Randomised Controlled Trials for the Evaluation of the CAP: Empirical Evidence about Acceptance by Farmers

Ulrich B. Morawetz
University of Natural Resources and Life Sciences Vienna, Austria

Christoph Tribl
Federal Institute of Agricultural Economics, Rural and Mountain Research, Vienna, Austria

Abstract

To conduct a randomised controlled trial (RCT) to evaluate the Common Agricultural Policy it would be necessary to exclude a random selection of farms from participation. This exclusion might limit the acceptance of RCTs. We assess the acceptance of an innovative alternative RCT called the 'unconditional payment RCT' (upRCT). UpRCTs allow for the evaluation of the impact of policy measures in which farmers receive a payment conditional on the adoption of farm management practices (e.g., agri-environment-climate measures). We surveyed Austrian farmers who participated in the 'refrain from silage' measure to compare the acceptance of a conventional RCT and an upRCT using thought experiments. The acceptance of the farmers was between 18% and 51%, and the treatment effects of both variants were of comparable size. Our survey suggests that acceptance of the upRCT is about twice as high as the acceptance of the conventional RCT. We discuss that upRCTs are useful when a new measure is introduced or when the upRCT is conducted for several years.

Key Words

EU Common Agricultural Policy (CAP); Randomised Controlled Trial (RCT); Evaluation

1 Introduction

European member states need to justify their expenditures for the Common Agricultural Policy (CAP). Currently, the evaluation of the CAP is mostly based on economic simulation models (e.g., KIRCHNER et al., 2015; SCHROEDER et al., 2015), econometric models (e.g., KIRCHWEGER et al., 2015; KLAIBER et al., 2017; CHABÉ-FERRET and SUBERVIE, 2013), case studies (e.g., MITTER et al., 2014) or qualitative approaches (e.g., DARNHOFER et al., 2017). One requirement of all the econometric evaluation approaches is a suitable control group. In evaluations of the CAP, it is often difficult to find an appropriate control group because i) many CAP measures are carefully designed to target specific sub-groups; ii) CAP measures are available, on a voluntary basis, to all applicants who fulfil the eligibility criteria; and iii) CAP measures are typically maintained for several programme periods, limiting the number of pretreatment observations. To secure an appropriate control group, randomised controlled trials (RCTs) have become a well-established evaluation method in labour and development economics. To the best of our knowledge, RCTs have not been used in evaluation studies of the CAP or in North America (COLEN et al., 2016; PALM-FORSTER et al., 2019; BEHAGHEL et al., 2019).

The guidelines of the European Evaluation Network for Rural Development (an evaluation expert network that operates under the responsibility of the European Commission's Directorate-General for Agriculture and Rural Development) consider RCTs to be a 'golden standard', although they are difficult to apply (EENRD, 2014: 87). Difficulties with RCTs are also common in other institutional settings. SHADISH et al. (2002) describe how the pilot studies or phase-in of a programme can be used to apply RCTs: a random subsample is treated earlier and compared with the untreated subsample. A similar strategy can be applied in cases of an over-subscription (i.e., with more applicants than can be supported); who is treated can be randomly determined. Another option for a random treatment is the 'encouragement design' where randomly selected eligible farms are encouraged (e.g., through targeted information) to participate in a measure. The intensity of the encouragement is then used as an instrumental variable in the evaluation (for an application, see, for example, LEÓN, 2017). While they are relatively straightforward to apply, we are not aware of any evaluations of this kind in the context of the CAP.

A related strand of literature uses RCTs to assess payments for ecosystem services (PES) by private insti-

DOI: 10.30430/69.2020.3.183-199

tutions. SMITH et al. (2019) used an RCT to show that the PES paid by a water company to farmers in central England reduced the metaldehyde concentration in treated water catchments. JAYACHANDRA et al. (2017) used an RCT to show that the decline of the tree cover in Uganda was reduced by half using PES paid by a nonprofit organisation. In comparison to the CAP, private companies and nonprofit organisations are not required to grant payments to all the eligible applications and can therefore randomise who is treated.

Because nobody can be forced to participate in a CAP measure, randomisation in conventional RCTs could only be achieved by excluding some eligible applicants from participation (i.e., randomly selecting eligible applicants who must not participate or who can participate only later in the case of a phase-in). According to the information from the Directorate-General for Agriculture and Rural Development of the European Commission, if an applicant for an agri-environment-climate measure is found to be eligible, the member state is obliged to pay the applicant in full (EUROPE DIRECT, 2019). Thus, it is impossible to use a conventional RCT as an integral part of a CAP measure under current EU regulations.

Furthermore, the successful application of RCTs is only possible when there is support by farmers, their associations and the managing authority. In this article, we elaborate on how evaluation studies of the CAP could be supplemented with RCTs by considering an innovative RCT variant that was first described by MORAWETZ (2014) and was referred to by several authors (BEHAGHEL et al., 2019; COLEN et al., 2016; THOYER and PRÉGET, 2019). This variant aims to increase the acceptance of RCTs by farmers but has not yet been tested. We call this variant an 'unconditional payment RCT' (upRCT) because the control group receives the CAP payments of the respective measure unconditionally. An upRCT is different from a conventional RCT and does not depend on a phase-in, over-subscription or encouragement design. An upRCT can be applied when the payment is conditional on some management practice. For example, the payments of agri-environment-climate measures are usually conditional on farm management that is more environmentally friendly. The payment intends to compensate for the additional costs or income foregone (e.g., because no pesticides are used). The key idea of an upRCT is that randomly selected eligible farms are granted support, but these farmers are free not to follow the management that usually comes as a condition of the support. Theoretically, this feature does not simply add a control group to CAP support but should increase the acceptance of such an RCT; the control group (i.e., those receiving the unconditional payment) are better off than the group without it and those who are treated are just as well off. Note that those receiving the *conditional* payment (i.e., the 'normal' participants of the measure) are referred to as the 'treated' group.

There is an argument from behavioural economics regarding why upRCTs might be more acceptable than RCTs: in an upRCT, those who are randomly selected into the control group 'gain' because the conditionality of payments is removed. In an RCT, those who are randomly selected into the control group 'lose' because they do not receive payments. In both cases, however, one group is better off, which might be considered to be unfair. From behavioural economics, it is well known that, for most people, losses have much larger psychological impacts than gains of the same magnitude (KAHNEMAN and TVERSKY, 1979). We would thus expect upRCTs to be more acceptable. In our acceptability assessment below, we survey the acceptance of the control group (in RCTs and upRCTs). We do not survey the acceptance of those farmers who were not selected into the (up)RCT control group or of other stakeholders.

This article makes four contributions to improve the understanding of the suitability of RCTs for the CAP. First, we describe how an RCT and an upRCT could be applied to CAP measures. Second, using a thought experiment in a survey, we assess whether the acceptance rates of upRCTs are higher than those of RCTs among farmers. Thought experiments have been used in the economic literature (dubbed 'contingent behaviour') to ask questions related to hypothetical behaviour (ENGLIN AND CAMERON, 1996). Our estimates are based on a survey among farmers who participated in the 'refrain from silage' agri-environment measure from the Austrian Rural Development Programme in the year 2017. Third, in the survey, we also test whether there is a difference between the 'stated average treatment effect' in an RCT and that in an upRCT. We call the effect 'stated' because it is based on the replies from the thought experiment. Fourth, the survey is also used to discuss the extent to which RCTs are useful for evaluating if farmers have already participated in the evaluated measure before the RCT is conducted. This issue is particularly relevant for the CAP where many measures are established for several programme periods.

2 Methodological Framework: RCTs for the Evaluation of the CAP

Typically, the objective of an econometric evaluation study is to estimate the effect of participation in a programme. In programmes with voluntary participation, the focus of an evaluation study is usually on the treatment effect on the treated, i.e., on estimating the outcome that would have happened if the treated had not been treated. The crucial point is that those who are treated can be systematically different from those who are not treated (otherwise they would not have voluntarily participated in the programme). If some of the programme outcome would have happened even without the programme, this is called a 'windfall gain' for the programme participants or the 'dead weight loss' of the programme (CHABÉ-FERRET and SUBERVIE, 2013).

RCTs solve the self-selection problem in voluntary programmes by randomly selecting who is treated (i.e., those participating in a programme) and who is not treated (i.e., the control group). Let ATT denote the average treatment effect on the treated, and let y be the outcome of interest. For the definition of the counterfactual, it is useful to define y_1 as the outcome of those treated and y_0 as the outcome of the control group. Let D be a dummy variable that is equal to 1 if there was in fact a treatment and zero otherwise. Thus, an outcome y_0 in combination with D = 1 is counterfactual; we cannot observe this outcome in reality. The outcome y_0 describes the value of the outcome of interest when not treated (e.g., y_0 can be a certain indicator for environmental quality when that particular farm does not participate in an agri-environment measure), but conditioning on D = 1 reflects treatment (e.g., the farm in fact did participate in the agri-environment measure). In contrast, y_0 combined with D = 0 is observable. The expected value of the estimated ATT is then

$$\mathrm{E}(\widehat{\mathrm{ATT}}) = \underbrace{\mathrm{E}(y_1|\mathrm{D}=1) - \mathrm{E}(y_0|\mathrm{D}=1)}_{\text{average treatment effect on the treated}} = \tag{1}$$

$$\underbrace{\mathrm{E}(y|\mathrm{D}=1) - \mathrm{E}(y|\mathrm{D}=0)}_{\text{observed difference in average outcome}} -$$

$$\underbrace{(\mathrm{E}(y_0|\mathrm{D}=1) - \mathrm{E}(y_0|\mathrm{D}=0))}_{\text{selection bias}}$$

where $y = y_0 + (y_1 - y_0)\,\mathrm{D}$.

If the treatment is randomly assigned, the selection bias disappears because the treated and control groups are not systematically different. Several assumptions for unbiasedness under random assignment are necessary: the effect of the treatment is due to the treatment and not to factors correlated with the treatment ('exclusion restriction'), there are no systematic missing observations ('attrition'), all participants receive the treatment to which they were assigned ('compliance') and there is no interference between participants ('stable unit treatment value assumption'); see GERBER and GREEN (2012) for a detailed discussion.

Another requirement (for all empirical evaluation methods) is that the outcome of interest must be observed for participants and non-participants. The outcome of interest can be some indicator of environmental quality or economic performance. Most CAP measures, however, are action–based and focus on the farm management itself (e.g., refraining from pesticides rather than increased biodiversity) (see BURTON and SCHWARZ, 2013). We therefore focus on the evaluation of action–based measures and farm management as outcomes.

2.1 Unconditional Payment Randomised Controlled Trials (upRCTs)

An upRCT is applicable when the payment of a programme is conditional on a certain farm management practice, as is typical for action-based measures. When applying the upRCT, a random selection of eligible farms is granted the payment unconditionally. Thus, these farms must not participate in the CAP measure (and are therefore the random control group) but they receive payments without having to comply with the conditions. Given that the recipients of unconditional payments manage their farms as if they were not participating (at least with respect to the outcome of interest), we observe $\mathrm{E}(y_0|\,\mathrm{D}=1)$. This allows for estimating the average bias as $\mathrm{E}(y_0|\,\mathrm{D}=1) - \mathrm{E}(y_0|\,\mathrm{D}=0)$ and the ATT as $\mathrm{E}(y_1|\,\mathrm{D}=1) - \mathrm{E}(y_0|\,\mathrm{D}=1)$.

The key hypothesis for the validity of an upRCT is the equality of the ATT derived from the upRCT and RCT. A first reason why this might fail is that the moral obligations of the unconditional payment recipients might influence their behaviour. The literature on experimental auctions analysed 'reciprocal obligation' (CORRIGAN and ROUSU, 2006). The idea is that the participant wants to repay something to the experimenter by bidding high. In our context, this practice

would mean that a farmer may voluntarily comply with the conditionality to repay the managing authority for the unconditional payment. The effect of 'reciprocal obligation' is more likely in experimental auctions than in an CAP measure evaluation because the experimenter in a face-to-face experimental auction is a real person whereas the unconditional payment is provided by a managing authority. Additionally, the costs of paying something back are typically low in experimental auctions (a few Euros), but the related costs can be high in the context of CAP measures.

The second – and main – reason that the key hypothesis may fail is that not being admitted to participate in the measure (i.e., becoming part of the control group) is perfectly correlated with the unconditional payment. The change in the budget constraint of the control group resulting from the unconditional payment can change the (optimal) farm management. If this is the case, an upRCT is not suitable. Generally, the larger the payment is, the more likely the unconditional payment is to affect the optimal farm management. For a theoretical analysis of how the changes in the budget constraint influence production decisions, see CHAU and GORTER (2005).

When statistically testing the hypothesis of equal ATTs, it is helpful to augment the notation introduced above. The perfect correlation between being randomly selected into the control group and receiving the unconditional payment can be represented by replacing the dummy variable D for treatment with two dummy variables. We can define A as a dummy variable that is A = 1 if an applying eligible farm is admitted to the programme and therefore has to comply with the conditions of the measure. Then, A = 0 when an applying and eligible farm is not admitted to the programme and therefore does not have to comply with the conditions. We define P as another dummy variable that is P = 1 if there is a payment to the farm, and P = 0 if there is no payment to the farm. Those applying and eligible farms that are randomly selected and receive unconditional payments (i.e., the control group) have A = 0 and P = 1. The applying and eligible farms that are not randomly selected (i.e., the treatment group) will have A = 1 and P = 1. Table 1 compares the expected values of the outcomes in the RCTs and upRCTs. The upRCT differs from the RCT, as those who were not admitted to participate (A = 0) receive unconditional payments (P = 1).

The 'exclusion restriction' requires that the effect of a treatment is due to the treatment and not to factors correlated with the treatment (GERBER and GREEN, 2012: 39). In our case of a randomly determined control group, the exclusion restriction means that being randomly selected into the control group must not be correlated with factors influencing the outcome. In the case of an upRCT, the unconditional payment is perfectly correlated with the random selection into the control group (P = 1 whenever farms are selected for non-admission A = 0). Thus, if the unconditional payment has an effect on the outcome, the exclusion restriction is not fulfilled. We can test if the unconditional payment makes a difference by comparing the outcome of interest between an RCT and an upRCT (see Table 1):

$$E(y_0 | A = 0, P = 1) = E(y_0 | A = 0, P = 0) \quad (2)$$

In the RCT literature, a usual assumption is that a policy measure is newly introduced. In the case of CAP evaluations, this is not always possible, since many measures have been offered with slight variations over many programme periods. In this case, the long-term commitments by farmers might have been made in expectation of payments from the measure. This expectation can be interpreted as a violation of the 'compliance' assumption because the random

Table 1. Expected values (E( )) of the outcomes of the treatment and the control group conditional on being admitted to a programme (A = 1), receiving a payment (P = 1) and being admitted and receiving a payment (D = 1)

	RCTs	upRCTs
Average outcome of the treatment group	$E(y_1 \mid A = 1, P = 1) = E(y_1 \mid D = 1)$	$E(y_1 \mid A = 1, P = 1) = E(y_1 \mid D = 1)$
Average outcome of the control group	$E(y_0 \mid A = 0, P = 0) = E(y_0 \mid D = 0)$	$E(y_0 \mid A = 0, P = 1)$
Selection bias in E(ATT)	$E(y_0 \mid D = 1) - E(y_0 \mid D = 0) = 0$	$E(y_0 \mid D = 1) - E(y_0 \mid D = 0) = 0$ if: $E(y_0 \mid A = 0, P = 1) = E(y_0 \mid A = 0, P = 0)$

Source: own calculations

control group was expecting to receive the treatment. If the RCT or upRCT is conducted for only a short period of time (e.g., a year), then changing the farm management might not be optimal, even if it would be in the longer run.

Therefore, if the measure is not new, it is necessary to conduct the (up)RCT for a longer time period (e.g., the time it takes to change delivery contracts, for investments to be profitable and to gain experience). Otherwise, there is a risk that farmers just continue doing what they have been doing in the previous period.

Finally, when randomly selecting who receives payments unconditionally as part of an upRCT, it is necessary to consider not only eligible applicants but also eligible non-applicants. Otherwise, risk-loving non-applicants might decide to apply just for the chance to receive unconditional payments. If practical reasons make this practice infeasible, it is necessary to stipulate that farmers cannot resign from the contract once they know whether they are in the upRCT control group.

3 Measuring the Acceptability of upRCTs to Evaluate the 'Refrain from Silage' Agri-environment Measure

Support for the evaluation of CAP measures based on an upRCT can only be expected if the method is well understood. In the following, we present an acceptability assessment among Austrian farmers using a thought experiment in a survey. Thought experiments, sometimes called 'contingent behaviour', are used to ask questions related to hypothetical behaviour (ENGLIN and CAMERON, 1996). The objective of our thought experiment is to test the acceptance of an RCT and an upRCT and to investigate the assumption from Equation (2) for the 'refrain from silage' agri-environment measure. This agri-environment measure is part of the Austrian 'Rural Development Programme' 2014-2020.

The objective of the 'refrain from silage' measure is to increase biodiversity and preserve traditional land management. Farms are compensated for the additional costs and income foregone due to the production of hay instead of silage. Since grass is expected to be cut later for hay production than for silage production, hay is expected to have a positive effect on biodiversity. Since the effect on biodiversity (as the result of a farm management) is difficult to measure, our outcome of interest is hay production, i.e., the management practice per se. The evaluation of the management practice is also interesting because, since the first introduction of the 'refrain from silage' measure in the year 2000, the market for hay-milk products has substantially expanded.

Hay and silage are both used as fodder for livestock. The production of hay requires the cut grass to dry before it is stored. However, when producing silage, the cut grass can be immediately wrapped into silage bales. Therefore, hay production is much more susceptible to weather risk. For the production of silage bales, a wrapper is necessary. This mobile machinery can be shared among farms. Once the silage bales have been produced, they can be conveniently stored outdoors. However, storing hay requires an indoor space, which means that storing hay is in general more expensive. The nutritional value of silage is higher than that of hay. When using only hay, fodder supplements (concentrated feed) need to be given. The weather risk of hay can be reduced if famers use hay ventilation, which allows for partly drying hay indoors. Hay ventilation also improves fodder quality, but it requires an investment, and the ventilation itself increases energy costs.

Thus, silage has some advantages with respect to production, feed quality and costs. One reason why farms refrain from using silage, independent from participating in the agri-environment measure, is that dairies pay a higher price for raw milk from cows that are fed without silage. This raw milk is used to produce 'hay milk' products that have higher consumer prices than conventional milk products and for the production of traditional hard cheese. Milk producing farms can sign a hay milk delivery contract with the dairy and thereby commit to not feeding silage. Terminating the contract with the dairy is usually possible within a lead time of a couple of months, but there is no guarantee that farms can re-join again later under similar conditions (WIENER ZEITUNG, 2017).

The 'refrain from silage' agri-environment measure in Austria requires farms to completely refrain from producing, using, storing and trading silage. The payment for cattle farms is 80 Euros per hectare per year. If these cattle farms produce raw milk, the payment is 150 Euros per hectare. There is no payment for farms without cattle.

In an online survey conducted in the spring of 2018 among farmers in Austria who participated in 2017 in the 'refrain from silage' measure, we explained the concept of RCTs and upRCTs and con-

ducted two different thought experiments by setting up two different hypothetical scenarios. i) In one thought experiment ('RCT'), the respondents received a hypothetical letter that explained that they cannot participate in the 'refrain from silage' measure in the next year. The respondents will not receive any payments and are free to either comply or not with the conditions of the measure. ii) In a second thought experiment ('upRCT'), the respondents received a hypothetical letter that explained that they cannot participate in the 'refrain from silage' measure in the next year but will still receive the payments that usually come with participating in the measure. Even though they receive the payments, they are not obliged to comply with the conditions of the measure. In the survey, each respondent was presented both thought experiments. The order of the two thought experiments was randomised.

First, we asked respondents whether they would accept either the RCT or the upRCT and their reasons. Second, in order to test the assumption from Equation (2), we asked the respondents for their hypothetical hay production in each of the thought experiments (measured as a share of the total mowing material) and the reasons for their answers. We also asked for some farm characteristics and were able to use additional farm-specific data from the Integrated Administration and Control System (IACS).

There is an extensive literature on the usefulness of hypothetical scenarios in surveys. A recent meta-analysis of hypothetical biases of PENN and HU (2018) in the context of valuation studies found that surveys systematically differ in the magnitude of the hypothetical bias. These findings include that, on average, questions related to public goods have a higher bias, there is no statistically significant difference between survey modes (personal surveys, lab, online surveys, etc.) and that certainty follow-up questions substantially reduced the hypothetical bias.

Compared to the valuation of a public good, the first hypothetical task that we ask farmers to perform is relatively easy (to consider if they would accept an (up)RCT). Additionally, there is no obvious strategic behaviour. The second hypothetical task (to estimate the percentage of hay production if selected in the control group) may be more difficult, but the within-respondent design (each respondent answers both the RCT and the upRCT) helped to reduce the number of potentially confounding factors across respondents.

We also applied a variant of the certainty follow-up question to check for an uncertainty related bias (see Appendix A1 for details); the results did not change substantially when using only those respondents who were certain about their replies. We are thus optimistic that having a hypothetical survey does not undermine our conclusions.

All data analysis was performed using the R software (R CORE TEAM, 2018), and the online survey was performed using LimeSurvey. The translated questionnaire is available from http://www.wiso.boku.ac.at/ulrichmorawetz.html.

3.1 Results of the Acceptability Assessment

In 2017, a total of 11,021 farms received payments from the 'refrain from silage' agri-environment measure. (The data were retrieved from the IACS database in the spring of 2018.) Of these farms, 5,451 farmers could not be contacted because no (or no valid) email address was recorded. Of those farms with an email address in the database (5,570 farms), 23% (1,250) completed and 4% (245) started but did not complete the survey. Table 2 shows that the respondents had, on average, more utilised agricultural area and more livestock units and received higher agri-environment payments and higher 'refrain from silage' payments than non-respondents and those for whom no email address was available.

Our survey is not representative with respect to these farm characteristics. Weighting respondents to be representative with respect to these observed characteristics is possible. However, representativeness with respect to our variables of interest (acceptance of (up)RCTs and percentage of hay production when participating in an (up)RCT) is unknown. It is thus unknown whether weighting would improve or worsen the representativeness. Nevertheless, in the appendix A2, we show that the results do not change substantially when we conduct weighting by post-stratifying (LOHR, 2009: 342) responses according to the payments for 'refrain from silage' (above and below the median), milk production and organic farming.

Based on the survey data, we find a pronounced order effect for acceptance (see Table 3). Among those who were first presented the RCT, the acceptance rate of an upRCT is 51%; and among those who were first presented the upRCT, the acceptance rate of an upRCT is 31%.

Table 2. Farm characteristics of Austrian farms in the survey based on the 'refrain from silage' measure

	Mean	St. dev.	Min.	Max.	Median	# Obs.
Utilised agricultural area (ha)	15.53	15.82	0.10	320.35	10.87	11,021
Respondents: survey completed	18.55	16.04	0.78	136.06	13.87	1,250
Respondents: survey not completed	17.39	15.50	0.59	128.83	13.20	245
Non-respondents	16.28	16.87	0.10	320.35	11.32	4,075
Not contacted (no email)	14.18 **	14.81	0.57	278.33	9.86	5,451
Livestock units	18.26	16.37	0.46	317.38	13.43	11,021
Respondents: survey completed	21.74	18.26	1.11	142.03	16.46	1,250
Respondents: survey not completed	19.42 *	16.34	0.52	106.74	14.63	245
Non-respondents	19.25	17.84	0.46	317.38	14.22	4,075
Not contacted (no email)	16.67 **	14.47	0.75	182.88	12.10	5,451
Agri-environment payments (€)	5,332.43	4,893.09	18.32	113,265.59	3,901.75	11,021
Respondents: survey completed	6,973.04	6,073.85	168.31	76,758.13	5,225.05	1,250
Respondents: survey not completed	5,766.85 **	4,381.56	104.36	27,223.36	4,625.50	245
Non-respondents	5,668.95	5,264.86	18.32	113,265.59	4,181.52	4,075
Not contacted (no email)	4,685.11 **	4,153.02	162.01	47,643.93	3,432.83	5,451
'Refrain from silage' payment (€)	1,365.52	1,359.88	0.64	20,146.09	901.60	11,021
Respondents: survey completed	1,711.69	1,638.03	3.24	10,742.08	1,166.10	1,250
Respondents: survey not completed	1,458.26 *	1,437.40	32.24	7,857.40	970.25	245
Non-respondents	1,444.42	1,432.10	0.64	20,146.09	963.08	4,075
Not contacted (no email)	1,222.99 **	1,202.45	2.13	11,415.79	790.74	5,451

Notes: all values are for the year 2017.
The significance levels for the mean difference between 'Respondents: survey completed' and 'Respondents: survey not completed' and between 'Respondents: survey not completed' and 'Non-respondents' and between 'Non-respondents' and 'Not contacted (no email)' are as follows: *5% level, and **1% level.
Source: IACS database (Federal Ministry for Sustainability and Tourism, Austria; Agrarmarkt Austria) and own calculations

Table 3. Acceptance of the RCT and the upRCT

	RCT presented first	upRCT presented first	All
Acceptance RCT	26%	18%	22%
Acceptance upRCT	51%	31%	41%

Number of respondents: 1,246 (some respondents did not reply to both questions)
Source: own calculations

The participants had the option to select multiple reasons why they would (or would not) accept an RCT or an upRCT (see Table 4). We find that 48% of the respondents who would accept participation in an RCT would do so because they think that it is important to demonstrate the effect of the measure. A proportion of 31% of the respondents would accept participation in an RCT because the received payment is so low that it does not make a difference whether they receive or do not receive the payment. Those who would not accept participation in an RCT find it unfair (52%), would be disadvantaged by participating (57%) or had counted on the payment (52%).

Referring to the upRCT, we find that 48% of the respondents who would accept participation do so because they think it is important that the effect of the measure will be proven. We find that 35% of those who would accept an upRCT would do so because it is advantageous for them. Among the 'other reasons' (26%), respondents emphasised the opportunity to produce silage in case of rain during harvest time. Among those who would not accept an upRCT, 31% consider the upRCT to be unfair. Hence, the upRCT is considered as fairer than the RCT (which is considered as unfair by 52%), but the majority would not accept an upRCT because, in their view, unconditional payments make no sense (71%). This point suggests that better explanations of the idea of upRCTs could increase acceptance. 'Other reasons' for not accepting (up)RCTs can be summarised as a preference for hay production, regardless of the incentive structures. These farmers listed the steepness of slopes that makes silage production more expensive, existing contracts with a dairy, their 'current farm manage-

Table 4. Reasons for accepting the RCT and the upRCT

RCT:		upRCT:	
Accept because …		**Accept because …**	
… it is important to prove the effect of the measure	48%	… it is important to prove the effect of the measure	48%
… I can easily forgo the payment (e.g., because it is so low)	31%	… it results in an advantage for me	35%
Other reasons	30%	Other reasons	26%
Do not accept because …		**Do not accept because …**	
… it is unfair	52%	… it is unfair	31%
… it results in a disadvantage for me	57%	… unconditional payments do not make sense	71%
… I have counted on the payment (e.g., for investments)	52%		
… I am generally against checking the effect of measures	3%	… I am generally against checking the effect of measures	3%
Other reasons	19%	Other reasons	26%

Note: more than one answer was possible.
Source: own calculations

ment plan', or the 'smell of silage as a problem in agri-tourism' as reasons why silage would not be an option in any case.

Regardless of whether they accepted the measure or not, we asked respondents to tell us how much hay they would produce in the RCT and the upRCT thought experiments (as a percentage of the total mowing material). The mean values of hay production varied between 90% and 93% for the RCT (depending on whether the RCT or the upRCT was presented first) and between 93% and 94% for the upRCT, as shown in Table 5. The black line in the top panel of Figure 1 shows the percent of hay production in the RCT scenario. The respondents are sorted by percentage of hay production in the RCT scenario. Maintaining the same order of respondents, the red dots show the percentage of hay production in the upRCT scenario. When the red dots are on the black line, the respondents produce the same percentage of hay in the RCT and the upRCT scenario. When the red dots are above the black line, the respondents produce more hay in the upRCT scenario. When the red dots are below the black line, the respondents produce less hay in the upRCT scenario. For the majority of the respondents, their hay production is identical in both scenarios (85% of respondents). For 11% of the respondents, their hay production is higher in the upRCT scenario. This result could be explained by moral reciprocity or the budget constraint making hay production sub-optimal without payment. Figure 1 also shows that some respondents (4%) state that they would produce more hay in the RCT than in the upRCT scenario. We do not have an economic explanation for these replies.

Table 5. Percentage of hay production. The stated average treatment effect on the treated is 100% minus the stated hay production

	RCT presented first	upRCT presented first	All
RCT			
Mean	90%	93%	92%
Median	100%	100%	100%
25th percentile	100%	100%	100%
upRCT			
Mean	93%	94%	94%
Median	100%	100%	100%
25th percentile	100%	100%	100%
Difference upRCT-RCT			
Mean	2.9 ***	1.2 **	2.0 ***

Note: for the mean difference from zero, ***represents 1% significance and **5% significance
Source: own calculations

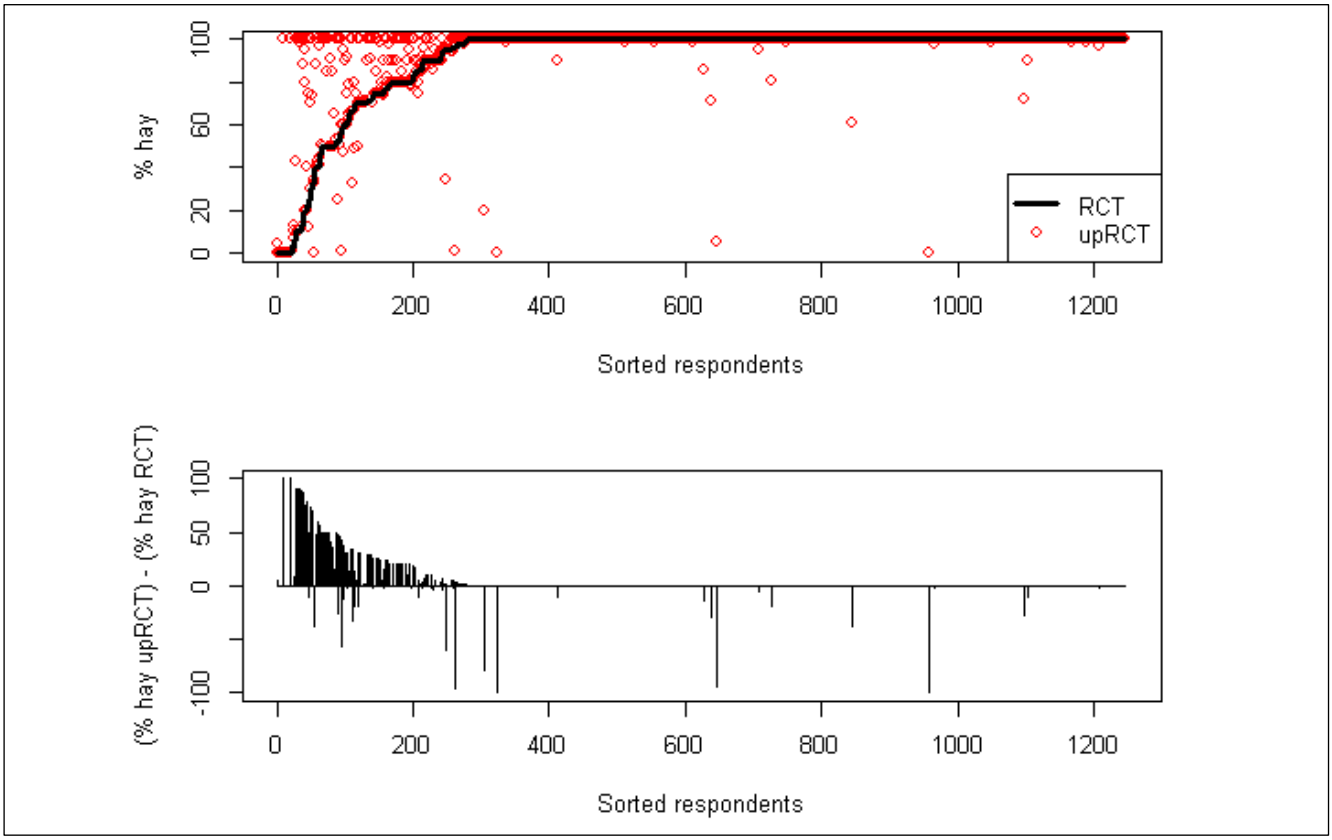
Using the stated hay production, the stated treatment effect on the treated can be derived by subtracting it from 100. The stated ATT is thus an 8% (RCT) and a 6% (upRCT) increase in hay production as a consequence of the 'refrain from silage' measure. We also analysed the results

for only those who would accept participation in an RCT and an upRCT. The results are not substantially different, as shown in Appendix A3.

The last row of Table 5 shows that there is a statistically significant difference from approximately 1 to 3 percentage points in the stated average treatment effect (i.e., between the mean hay production in the RCT and the upRCT scenarios). This small difference between the upRCT and RCT might be due to reciprocal obligations or changes in the budget constraint in the upRCT.

Using the replies of the respondents, we investigate these two reasons. In one of the debriefing questions of the questionnaire, respondents were asked whether the unconditional payment in the upRCT scenario was a reason for producing hay. The respondents who answered yes were asked more specifically if it was moral reciprocity or budget constraints that made the unconditional payment important in the decision to produce hay. We find that 7% (83 respondents) of those who produced hay in the upRCT thought experiment felt morally obliged to produce hay. To investigate the influence of the budget constraint, we did not rely on the responses to the debriefing question (5% said that the budget constraint was the reason for producing hay in the upRCT scenario) but instead compared the RCT scenario to the upRCT scenario. We tested if the change in the budget constraint due to the unconditional payment had an influence. We did so by comparing the stated hay production in the RCT thought experiment (without payment) to the stated hay production in the upRCT thought experiment (with unconditional payment). Since each respondent participated in both thought experiments (the order was randomised), this is a within-respondent design. We find that 11% of the respondents (134 respondents) stated that they would produce more hay in the upRCT thought experiment than what they stated in the RCT thought experiment. Of these 134 respondents, 15 also replied that they felt a moral obligation to produce hay under the upRCT. After omitting these 15 respondents, the percentage of those affected by budget constraints decreased to 10.5%. This figure reflects how many farms would increase their quantity of hay produced because the unconditional payment relaxes their budget constraint. However, many of these farms would only produce slightly more hay in the upRCT experiment than in the RCT thought experiment, which can be seen in the bottom panel of Figure 1.

Figure 1. Stated hay production in the upRCT and the RCT scenarios. Respondents sorted by the hay production percentage in the RCT scenario



Source: own calculations

All our respondents participated in the measure in the year before the survey took place (2017) and consequently produced 100% hay in that year. Thus, switching from hay to silage production for one year might not be possible or reasonable. The reasons for this include the following: among all the respondents, 60% have existing hay-milk delivery contracts, 51% lack silos or a silage bale press and wrapper, and 32% state that they have limited knowledge about silage production. At least one of these three limitations to switching to silage production in the short run applies to 79% of the respondents. For some of these respondents, the (up)RCT is not applicable because contracts, investments and experience are a consequence of having participated in the measure (one could say that the 'compliance' assumption is not fulfilled).

In Table 6, we use linear regressions to explain the hay production in the upRCT scenario minus the hay production in the RCT scenario as a percentage. The first column reproduces the results from the last row of Table 5; the constant is the average difference between the hay produced in the upRCT and RCT (last column, last row in Table 5). The dummy variable for respondents where the upRCT scenario was presented before the RCT scenario, 1.7 percentage points, is identical to the difference 2.9 − 1.2 in the bottom row of Table 5. In the second column of Table 6, we control for existing hay-milk contracts, lack of silage production facilities and limited silage production experience. The intercept decreases by 0.45 percentage point, confirming our main conclusion that a significant but small difference between the upRCT and RCT scenarios exists. The dummy for 'upRCT presented first' remains practically unchanged, the dummy for existing hay contracts is significant and positive (10% level), and the dummies for no silage production facilities and limited knowledge are insignificant. Together, the three dummies for restricted silage production possibilities are insignificant with an $F_{1212,3}$-statistic of 1.075 (p-value: 0.36). An alternative to conditioning the estimated coefficients using only those respondents who do not have silage production restrictions is dropping the restricted observations. After running the regression with only the 253 unrestricted observations, we find a decrease in the intercept of 0.61 percentage points compared to the unconditional regression. Again, this finding does not change our main conclusion. Interestingly, the dummy for 'upRCT presented first' becomes insignificant.

In this analysis, we focus on the difference between the RCT and upRCT. However, if the interest was in estimating the ATT, the selection which observations to disregard needs further considerations. Some farms did not use silage even before the measure was introduced: higher hay milk prices (traditional hard cheese production requires hay milk), higher silage production costs due to steep slopes and limited knowledge about silage production were already

Table 6. Effect of conditioning the difference between the upRCT and RCT scenarios on farms not restricted in the production of silage

	All observations	All observations	Unrestricted observations
Constant	2.907 ***	2.456 ***	2.297 ***
	(0.598)	(0.760)	(0.880)
upRCT scenario presented first (dummy)	-1.723 **	-1.751 **	-0.460
	(0.797)	(0.799)	(1.716)
Existing hay-milk contract (dummy)		1.301 *	
		(0.779)	
Lack of silage production facilities (dummy)		0.0002	
		(0.771)	
Limited experience in silage production (dummy)		-0.564	
		(0.868)	
Observations	1,215	1,215	253
R^2	0.004	0.006	0.0003

Note: * 10%, ** 5%, and ***1% significance

Dependent variable: hay production percentage in the upRCT scenario minus the hay production percentage in the RCT scenario. The unrestricted observations are those farms that do not have an existing hay-milk contract, do not lack silage production facilities and do not have limited experience in silage production. The robust standard errors are in parenthesis.

Source: own calculations

prevailing issues before the 'refrain from silage' measure was introduced. For those farms, the ATT is zero, and they should be included in the estimation of the ATT. If, instead, participation in the 'refrain from silage' measure in previous years made the farm commit itself to hay production (e.g., by signing a hay milk contract), the farm should not be included. Otherwise, the compliance assumption is not fulfilled. Thus, our 79% is the upper limit of the share of the farms where previous participation in the measure makes them unsuitable for the estimation of the ATT.

It is tempting to use the stated ATT for hay production elicited from the thought experiments as an estimate of the real ATT (i.e., from a non-hypothetical (up)RCT). We caution against this because the RCT and the upRCT scenarios were hypothetical. We did our best to ensure that our thought experiments were realistic, but our main focus was on testing acceptance and the assumptions from equation (2). To estimate the treatment effects, it would be necessary to pay unconditional payments for a longer time period (e.g., a seven-year CAP programming period). Doing so in a real experiment is possible, but doing so in a thought experiment is very difficult, as it involves complex hypothetical considerations for respondents. We thus consider a questionnaire to be unsuitable for estimating the treatment effects if long-term farm management decisions are involved. Our short-term thought experiment is sufficient to analyse whether there is a difference in the acceptance and the behaviour of farms in an RCT and in an upRCT but not to estimate the treatment effects.

4 Conclusions

Most CAP measures are voluntary, and all eligible applicants can participate. Consequently, a suitable control group is often not available. For the evaluation of a CAP measure where payments are conditional on a particular farm management practice, we describe a variant of RCTs: we define an unconditional payment RCT (upRCT) as an RCT in which a randomly selected group (the control group) receives payments unconditionally. An upRCT evaluation has the advantage that those who are randomly selected are better off than those not selected and are therefore more likely to accept this kind of evaluation method. In contrast, those who are randomly selected in an RCT are worse off than those not selected and are therefore likely to oppose it.

For upRCTs to be useful for an evaluation, the management decision of farms must not be influenced by the unconditional payment. Farmers might be influenced because of moral reciprocal obligations and the changes in their budget constraints from the unconditional payment. We investigated these two assumptions for the case of the Austrian 'refrain from silage' agri-environment measure. We use an online survey to conduct a thought experiment among the participants in the measure. The measure requires participants to refrain from producing, using, storing and trading silage. We find the acceptance of a hypothetical RCT to be between 18% and 26% and the acceptance of a hypothetical upRCT to be between 31% and 51%. The spread is explained by the order in which we presented the RCT and the upRCT in our within-respondent survey design. The responses also show that 71% of those who would not accept an upRCT do so because they do not understand the purpose of the unconditional payment. Non-familiarity with the evaluation methods suggest that the acceptance of an upRCT could be further increased by explaining the purpose of upRCTs more extensively.

We also asked farmers how much hay they would produce if the conditionality of the measure was lifted. We find that there is a statistically significant difference in the reduction of hay production in the RCT and the upRCT scenarios. The difference, however, amounts to only 2 percentage points on average. This finding suggests that – at least in our case – the difference in the stated treatment effect is relatively low, and the reciprocal obligations and the change in the budget constraint do not play major roles. Given the substantially higher acceptance among farmers and the relatively small difference in the estimated stated treatment effects, the results suggest that an upRCT has advantages over an RCT for the evaluation of the CAP.

The applicability of (up)RCTs as a method for the evaluation of the CAP depends on the acceptance of (up)RCTs by stakeholders. A future survey design could therefore be to survey farmers (treatment and control group), farmer representatives and the managing authority on the acceptance of (up)RCTs as an evaluation method. Future research also needs to investigate the heterogeneity in the acceptance of different measures, cost-benefit issues and legal issues related to RCTs. In the past, the European Commission has demonstrated that it is flexible when enabling innovative ideas in the CAP (TERWAN et al., 2016).

The main contribution of our study is to compare an upRCT to an RCT and assess their acceptance for the first time. Our empirical assessment is based on thought experiments. Acceptance may thus be overestimated. However, our finding that upRCTs are generally more accepted than RCTs is likely to hold because higher upRCT acceptance was found throughout all sub-samples, and it is based on a within-respondent design. Respondents' replies suggest that upRCTs are accepted because of higher benefits and because upRCTs are perceived as fairer than RCTs (possibly because of loss-aversion). We also emphasise that (up)RCTs must be conducted for longer periods of time for the measures already in place. The last point is relevant for the CAP because CAP measures often hardly change for several programme periods. It is important to keep in mind that the lack of pretreatment observations is a challenge for experimental and econometric evaluation methods alike. If no pretreatment evaluations are available, long-term upRCTs may be the only option to generate a reliable control group. This is particularly true for measures where a substantial share of participants have already been participating in the evaluated measure in the previous programme period.

References

BEHAGHEL, L., K. MACOURS and J. SUBERVIE (2019): How can randomised controlled trials help improve the design of the common agricultural policy? In: European Review of Agricultural Economics 46 (3): 473-93.

BURTON, R.J.F. and G. SCHWARZ (2013): Result-oriented agri-environmental schemes in Europe and their potential for promoting behavioural change. In: Land Use Policy 30 (1): 628-41.

CHABÉ-FERRET, S. and J. SUBERVIE. (2013): How much green for the buck? Estimating additional and windfall effects of French agro-environmental schemes by DID-matching. In: Journal of Environmental Economics and Management 65 (1): 12-27.

CHAU, N.H. and H. DE GORTER (2005): Disentangling the Consequences of Direct Payment Schemes in Agriculture on Fixed Costs, Exit Decisions, and Output. In: American Journal of Agricultural Economics 87 (5): 1174-1181.

COLEN, L., S. GOMEZ Y PALOMA, U. LATACZ-LOHMANN, M. LEFEBVRE, R. PRÉGET and S. THOYER (2016): Economic Experiments as a Tool for Agricultural Policy Evaluation: Insights from the European CAP. In: Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie 64 (4): 667-94.

CORRIGAN, J.R. and M.C. ROUSU (2006): The Effect of Initial Endowments in Experimental Auctions. In: American Journal of Agricultural Economics 88 (2): 448-457.

DARNHOFER, I., M. SCHERMER, M. STEINBACHER, M. GABILLET and K. DAUGSTAD (2017): Preserving permanent mountain grasslands in Western Europe: Why are promising approaches not implemented more widely? In: Land Use Policy 68 (Supplement C): 306-315.

EENRD (European Evaluation Network for Rural Development) (2014): Capturing the success of your RDP: guidelines for the ex-post evaluation of 2007-2013 RDPs. Available at http://enrd.ec.europa.eu/enrd-static/app_templates/enrd_assets/pdf/evaluation/epe_master.pdf: 212.

ENGLIN, J. and T.A. CAMERON (1996): Augmenting travel cost models with contingent behavior data. In: Environmental and Resource Economics 7 (2): 33-47.

EUROPE DIRECT (2019): Personal email from the Information Centre of the European Commission on 19 November 2019. Subject: 'Europe Direct - 101000546337'.

GERBER, A.S. and D.P. GREEN (2012): Field experiments: design, analysis, and interpretation. W. W. Norton, New York.

JAYACHANDRAN, S., J. DE LAAT, E.F. LAMBIN, C.Y. STANTON, R. AUDY and N.E. THOMAS (2017): Cash for carbon: A randomized trial of payments for ecosystem services to reduce deforestation. In: Science 357 (6348): 267-73.

KAHNEMAN, D. and A. TVERSKY (1979): Prospect Theory: An Analysis of Decision under Risk. In: Econometrica 47 (2): 263-291.

KIRCHNER, M., J. SCHMIDT, G. KINDERMANN, V. KULMER, H. MITTER, F. PRETTENTHALER, J. RÜDISSER, T. SCHAUPPENLEHNER, M. SCHÖNHART, F. STRAUSS, U. TAPPEINER, E. TASSER and E. SCHMID (2015): Ecosystem services and economic development in Austrian agricultural landscapes - The impact of policy and climate change scenarios on trade-offs and synergies. In: Ecological Economics 109: 161-74. Scopus.

KIRCHWEGER, S., J. KANTELHARDT and F. LEISCH (2015): Impacts on economic farm performance from government-supported investments in Austria. In: Agricultural Economics - Czech (Zemědělská ekonomika): 61 (8): 343-55.

KLAIBER, H.A., K. SALHOFER and S.R. THOMPSON (2017): Capitalisation of the SPS into Agricultural Land Rental Prices under Harmonisation of Payments. In: Journal of Agricultural Economics 68 (3): 710-726.

LEÓN, G. (2017): Turnout, political preferences and information: Experimental evidence from Peru. In: Journal of Development Economics 127: 56-71.

LOHR, S. (2009): Sampling: Design and Analysis. Cengage Learning, UK.

LUMLEY, T. (2010): Complex Surveys: A Guide to Analysis Using R. Wiley, Hoboken, New Jersey.

MITTER, H., M. KIRCHNER, E. SCHMID and M. SCHÖNHART (2014): The participation of agricultural stakeholders in assessing regional vulnerability of cropland to soil water erosion in Austria. In: Regional Environmental Change 14 (1): 385-400.

MORAWETZ, U.B. (2014): A concept for a randomized evaluation of agri-environment measures. In: Schmid, E. and S. Vogel (eds.): The Common Agricultural Policy in the 21st Century. facultas.wuv, Vienna, Austria: 113-130.

PALM-FORSTER, L.H., P.J. FERRARO, N. JANUSCH, C.A. VOSSLER and K.D. MESSER (2019): Behavioral and Experimental Agri-Environmental Research: Methodological Challenges, Literature Gaps, and Recommendations. In: Environmental and Resource Economics 73 (3): 719-742.

PENN, J.M. and W. HU (2018): Understanding Hypothetical Bias: An Enhanced Meta-Analysis. In: American Journal of Agricultural Economics 100 (4): 1186-1206.

R CORE TEAM (2018): R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available at http://www.R-project.org/.

SCHROEDER, L.A., A. GOCHT and W. BRITZ (2015): The Impact of Pillar II Funding: Validation from a Modelling and Evaluation Perspective. In: Journal of Agricultural Economics 66 (2): 415-441.

SHADISH, W.R., T.D. COOK and D.T. CAMPBELL (2002): Experimental and quasi-experimental designs for generalized causal inference. Houghton Mifflin, Boston.

SMITH, G., B. DAY, R. WELTERS, L. GILFOYLE and R. REYNOLDS (2019): Cash for clean water: a field experiment of payments for ecosystem services to improve water quality in the UK. Contributions to the Annual Conference of the European Association of Environmental and Resource Economists 2019 in Manchester: 32.

TERWAN, P., J.G. DEELEN, A. MULDERS and E. PEETERS (2016): The cooperative approach under the new Dutch agri-environment-climate scheme. Background, procedures and legal and institutional implications. Ministry of Economic Affairs, The Hague, The Netherlands.

THOYER, S. and R. PRÉGET (2019): Enriching the CAP evaluation toolbox with experimental approaches: introduction to the special issue. In: European Review of Agricultural Economics 46 (3): 347-366.

WIENER ZEITUNG (2017): 37 'Milchrebellen' ohne Liefervertrag. Wirtschaft Österreich - Wiener Zeitung Online. Available at https://www.wienerzeitung.at/nachrichten/wirtschaft/oesterreich/885613_37-Milchrebellen-ohne-Liefervertrag.html, accessed 16 March 2018.

Acknowledgement

We thank Michael Eder, Mona Friedrich, Josef Hambrusch, Otto Hofer, Stefan Kirchweger, Eva Krickler, Siegbert Linder, Hermine Mitter, Andreas Niedermayr, Melanie Ollinger, Thomas Resl, Martin Schönhart and Michael Weichselbaumer and the participants of the REECAP workshops in Montpellier, Angers and Vienna and the anonymous reviewers for their valuable comments. In addition, we thank Andreas Reindl for helping us to retrieve the data from the IACS database.

Contact author:
DR. ULRICH B. MORAWETZ
University of Natural Resources and Life Sciences Vienna
Feistmantelstr. 4, 1180 Vienna, Austria
e-mail: ulrich.morawetz@boku.ac.at

Appendix

A1. Certainty Follow-up

After asking for acceptance of (up)RCTs and after asking how much hay farmers would produce if they took part in an upRCT we asked participants to rate the statements 1) 'The thought experiment was easy to imagine' and 2) 'The percentage of hay was easy to estimate' (see Table A1.1). Using only those respondents who selected 'Agree' to the first statement we re-estimated the results from Table 3, see Table A1.2. Using only those respondents who selected 'Agree' to the second statement we re-estimated the results from Table 5, see Table A1.3. The difference in the results is small enough to suggest that difficulties in imagining the thought experiment and estimating the hay production would not undermine the main conclusions of our results.

Table A1.1. Responses to debriefing questions

	Agree	Rather agree	Rather not agree	Don't agree
The thought experiment was easy to imagine	57.46%	32.58%	8.11%	1.85%
The percentage of hay was easy to estimate	67.58%	24.80%	6.02%	1.61%
It was a pleasure supporting research on evaluation	64.29%	28.17%	6.02%	1.52%

Source: own calculations

Table A1.2. Acceptance of the RCT and the upRCT of respondents who considered the thought experiment as 'easy to imagine'

	RCT presented first	upRCT presented first	All
Acceptance RCT	27%	20%	23%
Acceptance upRCT	56%	33%	43%

Compare to Table 3 from the main text. Number of respondents: 716 (some respondents did not reply to all questions).
Source: own calculations

Table A1.3. Percentage of hay production for all respondents who considered it easy to estimate the percentage of hay production

	RCT presented first	upRCT presented first	All
RCT			
Mean	93%	95%	94%
Median	100%	100%	100%
25th percentile	100%	100%	100%
upRCT			
Mean	95%	95%	95%
Median	100%	100%	100%
25th percentile	100%	100%	100%
Difference upRCT-RCT			
Mean	2.1 ***	-0.1	0.9 **

The stated average treatment effect on the treated is 100% minus the stated hay production. Compare to Table 5 from the main text.
Source: own calculations

A2. Post-stratification

Our sample is not representative with respect to the farm characteristics utilised agricultural area, livestock units, agri-environment payments and 'refrain from silage' payments. To investigate if it makes a difference when we weight respondents according to the number of farms in the population they represent, we estimate our main results using post-stratification (LOHR, 2009: 342). Post-stratified estimates are approximately unbiased if within each post-stratum 1) each unit has the same probability of responding, 2) the response propensity is the same for every unit, or 3) the response is uncorrelated with the response propensity (LOHR, 2009: 343). Whether one of these requirements is fulfilled is untestable. Additionally, as a rule of thumb at least 20 responses per post-stratum are recommended and the response rate for each group should be 50% (LOHR, 2009: 343).

Table A2.4 shows the post-strata used. The farms are grouped as follows: whether the 'refrain from silage' payment is above or below the median of 902€, whether the farm is producing milk (payment for milk producers is 150€ per hectare instead of 80€ per hectare), whether the farm is an organic farm (farmers of organic farms may have a different attitude towards certain farm management practices). In total, this resulted in eight different post-strata where one respondent represents between 4.68 and 12.74 farms. As we have responses from 11% of the population, achieving the recommended representation factor of 2 is not feasible here. We use the R package 'survival' to derive the post-stratified weights and standard errors (LUMLEY, 2010).

Table A2.5, Table A2.6 and Table A2.7 show the re-estimated results from Table 3, Table 4 and Table 5 using post-stratified values. The comparison reveals that weighting observations does not substantially change the results.

Table A2.4. Strata used for post-stratification

'Refrain from silage' payment	Milk producer	Organic farm	Farms in population	Responding farms in sample	Farms represented by respondent
up to 902 €	No	No	3,051	241	12.66
above 902 €	No	No	244	21	11.62
up to 902 €	Yes	No	726	57	12.74
above 902 €	Yes	No	2,763	248	11.14
up to 902 €	No	Yes	1,475	197	7.49
above 902 €	No	Yes	449	96	4.68
up to 902 €	Yes	Yes	259	26	9.96
above 902 €	Yes	Yes	2,054	364	5.64

Population values were calculated from the IACS database.
Source: IACS database (Federal Ministry for Sustainability and Tourism, Austria; Agrarmarkt Austria) and own calculations

Table A2.5. Post-stratified acceptance of the RCT and the upRCT

	RCT presented first	upRCT presented first	All
Acceptance RCT	28%	18%	23%
Acceptance upRCT	52%	32%	42%

Compare to Table 3 from the main text.
Source: own calculations

Table A2.6. Post-stratified reasons for accepting the RCT and the upRCT

RCT:		upRCT:	
Accept because …		**Accept because …**	
… it is important prove the effect of the measure	46%	… it is important prove the effect of the measure	48%
… I can easily forgo the payment (e.g., because it is so low)	34%	… it results in an advantage for me	35%
Other reasons	29%	Other reasons	26%
Do not accept because …		**Do not accept because …**	
… it is unfair	52%	… it is unfair	31%
… it results in a disadvantage for me	57%	… it doesn't make sense to pay unconditional premiums	71%
… I have counted on the payment (e.g., for investments)	51%		
… I am generally against checking the effect of measures	4%	… I am generally against checking the effect of measures	3%
Other reasons	19%	Other reasons	19%

Note: More than one answer was possible
Compare to Table 4 from the main text.
Source: own calculations

Table A2.7. Post-stratified percentage of hay production

	RCT presented first	upRCT presented first	All
RCT			
Mean	90%	93%	92%
Median	100%	100%	100%
25th percentile	99%	100%	100%
upRCT			
Mean	93%	94%	94%
Median	100%	100%	100%
25th percentile	100%	100%	100%
Difference upRCT-RCT			
Mean	2.6 ***	1.2 **	1.9 ***

Note: for the mean difference from zero, *** represents 1% significance and **5% significance
The stated treatment effect on the treated is 100% minus the stated hay production. Compare to Table 5 from the main text.
Source: own calculations

A3. Sample Restricted to those Accepting RCT/upRCT

Independent of whether they did or did not accept the measures when an RCT/upRCT is applied, we asked participants to tell us how much hay they would produce in the RCT and the upRCT thought experiments (as a percentage of the total mowing material). In Table A3.8 we restricted the sample to observations where the respondent would accept the RCT or the upRCT. Comparing Table 5 to Table A3.8 reveals that the difference is not substantial.

Table A3.8. Percentage of hay production for respondents who would accept to participate in an upRCT (n=275) or an RCT (n=530)

	RCT presented first	upRCT presented first	All
RCT			
Mean	91%	91%	91%
Median	100%	100%	100%
25th percentile	100%	100%	100%
upRCT			
Mean	91%	91%	91%
Median	100%	100%	95%
25th percentile	99%	90%	100%
Difference upRCT-RCT			
Mean	2.4 ***	-0.3	1.6

Note: for the mean difference from zero, *** represents 1% significance and **5% significance
The difference in the last line is calculated using those who would accept both (n=179). The stated average treatment effect on the treated is 100% minus the stated hay production. Compare to Table 5 from the main text.
Source: own calculations