SUMMARY[1]

Introduction

In all OECD countries, policies are increasingly being implemented with the intention of improving the environmental performance of agriculture. Agricultural policies themselves also have important environmental impacts. But to what extent are the agri-environmental policies being evaluated to determine whether they are achieving their objectives and doing so in efficient ways? The purpose of the Workshop on Evaluating Agri-environmental Policies was to review the efforts underway in OECD member governments to evaluate their agri-environmental policies – to both share experiences and to assist the OECD deepen its analysis. This publication closely follows the structure of the Workshop.

Part I provides an introduction to the issue by considering why OECD and its member countries are interested in evaluating agri-environmental policies, which factors need to be taken into account and what methodologies could be used when undertaking evaluations. The nineteen country contributions are organised as chapters in Parts II-VII according to the type of agri-environmental policy measure evaluated: payments, taxes, regulatory requirements, advisory and institutional measures, and policy mixes.

Three specific objectives on which the Workshop focused were: (a) the tools and methods used for evaluation; (b) the cause-effect linkages between policies, farm management and outcomes; and (c) the conclusions on the effectiveness and/or efficiency of the policy measure. During the Workshop, three invited experts presented their views and assessments of what has been learnt with respect to one of the three objectives. The Workshop concluded with a panel discussion on the implications of the findings for the OECD's on-going work in monitoring and evaluating agri-environmental policies.

This *Summary* is organised around the three specific Workshop objectives, discussing some of the answers provided by the experts and the Workshop participants. The main recommendations for consideration by the OECD in thinking about its future work are also outlined, which provide additional insight into the issues raised. Further to these comments, valuable contributions by the Workshop discussants can be found at the beginning of each of the seven parts. Finally, to make the material contained within the country contributions more accessible, the *Summary Annex* contains one-page summaries for Chapters 4-22 (excluding Chapter 18 on community-based programmes in Australia), answering a checklist of questions developed to guide contributors.

---

1.   The Summary has been prepared under the responsibility of the OECD Secretariat and does not necessarily reflect the views of the OECD member countries or participants at the Workshop.

What have we learnt about the approaches, tools and methods used for the evaluation of agri-environmental policy measures?

A wide variety of different methodologies can be used to evaluate agri-environmental policies. It was generally considered that these should be seen as complements rather than substitutes, although some methods provide more information than others. It was also recognised that at the end of the day, the policy maker is required to make a judgement based of the evidence available, however incomplete that may be. Whatever methodology used, evaluators need to be transparent about the assumptions they use, and sensitivity analysis could be utilised to assess the most important assumptions.

Both *ex ante* and *ex post* evaluations have been used in the policy development process. For example, the results of the recent mid-term evaluation of agri-environmental programmes under the Rural Development Plans in the European Union member states played an important role in shaping the European Commission's recommended revisions to the Rural Development Programme.

Effective evaluation requires an inter-disciplinary approach, including linking economic modelling and biophysical process models. The contributions from Sweden, Switzerland and the United States Conservation Reserve Program (CRP) showed this. Further, stakeholder involvement can improve not only the evaluation but the success in implementing the policy.

Many of the evaluations are undertaken within governments. This may cause some bias in the results, perhaps illustrated by the fact that few negative evaluation results were presented at the Workshop, the Greek and Dutch contributions being notable exceptions.

One of the common limitations of the evaluations, noted in many instances by contributors themselves, was that the goals and objectives of agri-environmental policies are often not explicit, specific or measurable. Better articulation of both the explicit and implicit objectives is needed if a good evaluation is to be made.

A major challenge is to undertake the evaluations in a timely manner, yet the time spans involved vary greatly. On the one hand, deadlines set by policy makers are shortening and the speed of societal change is increasing. On the other, the environmental outcomes are part of natural processes which often take a long time for the effects to flow through.

While spatially distinct data and tools are needed to evaluate environmental effectiveness, the lack of good quality data was often noted. A number of countries such as Australia and France are explicitly incorporating monitoring and data collection into policy design to assist future evaluations, which are positive developments.

In addition, baseline and additionality issues are critical to good assessments, *i.e.* to establish what farmers would have done in the absence of specific policies. This is important because many policies interact to influence environmental outcomes. However, there are difficult to separate, and many studies have poorly defined baselines and monitoring data to do this. Surveys can be helpful and the German approach of statistically comparing between "similar farms" was considered to be very useful. Good examples of contributions that tried to disentangle the effects of the specific policy measure are provided by the Dutch manure and the US conservation compliance evaluations.

A number of different approaches have been used to assess environmental quality changes, including: expert opinion (*e.g.* buffer strips in Finland); statistical studies or actual observations (*e.g.* nitrogen pollution in Denmark); GIS data (*e.g.* Italy); research literature on environmental effectiveness of practices (*e.g.* Canadian Shelterbelt Program); and biophysical process models

(*e.g.* the Planetor model in Umbria, Italy and SOIL in Sweden). Models are commonly used in *ex ante* analysis but are now increasingly used in *ex post* evaluation.

Few studies include the evaluation of economic efficiency. Good cost estimates are necessary. However, many studies simply use programme costs and these do not reflect the full social costs of the measures. Further, relatively little benefit valuation (in monetary terms) are undertaken, the evaluation of the US CRP being one of the few examples. One possible method that could be useful is the benefits transfer method used in the Canadian shelterbelt study.

How has our understanding of the linkages between policies, farm practices and environmental outcomes improved?

The importance of understanding this linkage was stressed time and again. Evaluations should not only consider whether a policy objective has been met, but also to understand something about why and how the outcome was achieved – or not. This understanding is crucial to developing appropriate policy measures and to improving them. Evaluations need to consider: (a) the interplay of participants in the policy system; (b) administration, compliance and transaction costs: and (c) management adaptation. In this respect, many believe that these "soft effects" were not given the prominence that they merited in the OECD evaluation criteria.

The country contributions contain a number of different methods for measuring behavioural change. Some studies include information on changes in input use, for example, amounts of reduced pesticide application and treatment frequency in Denmark and Norway. Others include information on the number of farmers entering the programme, for example the UK Entry Level Stewardship programme, and the number of enrolees in the CRP. Farm surveys were used in others – such as in Belgium where farmers were asked if they would have started the practice without support. But how these link back to the policy and forward to the environmental outcome are less well analysed.

What insights have we gained into the environmental effectiveness and economic efficiency of agri-environmental policies?

The country contributions cover a good range of both the various policy measures and objectives of agri-environmental policies. On the policy side, there were a large number (ten) evaluating the provision of payments. However, the two contributions analysing regulations probably under represent the importance of this group in relation to the variety of measures actually impacting on farmers. This may reflect the greater requirement to evaluate policy measures associated with budgetary expenditure.

In terms of objectives, the majority of policies focus on limiting the harmful environmental effects of agriculture, in particular reducing water pollution from nutrients and pesticides, and the prevention of erosion. Fewer policy measures evaluated had a specific focus on biodiversity and landscape. This may reflect the existence of relatively fewer clear and precise indicators to measure progress in this area.

The evaluations tend to focus more on the environmental effectiveness of policy measures rather than on their economic efficiency. In terms of effectiveness, the evaluations generally conclude that the policy measures have been producing positive results. This appeared to be particularly so when clear targets or objectives had be set. In almost all cases, the contributions showed that progress was been made at a slower rate than initially anticipated.

However, there are some important issues about how effectiveness is measured. The evaluations illustrate the difficulty of establishing the counterfactual – what would have happened in the absence of the policy measure. This problem is compounded in the agricultural sector by the existence of many forms, and often at high levels, of agricultural support policies. There were only very few studies – Sweden and Finland being examples – which discussed the effects of such policies on the effectiveness and efficiency of the agri-environmental measures examined. A question was also raised as to whether there was a benchmark or standard that was trying to be reached, such as an appropriate level of soil erosion in the United States, or form and scale of agricultural landscape in the European Union.

The evaluation examples show that regulations and taxes have been effective in meeting their objectives but some questions remain about their use. For example, in order to be effective regulations need to be enforced – what is the cost of this for the government, and what are the compliance and administration costs imposed by regulations on producers? Taxes also appear to be relatively efficient, but why have countries not utilised this approach? The fact that these questions remain unanswered partly illustrates the lack of analysis regarding the economic efficiency of policies.

Recommendations for future OECD work

During the Workshop, a number of suggestions were made regarding work that could be undertaken by the OECD and in member countries in relation to evaluating agri-environmental policies. Suggestions included the following.

1. *Encourage the use of statistical methods for analysing the cause and effect linkages between policies and environmental outcomes* – a meeting of experts to discuss statistical methods and modelling tools was held in July 2005.

2. *Refine and further develop the OECD set of agri-environmental indicators* – this work has been valuable in assisting member countries develop their own indicators, an essential part of the evaluation process. Effort needs to focus on improving the quality of the data used in the indicator calculations and on using them in policy-relevant analysis.

3. *Undertake a Cost Benefit Analysis of data collection for evaluation* – a successful evaluation requires appropriate data, but this is often undervalued by policy makers and often few resources are made available to undertake this before, during and after the implementation of the programme. Participants believed that the cost of monitoring and evaluation was generally small in relation to the total budgetary costs of the programme, and provided considerable benefits to the policy making process. Such a CBA would strengthen the argument for developing appropriate monitoring efforts.

4. *Clearly identify the criteria used and the assumptions made when undertaking evaluations of agri-environmental policies* – the focus should not be on apportioning blame for failure but on assisting the development of appropriate policy measures.

5. *Develop and maintain a repository of agri-environmental evaluations that have been undertaken* – this would provide a reference tool for member countries as they further develop and refine their own evaluative techniques. In doing so, the Secretariat will need to develop an appropriate typology or classification system so that the appropriate evaluation examples can be easily located. A link could be drawn with the OECD *Inventory of Policy Measures Addressing Environmental Issues in Agriculture* (www.oecd.org/agr/env) and with

the list of questions developed to guide the preparation of country contributions to the Workshop.

6. *Provide an external review function to "evaluate the evaluations"* – based on the principle that the more external review provided, the better evaluations are likely to be. This could be done either during (*e.g.* comment on drafts) or after the evaluation process.

7. *Assist with the development and organisation of interdisciplinary approaches to agri-environmental policy evaluation* – recognising that the fuller the evaluation of such policies, the greater the range of participants required, including economists, natural scientists, sociologists and other disciplines.

8. *Build a "generic framework"/"guidelines for best practice" of issues and questions that should form part of an agri-environmental policy evaluation* – drawing on the work identified in the previous three bullet points. This could also identify, and perhaps prioritise, from both theory and experience, the key issues that need to be evaluated. For example, when undertaking a CBA, identifying the main cost and benefit variables to include such that they can contribute the most to the estimate of the total costs and benefits.

9. *Guide the valuation of environmental benefits* – this is required to assist in the progression of evaluations from a focus on environmental effectiveness to economic efficiency. This could include a review of valuation methods and/or the identification of values already placed on benefits within evaluations. These values may or may not be transferable to other countries, but would provide an indication of how results compare with other valuations for similar environmental benefits elsewhere.

10. *Host a similar Workshop in three years' time to consider the progress made by OECD member countries in evaluating their agri-environmental policies* – participants saw benefits in learning from others and from being able to "benchmark" their progress.

SUMMARY ANNEX

ONE-PAGE SUMMARIES OF COUNTRY EVALUATIONS

Chapter 4. Evaluation of Agri-environmental Measures in Flanders, Belgium

What policy measure(s) were evaluated?	Twelve agri-environmental measures providing payments to farmers for undertaking certain environmentally beneficial practices, provided under the Flanders Rural Development Plan 2000-06.
Who did the evaluation?	External evaluator (consortium of consultants and university) in close consultation with the administration.
How soon after implementation?	Evaluation was held in 2003, and concerned the 2000-02 implementation period.
What are the specific objectives of the policy?	Most of the measures have the objective of making farming practices more environmentally friendly, although some measures are also directed at nature and landscape management. The two most important measures, by area and budget, are payments to reduce fertiliser use and increase soil cover (to reduce nitrate pollution and erosion respectively).
What criteria were used to evaluate?	Progress of implementation, accuracy of support, soft effects (*e.g.* farmer attitudes).
What tools and methods were used?	Analysis of budget, discussion with administrators, farmer survey, interviews with experts, GPS mapping.
What baselines/scenarios were used?	"No policy situation" and "farmers not affected by the policy".
What methods were employed to disentangle the impact of policy measures?	None.
What were the main conclusions of the evaluation?	Measuring soft effects through a survey is very useful, certainly in the early stage of implementation when environmental data is limited. More specific targets should be introduced or improved for each measure.
What recommendations were made?	Measures are doing well in this early stage of implementation but already some adjustments could be made. Greater effort is needed by the administration to collect more and better quality environmental data.
What are the strengths and limitations of the evaluation methodologies used?	The farmer survey is very interesting but results should be carefully interpreted: *ex ante* or *ex post* will give different answers. It was also difficult to compare between the measures (some aim at public goods, other more at private goods).

Chapter 5. Evaluation of Agri-environmental Policies implemented in France from 2000-02 under the CTE Farm Contract Scheme

What policy measure(s) were evaluated?	Agri-environmental measures (AEM) implemented from 2000 to 2002 under the territorial farming contracts (CTE).
Who did the evaluation?	Eleven consultants chosen by tender for 21 contracts.
How soon after implementation?	The first five-year AEM contract began in 1999.
What are the specific objectives of the policy?	Primarily environmental, *i.e.* to help introduce or maintain farming methods that are compatible with the increased need to protect and enhance natural resources, landscape and biodiversity. Policies also have socio-economic goals (*e.g.* employment) and help to achieve territorial balance.
What criteria were used to evaluate?	Relevance, potential environmental effectiveness, potential economic efficiency, internal and external consistency, implementation (administrative management, costs etc.).
What tools and methods were used?	Analysis of achievements (*e.g.* location and characteristics of farms by AEM) with reference to background data. Interviews with those directly concerned by the implementation (*e.g.* local agricultural and environmental governments, farm advisory bodies) – approximately 550. Various types of surveys (*e.g.* face-to-face, telephone) of farmers (approx 1 200) both participating and not in AEM.
What baselines/scenarios were used?	The situation prior to the programme introduction.
What methods were employed to disentangle the impact of policy measures?	As it was a mid-term evaluation, estimates were made of the potential impacts based on the type of AEM implemented, geographic location etc.
What were the main conclusions of the evaluation?	Around 30% of French farmers and farmland is enrolled in the scheme. Environmental effectiveness has been mixed: maintaining amenities (*e.g.* biodiversity and landscape) in grazing and mixed crop/livestock farms, but working poorly in terms of reducing negative externalities (*e.g.* water quality). It has played an important part in raising environmental awareness among farmers.
What recommendations were made?	A number of recommendations, particularly to enhance the scheme's environmental effectiveness.
What are the strengths and limitations of the evaluation methodologies used?	*Strengths*: the large number of people involved allowed for a wide diversity of opinion. *Limitations*: data were often difficult to use and data sets were not exhaustive.

Chapter 6. The Implementation of Organic Farming: The Case of Peloponnese, Greece

What policy measure(s) were evaluated?	Organic Farming Scheme – providing per-hectare payments to support organic farming.
Who did the evaluation?	Consultant – European Enterprise Organisation Group (EEO) SA, on behalf of the Greek Ministry of Agriculture.
How soon after implementation?	Eight years – Organic Farming Scheme was first implemented in Greece under Regulation (EEC) 2078/92 in 1995.
What are the specific objectives of the policy?	Reduce pollution caused by agriculture; protect wild flora and fauna; protect biodiversity in rural ecosystems and the landscape; manage agricultural resources in a sustainable manner; protect public health and develop organic farming practices uniformly on a national level.
What criteria were used to evaluate?	Environmental effectiveness, based on the analysis of the EU common evaluation questions, and some elements of economic efficiency.
What tools and methods were used?	Questionnaire addressed to all the competent local authorities about the implementation of the scheme (mainly qualitative assessment). Statistical data regarding the implementation of the scheme (quantitative).
What baselines/scenarios were used?	The situation before the implementation of the scheme.
What methods were employed to disentangle the impact of policy measures?	Use of common evaluation questions (STAR Documents).
What were the main conclusions of the evaluation?	There is a constant interest from farmers in the scheme. However, a lack of sufficient personnel in the regions, combined with the lack of training and awareness of the farmers, is a hindrance to the successful implementation of the programme.
What recommendations were made?	The establishment of a network to distribute organic products is needed so that the market can play a greater role. Raising the awareness of the farmers, and training both farmers and agronomists who manage the programme in the regions would contribute to the successful development of organic farming in the area of Peloponnese and Greece in general.
What are the strengths and limitations of the evaluation methodologies used?	Lack of statistical data.

Chapter 7. The Evaluation of Agri-environmental Measures: A Survey of Different Methods used by Italian Regions

What policy measure(s) were evaluated?	Agri-environmental payments.
Who did the evaluation?	Universities, research institutes, private consultants.
How soon after implementation?	Three years.
What are the specific objectives of the policy?	Reduction of the use of chemicals, protection of biodiversity.
What criteria were used to evaluate?	Economic and environmental criteria.
What tools and methods were used?	Comparison of samples of farms, impact ratio indicators, models.
What baselines/scenarios were used?	Accounting data of farms applying and not applying the policy, located in homogenous agronomic and environmental conditions. G.I.S (contextual and administrative data). Administrative data.
What methods were employed to disentangle the impact of policy measures?	Test farms not applying the policy located in homogenous agronomic and environmental conditions.
What were the main conclusions of the evaluation?	The use of impact ratio indicators shows when target zones are not reached by the policy. Economic evaluation has given good information for better calculate incentives and increase cost effectiveness of the policy and on the sustainability of the change in the methods of production besides the economic incentive provided by the agri-environmental payment. Farm data has given information on changes in quantity and quality of chemical inputs.
What recommendations were made?	Relevance of monitoring and data collection. Need to integrate different methods and data sources. Need of interdisciplinary work. The evaluations have usually pointed out that the programmes succeeded in meeting their objectives.
What are the strengths and limitations of the evaluation methodologies used?	Difficult to estimate the environmental impact. Possibility of disentangle other variables effects when using different samples of farm data. The use of contextual and administrative data coupled with scientific data coming from literature or specific research can give good proxy evaluation at lower cost. Soil or water tests at farm level are very expansive and can be performed only at low scale level. To disentangle the effect of the policy from other factors test must also be done on comparable control farms. The crucial point is the availability of reference data and a monitoring system at local level. A detailed geographic information system is a precondition for the evaluating the impact of any environmental policy and it must be financed by public resources.

Chapter 8. Evaluating Agri-environmental Schemes in England

What policy measure(s) were evaluated?	The existing Environmentally Sensitive Areas (ESA) scheme and Countryside Stewardship Scheme (CSS), and the new Environmental Stewardship (ES) scheme. These provide payments to farmers for environmentally beneficial practices.
Who did the evaluation?	Three external consultants and the Department for Environment, Food and Rural Affairs (DEFRA).
How soon after implementation?	The ESA began in 1987 and the CSS in 1991. These were evaluated by two consultant and DEFRA over 2002-03. A pilot ES scheme was launched in February 2003 with the consultant evaluation completed by October 2003.
What are the specific objectives of the policy?	Primarily biodiversity (*e.g.* wildlife) and landscape.
What criteria were used to evaluate?	Environmental effectiveness and economic efficiency.
What tools and methods were used?	The environmental effectiveness evaluation analysed monitoring data and reviewed other external studies. The economic efficiency evaluation undertook a cost/benefit analysis, and examined payment rates and incentive options. DEFRA undertook a further literature review and public consultations on strengths and weaknesses of the two existing schemes and possible approaches to the new scheme. The evaluation of the ES pilot focussed on environmental effectiveness using a model and expert opinion.
What baselines/scenarios were used?	The economic evaluation tried to determine if the schemes were needed at all.
What methods were employed to disentangle the impact of policy measures?	None.
What were the main conclusions of the evaluation?	The environmental effectiveness study concluded that: ESS was particularly successful at conserving the historical environment but less at enhancing wildlife; CSS was successful for some birds (but not those dependent on a widespread habitat) but less at maintaining landscape and historical environment. The economic evaluation found high benefit/cost ratios. The study of the pilot ES predicted substantial environmental benefits.
What recommendations were made?	These evaluations were a major input into the design of the new Environmental Stewardship scheme which was launched in early 2005.
What are the strengths and limitations of the evaluation methodologies used?	*Strengths*: variety of methods (*e.g.* surveys, CBA), extensive environmental data sets, specific criteria to evaluate against. *Limitations*: time taken to feed back into scheme design.

Chapter 9. Alternative Approaches for Evaluating the Performance of Buffer Strip Policy in Finland

What policy measure(s) were evaluated?	3-metre and 15-metre buffer strips.
Who did the evaluation?	Environmental experts, scientists, economists.
How soon after implementation?	Field experiments began before implementation.
What are the specific objectives of the policy?	To reduce surface runoff of nitrogen, phosphorus and sediment. To promote biodiversity and landscape diversity by providing semi-natural habitats and linear landscape elements.
What criteria were used to evaluate?	Mainly environmental effectiveness but also to some extent economic efficiency.
What tools and methods were used?	Field experiments, field surveys, expert surveys, belief network modelling, economic analysis using farm-level financial data.
What baselines/scenarios were used?	*E.g.* surface run-offs of nutrients and sediment in the absence of buffer strips, species number in the normal (1 m) field edge (in the absence of buffer strips).
What methods were employed to disentangle the impact of policy measures?	Disentangling the effect of the policy measure is easy since in the absence of incentives buffer strips are not implemented as they result in loss of profits. However, in practice they are usually implemented in combination with other policies (*e.g.* fertiliser use restrictions and pesticide use regulations) which strengthen the environmental effectiveness of buffer strips. Farm level financial analysis was used to find out the effect of area payments on profitability of participating in the buffer scheme.
What were the main conclusions of the evaluation?	Buffer strips are environmentally effective policy. However, economic analysis revealed that the establishment of 15-m buffer strips is not profitable for farmer at current compensation level.
What recommendations were made?	Narrow (3-m) buffer strips perform well as an environmental cross-compliance measure and could be widened to 5 m in order to increase their environmental performance. Compensation level for wide (15-m) buffer strips should be increased or arable crop area payments should be decoupled to make buffer strip establishment more profitable for farmer.
What are the strengths and limitations of the evaluation methodologies used?	Spatial and temporal factors affect the performance of buffer strips in terms water quality protection and as a result field experiments should be conducted in different spatial and temporal settings. Expert surveys could be helpful in the absence of environmental outcome data. In terms of biodiversity the results from alternative evaluation techniques seems to be quite consistent whereas in terms of water quality protection results are more diverse.

Chapter 10. Assessing Long-term Impacts of Agri-environmental Measures in Germany

What policy measure(s) were evaluated?	Voluntary agri-environmental measures (AEM) like support for organic farming or other extensification measures, with co-funding of the EU according to Regulation 2078/92 and Regulation 1257/1999.
Who did the evaluation?	Institute of Rural Studies, Federal Agricultural Research Centre, Germany.
How soon after implementation?	More than 5 years (co-funded AEM were broadly introduced after 1992, the analysis includes data from 1989-2002).
What are the specific objectives of the policy?	Dominantly abiotic resource protection, *e.g.* reduction of nitrogen surplus, also maintenance of biodiversity and open landscapes. A problem for evaluation is the lack of precise and quantified targets.
What criteria were used to evaluate?	Environmental effectiveness (measured by changes in land use (grassland, intensive crops like forage maize, etc.) and intensity indicators (livestock density and milk production per hectare, monetary fertiliser and pesticide input per hectare, nitrogen surplus in kg N/ha)), and economic effects (productivity (cereal yields, milk per cow) and farm income).
What tools and methods were used?	Farm accounting data of identical farms for a time series, statistical cluster analysis for selection of similar farms with and without support for AEM, statistical analysis and comparison between these groups and over time ("with-without-comparison").
What baselines/scenarios were used?	Baseline is the group of farms without (or with little) support for AEM, based on empirical data.
What methods were employed to disentangle the impact of policy measures?	Cluster analysis for selection of similar farms with and without support for AEM.
What were the main conclusions of the evaluation?	Significant improvement of environmental performance of farms participating in AEM compared to non-participants, especially in organic farms. Also positive income effects were found. Obviously, incentives are needed for changes in farm management, leading both to environmental improvements and income effects.
What recommendations were made?	Long-term impacts of AEM can be evaluated using the presented method, which should be included in evaluations in future.
What are the strengths and limitations of the evaluation methodologies used?	*Strengths*: analysis of long-term effects, use of an empirical, largely un-biased reference group, statistical test based on large number of farms; *Limitations:* data availability is crucial, several environmental impacts dependent on farm management are not "visible" in farm accounts, *e.g.* in the fields of erosion and biodiversity.

Chapter 11. Sweden's Experience with Evaluating Agri-environmental Payments

What policy measure(s) were evaluated?	Agri-environmental payments for semi-natural pasture and for ley farming, in comparison to support payments for grazing animals (headage payment) and regional aid.
Who did the evaluation?	The Swedish Board of Agriculture in cooperation with the National Environmental Protection Agency and the National Heritage Board. Independent consultants were hired to do model simulations of the agricultural sector's reaction to the measures, and to calculate environmental effects of changes to the sector.
How soon after implementation?	Two years after implementation of the agri-environmental scheme.
What are the specific objectives of the policy?	Increased nature and cultural heritage values in the agricultural landscape, as well as reduced environmental load.
What criteria were used to evaluate?	Environmental effectiveness (nitrogen leaching, ammonia emission and pasture area) and efficiency in comparison with alternative measures.
What tools and methods were used?	An economic model for simulating farmers' reaction to policies combined with ecological models for analysing environmental impacts from an assumed change in agriculture. Supported by a farmer survey.
What baselines/scenarios were used?	Simulation of what the situation would have been in ten years time without the payment.
What methods were employed to disentangle the impact of policy measures?	Model simulation of potential situations with and without the measure.
What were the main conclusions of the evaluation?	All evaluated payments are related to grazing animals or to grassland. Hence they have a positive effect on the pasture area and on nature and cultural heritage values. At the same time, all the measures contribute to negative environmental load in the form of nitrogen leaching and ammonia emission. The most targeted form of support (support to pastures) showed a substantially higher effectiveness and efficiency in terms of obtained positive effects in relation to negative effects.
What recommendations were made?	Further targeting of agri-environmental payments should be analysed.
What are the strengths and limitations of the evaluation methodologies used?	The use of models is a strength, and necessary for isolating the effect of the policy from other effects, and for analysing relevant reference scenarios. Transaction costs were not included in the analysis, which is a weakness when analysing more targeted measures for which these costs could be high.

Chapter 12. Evaluation of Agri-environmental Measures in Switzerland

What policy measure(s) were evaluated?	Ecological payments, *e.g.* for extensive meadows; animal welfare payments; and cross-compliance, *e.g.* a balanced use of nutrients.
Who did the evaluation?	The studies were made by private and public research institutions, the Swiss Federal Veterinary Office and the Swiss Federal Office for Agriculture. An external expert integrates the results in a final report.
How soon after implementation?	The evaluation was started one year after implementation. It will be finalised in autumn 2005.
What are the specific objectives of the policy?	The objective of the evaluated measures is to promote an environmentally sound and animal friendly agriculture.
What criteria were used to evaluate?	Measurable criteria were defined in the areas biodiversity, nitrogen, phosphorus, pesticides, animal friendly housing systems and economic efficiency.
What tools and methods were used?	Appropriate to the different areas, a multiplicity of tools and methods were used.
What baselines/scenarios were used?	Baseline is the situation three years before the start of the policy measures. In several areas no data were available for this baseline. In these cases the researchers used the best alternative solution.
What methods were employed to disentangle the impact of policy measures?	A questionnaire based on a method of policy evaluation was developed; expert knowledge and results from other studies.
What were the main conclusions of the evaluation?	The ecological quality of certain ecological compensation areas does not differ from the ecological quality of the productive area. Other types of ecological compensation areas are especially favourable to biodiversity. Besides the type of habitats a minimal abundance, the interconnectedness of the areas and the proximity to habitats with rare species is also important for biodiversity. This favours measures that promote special ecological quality and the linking of semi-natural habitats. The input of N, P and pesticides has decreased substantially during the last decade but not recently. The present agri-environmental measures do not incite the farmers to further improve the efficiency of nutrient inputs.
What recommendations were made?	The final recommendations will be worked out in summer 2005.
What are the strengths and limitations of the evaluation methodologies used?	*Strengths*: examination at field, regional and national level; multidisciplinary; discussion of the main conclusions until all involved researchers can accept them. *Limitations*: costs; lack of baseline; and even with a huge effort it is not possible to examine all important interrelations.

Chapter 13. Conservation Policy and Agriculture in the US: Valuing the Impacts of the Conservation Reserve Program

What policy measure(s) were evaluated?	The Conservation Reserve Program (CRP) – a voluntary programme that offers long term rental payments, cost share assistance and technical assistance to farmers to convert farm land to vegetative cover, such as tame or native grasses, wildlife plantings, trees, filterstrips, or riparian buffers. The Program provides farmers an annual rental per hectare on land placed in a permanent cover under a 10 to 15-year contract; half of the cost of establishing a permanent land cover; and funds technical assistance and extension to help farms to adopt the Program.
Who did the evaluation?	The paper reviews a large number of evaluations of the environmental and economic/social impacts of the CRP, carried out by a range of researchers in academia and government.
How soon after implementation?	The CRP began in 1985. The evaluations covered in the paper date from 1990 to 2004.
What are the specific objectives of the policy?	To encourage farmers to plant long-term resource-conserving cover. The initial focus was on soil conservation (reducing erosion). Over time, other objectives such as habitat and water quality improvements; soil erosion reduction and productivity gains; air quality improvements and carbon sequestration water quality have been added. An environmental benefits index (EBI) is used to rank bids.
What criteria were used to evaluate?	Focused on studies that placed a monetary value on the environmental and economic/social impacts.
What tools and methods were used?	The studies use economic models to assess the changes.
What baselines/scenarios were used?	A variety.
What methods were employed to disentangle the impact of policy measures?	None.
What were the main conclusions of the evaluation?	Having reviewed the valuation exercise, the paper highlights two major data needs for future evaluations: (1) better measures of the actual environmental impacts of the CRP (*e.g.* land-use change is often used as a proxy for actual changes in bio-physical attributes) and (2) better measures of human interaction.
What recommendations were made?	None.
What are the strengths and limitations of the evaluation methodologies used?	Not applicable.

Chapter 14. The Use of Green Taxes in Denmark for the Control of the Aquatic Environment

What policy measure(s) were evaluated?	Economic instruments (taxes and tradable permits) for the control of nitrogen (N) and phosphorus (P) pollution.
Who did the evaluation?	Danish Ministry of Taxation.
How soon after implementation?	This was an *ex ante* evaluation conducted as preparatory work in 2003 for the Action Plan for the Aquatic Environment III.
What are the specific objectives of the policy?	To develop a simpler and more cost-effective system for N than the current regulatory regime and to consider a similar scheme for P.
What criteria were used to evaluate?	Environmental effectiveness (reduction in the N and P surpluses) and farmer adjustment costs.
What tools and methods were used?	An economic model was constructed to simulate the potential impacts of different tax rates and tradable quota systems. The model included empirically determined production functions and assumptions regarding nutrient absorption by crops and animals.
What baselines/scenarios were used?	The point of comparison was an estimate of N use, yield, number of animals etc. in 2001, the most recent period with sufficient data.
What methods were employed to disentangle the impact of policy measures?	It was assumed that as much as possible of the existing regulatory regime was abandoned, which means abolishing of the current norms and quotas for the allocation of N and the norms for the utilisation of N in livestock manure. However, certain EU regulations, *e.g.* livestock density, were maintained.
What were the main conclusions of the evaluation?	The closer the tax base is to the environmental problem (the N or P surplus) the lower the farmer adjustment costs. A given reduction in N surplus could be achieved with considerable lower costs to society as a whole, and particular to farmers, if the existing regulations were replaced by a tax on the nitrogen surplus especially if the tax was levied at a sector level or a national level.
What recommendations were made?	The Danish government decided to carry on the existing regulatory regime for use of nitrogen but decided to introduce a tax on mineral phosphorus. The Act has come into force from 1 April 2005.
What are the strengths and limitations of the evaluation methodologies used?	*Strengths*: even if the models are subject to a degree of uncertainty they are considered to be perfectly capable of ranking the various options in terms of tax level and cost. It also provides an indicator for the quantitative differences in environmental impacts and adjustment costs between the different options. *Limitations*: the analysis of different P taxes could not be based on theoretical models to anything like the same extent as a tax on nitrogen, since there is less information available including production functions.

Chapter 15. Taxes as a Tool to Reduce Health and Environmental Risk from Pesticide Use in Norway

What policy measure(s) were evaluated?	An area-based tax system banded by patterns of use and health and environmental criteria.
Who did the evaluation?	A working group at the Norwegian Agricultural Inspection Service.
How soon after implementation?	Four years.
What are the specific objectives of the policy?	The development of the new tax system was a follow-up of the Norwegian Action Plan for Pesticide Risk Reduction (1998-2002). The objective was to reduce use of the pesticides that represent the greatest risk to human health and the environment. This should be achieved by implementing a system where the products with the highest potential risk have the highest tax.
What criteria were used to evaluate?	Experiences with the new system from regulators, farmers and industry. Effects on farmers' costs. Trends in sales data and risk.
What tools and methods were used?	Results from farmer survey. Comments from regulators and pesticide companies. Assessment of the farmers' costs. Assessment of trends in sales data and risk indicators.
What baselines/scenarios were used?	In the assessment of sales data and risk trends, the years preceding the implementation were used as baselines.
What methods were employed to disentangle the impact of policy measures?	None.
What were the main conclusions of the evaluation?	The main conclusion so far is that the new tax system is a better system for human health and the environment than the old system based on import value. The evaluation was, however, performed too soon after implementation to draw clear conclusions about sales and risk trends.
What recommendations were made?	The recommendations concerned both the health and environmental criteria and the construction of the tax bands.
What are the strengths and limitations of the evaluation methodologies used?	*Strengths*: a wide range of methods employed. *Limitations*: data limitations (lack of time series, hoarding effects).

Chapter 16. The Regulation of Nutrient Losses in Denmark to Control Aquatic Pollution from Agriculture

What policy measure(s) were evaluated?	Action Plans I and II on the aquatic environment – the main nutrient-related measures are: mandatory fertiliser and crop rotation plans with limits on nitrogen (N) application, statutory norms on the utilisation of manure N, manure storage, application timing, and maximum livestock density. These have been reinforced by research, information campaigns and extension.
Who did the evaluation?	Various government agencies including the Danish Institute of Agricultural Sciences and the National Environmental Research Institute.
How soon after implementation?	Almost twenty years since the first regulations were introduced in 1985. Action Plan I began in 1987 and Action Plan II in 1998.
What are the specific objectives of the policy?	The target of both Action Plans was to achieve a 49% reducing of N-leaching from the mid 1990s level.
What criteria were used to evaluate?	Environmental effectiveness.
What tools and methods were used?	Farm-level indicators were N-surplus, N-efficiency and N-leaching. Action Plan I established a significant monitoring programme. The main aquatic indicators are the nitrogen content of water sampled at agricultural watersheds, rivers and coastal areas.
What baselines/scenarios were used?	As the requirements are mandatory for all Danish farms, a cross-farm comparison was not possible. A simple before (*i.e.* mid-1980s) and after (now) approach was used.
What methods were employed to disentangle the impact of policy measures?	None.
What were the main conclusions of the evaluation?	The measures have had a large effect in improving N-utilisation, reducing N-surplus and reducing nitrogen leaching at the farm-level. The impacts of these farm-level changes on the aquatic environment are slowing emerging. N-balances were a good indicator. Enforcement was important.
What recommendations were made?	The evaluation feed into the formation of Action Plan III, which, following the success of the policy measures in reducing nitrogen pollution has been extended to address phosphorus.
What are the strengths and limitations of the evaluation methodologies used?	*Strengths*: a long-term, country-wide monitoring programme provided evidence of aquatic changes. *Limitations*: it did not disentangle the effects of the various policy instruments used or estimate the costs, particularly on producers of regulations.

Chapter 17. Has Conservation Compliance Reduced Soil Erosion on US Cropland?

What policy measure(s) were evaluated?	Conservation compliance makes eligibility for many federal agricultural subsidy programmes contingent on implementation of an approved soil conservation systems on highly erodible cropland.
Who did the evaluation?	Roger Claassen, Economic Research Service, USDA.
How soon after implementation?	Conservation compliance was enacted in 1985. Producers had until 1995 to fully implement approved conservation plans. The study is based on data collected in 1982 and 1997.
What are the specific objectives of the policy?	To reduce soil erosion for the maintenance soil productivity, enhancement of water quality, and other purposes.
What criteria were used to evaluate?	Estimate the level of cropland soil erosion reduction that could be attributed directly to conservation compliance.
What tools and methods were used?	The National Resources Inventory (NRI) contains information on land use and land condition (including estimates of cropland soil erosion) for more than 800 000 points of land. These points have been sampled four times: in 1982, 1987, 1992, and 1997 allowing characterisation of changes in soil erosion. Using GIS techniques, data on federal payments to farms was merged with the NRI data to estimate soil erosion reduction on land subject to compliance.
What baselines/scenarios were used?	Numerical estimates were for soil erosion reductions likely to have taken place in the context of a conservation compliance-based soil conservation system. For these reductions to be additional, it would also be necessary to show that soil conserving technologies would not have been adopted in the absence of conservation compliance.
What methods were employed to disentangle the impact of policy measures?	Starting with an estimate of overall reduction cropland soil erosion, the NRI data was used to subtract components of erosion reduction not likely to have been prompted by compliance, *e.g.* land not subject to compliance, erosion reduction due to land use change (which) was not likely to have been caused by compliance, and land in farms not receiving payments subject to compliance.
What were the main conclusions of the evaluation?	As much as 25% of the decline in US cropland erosion could have been prompted by conservation compliance. However, soil erosion was reduced on all types of cropland, not just highly erodible land subject to compliance, suggesting that other factors such as technical change may have also contributed to erosion reduction.
What recommendations were made?	None.
What are the strengths and limitations of the evaluation methodologies used?	*Strengths*: simplicity and transparency. *Limitation*: the inability to more fully address the question of additionality (due to data limitations).

Chapter 19. The Canadian Shelterbelt Program: Economic Valuation of Benefits

What policy measure(s) were evaluated?	The Shelterbelt Program provides planting material (tree and shrub seedlings) free of charge, along with research, development and technology transfer activities related to shelterbelts.
Who did the evaluation?	Suren Kulshreshtha and Edward Knopf, University of Saskatchewan, with John Kort of the Shelterbelt Centre.
How soon after implementation?	The Shelterbelt Program was created in 1901.
What are the specific objectives of the policy?	To encourage the planting of trees and shrubs in the arid and semi-arid regions of the Prairie provinces to protect farmsteads, farm fields and public roads against wind and extreme cold weather.
What criteria were used to evaluate?	Environmental effectiveness (what social benefits are generated); economic efficiency (comparing benefits with Program costs).
What tools and methods were used?	A three step approach was used: 1) physical description of shelterbelts; 2) biophysical changes that generate various benefits (based on literature); 3) social valuation of benefits, further refined by using the method of benefit transfer (based on literature).
What baselines/scenarios were used?	The baseline situation was the absence of shelterbelts.
What methods were employed to disentangle the impact of policy measures?	The methodology used to assess social benefits significantly reduces any cross-effects with other programmes. No additional method was used to disentangle benefits.
What were the main conclusions of the evaluation?	Estimated social benefits ranged between CAD 105-600 million for trees planted during 1981-2001 (NPV, 1981). These social benefits are mainly from reduced GHG emissions/accumulation, reduced soil erosion and improved consumptive wildlife based recreational activities. Social benefits that could be estimated compare favourably with taxpayer costs. Estimated costs for the same period were between CAD 15 and 19 million. Many benefits could not be estimated because of the lack of data or methodology
What recommendations were made?	Recommendations were limited to how to improve future research work. No recommendation on Program design/implementation. Efforts should be made to collect data that could improve the evaluation of the Program. Availability of data was a significant limitation to conduct this evaluation. Future research of this nature should be conducted by an interdisciplinary team
What are the strengths and limitations of the evaluation methodologies used?	*Strengths*: comprehensive scope of assessment to take into account all different pathways of social benefits, and innovative method (three-step approach). *Limitations*: data deficiencies on bio-physical changes related to the planting of shelterbelts and economic valuation of benefits.

Chapter 20. Evaluation of the New Zealand Sustainable Farming Fund: A Work in Progress

What policy measure(s) were evaluated?	The Sustainable Farming Fund (SFF) provides funds to projects that are producer-led and representative of community interest. Funds are allocated to projects on a contestable basis, last for one to three years and are co-financed from private sector sources.
Who did the evaluation?	The *Early-Stage Evaluation* was undertaken by external consultants and MAF Internal Audit. *Evaluation 2004* was led by a private consultant company, with input from SFF staff.
How soon after implementation?	The *Early-Stage Evaluation* was conducted in late 2002, 18 months after implementation began. *Evaluation 2004* was completed two years later.
What are the specific objectives of the policy?	The objective of the policy is very broad: to support projects that will contribute to improving both the financial and environmental performance of the land-based sectors.
What criteria were used to evaluate?	The *Early-Stage Evaluation* task was to improve the financial, managerial and report systems. *Evaluation 2004* was more comprehensive and estimated the potential economic benefits.
What tools and methods were used?	The *Early-Stage Evaluation* adopted Bennett's Hierarchy as a methodological framework. *Evaluation 2004* also used this. Projects were identified according to type and the benefits produced. Potential economic benefits were estimated for seven project groups comprising 106 projects (out of a total of 257). Both evaluations conducted surveys and interviews.
What baselines/scenarios were used?	None.
What methods were employed to disentangle the impact of policy measures?	None.
What were the main conclusions of the evaluation?	The SFF was highly regarded by the rural community and agricultural sector. Community and stakeholder involvement created cohesion, increasing the likelihood of adoption. The type of projects had "matured", moving from a dominance of projects dealing with the transfer of existing knowledge to projects generating new knowledge. Potential economic and environmental benefits were identified but actual results are still to be realised.
What recommendations were made?	*Evaluation 2004* concluded that the key challenge was to develop further mechanisms and processes for post-project extension. Another evaluation will be carried out in three years time.
What are the strengths and limitations of the evaluation methodologies used?	*Strengths*: consistent framework used for the two evaluations. *Limitations*: only quantified potential economic benefits – environmental benefits still to be determined.

Chapter 21. Evaluation of Agri-environmental Policies in Japan

What policy measure(s) were evaluated?	Policy evaluation in MAFF is conducted for each policy area and for each policy programme. Agri-environmental policy areas subject to the evaluation are (a) environmental protection measures in crop and livestock and (b) promotion of biomass use. The agri-environmental measure "programme for resource recycling and co-operation between crop and livestock production" is being evaluation in JFY 2004.
Who did the evaluation?	MAFF officials. The results of the evaluation are then discussed by members of the Policy Evaluation Committee.
How soon after implementation?	The evaluation is conducted annually. The evaluated measures may be under implementation.
What are the specific objectives of the policy?	Varied.
What criteria were used to evaluate?	National target indicators are used to measure environmental effectiveness, effects of stakeholders' action and its extension. Economic efficiency evaluations are conducted for site-specific programmes.
What tools and methods were used?	Surveys are used as tools to evaluate policies. Site-specific programmes are evaluated with cost-benefit analysis.
What baselines/scenarios were used?	Where possible, trend estimates are used as baselines; otherwise, levels at the base year are used as baselines. Such baselines may imply the situation without the policies, making policy effects visible. No *ex ante* evaluation was made for the annual policy evaluation as a whole, but *ex ante* and *ex post* evaluations have been conducted for individual site-specific programmes.
What methods were employed to disentangle the impact of policy measures?	The effect of each policy measure is not separated, but the identified indicators show the results from various aspects of policy mixes.
What were the main conclusions of the evaluation?	The evaluation results are ranked A, B or C, and the overall evaluation with its results are discussed at the MAFF Policy Evaluation Committee, made up of non-MAFF officials.
What recommendations were made?	When a policy area is ranked C, relevant policy measures are examined for possible improvement, with proposal made for budgetary consideration.
What are the strengths and limitations of the evaluation methodologies used?	Strengths include the fact it was relatively easy to understand the effects of complex relevant policy measures. However, this approach is limited because the effect of each policy measure is not separated. No identification is made in terms of which target or indicator is most useful.

Chapter 22. Evaluation of the Dutch Manure and Fertiliser Policy 1998-2002

What policy measure(s) were evaluated?	1) Minerals Accounting System "MINAS", 2) system of Manure Transfer Contracts, 3) system of Production Quota, 4) Buying-up Scheme for Production Quota, 5) Nitrate Projects Action Scheme.
Who did the evaluation?	Four co-operating groups of research institutes.
How soon after implementation?	Measures were implemented in different years since 1997. This was the third evaluation since 1998.
What are the specific objectives of the policy?	Control national animal manure production and a balance between N and P inputs and outputs at the farm with the objective of reducing the loss of N and P to the environment.
What criteria were used to evaluate?	Achievement of environmental targets, effectiveness of policy instruments and efficiency of policy instruments.
What tools and methods were used?	Environmental monitoring, farm nutrient budgeting, Cost analysis on emission reduction, interviews with farmers and organisations responsible for implementation, surveys among farmers.
What baselines/scenarios were used?	None, the reference year was 1997.
What methods were employed to disentangle the impact of policy measures?	Correction of environmental data for effects of precipitation surplus and selecting sample sites with dominant agricultural influence. Quantification of effects of single policies on emissions, and reconstruction of emission trends without policy impact. Comparison of policy effects for subgroups of farmers.
What were the main conclusions of the evaluation?	Large reduction of in N and P losses to the environment but the environmental targets were not achieved everywhere. MINAS was very effective and efficient, particularly for dairy farmers. Production quotas are a simple, effective and efficient instrument to prevent growth of manure production, but Transfer Contracts were not. Buying up schemes are an effective but expensive instrument compared to feed efficiency measures, unless profits from selling vacant farm lots return to the public means.
What recommendations were made?	To lower loss standards for N and P. To maintain MINAS as a voluntary tool for farmers to provide evidence for increased nutrient efficiencies and too allow for some flexibility in enforcement of the new system of Application Standards.
What are the strengths and limitations of the evaluation methodologies used?	*Strengths*: a strong empirical basis; analysis of effectiveness based on a realistic approach of the complex cause-effect chain; unambiguous conclusions. *Limitations*: disentanglement of autonomic effects; insufficient account for buffer effects of phosphate; no quantification of ecological consequences of exceeding environmental quality targets.

Part I

EVALUATION DESIGN

TABLE OF CONTENTS