, 2 and 3 and in Bottazzi *et al.* (2018). In this article we focus on level 1 *Watershared* contracts, where landowners can receive $10/hectare/year equivalent, plus a $100-equivalent joining bonus, in return for agreeing not to deforest and not to allow cattle or other livestock to enter enrolled areas which must be forested land within 100m of a stream or other water body. Compliance monitoring was conducted yearly.

Households in the study area rely on water supplies with different underlying infrastructure. Commonly, water intakes (known locally as *tomas de agua*) are located in streams, or on top of springs (figure 4.1a); the water extracted then flows downhill (pumped water is very rare; the authors only encountered pumped water in one community of more than 120 visited) through sedimentation and/or filtration chambers (in some cases). It is then stored in a storage tank from where it is distributed to taps in the community, in the central square (figure 4.1b), at the school, and in private houses. Communities may have a single system supplying all members of the community, or may have multiple systems; in rare cases they may share water intakes with other communities.

Figure 4.1. Examples of typical water supply systems in communities of the ANMI Río Grande. a) Water intake, Estancia Huaico (Vallegrande municipality); b) Community tap, central square, Vado del Yeso (Vallegrande municipality). Photographs: Edwin Pynegar.

Chemical treatment such as chlorination is extremely rare and was only encountered in one community. Therefore the quality of water entering the water supply system strongly influences its quality at the tap. *Watershared* seeks to improve water quality in conserved areas by preventing cattle from defecating in or near protected intakes, by reducing turbidity in water through preventing cattle trampling and disturbing soil, and through lowered levels of deforestation for agriculture reducing erosion. There exists a widespread perception in the area that contamination of water due to cattle access has resulted in poor water quality and widespread waterborne disease burden (Paredes & Isurza 2012; Rojas Banegas 2012). Gastrointestinal illnesses are endemic throughout the area, especially in marginalized groups such as the very young, very old, and immuno-compromised people. Communities have been fencing off water intakes against livestock for many years, both endogenously and via collaborations with other non-governmental organizations such as the *Instituto de la Capacitación del Oriente* (ICO). There is thus a pre-existing local belief in the benefits of forest conservation and cattle exclusion for health and this focus on water quality benefits is a significant part of how the *Watershared* intervention is presented to members of communities being invited to join (figure 4.2).

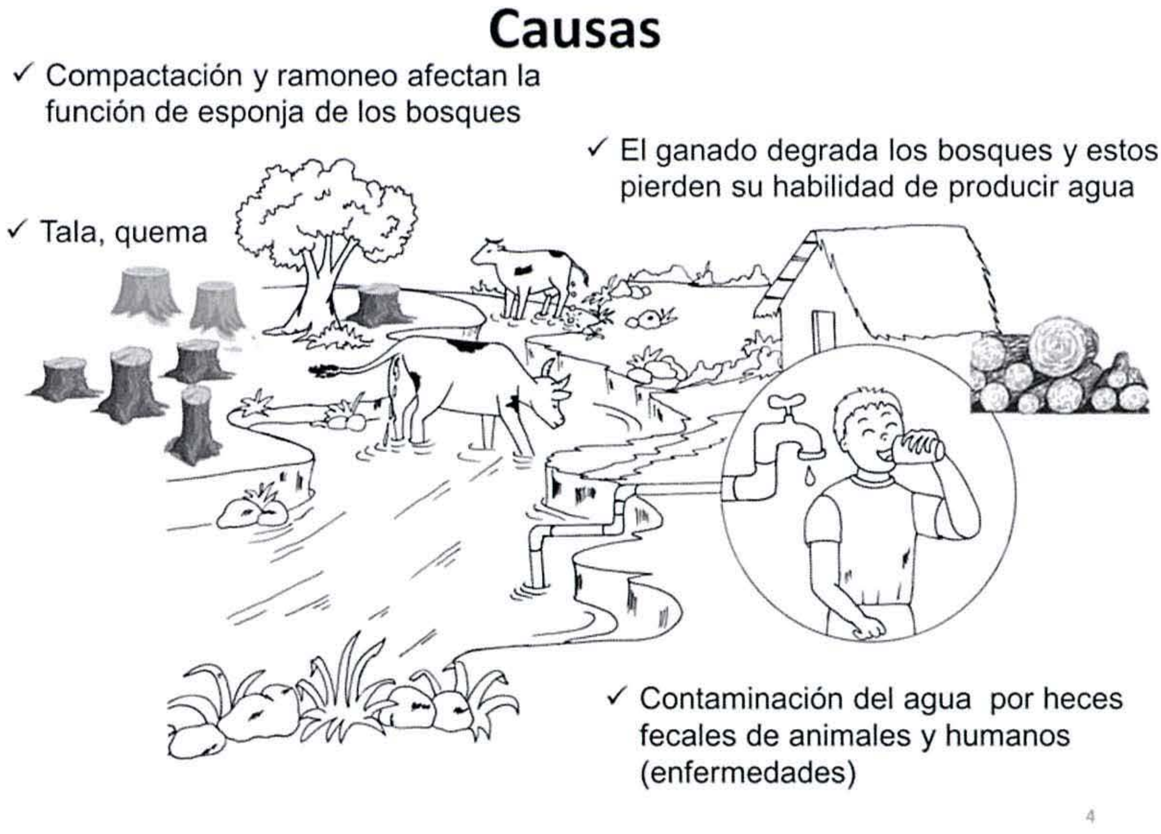


Figure 4.2. Effects of deforestation and extensive cattle grazing on water supplies, as presented by *Natura* when offering the *Watershared* agreements to community members.

4.4 Methods

4.4.1 Data collection and availability

The analyses presented in this paper are based upon two separate sets of data. The first is a household survey collecting endline data for an RCT-based evaluation of *Watershared*, including questions about incidence of diarrhoeal disease experienced by children under the age of 18. The second data set is measures of microbial water quality from a sub-set of water systems (intakes and the taps they supply) chosen specifically for this study (they were known to supply households with children). Extensive field work was required to ensure that it was possible to match up which household obtained water from which water system.

4.4.1.1 The household survey

The aim of the household survey was used to evaluate the success of the *Watershared* program in the *ANMI Río Grande* in achieving its intended socio-economic outcomes. It was a modified version of a baseline conducted in 2010-11 before *Watershared* was offered in treatment communities (see Grillos 2017). The survey was adapted by Bangor University and conducted by *Natura* technicians between October 2015 and June 2016. Heads of all locatable households within the 129 communities in the RCT were surveyed. More details are available in Bottazzi *et al.* (2018) and in the data repository http://reshare.ukdataservice.ac.uk/852623/. The survey contained a section relating to water supply, management and institutions. It also contained a section related to number of diarrhoeal disease episodes suffered by household members under the age of 18 over the past year. Ethical approval was granted by Bangor University. Total number of surveyed households was 1459.

4.4.1.2 Direct measurements of microbial water quality

In March and April 2016 Edwin Pynegar sampled water quality at a selected subset of the water intakes and their supplied taps within the communities of the *ANMI Río Grande*. He also established which households were supplied by which intake (so this data could be related to data from the household survey above). This was carried out in 44 selected water systems (water intakes and their supplied taps). We selected these systems based on knowing that there were children under the age of 16 in households supplied by these systems as well as accessibility from the town of Vallegrande. There is no reason to believe that this would have introduced a systematic bias in which kinds of water supply systems were sampled. We measured both physico-chemical and microbial indicators of water quality (see below) as well as features of the water system's infrastructure and presence of potential sources of contamination in the immediate environment and catchment upstream of the intake. We added some new indicators to the list of those monitored in 2015 (chapter 3), with a particular focus on land use in catchments above intakes. Details of these features and how they were measured can be found in table 4.S1.

The metric of faecal contamination used was *Escherichia coli* concentration in water samples. Many types of water-borne organism – bacterial, viral and protozoal – may cause similar or related symptoms (e.g. Ashbolt 2004). Of these, however, *Escherichia coli* is of particular interest because it may be both pathogenic *per se* and also an easily detectable indicator of broader pathogenic contamination. Some but not all strains of *E. coli* are harmful to humans (such as *E. coli* O157:H7; Leclerc *et al.* [2001]) and *E. coli* as a species is generally believed to exist in the environment exclusively from mammalian or avian faeces and therefore is an indicator of faecal contamination and presence of other organisms, such as rotavirus, which cause gastrointestinal diseases even more frequently (e.g. Leclerc *et al.* 2001). Hence its presence (or that of the closely related 'faecal' (thermotolerant) coliform group) has been widely used as an indicator of more general pathogenic contamination (Edberg *et al.* 2000; Ashbolt, Grabow & Snozzi 2001). A recent meta-analysis (Gruber, Ercumen & Colford 2014) supported the use of *E. coli* specifically by finding a significant link between *E. coli* concentration and relative risk of diarrhoeal infection, while no link was found between 'faecal'/thermotolerant coliform concentration and risk of infection.

E. coli concentration, along with that of other non-*E. coli* bacteria belonging to the coliform group, was enumerated using the Coliscan Easygel method (Micrology Labs, Goshen, IN, USA). Coliscan Easygel allows enumeration of coliforms as after incubation *E. coli* colonies appear purple, blue-purple or dark blue due to metabolism of both beta-galactosidase and beta-glucuronidase. Other non-*E. coli* coliforms are pink based upon metabolism of beta-galactosidase only. Colonies of a blue-green or sky blue colour (metabolism of beta-glucuronidase only) and white colonies were not counted (Micrology Labs 2016). It is likely that *Coliscan Easygel* is adequate to determine *E. coli* concentration, given that it works based upon metabolism of beta-galactosidase (all coliforms) and beta-glucuronidase (in the coliform group, >95% of these are specifically *E. coli*; Kilian & Bülow 1979; Rice, Allen & Edberg 1990).

Three separate samples of water at each of the selected intakes and each respective supplied tap were taken using sterile Coliscan sampling flasks of size 35ml and placed on ice within 1 hour of sampling. Within 6 hours of sampling (although generally within 4) we produced Easygel Petri dishes using 5ml of water from each flask as inoculum. After solidification we sealed the Petri dishes and incubated them for 36 hours at 35-37°C in a portable incubator (NQ28 model, Darwin Chambers, St Louis, MO, USA). In locations where no mains electricity was available we maintained a constant incubation temperature through use of a 12V vehicle power supply or supply from a car battery. After incubation we counted *E. coli* and other non-*E. coli* coliform colony forming units. Petri dishes and Coliscan bottles were subsequently sterilized by boiling for a minimum of 1 hour and then disposed of.[1]

We also measured at each site a number of physico-chemical parameters of this water: temperature, dissolved oxygen in mg/l and as per cent of saturation value, pH, salinity and conductivity with an HQ40d portable multi-parameter meter and IntelliCAL LDO101, PHC101 and CDC401 rugged probes respectively (HACH Company, Loveland, CO, USA). We measured turbidity in formazin attenuation units through the use of a DR/850 colorimeter following standard protocol (HACH Company, Loveland, CO, USA).

In communities with multiple water intakes supplying them, we asked community members for information on which households were supplied by which intakes. Separate ethical approval was granted for this part of the research by Bangor University. We obtained permission from community members to take all water samples after explaining the nature of the study and that it was conducted together with *Natura*. Any personally identifiable data were stored on encrypted hard drives. Results were passed to communities with explanations and possible actions to take to improve water quality (see document 4.S1), and were also passed to municipal governments in Vallegrande and Pucará.

4.4.2 **Analyses**

The first two analyses (exploring factors influencing *E. coli* contamination in water supply systems and the relationship between *E. coli* contamination and health outcomes) use only data associated with 44 selected water systems as specified above (both household survey responses from households supplied by those systems, and measures of microbial water quality at intakes and taps). The third analysis (exploring the factors associated with health outcomes) uses household survey data from the full RCT. All analysis was done in in the R programming language (R Development Core Team 2014).

4.4.2.1 Which factors influence E. coli *contamination in water supply systems?*

In chapter 3, we explored the influence of environmental context and potential sources of contamination around water intakes and the infrastructure of the water intake on *E. coli* concentration at intakes and taps based on data from a total of 124 water systems (intakes and taps) measured in 2015. Here we use our water quality data from 2016, collected in a similar way from 44 selected water systems (84 sites, i.e. intakes and taps), to check for consistency in our conclusions and also to examine whether the increased range of predictors measured also explained more about *E. coli* levels (see table S1). Predictors associated with intake sites (such as presence of faeces) were also attributed to the tap sites supplied by those intakes, as in the previous study.

[1] This paragraph is almost identical to that in chapter 3.

We evaluated which predictors (listed in table 4.S1) significantly predicted *E. coli* concentration in the measured water systems. We used the *glmmADMB* package in R as previously (Fournier *et al.* 2012; Skaug *et al.* 2016) to produce hierarchical GLMMs with log-links and negative binomial error structures in order to evaluate this, as in the previous chapter. We included water system throughout as a random effect. We began with a repetition of the best model from the previous chapter, in which we found that turbidity, intake or tap site, spring or stream-supplied intake, and presence of faeces at intake all significantly predict contamination. (Disturbance of sediment in intakes also predicted contamination in 2015, but we took care not to do this in 2016.) We then conducted a model selection process based upon minimization of the Akaike's Information Criterion, selecting potential models to include based on their potential to be both predictive and parsimonious. Once we had identified the models we considered most interesting and instructive (the model with lowest AIC, the simplest good model, and a repetition of the model including predictors from 2015), we calculated 95% coefficient confidence intervals and produced coefficient plots showing these using the *ggplot2* package in R (Wickham 2009). We produced normalized model weightings showing relative goodness of the models tested. As a number of the produced models had very similar AIC values, we also averaged across all models with independent predictors (i.e. not including each when they are measures of the same response) using the *MuMIn* package in R (Barton 2016) and similarly calculated 95% CIs and produced plots showing the coefficients of the conditional average of these models.

4.4.2.2 *To what extent does* E. coli *concentration in drinking water predict diarrhoeal disease levels in communities?*

We identified the water system (measured in March/April 2016) that supplies each household, and also that which supplies the school which any children attend. We could therefore join the household survey data with the water quality data for household supply (348 children in 190 households supplied by 36 water supply systems) and school supply (362 children in 200 households supplied by 24 water supply systems).

We used this data to investigate whether the *E. coli* concentration measured at taps affected levels of diarrhoeal disease as reported in the endline household survey. We used the *glmer* function in the *lme4* package in R (Bates *et al.*, 2015) to fit four different generalized linear mixed models, with two alternative metrics of illness level and referring to two different water supplies. Each of these four models included fixed-effect predictor variables of *E. coli* concentration at the tap in question and child age; the model also contained a random effect representing groups of children supplied by the tap in question. One metric of illness used was the number of times the child had been ill, in which case the model had a log-link and Poisson error structure; the other metric was whether or not that child had been ill, and thus a logit link and binomial error structure. The two water supplies whose *E.*

coli concentration were used as predictors were the household water supply for each child, and the school water supply. These could be the same, or could be different depending upon the layout of the community's supply system(s). We tested for significance of each predictor in each model, and again calculated 95% coefficient confidence intervals and produced coefficient plots for selected models using ggplot2, as well as plots showing the direct relationships between *E. coli* levels and levels of disease.

4.4.2.3 Which factors predict levels of diarrhoeal disease among children in the communities?

Using the full household survey data (for this analysis we were not restricted to the 44 water systems for which we had measured microbial water quality) we tested whether a number of survey responses relating to water supply and management predict diarrhoeal disease levels among children under the age of 18 within the communities of the RCT (see table 4.S2). Some of these responses relate closely to those in section 4.4.2.1, but others relate to separate factors relating to water management institutions and household water treatment. We reclassified some of these responses to combine them when very similar (see table 4.S2); we also tested a number of predictors derived from a question relating to level of protection of water source from cattle (coded as CE1-4 in table 4.S2), as we felt these to be particularly important. (Once we had established which of these predictors was best, we retained that in all other models.) For each child in the dataset we knew the household they belonged to and so we were able to associate them with predictor variables. We removed children from the dataset for whom we had missing data; the final dataset therefore consisted of 1012 children in 552 households belonging to 106 communities.

We produced GLMMs with a hierarchical structure using the *glmer* function in the *lme4* package in R to explore factors predicting number of episodes of diarrhoeal disease each child had suffered from over the past year (the response variable selected). We selected these models as potentially both predictive and parsimonious, using the predictor variables specified in table 4.S2 as fixed effects and a random effect of children nested within households. The models used a log link and Poisson error structure. We had planned to structure the model with children nested within households within communities, but we found initially that the community (*OTB*) part of the random effect was of negligible importance (zero variance explained).

We conducted model selection by means of the Akaike's Information Criterion. Once we had a most parsimonious model candidate, we separately tested significance of two further kinds of potential predictor based on minimization of the AIC. The first of these related to whether the household was supplied by water from an intake in a level 1 *Watershared* area and/or belonging to a treatment or control community (factors relating to the evaluation design, and in the second case an example of a standard *intention to treat* RCT analysis, but also factors independent from any predictors in the most parsimonious model); and the second related to perception of changes in water quality and quantity over the past 5 years.

We calculated 95% confidence intervals for the predictors in three of the resulting models (the global model, the minimal model, and the model found to be more parsimonious including the perception of water quality change), and used the *ggplot2* package (Wickham 2009) in R to produce coefficient plots showing these models. We conducted normalized model weighting for each of the three model selection processes. We also again conducted model averaging using *MuMIn* (Barton 2016), including all the models tested which had independent predictors, and calculated 95% confidence intervals and produced coefficient plots based on the conditional average.

4.5 Results

4.5.1 Which factors influence *E. coli* contamination levels in water supply systems?

Running the best predictive model from chapter 3 with the data described here (model 1.1) showed that type of water body which supplies the intake, and turbidity, both significantly predicted *E. coli* level as in 2015. Tap sites showed a tendency to lower levels than intake sites, as in chapter 3, but this was not significant at 95% CI. Unlike in the results of chapter 3, presence of cattle faeces around the water intake did not have any significant effect (figure 4.S1a).

Model selection based upon AIC showed that presence of cattle upstream and presence of agriculture upstream or uphill of the intake strongly predicted *E. coli* contamination in all cases, as did (negatively) the intake being fed from a spring rather than a stream. A number of models had very similar AIC values (difference <1) and normalized model weighting showed very little difference in which would be most likely. Some of these models showed an effect of faecal presence at water intake and/or of the intake being protected from cattle (figure 4.S1b). The model with fewest predictors, however, only contained those which were included in all models (figure 4.S1c). There was no effect in any of the most parsimonious models of the intake itself being located in a level 1 *Watershared* area. Model averaging showed that the conditional model only included as significant at 95% CI the same variables as that model with the fewest predictors (figure 4.3; also see figure 4.S1c). Model selection, including AIC values and normalized model weights, is shown in table 4.S1.

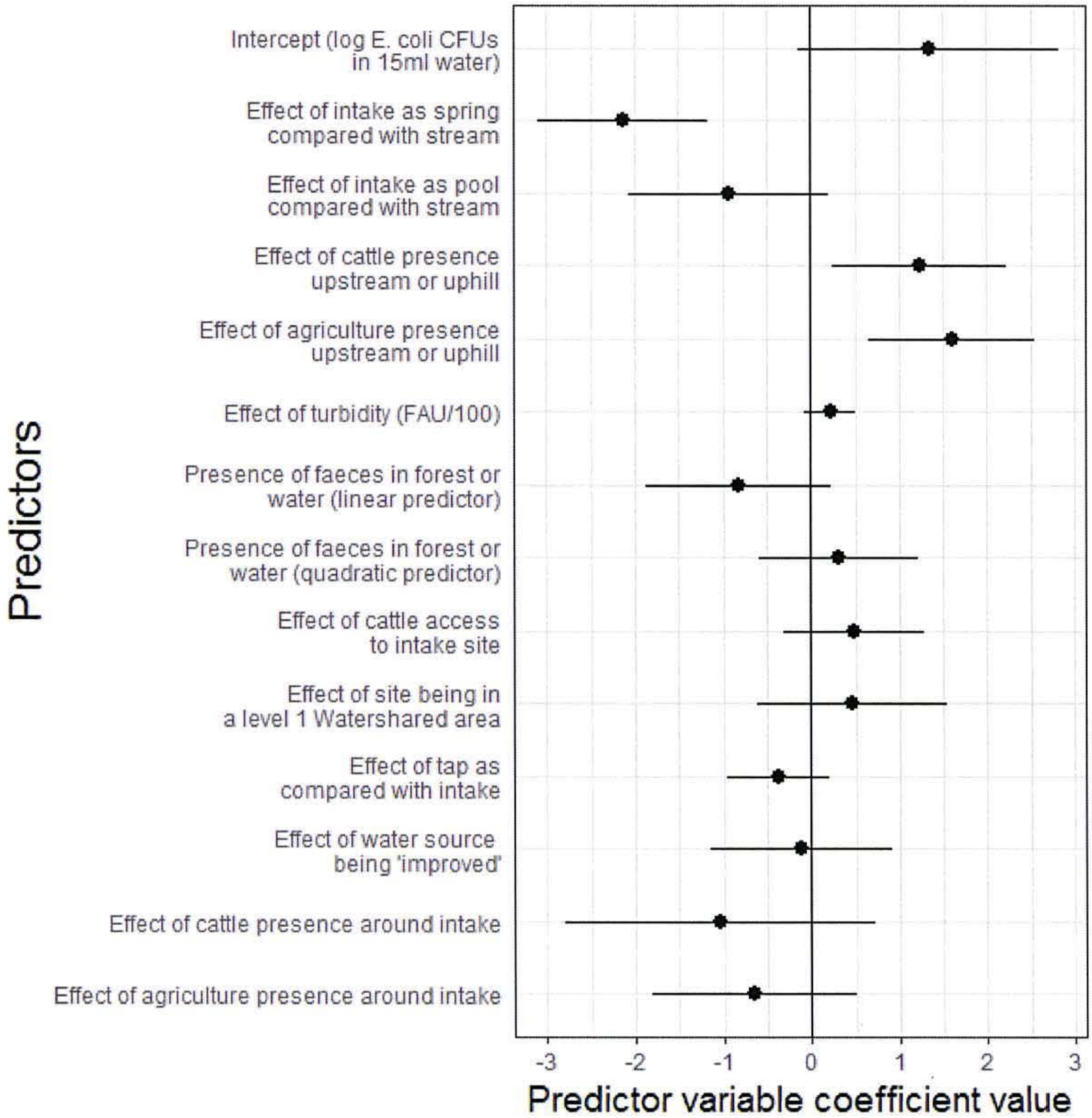


Figure 4.3. Conditional average of models testing predictors of *E. coli* levels in water supply systems (N=84; water system N=44).

4.5.2 To what extent does *E. coli* concentration in drinking water predict diarrhoeal disease levels in communities?

E. coli concentration in taps supplying households is a significant predictor of both number of diarrhoeal disease episodes over the previous 12 months (p=0.017) and presence or absence of these episodes (p=0.037); age of the child also highly significantly negatively predicts number of disease episodes (figure 4.4 and 4.5; coefficients in tables 4.S4a and 4.S4b). However, *E. coli* concentration at the tap of the school which the children in attend showed no significant effect (tables 4.S4c and 4.S4d). Direct relationships between *E. coli* and diarrhoeal disease levels are shown in figures 4.S4a (number of episodes) and 4.S4b (probability of >=1 episode in the past year).

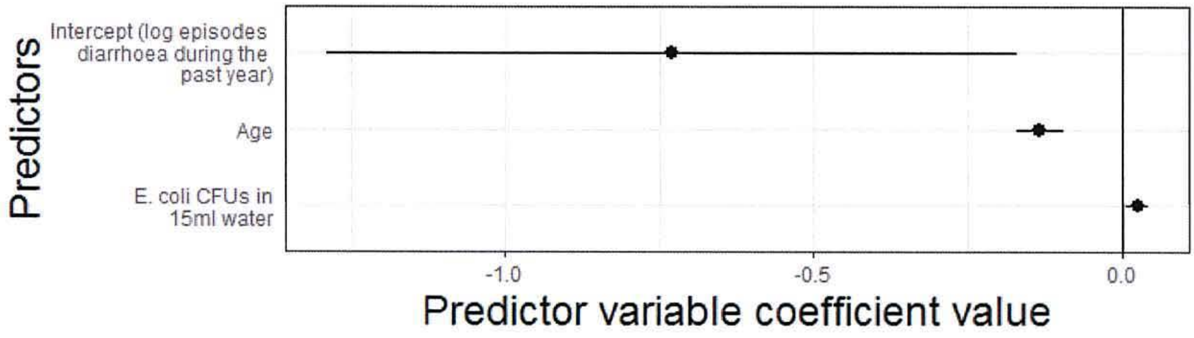


Figure 4.4. Coefficient plot (with 95% CI values) showing the effect of age and measured *E. coli* concentration in household water supplies on diarrhoeal disease levels in communities over the previous 12 months (model specified in table 4.S4a).

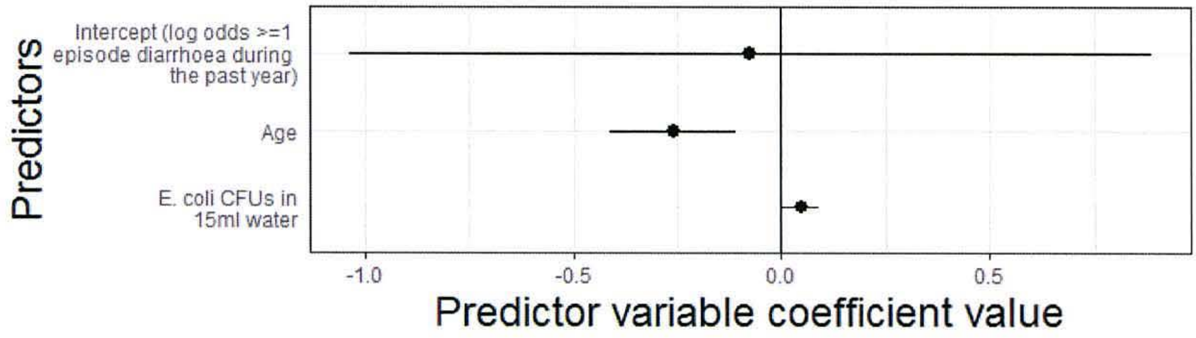


Figure 4.5. Coefficient plot (with 95% CI values) showing the effect of age and measured *E. coli* concentration in household water supplies on incidence of diarrhoeal disease over the previous 12 months in communities (model specified in table 4.S4b).

4.5.3 Which survey responses predict the level of diarrhoeal disease?

A number of the factors derived from responses to survey questions significantly predicted number of illness episodes, although none related to actions directly linked to *Watershared*. Age was highly significantly negatively associated with number of diarrhoeal disease episodes across all models. Boiling or chlorination within households was also a positive predictor of diarrhoeal disease (this may indicate households with particularly dirty water). Months without water was positively significant at 90% CI and was retained in the model with lowest AIC (model 2.9; figure 4.S3a; table 4.S5a). The global model, including all initially tested predictors (model 2.1), shows a tendency for increasing levels of cattle access to water sources to be associated with increasing diarrhoeal disease level; while this was not retained in the lowest-AIC model, it was retained in some of those with very similar (<1) AIC values (figure 4.S3b; table 4.S5a; see models 2.5 and 2.7). Neither the respondent's community belonging to the treatment group nor the location of their supplying intake in a level 1 *Watershared* area had any predictive effect on the disease level (table 4.S5b). We did however find that perception of decline in water quality over the past 5 years was predictive of increased diarrhoeal disease frequency (model 2.20; table 4.S5c, figure 4.S3c). However, as with all cross-sectional studies of this kind, this perception could itself be derived from observation of diarrhoeal episodes (i.e. the direction of causality is unclear). Averaging across all independent models we find only child age, boiling or chlorination, and perception of water quality decline to be significant at 95% CI, while months without water and level of protection of water sources from cattle only show a tendency towards a significant effect (figure 4.6).

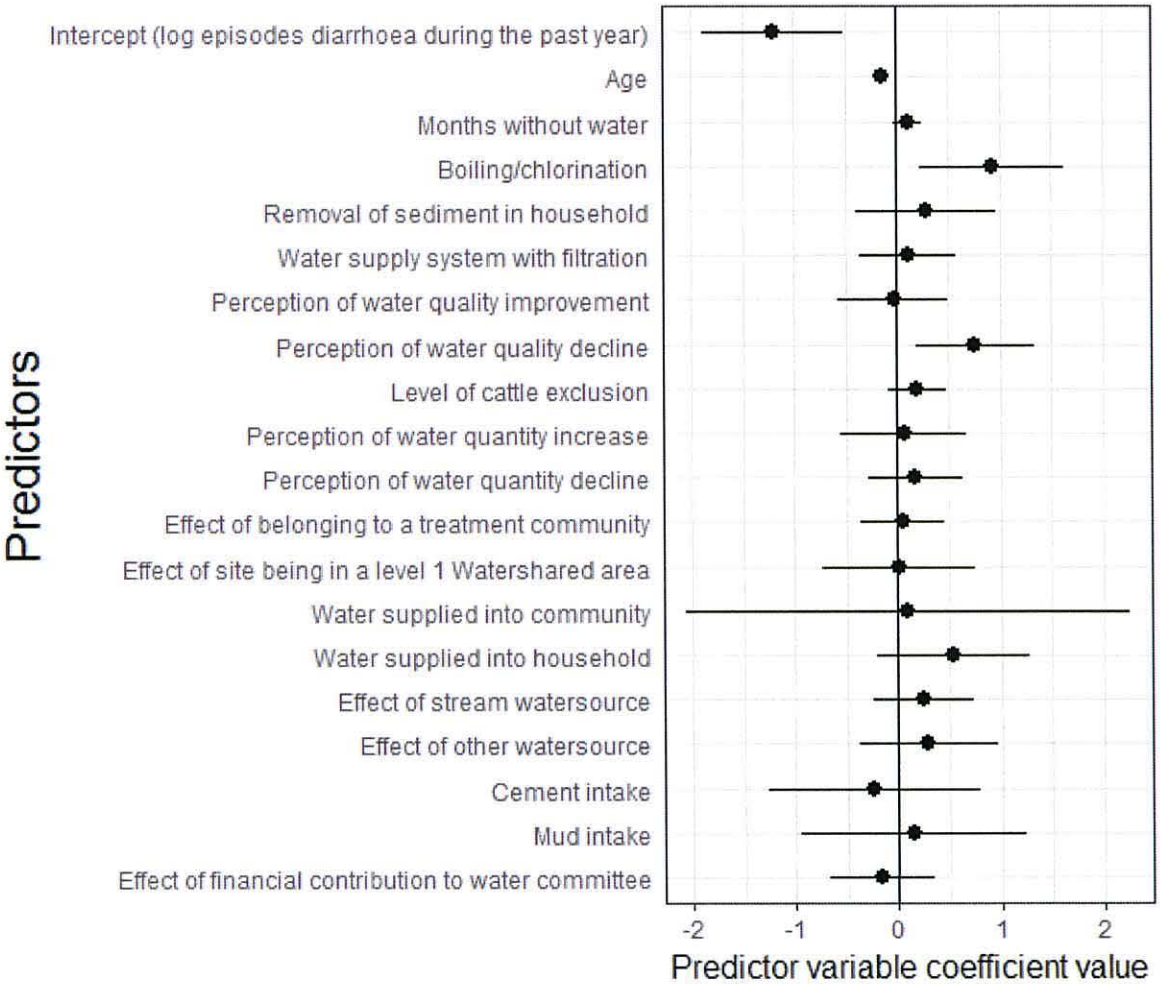


Figure 4.6. Coefficient plot (with 95% CI values) showing conditional average of models relating to water supply infrastructure and institutions and conservation actions on self-reported levels of diarrhoeal disease over the previous 12 months among children in communities (N=1012; household N=552).

4.6 Discussion

We investigated whether land use and watershed protection in general, and conservation actions associated with the *Watershared* payments for ecosystem services-like intervention in particular, could be found to have an effect on levels of faecal indicators of contamination and of diarrhoeal disease in communities within the *ANMI Río Grande – Valles Cruceños* protected area. We also directly evaluated whether the RCT setup provided evidence for the *Watershared* intervention having led to improved health outcomes in treatment communities.

4.6.1 Factors directly associated with the *Watershared* intervention do not have a measurable effect on microbial contamination of water supplies, but its goal (forest conservation) may do

In common with the analysis presented in chapter 3 (based on measures of water quality in a larger number of water systems in 2015), we found a number of characteristics of upstream land use and intakes significantly predict measured *E. coli* concentrations in water systems. The 2016 data presented in this chapter (collected from 44 water systems studied in greater depth) again showed that water from springs or pools over streams was cleaner than rivers (something implied by various examples in the literature; Howell et al., 1995; UNICEF, 2008; WaterAid, 2013) and the association of turbid waters with *E. coli*. Taps showed a tendency to be cleaner than intakes, as before, but the low water system *N* (44) may explain why this was not significant at 95% CI. In 2016 we collected data on a larger number of potential predictors of water quality including upstream land uses. Presence of cattle and/or agriculture upstream of the water intake was extremely strongly predictive of elevated *E. coli* concentrations. This supports the importance of considering catchments as an integrated unit if *Watershared* and similar programs are to be effective (see also Dodds & Oakes 2008). While faeces presence positively predicted *E. coli* levels in 2015, the 2016 data did not show this. We found no evidence to suggest the intake being located in a level 1 *Watershared* area to be significant as a predictor.

4.6.2 Faecal contamination (as measured by *E. coli* concentration) positively predicts diarrhoeal disease, despite methodological issues

We found clear evidence that *E. coli* concentration in water supplied to households was significantly positively associated with both incidence of diarrhoeal disease over the past year, and with the number of disease episodes experienced per child. This supports the relationship suggested by the published literature (Gruber, Ercumen & Colford 2014). It also demonstrates that *E. coli* concentration (measured with the Coliscan Easygel method) is an appropriate proxy for microbial water quality and has direct relevance to health and disease. *E. coli* concentration in water supplied to schools did not predict diarrhoeal disease; this is likely not only to be due to the fact that the number of independent data points (water systems) is lower in this case, but also that children will spend a majority of their time and therefore drink most water in the household rather than the school.

We were somewhat surprised to be able to detect this relationship, as *E. coli* measurements taken on one occasion may not be representative of concentration throughout the year, and levels are known to be highly variable (although do tend to increase in the wet season, when we sampled [Kostyla *et al.* 2015]). Also, the 12-month period for self-reported disease recall used in the household survey was much longer than the 2 weeks considered adequate in similar studies (see Fewtrell *et al.* 2005) and the period of a few days recommended by many researchers (Melo *et al.* 2007; Feikin *et al.* 2010; Zafar, Luby & Mendoza 2010; Arnold *et al.* 2013). We therefore expected substantial underestimates in recalled numbers of episodes, but decided to continue using that metric as it contains more information than simple presence/absence of diarrhoeal disease (this explains why our results were not reported as odds ratios as is conventional for cross-sectional diarrhoeal disease studies of this kind). We believe this decision to have been supported by the fact that *E. coli* concentration more strongly significantly predicted number of disease episodes than it predicted presence or absence of diarrhoeal disease. One-year retrospective recall periods have however been used in establishing incidence of other water-borne conditions. For example, Tayeh & Cairncross (1995) found that false positives and false negatives cancelled each other out.

4.6.3 Survey responses associated with the *Watershared* intervention did not predict reduced levels of diarrhoeal disease

We found a number of water-related characteristics from the household survey to be predictive of diarrhoeal disease frequency in the communities of the ANMI Río Grande. However even using this larger dataset, none of the predictors found to be significant were directly related to *Watershared*-associated actions. Age is well known to be a significant predictor of risk of diarrhoeal disease (e.g. Herrera *et al.* 2017), and this was strongly borne out by our data. Months without available water also positively predicted disease frequency in some models, as dry primary water sources during the dry season would likely lead to households being forced to supply themselves from secondary water sources which they might know to be contaminated, or which would have much less infrastructure to facilitate physical or chemical treatment. Additionally, households with several months without water likely live in communities with poor water availability, as otherwise they would be unlikely to construct water supply infrastructure in places without adequate water throughout the year. The fact that boiling or chlorination was positively associated with diarrhoeal disease was a surprise, but given the resources, time and knowledge required to consistently do this, we might hypothesise that households only do it in cases where the water supply is already known to be of poor quality. It is likely that not all water that children consume will have been treated in this way, so this treatment may serve as an indicator of poor quality water, rather than evidence of an effect of water treatment on diarrhoeal disease *per se*. Defensive behaviour, as this is known, is known to increase with perception of exposure to poor quality water (Alberini *et al.* 1996). This is an example of the familiar challenge of establishing causality in cross-sectional studies (e.g. Mann 2003).

4.6.4 Upstream land use can influence water quality and therefore health, but design of the *Watershared* intervention may need to be modified to achieve this.

We conclude that upstream land use and therefore (implicitly) conservation actions can affect *E. coli* concentration, which in turn affects disease levels in communities. This supports the recent findings of Herrera *et al.* (2017), who showed using a global dataset that forest loss or other declines in conservation status of a catchment increased probabilities of diarrhoeal disease for those children under 5 living in it. While these results support the use of forest conservation to improve health in similar contexts, and adds to a growing literature which arrives at similar conclusions (Abell 2017; Herrera *et al.* 2017), they do also suggest that the *Watershared* intervention as implemented may not be particularly effective in achieving this. Small parcels of land enrolled in conservation agreements, wherever the landowner wishes to enrol them, may imply low effectiveness due to mismatch in spatial scales between the action and the biophysical processes occurring – as the landowner may enrol land in other catchments entirely and/or may enrol very small proportions of land in parts of the catchment of lesser importance. Beyond the level 1/2/3 distinction, *Watershared* has no spatial targeting integrated into its design. Loading with coliforms can and does occur through a number of pathways (Pachepsky & Shelton 2011; Cho *et al.* 2016), and coliforms can persist in the environment for a substantial length of time (Meals, Dressing & Davenport 2010), especially in waterbody sediment (Anderson, Whitlock & Harwood 2005). This may explain how we continued to encounter *E. coli* and other coliforms even in intakes located within protected forests and which amount to improved water sources (such as that found in the community of *Paredones*; figure 4.7). Similarly, substantial differences in deforestation and thus in additional forest protection in catchments would likely be necessary in order to detect any changes in diarrhoeal disease levels, given that a much larger dataset only showed a 5% change in relative risk of disease from 30% deforestation of a catchment (Herrera *et al.*, 2017). We examine related issues and their implications in the following chapter and in the Discussion.

Figure 4.7. Improved water source supplying the community of Paredones.

4.6.5 **Suggestions for improving water quality and health outcomes in the study region**

Infrastructure and education, as well as healthcare availability, are important confounding factors which affect community health outcomes. Only 40% of the relative risk of diarrhoeal disease in children is estimated to be attributable to the quality of the supplied water (UNICEF 2008); the remainder relates to within-household water storage, sanitation system availability, functionality and access, and hygiene. Socio-economic status of households is also a determinant and cuts across many of these features (e.g. Kumi-Kyereme & Amo-Adjei 2016). While we have shown that communities sourcing water from well-conserved catchments may reduce the faecal contamination to which they are exposed through their water supply, which in turn reduces diarrhoeal disease levels in communities, achieving such changes through programs such as *Watershared* is but one of a number of kinds of action that could be taken in order to reduce waterborne disease burden (Fewtrell *et al.* 2005; Clasen *et al.* 2007b; Kremer *et al.* 2011). Cost-benefit analyses of the expected effects of such interventions would help to clarify whether 'conservation for health' could be a feasible and cost-effective kind of program in this context, while not forgetting that payments for ecosystem services-like conservation programs such as *Watershared* aim to change a wide range of environmental and socio-economic outcomes in the areas in which they are implemented (Ferraro & Kiss 2002).

Centralized water treatment systems, as exist in much of the industrialized world, are unlikely to be feasible in contexts such as rural Bolivia over the short to medium term. Therefore, quality of water consumed in communities will continue to depend on the ecological status of the area hydrologically upstream of the intake; the infrastructure of the intake and the system that supplies the water to users, including treatment should it exist; actions related to water storage and sanitation, both inside and outside the household; any household-level water treatment; and the institutions which are responsible for the management of these features. We therefore suggest that provision of good quality water to these, and similar, communities, will require a consideration of these supply systems as an integrated whole. Prioritizing infrastructure construction above other aspects will not provide a sustainable solution without institutional support for maintenance, cleaning, and repairs (see also the discussion of the case of Pucará in chapter 6), while existence of a water committee (Foster 2005) or other institution for local management. without adequate understanding of infrastructure or available financial resources may achieve little and can also act to disempower such communities (Chowns 2015). Thus the issue of supplying clean water in adequate quantity in the *ANMI Río Grande* and similar contexts in rural developing Latin America is a multifaceted one. Consideration of each of these aspects, as well as how they may function together as a whole, will be indispensable in developing long-term solutions.

4.7 Conclusions

Our findings suggest that the ecological status and land use in catchments is an important determinant of the (microbial) quality of water in those catchments, which in turn affects community health and disease outcomes when water supply is sourced from such catchments. However it may be the case that the *Watershared* intervention itself, as implemented in the *ANMI Río Grande*, is not necessarily effective in achieving such changes. We have no evidence to suggest that the *Watershared* actions directly achieve this intended outcome, and nor that individuals in treatment communities as a whole have lower numbers of disease episodes. This may be due to the relatively small scale of the level 1 protected areas and the possible requirement to conserve larger proportions of catchments than that often achieved by the agreements. We conclude that a larger-scale implementation of agreements in critical catchments or other identified areas would have more potential for achieving its goals, as part of an integrated consideration of water supply and treatment.

5 The devil in the detail: experiences from the implementation of a large-scale socio-ecological Randomised Control Trial

5.1 Abstract

Randomised Control Trials have been widely used for decades in many fields of applied science, and interest in them is growing in the field of conservation. We previously identified a number of issues which evaluators may be faced with when evaluating socio-ecological conservation interventions – those in which desired ecological changes are mediated by human behaviour changes – in ensuring that the evaluation is of high quality. In this chapter, we examine the consequences of some of these issues with reference to the outcomes of the RCT of *Watershared*, a Payments for Ecosystem Services-like intervention in the Bolivian Andes. We show how selection of randomisation unit can be very difficult when multiple outcomes are intended to be evaluated, and explore the implications of this for spillover effects between units. We consider how selecting appropriate sites for monitoring outcomes is challenging and can result in attrition when these sites are unavoidably moved, and how site selection may interact with other spatial aspects of the intervention to affect the outcomes actually measured. We also examine the consequences of intervention implementation being voluntary on the treatment effect and its consequences on measurable outcomes, as well as the potential for error and copying by members of control units. We conclude that undertaking a good-quality, internally valid RCT will require understanding a clear theory of change, substantial preparation and planning, and a focus on a small number of principal outcomes of interest, rather than expecting an RCT to be able to evaluate changes in many different outcomes.

5.2 Introduction

Randomised Control Trials (RCTs) are a rigorous method of determining whether, and to what extent, a cause-effect relationship exists between an intervention or program and an outcome of interest (e.g. Rubin 1974). RCTs therefore are widely promoted as a robust approach to impact evaluation (e.g. Independent Evaluation Group 2012). The principal advantage of the RCT method is that random allocation of the intervention across all units should ensure that there is no systematic bias in which units receive the intervention (White 2013); comparing outcomes subsequently between treatment and control units allows calculation of the effect size of the intervention. RCTs are the backbone of evidence-based medicine (e.g. Barton 2000) and have been used in many other fields, particularly in development microeconomics (e.g. Banerjee & Duflo 2011). There are increasing calls for wider use of RCTs in evaluating the impact of environmental management interventions (Greenstone & Gayer

2009; Pattanayak 2009; Miteva, Pattanayak & Ferraro 2012; Curzon & Kontoleon 2016; Börner *et al.* 2016, 2017).

There have long been concerns about difficulties in implementing RCTs in ways which ensure that the effect size measured in the experiment approximates the true effect of the intervention; that the evaluation is *internally valid* (e.g. Rubin 1974; White 2013). Internal validity is a modest challenge in the case of conservation or environmental management interventions which are primarily ecological in nature (chapter 2), but in socio-ecological interventions (where outcomes are strongly mediated by human actions or behaviour) it may present more substantial difficulties. Socio-ecological interventions are common in environmental management and include payments for ecosystem services, agri-environment schemes, environmental education programs, and ecotourism. The difficulties in these cases may be caused by confounding factors including attrition, spillover of biophysical processes or effects between treatment- and control-associated areas, and behavioural effects such as the well-known *Hawthorne* and *John Henry* effects, as well as copying of the intervention by individuals in control units (e.g. Glennerster & Takavarasha 2013; chapter 2). These issues may also be encountered in other behaviour-mediated interventions in public health and development economics, for example, and some have claimed that such effects may be large, potentially prejudicing the internal validity of an RCT (Bulte *et al.* 2014). The extent to which these effects may manifest themselves in differing contexts however remains unclear (see Bausell 2015).

Recent years have seen a small number of pioneering RCTs evaluating conservation interventions undertaken. RCTs evaluating payments for ecosystem services have been undertaken in Bolivia and Uganda (Asquith 2016; Jayachandran *et al.* 2017; this thesis) and a further RCT has evaluated the effect of unconditional cash transfers to communities around a national park in Sierra Leone on deforestation rates within that park (Kontoleon *et al.* 2016). However little has been published on the challenges associated with implementing RCTs in the context of conservation, and how implementers may maximise their quality, an increasingly pertinent issue given the fact that more are currently in progress or in preparation (see NSF award 1660481; also see www.3ieimpact.org/en/evidence/impact-evaluations/details/6264/).

In our previous work (chapter 2), we examined the theory surrounding RCT use in other fields, and applied it to RCT design in conservation. We developed a decision tree for use by evaluators illustrating the factors relevant to designing a good-quality and internally valid RCT and illustrated these factors with the case of the RCT of *Watershared*, a Payments for Ecosystem Services-like watershed conservation program undertaken by the non-governmental organization *Fundación Natura Bolivia* (*Natura*) in the Andean forests of Bolivia. The *Watershared* intervention aims to impact a number of outcomes of interest (forest cover, water quality, biodiversity and socio-economic well-being). In this chapter, we return to the *Watershared* RCT, reflecting on the consequences of design and implementation decisions on the ability of the RCT to reveal clear information about the intervention's effectiveness for multiple outcomes of interest. We expect that such learning will be useful for researchers and practitioners interested in evaluation of environmental interventions, especially those with multiple outcomes of interest, via high-quality randomised methods.

Specifically, we examine three principal topics, each of which relates to how spatial relationships between the intervention (in this case the PES-like *Watershared* program) and the design of the RCT complicate meaningful evaluation and risk prejudicing internal validity. Firstly, the selection of randomisation units suitable for one outcome of interest may not work well for other outcomes of interest, imposing boundaries which may be illogical due to the processes of change affecting those outcomes and leading to spillover effects. Secondly, decisions surrounding selection of sites for measuring some outcomes may add complexity given the spatially explicit nature of the theories of change of these outcomes, and can also lead to unavoidable attrition. Thirdly, the voluntary nature of PES and similar interventions complicates evaluation, and we show how inconsistent (and low) uptake of an intervention may make conventional 'intention to treat' RCT results difficult to interpret.

5.3 Methods

5.3.1 **Case study**

In 2010 *Natura* collected baseline data on biodiversity, water quality and socio-economic indicators from 129 communities in the *Area Natural de Manejo Integrado Río Grande – Valles Cruceños* protected area in the Bolivian Andes. These communities (units) were then randomly allocated to the treatment group (65) or control group (64). From 2011 landowners in the intervention group were offered in-kind incentives in the form of agriculture-related goods (fruit trees, barbed wire, cement, irrigation tubing, or other similar products) in return for agreeing to conserve forest and keep cattle away from watercourses under three-year (renewable) contracts. Incentives and enrolment were offered every six months; compliance monitoring took place yearly. More details are available in chapters 1, 2 and 3 and in Bottazzi *et al.* (2018). The intended outcomes of the intervention were reduced deforestation rates and thereby conservation of forest carbon stocks and forest biodiversity, improved drinking water quality in communities supplied by water intakes located in the forest, and increased socio-economic wellbeing of community members. Incentives were offered at three different rates:

'Level 1' - \$10 equivalent per hectare of enrolled land per year, plus \$100 upon joining; in return landowners were required to cease any activities causing deforestation and additionally to prevent cattle and other livestock from entering and contaminating watercourses. Forested land within 100m of a water body was eligible.

'Level 2' - $3 equivalent/ha of enrolled land per year. Eligibility was the same as for level 1 land, but landowners were only obliged to reduce cattle presence over the 3-year period, and then were not allowed to renew a level 2 contract.

'Level 3' - $1 equivalent/ha of enrolled land per year; in return landowners were obliged to cease deforestation but could continue allowing cattle access. All non-agricultural and non-built-up land was eligible.

Community-based environmental education programs on the importance of forest and water conservation were carried out in both control and intervention communities. By December 2014, 2206 hectares were under level 1 *Watershared* conservation agreements, with a total of 51673 ha under *Watershared* agreements of all levels.

5.3.2 **Data sources**

We used a number of separate data sources to conduct the analyses described in this chapter.

Community boundaries and locations: Locations of *Watershared* areas enrolled between August 2011 and December 2014 were taken from *Natura*'s GIS database. Community boundaries were drawn by David Crespo Rocha using the database of land ownership of the Bolivian government agency *Instituto Nacional de la Reforma Agraria* (INRA) as of 2016, with some local validation. Locations of centres of communities were derived by Edwin Pynegar from either school tap location in 2015 or community coordinates from a baseline survey conducted by *Natura* in 2010 and subsequently verified using Google Earth.

Forest: Forest classification was conducted as described in Wiik *et al.* (in review) from RapidEye and Sentinel 2 satellite imagery. The map used for analysis shows the result of this classification as the estimated forest cover in 2011 at the start of the RCT, with pixel resolution of 10m x 10m. Seven communities had to be excluded from forest cover analysis as parts of them lay outside the boundaries of the *ANMI Río Grande* and therefore also the satellite image analysed by Wiik *et al.* Therefore for analyses involving forest cover, N=122.

Stream and catchment delineation: This was conducted using ASTER GDEM v2 30m resolution digital elevation model data and the TauDEM 5.3.7 toolbox in ESRI ArcGIS 10.2.2 (David Tarboton, Utah State University). Streams were assumed to be present in those cells with flow accumulations of 50 cells or more ($\approx 0.45 km^2$) with this criterion based upon field observations. We calculated the catchments of all stream water intakes (N=70; treatment=33, control=31, other=6) measured in the water monitoring conducted between March and May of 2015. (Groundwater recharge zones associated with intakes that are springs cannot be straightforwardly or accurately determined using this method.)

Water quality monitoring (water intakes): Water intake selection for monitoring was conducted as described in chapter 3. Monitoring was conducted at the water intake which supplies the school tap or the majority of households in the community, and at the school tap in each community. This monitoring was conducted four times, in 2010 (baseline), 2012, 2014 and finally 2015 (endline). Coordinates were recorded by GPS for every monitored site throughout all rounds of monitoring.

Many communities have more than one intake. For a subset of the communities (Huertas, Masicurí, Saguintal, Falda de la Cebada and Chorrillos), the locations of all functional intakes as of October 2015, and the number of households they supply, were mapped by Patrick Bottazzi and David Crespo Rocha and the position recorded by hand-held GPS.

Biodiversity monitoring: Transects were established for the biodiversity monitoring in two locations, one in the closest forested land directly upstream or uphill of the water intake monitored in 2010 (the principal site) and a nearby secondary site 30m upstream or uphill of the end of the first set of transects or parcels (see Vidaurre & Gonzalez 2011 for a full description). Biodiversity surveys for amphibians and scarab beetles were conducted at both sites in each community in 2011 and 2016. For the spatial analysis presented we use the location of the principal site in each community only.

Socio-economic survey: A household survey covering livelihoods, community institutions and trust was conducted in all communities in 2010 before communities were allocated to control or intervention groups. A modified endline survey (with a greater focus on water supply systems and institutions for water management) was conducted between October 2015 and June 2016. Heads of all households found within the RCT communities were intended to be surveyed at both baseline and endline. Achieved follow-up rate was 55.6% (N=1459 out of 2623 in baseline, treatment community household N=884, control community household N=575). A more detailed description of the survey protocol is available in Bottazzi *et al.* (2018), and with the data in the repository http://reshare.ukdataservice.ac.uk/852623/.

Two specific questions were added to the endline survey to investigate the potential for existence of copying on the part of control community members. For those households whose water comes from an intake protected from cattle, we probed where the idea for protecting the water source came from.

Who suggested that you should protect this water source? (possible answers: it was our own idea; the community; *Natura*; *ICO* (*Instituto de la Capacitación del Oriente*, another NGO which has worked in the area since 1981); public authorities (national or municipal government); we saw it in a neighbouring community; other [specify]).

If the interviewee responded with "We saw it in a neighbouring community", we then asked the follow-up question:

Which community did you see it in? / A member of which community suggested it? (answer: [Community specified]).

If a head of household belonging to a control community were to name a treatment community here, we considered this evidence of copying phenomena.

5.3.3 **Analyses**

Spatial calculations and analysis were conducted throughout using ArcGIS 10.2.2 and QGIS 2.6. We estimated how much land was eligible for enrolment at the beginning of the experiment by calculating how much land was within 100m of a delineated stream and also forested, according to the 2011 classified satellite image. We then calculated the percentage of the catchments which are under level 1 *Watershared* agreements, and calculated the difference in levels of catchment conservation between catchments supplying treatment and control communities, as well as the proportions of forested land, level 1 eligible land, and level 1 enrolled land within the boundaries of each of the community boundaries. We calculated the level of biophysical spillover by noting which water intakes and principal biodiversity monitoring sites were located outside of the communities to which they related, and also calculating the proportions of catchments supplying monitored intakes which were not physically located within the communities which they supplied.

We calculated the number of water intakes that have remained at the same site throughout each of 4 rounds of water monitoring (in 2010, 2012, 2014 and 2015) in the 129 communities included in the RCT by taking the coordinates for these sites from the respective databases and calculating distances between them; if they were less than 2 times GPS error apart (approximately 60m) we accepted them as being the same site. This was verified by technicians from *Natura* who had conducted this monitoring and had personal experience of the sites in question.

We used QGIS 2.6 to produce maps showing the potential for incongruence of randomisation units and the consequences of the voluntary nature of the intervention, using the example of the community of Piraimirí. We also produced a map showing the problems associated with movement of intakes for the community of Huertas. Figures showing the distributions of relevant results from the spatial analyses by community or (in relevant cases) by catchments were produced using the ggplot2 package in R (Wickham 2009; R Development Core Team 2014).

5.4 Results

5.4.1 Multiple outcomes of interest make selection of a meaningful randomisation unit difficult

The *Watershared* intervention aimed to impact four principal outcomes of interest (forest cover, water quality, biodiversity and socio-economic well-being). Each of these has its own ideal randomisation unit (table 5.1).

The *Watershared* RCT selected community as the randomisation unit for the experiment (i.e. it was communities which were randomly allocated to the control or treatment groups). This worked well for socio-economic wellbeing. However for water quality, the appropriate unit is catchment (or hydrological recharge zone in the case of spring intakes). These are not necessarily congruent with community boundaries and may be partially or wholly outside of the community to which they supply, as well as being determined initially by the monitoring site (water intake supplying school). For measuring the impact of the intervention on deforestation or biodiversity, the appropriate unit would be forest cover within the boundaries of communities at baseline (for biodiversity, considered to be habitat).

In figure 5.1a we highlight the implications of the mismatch between the ideal randomisation unit for each outcome (table 5.1) and the compromise implemented in the *Watershared* RCT for forest cover, biodiversity, and water quality, using the example of the community of Piraimirí. 86.8% of the land area of this community was forested in 2011, so the ideal randomisation unit for forest cover and biodiversity outcomes is aligned with, but not identical to, the randomisation unit selected (community). In the case of water quality, the ideal unit (catchment above the community's water intake supplying the school tap) takes up a tiny proportion of the community's area (5.0%) and is not aligned with its borders (36.9% of the catchment in question is located within a neighbouring community, Monte Pablo). This means that the area in which land enrolled in *Watershared* agreements can influence measured water quality (and/or quantity) is extremely small, and part of the catchment of interest being in a neighbouring community will be likely to result in biophysical spillover, reducing the potential effect size of the intervention and breaching the assumption of independent units in RCTs (e.g. Rubin 1974).

Table 5.1. The selected randomisation unit – community – presented problems for many of the intended outcomes of the *Watershared* intervention. In RCTs of interventions with multiple kinds of outcome of interest, it is likely to prove difficult or impossible to select an appropriate randomisation unit for all outcomes.

Outcome of interest	**Ideal randomisation unit**	**Selected randomisation unit**	**Monitoring measure/site selected**	**Issues**
Forest cover	Forested land within boundaries of community	Community	Forested land within boundaries of community	Individuals belonging to one community were in practice able to enrol land physically located in another community, because of original lack of clarity over location of community boundaries. This therefore can lead to a biophysical spillover effect.
Biodiversity	Habitat for monitored species (scarab beetles and amphibians) within boundaries of community	Community	Transect of sites directly upstream/ uphill of monitored community water intake	Organisms may move within their habitat (e.g. scarab beetles have large ranges within forested patches [Zimmerman & Bierregaard 1986; Klein 1989; Dale *et al.* 1994]). Where habitat overlaps community boundaries – the spatial manifestation of the randomisation unit – such movement may lead to biophysical spillover. The choice of monitoring site proved problematic. Sites for monitoring had to be selected in a systematic fashion at baseline (before the intervention was offered). However many of these sites in treatment communities were never entered into a Watershared contract (as the intervention is voluntary and evaluators cannot know which areas of land community members will), meaning there was no reason to expect any change in biodiversity at monitored sites from the intervention.
Water quality	Watershed above intake at which water quality is measured	Community	Community water intake supplying school tap, school tap	In some communities the watershed of interest may fall entirely within the community's boundaries; in others it may extend into neighbouring communities, risking biophysical spillover. Monitoring site choice implicitly selects the randomisation unit, as this determines the watershed of interest. The school tap was selected as this water supply was thought to have greatest relevance for public health in the communities. However, there is not necessarily a link between implementation of the intervention and the watershed above the intake feeding that tap (land outside the watershed may be enrolled, but cannot possibly affect water quality); there may be multiple (or no) intakes in a community and/or sharing of intakes between communities; communities may move their intakes over time.
Socio-economic wellbeing	Community, or household within community (depending upon outcome of interest)	Community	All households within community	Heads of all households surveyed, but households may move away or households may cease to exist due to deaths, and new households may be formed during the time period of the evaluation, leading to dropout/attrition. 55.6% of households surveyed in the baseline survey were also located to be surveyed in the endline. The fact that multiple individual households exist within each randomisation unit means that the RCT is a cluster trial for household-level outcomes, so correct calculation of effect sizes on these outcomes requires more complex statistical analysis (Feng *et al.* 2001; Donner & Klar 2004).

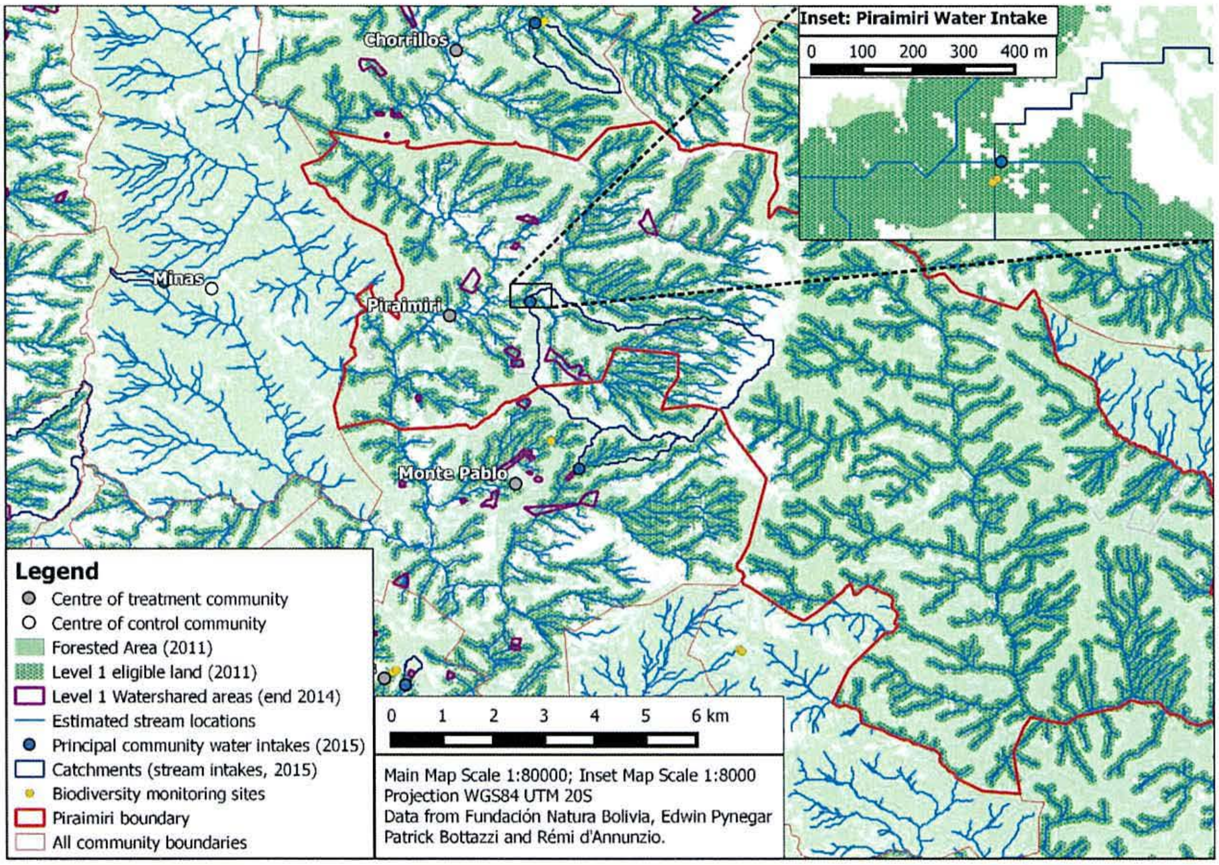


Figure 5.1. *Watershared* mapped in the community of Piraimirí. (a) Not all randomisation units are aligned and measurements of water quality may be at risk of spillover. (b) The voluntary nature of the intervention meant that only a small proportion of the eligible land was actually enrolled in agreements, including (for the water quality outcome) in some cases no land at all was enrolled in catchments above the monitored intake site.

This is a general issue across all communities. A mean of 71.5% of the land in each community for which we have data was forested, but this varied markedly from community to community (min 1.6%, max 94.1%; figures 5.2a and 5.2b). The catchments above the intake monitoring sites only covered very small proportions of the area of the communities to which they supplied water (mean 5.6%, min 0.0%, max 50.2%; figure 5.2c).

Incongruent randomisation units led to potential for biophysical spillover effects. Only 27 of 66 (40.9%) of the catchments above intakes supplying single communities – water quality randomisation units – fell entirely within the boundaries of the communities which they supplied, and in 11 of 66 (16.7%) of cases were entirely outside the community in question. The consequence of this was a biophysical spillover effect: of the 31 catchments above intakes supplying control communities, 6 of these (19.3%) contained at least some land belonging to members of a treatment community and enrolled in a Level 1 *Watershared* area.

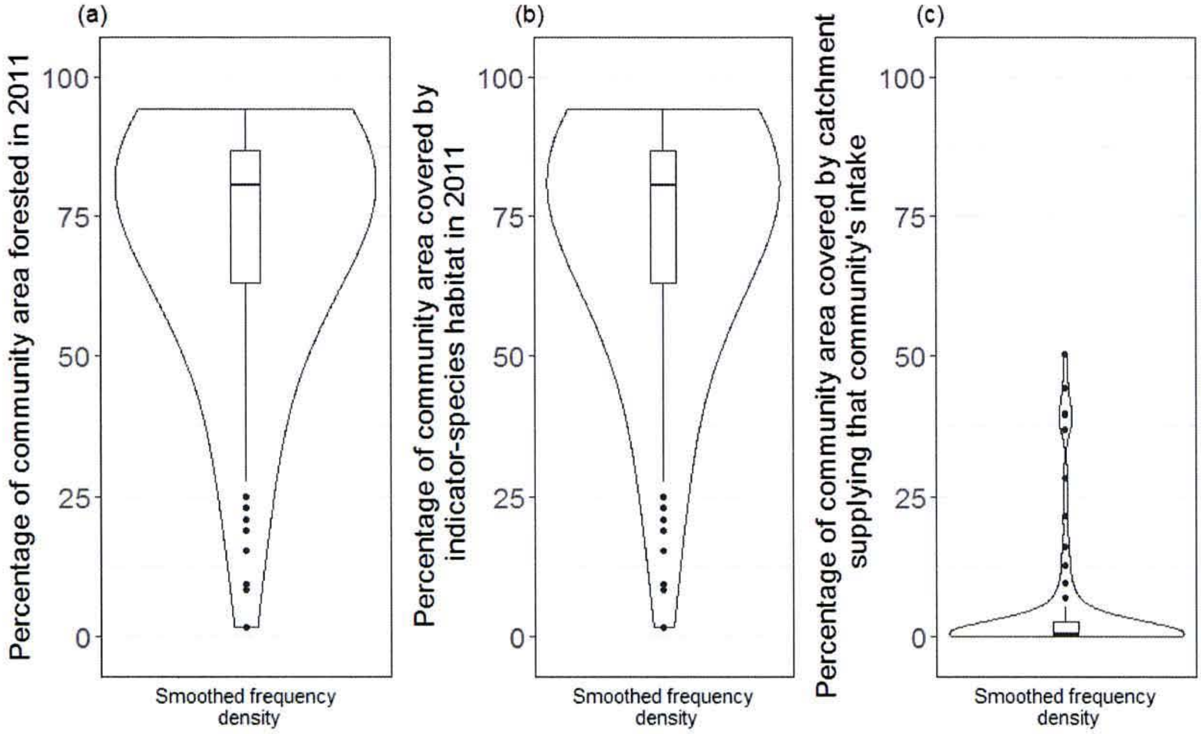


Figure 5.2. Randomisation units do not necessarily align for different outcomes of interest. Forested land area in each community (the randomisation unit for forest cover; [a]) and indicator-species habitat in each community (the randomisation unit for biodiversity; [b]) are identical as both are understood to be the forested area within community boundaries. Areas of catchments supplying monitored intakes (water quality randomisation unit; [c]) in general cover much smaller areas of the communities to which they supply. The area of this unit may also spill over into neighbouring communities. The width of the violin plot in each panel of the figure represents a smoothed, normalised density function of the number of units at each level of the respective y-axis.

5.4.2 Monitoring site selection procedure must be logical and consistent throughout all units, but real-world complexity can result in attrition

A priority for all RCTs is keeping the units at baseline within the trial up until the endline data has been collected, as dropout or attrition not only reduces sample size, increasing the minimum detectable effect size, but can also result in bias if there is a tendency for certain types of unit to drop out (e.g. Glennerster & Takavarasha 2013). In the *Watershared* RCT, the selection of sites at which to measure outcomes of interest has implications for dropout. Complete satellite images were not available for 7 of the communities so these were excluded from analyses requiring forest cover data, but there is no reason to expect that these contained any systematic bias, and 122 of 129 (94.6%) of the units remained included. A significant number of households could not be re-interviewed at endline (due to outmigration, death of the household head, or were simply not available to answer questions) resulting in attrition of 44.4% of the households monitored (Bottazzi *et al.* 2018).

The selection of appropriate monitoring sites for recording baseline and endline data posed real challenges for the biodiversity and water quality and outcomes. The challenge for defining consistent sites for monitoring water quality is firstly that the relationship between water intakes and communities was not one-to-one. In communities with more than one intake supplying them, RCT analysis remained feasible although monitoring results were inevitably only partially informative of drinking water quality in that community. However in three cases, one intake (or two intakes within 30m of each other in the same stream) supplied two communities. This meant that measurements from those intakes and the supplied taps were not attributable to treatment or control groups in the RCT analysis of water quality, and so these data had to be discarded, elevating the attrition rate. Secondly, some communities changed the intake supplying the school tap (or were forced to, due to problems with the existing intake) during the period over which the RCT ran. In such cases comparing baseline and endline measurements of water quality was not meaningful. The number of intakes which remained the same between the initial baseline water quality measures in 2010 (N=125) and endline in 2015 was only 47 (table 5.2). In figure 5.3 we show an example of the issues surrounding water quality monitoring site selection for the community of Huertas.

Table 5.2. Number of intakes feeding school tap which remained in the same location between each possible pair of rounds of water monitoring undertaken in the 129 communities included in the RCT. This shows that no comparison between pairs of years is complete, and that some pairs had less than half of the monitored intakes in common.

	2010	2012	2014	2015
2010		56	46	47
2012			100	99
2014				104
2015				

Biodiversity baseline monitoring, in 2011, was undertaken upstream or uphill of the water intakes monitored in 2010. The endline monitoring in 2016 was carried out at the same location, even if the nearby water intake was no longer used and regardless of whether the specific site had become covered by a *Watershared* agreement in the interim. While this does not lead to attrition, it does lead to problems for evaluation derived from the voluntary nature of the intervention (see below). Also, intakes being located outside of the communities they supplied meant that principal sites for biodiversity monitoring also were located outside of the community to which they were supposedly associated in 17 out of 127 (13.4%) of cases (a potentially avoidable kind of spillover effect, but also a consequence of a decision to monitor systematically following a particular protocol).

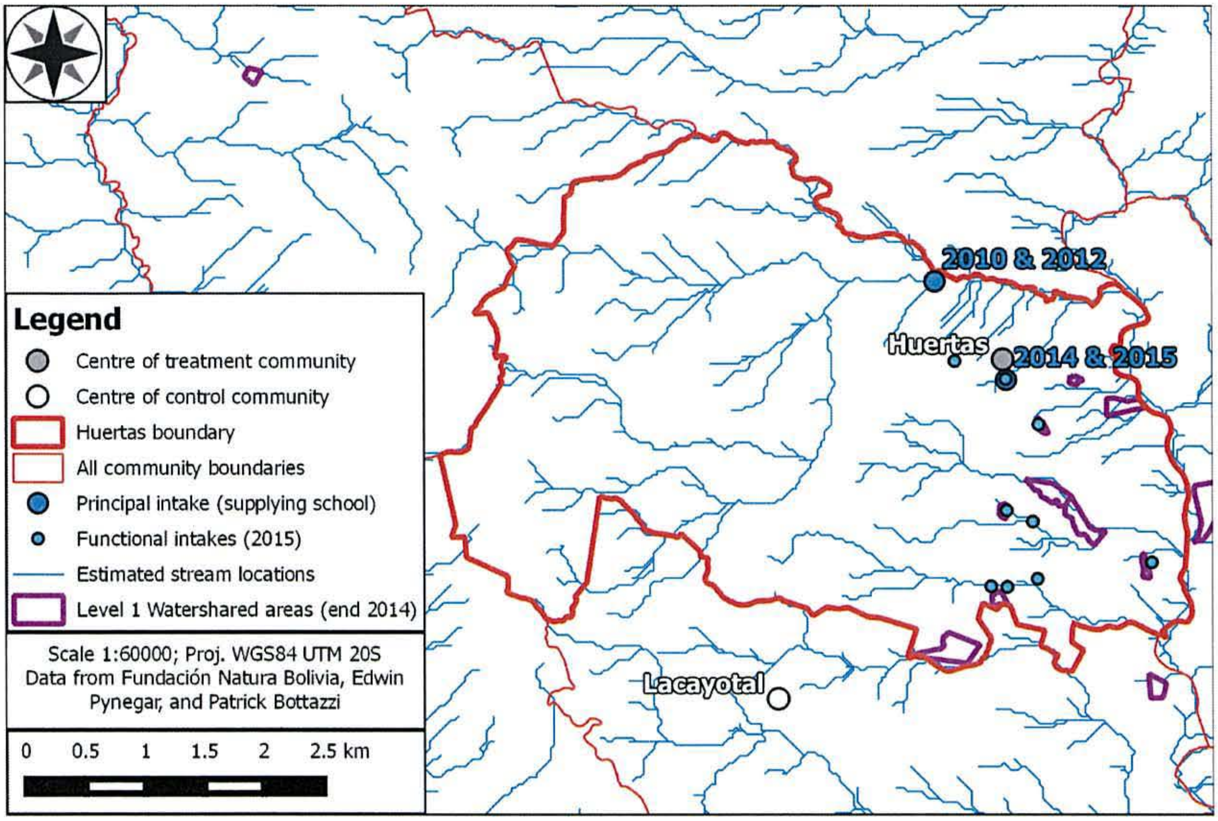


Figure 5.3. Multiple water intakes and water intake movement presented problems for monitoring in the community of Huertas. Nine water intakes supplied the community of Huertas as of October 2015. The intake currently supplying the school tap has only done so since 2014, and the intake used previously to that is no longer functional. This demonstrates how, in this community, comparisons of water quality cannot meaningfully be made between measurements taken in 2012 or earlier, and measurements taken in 2014 and later, because of the (necessary and unavoidable) movement of the principal intake.

5.4.3 Voluntary participation in an intervention poses challenges for evaluation by RCT

5.4.3.1 Uptake of voluntary interventions may be low, making evaluation challenging

Watershared is a voluntary intervention. Landowners in treatment communities may enrol any area of land larger than 0.5 ha, anywhere in their community, as long as it meets the eligibility criteria. We estimated that a mean of 1637 hectares (39.1%) of the area of each of the treatment communities was eligible for enrolment as level 1 *Watershared* areas (min 17.6 ha, max 11637 ha, N=62; proportionally min 1.4%, max 53.7%, N=62). The consequence of the intervention's voluntary nature for the RCT is that the amount of the intervention applied varies markedly between treatment communities. We found that a mean of only 3.8% of that eligible land was actually enrolled in level 1 agreements (min 0.0%, max 17.8%, N=62; fig. 5.4a). If we consider all land in communities, then the proportion of it enrolled in level 1 areas is even lower. In one treatment community, Candelaria, no one elected to enrol any land in level 1 contracts.

5.4.3.2 *Uptake will not necessarily occur in areas most likely to result in impact*

Enrolment of some areas of land is likely to be more important for achieving detectable changes in the biodiversity and water quality outcomes than others, because of the mechanisms by which we might expect the implementation of *Watershared* agreements to feed through to outcomes of interest.

The voluntary nature of the intervention means that implementers could not predict *which* areas would be enrolled in *Watershared* agreements. This therefore makes it difficult to know where to meaningfully monitor biodiversity at baseline as it is not possible where a change could be expected. The decision to monitor biodiversity indicators on a transect directly upstream of the community water intake was based upon the assumption that communities would be most likely to put such land into contracts because of the potential benefits in terms of water quality (N. Asquith, pers. comm.). However we found that of 64 principal biodiversity monitoring sites associated with treatment communities, only 14 of these (21.9%) were actually located within level 1 *Watershared* areas by the end of 2014 (and one of these was physically located in a different community to that with which it was supposedly associated). Therefore impacts of the intervention (enrolment in *Watershared* agreements) on biodiversity could only reasonably be expected to be observed in about a fifth of the communities where baseline data was collected and *Watershared* was offered.

An impact of the intervention on water quality might only be expected if a significant proportion of the catchment above the monitored intake is under *Watershared* agreements. However, this proportion was invariably low with an average of only 3.6% (N=33; min=0.0%; max=27.1%) in treatment communities (and the catchments above the intake supplying 17 out of 33 treatment communities having no level 1 land within them at all; figure 5.4b).

We also calculated the proportion of level 1 land enrolled associated with a community, located within the catchment feeding that community's school tap in 2015, and compared this with the total amount of enrolled land within the community. This was to establish whether there was evidence that members of treatment communities had targeted level 1 *Watershared* areas to be within catchments supplying their tap. Only 14 catchments had land within them enrolled by community members from that same supplied community, and we found no indication that land was preferentially enrolled in those catchments (n=25, p=1, two-tailed one-sample sign test).

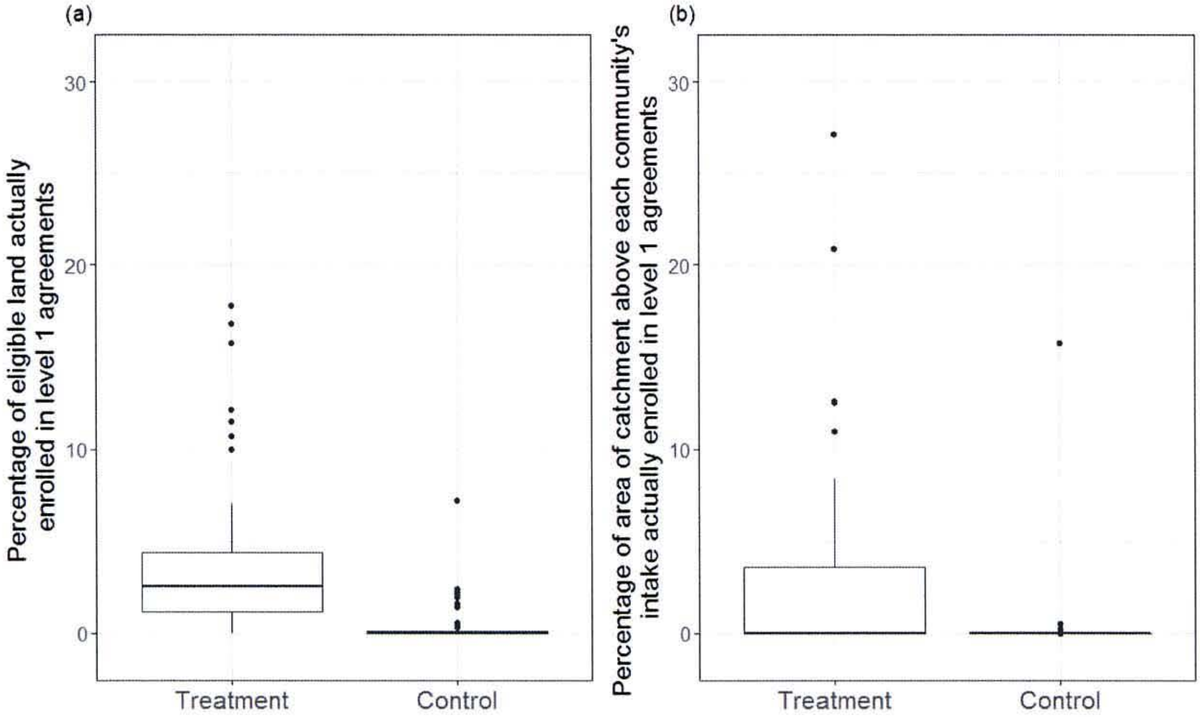


Figure 5.4. The voluntary nature of the intervention has consequences for evaluating the effectiveness using an RCT. (a) Only a small proportion of the eligible land was actually enrolled in level 1 *Watershared* agreements (enrolment is not zero in control communities due to people owning land in communities which are not those where they live). (b) Only a small (and highly variable) proportion of catchments supplying school taps in treatment communities were covered by Level 1 *Watershared* areas.

5.4.3.3 Real-world complexities (and human error) may reduce differences between control and treatment

People living in a treatment community (and therefore offered the chance to enrol their land in *Watershared*) may also own land in a control community. 25 out of the 63 control communities for which we had boundaries had at least some level 1 *Watershared* land located within them, and in total 4.6% (99.9 ha) of the L1 land enrolled between 2011 and 2014 was located within the boundaries of control communities (figure 5.4a). Also, as discussed above, catchments are not aligned with community boundaries. These factors combined meant that there was not zero enrolment of land in the catchments supplying control community water intakes; in fact a mean of 1.0% of land (N=31) in control communities' associated catchments was under level 1 agreements (figure 5.4b). However despite this the proportion of conserved land was significantly different between treatment and control community groups (N=64, p=0.009, Wilcoxon rank sum test).

These consequences of voluntary interventions can also be seen in figure 5.1b for the community of Piraimirí and its neighbours. In Piraimirí, we estimated that 5210 ha (47.4% of total area) of forest was eligible for enrolment in level 1 *Watershared* areas. However by December 2014 a mere 45.3 ha (0.9%) of that eligible land had actually been enrolled. This did not include the biodiversity monitoring sites. Of this 45.3 ha, 15.8 ha was located within the catchment supplying the intake, representing 1.8% of the catchment and 2.8% of the catchment located within the boundaries of the community itself. However the enrolled land located in the watershed is 34.9% of all the enrolled land within Piraimirí, representing a near six-fold over-representation of level 1 *Watershared* land within the most important watershed itself. However, in the neighbouring communities of Monte Pablo and Chorrillos, none of the *Watershared* areas in these communities are located within the catchments supplying them at all.

5.4.3.4 *Evidence of copying phenomena was found*

We encountered one case in which copying appeared to have occurred. One head of household belonging to the control community of La Senda (Moro Moro municipality) responded that they elected to protect forest or watercourses because they had seen it, or been told about it, in the neighbouring treatment community of Abra del Astillero (Moro Moro municipality) (total control community household N=575).

5.5 Discussion

We have examined a number of challenges we encountered when analysing a pioneering large-scale RCT of a socio-ecological Payments for Ecosystem Services-like intervention. Here we draw out some key principles, with reference to literature from a range of fields. We hope that these will be instructive for evaluators interested in using RCTs to evaluate socio-ecological interventions and will improve the quality of future RCTs.

5.5.1 **Selecting appropriate randomisation unit is challenging when there are multiple outcomes of interest**

Evaluators designing and undertaking RCTs must decide which unit should be used for randomisation. Theory suggests that the unit selected should logically reflect the theory of change behind the intervention; for example health-education programs may use schools or clinics as the randomisation unit as the intervention is most easily implemented at that level, and spillover based upon reinfection or copying based upon information sharing could occur if the intervention were randomised at an individual level (e.g. Milsom *et al.* 2006). However, conservation interventions often have a number of desired outcomes, for example, achieving improvements in both ecological indicators and human wellbeing has been the justification for interventions such as payments for ecosystem services

(Ferraro & Kiss 2002; Turner *et al.* 2012; Ingram *et al.* 2014; Kolinjivadi *et al.* 2015). Donors and other funders are also likely to look favourably on, or indeed demand, interventions promising improvements in both environmental and human welfare indicators of interest, given the perception of close links between conservation and development (e.g. Roe *et al.* 2013).

The randomisation unit selected in an RCT has to work for all outcomes of interest. However, drawing any boundary in an open system such as a landscape is informed by how problems and solutions are perceived (and these boundaries can then reinforce those same perceived problems and solutions in turn [Brown 2010; Santos de Lima, Krueger & García-Marquez 2017]). Implementers of interventions may in general be able to work around such issues, conceptualising different (appropriate) boundaries and different stakeholders for different outcomes. In the 'one-unit-must-fit-all' context of an RCT, this is not likely to be feasible. Different outcomes will inevitably have different ideal randomisation units associated with them, given the way the intervention is expected to lead through to desired outcomes for each (and this is a problem that may not be shared with quasi-experimental evaluation methods such as statistical matching, as in that case the evaluator may be able to construct separate controls for each outcome). Any randomisation unit eventually selected may open up the potential for biophysical spillover for any outcome of interest with an ideal randomisation unit not congruent with that selected. For example, in this case, level 1 *Watershared* areas belonging to treatment communities may be located within the catchment of a water intake supplying a control community. Worse, an RCT implemented with an inappropriate randomisation unit for a particular outcome may almost automatically result in the intervention not having any impact on that outcome, if the unit forces implementers to attempt to solve problems working with the 'wrong' individuals or in the wrong areas. It is worth noting that the only other currently published RCT of PES restricted itself to forest cover/carbon as its outcomes of interest (Jayachandran *et al.* 2017).

5.5.2 Selecting monitoring sites at which measurements adequately capture changes caused by the intervention can be challenging

High-quality monitoring is costly, and evaluators often find themselves having to make inferences on the true status of variables of interest based upon samples taken at sampled locations at specific points in time (Biggs, Carpenter & Brock 2009; Santos de Lima, Krueger & García-Marquez 2017). However, outcomes of interest must be measured in a consistent fashion so that changes can be detected if they exist (e.g. Sommerville, Milner-Gulland & Jones 2011; Glennerster & Takavarasha 2013). This presents evaluators with a decision to make surrounding *where* to monitor such that the status of the outcome being monitored is captured as fully and as relevantly as possible. Those outcomes with theories of change in which spatial layout of the enrolled areas is likely to be key to achieving changes, represent those for which selection of monitoring site can have a major influence on the evaluation. In the case of *Watershared*, these were the water quality and biodiversity outcomes.

Firstly, monitoring at a particular site within a randomisation unit may not fully capture the status of the outcome in question throughout that unit. We see this with biodiversity (the abundance and diversity of indicator taxa at one particular location located directly above the monitored water intake may or may not be representative, as that location may be more or less disturbed than the 'average' location in that community's associated area). In the case of water quality, a community having multiple water intakes (as was the case in the village of Huertas for example) meant that monitoring as conducted can only be partially informative of the status of this outcome. There was no location which could be monitored which represented water quality as supplied to the whole of the community (the 'ideal' one-to-one relationship between monitoring site and randomisation unit did not exist in all of the communities in the RCT).

Secondly, attrition may result from unavoidable changes in monitoring site, as we encountered in the case of water quality when communities elected to, or were forced to, change the main water intake. This attrition is a general issue with RCTs: while the method generally tolerates endogenous changes within randomisation units, based on the principle that these changes will be taking place at the same rate in units in treatment and control groups, this is not the case when the changes are affecting some feature which precludes measurement in a consistent fashion. Such site changes in practice equate to an enforced change in monitoring protocols, meaning earlier data has to be discarded.

Thirdly, monitoring site selection directly determines which areas' status can be important for causing a detectable change in the monitored outcome of interest: in the case of biodiversity the habitat patch or patches containing the monitoring site, and in the case of water quality the catchment or hydrological recharge zone above the monitored intake. This has clear implications given the voluntary nature of the intervention (see section 5.5.3 below) as this determines which of the *Watershared* areas enrolled actually could affect the outcomes of interest.

5.5.3 **Voluntary implementation of an intervention makes RCT evaluation particularly complex**

In practice, many socio-ecological interventions one might seek to evaluate with an RCT are voluntary, and individuals may fail to take up an intervention allocated to them through lack of interest, or may fail to complete the intervention's implementation for a whole host of reasons. This leads to dropout and thus attrition. Evaluators are familiar with the complications for RCT analysis derived from interventions being of a voluntary nature, and therefore analyses can consider both the impact on the *intention to treat* population (those offered the intervention) and *treatment on the treated* (those who took up the intervention in full) effect sizes (see Glennerster & Takavarasha [2013]). Almost all socio-ecological conservation interventions are voluntary; for example, farmers can choose whether to enrol land in an agri-environment scheme (Morris & Potter 1995), communities may choose to engage with community forest management or not (e.g. Dolisca *et al.* 2006), or to develop community-based ecotourism initiatives (e.g. Liu *et al.* 2014).

5.5.3.1 *Level of implementation of the intervention is highly variable, making RCT analysis difficult to interpret*

Incentive-based conservation adds a further level of complexity, as individuals offered an intervention decide themselves to what extent they wish to implement it. In the case of *Watershared*, landowners can theoretically enrol in the intervention anything from zero hectares to all eligible hectares of land that they own. Also, exactly the same action may have greater or lesser effects on each outcome of interest (as monitored) depending on *where* the action takes place. (In recognition of this, many agri-environment schemes target areas believed to be particularly important with the goal of improving cost-effectiveness [e.g. Uthes *et al.* 2010]).

While this issue is not specific to randomised trials, it does have specific consequences for RCT analyses. Conventional RCT-type evaluations, in which treatment or control is a dummy variable in a regression or linear model (see Glennerster & Takavarasha 2013), may not deal well with it. In a conventional *intention-to-treat* analysis, the treatment will have to be considered the *offer* of the intervention to be evaluated, but the *de facto* intervention directly affecting the outcomes of interest will be whatever the individuals in treatment groups decide to implement. Levels of 'treatment' in practice will therefore vary depending on how much, and where, landowners elect to conserve their land, and so simple treatment/control comparisons erase a great deal of important information.

Our results show that randomisation did succeed in imposing exogenous variation between the treatment and control groups in the *Watershared* RCT as intended, as there were highly significant differences in the amount of level 1 land enrolled in communities, and in the catchments above monitored water intakes, between the treatment and control community groups. However, variable implementation of the intervention did indeed occur. Over half of treatment-associated intakes had no level 1 land at all in their associated catchments (and thus *in practice* were untreated) while a not insignificant proportion (6 of 31) of control-associated intakes had at least some level 1 land in their catchments (and thus were partially treated, whether due to implementer error or to spillover). This (inevitable) variable implementation acts as a large source of error, and thus would greatly increase the minimum detectable effect size in any analysis. (Perhaps unsurprisingly, in chapter 3 we found no significant effect over time of the intervention on water quality as evaluated with an intention-to-treat RCT.)

5.5.3.2 *Inadequate incentives and absence of targeting result in low levels of implementation in key areas, and consequentially reduced likelihood of detecting a significant effect*

Overall level of conservation of eligible land and, of the catchments above the monitored intakes, was low. Even the catchments associated with water intakes in treatment communities had only a small proportion under level 1 contracts (mean 3.6%; over half contained no level 1 areas at all). We also found no evidence of a focus on enrolment of land in these particular catchments. This low level of uptake, combined with no targeting of contracts to the most important land such as that surrounding water intakes and/or land in the catchments of interest, may well explain why no significant effect of the intervention on water quality was found when analysed by a standard RCT method (see chapter 3).

Implicit here is a broader issue which will be of major importance when considering any voluntary intervention with spatial aspects associated with its theory of change. The relationship between level of implemented action and the effect on each respective outcome of interest will be key for understanding potential effectiveness (Wong *et al.* 2015; Ponette-González *et al.* 2015). In some cases, actions may achieve desired outcomes easily (where only a few actions result in a large change in the outcome of interest), or the reverse. This can be conceptualized in a graphical model of the relationship between actions and outcomes (figure 5.5). *Watershared* is intended to improve water quality through conservation of catchments above intakes, so the level of conservation required to achieve this change is key[1]. Forest conservation and cattle exclusion do lead to reductions in faecal contamination of watercourses (Line 2003; Vidon *et al.* 2008; Sunohara *et al.* 2012), but these studies evaluate contamination above and below fully conserved small parcels and so may be of little relevance for a voluntary intervention focused at the catchment scale. (In particular, we frequently encountered level 1 *Watershared* areas on one side of a watercourse, which represented the boundary of one farmer's land; another farmer's land extended on the other side, which was not under conservation. As a consequence, livestock could continue to enter the watercourse. The level 1 area in the catchment of the intake of Piraimirí in figure 5.1 is an example of this. This again highlights

[1] This is an example of a well-known, but inadequately studied, challenge with payments for watershed services programs (e.g. Wong *et al.* 2015). Participants may conserve forested land in the headwaters of catchments with the intention of maintaining or increasing water availability and/or quality downstream (e.g. Porras *et al.* 2008; Abell 2017); however, the question of how much of the catchment would need to be conserved, and in what conformation, to achieve a significant improvement in these indicators is context-dependent and is largely unclear (c.f. Bruijnzeel 2004; Le Tellier, Carrasco & Asquith 2009; Beck *et al.* 2013; Ponette-González *et al.* 2015).

the importance of spatial conformation for effectiveness and the consequent likely non-linear nature of the relationship between action and outcome.)

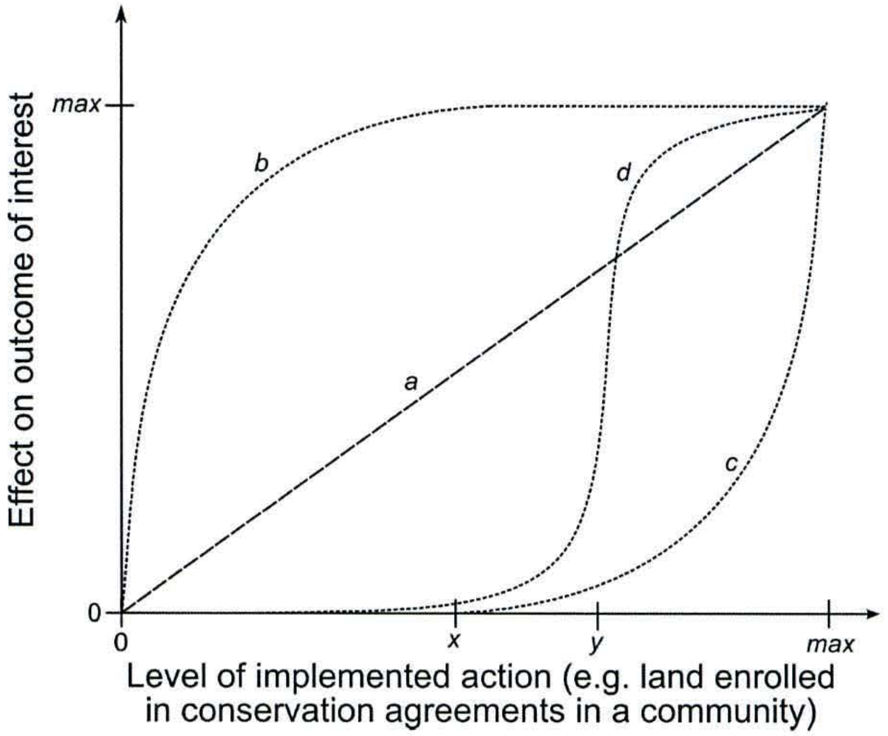


Figure 5.5. Conceptual model of the relationship between conservation actions and outcomes of interest. Effectiveness of a voluntary conservation intervention will vary depending upon how intensively the intervention is implemented, and for some kinds of outcome, the location of that implementation. ***a*** on this figure assumes that all instances of the implementation of the intervention have equal effects on the outcome of interest. ***b*** represents a case where the intervention is able to trigger major improvements in the outcome of interest with relatively low levels of the intervention, or where the implementation of the intervention is highly spatially targeted to critical locations. In ***c***, low levels of implemented actions do not improve the outcome of interest but once a high level of implemented action is reached, this changes. ***d*** represents a tipping point dynamic in which levels of action up to ***y*** have minimal effect, but then the system may shift into a more favourable state when the intervention is implemented beyond ***y***. This may be the case for species requiring a certain minimum habitat size or a minimum level of connectivity to survive or reproduce, in the case of a biodiversity outcome, or in the case of water quality, there may be a tipping point for proportion of land in the catchment enrolled, below which the effect is minimal, but above which the effect obtained is large (W. Buytaert, pers. comm.).

It is not clear from the existing literature what proportion of a catchment under conservation could be expected to achieve a significant improvement in water quality. However other measures of Andean stream health have shown that >70% of catchment area may need to be covered by good-quality forest (Iñiguez-Armijos *et al.* 2014), and that the status of the catchment as a whole, rather than solely riparian buffer or indeed the nearby reach, is key for determining ecological status (Allan 2004). Regardless it seems very likely that an adequate 'treatment' expected to improve water quality is well above the actual percentages of land conserved in catchments of interest.

Uptake of the intervention (in general) is low likely because the incentive is insufficiently attractive, which even for level 1 areas compares unfavourably in financial terms with some other payments for watershed services-type programs in Latin America (Mexico's PSA-H pays 27 USD/hectare/year for primary forest and 36 USD/ha/yr for cloud forest [Muñoz-Piña *et al.* 2008]; Costa Rica's national PES pays 45 to 163 USD/hectare/year [Wunder, Engel & Pagiola 2008]). The Ugandan PES program analysed in the RCT by Jayachandran *et al.* (2017) paid landowners 28 USD per year per hectare of forest. Bottazzi *et al.* (2018) show that those signing *Watershared* agreements have multiple motivations for doing so, of which the delivered incentive is often not the most important, and indeed the theoretical framing of *Watershared* is that of reciprocity and 'nudges' rather than opportunity costs and ecosystem service 'sales' (Asquith 2016). As a result *Natura* claim that the low level of incentive is of relatively low importance. However, Bottazzi *et al.* also show that the program may *facilitate* conservation actions which landowners had a pre-existing motivation to make anyway, but that therefore it may not be sufficient to cause landowners to place large areas in conservation for the benefit of the community as a whole. Both theory (Martin Persson & Alpízar 2013) and empirical data (Arriagada *et al.* 2009) do predict that low incentives lead to low participation, or (at best) enrolment of non-additional areas. Implementers may be able to achieve increased participation in those key areas through increasing the amount of incentives offered (see Börner *et al.* 2017). Alternatively, the intervention could have included spatial targeting towards those catchments or areas considered important (Ezzine-de-Blas *et al.* 2016), and/or an agglomeration incentive to ensure high levels of implementation in those areas (Parkhurst & Shogren 2007; Engel 2016). While this may be more cost-effective, and spatial targeting is known as a predictor of program effectiveness (Ezzine-de-Blas *et al.* 2016), different key areas may need to be targeted for different outcomes of interest, meaning the intervention's design would become substantially more complex, which in turn increases transaction costs.

Consequently, the low levels of implementation of the intervention in key catchments leads to a minimal influence of the intervention on the outcome of interest, and very little potential for the effect of the intervention to be disaggregated from other biotic, abiotic, and anthropic factors which influence outcomes (c.f. Barnaud & Antona 2014; Santos de Lima, Krueger & García-Marquez 2017). This relative lack of success of the intervention as designed shows the importance of formative evaluation – or trialling and progressive improvement – of the intervention (Rossi, Lipsey & Freeman 2004) before the undertaking of a summative evaluation as represented by an RCT. If the design of the intervention's implementation had resulted in treatment-intake catchments being largely conserved, the problem of variable implementation and similar implementation levels in treatment and control discussed in section 5.5.3.1 would not have arisen to such an extent. In that situation a straightforward treatment-control comparison would have been much more meaningful.

5.5.3.3 Copying (and simple implementer error) may prejudice internal validity

Copying is a phenomenon familiar to RCT implementers, in which individuals belonging to control units may see or hear about an intervention being undertaken in treatment units and implement aspects of the intervention in their own communities, believing that the intervention to be beneficial (e.g. Glennerster & Takavarasha 2013; chapter 2). While this may be desirable for implementers of interventions, as a 'positive' type of spillover, it goes against the assumption of independent units integral to an RCT (Rubin 1974), and risks resulting in an erroneously low estimate of treatment effect size when the actions implemented in control communities improve the outcome of interest. Evaluators therefore generally try to avoid it.

In the *Natura* RCT, we only encountered one case of direct copying, leading us to conclude that this potential confounder could not have had any significant impact on outcomes. However the fact that it did exist confirms our view that a great deal more work remains to be done on understanding whether, and in which circumstances, such behavioural responses can exist in conservation RCTs. Our experience when monitoring water quality in these communities in 2015 (see chapter 3) suggested that control community members were well aware of the actions of *Natura* in neighbouring treatment communities and of the fact that their neighbours, but not they, had been offered conservation agreements. Totally independent units are unlikely to be a realistic goal in any such voluntary, community-based, human-action-mediated RCT.

5.6 Conclusions

RCTs may appear conceptually simple, especially when compared with other approaches to quantitative impact evaluation such as propensity score matching, regression-discontinuity, or instrumental variables methods (see Butsic *et al.* 2017). RCTs of interventions where narrow outcomes of interest can be defined in advance (such as a new drug regime, or fluoridation of water), where the intervention itself is very clearly defined, and (to some extent) when the intervention may be expected to have relatively similar outcomes wherever it is implemented (high *external validity*) are indeed more straightforward. Few doubt the value of RCTs for evaluating the impact of such interventions (e.g. Barton 2000); however, such situations are encountered principally in the medical field (Auerswald 2011). It is no coincidence that as the method started to be applied beyond such settings, and be used to evaluate the impact of development interventions for example, that controversy about RCTs' value has erupted (e.g. Ravallion 2009; Picciotto 2012). Increasingly, the sorts of interventions being promoted in conservation and environmental management seek to have positive impacts on both environmental and social outcomes (Ferraro & Kiss 2002; Kolinjivadi *et al.* 2015) and operate through aiming to change behaviour (though incentives and/or social pressure). While the conservation community has been calling out for more high-quality evidence about what works (e.g. Segan *et al.* 2011; Baylis *et al.* 2016; Curzon & Kontoleon 2016), including from RCTs, it should be no surprise that designing and implementing good-quality RCTs of socio-ecological interventions is challenging. Our experiences with one of the first RCTs to evaluate the impact of a socio-ecological conservation intervention demonstrate their associated difficulties in undertaking a high-quality evaluation, especially when the intervention seeks to affect multiple outcomes of interest. Whether these challenges are more substantial than those encountered when using alternative evaluation methods remains unclear, and this issue merits much more research effort. We suggest that when Randomised Control Trials are used to evaluate the impact of environmental management interventions, evaluators should prioritise designs which allow high quality evaluation for a few key outcomes, rather than expect any RCT to be able to simultaneously answer many questions.

6 Land use and conservation of catchments supplying water to communities: two case studies

6.1 Abstract

The previous chapters of this thesis explore the effects of the *Watershared* program on water quality and related health outcomes, as well as consequences of robustly evaluating such effects via a randomised control trial. In this chapter I present other evidence supporting my findings, which also leads into the conclusions presented in the Discussion. Here I show that data from health centres in the largest three communities of the *ANMI Río Grande – Valles Cruceños* supports the previous results relating to quality of water, selection of water sources, and their implications for community health outcomes (records show a three- to five-fold increase in per capita numbers of cases of disease attended in those communities with *E. coli* presence in their water supplies than in that community without). I also studied a 24-year-old forest conservation area – representing the implicit end goal of the level 1 *Watershared* areas – and examined its effect on the water quality of the community which is supplied from within it. I found that 20% of a catchment under conservation was not sufficient to result in water of good quality, at least in the wet season when I sampled, and local knowledge suggested that the contamination likely derived from agricultural activities in the upper catchment. This has implications for understanding of the lack of success of *Watershared* agreements in improving water quality at the community scale.

6.2 Introduction

During fieldwork conducted in 2016 I obtained more information which contributes to my understanding of the relationship between land management and water quality and therefore has implications for understanding the outcomes and impact of *Watershared*. This was not planned to be collected and nor was it obtained a systematic fashion or as part of a larger investigation. However I believe it to be instructive and illustrative when considered together with the results from the planned studies and the pre-existing literature on catchment conservation and water source protection.

Specifically, it shows how ecological context and type of intakes may interact with built infrastructure and community institutions to determine water quality supplied to communities and health outcomes within those communities. It also strongly supports the importance for water quality of considering catchments as an integrated whole (something also shown in chapter 4 by the significant predictive effects of cattle and agriculture in the areas upstream of monitored intakes on *E. coli*).

6.3 Health records, water quality, and water supply systems: further evidence of the influence of land use on water quality and health

I further examined the links between environmental context and land use, water quality in supply systems, and health outcomes in the three largest communities of the *ANMI Río Grande – Valles Cruceños* protected area (Moro Moro, Pucará and Postrervalle). I obtained data on the number of cases of diarrhoeal disease attended at these communities' health centres during the years of 2014 and 2015 (table 6.1); in the cases of Pucará and Postrervalle, these data were available disaggregated by month. (Such records may be more reliable than the self-reported survey responses, particularly due to the excessively long 12-month recall period associated with these, but they only record the cases that are attended at the clinic. These are likely to be the most severe cases, and the true level may be three or four times that recorded [N. Araujo, pers. comm.]) I combined this data with water quality measurement data from 2016 (that used in the analysis in chapter 4) and local knowledge of the context, land use and existing institutions.

6.3.1 *E. coli* presence measured is associated with both type of water body supplying the water system and number of cases attended at community health centres

Moro Moro is supplied by 3 stream intakes; Pucará by 1 stream intake which changed in August 2015 (see section 6.3.2 below), and Postrervalle by 1 spring which has been placed in a concrete tank, thus being an 'improved' water source according to the UN MDG definition. We measured water quality and potential sources of contamination in these water systems as part of the study described in chapter 4, using the protocol described therein. The only difference with the specified protocol there was that in these three communities we also measured water quality in storage tanks before distribution to the communities.

The results are instructive in supporting my conclusions about the importance of water source and quality on health. As shown in table 6.1, Moro Moro and Pucará, with water supplied from streams – all of which were contaminated with *E. coli* – have approximately four times the per capita level of diarrhoea of those in Postrervalle, which was *E. coli*–free (although note: the water there does still not comply with the standard in Bolivia, *Norma Boliviana 512* [Ministerio de Medio Ambiente y Agua 2010], due to the presence of other coliforms). Results are available in table 6.S1. Unlike in Moro Moro and Pucará, the water committee of Postrervalle also employs someone to clean and maintain the system, which was reported to be sterilized monthly with chlorine bleach. This supports the idea that institutions such as water committees (common in rural Latin America: [Foster 2005]) require substantial technical and financial resources if they are to be effective in managing water supplies (Chowns 2015).

Table 6.1. Cases of diarrhoeal disease attended at community health centres in 2014 and 2015 in the three largest communities of the *ANMI Río Grande* area. The increase in Pucará in 2015 is likely related to the switch of water supply (see section 6.3.2 below). Health data derived from the Bolivian *Servicio Nacional Integral de Salud*, data delivered by medical staff working at health centres in Moro Moro, Pucará and Postrervalle. Population data derived from Antúnez *et al.* (2009).

	Population (2008)	2014 cases	2014 cases per 1000 inhabitants	2015 cases	2015 cases per 1000 inhabitants
Moro Moro	743	206	277	236	318
Pucará	652	109	167	172	263
Postrervalle	1208	74	61	71	59

6.3.2 Changes in water supply source can be associated with spikes in infection levels: the case of Pucará

Pucará is a community of approximately 600 people in the western part of the ANMI Río Grande area. The community was previously supplied by a water intake some 2.5 kilometres from the village. However, in the years leading up to 2015 the quantity of water available began to perceptibly drop, to the extent that taps frequently ran dry in the dry season (roughly July to October). Community members blamed the government-run telecommunications company *ENTEL* for this, as it had constructed a mast on the summit of the hill above the intake. This construction required the opening of roads and other pathways which subsequently were taken advantage of by local people to then plant potatoes and clear land for pasture. This was perceived to have led to the drying-out of local water sources including the main community intake (P. Bottazzi, pers. comm).

Meanwhile the Bolivian national government had implemented a number of programs known as MIAGUA (see Ministerio de Medio Ambiente y Agua 2015), which provided funds for the construction of new infrastructure for rural water supplies, both for household supply and for irrigation. The municipal government of Pucará received some of this money (over 600,000 USD) and used it to construct a new, also gravity-fed, water supply system with an intake located in a stream several kilometres away (actually in the neighbouring community of El Cruce) which supplies adequate quantities of water throughout the entire year. This system was finished in 2015 and the water supply was switched to the new system in August of that year (figure 6.1).

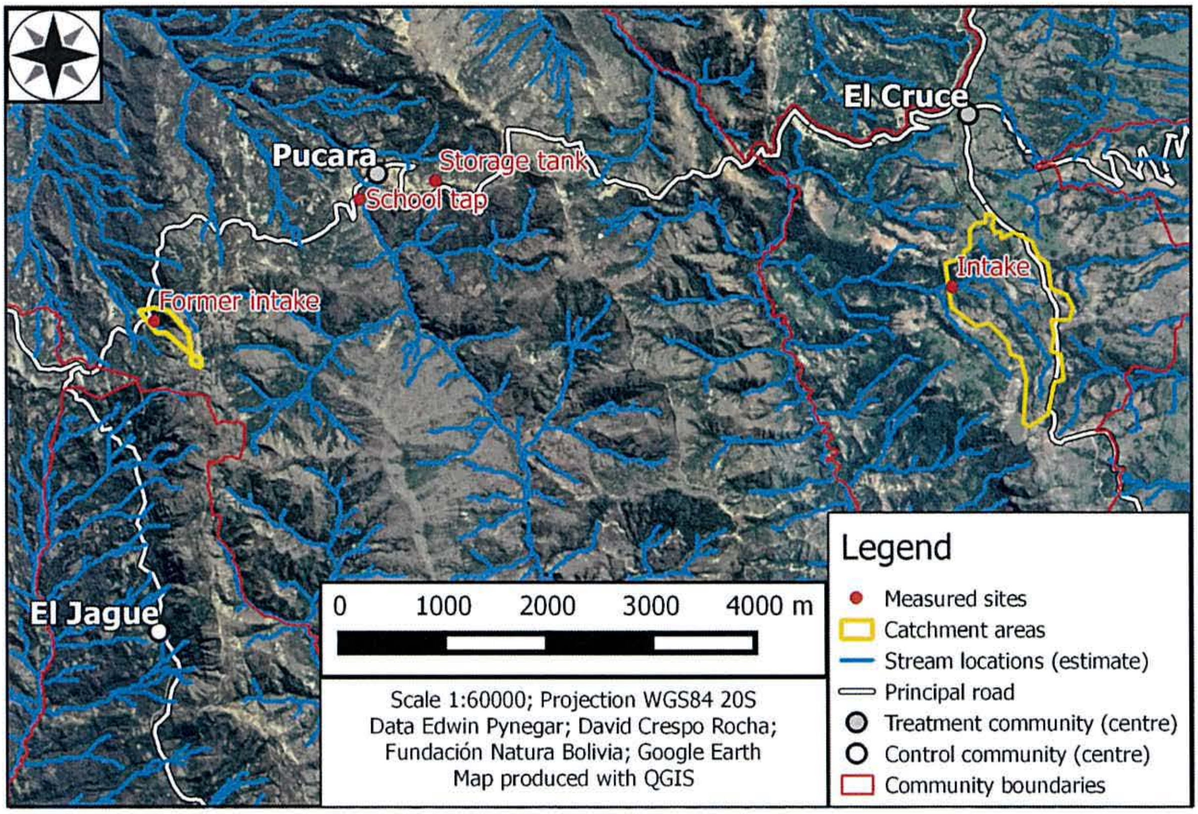


Figure 6.1. Locations of current and former water intakes supplying Pucará, catchments, and sites measured in 2016.

Immediately after this occurred, people began to complain of gastro-intestinal problems including bloody diarrhoea. This phenomenon was new and affected all parts of the community (R. Calzadilla [mayor of Pucará], pers. comm.). Cases were particularly frequent following rainfall events (Dr. M. Bustillos [doctor in charge of health centre], pers. comm). Presentation of cases of diarrhoeal disease at the community health centre spiked in the months following this changeover (figure 6.2; also table 6.1 above).

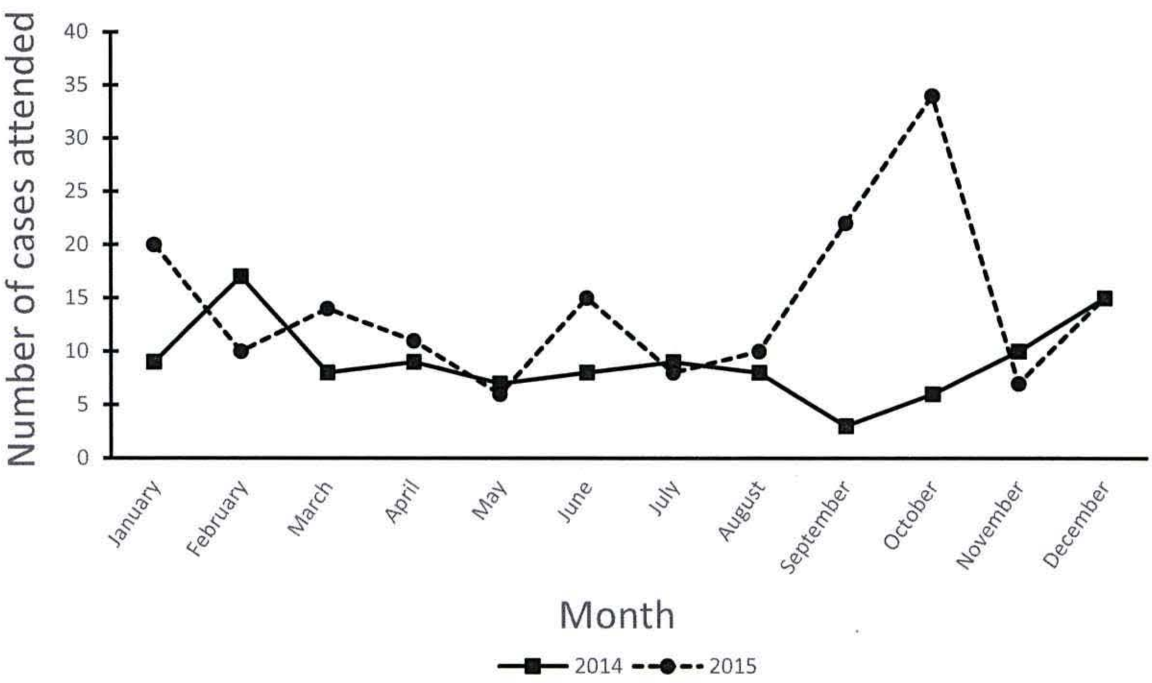


Figure 6.2. Cases of diarrhoea attended at the health centre in Pucará, by month, in 2014 and 2015. Supply was switched to the new intake in August of 2015.

Analysis of water quality from the new water source in 2016 clearly demonstrated that the new water source was contaminated with *E. coli* (table 6.S1). The new intake site was located in a catchment of 116 hectares (figure 6.1) which is private property and is used for rough grazing for cattle (*estancia*). Cattle faeces were observed above the intake and reported to be present throughout the catchment. Additionally, a road (that linking Vallegrande with Alto Seco and other smaller communities) is located in the headwaters of this catchment (figure 6.1). The previous intake was also found to be contaminated with *E. coli* (table 6.S1), although this may be because the protection previously existing there had become derelict since the changeover, and in some cases removed. I was also informed that the cattle density there was substantially lower than at the new intake. This therefore is a clear demonstration of the consequences of sourcing community water supplies from catchments with heavy livestock densities; there is also little forest land within this catchment and none of the watercourses upstream are under conservation. This example was used by the Nature Conservancy (Abell 2017).

6.4 How much of a catchment needs to be conserved to achieve improvements in water quality? Evidence from the *Reserva de Patrimonio Natural* (*REPANA*) in La Aguada

The community of La Aguada, in Trigal municipality, is located around 25km to the north of the *ANMI Río Grande*. While it is outside of the protected area, both its ecology and socio-economic context are very similar to many communities within the *ANMI* area itself. Approximately 100 people live in 31 households in the community, and all are supplied with water by an intake located in a stream a small distance from the centre of the community. Unusually, and most interestingly for the study of the effects of forest conservation in the region, in 1992 members of that community decided to conserve forest upstream of their water intake through cattle exclusion with the support of the NGO *Instituto de la Capacitación del Oriente* as an example of a *REPANA*, a previously implemented model for forest conservation and cattle exclusion in the Santa Cruz Valleys region (Asociación Zabalketa de Cooperación y Desarrollo 2008). The land cover found in the enclosed area (size: 152 hectares) is obviously different to that outside (figure 6.3). Previously published literature had stated that this action had improved water quality to the community, although this was based on anecdotal evidence only and contained no quantitative data (Robertson & Wunder 2005; Martin-Ortega, Ojea & Roux 2013).

On March 12th 2016 (during the rainy season in Bolivia) I went to La Aguada to take measurements of *E. coli* and coliform concentration in the water supply system (under the protocol presented in chapter 4). I measured at three sites: the intake, the principal storage tank, and a representative tap in the community (that of the house of Robert Rueda Villarroel). The water in all sites was notably turbid and a highly elevated concentration of coliforms and *E. coli* was encountered (table 6.S2).

I presented these results at the subsequent meeting of the community's water committee. The subsequent discussion established that there were only two realistic possible causes of such contamination: firstly from wildlife present within the 152 ha protected area, and/or secondly from agriculture and cattle grazing in the remainder of the catchment. It was considered by most members of the water committee that the second of these possibilities was substantially more likely to be the dominant factor.

Figure 6.3. Boundary between *REPANA* area (forested land, background) and non-*REPANA* area (non-forested degraded land in foreground).

Further spatial analysis showed that the 152-hectare protected area makes up only 20% of the 760-hectare catchment (figure 6.4). In the areas upstream of the protected area within the same catchment, satellite imagery shows fields and houses (local information shared suggests that maize and potatoes are grown, and extensive cattle grazing is conducted). This is therefore likely to be the source of the contamination. Further studies on this area will be extremely beneficial in establishing the source of the contamination (e.g. with DNA analysis of the *E. coli* colonies [e.g. Carson *et al.* 2001] or more straightforwardly by comparing relative concentrations of *E. coli* entering the REPANA area with that at the intake or at the point where the water leaves the area), the relative hydrological behaviour of the watershed containing the *REPANA* when compared with neighbouring watersheds, and a counterfactual analysis of the levels of contamination within that watershed. This therefore would represent a formative-type evaluation of the effectiveness of the actual *action* associated with the level 1 *Watershared* areas over the long term, as the La Aguada *REPANA* represents the (implicit) intended long-term outcome of the Level 1 areas of land in agreements.

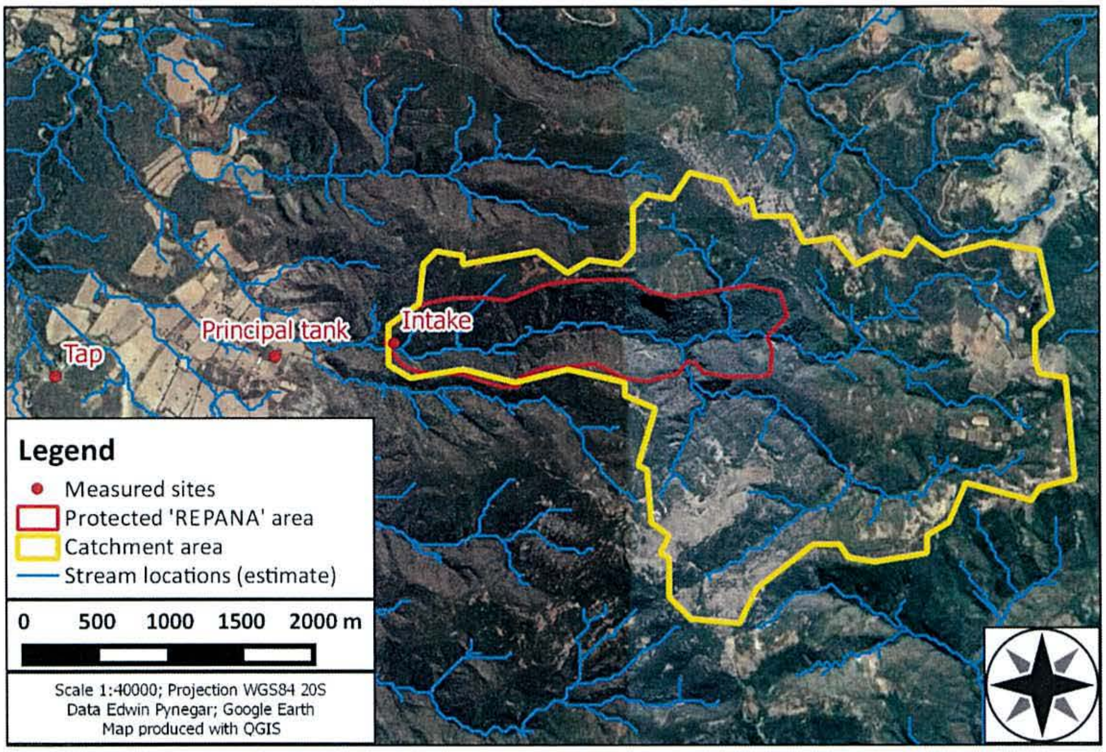


Figure 6.4. Map of La Aguada, the *REPANA* area, and the catchment within which it is located.

I obviously cannot say from this work whether the contamination levels encountered were significantly lower than they would have been without the forest conservation, whether the levels were lower than they were in an equivalent season before forest conservation had been implemented (i.e. in 1991), or whether the levels were lower than in neighbouring comparable watersheds (the counterfactual attribution problem in a nutshell again). However what is self-evidently clear is that this forest conservation, despite having been in place for 24 years, itself had not resulted in provision of water of good quality at all times.

6.5 Conclusions

I interpret the evidence presented here to support the conclusions of this thesis's principal chapters. The approximately four-fold difference in diarrhoeal disease levels between Moro Moro and Pucará, and Postrervalle, may well be attributable to the differential in *Escherichia coli* levels in those water supply systems and in the water bodies which supply them, supporting both the literature (Gruber, Ercumen & Colford 2014) and my own findings. The presence of *Escherichia coli* in the stream-fed systems supplying Moro Moro and Pucará, and its absence in the spring-fed system supplying Postrervalle, also supports the conclusions relating to the relative contamination of different types of water body from the 2015 and the 2016 data (chapters 3 and 4).

The evidence from Pucará of the effect of change of water systems is not as clear-cut as would ideally be the case, due to the fact that both the current intake and the previous intake contained high levels of *E. coli* contamination. However, local knowledge of the environmental context of the new water source and the (new) symptoms suffered by community members, as well as the data from the health centre, all do strongly support the cause of the problem having been the contamination from cattle in the catchment of interest. The 'spike' subsequent to the changeover and the novel symptoms may well also be related to exposure of the population to different kinds of pathogen originating there and so being without acquired immunity. Local knowledge of this kind may provide understanding of such issues to external researchers much more quickly than conventional scientific investigations (Santos de Lima, Krueger & García-Marquez 2017). We discuss the value of such knowledge and the implications of the fact that this represents non-RCT-derived learning in section 2 of the Discussion.

The La Aguada example is particularly important when considering the design of *Watershared* and similar payments for ecosystem services-type interventions. The findings presented in chapter 5 show that the 20% of a catchment conserved in La Aguada is a greater proportion than in almost any of the communities in which *Watershared* was implemented. Conservation actions may not be sufficiently effective if implemented at a scale which does not have the potential to remove the threat, as here with sources of contamination in the headwaters of the catchment (e.g. Dodds & Oakes 2008). We discuss further what this may imply for *Watershared* and the nature of action-outcome relationships in section 3 of the Discussion.

7 Discussion

RCTs have been used to evaluate public policy instruments and interventions for many decades (Glennerster & Takavarasha 2013; Council of Economic Advisers 2014) and there exists a large academic and non-academic literature about their use as a method of impact evaluation (Ravallion 2009; Banerjee & Duflo 2011; Picciotto 2012; Glennerster & Takavarasha 2013; Pritchett 2014). However RCTs have been little used in evaluating environmental management and conservation interventions, especially socio-ecological interventions whose effectiveness is mediated by the actions of human agents (see Butsic *et al.* 2017). There is great interest in the effectiveness of economic incentives for conservation such as Payments for Ecosystem Services schemes (e.g. Sánchez-Azofeifa *et al.* 2007; Alix-Garcia, Shapiro & Sims 2012; Samii *et al.* 2014; Börner *et al.* 2016, 2017). The existence of the RCT setup to allow robust evaluation of *Watershared* (a payments for ecosystem services-like intervention in the Bolivian Andes) therefore offered a valuable opportunity to learn lessons both about what RCTs can contribute to impact evaluation in conservation, and about the effectiveness of that specific intervention.

In this discussion I first lay out my key findings. I discuss the implications of this research for those interested in using RCTs in impact evaluation. I go on to discuss the implications for those interested in improving the design of *Watershared* or other similar PES-like interventions. Finally I describe how this research is being used to inform the design and implementation of *Watershared* in a new part of Bolivia (the Chaco region of Tarija department in the south-east of Bolivia) and an associated RCT to answer new questions relating to *Watershared*'s effectiveness in different contexts.

7.1 Key findings

7.1.1 RCTs have the potential to play a more significant role in conservation impact evaluation, but ensuring they generate good quality evidence of the effectiveness of an intervention requires substantial technical knowledge in their design and implementation

A number of factors strongly affect the relevance, feasibility and quality of any RCT setup. Implementing any high-quality RCT which is internally valid and able to meaningfully evaluate effectiveness of an intervention on a number of different kinds of outcomes of interest (which PES programs in general attempt to achieve) requires careful and detailed planning and high levels of conceptual and technical knowledge. We examined these factors in chapter 2, with reference to the PES-like *Watershared* program and the decisions that implementers made relating to its design. In chapter 5 we considered how these decisions affected the evaluation's validity. RCT designers must consider spatial and temporal aspects of multiple inter-related theories of change, as well as the interaction of the intervention's design with that of the evaluation. Developing effective and cost-effective monitoring protocols in a non-linear and spatially and temporally dynamic system is also likely to be challenging. Evaluators should also be aware of the potential for undesirable behavioural effects to affect outcomes, and much more work is required to understand the magnitude of these effects in RCTs.

7.1.2 *Watershared*, as implemented in the *ANMI Río Grande – Valles Cruceños* protected area, did not result in a measurable improvement in water quality supplied to communities (although consequences of incentivized actions, such as absence of faeces near water sources, did improve water quality locally)

In chapter 3 we find that the *Watershared* program did not achieve significant improvement in water quality (as measured by *Escherichia coli* contamination in community water supplies). However we found evidence that outcomes of actions *Watershared* seeks to incentivise (e.g. keeping cattle out of water sources) was indeed associated with improved water quality (presence of cattle faeces is, perhaps unsurprisingly, a predictor of *E. coli* contamination). Hence while the conservation actions *Watershared* seeks to incentivise may be important, the implementation of the program may result in their effects not being detectable at the community scale. This is partly because communities (including control communities) have implemented some conservation actions (such as fencing cattle out of key water sources) independently of the *Watershared* program.

7.1.3 Links between environmental context, water quality, and community health exist, but may not have been well captured by *Watershared*

In chapter 4, while we did not detect a direct link between *Watershared*-related actions and levels of diarrhoeal disease in communities, we did find disease levels to be predicted by *E. coli* levels in water supplies. *E. coli* levels themselves are predicted by a number of factors including types of water intake and factors which the *Watershared* programme seeks to influence (such as land use in catchments supplying those intakes). This shows that land use and management is important for watershed ecosystem service provision and consequently wellbeing, but that *Watershared* as it was implemented may not be effective in producing ecosystem services at the community scale. These conclusions are supported by the findings described in chapter 6.

7.2 What does the experience of evaluating *Watershared* tell us about the value of RCTs in conservation impact evaluation?

While RCTs have often been referred to as a 'gold standard' for evaluation (Barton 2000; Cartwright 2010), this may be in a more theoretical than a practical sense, as challenges relating to internal validity and establishing a clear and unambiguous cause-effect relationship are well known (Rubin 1974; Ferraro & Hanauer 2014). Our evaluation of *Watershared* highlighted a number of points surrounding the usefulness of RCTs in conservation, as well as their limitations. These go well beyond the examination of external validity in the second chapter. Measurement and monitoring actions associated with RCT evaluation bring their own distinct benefits for knowledge production and contextualisation, but both technical challenges and trade-offs between monitoring effort and completeness and breadth of data obtained will be unavoidable (we discuss this below). RCT results may be challenging to usefully interpret without further information, and the requirement for such an evaluation to be externally valid will inevitably restrict the kinds of intervention that are feasible to evaluate meaningfully. Additionally, their relevance for practitioners and policymakers should not be taken for granted; the most useful evaluations will be conducted with local understanding and contributions throughout the process; and researchers and practitioners should expect practice- and policy-relevant evidence to come in a diversity of forms.

7.2.1 High-quality measurement is necessary and brings multiple benefits, but will often be challenging

Undertaking an RCT requires evaluators to conduct systematic high-quality measurements of outcomes of interest in order to be able to calculate effect sizes of the intervention on those outcomes. This would represent, in and of itself, a major step forward for conservation science, quite separately from the potential benefits of conducting RCTs. High-quality monitoring of conservation programs has in the past not been widely conducted, especially the monitoring of outcomes (such as ecosystem services) rather than actions (Naeem *et al.* 2015). Such data also has its own value for understanding and wider knowledge creation (Naeem *et al.* 2015). However adequate measurement will normally not be straightforward, and high levels of technical knowledge, organizational and logistical capabilities, and financial inputs will be required to adequately conduct it (Jack, Kousky & Sims 2008; Muradian *et al.* 2010; Santos de Lima, Krueger & García-Marquez 2017). Monitoring and analysis is also inevitably imprecise, quite separately from stochasticity in outcomes (Santos de Lima, Krueger & García-Marquez 2017). Many of our most substantial challenges relating to analysis and its subsequent interpretation in this thesis related to data quality, and reducing such error wherever possible will be key to undertaking a good-quality RCT.

Ponette-González *et al.* (2015) point out that in most payments for watershed services programs, “the opportunity to measure water flows often arises simultaneously with the opportunity for intervention”. In the context of a well-designed RCT, however, baseline measurements of metrics of interest will be made well before intervention implementation, and will continue in control units throughout the trial. General data collection also will often allow other informative analyses to be conducted beyond an intention-to-treat or treatment-on-the-treated effect size calculation (as many of the quantitative analyses in this thesis in fact are).

Design and management of monitoring will often require expert input, and it may not be possible for technicians to conduct high quality monitoring without scientific oversight (Santos de Lima, Krueger & García-Marquez 2017). This was our experience with water quality monitoring as conducted by *Natura*. A number of rounds of water quality monitoring had been conducted before 2015, all of which had associated with them substantial problems with the protocols as designed and then their execution. Coliscan Easygel was selected for *E. coli* measurement as it can be incubated at room temperature and thus no incubator would be necessary (a great advantage in the context of mobile field sampling). However, the manufacturer means 22-25^{o}C by "room temperature" (Micrology Labs 2016), something which was not able to be complied with in the field. In the rounds of water quality monitoring previous to 2015, water samples were often kept (unrefrigerated) for several hours and in some cases even a number of days before being plated out. This made sense given the logistical challenges of the field context, but was likely detrimental to the quality of the data. In 2015, when I was involved, we purchased portable incubators to keep Easygel Petri dishes at 35-37^{o}C, and also ensured water samples were kept on ice and plated a maximum of six hours after collection.

Evaluators will also be faced with trade-offs between cost of measurement effort and data quality and/or richness. We encountered this in the selection of indicators of water quality in the evaluation of *Watershared*: *E. coli* and other coliforms represent only a partial measure of water quality, and single measurements represent contamination at only one point in time, leaving substantial observational uncertainties (e.g. McMillan, Krueger & Freer 2012). A more comprehensive measurement procedure involving multiple indicators, with more than one measurement per round of monitoring, would serve to reduce error and increase precision and therefore reduce the likelihood of a false negative. In the first year, *Natura* technicians collected data on macroinvertebrate communities at their study sites using the Biological Monitoring Working Party scores (Hawkes 1998) modified for use in Bolivia (Ministerio de Medio Ambiente y Agua 2011). This required identification of over 100,000 aquatic macroinvertebrates. *Natura* subsequently decided not to repeat this monitoring at endline simply because the associated effort of the (highly technically skilled) labour required would not bring sufficiently informative benefits. Such implicit or explicit cost-benefit analysis will inevitably be necessary, and will depend upon the goals of the intervention and/or RCT as well as the resources available. Evaluators may minimize measurement costs by developing clear theories of change and selecting indicators which are appropriate for detecting expected changes in the steps in those theories of change and outcomes of interest, and in the case of RCTs, basing planned measurements on power calculations for each of the variables in question.

7.2.2 RCTs may not be required to learn from an intervention, and RCT results may require a substantial amount of further interpretation to be useful

Much of what we learned about *Watershared*'s functioning and effectiveness was obtained outside of the RCT evaluation structure *per se*. *Watershared* did not significantly improve water quality in treatment community water supplies and did not directly result in improved health outcomes in treatment community households. Analyses of potential predictors of water quality using the 2015 and (narrower) 2016 water quality data, and the study of predictors of diarrhoeal disease (both microbial and survey-based) found a number of these to significantly predict both *E. coli* concentration and self-reported health outcomes. However, whether the community belonged to the treatment or control group was unimportant. These results allowed us to learn a substantial amount about the functioning and effectiveness of the *Watershared* areas, and the intervention as a whole, but did not depend upon the intervention being set up as a randomised trial (although issues of potential bias and reversed causality may make this evidence less convincing). We also interpret the results presented in the previous chapter to support our conclusions, while obviously also being unrelated to the RCT.

Negative results of RCTs are, as ever in scientific investigations, overdetermined: it is not necessarily clear if the negative result was due to the intervention being inherently ineffective in achieving its objectives, the measurement techniques being inadequate, the experiment underpowered, or all three (Kerry *et al.* 2012). Given the relative lack of formative evaluation of the version of *Watershared* implemented in the *ANMI Río Grande* and the problems with measurement described above, it is difficult to immediately eliminate any of these possibilities. Other pieces of evidence such as the small proportions of catchments of interest conserved and the absence of evidence strongly supporting the effectiveness of level 1 *Watershared* areas suggest that formative evaluation of intervention design may have been lacking – which may go a long way to explain why the intervention did not achieve its intended outcomes at the community scale. Without these other analyses, however, drawing any kind of meaningful conclusion about *Watershared* and its effectiveness from the RCT results alone would have been very challenging.

7.2.3 In some cases, conducting an intervention in an RCT evaluation may actively hinder successful implementation of that intervention

Conducting an RCT, especially one with a relatively long associated timescale as most conservation RCTs will have, can hinder successful implementation of some kinds of intervention. Specifically, RCTs are likely incompatible with any process of implementation involving adaptive management. Implementers of interventions will inevitably learn throughout the process of implementation, whether from conventional scientific studies, simple observations, or from local and indigenous knowledge, which aspects of an intervention do and do not function as intended (Krueger *et al.* 2012).

This understanding will suggest feasible, and in some cases newly obvious, improvements to the intervention (Williams & Brown 2016). This iterative implementation of PES and similar interventions is something that practitioners should embrace to improve intervention effectiveness (Santos de Lima, Krueger & García-Marquez 2017), but this will often be in conflict with the desire to conduct a meaningful and externally valid evaluation.

This problem may particularly affect PES programs focused on provision of hydrological services, as these have specifically been conceptualized as a means of ecological fine-tuning (Perrot-Maître 2006; Kolinjivadi, Adamowski & Kosoy 2014). However this being realistic implies visible and/or measurable changes in the outcomes of interest with short enough feedback loops for implementers and beneficiaries to be able to associate the action with the outcome (Biggs, Carpenter & Brock 2009; Lele 2009; Ponette-González *et al.* 2015). It also again implies a community-based, adaptive management-type approach to the problem, with enough institutional flexibility to be able to swiftly react to perceived or detected changes in outcomes (Perrot-Maître 2006). This therefore makes undertaking a meaningful RCT – as in, an RCT with a fixed treatment – extremely problematic. Whether implementers prioritize modifications to the intervention design based on early experiences or continue with the initial intervention design may depend on whether success of the intervention in achieving its intended outcomes, or its evaluation, is implementers' principal priority.

7.2.4 Evaluators should not automatically expect RCT results to be useful to policymakers

Finally, extra information on the quantitative impact of interventions, as evaluated by an RCT or through other methods, is not automatically beneficial and valuable (see Pritchett 2014). The 'value of (new) information' (Williams & Brown 2016) is something that stakeholders must consider and debate, which also implies an open and transparent discussion surrounding uncertainty (Hamel & Bryant 2017). More results from RCTs or other studies are not in and of themselves a solution to uncertainty and lack of policy-relevant knowledge; uncertainty does not disappear with more science (Brown 2010). Quantitative results will inevitably have uncertainty associated with them, which is rarely well understood by decision makers (Stirling 2010). Such results inevitably suggest further supposedly necessary research. Obtaining this kind of knowledge may thus present unintended consequences for evidence-informed policymaking by serving to emphasise epistemic gaps (Gross 2010). Also, impact evaluation using RCTs may be perceived as an external imposition rather than a participatory process (c.f. Baele 2013), and both the evaluation itself and the kinds of results it produces risk potentially being seen in the local context as illegitimate and/or irrelevant to decision-making (c.f. Cook *et al.* 2013).

Even if the RCT does not suffer from these problems and is accepted locally, the marginal utility of new information, especially estimates of the average treatment effects of interventions, may be modest (Miteva, Pattanayak & Ferraro 2012; Börner *et al.* 2016). This is not to say RCTs do not have a useful role, but that evaluators may wish to consider designing them such that they can also answer other, more fundamental, questions on effectiveness of interventions by varying contextual conditions under which an intervention is implemented. (This *is* feasible: see the section discussing the 'new *Watershared*' in the Chaco below). This is close to the 'realist' model of RCT implementation (Bonell *et al.* 2012; Jamal *et al.* 2015) and its suggested solution to the external validity problem by using the evaluation to test links in a hypothesised causal chain.

7.3 How can *Watershared*-like programs be implemented in a way more likely to achieve their intended goals?

Payments for Watershed Services-type programs have been implemented across the world, with a particular focus on Latin America (Martin-Ortega, Ojea & Roux 2013; Grima *et al.* 2016). Initial evaluations suggest that their effectiveness is mixed (Grima *et al.* 2016). Previous programs have tended not to be designed based upon scientific evidence, but there are calls for this to change (Lele 2009; Naeem *et al.* 2015; Ponette-González *et al.* 2015).

Implementation of *Watershared* in treatment communities did not achieve significant improvements in water quality in the systems supplying those communities, and did not improve waterborne health metrics at the community level. Nor do we have evidence that it increased household income or other metrics of wellbeing. However *Watershared* areas were widely implemented throughout treatment communities, a substantial proportion of those areas represented additional conservation (Bottazzi *et al.* 2018) and preliminary results suggest that deforestation levels were significantly lower in treatment communities (Wiik *et al.*, in review). Despite these mixed results at the community scale – also the RCT treatment and intervention implementation scale – aspects of the intervention were found to have significant effects on water-related outcomes of interest at a more local scale. Thus the intervention may modify ecosystem functions in the fashion intended, but at a scale not always able to provide the desired ecosystem services to communities. In this section we summarise what lessons can be learnt from this for design and implementation of payments for watershed services-type programs such as *Watershared*.

7.3.1 Designing PES programs with action-outcome relationships in mind is key for success in changing those outcomes

The relationship between action and outcome is a key concept on which the design of any payments for ecosystem services program should be based (Börner *et al.* 2017). While actions enhancing functions should increase provision of ecosystem services, the way in which they do so will differ enormously depending upon the action and the type of desired outcome (e.g. Ponette-González *et al.* 2015). Hence, effective programs should be designed to achieve desired changes in outcomes of interest according to the best understanding of the action-outcome relationship which exists. The fact that many ecological processes incorporate non-linear and threshold-type responses will add to this complexity.

We encountered these issues extremely clearly in the case of *Watershared.* As discussed in chapters 4, 5, and 6, detectable and significant improvements in water quality likely rely on a large proportion of land enrolled in level 1 *Watershared* agreements in the catchments supplying the water intakes (also see Jawson *et al.* 1982; Tiedemann *et al.* 1987; Dodds & Oakes 2008). Non-linear threshold effects likely exist in these catchments, and so a certain proportion of land may have to be conserved before evaluators can expect effects on water quality to become detectable (Allan 2004; W. Buytaert, pers. comm.). Maintaining ecological health of Andean streams may require more than 70% native vegetation cover (Iñiguez-Armijos *et al.* 2014), and land cover of catchments as a whole is likely more important in determining statuses of ecological indicators than more localised-scale predictors (Death & Collier 2010). It is not clear whether small forest parcels can 'reset' water quality in otherwise highly modified catchments (e.g. Harding, Claassen & Evers 2006; de F. Fernandes, de Souza & Tanaka 2014). Additionally, the La Aguada case is not the only example of conservation downstream being unable to counteract the effects of damaging land use practices in catchment headwaters (Dodds & Oakes 2008).

If this is true, it immediately presents a mismatch in *scale* between actions and outcomes, as catchments are often covered by multiple separate properties. *Watershared* agreements are implemented between the NGO and the landowner, but water quality at the intake site is a cumulative consequence of land use in the whole catchment (as well as factors well beyond the control of any of the involved stakeholders such as climate and geology). Hence incentivised actions are at the property scale, while outcomes occur at the catchment scale (Barnaud & Antona 2014). Effects of the intervention are thus immediately and inherently difficult to disaggregate from other actions (Santos de Lima, Krueger & García-Marquez 2017). (The only exceptions to this problem are the case of *self-environmental service provision* [Bottazzi *et al.* 2018] in which individuals use the incentives provided to conserve their *own* water supplies through conserving their own land.) Such spatial mismatches will occur in any case in which multiple actors are required to cooperate towards an external objective, thus adding a collective-action problem (Engel 2016; also Santos de Lima, Krueger & García-Marquez 2017).

PES designers have a number of available tools to achieve these minimum necessary actions and solve the scale-mismatch issue, including spatial targeting and differentiated payments, agglomeration bonuses, or minimum proportions of eligible land required before any payments are disbursed, as well as simply raising payment levels. Spatial targeting and payment differentiation have been found to be significantly positively associated with program effectiveness (Ezzine-de-Blas *et al.* 2016). A *Watershared* intervention more likely to achieve its goal of improving water quality would integrate these features into its design, accounting for the action-outcome relationship. Differentiated payments would likely not have been feasible, not only due to the extremely high associated informational and transaction costs, but also because such attempts to maximise economic efficiency go against the reciprocal and incentive-based nature of the scheme. Spatial targeting, by contrast, would likely have been beneficial, involving targeting towards key catchments, or to the riparian buffers within those catchments (Allan 2004), combined with an increase in the incentives offered for conservation of those areas. An agglomeration bonus, in which increased payments are offered once a certain area or proportion of land and/or enrolled land in a particular configuration (Drechsler *et al.* 2016; Engel 2016), or a minimum proportion of each eligible catchment required to be enrolled, may have been appropriate to overcome the implicit collective-action problem. This would reflect the widely known example of the protection of the Catskill watersheds in upstate New York, in which generally highly individualistic farmers had to agree to implementation of best management practices on at least 85% of farms before any payments could be disbursed (e.g. Ashendorff *et al.* 1997).

Certain social contexts may also favour implementation of effective payments for watershed services-type programs such as *Watershared*. The very nature of water quality maintenance and hydrological regulation as ecosystem services – non-rival but excludable – makes them very suitable for communal resource management via socially managed PWS as a toll or club good, but highly unsuitable for individualist market-based PES-type interventions (Kolinjivadi, Adamowski & Kosoy 2014). PWS may only work as a policy tool when it is applied taking into account the particular social and spatial configurations in any social-ecological system (Kolinjivadi, Adamowski & Kosoy 2014). Land tenure in catchments which lack spatial scale mismatch, in that decision-making takes place at the same scale as the required action, may also facilitate implementation of the intervention at the scale required (see Barnaud & Antona 2014). This may suggest a reciprocal agreement with a single private landowner, or in the case of communal land tenure – almost unknown in the *ANMI Río Grande* but widespread in other parts of Bolivia – an agreement either within their own community or a reciprocal agreement with another community. The latter of these cases would represent a further example of self-environmental service provision, but at a community scale.

Finally, *Watershared* may have achieved its intended goal of significantly slowing deforestation (Wiik *et al.*, in review), confirming the finding that a significant proportion of the *Watershared* areas represent additional conservation (Bottazzi *et al.* 2018). Jayachandran *et al.* (2017) obtained a similar result. The difference between the outcomes relating to deforestation and to water quality is instructive. The simpler theory of change associated with reducing deforestation, the closer links between action and outcome, the lower importance of spatial coordination and non-linear ecological processes, and the potential for the action to be able to take effect at the (individual) level at which the intervention was implemented all may have made the intervention more likely to achieve its goals. Hence simple and direct theories of change may be likely to be associated with successful PES programs; this question merits further examination in future studies.

7.3.2 Transaction costs of establishing action-outcome relationships are likely to be high, but may not be necessary to implement an intervention

If establishing action-outcome relationships is genuinely a requirement for a successful PES program, then as a transaction cost it is highly significant (Muradian *et al.* 2010). Some authors have argued that clear and widely understood linkages between the action incentivized by a PES program and changes in the desired outcome of interest of that program are important for project success, especially over the long term, as otherwise participants will inevitably lose confidence as they see that their actions are not bringing about the intended changes in ecosystem service provision (Ponette-González *et al.* 2015). This assumes that participants will be able to detect any change over time and that that change can be attributed to the intervention (the impact evaluation problem in a nutshell again).

Paradoxically, empirical findings strongly contest this claim: it has long been known that the relationship between land cover/use and hydrology is extremely complex and is also very specific to individual contexts, and is rarely understood even broadly in a context before PES is implemented (Kosoy *et al.* 2007; Lele 2009). Indeed, most operate on the assumption that more forest cover results in more and/or better quality water, which is doubtful (while not denying that forest conservation does provide watershed services [Ponette-Gonzalez *et al.* 2014; Herrera *et al.* 2017]). Furthermore, uncertainty, stochastic processes such as weather, inevitable monitoring, data processing and analysis errors and resources required to conduct studies mean that implementers will never have as much understanding of such processes as they might ideally want.

Watershared agreements have been signed in the area for a number of years without any such studies being conducted, and *Watershared*'s implementers have little understanding of its associated action-outcome relationships, particularly at the catchment scale. Cattle exclusion from well over 20% of catchments may be required to achieve an adequate improvement in water quality (as discussed in chapter 6) and our baseline data was not of sufficient quality to establish the effect of exclusion over time on water quality at spring sites. Part of *Natura*'s success in achieving land enrolment is therefore predicated on pre-existing understandings and local knowledge. More forests are understood to provide more water (this may be true in the dry season when water scarcity is an issue; also see Murtinho *et al.* [2013]) and fencing off water bodies improves their water quality (again partially supported). Cattle exclusion has been practised for decades by many local communities without any incentives (e.g. Asociación Zabalketa de Cooperación y Desarrollo 2008), suggesting that it anyway warrants substantial investment of time and resources. Consequently, it would seem clear that a lack of robust evidence of changes in outcomes is no bar to implementation and spread of PES programs (also see Porras *et al.* 2013), especially when implementers can 'piggy-back' on pre-existing understandings.

Obtaining the desired outcomes from these actions though is a very separate issue. The greater the uncertainty about the action-outcome link, and the more difficult this is to monitor, the more difficult it may be for implementers to secure the efficiency gains available from payment by outcomes (Gibbons *et al.* 2011).

7.3.3 Program designs predicated upon action-outcome relationships may not be feasible in some contexts

When an intervention's theory of change contains spatial agglomeration, spatial differentiation or threshold effects, designers must consider the potential for the existence of a contradiction between ecological necessity (the requirement to achieve a certain level of implementation of the intervention before changing the outcome of interest) and social reality (that feasible to implement in that social context, including accounting for pre-existing norms of natural resource management). *Watershared* in the ANMI Río Grande encountered this contradiction in that the land tenure system is largely individual, the social structure is relatively individualist, and farmers may need to continue using their lands located in the catchments of interest; these factors may all prevent sufficient areas of land being enrolled in those catchments. Similar examples have been encountered in Colombia (Santos de Lima, Krueger & García-Marquez 2017): a payments for watershed services program intended to be established in the Cauca Valley used a modelling tool to identify priority areas for provision of hydrological services. However it was impossible to enrol land in agreements or contracts or even conduct field measurements in those areas because of extremely high levels of distrust between involved actors in those locations, a consequence of the only recently concluded civil war which impacted that part of Colombia particularly heavily. Such implicit contradictions associated with PES programs deserve substantially more attention than they have up until now received.

7.4 Implementation of *Watershared* as an RCT in the Bolivian Chaco

Natura is undertaking another RCT of *Watershared* in four municipalities making up the Chaco region of Tarija department in the south of Bolivia. I contributed to the design and implementation of this new RCT based on findings in this PhD thesis, redesigning the *Watershared* intervention in ways I believe likely to increase the chance of it achieving its intended goals, and suggesting modifications to the RCT setup to increase its internal validity and the usefulness and breadth of its results. In particular, I tried to solve the problem of any selected randomisation unit not being suitable for all intended outcomes of interest, and I explain our decisions as a reference for other evaluators.

7.4.1 The design of the intervention will be modified to better account for expected action-outcome relationships

In the Chaco, *Watershared* agreements will be implemented by catchment (or hydrological recharge area) and implementers will have to agree to conserve a minimum area of that catchment before any agreement can be signed and any incentives can be distributed. The proportion is yet to be determined, but will probably be >70% with a particular focus on headwaters and the reach just above the water intake. Such intervention designs are feasible because most of the land in this area is under a communal land tenure system, such that communities can agree amongst themselves to conserve their catchments, in return for incentives delivered on a communal basis. In cases where that is not the case, a negotiation will be mediated by *Natura* in which stakeholders will have to agree to conserve the minimum area between them, and no incentives will be delivered to any of them if they cannot agree to do so. This therefore demonstrates again the need for consideration of land tenure systems and the potential clash between ecological necessity and social reality.

The value of the incentives offered will be determined by a number of factors, not least the budget available and the areas of target catchments. However we also intend to test 'willingness to accept' to ensure that intervention uptake (and thus in terms of the RCT, units treated) will be likely to be at a reasonable level.

7.4.2 The design of the RCT will be modified to improve internal validity and to answer broader questions

The RCT will evaluate both the effectiveness of *Watershared* on a number of outcomes of interest, and will answer more fundamental questions about implementation of such programs through varying contextual conditions of incentive delivery. Researchers will investigate the effect of introducing the program in different ways on wellbeing, participation, and other outcomes of interest. Therefore, randomisation units are to be divided into three groups: two treatment groups and a control. The two treatment groups differ in that in one, the program will be introduced as a 'classic' Payments for

Ecosystem Services program (or as closely as possible to a Payments for Ecosystem Services program in the Bolivian political context); in the other it will be introduced as a reciprocal water agreement, with heavy emphasis on reciprocity and altruism rather than financial inducement. Results will be used to modify the implementation of *Watershared* in future. This shows how careful consideration of RCT design can allow evaluators to simultaneously measure intervention effect sizes, and also answer other types of questions (Bonell *et al.* 2012).

Additionally, the initial diagnostic survey posed questions which relate closely to each step in the intended theory of change, such as whether cattle were present in the catchment upstream of the intake and whether the intake was protected (features intended to be changed by the intervention). Follow-up and endline surveys will ask the same questions. This will therefore not only let us evaluate effectiveness of the intervention but also how the theory of change feeds through from one stage to another. Careful design of the evaluation with focused questions will therefore provide more informative outputs than a 'black box' RCT, in the way that we discuss in section 7.2.4 above.

7.4.2.1 Solutions to the randomisation-unit problem, and the interaction between the intervention and the evaluation

In the planning and design phase, we again encountered the problem of scale mismatch and incongruent spatial relationships between intended randomisation unit and natural units of change. Randomisation units were intended to be communities. However the diagnostic survey and baseline water quality monitoring showed that many communities shared water intakes, or there were multiple intakes supplying one community (a number of different layouts were possible). Conservation of such shared water intake catchments could clearly be implemented by the community in which the catchment fell, but would not be able to straightforwardly answer the questions relating to the effects of differential contextualisation (PES or RWA) of the program.

We therefore classified communities by their water system layout, and then stratified communities by this classification (amongst others) before randomisation. In cases where one or more catchments supply one single community, each community would be treated as its own randomisation unit and randomisation of incentive allocation would be conducted as planned. If the community had more than one catchment, then communities will have to agree to conserve each of their catchments as part of the *Watershared* agreement. In cases where multiple communities are supplied by one or more intakes, these communities will be aggregated and treated as a kind of 'mega-community' and as one randomisation unit. While incentives will still be delivered in return for signing *Watershared* agreements, these will not be included in the part of the evaluation relating to program contextualisation. The stratification is intended to allow removal of these communities from those analyses without introducing biases, and to prevent there being systematic differences between groups if communities behave differently when the intervention involves implementation of agreements in one or multiple catchments, thus accounting for within-group heterogeneity (see Glennerster & Takavarasha 2013). This again shows the importance of planning, understanding of the local context, and consideration of spatial relationships in order to ensure that the RCT setup is logical, meaningful, and internally valid.

7.4.2.2 Monitoring and measurement

Site selection for monitoring of water quality and related outcomes (community water intakes) will be determined by those catchments where the *Watershared* agreements are offered (as discussed above this will be determined by the type of community). Changes to intake locations will unavoidably result in attrition, although we expect this to be relatively low especially if *Watershared* agreements are implemented upstream of the intakes in question.

We plan to conduct monitoring of water quality annually during the wet season. The method used will continue to be based on Coliscan Easygel, as many other products on the market do not allow rapid, mobile monitoring of large numbers of sites in the field (see Bain *et al.* 2012). The protocol itself is based on that described in chapter 4 of this thesis; I trained technicians in person in the method's use to ensure that it will be applied consistently throughout the process. Hence both design and implementation do require experience if not expert knowledge.

7.5 Conclusions

In this thesis we examined the potential for evaluation of conservation interventions using Randomised Control Trials. We contend that RCTs are potentially powerful tools for impact evaluation in conservation, but while there may be a great deal of interest in their wider use, there will be many situations in which they will not be feasible or appropriate. We also conclude that there are a large number of pitfalls that implementers of RCTs must be aware of in order to conduct a high-quality trial. These requirements may also restrict or prevent certain kinds of conservation intervention from being evaluated meaningfully via an RCT, and implementers must anyway closely consider the way in which the intervention may be expected to cause measurable changes in the outcomes to be evaluated to ensure that the trial as designed has a reasonable probability of detecting an effect of the intervention, should it exist. The conceptual and technical knowledge required to execute an RCT well is therefore substantial, and implementers should not expect to be able to conduct a high-quality example cheaply or without a high level of pre-implementation planning and information gathering. The transdisciplinary nature of RCTs in conservation should be embraced rather than seen as a challenge, such that links between researchers in different disciplines as well as between researchers and practitioners be fostered if RCTs are to be more widely used.

We have also shown that payments for watershed services-like conservation interventions such as *Watershared* do have a great deal of potential for improving water quality and health outcomes, especially in rural developing-world contexts where water treatment infrastructure is far from ubiquitous (McDonald *et al.* 2014; Herrera *et al.* 2017). However, implementation of an effective program requires close consideration of action-outcome relationships (Biggs, Carpenter & Brock 2009; Kolinjivadi, Adamowski & Kosoy 2014; Ponette-González *et al.* 2015) to ensure that the intervention has a genuine possibility of achieving the intended changes. Many factors, including land tenure regimes, will affect feasibility of implementation of such interventions (Santos de Lima, Krueger & García-Marquez 2017), and close consideration must be paid to ecological processes, social dynamics and pre-existing institutions, and the interactions between the two, if interventions are to achieve their intended outcomes and goals.

Reference List

Abell, S. (2017) *Beyond the Source: The Environmental, Economic and Community Benefits of Source Water Protection*. The Nature Conservancy, Arlington, VA, USA.

Adams, W.M. & Sandbrook, C. (2013) Conservation, evidence and policy. *Oryx*, **47**, 329–335.

Ahmed, W., Hodgers, L., Sidhu, J.P.S. & Toze, S. (2012) Fecal indicators and zoonotic pathogens in household drinking water taps fed from rainwater tanks in Southeast Queensland, Australia. *Applied and Environmental Microbiology*, **78**, 219–226.

Alberini, A., Eskeland, G.S., Krupnick, A. & McGranahan, G. (1996) Determinants of diarrheal disease in Jakarta. *Water Resources Research*, **32**, 2259–2269.

Alexander, K.A., Carzolio, M., Goodin, D. & Vance, E. (2013) Climate change is likely to worsen the public health threat of diarrheal disease in Botswana. *International Journal of Environmental Research and Public Health*, **10**, 1202–1230.

Alix-Garcia, J.M., Shapiro, E.N. & Sims, K.R.E. (2012) Forest Conservation and Slippage: Evidence from Mexico's National Payments for Ecosystem Services Program. *Land Economics*, **88**, 613–638.

Allan, J.D. (2004) Landscapes and riverscapes: the influence of land use on stream ecosystems. *Annual Review of Ecology, Evolution, and Systematics*, **35**, 257–284.

Alpízar, F., Nordén, A., Pfaff, A. & Robalino, J.A. (2017) Spillovers from targeting of incentives: Exploring responses to being excluded. *Journal of Economic Psychology*, **59**, 87–98.

Andam, K.S., Ferraro, P.J., Pfaff, A., Sanchez-Azofeifa, G.A. & Robalino, J.A. (2008) Measuring the effectiveness of protected area networks in reducing deforestation. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 16089–16094.

Anderson, K.L., Whitlock, J.E. & Harwood, V.J. (2005) Persistence and differential survival of fecal indicator bacteria in subtropical waters and sediments. *Applied and Environmental Microbiology*, **71**, 3041–3048.

Angelsen, A. (2017) REDD+ as Result-based Aid: General Lessons and Bilateral Agreements of Norway. *Review of Development Economics*, **21**, 237–264.

Angrist, J.D. & Pischke, J.-S. (2010) The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. *Journal of Economic Perspectives*, **24**, 3–30.

Antúnez, N.A., Zeballos, D.M., Fuente, J., Vargas, M.T., Farell, M.E., Guerra, A., Vargas, I., Azurduy, H., Delgadillo, G., Siles, P. & Céspedes, L. (2009) *Estudio Socioeconómico Y Antropológico Del ANMI Río Grande - Valles Cruceños*. Centro para el Desarrollo Institucional, Humano y Ecológico, Santa Cruz de la Sierra, Bolivia.

Arnold, B.F., Galiani, S., Ram, P.K., Hubbard, A.E., Briceño, B., Gertler, P.J. & Colford, J.M. (2013) Optimal recall period for caregiver-reported illness in risk factor and intervention studies: A multicountry study. *American Journal of Epidemiology*, **177**, 361–370.

Arriagada, R.A., Sills, E.O., Pattanayak, S.K. & Ferraro, P.J. (2009) Combining Qualitative and Quantitative Methods to Evaluate Participation in Costa Rica's Program of Payments for Environmental Services. *Journal of Sustainable Forestry*, **28**, 343–367.

Arsel, M. & Büscher, B. (2012) Nature™ Inc: Changes and Continuities in Neoliberal Conservation and Market-based Environmental Policy. *Development and Change*, **43**, 53–78.

Ashbolt, N.J. (2004) Microbial contamination of drinking water and disease outcomes in developing regions. *Toxicology*, **198**, 229–238.

Ashbolt, N., Grabow, W. & Snozzi, M. (2001) Indicators of microbial water quality. In *Water Quality: Guidelines, Standards and Health* (eds. L. Fewtrell & J. Bartram), pp. 289–316.

Ashendorff, A., Principe, M.A., Seeley, A., LaDuca, J., Beckhardt, L., Faber, W. & Mantus, J. (1997) Watershed protection for New York City's supply. *Journal of the American Water Resources Association*, **89**, 75–88.

Asociación Zabalketa de Cooperación y Desarrollo. (2008) *Establecimiento de Reservas Comunales Para La Conservación Y Gestión de Fuentes de Agua*. Imprenta Arenas, Getxo, Basque Country.

Asquith, N.M. (2016) *Watershared: Adaptation, mitigation, watershed protection and economic development in Latin America.* Climate & Development Knowledge Network, London.

Asquith, N.M., Vargas, M.T. & Wunder, S. (2008) Selling two environmental services: In-kind payments for bird habitat and watershed protection in Los Negros, Bolivia. *Ecological Economics*, **65**, 675–684.

Asquith, N.M. & Wunder, S. (2008) *Payments for Watershed Services: The Bellagio Conversations*. Fundación Natura Bolivia, Santa Cruz de la Sierra, Bolivia.

Auerswald, P. (2011) Why Randomized Controlled Trials work in public health...and not much else, http://thecomingprosperity.blogspot.co.uk/2011/09/why-randomized-controlled-trials-work.html. Accessed 19th December 2017.

Babad, E.Y., Inbar, J. & Rosenthal, R. (1982) Pygmalion, Galatea, and the Golem: Investigations of biased and unbiased teachers. *Journal of Educational Psychology*, **74**, 459–474.

Baele, S.J. (2013) The ethics of New Development Economics: is the Experimental Approach to Development Economics morally wrong? *Journal of Philosophical Economics*, **7**, 2–42.

Bain, R., Bartram, J., Elliott, M., Matthews, R., McMahan, L., Tung, R., Chuang, P. & Gundry, S. (2012) A summary catalogue of microbial drinking water tests for low and medium resource settings. *International Journal of Environmental Research and Public Health*, **9**, 1609–1625.

Bain, R., Cronk, R., Hossain, R., Bonjour, S., Onda, K., Wright, J., Yang, H., Slaymaker, T., Hunter, P., Prüss-Ustün, A. & Bartram, J. (2014a) Global assessment of exposure to faecal contamination through drinking water based on a systematic review. *Tropical Medicine & International Health*, **19**, 917–927.

Bain, R., Cronk, R., Wright, J., Yang, H., Slaymaker, T. & Bartram, J. (2014b) Fecal Contamination of Drinking-Water in Low- and Middle-Income Countries: A Systematic Review and Meta-Analysis ed P.R. Hunter. *PLoS Medicine*, **11**, e1001644.

Balmford, A., Bruner, A., Cooper, P., Costanza, R., Farber, S., Green, R.E., Jenkins, M., Jefferiss, P., Jessamy, V., Madden, J., Munro, K., Myers, N., Naeem, S., Paavola, J., Rayment, M., Rosendo, S., Roughgarden, J., Trumper, K. & Turner, R.K. (2002) Economic reasons for conserving wild nature. *Science*, **297**, 950–3.

Banerjee, A., Chassang, S. & Snowberg, E. (2016) *Decision Theoretic Approaches to Experiment Design and External Validity*. NBER Working Paper No. 22167, Cambridge, MA.

Banerjee, A. V. & Duflo, E. (2011) *Poor Economics*. PublicAffairs, New York, NY.

Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Pariente, W., Shapiro, J., Thuysbaert, B. & Udry, C. (2015) A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science*, **348**, 1260799.

Barnaud, C. & Antona, M. (2014) Deconstructing ecosystem services: Uncertainties and controversies around a socially constructed concept. *Geoforum*, **56**, 113–123.

Barton, S. (2000) Which clinical studies provide the best evidence? *BMJ*, **321**, 255–256.

Barton, K. (2017). MuMIn: Multi-Model Inference. R package version 1.40.0. https://CRAN.R-project.org/package=MuMIn

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, **67**(1), 1-48.

Bauch, S.C., Birkenbach, A.M., Pattanayak, S.K. & Sills, E.O. (2015) Public health impacts of ecosystem change in the Brazilian Amazon. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 7414–7419.

Bausell, R.B. (2015) *The Design and Conduct of Meaningful Experiments Involving Human Participants: 25 Scientific Principles*. Oxford University Press, New York, NY.

Baylis, K., Honey-Rosés, J., Börner, J., Corbera, E., Ezzine-de-Blas, D., Ferraro, P.J., Lapeyre, R., Persson, U.M., Pfaff, A. & Wunder, S. (2016) Mainstreaming Impact Evaluation in Nature Conservation. *Conservation Letters*, **9**, 58–64.

Baylis, K., Peplow, S., Rausser, G. & Simon, L. (2008) Agri-environmental policies in the EU and United States: A comparison. *Ecological Economics*, **65**, 753–764.

Beck, H.E., Bruijnzeel, L.A., van Dijk, A.I.J.M., McVicar, T.R., Scatena, F.N. & Schellekens, J. (2013) The impact of forest regeneration on streamflow in 12 mesoscale humid tropical catchments. *Hydrology and Earth System Sciences*, **17**, 2613–2635.

Bétrisey, F. & Mager, C. (2015) Les paiements pour services environnementaux de la Fondation Natura Bolivia entre logiques réciprocitaires, redistributives et marchandes. *Revue Française de Socio-Économie*, **15**, 39–58.

Biggs, R., Carpenter, S.R. & Brock, W.A. (2009) Spurious Certainty: How Ignoring Measurement Error and Environmental Heterogeneity May Contribute to Environmental Controversies. *BioScience*, **59**, 65–76.

Blanchard, L., Vira, B. & Briefer, L. (2015) The lost narrative: Ecosystem service narratives and the missing Wasatch watershed conservation story. *Ecosystem Services*, **16**, 105–111.

Bloom, H.S. (2008) The Core Analytics of Randomized Experiments for Social Research. In *The SAGE Handbook of Social Research Methods* (eds P. Alasuutari, L. Bickman, & J. Brannen), pp. 115–133. SAGE Publications Ltd, London.

Bonell, C., Fletcher, A., Morton, M., Lorenc, T. & Moore, L. (2012) Realist randomised controlled trials: A new approach to evaluating complex public health interventions. *Social Science & Medicine*, **75**, 2299–2306.

Börner, J., Baylis, K., Corbera, E., Ezzine-de-Blas, D., Ferraro, P.J., Honey-Rosés, J., Lapeyre, R., Persson, U.M. & Wunder, S. (2016) Emerging Evidence on the Effectiveness of Tropical Forest Conservation. *PLOS ONE*, **11**, e0159152.

Börner, J., Baylis, K., Corbera, E., Ezzine-de-Blas, D., Honey-Rosés, J., Persson, U.M. & Wunder, S. (2017) The Effectiveness of Payments for Environmental Services. *World Development*, **96**, 359–374.

Bottazzi, P., Wiik, E., Crespo, D. & Jones, J.P.G. (2018). Payment for Environmental "Self-Service": Exploring the Links Between Farmers' Motivation and Additionality in a Conservation Incentive Programme in the Bolivian Andes. *Ecological Economics*, In press.

Brody, H. (2012) A critique of clinical equipoise. In *The Ethical Challenges of Human Research* (ed. F.G. Miller), pp. 199–216. Oxford University Press, New York, NY.

Brown, J.D. (2010) Prospects for the open treatment of uncertainty in environmental research. *Progress in Physical Geography*, **34**, 75–100.

Bruijnzeel, L.A. (2004) Hydrological functions of tropical forests: Not seeing the soil for the trees? *Agriculture, Ecosystems and Environment*, **104**, 185–228.

Bulte, E., Beekman, G., Di Falco, S., Hella, J. & Lei, P. (2014) Behavioral Responses and the Impact of New Agricultural Technologies: Evidence from a Double-blind Field Experiment in Tanzania. *American Journal of Agricultural Economics*, **96**, 813–830.

Butler, R. (2012) Bees, Trees, and Mountain Streams: Bolivia Redefines Watershed Payments, http://www.ecosystemmarketplace.com/articles/bees-trees-and-mountain-streams-br-bolivia-redefines-watershed-payments/. Accessed 19th December 2017.

Butsic, V., Lewis, D.J., Radeloff, V.C., Baumann, M. & Kuemmerle, T. (2017) Quasi-experimental methods enable stronger inferences from observational data in ecology. *Basic and Applied Ecology*, **19**, 1–10.

Calvet-Mir, L., Corbera, E., Martin, A., Fisher, J. & Gross-Camp, N. (2015) Payments for ecosystem services in the tropics: A closer look at effectiveness and equity. *Current Opinion in Environmental Sustainability*, **14**, 150–162.

Carson, C.A., Shear, B.L., Ellersieck, M.R. & Asfaw, A. (2001) Identification of Fecal Escherichia coli from Humans and Animals by Ribotyping. *Applied and Environmental Microbiology*, **67**, 1503–1507.

Cartwright, N. (2010) What are randomised controlled trials good for? *Philosophical Studies*, **147**, 59–70.

Chassang, S., Padró i Miquel, G. & Snowberg, E. (2012) Selective Trials: A Principal-Agent Approach to Randomized Controlled Experiments. *American Economic Review*, **102**, 1279–1309.

Cho, K.H., Pachepsky, Y.A., Oliver, D.M., Muirhead, R.W., Park, Y., Quilliam, R.S. & Shelton, D.R. (2016) Modeling fate and transport of fecally-derived microorganisms at the watershed scale: State of the science and future opportunities. *Water Research*, **100**, 38–56.

Chowns, E. (2015) Is Community Management an Efficient and Effective Model of Public Service Delivery? Lessons from the Rural Water Supply Sector in Malawi. *Public Administration and Development*, **35**, 263–276.

Chuang, P., Trottier, S. & Murcott, S. (2011) Comparison and verification of four field-based microbiological tests: H2S test, Easygel®, Colilert®, Petrifilm™. *Journal of Water Sanitation and Hygiene for Development*, **1**, 68–85.

Clasen, T. (2009) *Scaling Up Household Water Treatment Among Low-Income Populations*. World Health Organization, Geneva, Switzerland.

Clasen, T., Cairncross, S., Haller, L., Bartram, J. & Walker, D. (2007a) Cost-effectiveness of water quality interventions for preventing diarrhoeal disease in developing countries. *Journal of Water and Health*, **5**, 599-608.

Clasen, T., Schmidt, W.-P., Rabie, T., Roberts, I. & Cairncross, S. (2007b) Interventions to improve water quality for preventing diarrhoea: systematic review and meta-analysis. *British Medical Journal*, **334**, 782–785.

Coase, R.H. (1960) The Problem of Social Cost. *The Journal of Law and Economics*, **3**, 1–44.

Cook, C.N., Mascia, M.B., Schwartz, M.W., Possingham, H.P. & Fuller, R.A. (2013) Achieving conservation science that bridges the knowledge-action boundary. *Conservation Biology*, **27**, 669–678.

Costanza, R., de Groot, R., Sutton, P., van der Ploeg, S., Anderson, S.J., Kubiszewski, I., Farber, S. & Turner, R.K. (2014) Changes in the global value of ecosystem services. *Global Environmental Change*, **26**, 152–158.

Council of Economic Advisers. (2014) Evaluation as a tool for improving federal programs. *Economic Report of the President, Together with the Annual Report of the Council of Economic Advisors*, pp. 269–298. U.S. Government Printing Office, Washington DC.

Crane, S.R., Moore, J.A., Grismer Jr., M.E. & Milner, J.R. (1983) Bacterial pollution from agricultural sources: A review. *Transactions of the American Society of Agricultural Engineers*, **26**, 858–866.

Crittenden, J.C., Trussell, R.R., Hand, D.W., Howe, K.J. & Tchobanoglous, G. (2012) *Water Treatment: Principles and Design*. 3rd edition. John Wiley & Sons, Hoboken, NJ.

Curzon, H.F. & Kontoleon, A. (2016) From ignorance to evidence? The use of programme evaluation in conservation: Evidence from a Delphi survey of conservation experts. *Journal of Environmental Management*, **180**, 466–475.

Dale, V.H., Pearson, S.M., Offerman, H.L. & O'Neill, R. V. (1994) Relating Patterns of Land-Use Change to Faunal Biodiversity in the Central Amazon. *Conservation Biology*, **8**, 1027–1036.

Daly, H.E. & Farley, J. (2010) *Ecological Economics: Principles and Applications*. 2nd edition. Island Press, Washington DC.

Davey, C., Aiken, A.M., Hayes, R.J. & Hargreaves, J.R. (2015) Re-analysis of health and educational impacts of a school-based deworming programme in western Kenya: a statistical replication of a cluster quasi-randomized stepped-wedge trial. *International Journal of Epidemiology*, **44**, 1581–1592.

Death, R.G. & Collier, K.J. (2010) Measuring stream macroinvertebrate responses to gradients of vegetation cover: When is enough enough? *Freshwater Biology*, **55**, 1447–1464.

Deaton, A. (2010) Instruments, Randomization, and Learning about Development. *Journal of Economic Literature*, **48**, 424–455.

Deaton, A. & Cartwright, N. (2016) *Understanding and Misunderstanding Randomized Controlled Trials*. NBER Working Paper No. 22595, Cambridge, MA.

Dirección de Áreas Protegidas del Gobierno Autónomo de Santa Cruz, Fundación Natura Bolivia & Centro para el Desarrollo Institucional Humano y Ecologico. (2009) *Plan de Manejo Área Natural de Manejo Integrado Río Grande Y Valles Cruceños*. Dirección de Áreas Protegidas del Gobierno Autónomo de Santa Cruz, Santa Cruz de la Sierra, Bolivia.

Dodds, W.K. & Oakes, R.M. (2008) Headwater influences on downstream water quality. *Environmental Management*, **41**, 367–377.

Dolisca, F., Carter, D.R., McDaniel, J.M., Shannon, D.A. & Jolly, C.M. (2006) Factors influencing farmers' participation in forestry management programs: A case study from Haiti. *Forest Ecology and Management*, **236**, 324–331.

Donnelly, C.A., Woodroffe, R., Cox, D.R., Bourne, F.J., Cheeseman, C.L., Clifton-Hadley, R.S., Wei, G., Gettinby, G., Gilks, P., Jenkins, H., Johnston, W.T., Le Fevre, A.M., Mclnerney, J.P. & Morrison, W.I. (2005) Positive and negative effects of widespread badger culling on tuberculosis in cattle. *Nature*, **439**, 843–846.

Donner, A. & Klar, N. (2004) Pitfalls of and Controversies in Cluster Randomization Trials. *American Journal of Public Health*, **94**, 416–422.

Drechsler, M., Smith, H.G., Sturm, A. & Wätzold, F. (2016) Cost-effectiveness of conservation payment schemes for species with different range sizes. *Conservation Biology*, **30**, 894–899.

Edberg, S.C., Rice, E.W., Karlin, R.J. & Allen, M.J. (2000) Escherichia coli: the best biological drinking water indicator for public health protection. *Journal of Applied Microbiology*, **88**, 106–116.

Engel, S. (2016) The devil in the detail: A practical guide on designing payments for environmental services. *International Review of Environmental and Resource Economics*, **9**, 131–177.

Engel, S., Pagiola, S. & Wunder, S. (2008) Designing payments for environmental services in theory and practice: An overview of the issues. *Ecological Economics*, **65**, 663–674.

Enger, K.S., Nelson, K.L., Rose, J.B. & Eisenberg, J.N.S. (2013) The joint effects of efficacy and compliance: A study of household water treatment effectiveness against childhood diarrhea. *Water Research*, **47**, 1181–1190.

Ezzine-de-Blas, D., Wunder, S., Ruiz-Pérez, M. & Del Pilar Moreno-Sanchez, R. (2016) Global patterns in the implementation of payments for environmental services. *PLoS ONE*, **11,** e0149847.

de F. Fernandes, J., de Souza, A.L.T. & Tanaka, M.O. (2014) Can the structure of a riparian forest remnant influence stream water quality? A tropical case study. *Hydrobiologia*, **724**, 175–185.

Feikin, D.R., Audi, A., Olack, B., Bigogo, G.M., Polyak, C., Burke, H., Williamson, J. & Breiman, R.F. (2010) Evaluation of the optimal recall period for disease symptoms in home-based morbidity surveillance in rural and urban Kenya. *International Journal of Epidemiology*, **39**, 450–458.

Feng, Z., Diehr, P., Peterson, A. & McLerran, D. (2001) Selected Statistical Issues in Group Randomized Trials. *Annual Review of Public Health*, **22**, 167–187.

Ferraro, P.J. & Hanauer, M.M. (2014) Advances in Measuring the Environmental and Social Impacts of Environmental Programs. *Annual Review of Environment and Resources*, **39**, 495–517.

Ferraro, P.J. & Kiss, A. (2002) Direct Payments to Conserve Biodiversity. *Science*, **298**, 1718–1719.

Ferraro, P.J. & Pattanayak, S.K. (2006) Money for Nothing? A Call for Empirical Evaluation of

Biodiversity Conservation Investments. *PLoS Biology*, **4**, e105.

Ferraro, P.J. & Simpson, R.D. (2002) The Cost-Effectiveness of Conservation Payments. *Land Economics*, **78**, 339–353.

Fewtrell, L., Kaufmann, R.B., Kay, D., Enanoria, W., Haller, L. & Colford Jr., J.M. (2005) Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: a systematic review and meta-analysis. *Lancet Infectious Diseases*, **5**, 42–52.

Fisher, R.A. (1935) *The Design of Experiments*. Oliver and Boyd, Edinburgh, Scotland.

Fisher, B., Turner, R.K. & Morling, P. (2009) Defining and classifying ecosystem services for decision making. *Ecological Economics*, **68**, 643–653.

Folke, C. (2006) The economic perspective: Conservation against development versus conservation for development. *Conservation Biology*, **20**, 686–688.

Foster, V. (2005) *Ten Years of Water Service Reforms in Latin America: Toward an Anglo-French Model*. Water Supply and Sanitation Sector Board discussion paper series No. 3. World Bank, Washington DC.

Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A. & Sibert, J. (2012) AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, **27**, 233–249.

Freeman, M.C., Trinies, V., Boisson, S., Mak, G. & Clasen, T. (2012) Promoting Household Water Treatment through Women's Self Help Groups in Rural India: Assessing Impact on Drinking Water Quality and Equity. *PLoS ONE*, **7**, e44068.

Gibbons, J.M., Nicholson, E., Milner-Gulland, E.J. & Jones, J.P.G. (2011) Should payments for biodiversity conservation be based on action or results? *Journal of Applied Ecology*, **48**, 1218–1226.

Glennerster, R. & Takavarasha, K. (2013) *Running Randomized Evaluations: A Practical Guide*. Princeton University Press, Princeton, NJ.

Gobierno Autónomo Departamental de Santa Cruz. (2012) *Decreto Departamental N° 161*. Santa Criz de la Sierra, Bolivia.

Gobierno Autónomo Departamental de Santa Cruz. (2017) Sistema Departamental de Áreas Protegidas - Área Natural de Manejo Integrado Río Grande Valles Cruceños, http://www.santacruz.gob.bo/turistica/medioambiente/recursos/areasprotegidas/dptal/conte nido.php?IdNoticia=3007&IdMenu=30002220. Accessed 19th December 2017.

Gómez-Baggethun, E., de Groot, R., Lomas, P.L. & Montes, C. (2010) The history of ecosystem services in economic theory and practice: From early notions to markets and payment schemes. *Ecological Economics*, **69**, 1209–1218.

Goodwin, B.K. & Smith, V.H. (2003) An ex post evaluation of the conservation reserve, federal crop insurance, and other government programs: Program participation and soil erosion. *Journal of Agricultural and Resource Economics*, **28**, 201–216.

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

Grieg-Gran, M., Porras, I. & Wunder, S. (2005) How can market mechanisms for forest environmental services help the poor? Preliminary lessons from Latin America. *World Development*, **33**, 1511–1527.

Grillos, T. (2017) Economic vs non-material incentives for participation in an in-kind payments for ecosystem services program in Bolivia. *Ecological Economics*, **131**, 178–190.

Grima, N., Singh, S.J., Smetschka, B. & Ringhofer, L. (2016) Payment for Ecosystem Services (PES) in Latin America: Analysing the performance of 40 case studies. *Ecosystem Services*, **17**, 24–32.

Gross, M. (2010) *Ignorance and Surprise - Science, Society, and Ecological Design*. MIT Press, Cambridge, MA.

Gruber, J.S., Ercumen, A. & Colford, J.M. (2014) Coliform Bacteria as Indicators of Diarrheal Risk in Household Drinking Water: Systematic Review and Meta-Analysis. *PLoS ONE*, **9**, e107429.

Hamel, P. & Bryant, B.P. (2017) Uncertainty assessment in ecosystem services analyses: Seven challenges and practical responses. *Ecosystem Services*, **24**, 1–15.

Hamilton, L.S. (2008) *Forests and Water: A Thematic Study Prepared in the Framework of the Global Forest Resources Assessment, 2005*. Food and Agriculture Organization of the United Nations, Rome, Italy.

Harding, J.S., Claassen, K. & Evers, N. (2006) Can forest fragments reset physical and water quality conditions in agricultural catchments and act as refugia for forest stream invertebrates? *Hydrobiologia*, **568**, 391–402.

Hawkes, H.A. (1998) Origin and development of the biological monitoring working party score system. *Water Research*, **32**, 964–968.

Haynes, L., Service, O., Goldacre, B. & Torgerson, D. (2012) *Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials*. UK Government, Cabinet Office Behavioural Insights Team, London.

Herrera, D., Ellis, A., Fisher, B., Golden, C.D., Johnson, K., Mulligan, M., Pfaff, A., Treuer, T. & Ricketts, T.H. (2017) Upstream watershed condition predicts rural children's health across 35 developing countries. *Nature Communications*, **8**, 811.

Howell, J.M., Coyne, M.S. & Cornelius, P. (1995) Fecal Bacteria in Agricultural Waters of the Bluegrass Region of Kentucky. *Journal of Environment Quality*, **24**, 411.

Hunter, P.R. (2009) Household Water Treatment in Developing Countries: Comparing Different Intervention Types Using Meta-Regression. *Environmental Science & Technology*, **43**, 8991–8997.

Ibisch, P.L., Beck, S.G., Gerkmann, B. & Carretero, A. (2003) Ecoregiones y ecosistemas. In *Biodiversidad: la riqueza de Bolivia* (eds P.L. Ibisch & G. Mérida), pp. 47–88. Editorial Fundación Amigos de la Naturaleza, Santa Cruz de la Sierra, Bolivia.

Independent Evaluation Group. (2012) *World Bank Group Impact Evaluations: Relevance and Effectiveness*. World Bank, Washington DC.

Ingram, J.C., Wilkie, D., Clements, T., McNab, R.B., Nelson, F., Baur, E.H., Sachedina, H.T., Peterson, D.D. & Foley, C.A.H. (2014) Evidence of Payments for Ecosystem Services as a mechanism for supporting biodiversity conservation and rural livelihoods. *Ecosystem Services*, **7**, 10–21.

Iñiguez-Armijos, C., Leiva, A., Frede, H.G., Hampel, H. & Breuer, L. (2014) Deforestation and benthic indicators: How much vegetation cover is needed to sustain healthy Andean streams? *PLoS ONE*, **9**, e105869.

Jack, B.K., Kousky, C. & Sims, K.R.E. (2008) Designing payments for ecosystem services: Lessons from previous experience with incentive-based mechanisms. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 9465–9470.

Jakovljevic, M. (2014) The placebo–nocebo response: Controversies and challenges from clinical and research perspective. *European Neuropsychopharmacology*, **24**, 333–341.

Jamal, F., Fletcher, A., Shackleton, N., Elbourne, D., Viner, R. & Bonell, C. (2015) The three stages of building and testing mid-level theories in a realist RCT: a theoretical and methodological case-example. *Trials*, **16**, 466.

Jawson, M.D., Elliott, L.F., Saxton, K.E. & Fortier, D.H. (1982) The effect of cattle grazing on indicator bacteria in runoff from a pacific northwest watershed. *Journal of Environment Quality*, **11**, 621–627.

Jayachandran, S., de Laat, J., Lambin, E.F., Stanton, C.Y., Audy, R. & Thomas, N.E. (2017) Cash for carbon: A randomized trial of payments for ecosystem services to reduce deforestation. *Science*, **357**, 267–273.

Jemio, M.T. (2017) Tres claves para entender por qué las comunidades se oponen a la hidroeléctrica Rositas en Bolivia, https://es.mongabay.com/2017/11/tres-claves-entender-las-comunidades-se-oponen-la-hidroelectrica-rositas-bolivia/. Accessed 19th December 2017.

Keeler, B.L., Polasky, S., Brauman, K.A., Johnson, K.A., Finlay, J.C., O'Neill, A., Kovacs, K. & Dalzell, B. (2012) Linking water quality and well-being for improved assessment and valuation of ecosystem services. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 18619–24.

Kerry, R., Eriksen, T.E., Lie, S.A.N., Mumford, S.D. & Anjum, R.L. (2012) Causation and evidence-based practice: An ontological review. *Journal of Evaluation in Clinical Practice*, **18**, 1006–1012.

Kilian, M. & Bülow, P. (1979) Rapid Identification of Enterobacteriaceae. *Acta Pathologica Microbiologica Scandinavica Section B Microbiology*, **87B**, 271–276.

Klein, B.C. (1989) Effects of Forest Fragmentation on Dung and Carrion Beetle Communities in Central Amazonia. *Ecology*, **70**, 1715–1725.

Kolinjivadi, V., Adamowski, J. & Kosoy, N. (2014) Recasting payments for ecosystem services (PES) in water resource management: A novel institutional approach. *Ecosystem Services*, **10**, 144–154.

Kolinjivadi, V., Grant, A., Adamowski, J. & Kosoy, N. (2015) Juggling multiple dimensions in a complex socio-ecosystem: The issue of targeting in payments for ecosystem services. *Geoforum*, **58**, 1–13.

Kontoleon, A., Conteh, B., Bulte, E., List, J.A., Mokuwa, E., Richards, P., Turley, T. & Voors, M. (2016) *The Impact of Conditional and Unconditional Transfers on Livelihoods and Conservation in Sierra Leone, 3ie Impact Evaluation Report 46.* International Initiative for Impact Evaluation (3ie), New Delhi, India.

Kosoy, N. & Corbera, E. (2010) Payments for ecosystem services as commodity fetishism. *Ecological Economics*, **69**, 1228–1236.

Kosoy, N., Martinez-Tuna, M., Muradian, R. & Martinez-Alier, J. (2007) Payments for environmental services in watersheds: Insights from a comparative study of three cases in Central America. *Ecological Economics*, **61**, 446–455.

Kostyla, C., Bain, R., Cronk, R. & Bartram, J. (2015) Seasonal variation of fecal contamination in drinking water sources in developing countries: A systematic review. *Science of the Total Environment*, **514**, 333–343.

Kremer, M., Leino, J., Miguel, E. & Zwane, A.P. (2011) Spring cleaning: Rural water impacts, valuation, and property rights institutions. *Quarterly Journal of Economics*, **126**, 145–205.

Krueger, T., Page, T., Hubacek, K., Smith, L. & Hiscock, K. (2012) The role of expert opinion in environmental modelling. *Environmental Modelling and Software*, **36**, 4–18.

Kumi-Kyereme, A. & Amo-Adjei, J. (2016) Household wealth, residential status and the incidence of diarrhoea among children under-five years in Ghana. *Journal of Epidemiology and Global Health*, **6**, 131–140.

Leamer, E.E. (1983) Let's take the con out of econometrics. *American Economic Review*, **73**, 31–43.

LeChevallier, M.W., Evans, T.M. & Seidler, R.J. (1981) Effect of turbidity on chlorination efficiency and bacterial persistence in drinking water. *Applied and Environmental Microbiology*, **42**, 159–167.

Leclerc, H., Mossel, D.A., Edberg, S.C. & Struijk, C.B. (2001) Advances in the bacteriology of the coliform group: their suitability as markers of microbial water safety. *Annual Review of Microbiology*, **55**, 201–34.

Lele, S. (2009) Watershed services of tropical forests: from hydrology to economic valuation to integrated analysis. *Current Opinion in Environmental Sustainability*, **1**, 148–155.

Levitt, S.D. & List, J.A. (2009) Field experiments in economics: The past, the present, and the future. *European Economic Review*, **53**, 1–18.

Levitt, S.D. & List, J.A. (2011) Was There Really a Hawthorne Effect at the Hawthorne Plant? An Analysis of the Original Illumination Experiments. *American Economic Journal: Applied Economics*, **3**, 224–238.

Lindenmayer, D.B. & Likens, G.E. (2009) Adaptive monitoring: a new paradigm for long-term research and monitoring. *Trends in Ecology & Evolution*, **24**, 482–486.

Line, D.E. (2003) Changes In a Stream's Physical and Biological Conditions Following Livestock Exclusion. *American Society of Agricultural Engineers*, **46**, 287–293.

List, J.A. & Rasul, I. (2011) Field Experiments in Labor Economics. In *Handbook of Labor Economics* (eds. O. Ashenfelter & D. Card), pp. 104–228. North Holland, Amsterdam, The Netherlands.

Liu, L., Johnson, H.L., Cousens, S., Perin, J., Scott, S., Lawn, J.E., Rudan, I., Campbell, H., Cibulskis, R., Li, M., Mathers, C. & Black, R.E. (2012) Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *The Lancet*, **379**, 2151–2161.

Liu, J., Qu, H., Huang, D., Chen, G., Yue, X., Zhao, X. & Liang, Z. (2014) The role of social capital in encouraging residents' pro-environmental behaviors in community-based ecotourism. *Tourism Management*, **41**, 190–201.

Lubell, M., Schneider, M., Scholz, J.T. & Mete, M. (2002) Watershed partnerships and the emergence of collective action institutions. *American Journal of Political Science*, **46**, 148–163.

Millennium Ecosystem Assessment. (2005) *Ecosystems and Human Well-Being: Synthesis.* Island Press, Washington DC.

Mann, C.J. (2003) Observational research methods. Research design II: cohort, cross sectional, and case-control studies. *Emergency Medicine Journal*, **20**, 54–60.

Margoluis, R., Russell, V., Gonzalez, M., Rojas, O., Magdaleno, J., Madrid, G. & Kaimowitz, D. (2001) *Maximum Yield? Sustainable Agriculture as a Tool for Conservation*. Biodiversity Support Program, Washington DC.

Margoluis, R., Stem, C., Salafsky, N. & Brown, M. (2009) Design alternatives for evaluating the impact of conservation projects. *New Directions for Evaluation*, **122**, 85–96.

Martin-Ortega, J., Ojea, E. & Roux, C. (2013) Payments for water ecosystem services in Latin America: A literature review and conceptual model. *Ecosystem Services*, **6**, 122–132.

Martin Persson, U. & Alpízar, F. (2013) Conditional Cash Transfers and Payments for Environmental Services—A Conceptual Framework for Explaining and Judging Differences in Outcomes. *World Development*, **43**, 124–137.

McDonald, R.I., Weber, K.F., Padowski, J., Boucher, T. & Shemie, D. (2016) Estimating watershed degradation over the last century and its impact on water-treatment costs for the world's large cities. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, 9117–9122.

McDonald, R.I., Weber, K., Padowski, J., Flörke, M., Schneider, C., Green, P.A., Gleeson, T., Eckman, S., Lehner, B., Balk, D., Boucher, T., Grill, G. & Montgomery, M. (2014) Water on an urban planet:

Urbanization and the reach of urban water infrastructure. *Global Environmental Change*, **27**, 96–105.

McMillan, H., Krueger, T. & Freer, J. (2012) Benchmarking observational uncertainties for hydrology: Rainfall, river discharge and water quality. *Hydrological Processes*, **26**, 4078–4111.

Meals, D.W., Dressing, S.A. & Davenport, T.E. (2010) Lag time in water quality response to best management practices: a review. *Journal of Environmental Quality*, **39**, 85–96.

Melo, M.C.N. De, Taddei, J.A.D. a C., Diniz-Santos, D.R., May, D.S., Carneiro, N.B. & Silva, L.R. (2007) Incidence of diarrhea: poor parental recall ability. *Brazilian Journal of Infectious Diseases*, **11**, 571–579.

Michalopoulos, C., Bloom, H.S. & Hill, C.J. (2004) Can Propensity-Score Methods Match the Findings from a Random Assignment Evaluation of Mandatory Welfare-to-Work Programs? *Review of Economics and Statistics*, **86**, 156–179.

Micrology Labs. (2016) Detection of Waterborne Coliforms and E. coli with Coliscan Easygel, https://www.micrologylabs.com/page/95/Instructions. Accessed 19th December 2017.

Miguel, E. & Kremer, M. (2004) Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities. *Econometrica*, **72**, 159–217.

Milsom, K., Blinkhorn, A., Worthington, H., Threlfall, A., Buchanan, K., Kearney-Mitchell, P. & Tickle, M. (2006) The effectiveness of school dental screening: a cluster-randomized control trial. *Journal of Dental Research*, **85**, 924–8.

Ministerio de Medio Ambiente y Agua. (2010) *Norma Boliviana 512*. La Paz, Bolivia.

Ministerio de Medio Ambiente y Agua. (2011) *Guía Para La Evaluación de Las Condiciones Biológicas de Cuerpos de Agua Utilizando Macroinvertebrados Bentónicos*. La Paz, Bolivia.

Ministerio de Medio Ambiente y Agua. (2015) *Más Inversión Para Agua - MIAGUA*. La Paz, Bolivia.

Mintz, E.D., Bartram, J., Lochery, P. & Wegelin, M. (2001) Not just a drop in the bucket: Expanding access to point-of-use water treatment systems. *American Journal of Public Health*, **91**, 1565–1570.

Miteva, D.A., Pattanayak, S.K. & Ferraro, P.J. (2012) Evaluation of biodiversity policy instruments: What works and what doesn't? *Oxford Review of Economic Policy*, **28**, 69–92.

Mokondoko, P., Manson, R.H. & Pérez-Maqueo, O. (2016) Assessing the service of water quality regulation by quantifying the effects of land use on water quality and public health in central Veracruz, Mexico. *Ecosystem Services*, **22**, 161–173.

Montgomery, M.A. & Elimelech, M. (2007) Water and Sanitation in Developing Countries: Including Health in the Equation. *Environmental Science & Technology*, **41**, 17–24.

Morris, C. & Potter, C. (1995) Recruiting the new conservationists: Farmers' adoption of agri-environmental schemes in the U.K. *Journal of Rural Studies*, **11**, 51–63.

Muñoz-Piña, C., Guevara, A., Torres, J.M. & Braña, J. (2008) Paying for the hydrological services of Mexico's forests: Analysis, negotiations and results. *Ecological Economics*, **65**, 725–736.

Muradian, R., Arsel, M., Pellegrini, L., Adaman, F., Aguilar, B., Agarwal, B., Corbera, E., Ezzine de Blas, D., Farley, J., Froger, G., Garcia-Frapolli, E., Gómez-Baggethun, E., Gowdy, J., Kosoy, N., Le Coq, J.F., Leroy, P., May, P., Méral, P., Mibielli, P., Norgaard, R., Ozkaynak, B., Pascual, U., Pengue, W., Perez, M., Pesche, D., Pirard, R., Ramos-Martin, J., Rival, L., Saenz, F., Van Hecken, G., Vatn, A., Vira, B. & Urama, K. (2013) Payments for ecosystem services and the fatal attraction of win-win solutions. *Conservation Letters*, **6**, 274–279.

Muradian, R., Corbera, E., Pascual, U., Kosoy, N. & May, P.H. (2010) Reconciling theory and practice: An alternative conceptual framework for understanding payments for environmental services. *Ecological Economics*, **69**, 1202–1208.

Murtinho, F., Tague, C., de Bievre, B., Eakin, H. & Lopez-Carr, D. (2013) Water Scarcity in the Andes: A Comparison of Local Perceptions and Observed Climate, Land Use and Socioeconomic Changes. *Human Ecology*, **41**, 667–681.

Myers, S.S., Gaffikin, L., Golden, C.D., Ostfeld, R.S., H. Redford, K., H. Ricketts, T., Turner, W.R. & Osofsky, S.A. (2013) Human health impacts of ecosystem alteration. *Proceedings of the National Academy of Sciences*, **110**, 18753–18760.

Naeem, S., Ingram, J.C., Varga, A., Agardy, T., Barten, P., Bennett, G., Bloomgarden, E., Bremer, L.L., Burkill, P., Cattau, M., Ching, C., Colby, M., Cook, D.C., Costanza, R., DeClerck, F., Freund, C., Gartner, T., Goldman-Benner, R., Gunderson, J., Jarrett, D., Kinzig, A.P., Kiss, A., Koontz, A., Kumar, P., Lasky, J.R., Masozera, M., Meyers, D., Milano, F., Naughton-Treves, L., Nichols, E., Olander, L., Olmsted, P., Perge, E., Perrings, C., Polasky, S., Potent, J., Prager, C., Quetier, F., Redford, K., Saterson, K., Thoumi, G., Vargas, M.T., Vickerman, S., Weisser, W., Wilkie, D. & Wunder, S. (2015) Get the science right when paying for nature's services. *Science*, **347**, 1206–1207.

Naughton-Treves, L. & Wendland, K. (2014) Land Tenure and Tropical Forest Carbon Management. *World Development*, **55**, 1–6.

Nelson, F. (2008) *Developing Alternative Frameworks for Community-based Conservation: Piloting Payments for Environmental Services (PES) in Tanzania's Simanjiro Plains.* USAID TransLinks No. 35. United States Agency for International Development, Washington DC.

Norgaard, R.B. (2010) Ecosystem services: From eye-opening metaphor to complexity blinder. *Ecological Economics*, **69**, 1219–1227.

Oliver, D.M., Page, T., Hodgson, C.J., Heathwaite, A.L., Chadwick, D.R., Fish, R.D. & Winter, M. (2010) Development and testing of a risk indexing framework to determine field-scale critical source areas of faecal bacteria on grassland. *Environmental Modelling and Software*, **25**, 503–512.

Ongley, E.D. (1996) *Control of Water Pollution from Agriculture - FAO Irrigation and Drainage Paper 55*. Food and Agriculture Organization of the United Nations, Rome, Italy.

Orne, M.T. (1962) On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, **17**, 776–783.

Pachepsky, Y.A. & Shelton, D.R. (2011) Escherichia coli and Fecal Coliforms in Freshwater and Estuarine Sediments. *Critical Reviews in Environmental Science and Technology*, **41**, 1067–1110.

Pagiola, S., Arcenas, A. & Platais, G. (2005) Can Payments for Environmental Services help reduce poverty? An exploration of the issues and the evidence to date from Latin America. *World Development*, **33**, 237–253.

Paredes, S. & Isurza, J.L. (2012) *Estudio de Impacto de La Ganadería en el ANMI Río Grande - Valles Cruceños*. Fundación Natura Bolivia, Santa Cruz de la Sierra, Bolivia.

Parkhurst, G.M. & Shogren, J.F. (2007) Spatial incentives to coordinate contiguous habitat. *Ecological Economics*, **64**, 344–355.

Pattanayak, S.K. (2009) *Rough Guide to Impact Evaluation of Environmental and Development Programs*. South Asian Network for Development and Environmental Economics, Kathmandu, Nepal.

Pattanayak, S.K. & Wendland, K.J. (2007) Nature's care: Diarrhea, watershed protection, and biodiversity conservation in Flores, Indonesia. *Biodiversity and Conservation*, **16**, 2801–2819.

Pattanayak, S.K., Wunder, S. & Ferraro, P.J. (2010) Show me the money: Do payments supply environmental services in developing countries? *Review of Environmental Economics and Policy*,

4, 254–274.

Pearce, D.W. & Turner, R.K. (1990) *Economics of Natural Resources and the Environment*. 6th edition. Johns Hopkins University Press, Baltimore, MD.

Perrot-Maître, D. (2006) *The Vittel Payments for Ecosystem Services: A 'perfect' PES Case*. International Institute for Environment and Development, London.

Picciotto, R. (2012) Experimentalism and development evaluation: Will the bubble burst? *Evaluation*, **18**, 213–229.

Ponette-González, A.G., Brauman, K.A., Marín-Spiotta, E., Farley, K.A., Weathers, K.C., Young, K.R. & Curran, L.M. (2015) Managing water services in tropical regions: From land cover proxies to hydrologic fluxes. *Ambio*, **44**, 367–375.

Ponette-Gonzalez, A.G., Marin-Spiotta, E., Brauman, K.A., Farley, K.A., Weathers, K.C. & Young, K.R. (2014) Hydrologic Connectivity in the High-Elevation Tropics: Heterogeneous Responses to Land Change. *BioScience*, **64**, 92–104.

Porras, I., Grieg-Gran, M. & Neves, N. (2008) *All that glitters: a review of payments for watershed services in developing countries*. Natural Resource Issues No. 11. International Institute for Environment and Development, London.

Porras, I., Barton, D.N., Miranda, M. & Chacón-Cascante, A. (2013) *Learning from 20 Years of Payments for Ecosystem Services in Costa Rica*. International Institute for Environment and Development, London.

Pritchett, L. (2014) An Homage to the Randomistas on the Occasion of the J-PAL 10th Anniversary: Development as a Faith-Based Activity, https://www.cgdev.org/blog/homage-randomistas-occasion-j-pal-10th-anniversary-development-faith-based-activity. Accessed 19th December 2017.

Prüss-Ustün, A., Bartram, J., Clasen, T., Colford, J.M., Cumming, O., Curtis, V., Bonjour, S., Dangour, A.D., De France, J., Fewtrell, L., Freeman, M.C., Gordon, B., Hunter, P.R., Johnston, R.B., Mathers, C., Mäusezahl, D., Medlicott, K., Neira, M., Stocks, M., Wolf, J. & Cairncross, S. (2014) Burden of disease from inadequate water, sanitation and hygiene in low- and middle-income settings: a retrospective analysis of data from 145 countries. *Tropical Medicine & International Health*, **19**, 894–905.

Pullin, A.S. & Knight, T.M. (2001) Effectiveness in conservation practice: Pointers from medicine and public health. *Conservation Biology*, **15**, 50–54.

Pullin, A.S., Knight, T.M., Stone, D.A. & Charman, K. (2004) Do conservation managers use scientific evidence to support their decision-making? *Biological Conservation*, **119**, 245–252.

Pullin, A.S., Sutherland, W., Gardner, T., Kapos, V. & Fa, J.E. (2013) Conservation priorities: Identifying need, taking action and evaluating success. In *Key Topics in Conservation Biology 2* (eds. D. Macdonald & K. Willis), pp. 1–22. John Wiley & Sons, Oxford, UK.

R Development Core Team. (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rasolofoson, R.A., Ferraro, P.J., Jenkins, C.N. & Jones, J.P.G. (2015) Effectiveness of Community Forest Management at reducing deforestation in Madagascar. *Biological Conservation*, **184**, 271–277.

Ravallion, M. (2009) Should the Randomistas Rule? *The Economists' Voice*, **6**, 8–12.

Rawlings, L.B. (2005) Evaluating the Impact of Conditional Cash Transfer Programs. *The World Bank Research Observer*, **20**, 29–55.

Redford, K.H., Padoch, C. & Sunderland, T. (2013) Fads, Funding, and Forgetting in Three Decades of Conservation. *Conservation Biology*, **27**, 437–438.

Rice, E.W., Allen, M.J. & Edberg, S.C. (1990) Efficacy of beta-glucuronidase assay for identification of Escherichia coli by the defined-substrate technology. *Applied and environmental microbiology*, **56**, 1203–5.

Richards, S., Paterson, E., Withers, P.J.A. & Stutter, M. (2016) Septic tank discharges as multi-pollutant hotspots in catchments. *Science of the Total Environment*, **542**, 854–863.

Roberts, L., Chartier, Y., Chartier, O., Malenga, G., Toole, M. & Rodka, H. (2001) Keeping clean water clean in a Malawi refugee camp: a randomized intervention trial. *Bulletin of the World Health Organization*, **79**, 280–7.

Robertson, N. & Wunder, S. (2005) *Fresh Tracks in the Forest: Assessing Incipient Payments For Enivronmental Services Initiatives in Bolivia*. Center for International Forestry Research, Bogor, Indonesia.

Rode, J., Gómez-Baggethun, E. & Krause, T. (2015) Motivation crowding by economic incentives in conservation policy: A review of the empirical evidence. *Ecological Economics*, **117**, 270–282.

Rodríguez-de-Francisco, J.C. & Budds, J. (2015) Payments for environmental services and control over conservation of natural resources: The role of public and private sectors in the conservation of the Nima watershed, Colombia. *Ecological Economics*, **117**, 295–302.

Roe, D. & Elliott, J. (2004) Poverty reduction and biodiversity conservation: rebuilding the bridges. *Oryx*, **38**, 137–139.

Roe, D., Mohammed, E.Y., Porras, I. & Giuliani, A. (2013) Linking biodiversity conservation and poverty reduction: de-polarizing the conservation-poverty debate. *Conservation Letters*, **6**, 162–171.

Rojas Banegas, S. (2012) *Línea Base Biofísica: Área Natural de Manejo Integrado Rio Grande Valles Cruceños*. Fundación Natura Bolivia, Santa Cruz de la Sierra, Bolivia.

Rosenthal, R. & Jacobson, L. (1968) Pygmalion in the classroom. *The Urban Review*, **3**, 16–20.

Rossi, P., Lipsey, M. & Freeman, H. (2004) *Evaluation: A Systematic Approach*. 6thedition. SAGE Publications, Thousand Oaks, CA.

Rubin, D.B. (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, **66**, 688–701.

Rundlöf, M., Andersson, G.K.S., Bommarco, R., Fries, I., Hederström, V., Herbertsson, L., Jonsson, O., Klatt, B.K., Pedersen, T.R., Yourstone, J. & Smith, H.G. (2015) Seed coating with a neonicotinoid insecticide negatively affects wild bees. *Nature*, **521**, 77–80.

Salafsky, N., Salzer, D., Stattersfield, A.J., Hilton-Taylor, C., Neugarten, R., Butchart, S.H.M., Collen, B., Cox, N., Master, L.L., O'Connor, S. & Wilkie, D. (2008) A Standard Lexicon for Biodiversity Conservation: Unified Classifications of Threats and Actions. *Conservation Biology*, **22**, 897–911.

Samii, C., Lisiecki, M., Kulkarni, P., Paler, L. & Chavis, L. (2014) Effects of Payment for Environmental Services (PES) on Deforestation and Poverty in Low and Middle Income Countries: A Systematic Review. *Campbell Systematic Reviews*, **10**.

Sánchez-Azofeifa, G.A., Pfaff, A., Robalino, J.A. & Boomhower, J.P. (2007) Costa Rica's payment for environmental services program: Intention, implementation, and impact. *Conservation Biology*, **21**, 1165–1173.

Santos de Lima, L., Krueger, T. & García-Marquez, J. (2017) Uncertainties in demonstrating environmental benefits of payments for ecosystem services. *Ecosystem Services*, **27**, 139–149.

Saretsky, G. (1972) The OEO PC experiment and the John Henry effect. *Phi Delta Kappan*, **53**, 579–581.

Scriven, M. (1967) The methodology of evaluation. In *Perspectives of curriculum evaluation* (eds. R.W. Tyler, R.M. Gagne, & M. Scriven), pp. 39–83. Rand McNally, Chicago, IL.

Scriven, M. (2008) A summative evaluation of RCT methodology: and an alternative approach to causal research. *Journal of Multidisciplinary Evaluation*, **5**, 11–24.

Segan, D.B., Bottrill, M.C., Baxter, P.W.J. & Possingham, H.P. (2011) Using Conservation Evidence to Guide Management. *Conservation Biology*, **25**, 200–202.

Senn, S. (2013) Seven myths of randomisation in clinical trials. *Statistics in Medicine*, **32**, 1439–1450.

Shannon, M.A., Bohn, P.W., Elimelech, M., Georgiadis, J.G., Mariñas, B.J. & Mayes, A.M. (2008) Science and technology for water purification in the coming decades. *Nature*, **452**, 301–310.

Silvertown, J. (2015) Have Ecosystem Services Been Oversold? *Trends in Ecology & Evolution*, **30**, 641–648.

Skaug, H., Fournier, D., Bolker, B., Magnusson, A. & Nielsen, A. (2016) Generalized Linear Mixed Models using AD Model Builder. R package version 0.8.3.3.

Sommerville, M.M., Jones, J.P.G. & Milner-Gulland, E.J. (2009) A Revised Conceptual Framework for Payments for Environmental Services. *Ecology and Society*, **14**, 34.

Sommerville, M.M., Milner-Gulland, E.J. & Jones, J.P.G. (2011) The challenge of monitoring biodiversity in payment for environmental service interventions. *Biological Conservation*, **144**, 2832–2841.

Spears, D., Ghosh, A. & Cumming, O. (2013) Open Defecation and Childhood Stunting in India: An Ecological Analysis of New Data from 112 Districts. *PLoS ONE*, **8**, e73784.

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R. & Befani, B. (2012) *Broadening the Range of Designs and Methods for Impact Evaluations*. UK Government, Department for International Development. London.

Stirling, A. (2010) Keep it complex. *Nature*, **468**, 1029–1031.

Sunohara, M.D., Topp, E., Wilkes, G., Gottschall, N., Neumann, N., Ruecker, N., Jones, T.H., Edge, T.A., Marti, R. & Lapen, D.R. (2012) Impact of riparian zone protection from cattle on nutrient, bacteria, F-coliphage, and loading of an intermittent stream. *Journal of Environmental Quality*, **41**, 1301–14.

Sutherland, W.J., Pullin, A.S., Dolman, P.M. & Knight, T.M. (2004) The need for evidence-based conservation. *Trends in Ecology and Evolution*, **19**, 305–308.

Tacconi, L. (2012) Redefining payments for environmental services. *Ecological Economics*, **73**, 29–36.

Tayeh, A. & Cairncross, S. (1995) The reliability of retrospective studies using a one-year recall period to measure dracunculiasis prevalence in Ghana. *International Journal of Epidemiology*, **24**, 1233–1239.

Le Tellier, V., Carrasco, A. & Asquith, N. (2009) Attempts to determine the effects of forest cover on stream flow by direct hydrological measurements in Los Negros, Bolivia. *Forest Ecology and Management*, **258**, 1881–1888.

Tiedemann, A.R., Higgins, D.A., Quigley, T.M., Sanderson, H.R. & Marx, D.B. (1987) Responses of Fecal Coliform in Streamwater to Four Grazing Strategies. *Journal of Range Management*, **40**, 322–329.

Turner, W.R., Brandon, K., Brooks, T.M., Gascon, C., Gibbs, H.K., Lawrence, K.S., Mittermeier, R.A. & Selig, E.R. (2012) Global Biodiversity Conservation and the Alleviation of Poverty. *BioScience*, **62**, 85–92.

UK Government. Guide to Cross-Compliance in England: 2016. https://www.gov.uk/guidance/guide-to-cross-compliance-in-england-2016/gaec-1-establishment-of-buffer-strips-along-watercourses. Accessed 19th December 2017.

UNEG Impact Evaluation Task Force. (2013) *Impact Evaluation in UN Agency Evaluation Systems: Guidance on Selection, Planning and Management*. United Nations, New York, NY.

UNICEF. (2008) *UNICEF Handbook on Water Quality*. United Nations Children's Fund, New York, NY.

United Nations. (2015) *Sustainable Development Goals: 17 Goals to Transform Our World.* http://www.un.org/sustainabledevelopment/sustainable-development-goals/. Accessed 19th December 2017.

USAID. (2016) *Evaluation: Learning from Experience. USAID Evaluation Policy*. United States Agency for International Development, Washington DC.

Uthes, S., Matzdorf, B., Müller, K. & Kaechele, H. (2010) Spatial targeting of agri-environmental measures: Cost-effectiveness and distributional consequences. *Environmental Management*, **46**, 494–509.

Vidaurre, T. & Gonzalez, L. (2011) *Protocolo Línea Base Para El Monitoreo de Biodiversidad (Escarabajos Copráfagos Y Anfíbios) En El ANMI RG-VC*. Fundación Natura Bolivia, Santa Cruz de la Sierra, Bolivia.

Vidon, P., Tedesco, L.P., Wilson, J., Campbell, M.A., Casey, L.R. & Gray, M. (2008) Direct and Indirect Hydrological Controls on Concentration and Loading in Midwestern Streams. *Journal of*

Environmental Quality, **37**, 1761.

Walsh, J.C., Dicks, L. V. & Sutherland, W.J. (2015) The effect of scientific evidence on conservation practitioners' management decisions. *Conservation Biology*, **29**, 88–98.

WaterAid. (2013) Gravity-fed Schemes, www.wateraid.org/~/media/Publications/Gravity-fed-schemes.pdf. Accessed 19th December 2017.

Weaver, R.W., Entry, J.A. & Graves, A. (2005) Numbers of fecal streptococci and Escherichia coli in fresh and dry cattle, horse, and sheep manure. *Canadian Journal of Microbiology*, **51**, 847–851.

White, H. (2013) An introduction to the use of randomised control trials to evaluate development interventions. *Journal of Development Effectiveness*, **5**, 30–49.

White, H. & Phillips, D. (2012) *Addressing Attribution of Cause and Effect in Small N Impact Evaluations: Towards an Integrated Framework*. International Initiative for Impact Evaluation (3ie), New Delhi, India.

Whitmee, S., Haines, A., Beyrer, C., Boltz, F., Capon, A.G., De Souza Dias, B.F., Ezeh, A., Frumkin, H., Gong, P., Head, P., Horton, R., Mace, G.M., Marten, R., Myers, S.S., Nishtar, S., Osofsky, S.A., Pattanayak, S.K., Pongsiri, M.J., Romanelli, C., Soucat, A., Vega, J. & Yach, D. (2015) Safeguarding human health in the Anthropocene epoch: Report of the Rockefeller Foundation-Lancet Commission on planetary health. *The Lancet*, **386**, 1973–2028.

WHO. (2013) *WHO Evaluation Practice Handbook*. World Health Organization, Geneva, Switzerland.

WHO/UNICEF Joint Monitoring Program. (2017) *Safely managed drinking water - thematic report on drinking water 2017.* World Health Organization, Geneva, Switzerland.

Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, NY.

Wiik, E., D'Annunzio, R., Pynegar, E.L., Bottazzi, P., Crespo, D. & Jones, J.P.G. Can Payments for Environmental Services reduce deforestation? Results from a Randomized Control Trial experiment in the Río Grande catchment of Bolivia. *Conservation Letters,* In review.

Williams, B.K. & Brown, E.D. (2016) Technical challenges in the application of adaptive management. *Biological Conservation*, **195**, 255–263.

Wong, C.P., Jiang, B., Kinzig, A.P., Lee, K.N. & Ouyang, Z. (2015) Linking ecosystem characteristics to final ecosystem services for public policy. *Ecology Letters*, **18**, 108–118.

Woodcock, B.A., Bullock, J.M., Shore, R.F., Heard, M.S., Pereira, M.G., Redhead, J., Ridding, L., Dean, H., Sleep, D., Henrys, P., Peyton, J., Hulmes, S., Hulmes, L., Sárospataki, M., Saure, C., Edwards, M., Genersch, E., Knäbe, S. & Pywell, R.F. (2017) Country-specific effects of neonicotinoid pesticides on honey bees and wild bees. *Science*, **356**, 1393–1395.

Wunder, S. (2008) Payments for environmental services and the poor: concepts and preliminary evidence. *Environment and Development Economics*, **13**, 279–297.

Wunder, S. (2015) Revisiting the concept of payments for environmental services. *Ecological Economics*, **117**, 234–243.

Wunder, S., Engel, S. & Pagiola, S. (2008) Taking stock: A comparative analysis of payments for environmental services programs in developed and developing countries. *Ecological Economics*, **65**, 834–852.

Xu, Z., Bennett, M.T., Tao, R. & Xu, J. (2004) China's sloping land conversion programme four years on: current situation and pending issues. *International Forestry Review*, **6**, 317–326.

Zafar, S.N., Luby, S.P. & Mendoza, C. (2010) Recall errors in a weekly survey of diarrhoea in Guatemala: determining the optimal length of recall. *Epidemiology and infection*, **138**, 264–269.

Zimmerman, B.L. & Bierregaard, R.O. (1986) Relevance of the Equilibrium Theory of Island Biogeography and Species-Area Relations to Conservation with a Case from Amazonia. *Journal of Biogeography*, **13**, 133.

Appendix A. List of communities randomly allocated to treatment and control groups.

Four communities were excluded from the RCT: La Tranca and Moro Moro in Moro Moro municipality, Postrervalle in Postrervalle municipality, and Alto la Laja in Vallegrande municipality. Communities lying within the ANMI Río Grande but in Cabezas and Gutiérrez municipalities were not included as the logistics of access from the town of Vallegrande made working in them impractical.

TREATMENT COMMUNITIES		CONTROL COMMUNITIES	
Municipality	*Community*	*Municipality*	*Community*
Moro Moro	Abra del Astillero	Moro Moro	Alto del Veladero
Moro Moro	Añapanco	Moro Moro	Buena Vista
Moro Moro	Candelaria	Moro Moro	El Tholar
Moro Moro	Juan Ramos	Moro Moro	La Higuera
Moro Moro	La Senda	Moro Moro	Lagunitas
Moro Moro	Pampanegra	Moro Moro	Potrerillos
Moro Moro	Saguintal	Moro Moro	Torrecillas
Postrervalle	Llorenti	Postrervalle	Mosquera
Postrervalle	Los Churcos	Postrervalle	Pampas
Postrervalle	Mosquerilla	Postrervalle	Río Vilcas
Postrervalle	Quebrada el Palo	Postrervalle	San Marcos
Postrervalle	Vilcas	Postrervalle	San Miguel
Pucará	Abra del Picacho	Pucará	El Cerro
Pucará	El Cruce	Pucará	El Jague
Pucará	El Estanque	Pucará	El Quiñal
Pucará	El Potrero	Pucará	La Higuera
Pucará	El Pujio	Pucará	La Torre
Pucará	El Tipal	Pucará	Lacayotal
Pucará	Huertas	Pucará	Mizquiloma
Pucará	Loma Larga	Pucará	Salsipuedes Chico
Pucará	Pucará	Pucará	Zapallar
Pucará	Salsipuedes Grande	Samaipata	Alto Florida
Samaipata	Achira	Samaipata	Bella Vista
Samaipata	Agua Rica	Samaipata	Bermejo
Samaipata	Alto la Yuruma	Samaipata	Cuevas
Samaipata	Bicoquin	Samaipata	El Pacay
Samaipata	Capa Rosa	Samaipata	La Coca
Samaipata	Floripondio	Samaipata	La Junta
Samaipata	La Negra	Samaipata	La Laja
Samaipata	Lagunillas	Samaipata	La Pajchita
Samaipata	Las Chacras	Samaipata	Lagunitas
Samaipata	Miscas	Samaipata	Los Alisos
Samaipata	Paredones	Samaipata	Palermo

Samaipata	San Juan del Rosario	Samaipata	Petacas
Samaipata	Sivingalito	Vallegrande	Agua de Oro
Vallegrande	Algodonales	Vallegrande	Aguaditas
Vallegrande	Alto Citano	Vallegrande	Alto Seco
Vallegrande	Chapitas	Vallegrande	Apocentillo
Vallegrande	Chirguanañan	Vallegrande	Arenales
Vallegrande	Chorrillos	Vallegrande	Chapas
Vallegrande	Churo de la Collpa	Vallegrande	Chujllas
Vallegrande	El Ojito	Vallegrande	Churo Citano
Vallegrande	Falda de la Cebada	Vallegrande	Cuevas
Vallegrande	Hornos	Vallegrande	El Palmar (Oeste)
Vallegrande	Huantas	Vallegrande	El Pino
Vallegrande	Islas	Vallegrande	El Rodeo
Vallegrande	Javoncillo	Vallegrande	Estancia Huaico
Vallegrande	Lajas Toco	Vallegrande	Fernandez
Vallegrande	Loma Larga	Vallegrande	Guayabo
Vallegrande	Masicurí	Vallegrande	Kallana
Vallegrande	Mataralcito	Vallegrande	Khasamonte
Vallegrande	Molleaguada	Vallegrande	La Ceja
Vallegrande	Monte Pablo	Vallegrande	La Cruz
Vallegrande	Palmar (Peñones)	Vallegrande	La Hoyada
Vallegrande	Palmitas	Vallegrande	Manchones
Vallegrande	Piraimirí	Vallegrande	Minas
Vallegrande	Plan Citano	Vallegrande	Naranjal
Vallegrande	Potrerillos	Vallegrande	Naranjo
Vallegrande	Pucarillo	Vallegrande	Pampillas
Vallegrande	Rancho Novillero	Vallegrande	Pata Estancia
Vallegrande	Santa Elena	Vallegrande	Quebrada Peñones
Vallegrande	Temporal	Vallegrande	Santa Ana
Vallegrande	Toco Citano	Vallegrande	Torneado Grande
Vallegrande	Torneado Chico	Vallegrande	Vado del Yeso
Vallegrande	Villa Nueva		

Appendix B. Example of the text of a *Watershared* agreement between a landowner and *Natura*

CONTRATO PRIVADO PARA LA CONSERVACIÓN DE VERTIENTES DE AGUA Y BOSQUE

Nivel 1

El presente contrato es suscrito bajo conocimiento del H. Alcalde municipal, comité de vigilancia y los respectivos concejales del municipio de Postrervalle

Primera. Naturaleza Jurídica

El presente contrato se inscribe al amparo del Código Civil Boliviano según lo establecido en el artículo 451, que determina que la voluntad de las partes de asumir obligaciones *de hacer o no hacer*, se convierte en ley entre ellos, mientras no violente el ordenamiento jurídico vigente. Los artículos 519 y 520 del Código Civil indican que el contrato tiene fuerza de ley entre las partes contratantes y no puede ser disuelto sino por consentimiento mutuo o por causas autorizadas por la ley. Se presume que el mismo es ejecutado de buena fe y obliga no solo a lo expresado en él, sino también a todos los efectos que deriven de su naturaleza.

Segunda. De las partes

La Fundación Natura Bolivia, representada por su Directora Ejecutiva, Lic. María Teresa Vargas, con C.I. [xxxxxxxx] SC, con mandato suficiente según poder notarial N° 385/2009 y acuerdo de su directorio, en adelante denominada como **La Fundación.**

El Señor [NAME REDACTED] con C.I. N° [xxxxxxxx] **Sc**, mayor de edad, hábil por ley, vecino de la localidad de [COMMUNITY] del municipio de Postrervalle, provincia Vallegrande del Departamento de Santa Cruz, quien declara ser dueño de una propiedad denominada "[PROPERTY NAME]" ubicada en la comunidad de [COMMUNITY], con una superficie total de [SIZE] según plano de ubicación y documento adjuntos al presente contrato, la cual colinda al norte con la propiedad de [NAME], al sur con El Camino Vecinal, al este con la propiedad de [NAME] y al oeste con la propiedad de [NAME], en adelante denominado como **El propietario.**

Tercera. Objeto
El propietario de manera voluntaria, se compromete a colocar bajo conservación un área de [SIZE] de bosque de la totalidad de su predio, para la conservación de vertientes de agua y bosque.

Cuarta. Duración
El presente acuerdo tiene una duración de 3 años, contados a partir del día de su firma, con posibilidades de renovación, si así lo deciden las partes firmantes.

Quinta. De las compensaciones

La Fundación Natura Bolivia y el **Gobierno Municipal de Postrervalle** se comprometen a entregar, alambres de púa y grampas, a favor de **El propietario.** A cambio de la conservación de vertientes de agua y bosque, las que serán entregadas de la siguiente manera:

a la firma del acuerdo, se entregará un incentivo consistente en; 1 rollo de alambre y 1 kilo de grampa, además de materiales para solucionar el acceso al agua consistente en 2 rollos de alambre y 1 kilo de grampa, según consta en acta de entrega adjunta al presente acuerdo.
a la finalización del segundo año, se entregara un incentivo equivalente a [xx] dólares americanos, este monto será entregado en especie, según consta en acta de entrega adjunta al presente acuerdo
a la finalización del tercer año, se entregara un incentivo equivalente a [xx] dólares americanos, este monto será entregado en especie, según consta en acta de entrega adjunta al presente acuerdo
Se deja expreso en el presente acuerdo que los insumos entregados el primer año serán utilizados en la comunidad de [COMMUNITY]. Además el propietario se compromete a reducir el ingreso del ganado al bosque en conservación, según compromiso del propietario.

Sexta. Obligaciones de las partes

El propietario se compromete a cuidar la totalidad del área de [SIZE] de bosque en conservación que se identifican en el plano adjunto al presente contrato y se obliga a evitar:

1) Los incendios forestales
2) La cacería
3) El chaqueo
4) Evitar el ingreso de ganado al bosque en conservación
5) Apertura de caminos
6) Cualquier actividad extractiva de los recursos naturales del área
7) Cualquier otra actividad que ponga en riesgo el área en conservación
8) Utilizar los insumos para cualquier actividad no prevista en la cláusula quinta

La Fundación Natura Bolivia y el **Gobierno Municipal de Postrervalle** se obligan a: 1) entregar los incentivos al propietario en los plazos acordados 2) realizar el monitoreo e inspección del área en conservación una vez al año para determinar el fiel y estricto cumplimiento del presente acuerdo y la utilización de los insumos según el compromiso del **propietario** en la cláusula quinta.

Séptima. Respeto al derecho propietario

El presente acuerdo no otorga ningún derecho real ni usufructo a **la Fundación** sobre el área puesta bajo conservación. Tampoco faculta a ninguna persona particular ni institución pública ni privada, a iniciar cualquier trámite de expropiación o modificación de derechos de los legítimos propietarios del área antes mencionada.

Octava. Solución de controversias

En caso de controversias las partes se comprometen resolver las mismas de forma amigable a través de reuniones formales en un plazo no mayor a 30 días. Si finalmente no se logra llegar a algún acuerdo se procederá de acuerdo a la cláusula novena y según lo que establezca la legislación aplicada al caso.

Novena. Del incumplimiento

En caso de incumplimiento del presente acuerdo y agotada las instancias de solución de controversias establecidas en la cláusula octava, el propietario se compromete a devolver en efectivo o especies, a la OTB local y en beneficio de la comunidad, el monto total recibido por los beneficios y los insumos, según lo establecido en las actas adjuntas al presente acuerdo.

Décima. Disolución

Este contrato se disolverá por las siguientes causas:

1) Por mutuo acuerdo entre las partes intervinientes.

2) Cuando una de las partes incumpla con las condiciones establecidas dentro del presente acuerdo.

Décima primera. Aceptación

Las partes firmantes declaramos nuestra conformidad con todo lo establecido en el presente acuerdo, el cual surtirá todos los efectos de ley, comprometiéndonos a su fiel y estricto cumplimiento, por lo que firmamos en cuatro ejemplares y para un solo efecto legal. Es dado en la comunidad de [COMMUNITY] a los [DATE].

Sr. [NAME REDACTED]

CI [xxxxxxx] SC

Lic. Maria Teresa Vargas Ríos

Directora Ejecutiva

Sr. ....

Presidente de OTB

CI..

....

Gobierno Municipal de
Postrervalle

Appendix C. List of incentives available to those signing *Watershared* agreements

fundaciónnatura
BOLIVIA

COTIZACIÓN MATERIALES INSUMOS

Nro	Rubro	Material	Unidad	Precio (Bs)	OBS
1	**Ganadería**	Alambre liso de 2.2 mm (rollo de 1000 mtrs)	Rollo	515	
		Alambre de púas de alta resistencia	Rollo	279.3	
		Alambre de púas	Rollo	256	
		Alambre liso ponei	Rollo	548.8	
		Alambre liso eléctric	Rollo	485	
		Aisladores	Unidad	2.4	
		Tesadores/Catraca fabricada manualmente	Unidad	15	
		Tesadores/Catraca prefabricada rectangular	Unidad	35.8	
		Tesadores/Catraca prefabricada cuadrada	Unidad	28.2	
		Llave cácula (Llave de corte)	Unidad	37.8	
		Kit: Electrificador+Panel Solar (40 km)	Unidad	4002	
		Semilla de pasto (Kg) Decumbem	Kgr	90.5	
		Semilla de pasto (Kg) Brachiaria	Kgr	65	
		Semilla de pasto (Kg) Brachiaria, Gatum panicum	Kgr	75	
		Bebederos plasticos para ganado vacuno de 250Ltr	Unidad	452.4	
		Bebederos plasticos para ganado vacuno de 350 ltr	Unidad	600	

		Bebederos plasticos para ganado vacuno de 500 ltr	Unidad	730.8	
		Bebederos plasticos para ganado vacuno de 650 ltr	Unidad	1000	
		Grapas	Kgr	11.14	
2	Riego tecnificado	Motobomba de 2"	Unidad	2789	
		Motobomba de 3"	Unidad	3158	
		Bombas eléctricas para agua de 2HP	Unidad	1832	
		Cañería de 6 mtrs de 3/4	Metros	23.5	
		Cañería de 6 mtrs de 1/2	Metros	17.5	
		Rollo Politubo bicapa de 1/2"	Rollo	155	
		Rollo Politubo bicapa de 3/4"	Rollo	320	
		Rollo Politubo bicapa de 1"	Rollo	340	
		Rollo Politubo bicapa de 1 1/2"	Rollo	575	
		Rollo Politubo bicapa de 2"	Rollo	825	
		Tanque plasticos de almacenamiento de agua de 300 lt	Unidad	389	
		Tanque plasticos de almacenamiento de agua de 600 ltr	Unidad	591.6	
		Tanque plasticos de almacenamiento de agua de 900 ltr	Unidad	835.2	
		Tanque plasticos de almacenamiento de agua de 1200 ltr	Unidad	1078.8	
		Tanque plasticos de almacenamiento de agua de 5000 ltr	Unidad	5568	
		Cintas para riego por goteo de 1000 mtrs	Metros	1052.63	
		sello anillos de goma PVC/PE	Unidad	2.1	
		Conector linea cinta 16*Barb 16	Unidad	2.6	
		final linea cinta 16mm + anillo	Unidad	3.2	
3		Cemento de 50 Kgrs	Unidad	54	

	Sistemas de agua consumo humano	Tuberia galvanizada de 2"	Metros	330	
4	**Fruticultura**	Plantines de Cítricos (Mandarina, Naranja, limón, pomelo, kinoto y lima)	Unidad	15	
		Plantines de Durazno (Jade semitemplanero y tricocilio templanero)	Unidad	18	
		Plantines de Manzana	Unidad	18	
		Plantines de Uva/chirimoya	Unidad	18	
5	**Apicultura**	Guantes	Unidad	70	
		Casco Plástico	Unidad	80	
		Máscara protectora	Unidad	85	
		Overol	Unidad	240	
		Overol Mameluco con casco incluido	Unidad	268	
		Centrifugadora en plancha de acero inoxidable	Unidad	2050	
		Pinza con Harcas, Espatula	Unidad	105	
		Caja apicola	Unidad	670	
		Cuchillos	Unidad	120	
		Ahumador	Unidad	120	
		Cera estampada	Lamina	11	
6	**Sist. de Riego por Aspersion de acuerdo a propuesta de Valley (sin considerar linea principal ni linea secundaria)**	Aspersor Meganet 550 LH	1 Hectárea / maiz, mani	3200	**Precio total por ha (tomando en cuenta el material a requerir) Bs 11929.5**
		Conect.inicial Meganet 12mm hembra		120	
		Alojamiento Aspersor Meganet 1/2pulg x 12mm		480	
		Conect.inicial Meganet 12mm macho		120	
		Adaptador de varilla 8mm de Meganet		200	
		Estaca 1.2x8mm Meganet		400	
		Adaptador de varilla 8mm de Meganet		400	
		POLITUBO NEGRO 1 1/2" C-6		4600	
		REDUCCION BUJE DE 2" 1 1/2"		137.76	
		CODO PVC 1 1/2" R		230.64	

		Tapón hembra PVC de 1 1/2"		64	
		NIPLE EXAGONAL PVC 1 12" ROSCA		86.4	
		NIPLE EXAGONAL PVC 2" ROSCA		67.84	
		UNION PATENTE PVC 2"		384	
		LLAVE PASO PVC 1 1/2" ROSCA		1136	
		TEE PVC 2" R		239.76	
		CODO PVC 2" R		63.1	
7	**Sistema de Riego por Aspersión. Propuesta La casa del Riego**	Aspersor de 3/4	1 Hectárea / maiz, mani	498	**Precio total por ha Bs 3340**
		Llave de paso para maguera Lluvia		498	
		Tripode		360	
		Manguera Laifla		1984	
8	**Otros**	Perforador de Suelos	Unidad	11550	
		Teja colonial	mt2	37.8	

NOTA:

Los precios en la presente planilla son referenciales unitarios. Fue realizado en el mes de marzo del 2016 y en función a la cantidad final a comprar en cada acto de compensación los mismos varían de precio. Asimismo, es un listado general, de acuerdo a los requerimientos de las comunidades el quipo técnico puede solicitar un producto que no se encuentre en el presente listado

Appendix D. Chapter 3 Supplementary Information

Table 3.S1. Details of the intervention available to treatment-community landowners.

	Level 1	Level 2	Level 3
Eligible land	Forested land within 100m of a watercourse	Forested land within 100m of a watercourse	All non-agricultural and non-developed land
Principal required/ prohibited actions	Cattle removed/absent; no deforestation	Phased reduction/ removal of cattle; no deforestation	No land clearance
Value of incentives (US$/ha/year, equivalent in kind)	10, plus 100 one-off payment on joining	3	1
Timescale for action	Immediately	Deforestation immediately; reduction/removal of cattle over 3 years	Immediately
Compliance monitoring	Yearly; in person; transects to inspect for cattle presence and signs of deforestation	Yearly; in person; transects to inspect for cattle presence and signs of deforestation	Classified satellite images
Total area under conservation (as of end 2014, ha)	2206	1784	47683

Table 3.S2. Sites used in analyses.

	2010 Monitoring	2015 Monitoring*	Sites in Common (RCT)	2015 Sites in Biophysical Model†	2015 Sites in RCT Model
Intake Sites	126	141	47	123	118
Tap Sites	115	108	36	96	93
Total Sites	241	249	83	219	211
Water Systems	131	142	47	124	119
Communities	125	127	47	112	107

*In 2015 we monitored 2 water intakes in some communities (1 former intake, 1 current intake).

†Not all sites monitored in 2015 were able to be included in the 2015-only models, as some lacked a full set of predictors.

Table 3.S3. Variables monitored describing site condition or characteristics.

Disturbance	Categories	Location of monitoring
Black sulphurous mud in intake	Present/Absent	Intake
Substrate in intake	Rocky/With Sand/With Mud	Intake
Filamentous algae in intake	Present/Absent	Intake
Faeces in water or on riverbank	Present/Absent	10m Transect
Faeces in riparian forest	Present/Absent	10m Transect
Litter	None; 1-5 items; 6-10 items; 11+ items	10m Transect
Extractive activity	Present/Absent (if present, type)	10m Transect
Cattle	Present/Absent	10m Transect
Agriculture	Present/Absent	Intake
Forest cover	>80%; 50-80%; 10-50%; <10%	Intake
Forest connectivity	>75%; 50-75%; <50%	Intake
Fencing to prevent cattle access	Yes; No; No, but cattle cannot enter due to topography; Yes, but broken	Intake
Type of water source of intake	Stream; Spring; Roof rainwater collection	Intake

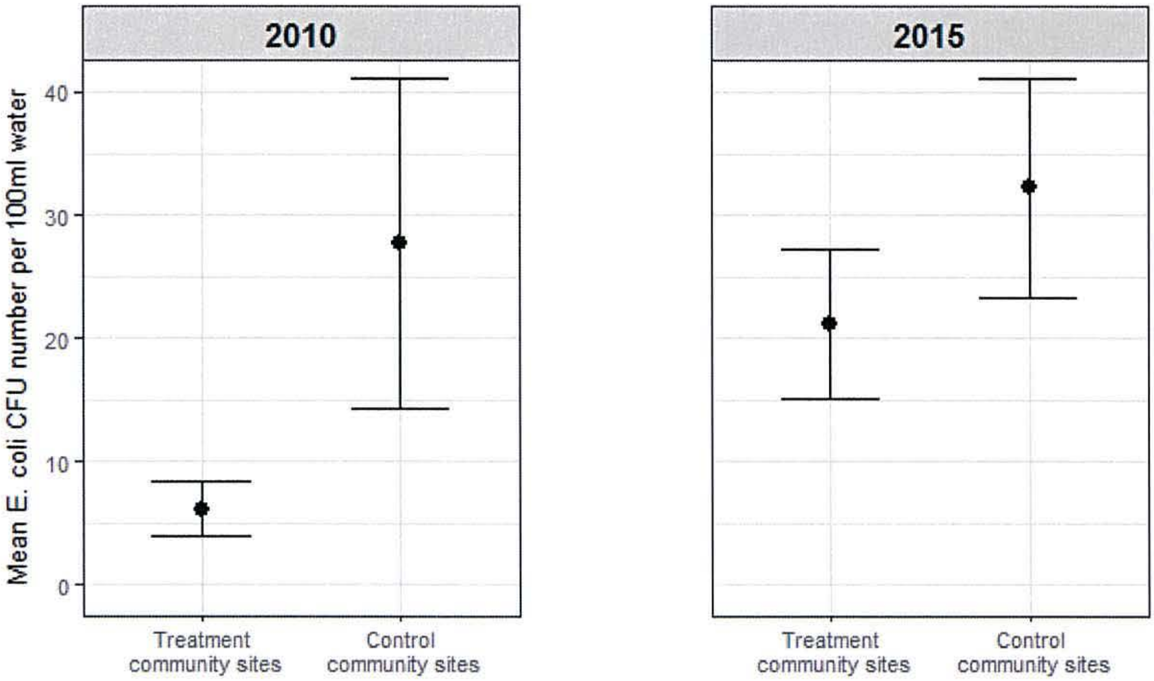


Figure 3.S1. Relative *E. coli* levels in treatment and control community sites in 2010 and 2015; shows data from all intake and tap sites. *E. coli* CFU numbers shown per 100ml sample equivalent.

Table 3.S4a. 95% confidence intervals of predictor coefficients in model representing comparison of treatment and control community site *E. coli* concentration in 2015 while accounting for levels in 2010 (site N=83; water system N=47; GLMM).

Predictor (interpretation in model)	2.5%	50%	97.5%
Intercept (log-transformed E. coli number per 5ml equivalent in control sites)	-0.81	0.062	0.94
2010 E. coli concentration (mean difference between E. coli concentration at sites in 2010 and 2015)	-0.30	0.22	0.74
RCT Status (effect of being a treatment community site)	-2.27	-1.22	-0.17
Interaction (difference in change in E. coli concentration between 2010 and 2015 between treatment and control sites)	-0.28	0.86	1.99

Table 3.S4b. Model selection table for GLMMs testing for significance of water system RCT status and 2010 *E. coli* concentration as predictors of 2015 *E. coli* concentration. K refers to the number of parameters estimated in the model.

Model	K	AIC	ωAIC
1. 1|Water System + RCT Status * 2010 E. coli concentration	4	371.70	0.372

2. 1|Water System + RCT Status + 2010 E. coli concentration	3	372.06	0.311
3. 1|Water System + RCT Status	2	373.75	0.134
4. 1|Water System + 2010 E. coli concentration	2	373.49	0.152
5. 1|Water System	1	376.66	0.031

Table 3.S5a. Model selection table for GLMMs testing for significance of biophysical variables as predictors of 2015 *E. coli* concentration (Site N=219, water system N=124). Codes are given in Table 3.2.

Model	K	AIC	ωAIC	ΔAIC
7. 1|Water System + SD + IC + ST + Tu	5	919.49	0.5963	0
6. 1|Water System + SD + IC + ST + C + Tu	6	921.33	0.2379	1.84
5. 1|Water System + SD + IC + ST + C + A + Tu	7	923.31	0.08865	3.82
4. 1|Water System + SD + IC + ST + C + A + Tu + S + pH	9	925.50	0.02957	6.01
8. 1|Water System + IC + ST + Tu	4	926.90	0.01468	7.41
3. 1|Water System + SD + IC + ST + C + A + Tu + Te + S + pH	10	927.02	0.01381	7.53
9. 1|Water System + SD + IC + Tu	4	927.28	0.01218	7.79
2. 1|Water System + SD + IC + ST + IS + C + A + Tu + Te + S + pH	12	929.29	0.00444	9.8
11. 1|Water System + SD + IC + ST	4	932.03	0.001132	12.54
1. 1|Water System + SD + IC + ST + IS + C + A + Tu + Te*S*pH	16	932.53	0.000881	13.04
10. 1|Water System + SD + ST + Tu	4	933.81	0.000463	14.32

Table 3.S5b. Model selection table for GLMMs testing for significance of all variables as predictors of 2015 *E. coli* concentration (site N=211, water system N=119).

Model	K	AIC	ωAIC	ΔAIC
16. 1|Water System + SD + IC + ST + Tu + F	7	917.34	0.3749	0
13. 1|Water System + SD + IC + ST + Tu + CA + F	8	919.09	0.1563	1.75
14. 1|Water System + SD + IC + ST + Tu + ARA + F	8	919.15	0.1517	1.81
7. 1|Water System + SD + IC + ST + Tu	5	919.49	0.1276	2.15
12. 1|Water System + SD + IC + ST + Tu + CA + ARA + F	9	920.92	0.06266	3.58
18. 1|Water System + SD + IC + ST + Tu + ARA	6	921.06	0.05843	3.72
17. 1|Water System + SD + IC + ST + Tu + CA	6	921.49	0.04693	4.15
15. 1|Water System + SD + IC + ST + Tu + CA + ARA	7	923.06	0.02149	5.72

Table 3.S5c. 95% confidence intervals of predictor coefficients in most likely model predicting *E. coli* concentrations in 2015 (model 16).

Predictor (interpretation in model)	2.5%	50%	97.5%
Intercept (log-transformed E. coli concentration)	1.44	2.30	3.17
Sediment (no disturbance of sediment)	-2.21	-1.43	-0.65
Site type (tap compared with intake)	-1.10	-0.67	-0.23
Intake category (spring compared with stream)	-1.84	-1.25	-0.67
Turbidity (per 100 FAU)	0.39	1.13	1.87
Feces presence (in forest compared with absent)	-0.50	0.47	1.43
Feces presence (in water compared with absent)	0.25	1.91	3.57

Table 3.S6. Model selection table for GLMMs of most likely model predicting 2015 *E. coli* concentration with and comparable model including water system RCT status (site N=211; water system N=119).

Model	K	AIC	ωAIC	ΔAIC
20. 1|Water System + SD + IC + ST + Tu + F + RCT	8	883.29	0.753	0
19. 1|Water System + SD + IC + ST + Tu + F	7	885.51	0.247	2.22

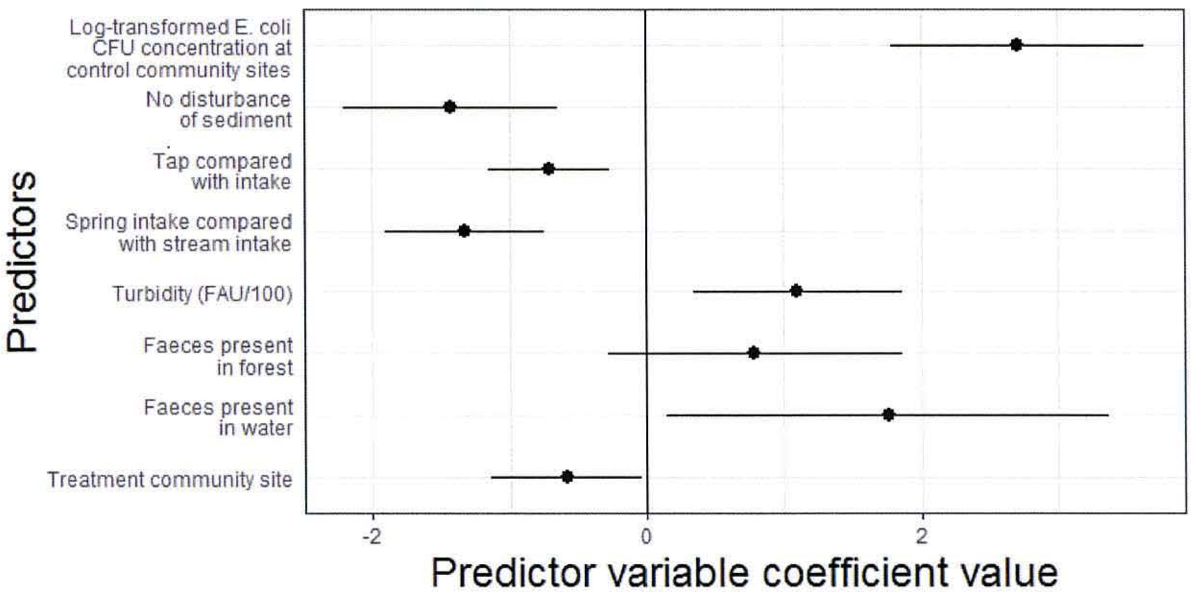


Figure 3.S2. Model coefficients of site features which are predictors of 2015 *E. coli* concentration in the most likely GLMM with RCT status included (model 20). Error bars show 95% CIs.

Appendix E. The use of Coliscan Easygel to conduct water monitoring in low-resource contexts.

As described in chapters 3, 4, and 6 of this thesis, we used the Coliscan Easygel method (Micrology Labs, Goshen, IN, USA) combined with a portable incubator and icebox in order to enumerate *E. coli* and non-*E. coli* coliform concentrations in community water supplies. This decision was taken by Edwin Pynegar, Nigel Asquith, and *Natura* technicians based upon available budget and the logistics of conducting water quality monitoring for relatively low budgets in a very low-resource setting (such as in many of the communities of the *ANMI Río Grande – Valles Cruceños* protected area). We therefore include this appendix as a reference for other evaluators and practitioners interested in how protocols may be modified for use in similar low-resource contexts.

We established that the method selected had to be able to achieve the following:

1) Samples had to be processed within 4 to 6 hours if kept on ice, or within 1 hour if not, and enough time had to be left for plates to solidify (if required) before moving them.

2) Samples had to able to be incubated at an appropriate temperature.

3) The method selected had to be able to monitor on average 3 to 4 communities per day, representing 6 to 8 sites and thus 24 to 32 samples (as we planned to take 4 samples per site in 2015). The method would ideally allow for more than this in cases where it was possible to monitor 5 or more communities within a day, or in which a community had more than 2 sites (1 intake, 1 tap) associated with it.

4) Consumables had to be importable or otherwise straightforwardly obtainable in Bolivia, ideally in the Santa Cruz Valleys region but otherwise in the city of Santa Cruz de la Sierra.

5) The selected method had to be functional in locations without reliable mains electricity.

6) Waste had to be non-toxic and disposable within the context.

7) The selected method had to be affordable.

These criteria determined possible methods for the context:

1) implies that a transportable method is required. Logistics would not allow return to Vallegrande, where a water quality testing laboratory already exists, as the requirement to be able to deal with samples within 6 hours would mean that very few sites could be sampled before needing to return.

Some communities are themselves 6 hours' drive (or more, depending upon road conditions) away from Vallegrande. Transportability, plus the requirement for an incubator [see 2)] result in 5).

Some methods, such as those based on the H_2S test for presence of *E. coli*, do not require incubation at a constant temperature (Bain *et al.* 2012). However most available tests, including the vast majority of those based on most-probable-number approaches as well as membrane-filtration-based enumeration techniques, do (Bain *et al.* 2012). Hence as we required more information than presence/absence, an incubator would be required which would thus have to be portable [based on 1)]. The assumption by *Natura* technicians that ambient temperature would be close enough to the 22-25^{0}C assumed "room temperature" specified by Micrology Labs was not borne out.

A number of integrated portable apparatuses are available manufactured specifically for this purpose, such as the Potatest 2 kit (Palintest Ltd, Gateshead, UK). However these kits are costly (approximately $3000 as of 2014, before import and customs costs of shipping to Bolivia), causing problems with criterion 7). Additionally, we expected that such products would still be inadequate as the amount of space available in the in-built incubators – there are two in the Palintest 2, with the 37^{0}C one intended for identification of total coliforms and the 44^{0}C for identification of thermotolerant coliforms (mostly *E. coli*). Taking multiple samples per site, for example 3 or 4, would quickly result in the incubators (20 each) to become full up. Thus such apparatuses fail on point 3). Additionally, the Petri dishes used in such situations are made of aluminium, reusable, specifically sized, and are required to be used as specified, as otherwise heat distribution within the incubators themselves would be uneven. This means they have to be cleaned in the field, and this not only takes a substantial amount of time but also specifically requires the obtaining and use of methanol, whose partial combustion produces formaldehyde which sterilises the dishes. However this not only appeared impractical in a high-speed field context, but also obtaining methanol in Santa Cruz de la Sierra, never mind in the Santa Cruz Valleys, was by no means guaranteed. Finally it was unclear how feasible it would be to make up medium in a sterile fashion for bacterial cultivation, and how time-consuming this would be. It was also unclear how used medium would adequately be disposed of.

Hence we dismissed such filter-based protocols as impractical, despite them meeting the World Health Organization's 100ml sample size criterion for *E. coli* monitoring. Similarly, we dismissed most-probable-number approaches as being insufficiently informative and complex as well as often requiring a specific non-mobile incubator. Products which the literature described as being in development which avoided these problems were not available at the time (R. Bain, personal communication).

We thus returned to Coliscan Easygel as our selected method, combined with a portable Darwin Chambers NQ-28 incubator (Darwin Chambers Co., St Louis, MO, USA). Coliscan Easygel samples were to be incubated at 35-37^{o}C in that incubator, which was able to comfortably contain well over 100 Petri dishes at any time. The incubator is manufactured as intended to be portable, with both a mains plug and 12V car power supply available. A model including rechargeable Li-ion batteries is also available although substantially more expensive. Consequently we instead purchased a car battery and used it in locations without mains electricity when the vehicle was switched off. At a maximum of 20-25 watt power consumption there was no issue with the battery going flat between trips to the field. Cost of the car battery, pincers, and battery charger cost approximately 50% of the equivalent Li-ion add-on to the incubator (Figure E.1).

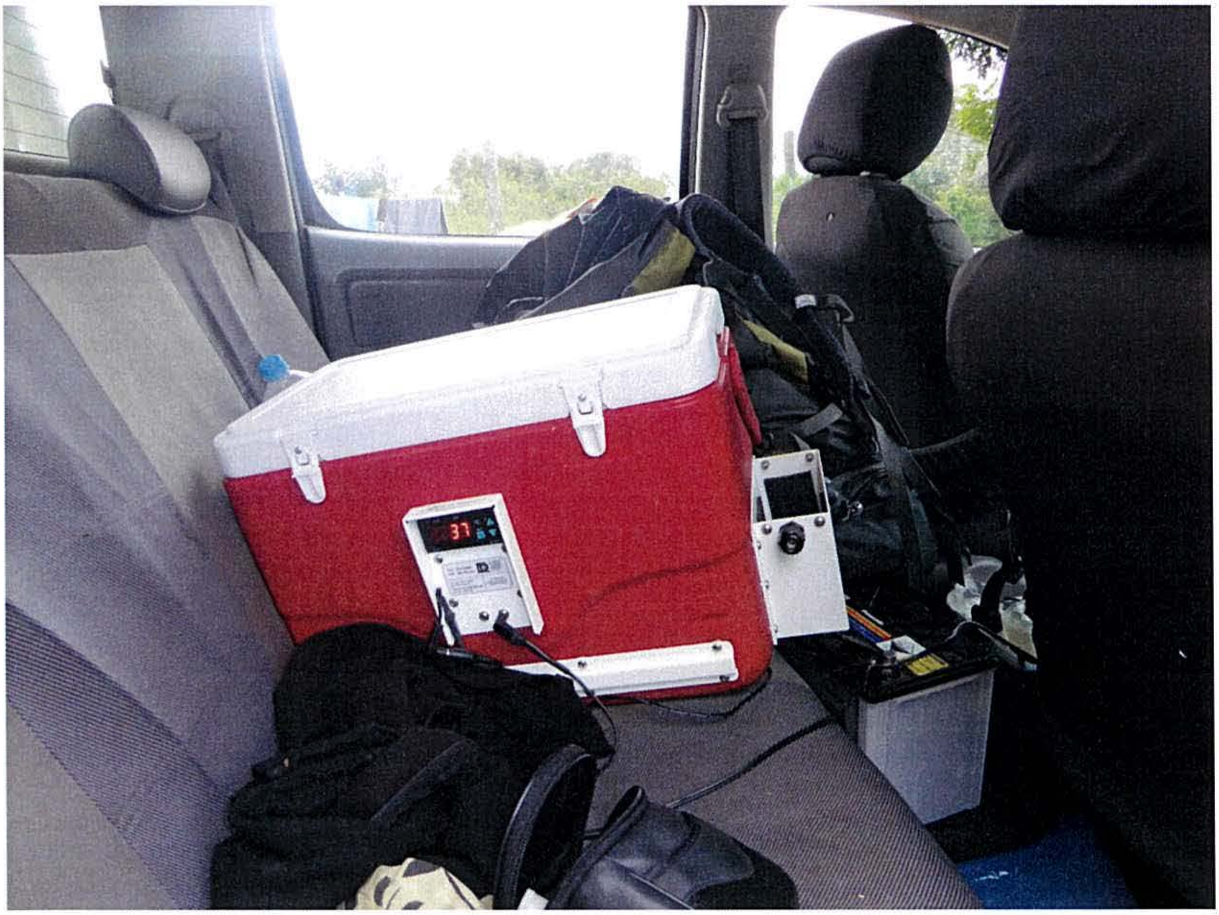


Figure E.1. NQ-28 portable incubator being powered by car battery in field location.

Samples were placed on ice in a simple polystyrene box. Obtaining ice for this could be a challenge in the field and we ended up either buying ice from markets in local towns or having to ask local people to freeze ice bags in their freezers overnight. In locations with no mains electricity (and hence obviously no freezers) we had to bring enough ice with us such that it would remain frozen for the time period of the trip.

The 4-hour limit (maximum 6 hours) for plating out of samples has implications for the organisation of the day's work. Often we would sample in the morning, plate out samples in the early afternoon, continue to sample, and then plate out again in the evening. Alternatively we would travel to the most distant site (to which we might arrive only at 12 noon or 1p.m.) and then sample on the return, requiring only one round of plating out per day in the early evening. We would do this plating out wherever possible – in communities, on the road (Figure E.2) or at the *Natura* office in Vallegrande if possible. This logistical calculation requires flexibility on the part of technicians collecting data depending upon the times and distances involved and the spatial layout of the sites to be monitored, as well as other work that required completion on that day. We counted CFU numbers as closely to 24 or 36 hours as possible after plating out in 2015 and 2016 respectively (from observation we learned that a number of colonies did not emerge until after 24 hours from plating out).

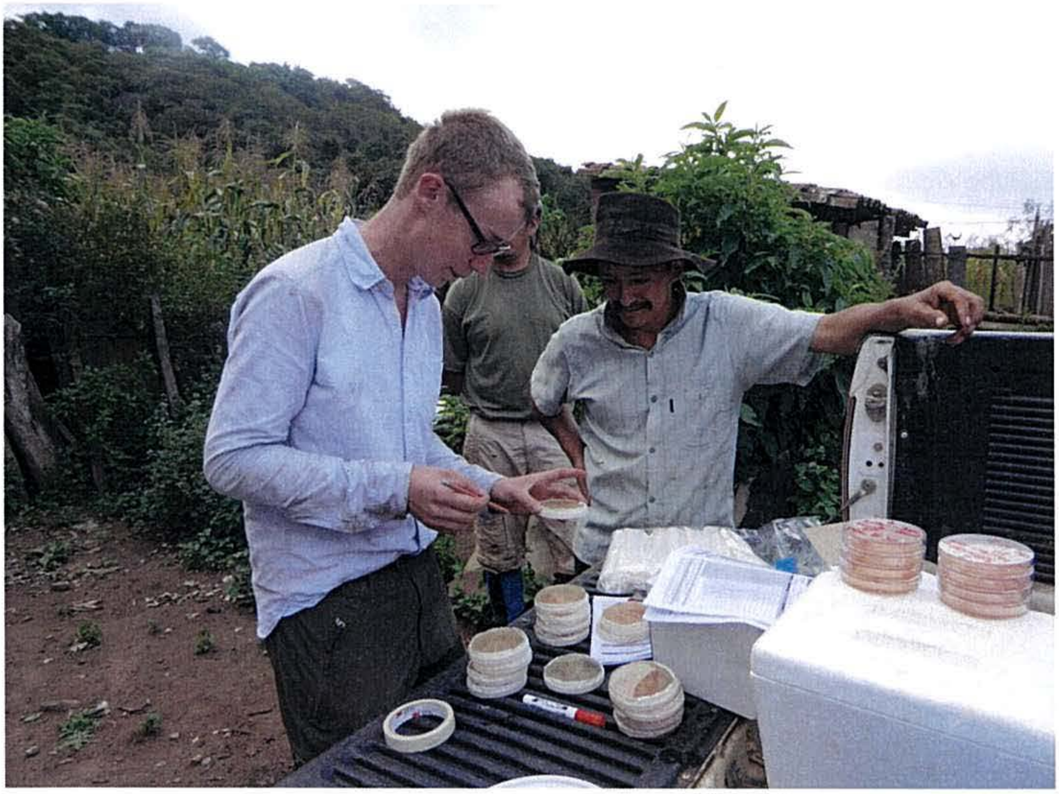


Figure E.2. Plating out and enumeration of *E. coli* and other coliforms in the field.

Criterion 6) relates to disposal of waste. While *E. coli* are mostly not pathogenic, some strains can be, and thus disposal requires careful consideration. Previously, *Natura* technicians had opened the used Petri dishes and added 5ml chlorine bleach to each one. I did not consider this an acceptable solution as I found the vapour was irritating to the respiratory system as well as the fact that sterilization of several hundred Petri dishes would result in the discharge of a number of litres of bleach into the local environment. Autoclaving the used Petri dishes was similarly out of the question. As a result we purchased a large saucepan and obtained a hotplate, and sterilised used Petri dishes via boiling for at least one hour.

With this, we show how water monitoring over a large scale can be rapidly and relatively cheaply conducted, even in low-resource contexts such as the *ANMI Río Grande*, with the critical caveat that the method selected did not meet the WHO's 100ml standard for water quality monitoring.

Appendix F. Chapter 4 Supplementary Information

Table 4.S1. Predictors and response variables used in modelling process, including coding and reclassification if required. Predictors with asterisks are those which were added in 2016 to those monitored previously in 2015 (see chapter 3).

Code	Description	Responses/Options	Type of variable	Measurement method/site
EC	*E. coli* colony forming units in 15ml sample (response variable)	Integer	Number	Coliscan Easygel
SYS	Water system – unique code	Numerical code	Factor	N/A
ST	Site type (intake/tap site)	"A" = intake; "B" = tap	Factor	N/A
ICN	Intake category (new classification)	"1" = spring; "2" = stream or river; "3" = roof collection; "4" = pool	Factor	Observation
IC	Intake category (previous classification)	"1" = spring or pool; "2" = stream or river	Factor	Observation
IWS*	Improved water source (see WHO/UNICEF Joint Monitoring Program 2017)	"No" = unimproved source; "Si" = improved source	Factor	Observation based on improved water source definition
F1	Faeces presence upstream/uphill of intake	"0" = absence; "1" = presence in forest; "2" = presence in water or riverbank	Factor	10m transect upstream/uphill of intake
F2	Faeces presence upstream/uphill of intake	"0" = absence; "1" = presence in forest; "2" = presence in water or riverbank	Ordered factor (0 – 2)	10m transect upstream/uphill of intake
Tu	Turbidity (formazin attenuation units/100)	Integer/100	Number	Intake (+ tap)
C	Cattle presence at/around intake	"No" = absence; "Si" = presence	Factor	Intake
A	Agriculture presence at/around intake	"No" = absence; "Si" = presence	Factor	Intake
CA	Cattle access to intake	"No" = no access; "Si" = access	Factor	Intake
UPC*	Cattle presence in catchment or hydrological recharge zone of intake	"No" = absence; "Si" = presence	Factor	Local knowledge
UPA*	Agriculture presence in catchment or	"No" = absence; "Si" = presence	Factor	Local knowledge

	hydrological recharge zone of intake			
ARA	Water source located within level 1 *Watershared* area	"0" = not in level 1 area; "1" = in level 1 area	Factor	Local knowledge and *Natura* databases

Table 4.S2. Predictor variables (and response as episodes of diarrhoeal disease) used in subsequent modelling process, including coding and reclassification if required. The bottom four variables refer to characteristics tested in the later model selection processes (results in tables S5b and S5c).

Code	Description	Responses/Options	Type of variable
TD	Episodes of diarrhoeal disease in the past year (response variable)	Integer	Number
HH	Household – unique code	Code	Factor
OTB	Community – unique code	Code	Factor
A	Age of child	Integer	Number
WS	Type of water source supplying household	"1" = spring or pool; "2" = stream or river; "3" = other	Factor
MW	Months without water	Integer (from 0 to 12)	Number
WT	Water treatment actions (this refers to what the household themselves do to treat their water before drinking)	"0" = no treatment; "1" = boiling/chlorination in households; "2" = removal of sediment in households; "3" = filtration in water supply system (no further treatment in households)	Factor
TI	Type of water intake	"0" = no infrastructure at all; "1" = concrete construction; "2" = 'rustic'/mud construction	Factor
CE1	Level of cattle exclusion (4 categories)	"1" = no access to cattle; "2" = no cattle in intake but present around it; "3" = cattle can drink in intake but not enter; "4" = full access to cattle	Number
CE2	Level of cattle exclusion (4 categories)	"1" = no access to cattle; "2" = no cattle in intake but present around it; "3" = cattle can drink in intake but not enter; "4" = full access to cattle	Ordered factor (1 – 4)
CE3	Level of cattle exclusion (3 categories)	"1" = no access to cattle; "2" = partial access of cattle to intake site; "3" = full access to cattle	Number
CE4	Level of cattle exclusion (3 categories)	"1" = no access to cattle; "2" = partial access of cattle to intake site; "3" = full access to cattle	Ordered factor (1 – 3)
BOL	Monthly financial contribution to water committee	"0" = no contribution; "1" = contribution	Factor

INF	Water infrastructure available to households	"0" = no piped water from intake; "1" = piped into community; "2" = piped into house	Factor
RCT	RCT status (indicates whether *Watershared* had been offered in the community or not)	"1" = treatment community; "0" = control community	Factor
ARA	Water source located within level 1 *Watershared* area	"0" = not in level 1 area; "1" = in level 1 area	Factor
QN	Perception of change in water quantity in past 5 years	"0" = same; "1" = increased; "2" = decreased	Factor
QL	Perception of change in water quality in past 5 years	"0" = same; "1" = improved; "2" = declined	Factor

Table 4.S3. Model selection to test effects of land use context and water supply infrastructure on measured *E. coli* concentrations in community water supply systems (site N=84; water system N=44).

Code	Structure (predictor variables)	K	AIC	ΔAIC	ωAIC
1.9	(1|SYS)+ICN+Tu+F2+CA+UPC+UPA	9	478.224	0	0.142
1.12	(1|SYS)+ICN+Tu+F2+UPC+UPA	8	478.268	0.044	0.139
1.15	(1|SYS)+ICN+Tu+UPC+UPA	6	478.514	0.290	0.123
1.7	(1|SYS)+ST+ICN+Tu+F2+CA+UPC+UPA	10	478.586	0.362	0.119
1.14	(1|SYS)+ICN+UPC+UPA	5	478.696	0.472	0.112
1.10	(1|SYS)+ICN+F2+CA+UPC+UPA	8	479.598	1.374	0.072
1.16	(1|SYS)+ICN+CA+UPC+UPA	6	479.764	1.540	0.066
1.20	(1|SYS)+ICN+UPC+UPA+ARA	6	479.934	1.710	0.060
1.11	(1|SYS)+ICN+Tu+CA+UPC+UPA	7	480.024	1.800	0.058
1.6	(1|SYS)+ST+ICN+IWS+Tu+F2+CA+UPC+UPA	11	480.562	2.338	0.044
1.3	(1|SYS)+ST+ICN+IWS+Tu+F1+C+A+CA+UPC+UPA	13	481.872	3.648	0.023
1.5	(1|SYS)+ST+ICN+IWS+Tu+F2+C+A+CA+UPC+UPA	13	481.872	3.648	0.023
1.18	(1|SYS)+ICN+UPA	4	482.808	4.584	0.014
1.2	(1|SYS)+ST+IC+IWS+Tu+F1+C+A+CA+UPC+UPA	12	487.626	9.402	0.001
1.4	(1|SYS)+ST+IC+IWS+Tu+F2+C+A+CA+UPC+UPA	12	487.626	9.402	0.001
1.19	(1|SYS)+ICN+UPC	4	489.150	10.926	0.001
1.1	(1|SYS)+ST+IC+Tu+F1	6	491.270	13.046	0
1.8	(1|SYS)+ST+ICN+Tu+F2+C+A+CA	10	491.616	13.392	0
1.13	(1|SYS)+Tu+F2+CA+UPC+UPA	7	493.258	15.034	0
1.17	(1|SYS)+UPC+UPA	3	495.678	17.454	0

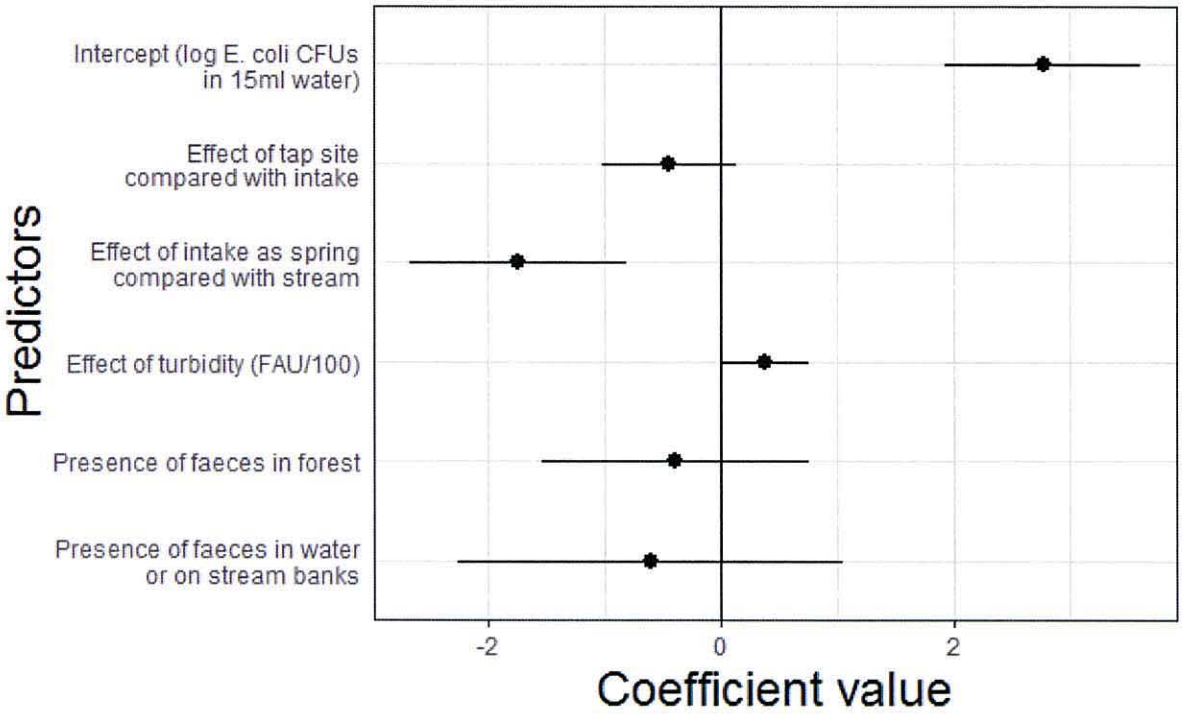


Figure 4.S1a. Coefficient plot (with 95% CI values) testing predictive model of *E. coli* concentration from 2015 data; predictions on measured *E. coli* concentration in water supplies in 2016 (model 1.1).

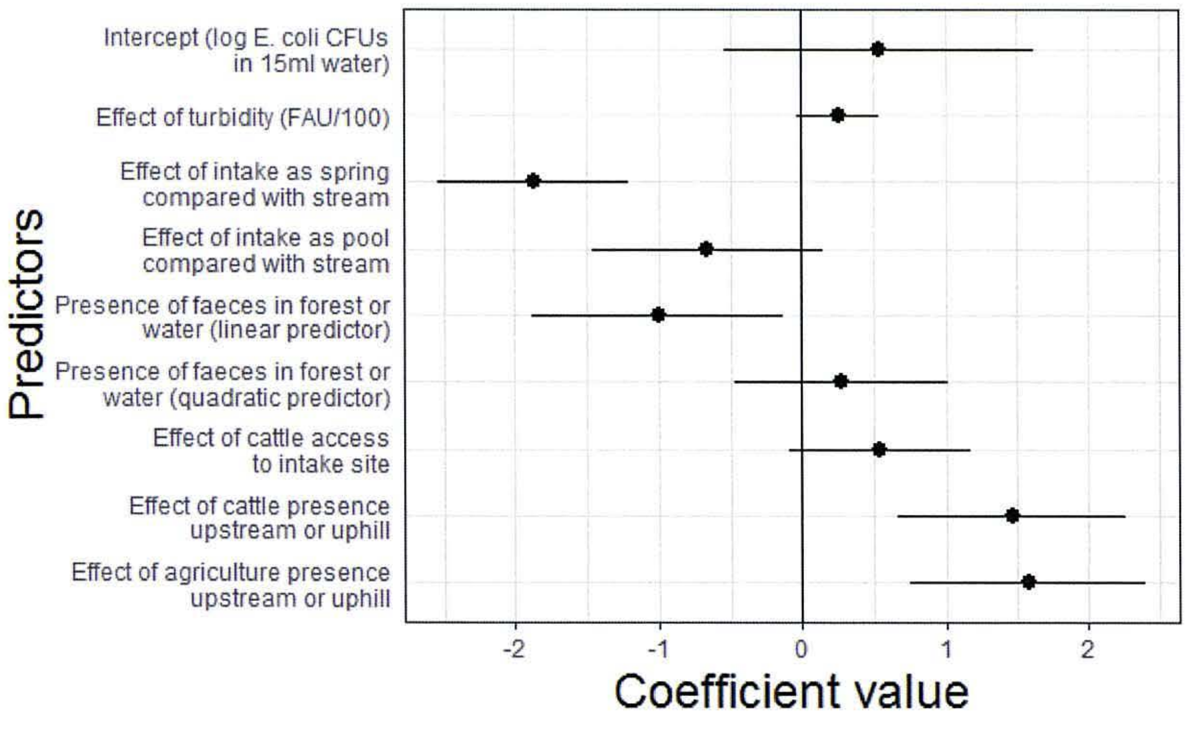


Figure 4.S1b. Coefficient plot (with 95% CI values) showing effects of land use context and water supply infrastructure on measured *E. coli* concentration in 44 water systems (model 1.9; model with lowest AIC value).

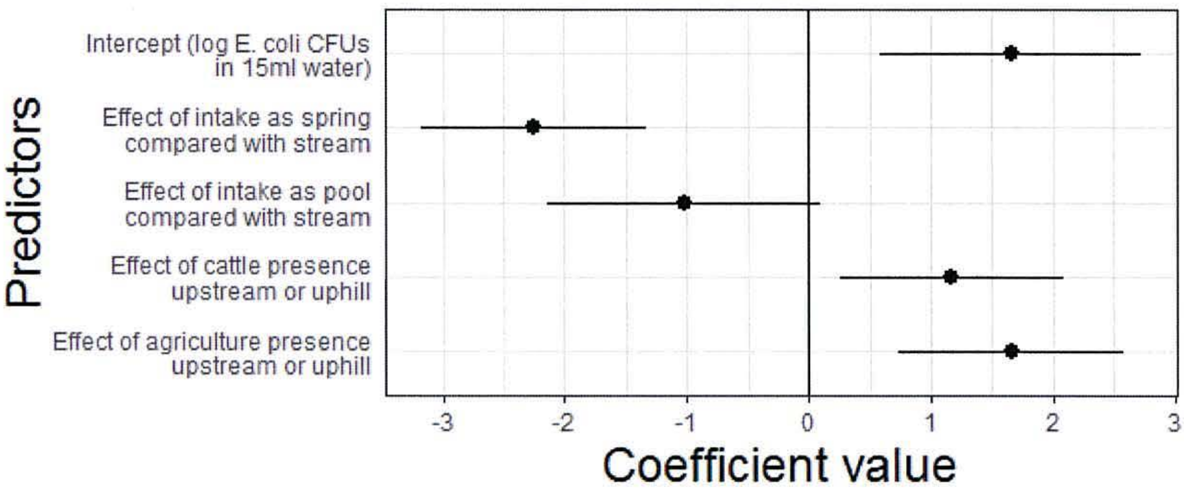


Figure 4.S1c. Coefficient plot (with 95% confidence intervals) showing effects of land use context and water supply infrastructure on measured *E. coli* concentration in 44 water systems (model 1.14; simplest good model).

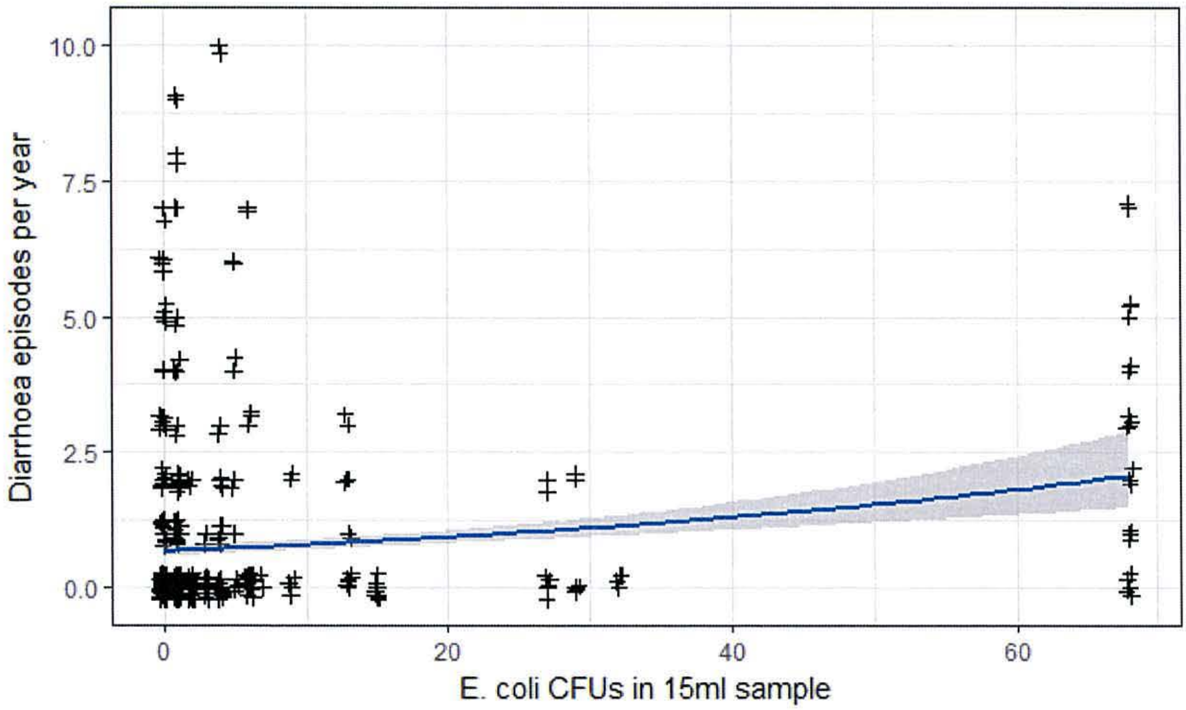


Figure 4.S2a. Relationship between *E. coli* colony forming unit concentration in household water supply and number of episodes of diarrhoeal disease.

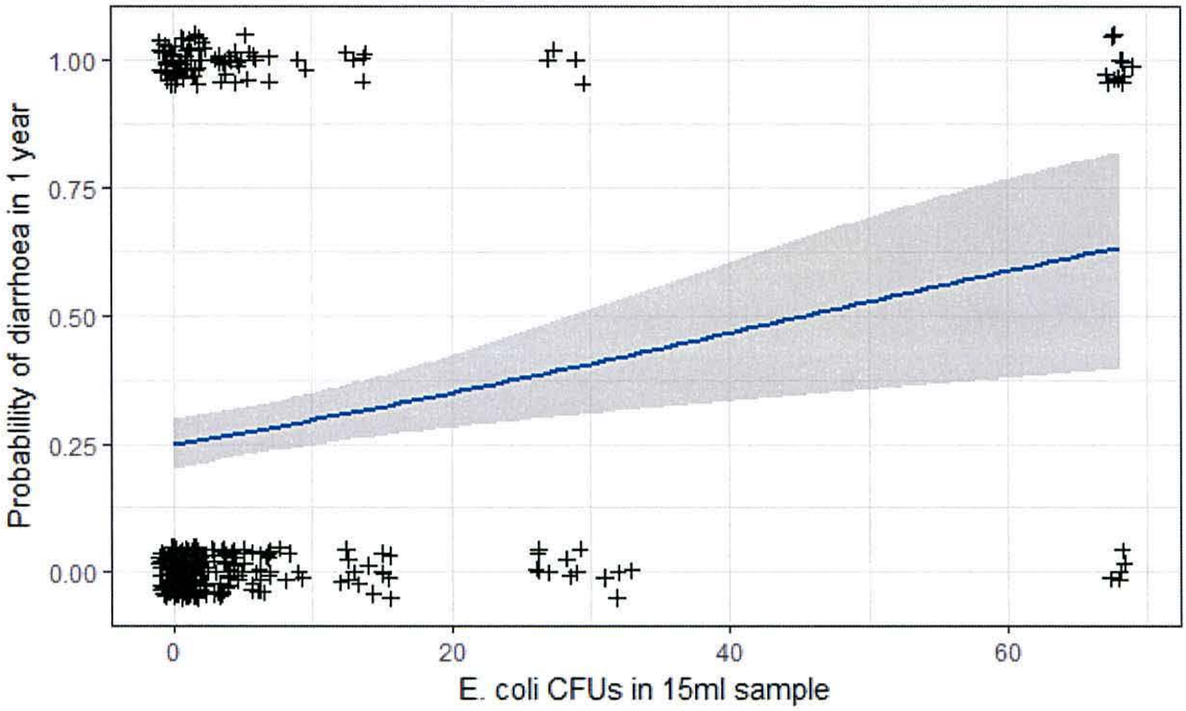


Figure 4.S2b. Relationship between *E. coli* colony forming unit concentration in household water supply and probability of any episode of diarrhoeal disease.

Table 4.S4a. Coefficients (Wald estimator-based 95% CI) of model testing relationship between number of episodes of diarrhoeal disease and *E. coli* concentration in household water supply.

	2.5%	50%	97.5%
Intercept	-1.290	-0.730	-0.170
Age	-0.172	-0.134	-0.0961
E. coli number in 15ml	0.00411	0.0232	0.0423

Table 4.S4b. Coefficients (Wald estimator-based 95% CI) of model testing relationship between probability of any episode of diarrhoeal disease and *E. coli* concentration in household water supply.

	2.5%	50%	97.5%
Intercept	-1.037	-0.0743	0.888
Age	-0.411	-0.260	-0.108
E. coli number in 15ml	0.00271	0.0465	0.0902

Table 4.S4c. Coefficients (Wald estimator-based 95% CI) of model testing relationship between number of episodes of diarrhoeal disease and *E. coli* concentration in school water supply.

	2.5%	50%	97.5%
Intercept	-1.001	-0.470	0.0614
Age	-0.176	-0.138	-0.101
E. coli number in 15ml	-0.0905	-0.0113	0.0678

Table 4.S4d. Coefficients (Wald estimator-based 95% CI) of model testing relationship between probability of any episode of diarrhoeal disease and *E. coli* concentration in school water supply.

	2.5%	50%	97.5%
Intercept	-0.680	0.277	1.234
Age	-0.356	-0.240	-0.123
E. coli number in 15ml	-0.130	-0.00919	0.112

Table 4.S5a. Model selection process to test effects of household survey responses relating to water supply infrastructure and institutions and conservation actions on reported diarrhoeal disease levels in communities (N=1012, household N=552).

Code	Structure (predictor variables)	K	AIC	ΔAIC	ωAIC
2.9	(1|HH)+A+MW+WT	6	1709.765	0	0.159
2.7	(1|HH)+A+MW+WT+CE3	7	1710.106	0.341	0.134
2.5	(1|HH)+A+MW+WT+CE1	7	1710.577	0.812	0.106
2.13	(1|HH)+A+WT	5	1710.660	0.895	0.102
2.11	(1|HH)+A+WT+CE3	6	1711.046	1.281	0.084
2.16	(1|HH)+A+A^2+MW+WT	7	1711.571	1.806	0.064
2.8	(1|HH)+A+MW+WT+CE4	8	1711.649	1.884	0.062
2.12	(1|HH)+A+MW	3	1711.688	1.923	0.061
2.10	(1|HH)+A+MW+CE3	4	1711.750	1.985	0.059
2.4	(1|HH)+A+MW+WT+CE3+INF	9	1711.940	2.175	0.054
2.14	(1|HH)+A	2	1712.519	2.754	0.040
2.3	(1|HH)+A+WS+MW+WT+CE3	9	1712.705	2.940	0.036
2.6	(1|HH)+A+MW+WT+CE2	9	1713.626	3.861	0.023
2.2	(1|HH)+A+WS+MW+WT+CE3+INF	11	1714.535	4.770	0.015
2.15	(1|HH)+A^2+MW+WT	6	1718.989	9.224	0.002
2.1	(1|HH)+A+WS+MW+WT+TI+CE3+BOL+INF	14	1719.048	9.283	0.002

Table 4.S5b. Model selection process to test effects of level 1 *Watershared* area and RCT status on reported diarrhoeal disease levels in communities.

Code	Structure (predictor variables)	K	AIC	ΔAIC	ωAIC
2.9	(1|HH)+A+MW+WT	6	1709.765	0	0.573
2.18	(1|HH)+A+MW+WT+RCT	7	1711.715	1.950	0.216
2.17	(1|HH)+A+MW+WT+ARA	7	1711.764	1.999	0.211

Table 4.S5c. Model selection process to test effects of perceived water quality and quantity changes over 5 years on reported diarrhoeal disease levels in communities.

Code	Structure (predictor variables)	K	AIC	ΔAIC	ωAIC
2.20	(1|HH)+A+MW+WT+QL	8	1706.965	0	0.678
2.9	(1|HH)+A+MW+WT	6	1709.765	2.800	0.167
2.19	(1|HH)+A+MW+WT+QL+QN	10	1710.584	3.619	0.111
2.21	(1|HH)+A+MW+WT+QN	8	1712.466	5.501	0.043

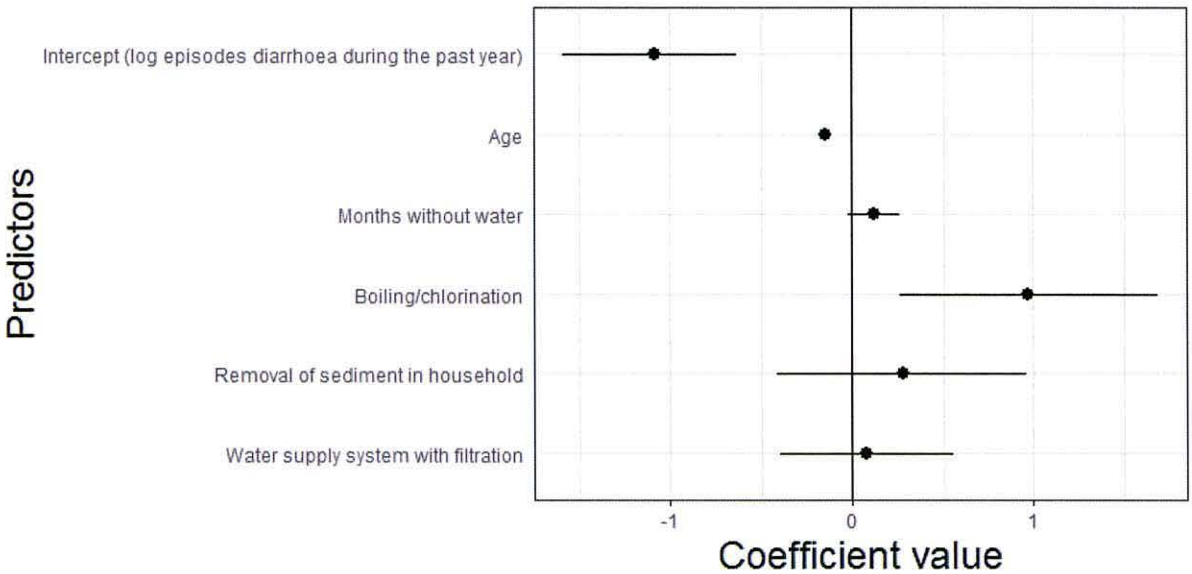


Figure 4.S3a. Coefficient plot (with 95% CI values) showing effects of household survey responses for model with lowest AIC value (model 2.9).

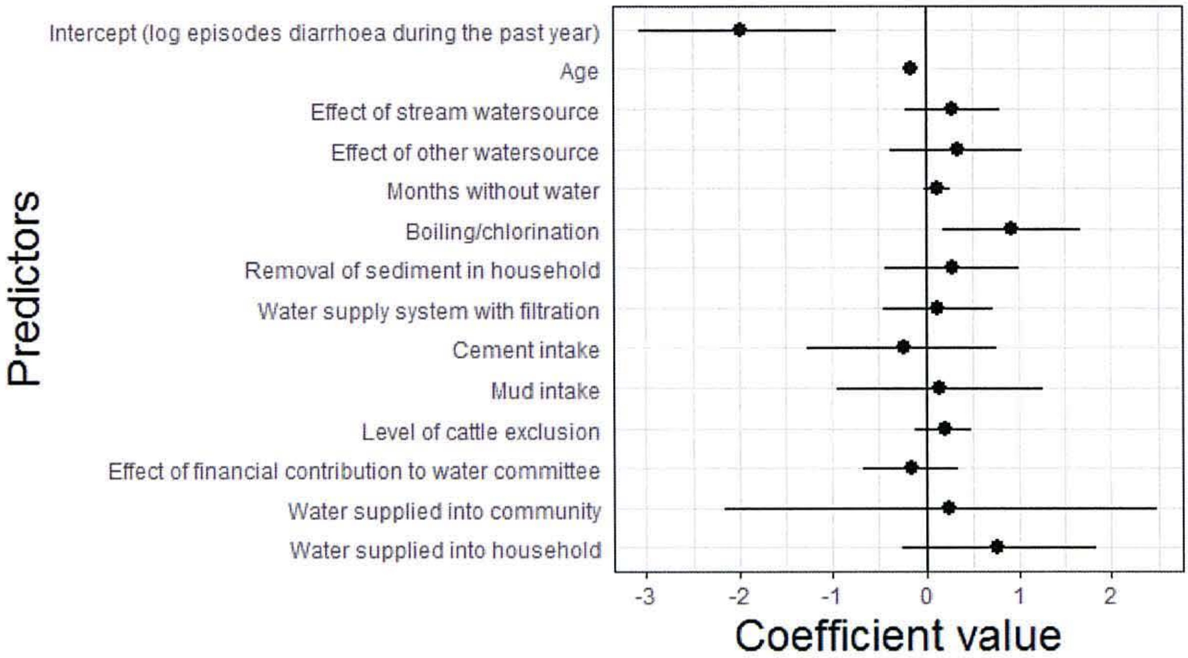


Figure 4.S3b. Coefficient plot (with 95% CI values) for global model including all initially tested potential predictors derived from survey responses (model 2.1).

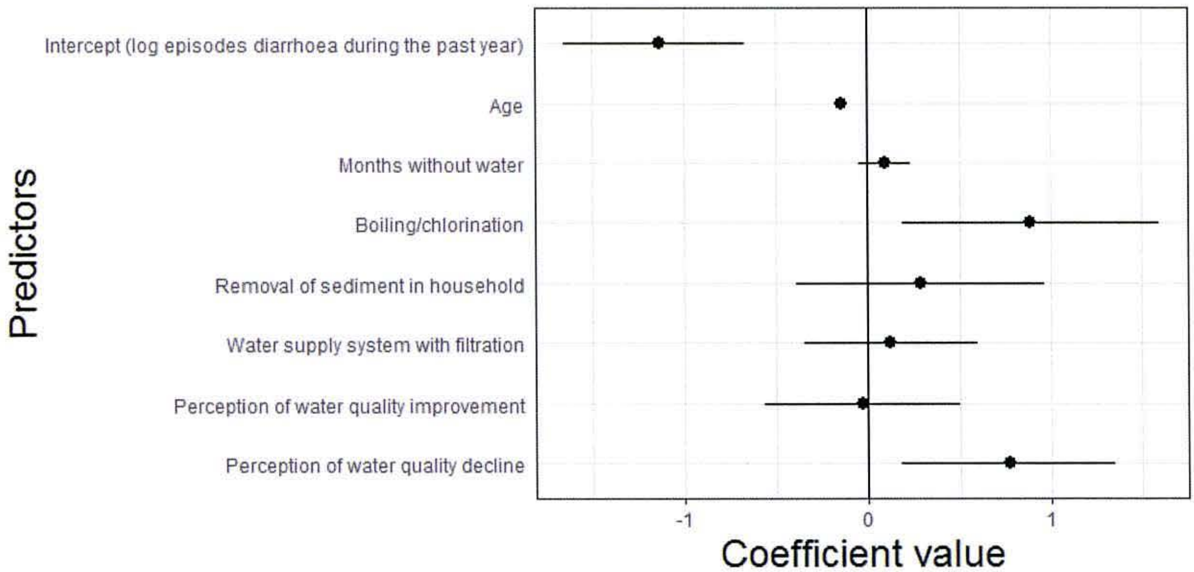


Figure 4.S3c. Coefficient plot (with 95% CI values) for lowest AIC model testing predictors derived from household survey responses, including perception of water quality and quantity changes over 5 years (model 2.20).

Document 4.S1. Example results report delivered to communities.

fundaciónnatura
BOLIVIA

Informe de la calidad biofísica del agua de la comunidad de Cuevas

Origen Muestra y coordenadas	Toma de agua (421782 X 7987741 Y); Grifo del colegio (421803 X 7988179 Y); Municipio de Samaipata
Responsable	Lic. Edwin Pynegar
Fecha y hora de Muestreo	08 – Abril – 2016; 13:00
Caracterización Muestra	Agua de la toma y de grifo
Muestreo Fisicoquímico y equipo	Puntual, Equipo portátil línea HACH
Muestreo Microbiológico y equipo	Cultivo in Vitro, "kit" específico Coliscan® Easygel®.

Resultado de Análisis

Tabla 1. Características bacteriológicas, fisicoquímicas y sus límites permisibles de los puntos muestreados.

VARIABLES	LIMITES PERMISIBLES [1 2]	RESULTADOS			
		Toma	Calificación	Grifo	Calificación
Características fisicoquímicas					
Temperatura (°C)	----------	23,1	*Buena*	24	*Buena*
Oxígeno en el agua (mg/l)	----------	7,01	*Buena*	7,51	*Buena*
pH	*6,5 – 9,0*	8,48	*Buena*	7,96	*Buena*
Conductividad (µS/cm)	*1500 µS/cm*	144,5	*Buena*	150	*Buena*
Salinidad (‰)	-----------	0,07	*Buena*	0,07	*Buena*
Características bacteriológicas					
Escherichia coli	*0,0 UFC/15 ml*	44	*Mala*	10	*Mala*
Otros coliformes	*0,0 UFC/15 ml*	600	*Mala*	55	*Mala*

1 2 Mediante Reglamento Nacional para el Control de la Calidad del Agua para Consumo Humano; (**CAPÍTULO III**) del Control de la Calidad del Agua para Consumo Humano, **Articulo 18** Parámetros de Control Mínimo. Los parámetros de Control Mínimo de la calidad del agua para consumo humano que se presentan en la Tabla

Conceptos de apoyo

- **Temperatura:** Temperaturas elevadas implican aceleración de la putrefacción, esto aumenta la demanda biológica de oxigeno (DBO) y disminuye el oxígeno disuelto.
- **Oxígeno Disuelto:** Niveles bajos o ausencia de oxígeno en el agua, puede indicar contaminación elevada, condiciones sépticas de materia orgánica o una actividad bacteriana.

- **pH:** Los niveles normales de ácidos y bases (pH) tienen la capacidad de neutralizar variaciones en el agua provocadas por la adición de uno u otro componente.
- **Conductividad:** Mide la capacidad del agua para transportar la corriente eléctrica (por sales disueltas) presentes en el agua.
- **Salinidad:** Los niveles de concentración de sales minerales afecta procesos y propiedades físicas del agua: densidad, viscosidad, tensión superficial, presión osmótica, punto de fusión, punto de ebullición y solubilidad de gases.
- **Escherichia coli (*E. coli*):** Microorganismos que se transmiten por medio de excrementos humanos y animales, que son depositados en el agua y se transmiten comúnmente por la ingestión o el contacto con agua contaminada. Tienen la potencial de ser dañinas a la salud humana; por esto, no deberían existir en agua destinada a consumo humano.
- **Otros coliformes:** Se encuentran en el intestino del hombre y de los animales, pero también en otros ambientes: agua, suelo, plantas, cáscara de huevo, etc. No son directamente dañinas pero un número elevado indica agua sucia.

Conclusiones

El agua de la comunidad Cuevas está:

1. Según los parámetros fisicoquímicos el agua se encuentra dentro de los límites.
2. Según los parámetros microbiológicos existe una **CONTAMINACIÓN** del agua por la presencia de *E. coli* y otros coliformes.
3. El agua **no** cumple con las normas para consumo humano por la presencia de *E. coli* en ella, para que pueda ser consumida seguramente se debe desinfectarse antes de su consumo.

Recomendaciones

Recomendaciones generales a nivel comunitario

- **Una vigilancia, control y limpieza periódica** de la estructura de la toma y del tanque de agua.
- **Limpieza regular** de la infraestructura de filtración.
- **Determinar lugares potenciales alternativos**, para tomas de agua (temporales o fijas).
- **Construcción de una galería filtrante** o una cámara de filtración entre la toma y los grifos.
- **Encerramiento de la toma y la cuenca arriba** para protegerla del ganado.
- **Los esquemas de Acuerdos Recíprocos por Agua (ARA)** propuestos por la Fundación Natura pueden apoyar en la conservación y preservación de la calidad de sus fuentes de agua.
- **Cosecha de agua de lluvia:** se recibe el agua de la lluvia que cae en los techos por medio de un tubo en un contenedor (turril) para así poder utilizar luego.

Recomendaciones especiales de tratamientos alternativos de agua en el hogar

- **Hervir el agua**: las altas temperaturas tienden a matar todas las clases de microorganismos incluyendo las bacterias.
- **Filtrar el agua:** la filtración remueve sedimento del agua, es en el sedimento que vive la mayoría de las bacterias.
- **Dejar el agua para que asiente el sedimento** antes de consumirla; para la misma razón como la filtración.
- **Desinfección solar del agua**: Para consumo humano, se recomienda dejar botellas de plástico transparentes expuestas al sol por algunas horas (6 a 8 horas de exposición es suficiente), este método usa los rayos UV del sol para destruir las bacterias que causan enfermedades transmitidas por el agua contaminada.

Appendix G. Chapter 6 Supplementary Information

Table 6.S1. Water quality measurements from Moro Moro, Pucará and Postrervalle. These data are a subset of those used for the analysis in chapter 4 and thus were collected using the methods described therein.

Community	Site	Date measured	Water body type	E. coli per 15ml sample	E. coli per 100ml equivalent	Other coliforms per 15ml sample	Other coliforms per 100ml equivalent	Turbidity (FAU)
Moro Moro	Intake 1	5/4/2016	Stream	35	233.33	464	3093.33	32
Moro Moro	Intake 2	5/4/2016	Stream	23	153.33	133	886.67	6
Moro Moro	Intake 3	5/4/2016	Stream	5	33.33	367	2446.67	10
Moro Moro	Storage tank	5/4/2016		28	186.67	197	1313.33	6
Moro Moro	Tap (school)	5/4/2016		5	33.33	98	653.33	7
Pucará	Intake	4/4/2016	Stream	12	80	343	2286.67	5
Pucará	Storage tank	4/4/2016		3	20	102	680	4
Pucará	Tap (school)	4/4/2016		1	6.67	38	253.33	5
Pucará	Former intake	4/4/2016	Small stream	11	73.33	278	1853.33	16
Postrervalle	Intake	31/3/2016	Spring	0	0	73	486.67	0
Postrervalle	Storage tank	31/3/2016		0	0	5	33.33	0
Postrervalle	Tap (primary school)	31/3/2016		0	0	4	26.67	0
Postrervalle	Tap (secondary school)	1/4/2016		0	0	13	86.67	0

Table 6.S2. Water quality measurements from La Aguada. These data were also collected using the method described in chapter 4.

Community	Site	Date measured	Water body	E. coli per 15ml sample	E. coli per 100ml equivalent	Other coliforms per 15ml sample	Other coliforms per 100ml equivalent	Turbidity (FAU)
La Aguada	Intake	12/3/2016	Stream	19	126.67	470	3133.33	49
La Aguada	Principal storage tank	12/3/2016		12	80	648	4320	87
La Aguada	Tap (house of Robert Rueda Villarroel)	12/3/2016		14	93.33	924	6160	120