Case Study

Farm questionnaires for monitoring genetically modified crops: a case study using GM maize

Kerstin SCHMIDT^{1*}, Ralf WILHELM², Jörg SCHMIDTKE¹, Lutz BEISSNER^{2,3}, Wenke MÖNKEMEYER¹, Petra BÖTTINGER², Jeremy SWEET⁴ and Joachim SCHIEMANN²

¹ BioMath GmbH, Schnickmannstraße 1, 18055 Rostock, Germany

² Julius Kühn-Institut, Bundesforschungsinstitut für Kulturpflanzen, Institut für Sicherheit in der Gentechnik bei Pflanzen, Messeweg 11/12, 38104 Braunschweig, Germany

³ Current address: KWS Saat AG, Grimsehlstr. 31, 37574 Einbeck, Germany

⁴ Environmental Consultant, 6 Green Street, Cambridge, CB4 5JA, UK

Monitoring is a statutory requirement for the cultivation of genetically modified (GM) crops in the European Community. Questionnaires for farmers to report on observations of effects linked with the cultivation of GM crops can form a useful part of a monitoring regime. A questionnaire for GM maize (*Zea mays* L.) was designed, with questions focusing on potential effects related to the GM maize grown, as well as on background information about cultivation methods and on individual field situations. In this paper we present the methodological approach of the monitoring regime, the structuring of the data, and the contents and structure of the questionnaire. The statistical requirements and background for an appropriate evaluation and interpretation of the data are described. Results of interviews made from 2001 to 2005 are also presented. It is envisaged that this approach will be developed for monitoring other cultivated GM plants and traits, and may be applicable in monitoring certain non-farmed environments.

Keywords: genetically modified plants / GMO / GMP / maize / monitoring / general surveillance / data acquisition / farm questionnaires / monitoring character / influencing factor

INTRODUCTION

The precautionary approach of Directive 2001/18/EC (Official Journal of the European Communities, 17 April 2001, 30 July 2002, 18 October 2002) and the EC-Regulation 1829/2003 (Official Journal of the European Communities, 18 October 2002) requires an environmental risk assessment (ERA) of the impact of a genetically modified organism (GMO) and its use. The evaluation of environmental effects of the GMO includes evaluations of indirect, unexpected or delayed effects, and impacts of changes in agricultural practice. Post-market environmental monitoring (PMEM) of GMOs is also a requirement of the directives, and has the aim to further protect "human health and the environment" from potential harm, adopting the precautionary approach as a guide.

Hence a key component of every application for the introduction of a GMO is a post market environmental monitoring plan, which comprises:

• *general surveillance* for unanticipated adverse effects affecting human health and the environment that were not identified in the ERA. These adverse effects could include adverse consequences of long-term and large-scale use of the GMO which could not be tested for prior to release.

Potential adverse environmental effects must be reported to the competent authorities, and the results of PMEM can be used to inform risk managers when making decisions on whether to continue or cancel the marketing approval of a GMO.

Up to now, "environmental harm" (caused by adverse environmental effects) has not been clearly described from a legal point of view. Available definitions are very general. Therefore, there is no specific guidance on the parameters that should be monitored

[•] *case-specific monitoring* to investigate distinct hypotheses about potential adverse environmental effects caused by the GMO or its use, which have been identified in the ERA in a conclusive manner;

^{*} Corresponding author: kerstin.schmidt@biomath.de

Article published by EDP Sciences and available at http://www.ebr-journal.org or http://dx.doi.org/10.1051/ebr:2008015

Protection goals	Information/Observations made by farmers
Soil function/quality	Use of fertilizers, performance of plant growth, yield as expression of soil fertility
Plant health	Abundance and intensity of diseases, pests (beneficials), efficacy of crop protection programs/management
Sustainable agriculture	(Annual records of) cultivation practices; crop protection measures, weed man- agement, change in soil cultivation patterns; GM crop persistence (volunteers); unusual events
Biodiversity	Focus on known species in fields and vicinity; invasiveness and weediness of GM crops, weed spectra, pesticide management regimes, occurrence of beneficials
Animal health	Effects of feeding GM crops

Table 1. Protection goals related to agricultural activities and associated information/data collected by farmers that are of relevance for monitoring GM varieties (based on Wilhelm et al., 2003).

or the monitoring methods. However, there are proposals that monitoring requirements should be determined by the legally-defined environmental protection goals (Bund/Länder AG, 2003; Sanvido et al., 2004, 2005; Wilhelm et al., 2002, 2003). Protection goals related to agriculture are given in Table 1 and can also be derived from Annex II of Directive 2001/18/EC.

Farmers record a range of agronomic information, and are the most frequent and consistent observers of crops and fields. For example, they collect field-specific records of seeds, tilling methods, physical and chemical soil analysis, fertilizer application, crop protection measures, yields and quality. Additionally, farmers hold in their "farm files" historical records of their agricultural land and its management. These provide background knowledge and experience that can be used as a baseline for assessing deviations from what is normal for their cultivation areas.

In order to provide a more systematic approach to collecting monitoring data that allows statistical analysis of observations, Wilhelm et al. (2004) and EFSA (2006a, 2006b) proposed a system of questionnaires and interviews for farms cultivating GM crops. This approach represents a simple (and economic) method of collecting data for monitoring purposes. Farm questionnaires utilize first-hand observations of mostly impartial observers, and also exploit farmers' knowledge and experience of their local agricultural environments, comparative crop performance and other factors that may be influencing events on their land.

The environmental monitoring of GM crops should mainly focus on the cultivation area and its surroundings, since measurable effects are to be expected there first. Farmers are the best source of this kind of information since they continually observe their fields and can readily incorporate additional records of GM crops into their existing farm recording systems. The defined protection goals related to agriculture (Tab. 1) indicate which parameters should be recorded at the cultivation sites of the GM crops.

Wilhelm et al. (2002, 2003) proposed a science-based selection of parameters relevant to monitoring for GM maize based on ecological and agronomic knowledge. In this paper we describe a pilot study from 2001 to 2005, in which these parameters were tested and refined at a total of 186 cultivation areas of GM maize, and 183 cultivation areas of conventional maize. Questionnaires were distributed to farmers, and maize breeders assisted them with collecting data. The objective of the study was to define appropriate monitoring parameters, to develop a methodology for surveying and statistical analysis, and to test the applicability, practicality and data quality of questionnaires. The study also analyzed the statistical validity of the information obtained and concludes on the value of the methodology as a tool for general surveillance (Schmidt et al., 2006).

RESULTS

Definition of monitoring parameters and baselines

Monitoring parameters are derived from protection goals. They have to be suitable to detect emerging effects and be practical to record. For example, a comparative assessment of plant growth and development gives information on trait stability, soil function and fertility. The parameters that are recorded should be characters that most likely can show a detectable effect in correlation to the cultivation of the GM crop (= "*monitoring characters*", see Tab. 2).

It is important to determine scales for recording data. At first sight, a strict quantitative approach seems to give the most precision. However, agricultural systems

Protection goals	Topics	Monitoring characters
Soil function	Plant development	Germination
		Plant growth and development
		Yield*
Plant health	Diseases	Occurrence of diseases
		Ustilago maydis infestation
		Rust, Fusarium infestation
		Corn borer larvae on Bt maize*
	Beneficials	Occurrence of beneficials
	Pests	Occurrence of pests
	Game	Damage by game
	Plant development	Time of flowering
		Stalk/root lodging
		Harvest date
Sustainable agriculture	Herbicide performance	Weed populations, problem weeds**
		Crop safety of herbicides**
		Efficacy of herbicides**
		Reaction of Bt maize to herbicides*
	General	Unusual events
		Particularities during growth
Biodiversity	GM crop	Presence of GM crops in the vicinity
	Other populations	Weeds (abundance)**
		Presence of beneficials
		Data on pesticide applications (kind, frequency, amount)
	General	Extraordinary observations

Table 2. Definition of monitoring characters to be surveyed in the pilot study with farm questionnaires to observe the impact of GM maize on local agricultural environments.

* Only for corn-borer-resistant varieties.

** Only for herbicide-tolerant varieties.

are complex, with poorly defined baselines and thresholds, so many effects would not be detected quantitatively. In addition, measuring for quantitative data is timeconsuming and costly – and unrealistic, as farmers would not agree to do so.

The most appropriate baselines for monitoring GM crops are either data from previous cultivation of non-GM crops on the monitored land or comparison with conventional crop comparators grown simultaneously on adjacent land on the same farm or in rotation with the GM crop.

Therefore, for most monitoring characters, qualitative and comparative scales are considered to be most applicable, and methods for analyzing these qualitative data need to be developed. The empirical data gathered from farmers are qualitative values based on their comparison with non-GM cultivation. Therefore they are categorized in a way to enable farmers to express the tendency of deviations from non-GMO cultivation, *e.g.*: "better – normal – worse", "earlier – normal – later" or "less – normal – more". An adverse effect would be suspected if numerous answers differing from "normal" were obtained, but in the "worse" direction (*e.g.* "more diseases" or "less beneficials" or "worse germination"). It is important to offer at least three comparative categories for possible answers, even if only adverse effects are of interest. Offering only "equal" or "different" as possible answer categories would distort negative and positive effects of cultivation, and does not correspond to the needs of GMO monitoring. Additionally, it is necessary to ask for specifications of each deviation from "normal", since only then can a possible connection with GMO cultivation be detected.

Quantitative thresholds or minimum effect sizes have to be defined to test for significance. From agricultural experience and from the results of this pilot study, a response probability of 10% worse situations was considered as a threshold for suspecting negative GM effects, and was subsequently used in the statistical analysis.

However, agro-ecosystems and the defined protection goals are influenced by numerous and manifold factors that are interdependent. For example, changes in agricultural practices could lead to changes in farmland biodiversity, so GM plant cultivation is only one among many influencing factors. Therefore numerous background data on the environment (*e.g.* weather conditions) and soil characteristics of the site and the established cultivation practices (*e.g.* sowing date) need to be identified and recorded in order to determine the cause of observed variation. These data represent the expression of "*influencing factors*" or *covariates*, while GM cultivation is the "*test factor*" of monitoring (Tab. 3). In summary, this practical approach to GM crop monitoring proposes:

- to record parameters related to GM crops on a qualitative and comparative scale, where the frequencies of character values indicating an (adverse) effect may be compared to an appropriate baseline;
- to record factors that are not directly associated with GM crop cultivation, but could influence the monitoring characters related to GM crops, in order to allow analysis of causes of the observed effects.

Modeling the monitoring information and definition of the monitoring object

Statistical analysis has the task of detecting potential real effects within the monitoring characters, and to separate GMO effects from influences due to other environmental and cultivation factors. Consequently, monitoring has to take into account characteristics that might directly relate to possible negative effects of a GMO (*"monitoring characters"*), as well as other factors, which can influence their expression (*"influencing factors"*, see Tab. 3). Figure 1 provides an overview of the structure of the monitoring information and of the factors influencing the expression of the characters recorded. This structure is the basis for the statistical model for analysis of the monitoring data.

Fields represent the smallest unit for recording the monitoring characters in relation to other factors (*e.g.* GM *vs.* non-GM cultivation), and where the influencing factors, especially cultivation practices, are assumed to be equal. Therefore the field where a specific variety is grown is defined as the "*monitoring unit*". All parameters measured/observed should be related to these units, *i.e.* the data in a questionnaire should refer to the field in which a GM crop is cultivated.

Concept and performance of a farm questionnaire for gathering monitoring information

Methodologies established in medicine or social sciences for the development, structuring, use and analysis of questionnaires helped in developing questionnaires for GMO monitoring (Diekmann, 2006; Wengraf, 2006). Questionnaires need clear structures reflecting the aims of the underlying study. The aim of GMO monitoring is to gather factual information on any environmental effect of GM crop cultivation, not to judge the cultivation practices of farmers. Farm questionnaires – dependent on the monitored plant/event – should be divided into thematic parts so that there is a clear separation between the section on *comparative GMO-specific* monitoring characters and the section on *general* influencing factors.

Also, the wording of the questions in relation to possible answers has to be considered carefully. Farmers are asked for their evaluation of the situation compared with non-GM cultivation. The responses "same" or "normal" refer to the baselines at each site. The questions and the possible answers need to be formulated considering the reasonable range of values of monitoring characters and influencing factors, both with regard to data acquisition and statistical analysis. Accordingly defined categories for the answers should be offered to farmers. However, since not all possible categories can be foreseen, space for free remarks should be provided to either add any category or to register any unexpected event. Questions and answers should avoid leading respondents to provide "wrong" answers, and options like "uncertain" or "no comment" have to be possible.

It is useful to perform a pre-test to check the practicability and quality of the questionnaire. For the analysis, it is indispensable to clearly define the types of characters, to collect the data systematically from representative samples, and to use statistical methods with adequate power. Experiences with data of monitoring characters and influencing factors that were collected/recorded during the initial phase of the pilot study were incorporated into the further development of the questionnaire, to obtain appropriate data, and for the subsequent analysis.

Results of a pilot study gathered by farm questionnaires

A pilot project, studying the use of farm questionnaires for monitoring GM maize cultivated during the period 2001–2005, was conducted in a total of 373 field sites spread over Germany. Of these, 188 fields were planted with varieties carrying transgenic traits, 185 with conventional maize varieties. Six herbicide-tolerant varieties, harboring event T25, were grown on 62 sites, and 16 corn-borer-resistant varieties – either with the BT176

166

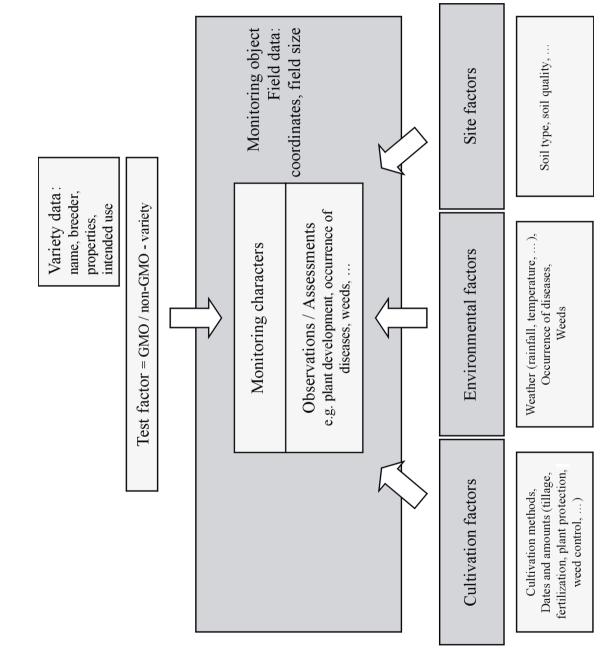


Figure 1. Structure of the monitoring information: monitoring characters and factors affecting their expression.

K. Schmidt et al.

Topic	Characters
Site	Soil type, category of field quality, humus content, crop rotation
Environment	Annual rainfall, corn borer infestation, degree of corn-borer infestation compared to previous year
Cultivation	Implements, tillage time, date of sowing, type of cultivation, planting method, weed management, treatment against corn borer
Variety	Variety name, features, intended use

Table 3. Topics and characters of background data (influencing factors) surveyed in the pilot study.

Table 4. Number of fields analyzed during the pilot study.

		GMO trait		Total
	Year of survey	Herbicide-tolerant	Corn-borer-resistant	
	2001	23	27	50
	2002	20	31	51
	2003	18	23	41
	2004		31	31
	2005		13	13
Total		61	125	186

or the MON810 event – were grown on 126 sites. Fortytwo conventional varieties were tested additionally. After quality control, the data of four fields (two GMO (one T25, one MON810), two conventional) were excluded from analysis because of poor quality. Therefore data from 186 fields with GM maize and 183 fields with conventional maize were analyzed further (Tab. 4).

Analysis of monitoring characters

The recorded monitoring characters (Tab. 2) mainly assessed the performance of GM varieties (e.g. plant development, occurrence of diseases, pests and beneficial (non-target) organisms) compared to non-GM maize cultivation. The answers were first analyzed descriptively for their frequencies, to form a certain pattern (Tabs. 5 and 6). The patterns for all monitoring responses show that for about 80–100% of the GM fields, the situations were assessed to be the same/normal or better. To evaluate whether the monitoring data directly indicated an adverse effect, the answers different from "normal" were assessed on a case-by-case basis, incorporating the specifications given by the farmers for those deviations. By doing this, cases with "positive" effects could be separated from cases with potential adverse effects. The answers "normal" and "better" were merged to create binomial characters ("normal/better" - "worse"), and used for testing the probability of "worse" cases against a threshold (Berensmeier et al., 2006).

From agriculture experience a baseline of 10% was fixed as an acceptable response probability for worse situations. It served as a threshold for an exact binomial test (null hypothesis H₀: probability P_{worse} \geq 10% against alternative hypothesis H_A: P_{worse} < 10%) in statistical analysis. The results are shown in Table 7. For most monitoring characters, the null hypotheses could be rejected on a significance level less than 0.01, *i.e.* with a high significance that there were no adverse deviations from normal or better situations. For a small number, the null hypothesis could not be rejected because of insufficient sample sizes, although no "worse" answers were recorded in these cases.

The frequencies of the monitoring character values were also analyzed for changes over 5 years (Fig. 2). Since monitoring was a continuous process over several years, this sequential approach is particularly suitable to analyze each year's data for any deviations, and to see whether there were tendencies emerging over the intervening years. For quick assessment of the frequencies, upper 95% probability tolerance limits were calculated. Like in process quality control, they give an upper value for the probability P_{worse} to be tolerated when comparing the data with a baseline threshold of 10%. From the data, an upper tolerance limit of about 13.5% was calculated (n = 183). Only in one case did data exceed the upper tolerance limit. From the background data, it could be reflected that this was due to a warm April, and therefore earlier sowing and germination.

Monitoring character	"positive deviation"	"standard situation"	"negative deviation"	valid answers
	n (%)	n (%)	n (%)	
Germination	9 (5.0)	168 (93.9)	2 (1.1)	179
Plant growth and development	27 (15.3)	145 (81.9)	5 (2.8)	177
Yield*	58 (63.1)	34 (36.9)	0 (0.0)	92
Occurrence of diseases	19 (11.7)	142 (87.7)	1 (0.6)	162
Ustilago maydis infestation	10 (9.4)	93 (86.9)	4 (3.7)	107
Time of rust appearance	1 (1.1)	89 (98.9)	0 (0.0)	90
Fusarium infestation	16 (16.2)	83 (83.8)	0 (0.0)	99
Corn borer larvae on Bt maize*	112 (92.6)	9 (7.4)	0 (0.0)	121
Occurrence of beneficials	2 (1.2)	162 (98.2)	1 (0.6)	165
Occurrence of pests	33 (19.9)	133 (80.1)	0 (0.0)	166
Damage by game	21 (12.6)	144 (86.2)	2 (1.2)	167
Time of flowering	23 (12.9)	144 (80.9)	11 (6.2)	178
Stalk/root lodging	44 (24.6)	133 (74.3)	2 (1.1)	179
Harvest date	26 (14.8)	138 (78.4)	12 (6.8)	176
Response of Bt maize to herbicides*	0 (0.0)	124 (100)	0 (0.0)	124
Noticeable events	13 (12.7)	84 (82.4)	5 (4.9)	102
Particularities during growth	4 (3.4)	102 (86.4)	12 (10.2)	118

Table 5. Pilot study results of the frequency analyses of monitoring characters.

n = Absolute numbers of entries, percentage in parenthesis.

* Only for corn-borer-resistant varieties (125 cases).

Table 6. Pilot study results of the frequency analyses of monitored herbicide characteristics.

Monitoring character	very good	good	critical	unacceptable	no statement	valid answers
	n (%)	n (%)	n (%)	n (%)		
Crop safety of Liberty Solo*	43 (89.6)	5 (10.4)	- (0.0)	-(0.0)	13	48
Crop safety of Liberty Plus*	10 (62.5)	2 (12.5)	4 (25.0)	-(0.0)	45	16
Crop safety of standard herbicide*	9 (52.9)	6 (35.3)	2 (11.8)	-(0.0)	44	17
Efficacy of Liberty Solo*	20 (41.7)	23 (47.9)	3 (6.3)	2 (4.1)	13	48
Efficacy of Liberty Plus*	12 (75.0)	4 (25.0)	- (0.0)	-(0.0)	45	16
Efficacy of standard herbicide*	5 (29.4)	11 (64.7)	1 (5.9)	- (0.0)	44	17

n = Absolute numbers of entries, percentage in parenthesis.

* Only for herbicide-tolerant varieties (61 cases).

Analysis of influencing factors

This type of data (Tab. 3) was mainly collected to describe possible influencing factors like soil quality, weather conditions, cultivation techniques, etc., to determine whether GM and conventional maize were cultivated under similar, comparable conditions. Descriptive analysis and subsequent tests (t-test, Chi²-test) showed a balanced distribution of the factor levels between these two groups (Tab. 8).

The mean values of the cultivation data did not differ as to the "Ackerzahl", the German rating system of soil quality, (GMP: 54.05 *vs.* conv. crop: 54.45) humus content (2.43 *vs.* 2.45) and annual rainfall (624 mm *vs.* 630 mm). There was no difference in soil texture for both groups; the maize was mainly grown on sandy loam. In data on cultivation and crop rotation, there was almost full correspondence between the two groups regarding pre-previous, previous, and succeeding crops in rotation.

No significant differences appeared regarding the details of the cultivation of maize (basic soil tillage, equipment, sowing methods, seeding dates, cultivation method, and weed treatment). Tillage was mainly conducted in

K. Schmidt et al.

Monitoring character	Estimated probability for	P -value ¹ for probability _{worse} ≥ 0.10
	"worse situation" in percent (%)	
Germination	1.12	0.00000
Plant growth and development	2.82	0.00024
Yield*	0.00	0.00006
Occurrence of diseases	0.62	0.00000
Ustilago maydis infection	3.74	0.01450
Time of rust infection	0.00	0.00008
Fusarium infection	0.00	0.00003
Corn-borer larvae on Bt maize*	0.00	0.00000
Occurrence of beneficials	0.61	0.00000
Occurrence of pests	0.00	0.00000
Damage by game	1.20	0.00000
Time of flowering	6.18	0.05095
Stalk/root lodging	1.12	0.00000
Harvest date	6.82	0.09536
Reaction of <i>Bt</i> maize to herbicide*	0.00	0.00000
Unusual events	4.90	0.05098
Particularities during growth	10.17	0.60017
Crop safety of Liberty Solo**	0.00	0.00636
Crop safety of Liberty Plus**	0.00	0.18530
Crop safety of standard herbicide**	0.00	0.16677
Efficacy of Liberty Solo**	4.17	0.12890
Efficacy of Liberty Plus**	0.00	0.18530
Efficacy of standard herbicide**	0.00	0.16677

Table 7. Pilot study results of the exact binomial test for the monitored characters.

* Only for corn-borer-resistant varieties.

** Only for herbicide-tolerant varieties.

¹ The *P*-value gives the probability for the estimated probability of "worse" answers if the null hypothesis (that the probability for "worse" answer is larger than 10%) is valid. *P*-values smaller than 0.05 indicate that it is very improbable that the null hypothesis is true, therefore it will be rejected.

Due to insufficient sample sizes of the monitoring characters "Crop safety" and "Efficacy" of Liberty Plus and of standard herbicide, the null hypothesis could not be rejected, although no "worse" answers were recorded.

autumn, and by mouldboard plough. Sowing was mostly by precision drilling (single-kernel) and weeds were controlled similarly by spraying in *Bt* and non-*Bt* varieties. Cultivation characteristics and conditions did not show statistically significant differences in direct comparison.

Relation between influencing factors and monitoring parameters

Although a direct comparison of influencing factors did not show any differences, a logistic regression analysis was performed to investigate the relationship between the monitoring characters and influencing factors. This would be important if influencing factors differed between GM and non-GM plant cultivation, and/or if monitoring characters showed an adverse effect, which might be explained by factors other than GM cultivation. Logistic regression effect coefficients might be estimated to assess the influence of the analyzed factors on the monitoring characters. Logistic regression leads to a model estimating the odds (ratio of the probabilities p/(1 - p), where in this case p is the probability of "worse" and therefore (1 - p) is the probability of "better/normal" answers) of binary distributed characters dependent on independent regressor variables, *i.e.* it calculates the odds of the monitoring character values ("worse" or "better/normal") dependent on the influencing factor values. The regression parameters for the influencing factors were analyzed for their significance, *i.e.* whether the

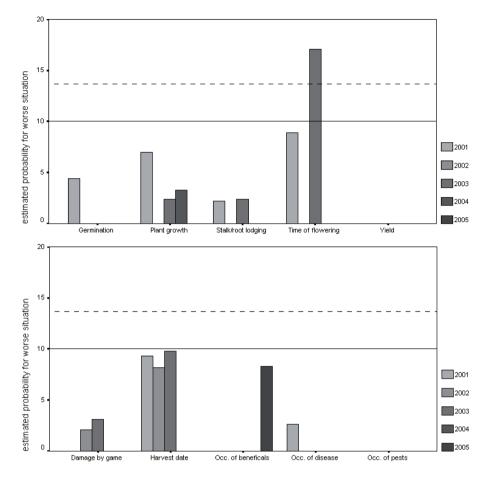


Figure 2. Yearly frequencies of monitoring characters of GM maize compared to conventional maize 2001–2005. The black line marks the 10% baseline threshold (less than 10% of "worse" or more than 90% of "normal/better" answers are accepted to be within the normal biological variation), the dashed line marks the 95% probability upper tolerance limit for "worse" answers.

factors have a significant influence on the monitoring character. We performed this logistic regression only for reasonable monitoring characters and influencing factors respectively. In our study, none of the influencing factors were identified as influencing any of the monitoring variables significantly.

Analysis of specific observations made on herbicidetolerant maize

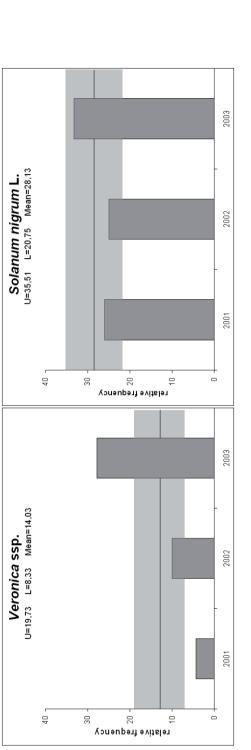
The herbicide applied to HT-maize cultivation areas was glufosinate ammonium, formulated as either 'Liberty Solo' or 'Liberty Plus'. These treatments were compared to the herbicides applied to fields of non-GM maize, which varied in active ingredients, dosage and rates.

Six HT-maize varieties were investigated from 2001 to 2003 at 18 to 23 sites in each year. At each field, the most frequent weeds were recorded (Bayer, 1992).

Altogether, the frequencies of 30 species or groups of weeds were documented. The most frequent weeds observed were goosefoot (*Chenopodium* ssp.), camomile species (*Matricaria* ssp.), cleavers (*Galium aparine*), black nightshade (*Solanum nigrum*), chickweed (*Stellaria media*), smart/knotweed (*Polygonum* ssp.) and millet species (*Setaria* ssp., *Digitaria* ssp.), *Panicum* ssp.).

Although only few fields could be investigated during this pilot study, because of the limited cultivation of HT maize, it was shown that an extended survey might document possible drifts of weed populations in HT maize cultivation regimes.

Unintended impacts/problems in weed or herbicide management can be analyzed by considering trigger values in weed frequencies, *e.g.*, upper or lower 90% probability tolerance limits were calculated (n = 61) for assessing weed development over several years (Fig. 3). They give upper and lower values for the frequencies of weeds over the years, and the mean frequencies were





K. Schmidt et al.

Character	GMO	Conventional	Significance
	Mean/Mode	Mean/Mode	$level^1$
Soil type	sandy loam	sandy loam	0.983
Category of soil quality	54.05	54.45	0.822
Humus content (%)	2.43	2.45	0.897
Crop rotation	cereal - cereal	cereal - cereal	1.000
Annual rainfall (mm)	624.34	630.44	0.679
Implementation	tilling	tilling	0.718
Tillage time	autumn	autumn	0.984
Date of sowing (2001)	05 May	05 May	0.968
Date of sowing (2002)	30 April	30 April	0.820
Date of sowing (2003)	27 April	27 April	0.994
Date of sowing (2004)	23 April	23 April	0.965
Date of sowing (2005)	29 April	29 April	0.979
Type of cultivation	conventional	conventional	0.794
Planting method	single kernel	single kernel	0.745
Weed management Bt	chemical	chemical	0.942

Table 8. Pilot study results of the descriptive data analysis of the influencing factors to compare the conditions between GM and conventional cultivation.

¹ The significance level gives the probability for the test statistic if the null hypothesis (that the means/modes are equal) is valid. Significance levels smaller than 0.05 indicate that it is very improbable that the null hypothesis is true, therefore it will be rejected; for large significance levels the null hypothesis will not be rejected (see text).

taken as the center points of the tolerance intervals. The data showed some slight undershooting of the lower or overshooting of the upper tolerance limits in some years, but no significant effects. In long-term analyses, cumulative overshooting might indicate environmental impacts.

Farmers were also asked to assess the selectivity and efficacy of weed control by Liberty Solo/Plus in comparison to standard herbicides. These data may also give information on effects on weed populations. The estimated probability patterns of the answers are illustrated in Figure 4, and were compared by Chi² tests. The crop safety of 'Liberty Solo' was significantly better than that of standard herbicides (sign. P = 0.001), the crop safety of 'Liberty Plus' didn't differ from that of the comparator herbicides (sign. P = 0.43). The analysis of weed frequencies in HT maize recorded in the questionnaires may also provide data of value for monitoring herbicide effects for stewardship and management purposes.

Analysis of specific observations made on maize with corn borer resistance

In addition to data on corn borers (target organisms), the yield, herbicide applications and variety-specific characteristics of the *Bt* maize were collected, in order to record possible interactions with agricultural cultivation practice. A total of 125 cultivation areas of *Bt* maize were

analyzed. 43.8% of the farmers judged corn borer attacks in maize to be increasing, 46.1% regarded infestation levels as stabilized.

The incidence of viable corn borers on Bt maize was recorded as a first step in studying resistance development. However, corn borers were only found on nine Btmaize sites, and larvae were found on less than 1% of the plants at these sites. It was later determined that, due to the quality of the seeds (for technical reasons), up to 2% of the plants were not expressing the Bt toxin. No resistant corn borers were reported.

Many different herbicide mixtures were used on different farms, but the same treatment was given to both Btand non-Bt crops on each farm. The plants were treated between weed stages 1 and 9 (see BBCH Code; Meier, 2001). Weed control in the Bt crops was not different from that in conventional crops. 98.4% of the farmers reported that the response of the Bt maize to the herbicide treatment was the same as that of the conventional maize (1.6% – not known).

In the questionnaire's free text fields, farmers were asked to comment on their observations of plant growth and development, reporting any possible adverse effects. The free text entries were listed and evaluated for reports of any kind of adverse effects. In general, farmers tended to make comments that were mostly positive about *Bt* maize, saying that it was better than conventional varieties. Reported characteristics, during growth weren't

K. Schmidt et al.

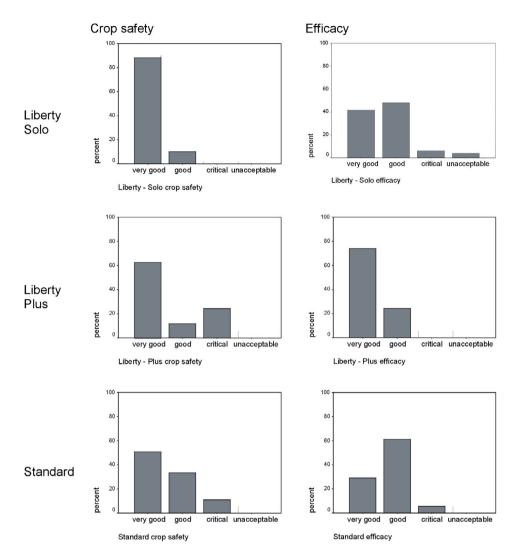


Figure 4. Crop safety and efficacy of the herbicides used for weed control during the study.

related to the GM characteristic and no indications of potential harmful effects were reported.

DISCUSSION

Method of gathering data

The questionnaire was launched both as a printed version and as an identical database-supported digitalized version. The different versions were used according to onsite conditions: the digital version proved to be less time consuming, but not available at all places. Additionally, although it saved one work step, it could not completely replace data quality control. In the future, both versions will be used, though with larger surveys done more effectively with telephone interviews and well-trained interviewers, the digital version might be compulsory.

Company representatives or advisors making interviews collected the data. When questionnaires were completed with an experienced interview partner, the answers were clearer. One of the reasons for this may be that the questionnaire was rather comprehensive – it took about half an hour to complete, and farmers are already overloaded with paperwork. Another reason is the specificity and complexity of the monitoring issues. Therefore, experienced, briefed interviewers, preferably with technical knowledge in agriculture, and familiar with the aims of GMO monitoring, must carry out surveys. Otherwise, the questionnaire would not get the necessary attention to detail and quality of answers from the farmers. Farmers who participated in the study gave positive comments on the questionnaire.

The questionnaire was improved from year to year, due to experience and technical expertise. At first, it offered a lot of open questions and space for individual comments, to get an overview on the possible answer spectra and to test the farmers' understanding. For example, in the beginning some data (soil type, coding of weed stages) were entered according to different systems, then they were harmonized, and one system was used in subsequent questionnaire versions. However, evaluation of the questionnaires showed that it is not possible to have the data results transferred into a database by untrained staff, due to technical terms, abbreviations and logical correlations.

Data quality control

To check the completeness and correctness of the data, quality and plausibility control needs to be established, operating at different stages during the procedures: *i.e.* during data reception (both of paper and digital questionnaires), data coding, within the database and in some parts of the statistical software. It consists of:

- A completeness check: Certain database fields should be defined as mandatory. Several items are indispensable for the analysis, so the corresponding fields have to be completed, for example the questions on monitoring characters.
- *A quality check*: Several quality control methods are available to check that data values are correct. Quantitative data may take on only certain values within a minimum-maximum range, and qualitative values may take on only certain categories.
- *A plausibility check*: Data values have to be logical. For example the harvest date has to be after the sowing date. Possible logical connections between the questions have to be identified and inquired. In case of missing or wrong data, queries can be sent to the farmer in order to complete the database.

Beside this technical quality control, it is important to ensure that common quality criteria of empirical survey research are met. Since the main parameters of monitoring are records of observations of the evaluation of situations compared with non-GM cultivation, the "grade of measurement" is often criticized as being unscientific or subjective. However, the implementation of the scientific methods and tools of empirical data research will provide inter-subjective verifiability of the empirical statements. The most important one is to carefully document all single methodological steps (*e.g.* formulation of questions, definitions, methods of measurement, kind and size of samples, statistical procedures). The quality of a survey with farm questionnaires should be set up and measured on the basis of its objectivity (results independent of the monitoring system), reliability (results reproducible) and validity (results exact). Quality criteria and standards, which have been defined for example by the German Research Foundation (DFG, 1999), must be applied to monitoring surveys by farm questionnaires (see Berensmeier and Schmidt, 2007).

Sample size determination

In this study, the number of questionnaires completed was dependent on the number of GM maize fields cultivated. In the future, a procedure for determining the number or proportion of GM crop areas to be monitored in each region in each season is required to give sufficient statistical power. It is commonly known that seasonal effects on monitoring characters such as pest and disease levels, crop yield and crop quality can be high. In order to fulfil statistical demands, adequate numbers of questionnaires are required each year to take account of areas cultivated and the geographic distribution of the GM crop.

As described above, the pilot study showed that the parameters recorded to determine whether there were any unforeseen effects on the environment may all be transformed to binary distributed characters. There were several monitoring characters, but for the sample size determination they are all treated equally.

The experimental design for the tests with the given hypotheses was done for determined values for the probabilities of type I and type II errors (Rasch et al., 2007a). In each statistical test, decisions for or against the null hypothesis were fixed as α (type I) and β (type II) error probabilities, respectively, where α is the probability of rejecting H_0 although it is true (type I error) and β gives the probability of accepting H₀ although it is wrong (type II error). In monitoring, it is important to detect adverse deviations from "normality", i.e. to accept H_0 with high probability if it is correct (small α value). On the other hand, a wrong rejection of H₀ could have far-reaching consequences, and therefore should be possible only with small β -probability. The sample size was planned with $\alpha = 0.01$ and $\beta = 0.01$ (power = 0.99) for an effect size $\delta = 3\%$ (see tolerance limit calculation in Materials and Methods), which results in a minimum sample size of 2436 questionnaires (CADEMO light, 2006). These conditions are very stringent to avoid false test decisions as far as possible. Considering losses of data due to poor data quality or missing questionnaires, we recommend planning to survey 2500 questionnaires evenly spread over the monitoring period of 10 years, *i.e.* 250 new sites per year on average (with post-cultivation monitoring for volunteers at each site for at least one year).

Links with other data collections

Farm questionnaires are an important tool for monitoring. However, other data of agricultural relevance from various existing monitoring programs need to be taken into account as well. This will broaden the analysis, not only for data on facts that are outside the farmers' responsibility, but also for validating the farmers answers (Schiemann et al., 2006). For example, there are data gathered by the Federal Plant Protection Services in Germany recording the incidence of pest populations or plant pathogens for pest and disease management purposes. The Bundesländer (the federal regions of Germany) intend to link such information and make it available on the Internet (e.g., http://www.isip.de). This provides the opportunity to obtain, or exchange, independent data for comparison with the questionnaire results (for example, whenever pests appear).

In addition, some plant breeders collect their own data from their extension or support services (stewardship programs) for new crop varieties. This information can also be used for comparison with data from questionnaires. For example, there are well-developed systems for monitoring sugar beet in Germany (Merkes et al., 2003).

To make efficient use of these systems for monitoring purposes, they at least have to fulfill the following criteria (Mönkemeyer et al., 2006):

- high quality of the data (design of the program, sampling, analysis and reporting methods and transparency), and a possibility of data transfer;
- data with relevance to GMO cultivation;
- representativeness of the data (quantity of the data: time, frequency and scale of data collection).

Information from these existing monitoring programs should be linked with the data resulting from the questionnaires. This could be done by adding their raw data to the questionnaire database *via* spatial and temporal coordinates, or by asking for the program reports to check them for any adverse effects that could be caused by GMO cultivation. Which approach is better is still under examination. There are proposals to establish a Central Reporting Office to collect reports from relevant networks, and assess them for GMO-caused effects (Sanvido et al., 2005; Schmidtke and Schmidt, 2007).

Prospects and conclusions

Various proposals have been made for collecting data for general surveillance of GM crops, often without assessing their practical value. We have shown that farm questionnaires represent a practical tool for collecting data from the sites where GM crops and comparators are grown, asking people with the greatest experience to collect the data, *i.e.* the farmers. If questionnaires are carefully designed, and consider all the characteristics that are required for statistical reliability, it will be possible to identify unanticipated effects and to assess whether they are potentially harmful and result from growing GM varieties.

In order to generate data of sufficient statistical power, the questionnaires should be completed in sequential seasons from a representative sample of sites. It is considered that an average of 250 questionnaires per year will give sufficient statistical reliability over a ten-year monitoring period. This will provide sufficient data to allow determination of factors specific to a certain region, as well as long-term or cumulative effects. It will also allow evaluation of the scale of an effect, from single farms to large-scale effects (landscapes, regions).

In this study, the limited cultivation of GM maize in Germany determined the time and spatial scale of the application of questionnaires and the amount of collected data. In the future, the cultivation of approved GM maize and other crop varieties will require that the submitted monitoring plans are representative of the planned cultivation of the GM crop over the ten-year period. The recent EFSA guidance on post-market environmental monitoring of GM crops (EFSA, 2006b) proposes that farm questionnaires should be conducted as part of general surveillance. In addition, most monitoring plans submitted contain plans for conducting farm-based surveys using questionnaires. The experiences reported in this study can be used as a basis for developing farm questionnaires for a range of GM crops and for evaluating monitoring plans submitted in applications to commercialize GM crops (Beißner et al., 2006).

MATERIALS AND METHODS

Contents and volume of the questionnaire

A questionnaire collecting information on monitoring characters and influencing factors from farmers was developed and used to monitor the cultivation of GM herbicide tolerant (HT) or corn-borer-resistant (Bt) maize in comparison with conventional maize. The questionnaire focused on four sections:

- 1. Technical data on the farm
- 2. General observations on GM maize cultivation
- 3. Specific data/observations on HT maize (T25) and comparator varieties
- 4. Specific data/observations on *Bt* maize (MON810, BT176) and comparator varieties.

Section 1

Farmers' names and addresses were recorded, but were kept anonymous by providing them with an identifier (ID) number in the database. The farm size (ha), the maize cultivation area (ha), utilization for silage, corn, and corncob mix (CCM), the type of farm, and the agricultural consulting office were registered.

The maize varieties grown on each farm were recorded. Since the observational unit for monitoring and statistical analysis is a field, data sets were established for each monitored field containing GM maize or a comparator. On most farms, both GM maize and conventional maize were cultivated, though in a few cases farmers grew either exclusively GM or conventional varieties.

Section 2

The influencing factors of general cultivation and background environmental data of the maize fields were recorded. In addition, the monitoring characters included observations or assessments of plant development, occurrence of diseases, pests, beneficial insects, and the impact of wild animals or game.

Section 3

This section considered only the HT GM varieties and their comparators, and the monitoring factors requested included observations on weed control regimes and their efficacy. In addition, observations on weed populations and opinions on the efficacy and crop safety of the herbicides used in GM and non-GM cultivation were requested.

Section 4

This section requested observations on corn borer infestation and the associated plant protection measures carried out, such as insecticides, use of *Trichogramma*, mechanical methods or *Bt* maize. The monitoring characters requested for observation and assessment were: growth, unusual effects, and yield. Furthermore, influencing factors such as the herbicide applications to *Bt* maize were recorded.

The complete questionnaire (version of 2004) is available at http://www.jki.bund.de/ cln_045/nn_902382/SharedDocs/17_PS/Publikationen/ arbgrmonit/questgmmaize040823.pdf.html.

Data collation and quality control

For a period of five years (2001 to 2005), maize breeders collected monitoring data from all farms in Germany cultivating GM maize. Because almost all farmers cultivated both GM and non-GM maize, they could compare GM maize directly with conventional maize.

Questionnaires were completed by farmers with assistance from agricultural consults, and collected by staff members of participating breeding companies. A digitalized data capture system in connection with a database (Microsoft Access file) was used for anonymous data entry of all information gathered. The data were transferred or entered into a Paradox database that was specifically developed for this pilot study (Schmidtke and Schmidt, 2006). Qualitative categories were coded (e.g. 1 = yes, 2 = no). A quality and plausibility control check was performed. It contained the verification of the completeness and validity of the data (only valid categories, within credible minimum/maximum ranges). Additionally, crosschecks were conducted to control logical contexts (like the completeness and correctness of specifications for monitored effects, *i.e.* farmers had to give a specification in each case they stated a monitoring character to be "better" or "worse").

From the database, an SPSS data file (SPSS for Windows) for a subsequent detailed statistical analysis and evaluation was exported, so that a single data set of the SPSS file referred to one cultivation area and therefore to one maize field (monitoring object). Comparative data (most monitoring characters) were assigned to the corresponding GM data set only.

Statistical analysis

The evaluation started with a description of the situation, both for the GM and conventional maize data (Schmidt et al., 2004). Frequency analyses were made for categorical data, while for metric variables, statistical measures were calculated (*i.e.* number, number of valid cases, mean, median, minimum, maximum, standard deviation). The distribution of the most important variables was shown by bar diagrams. Categorical monitoring characters – where reasonable – were transformed into binary distributed data to test for adverse effects. The frequencies of the "worse" answers of the monitoring characters were compared to a threshold of 10% by exact binomial tests.

Additionally, the frequencies over the years were estimated and 95% probability upper tolerance limits for any trends were calculated.

Since commercial cultivation is not an experiment, the sites for GM and non-GM cultivation were not randomized but chosen by the farmers, so the assignment of a variety to a field can be regarded as a random event whose distribution depends on several characteristics of the cultivation site and the farm (which align with the influencing factors, also referred to as covariates). As already mentioned, these covariates can generally influence the monitoring characters, and thus the direct comparison for the test factor GM cultivation could be biased. One of the main problems of the analysis is to balance the influence of the covariates on the monitoring characters, and so to get an unbiased assessment of the monitoring characters of GM crop cultivation. In order to detect any bias, the influencing factors were first compared for any differences between GM and non-GM cultivation by statistical tests (t-test, Chi²-test; Rasch, 2007b), and then analyzed for their potential influence on the monitoring characters by logistic regression analysis (Schneider, 2001). It was ascertained that in our data the distributions of the influencing factors were similar in GM and non-GM crops, since in this pilot study farmers commonly cultivated both GM and non-GM maize on their farms, and therefore most influencing factor levels didn't differ between GM and non-GM fields.

In addition to the statistical tests of the relevant characters, individual written remarks concerning apparent GM effects were listed. The causal dependencies with GM maize were evaluated and the degree of any potential harm was assessed.

The statistical analysis was performed with SPSS 10 and 14.

ACKNOWLEDGEMENTS

The authors are grateful to Bayer CropScience, KWS Saat AG, Monsanto, Pioneer and Syngenta for their cooperation, to the Federal Ministry for Agriculture, Food and Consumer Protection for providing the funds for Dr. R. Wilhelm, and to the Federal Ministry for Education and Research for providing the funds for Dr. L. Beißner, Dr. P. Böttinger, K. Schmidt and J. Schmidtke within the joint project "Methodenentwicklung für ein anbaubegleitendes Monitoring von GVP im Agrarökosystem".

Received February 2nd, 2008; accepted June 16, 2008.

REFERENCES

- **Bayer AG** (1992) Important crops of the world and their weeds: their scientific names, synonyms and computer codes according to WSSA/WSSJ; Leverkusen
- **Beißner L, Wilhelm R, Schiemann J** (2006) Current research activities to develop and test questionnaires as a tool for the General Surveillance of important crop plants. *J. Verb. Lebensm.* **1**: 95–97

- **Berensmeier A, Schmidt K** (2007) "Good Monitoring Practice" Quality control measures for farm questionnaires. *J. Verb. Lebensm.* **2**: 56–58
- Berensmeier A, Schmidt K, Beißner L, Schiemann J, Wilhelm R (2006) Statistical analysis of farm questionnaires to search for differences between GM- and non-GM maize. *J. Verb. Lebensm.* 1: 80–84
- Bund/Länder Arbeitsgruppe "Monitoring von Umweltwirkungen gentechnisch veränderter Pflanzen" (2003) Konzept für das Monitoring gentechnisch veränderter Organismen (GVO), September 2002. In Monitoring von gentechnisch veränderten Pflanzen: Instrument einer vorsorgenden Umweltpolitik, Texte 23/03, Umweltbundesamt, Berlin 2003, pp 169–209
- CADEMO for Windows 3.27 (2006) BioMath GmbH, Rostock, Germany
- **DFG Deutsche Forschungsgesellschaft** (1999) Qualitätskriterien der Umfrageforschung (Quality criteria for survey research), Akademie Verlag, Berlin
- **Diekmann A** (2006) Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. Rowohlt, Reinbek bei Hamburg
- **EFSA** (2006a) Guidance Document of the Scientific Panel on Genetically Modified Organisms for the Risk Assessment of Genetically Modified Plants and Derived Food and Feed. *EFSA Journal* **99**: 1–94
- **EFSA** (2006b) Opinion of the Scientific Panel on Genetically Modified Organisms on the Post Market Environmental Monitoring (PMEM) of Genetically Modified Plants. *EFSA Journal* **319**: 1–27
- Meier (ed) (2001) Growth Stages of mono- and dicotyledonous Plants. BBCH Monograph., 2nd Edition, Federal Biological Research Centre for Agriculture and Forestry, publicy available under: http://www.jki.bund.de/cln_045/nn_805656/DE/veroeff/bbch/ bbch_node.html_nnn=true
- Merkes R, Coenen H, Hesse F, Schütz G (2003) Stand der Produktionstechnik zu Zuckerrüben – Ergebnisse der Umfrage 2002. Zuckerindustrie 128: 425–433
- Mönkemeyer W, Schmidt K, Beißner L, Schiemann J, Wilhelm R (2006) A critical examination of the potentials of existing German networks for GMO-monitoring. *J. Verb. Lebensm.* 1: 67–71
- **Official Journal of the European Communities** (17 April 2001) Directive 2001/18/EC of the European Parliament and of the Council of 12 March 2001 on the deliberate release into the environment of genetically modified organisms and repealing Council Directive 90/220/EEC; L 106/1
- **Official Journal of the European Communities** (30 July 2002) Commission Decision of 24 July 2002 establishing guidance notes supplementing Annex II to Directive 2001/18/EC of the European Parliament and of the Council on the deliberate release into the environment of genetically modified organisms and repealing Council Directive 90/220/EEC (notified under document number C (2002) 2715); L 200/22
- Official Journal of the European Communities (18 October 2002) Council Decision of 3 October 2002 establishing guidance notes supplementing Annex VII to Directive 2001/18/EC

of the European Parliament and of the Council on the deliberate release into the environment of genetically modified organisms and repealing Council Directive 90/220/EEC; L 280/27

- **Official Journal of the European Communities** (18 October 2002) Regulation (EC) No 1829/2003 of the European Parliament and of the Council of 22 September 2003 on genetically modified food and feed; L 268/1
- Rasch D, Herrendörfer G, Bock J, Victor N, Guiard V (2007a) Verfahrensbibliothek Versuchsplanung und -auswertung. Oldenbourg Verlag, München
- Rasch D, Verdooren LR, Gowers JI (2007b) The Design and Analysis of Experiments and Surveys. Oldenbourg Verlag, München
- Sanvido O, Bigler F, Widmer F, Winzeler M (2004) Umweltmonitoring gentechnisch veränderter Pflanzen in der Schweiz. *Schriftenreihe der FAL / Les Cahiers de la FAL* 51. Agroscope FAL Reckenholz (in German)
- Sanvido O, Widmer F, Winzeler M, Bigler F (2005) A conceptual framework for the design of environmental postmarket monitoring of genetically modified plants. *Environ. Biosafety Res.* 4: 13–27
- Schiemann J, Wilhelm R, Beißner L, Schmidtke J, Schmidt K (2006) Data acquisition by farm questionnaires and linkage to other sources of data. J. Verb. Lebensm. 1: 26–29
- Schmidt K, Schmidtke J, Wilhelm R, Beißner L, Schiemann J (2004) Biometrical analysis of farmers' questionnaires for the monitoring of the cultivation of genetically modified maize varieties statistical evaluation of monitoring objectives. *Nachrichtenblatt des Deutschen Pflanzenschutzdienstes* 56: 206–212 (in German)
- Schmidt K, Beißner L, Schiemann J, Wilhelm R (2006) Methodology and tools for data acquisition and statistical analysis. J. Verb. Lebensm. 1: 21–25

- Schmidtke J, Schmidt K (2006) Data management and data base implementation for GMO monitoring. J. Verb. Lebensm. 1: 92–94
- Schmidtke J, Schmidt K (2007) Use of existing networks for the General Surveillance of GMP? Proposal of a Reporting System and a Central Reporting Office. J. Verb. Lebensm. 2: 79–84
- Schneider B (2001) Methoden der Planung und Auswertung klinischer Studien. In Rasch D (Hrsg.), Anwendungen der Biometrie in Medizin, Landwirtschaft und Mikrobiologie, BioMath GmbH, Rostock
- SPSS for Windows, Rel. 10.1.3. 2001. Chicago: SPSS Inc.
- Wengraf T (2006) Qualitative Research Interviewing. SAGE Publ., London
- Wilhelm R, Beißner L, Schiemann J (2002) Gestaltung des Monitoring der Auswirkungen gentechnisch veränderter Pflanzen im Agrarökosystem. *Gesunde Pflanzen* 54: 194–206 (in German)
- Wilhelm R, Beißner L, Schiemann J (2003) Concept for the Realisation of a GMO Monitoring in Germany. Updated English version of "Konzept zur Umsetzung eines GVO-Monitoring in Deutschland". *Nachrichtenblatt des Deutschen Pflanzenschutzdienstes* **55**: 258–272
- Wilhelm R, Beißner L, Schmidt K, Schmidtke J, Schiemann J (2004) Monitoring of the cultivation of genetically modified plants Use of farmers' questionnaires. *Nachrichtenblatt des Deutschen Pflanzenschutzdienstes* 56: 184–188 (in German), with supplement: http://www.jki.bund.de/cln_045/nn_902382/SharedDocs/ 17_PS/Publikationen/arbgrmonit/questgmmaize040823. pdf.html