SCIENTIFIC REPORT submitted to EFSA

Statistical Evaluation of the Achievements by Member States of the EU *Salmonella* Reduction Targets in Animal Populations[1]

Prepared by: Sotto C., Litière S., Aerts M.

CenStat, Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BIOSTAT),
Hasselt University & Katholieke Universiteit Leuven, Agoralaan 1, Bldg. D, 3590 Diepenbeek, Belgium

ABSTRACT

EU Member States are required to collect, evaluate and report data on zoonoses, zoonotic agents, antimicrobial resistance and food-borne outbreaks to the European Commission each year. The European Food Safety Authority is responsible for examining, analyzing and summarizing these data, and for publishing the results in the European Union Summary Report. The identification of trends in the occurrence of the zoonotic agents and the sources of human infections, in order to study the likelihood of Member States to achieve the European Union reduction targets, is one of the key analyses in the Summary Report. In this report, particular interest lies on the *Salmonella* European Union reduction targets in animal populations, particularly in flocks of breeding and laying hens of *Gallus gallus*. The main objective of this investigation was to explore and assess appropriate statistical methodologies enabling to evaluate the achievement by Member States of the *Salmonella* European Union reduction targets in animal populations. *Salmonella* flock prevalence data in breeding and laying hens of *Gallus gallus* at two levels – aggregated country-level monitoring data for all Member States, as well as non-aggregated, detailed, sample-level data for a number of Member States – were used for the investigation. For the aggregated-level data the extremely short time sequence available of minimum four annual time points implied that only simplistic models could be considered. It was concluded that reliable trend analyses could not be established based on such very limited amount of information. In contrast, for the sample-level data, a number of modelling approaches proved meaningful and stable enough to provide insight into the progress made by Member States towards the achievement of the *Salmonella* reduction targets in breeding and laying hens. For more reliable and informative trends analyses based on aggregated-level data it is recommended that Member States would provide quarterly periods or monthly prevalence data, rather than yearly values.

KEY WORDS

Salmonella spp., food-borne pathogens, European Union reduction target, logistic regression, generalized estimating equations.

[1] **(Question No EFSA-Q-2009-00831). Accepted for Publication on 30 November 2011.**

SUMMARY

The identification of trends in the occurrence of the zoonotic agents and the sources of human infections, in order to study the likelihood of Member States to achieve the European Union reduction targets, is an important aspect of the European Union Summary Report. In this report, several statistical methodologies useful for the evaluation of the *Salmonella* reduction targets in breeding and laying hens of *Gallus gallus* have been presented and assessed using data aggregated at country-level, as well as sample-level data.

For the aggregated-level monitoring data provided by all Member States, time trend analyses for *Salmonella* prevalence in breeding and laying hens of *Gallus gallus* were explored for those Member States having at least four years of prevalence data. These analyses consisted primarily of logistic regression models incorporating a linear, as well as a quadratic, trend in time. For some Member States, such simplistic models were sufficient to estimate trends in the observed *Salmonella* prevalence. However, for many other Member States, the said models were insufficient to establish a reliable trend analysis. Additional data for other years may significantly improve the fit of these models. Moreover, due to the fairly short period on which monitoring data was available (2004-2009), more complex model structures, though explored, did not yield very meaningful results. Finally, alternative modelling strategies, which were not feasible given the current data limitations but which might nevertheless be useful for the evaluation of the achievement of the reduction targets, were also proposed and described.

For the detailed sample-level data provided voluntarily by six Member States, time trend analyses for *Salmonella* prevalence were again considered. Prevalence data were available for one, two or three years, on a monthly basis, as opposed to on a yearly basis for the aggregated-level data. Moreover, detailed information on the membership of flocks to a particular holding was also available. Hence, a much more extensive range of models could be considered for the time trend analyses of the sample-level data in comparison to that for the aggregated-level data. Simplistic logistic regression models were initially considered, and these were further extended to account for possible clustering effects induced by flocks belonging to the same holding by means of the generalized estimating equations approach. In most cases, the generalized estimating equation models proved sufficient in capturing the observed prevalence trends. In addition, for Member States with more complex observed prevalence trends, more flexible modelling strategies (e.g., generalized estimating equations with splines) were further explored. The various methodologies were also compared based on their respective predictions for *Salmonella* prevalence. As was done for the aggregated-level data, other approaches that were not applicable to the data at hand but could be meaningful were further described.

For more reliable and informative trends analyses based on aggregated-level data it is recommended that Member States would provide quarterly periods or monthly prevalence data, rather than yearly values.

An extensive examination of the merits of the various methodologies considered is provided in the Discussion and Conclusions section, and recommendations for the modelling approaches used here, as well as for the proposed alternative modelling strategies, are provided in the Recommendations section.

Table of Contents

Background

The European Union (EU) system for monitoring and collection of information on zoonoses is established by Directive 2003/99/EC on the monitoring of zoonoses and zoonotic agents[2]. This Directive requires the Member States (MSs) to collect, evaluate and report data on zoonoses, zoonotic agents, antimicrobial resistance and food-borne outbreaks to the European Commission (EC) each year. The monitoring system used is based on that of the Members States, and in few cases it is harmonised by the EU legislation to the extent that the results from the monitoring are directly comparable between the MSs.

According to the Directive, the MSs have to send their report on zoonoses to the EC each year by 31st May. The EU Summary Report (EUSR) is prepared by the European Food Safety Authority (EFSA) in close collaboration with the European Centre for Disease Prevention and Control (ECDC) and EFSA´s Zoonoses Collaboration Centre. In this report, the information received from the MSs is analysed and summarised specifically to identify trends in the occurrence of the zoonotic agents and the sources of human infections.

Regulation (EC) No 2160/2003[3] on the control of *Salmonella* and other specified zoonotic agents provides for the setting of EU targets for reducing the prevalence of *Salmonella* serovars with public health significance in animal populations. Covered by EU-wide *Salmonella* prevalence baseline surveys reduction targets are being set for the reduction of certain *Salmonella* serovars in different poultry populations. As regards breeding hens of *Gallus gallus,* Regulation (EC) No 1003/2005[4] transitionally sets a target for reduction being 1% or less flocks remaining positive for *Salmonella* Enteritidis, *Salmonella* Typhimurium, *Salmonella* Hadar, *Salmonella* Infantis or *Salmonella* Virchow by the end of 2009. As regards laying hens (*Gallus gallus*), Regulation (EC) No 1168/2006[5] sets a general EU target for reduction of the maximum percentage to 2 % or less flocks remaining positive for *Salmonella* Enteritidis or *Salmonella* Typhimurium, or an annual Member State-specific minimum percentage of reduction of flocks remaining positive for *Salmonella* Enteritidis or *Salmonella* Typhimurium, during a transitional period until 1 February 2011.

These regulations also harmonise the monitoring of the *Salmonella* prevalence in those poultry populations in all EU MSs in order to verify the achievement of EU reduction targets. These prevalence data are reported by MSs in their national zoonoses reports in accordance with Directive 2003/99/EC[2] and this information is summarised and analysed in the EUSRs on zoonoses that are published on EFSA web-site.

In order to verify whether the mentioned EU targets for reduction are met by MSs, the monitoring data collected by the MSs should be statistically analyzed.

[2] Directive 2003/99/EC of the European Parliament and of the Council of 17 November 2003 on the monitoring of zoonoses and zoonotic agents, amending Council Decision 90/424/EEC and repealing Council Directive 92/117/EEC. OJ L 325, 12.12.2003.

[3] Regulation (EC) No 2160/2003 of the European Parliament and of the Council of 17 November 2003 on the control of *Salmonella* and other specified food-borne zoonotic agents. OJ L 325, 12.12.2003, p. 1.

[4] Regulation (EC) No 1003/2005 implementing Regulation (EC) No 2160/2003 as regards a Community target for the reduction of the prevalence of certain *Salmonella* serotypes in breeding flocks of *Gallus gallus* and amending Regulation (EC) No 2160/2003. OJ L 170, 1.7.2005, p. 12.

[5] Regulation (EC) No 1168/2006 implementing Regulation (EC) No 2160/2003 as regards a Community target for the reduction of the prevalence of certain *Salmonella* serotypes in laying hens of *Gallus gallus* and amending Regulation (EC) No 1003/2005. OJ L 211, 1.8.2006, p. 4.

Terms of reference

The overall objectives of the project were;

- To investigate the appropriateness of different statistical methodologies to evaluate the progress made by MS towards, or the achievement of, the *Salmonella* reduction targets in breeding flocks and laying hens, based on the *aggregated* prevalence data forwarded by MS in their annual zoonoses reports.
- To investigate the appropriateness of different statistical methodologies to evaluate the progress made by MS towards, or the achievement of, the *Salmonella* reduction targets in breeding flocks and in laying hens, based on more detailed, *sample-level* data, provided by at least one MS.
- To compare the consistency of the results of the aforementioned evaluations.

The specific objectives of the project were;

- To formulate recommendations regarding the design of monitoring schemes to verify the achievement of *Salmonella* reduction targets in animal populations
- MS-specific aggregated prevalence
  - Propose appropriate statistical methodologies enabling the evaluation of the progress made by the MS towards, or the achievement of, the *Salmonella* reduction targets, based on the aggregated prevalence data submitted by MS to EFSA.
  - Evaluate quantitatively by appropriate statistical methodologies the MS-specific likelihood of achievement of the *Salmonella* reduction targets in breeding and in laying hens, based on the aggregated prevalence data forwarded submitted by MS to EFSA.
  - Investigate the impact of different monitoring testing schemes with imperfect sensitivity and specificity on the assessment of achievement of targets.
  - Indicate the advantages and disadvantages of the different appropriate available statistical methodologies while clearly specifying the underlying assumptions.
- MS-specific sample-level prevalence
  - Indicate appropriate statistical methodologies enabling the evaluation of the progress made by one or more MS towards, or the achievement of, the *Salmonella* reduction targets, based on the non-aggregated, sample-level prevalence data submitted by one or more MS to EFSA.
  - Evaluate quantitatively by appropriate statistical methodologies the MS-specific likelihood of achievement of the *Salmonella* reduction targets in breeding and in laying hens, based on the non-aggregated, sample-level prevalence data submitted by one or more MS to EFSA.
- Comparison
  - Compare and interpret the results of both evaluations

Acknowledgements

The authors would like to thank the Members of EFSA's Task Force of Zoonoses Data Collection for their kind collaboration and review, and the EFSA's staff members: Giusi Amore and Frank Boelaert for the support provided to this scientific report submitted to EFSA. The six Member States' representatives who kindly provided sample-level *Salmonella* monitoring data in flocks of breeding and laying hens on a voluntary, confidential basis are also acknowledged.

This contract was awarded by EFSA to:

Center for Statistics, Hasselt University, Diepenbeek, Belgium.

Contractor: Center for Statistics, Hasselt University, Agoralaan, Building D, B – 3590 Diepenbeek , Belgium.

Contract title: Request for assistance in statistical evaluation of the achievement by Member States of the EU *Salmonella* reduction targets in animal populations.

Contract number: Specific Contract No 05, implementing Multiple Framework Contract CT/EFSA/AMU/2007/01-06

Introduction and Objectives

INTRODUCTION

Data collection on zoonoses and zoonotic agents in animals, food and feed, as laid down in Directive 2003/99/EC[6], is mainly based on the systems in place in MSs. The data collected should be relevant and comparable in order to identify and characterise risks and to assess exposures related to zoonoses and zoonotic agents. In order to verify whether the mentioned EU targets for *Salmonella* reduction are met by MSs, the monitoring data collected by the MSs should be statistically analyzed. A major objective of this statistical analysis is to identify trends in the occurrence of the zoonotic agents and the sources of human infections, in order to study the likelihood of EU MS to achieve the EU reduction targets.

In this project, the animal populations under study are the breeding flocks and laying hens of *Gallus gallus*. As regards the breeding hens, the EU target for reduction was set at 1% or less flocks remaining positive for *Salmonella* Enteritidis, *Salmonella* Typhimurium, *Salmonella* Hadar, *Salmonella* Infantis or *Salmonella* Virchow by 31 December 2009. The year 2009 was the third year where MSs were obliged to implement *Salmonella* control programmes in breeding flocks of *Gallus gallus* in accordance with Regulation (EC) No 2160/2003[7]. These control programmes aim to meet the *Salmonella* reduction target set by Regulation (EC) No 1003/2005[8], where the *Salmonella* reduction target in breeders covers the following serovars: *S.* Enteritidis, *S.* Typhimurium, *S.* Infantis, *S.* Virchow and *S.* Hadar. The target was set for all adult breeding flocks, during the production period, comprising at least 250 birds.

As regards the laying hens, in 2008 and 2009, MSs implemented *Salmonella* control programmes in laying hen flocks of *Gallus gallus* providing eggs intended for human consumption in accordance with Regulation (EC) No 2160/2003[6]. The control programmes consist of proper and effective measures of prevention, detection, and control of *Salmonella* at all relevant stages of the egg production line, particularly at the level of primary production, in order to reduce *Salmonella* prevalence and the risk to public health. In Regulation (EC) No 1168/2006[9], the target in laying hens is defined as an annual minimum percentage of reduction in the number of adult laying hen flocks (i.e. in the production period) remaining positive to the targeted serovars *S.* Enteritidis or *S.* Typhimurium, by the end of the previous year. The annual MS-specific targets are proportionate depending on the prevalence in the preceding year. For the most advanced MSs, the EU target is defined as a maximum percentage of flocks remaining ultimately positive of 2%, during a transitional period until 1 February 2011. For MSs with less than 50 flocks of adult laying hens, not more than one adult flock may remain positive.

OBJECTIVES

The main objective of this assignment is to provide assistance in the investigation of appropriate statistical methodologies enabling EFSA to evaluate the achievement of the *Salmonella* EU reduction targets in animal populations by MS.

---

[6] Directive 2003/99/EC of the European Parliament and of the Council of 17 November 2003 on the monitoring of zoonoses and zoonotic agents, amending Council Decision 90/424/EEC and repealing Council Directive 92/117/EEC. OJ L 325, 12.12.2003.

[7] Regulation (EC) No 2160/2003 of the European Parliament and of the Council of 17 November 2003 on the control of *Salmonella* and other specified food-borne zoonotic agents. OJ L 325, 12.12.2003, p. 1.

[8] Regulation (EC) No 1003/2005 implementing Regulation (EC) No 2160/2003 as regards a Community target for the reduction of the prevalence of certain *Salmonella* serotypes in breeding flocks of *Gallus gallus* and amending Regulation (EC) No 2160/2003. OJ L 170, 1.7.2005, p. 12.

[9] Regulation (EC) No 1168/2006 implementing Regulation (EC) No 2160/2003 as regards a Community target for the reduction of the prevalence of certain *Salmonella* serotypes in laying hens of *Gallus gallus* and amending Regulation (EC) No 1003/2005. OJ L 211, 1.8.2006, p. 4.

1. MATERIALS AND METHODS

The outcome variable that was considered was positivity, as a binary outcome variable (positive/negative), for (groups) of *Salmonella* serovars of public health and epidemiological significance.

- In flocks of breeding hens of *Gallus gallus*
  - positivity for at least one of the five *Salmonella* serovars covered by the *Salmonella* reduction target: *S.* Enteritidis, *S.* Typhimurium, *S.* Infantis, *S.* Virchow and *S.* Hadar, and
- In flocks of laying hens of *Gallus gallus*
  - positivity for at least one of the two *Salmonella* serovars covered by the *Salmonella* reduction target: *S.* Enteritidis and *S.* Typhimurium

Outcomes are analyzed at flock level, separately for breeding and laying hens. A flock is classified as positive if at least one sample (either boot swab samples or other faecal material and dust samples) is positive and negative if all samples are negative.

Two sets of data were available to study the progress of MSs towards achieving the *Salmonella* reduction targets. For all MSs and two non-MSs, aggregated (summarized) monitoring data were available since 2004. These consist of the number of existing flocks, the number of tested flocks and the number of *Salmonella* spp. positive flocks. For these data, no individual unit information is available. An overview of the MSs' *Salmonella* surveillance programmes in breeding and laying hen flocks, and of the minimum legal EU-monitoring and reporting requirements can be found in the Appendix tables of the European Union Summary Report on Trends and Sources of Zoonoses (EFSA, 2011).
For six MSs, non-aggregated, detailed data on the sample-level were available, as well as additional information on the date of setting up of the flock, the date of outcome, the date of slaughtering of negative flocks, and some historical data. These data were reported by the six MSs on a voluntary and confidential basis.

In the following section, descriptive tables and graphs are presented for the countries, followed by an overview of the various statistical methodologies considered for this report. As different approaches can be used for the aggregated-level and sample-level data, the statistical methods applicable to these two types of data are presented separately in respective subsections.

In order to guarantee confidentiality the country names were made anonymous by randomly attributing numbers (aggregated data set) and alphabetic letters (sample-level data set) to each of the MSs. This was done by using a random generator function of the statistical software R.

1.1. Aggregated-level data

1.1.1. Data

For all countries, aggregated (summarized) monitoring data were available since 2004. These consist of the number of existing flocks, the number of tested flocks and the number of *Salmonella* spp. positive flocks. No variables characterizing individual unit data were available.

An overview of the available information in terms of the percentage of positive flocks per year for each country is provided in Table 1 for breeding flocks and in Table 2 for laying flocks. Empty cells

represent missing information, which indicates that complete information was not available for all countries. In order to be able to fit and illustrate the methodology presented in the upcoming sections, a minimum amount of information is required. It was therefore decided that only countries which contributed information on at least four time points (years) would be used in the analysis, i.e. 15 countries for the breeding flocks and 18 countries for the laying flocks. Countries which will therefore not be included in the analysis are marked in light grey in Table 1 and Table 2.

For breeding hens, the percentage of breeding flocks positive to the five target *Salmonella* serovars for all years are generally low (<5%). Of particular note is the 100% positivity rate for country 1 for 2004. This, however, is based only on a single flock tested, and hence, is not informative as such. Country 25 was observed to have relatively high *Salmonella* positivity for 2004 and 2005 (>10%), while country 8 had moderately high rates in the first 3 years (>10%). In addition, a few countries have positivity rates consistently below the reduction target of 1% all throughout the six years (e.g. countries 17 and 22).

The percentage of laying flocks positive to the two target *Salmonella* serovars was also generally low (>1% to 10%) to very low (0.1% to 1%) for most countries, with the exception of countries 7, 8, 9 and 13 (in 2005). Countries 3, 11, 19, 23, 24, 26 and 28 registered prevalence rates that were lower than the EU reduction target of 2% for all the years for which these countries provided data for.

Table 1. Percentage of breeding flocks positive to the five target *Salmonella* serovars and number of tested breeding flocks, per year, by country, aggregated data, EUSR 2004-2009.

Country	2004	2005	2006	2007	2008	2009
1	100.0% (1)		0.3% (696)	1.0% (597)	0.0% (249)	2.3% (129)
3				0.7% (138)	0.0% (148)	0.0% (162)
4	2.8% (472)		7.3% (96)	0.9% (2,164)	0.5% (2,204)	2.7% (714)
5	3.6% (56)	6.7% (45)		0.0% (118)	0.7% (151)	0.0% (155)
6			0.0% (2,628)	0.2% (6,365)	0.8% (31,635)	3.4% (5,659)
7	6.3% (16)				0.0% (26)	0.0% (25)
8	16.6% (517)	15.8% (531)	19.9% (648)	2.3% (855)	2.5% (1,304)	3.3% (1,266)
9	4.2% (403)	7.8% (218)		13.2% (38)	0.0% (72)	7.0% (272)
10	4.5% (89)	1.1% (90)	0.7% (291)	5.1% (552)	1.1% (557)	1.0% (620)
11					0.0% (182)	0.0% (187)
13			3.5% (1,382)		0.0% (108)	4.1% (73)
14	1.7% (58)	3.3% (60)	0.0% (61)		0.0% (52)	0.8% (120)
15	3.0% (33)				0.0% (35)	1.8% (55)
16	0.5% (605)	0.5% (613)		1.2% (498)	0.9% (550)	0.0% (526)
17	0.3% (1,058)	0.7% (1,054)	0.3% (935)	0.6% (1,177)	0.5% (1,103)	0.2% (1,480)
18	5.4% (936)	7.8% (614)	5.3% (1,072)		5.4% (1,069)	2.7% (1,056)
19					0.0% (119)	0.0% (93)
20				0.1% (1,633)	0.5% (1,636)	0.1% (1,637)
21			60.0% (10)		0.0% (6)	0.0% (3)
22	0.7% (282)		0.8% (1,531)	0.9% (1,172)	0.6% (1,164)	0.5% (850)
23						0.6% (325)
24				0.0% (489)	0.5% (203)	0.0% (129)
25	57.3% (75)	26.0% (100)	5.3% (19)	15.4% (117)	5.7% (209)	0.5% (219)
26	0.7% (149)				0.0% (175)	0.0% (172)
27					0.0% (85)	0.9% (2,193)
28	1.2% (164)		0.7% (293)	1.1% (270)	0.6% (317)	1.2% (249)
29	1.0% (97)	4.2% (71)			0.0% (429)	1.6% (512)

Table 2. Percentage of laying flocks positive to the two target *Salmonella* serovars and number of tested laying flocks, per year, by country, aggregated data, EUSR 2004-2009.

Country	2004	2005	2006	2007	2008	2009
1	2.7% (219)	12.6% (309)	1.9% (1150)	1.7% (1032)	7.2% (138)	6.5% (155)
2				25.0% (8)	14.3% (7)	0.0% (7)
3	0.3% (772)	0.1% (859)	0.1% (670)	0.3% (590)	0.4% (724)	0.1% (904)
4			2.2% (417)		8.7% (866)	3.8% (887)
5	0.9% (112)	5.6% (107)	0.6% (165)	5.0% (179)	8.7% (172)	3.3% (209)
6	1.4% (4,707)	1.0% (4,873)	0.8% (2,764)	1.6% (5,105)	2.7% (6,304)	4.8% (4,399)
7		61.1% (18)	36.4% (11)	20.5% (73)	14.5% (69)	9.9% (71)
8	38.1% (21)	51.5% (485)	13.2% (1,125)	11.8% (771)	15.6% (845)	7.2% (1,511)
9	17.6% (85)				14.3% (112)	3.4% (327)
10	3.3% (90)			23.9% (426)	7.6% (449)	10.9% (467)
11					0.0% (1,080)	0.0% (1,031)
12						0.0% (48)
13	0.2% (663)	38.5% (13)			0.0% (13)	6.2% (81)
14	0.9% (1,896)	1.3% (3,488)	2.3% (2,419)	3.0% (2,565)	1.4% (1,966)	2.5% (2,578)
15					0.0% (40)	4.3% (92)
16				3.4% (378)	3.7% (649)	3.8% (763)
17	2.7% (3,359)		3.9% (3,099)	3.9% (2,960)	3.2% (3,067)	2.0% (3,657)
18	3.8% (1,838)	3.5% (1,865)	4.3% (1,819)	0.2% (3,814)	10.6% (1,533)	9.4% (1,718)
19		0.5% (1,631)	0.2% (1,828)	0.6% (521)	0.7% (306)	
20		7.9% (454)			1.0% (5,523)	0.3% (4,466)
21			4.0% (25)	1.6% (61)	1.9% (52)	0.0% (48)
22		3.4% (1,952)	4.4% (2,055)	5.3% (4,031)	2.6% (2,346)	1.5% (2,240)
23						0.2% (420)
24		1.4% (217)	0.3% (340)		0.3% (326)	0.0% (375)
25	55.6% (9)				10.6% (227)	6.4% (251)
26		0.1% (1827)		0.2% (626)	0.1% (950)	0.2% (900)
27					0.0% (119)	8.9% (101)
28	0.5% (641)	0.9% (658)	0.2% (565)	0.8% (510)	0.4% (508)	1.8% (454)
29	5.8% (434)	2.4% (490)	3.0% (332)	1.9% (1024)	6.8% (821)	5.6% (921)

The percentages of positive breeding and laying flocks for the most recent year, 2009, along with 95% confidence intervals (CIs[10]), are illustrated in Figure 1 and Figure 2, respectively. The highest prevalence values for 2009 were observed for countries 9 and 13 for breeding flocks (though both were below 8%) and for countries 7 and 10 for laying flocks (both below 15%). The prevalence values in relation to the reduction target (green dashed line) can also be seen from these figures. In addition to fairly differing levels of *Salmonella* prevalence across countries, the variability of these prevalence values also differ largely across MSs. Countries 1, 9, 13 and 15, for instance, exhibit quite some variability in the percentages of positive breeding flocks, while countries 1, 7, 13, 15 and 27 exhibit the same for laying flocks. The latter observation, however, might be attributed to the differences across countries in the number of flocks tested. A larger number of flocks sampled would naturally lead to smaller variability (or more narrow CIs).

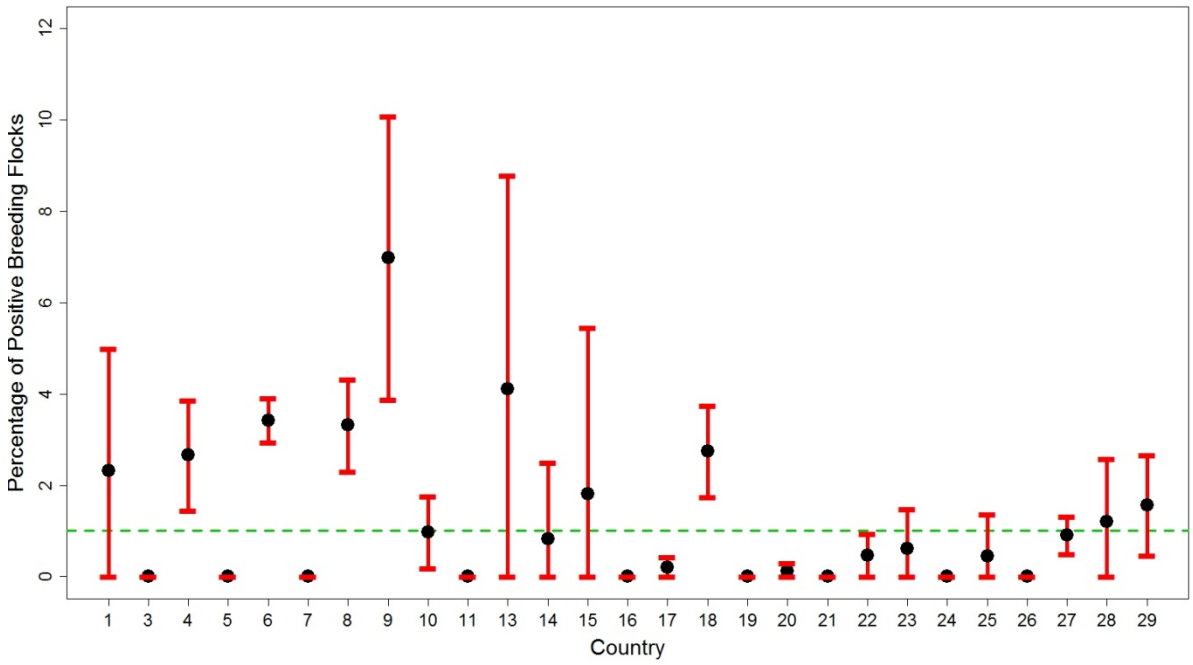


Figure 1. Percentage of breeding flocks positive to the five target *Salmonella* serovars*, with 95% Confidence Intervals, by country, aggregated data, EUSR 2009.

* *Salmonella* reduction target (dashed green line) at 1%

[10] Although in epidemiological terms, a census of flocks (all flocks) was sampled, in the present statistical report sampling variation was taken account of because not all hens within the flocks were sampled.

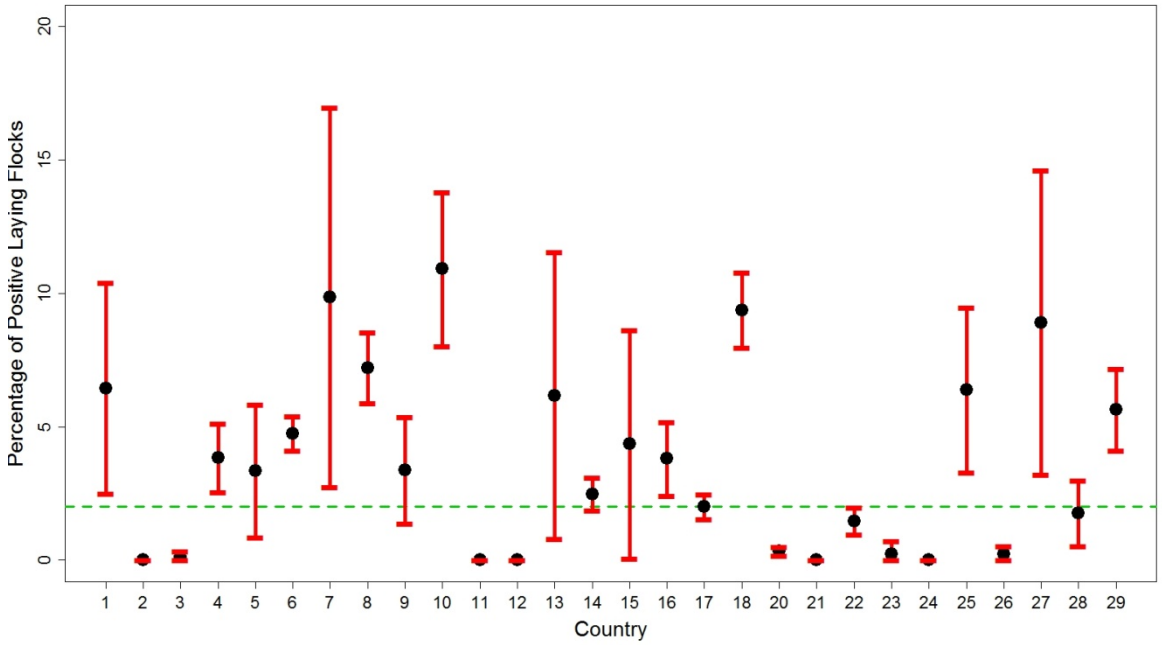


Figure 2. Percentage of laying flocks positive to the two target *Salmonella* serovars*, with 95% Confidence Intervals, by country, aggregated data, EUSR 2009.

* *Salmonella* overall reduction target (dashed green line) at 2%.

To provide an idea of the sample size for each country included in the analysis, the average number of units tested per year is provided in Figure 3 for breeding hens and in Figure 4 for laying hens.

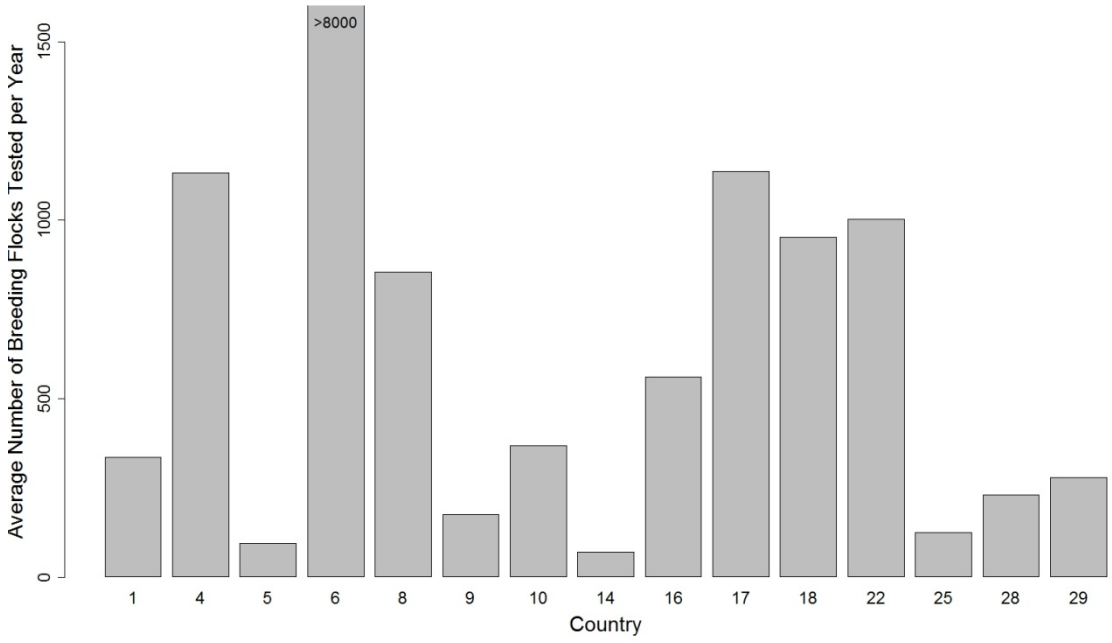


Figure 3. Average number of breeding flocks tested per year, by country, EUSR 2004-2009.

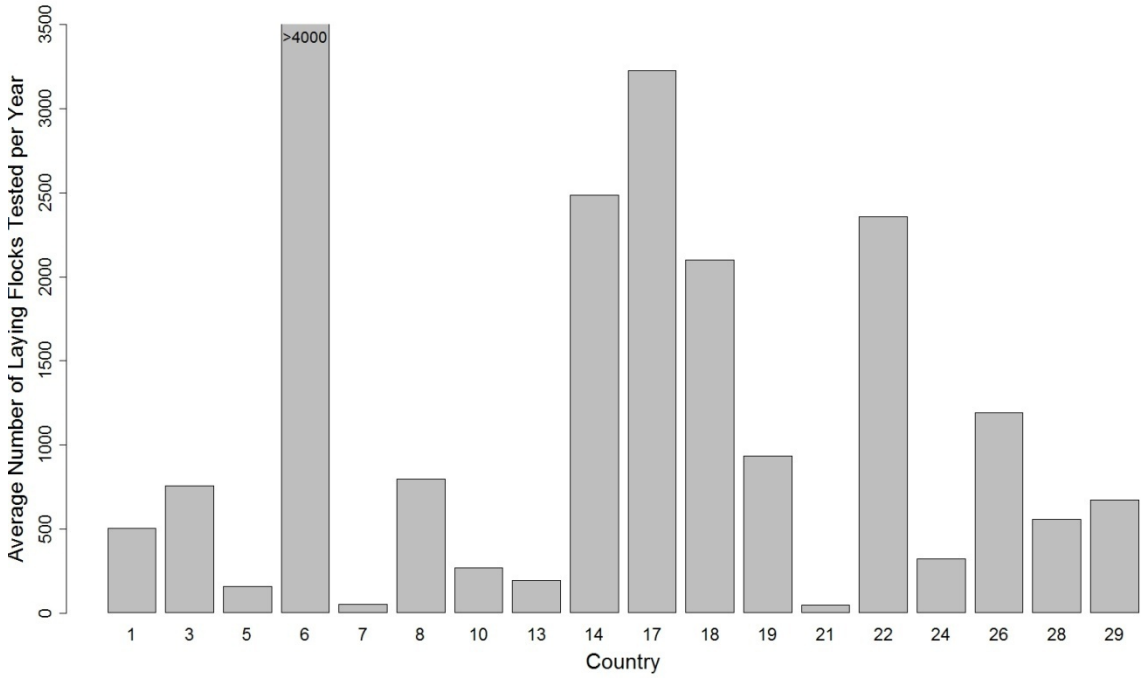


Figure 4. Average number of laying flocks tested per year, by country, aggregated data, EUSR 2004-2009.

1.1.2. Methodology

General Models for Statistical Trends

The most basic approach consists of fitting a ***logistic regression model* with a linear trend** to the available data points. According to the technical specifications, interest goes to the prevalence at the flock level. If the level of harmonisation of the monitoring scheme for the aggregated-level data is believed to be sufficiently high to allow valid inferences, then these can be obtained from the binomial distribution.

Let π_i be the probability for a flock to be positive, let n_{it} be the number of flocks at time point *t* from country *i*. Starting point for inference on the 'flock prevalence' is the binomial distribution for the number of positive flocks y_{it} at time point *t* in country *i*:

$$y_{it} \sim \mathrm{B}(n_{it}, \pi_{it}). \qquad (1)$$

A time trend can be included by considering a link function, $g(\cdot)$. Then, $g(\pi_i)$ can be expressed as a function of time, in this case (a transformation of) year of reporting, x_t:

$$g(\pi_{it}) = \beta_{0i} + \beta_{1i} x_t, \qquad (2)$$

where β_{0i} and β_{1i} respectively refer to MS specific intercepts and the time effects. If some curvature is present in the observed prevalence trends over time, Model (2) can be further extended to include a **quadratic effect for time**, i.e.

$$g(\pi_{it}) = \beta_{0i} + \beta_{1i}x_t + \beta_{2i}x_t^2. \quad (3)$$

If the observed trend is even more complex, one can consider going the route of a fully **nonlinear modelling approach**. Such a model might be of the form

$$g(\pi_{it}) = \beta_{0i} + \beta_{1i}x_t^{\beta_{2i}}. \quad (4)$$

For the aggregated-level data, a logistic regression with a linear time trend was first considered. If the latter was not appropriate or did not yield a good fit, the time trend was extended to a quadratic type to achieve a better fit. Finally, if such was still not appropriate, non-linear models were also further considered.

Note that the above model specifications express a transformation of the response in terms of the variable time. Other important factors or variables can also be included as terms within any of the above models. For instance, changes implemented in the legal reporting requirements, which could invalidate the assumption of a harmonized monitoring scheme, can be corrected for in (2) by considering additional covariates that would allow different evolutions of trends before and after the change.

Several link functions can be used – the most common of which for this type of data are the *logit* and the *probit* link functions. The *logit* link function is just the logarithm of the odds of the probability, $\log\left(\frac{\pi}{1-\pi}\right)$. The *probit* link function uses the inverse of the normal cumulative distribution function (Φ), and is denoted as $\Phi^{-1}(\pi_i)$.

Transformation of the resulting estimate from a logistic regression onto the probability scale can be done using the formula given below

$$\text{Logit}(\pi_j) = \ln\left(\frac{\pi_j}{1-\pi_j}\right) = \beta_0 + \beta_1 t_j \quad \Leftrightarrow \quad \pi_j = \frac{e^{\beta_0+\beta_1 t_j}}{1+e^{\beta_0+\beta_1 t_j}}, \quad (5)$$

where, for simplicity of notation, indices referring to country and holding have been omitted. Observe that β_0 and β_1 respectively correspond to the intercept and time parameter.

For a sample with independent observations, this approach would lead to an accurate estimate of the linear evolution of *Salmonella* prevalence over time for each country, and could hence be used to assess the likelihood of achieving *Salmonella* reduction targets. The main complication here is that the assumptions on the binomial distribution may be violated either by:

- violation of independence: outcomes from the same holding are expected to be more alike (correlated) as compared to outcomes from a different holding (hierarchical correlation structure), or,
- violation of constant probability: samples, even from the same holding or even from the same flock, might have different probabilities to be infected (heterogeneity of probability).

In the aggregated data, no holding-level information is available to assess the impact of these violations on the final inferences. In this setting, the best approach may therefore be to ignore the correlation and use Model (1)-(2) with a linear time trend, assuming independent observations. While this typically leaves the consistency of point estimation intact, the same is not true for measures of precision. In case of a 'positive' correlation (i.e., samples within a holding are more alike than between holdings), then ignoring this aspect of the data, just as ignoring overdispersion, overestimates precision, and hence, underestimates standard errors and lengths of CIs.

Exact Inference for Logistic Regression

Once the model is fitted, inferences (e.g., tests for significance) for the model parameters are typically done using a Wald statistic. Tests based on the latter are performed using large-sample approximations of the distribution of the test statistic and thus work fairly well for very large samples. For small samples, however, such large-sample approximations may breakdown, yielding unreliable conclusions regarding the significance of the parameters. In such cases, one can consider exact inference for the parameters of the logistic regression model. Exact conditional inference is based on generating the conditional distribution for the sufficient statistics of the parameters of interest. This distribution is called the permutation or exact conditional distribution. Significance tests based on the latter are then much more reliable when the sample size is quite modest or even small.

For the aggregated-level data, the large-sample assumptions should be reasonably met, since sample sizes per year for the different countries are fairly large (Figure 3 and Figure 4). For the sample-level data, however, as can be observed from Table 7 to Table 11 the sample sizes per month for some MSs are quite small. Hence, exact inference for the logistic model might be useful in this case.

1.1.3. Evaluation of the Proposed Methodology

Next to providing an overview of the state of the art of available methodology for the analysis of statistical trends in the progress of MS towards achievement of the *Salmonella* reduction targets, it is also important to investigate the appropriateness of these models. For the purpose of the statistical evaluation performed in this report for the achievement of the *Salmonella* reduction targets in laying hen flocks, only the general EU target being defined as a maximum percentage of flocks remaining ultimately positive of 2% (Regulation 1168/2006) has been taken into account for all MSs. Nevertheless, it is to be noted that this general EU target is not, for some MSs, the one established in the Regulation.

Goodness-of-Fit for Logistic Regression

The appropriateness of a statistical model can often be assessed by means of measures of so-called *goodness-of-fit*. For simple models such as a logistic regression, several goodness-of-fit statistics are available (Agresti, 2007). For instance, one can consider the Deviance statistic, G^2, and/or the Pearson statistic, X^2. Let us first introduce these statistics for a model with only categorical variables. In this setting, the data can be summarized as counts in a contingency table, for which expected frequencies can be estimated. The Deviance statistic compares the observed and expected values using the following expression

$$G^2 = 2\sum \text{observed}\left[\log\left(\frac{\text{observed}}{\text{expected}}\right)\right].$$

On the other hand, the Pearson statistic is based on the difference between the observed and expected values, as illustrated by the following expression

$$X^2 = \sum \frac{(\text{observed}-\text{fitted})^2}{\text{fitted}}.$$

Both have approximate chi-square distributions (when the expected frequencies are all above 5). The degrees of freedom are obtained as the difference between the number of parameters in the fitted model and the number of parameters in the saturated model. The null hypothesis for a goodness-of-fit

test (based on the Deviance or Pearson statistic) is good fit and a significant *P*-value $(p < 0.05)$ for these tests would provide evidence of a poor model fit.

In our case, time is considered as a continuous variable to study the evolution of *Salmonella* prevalence over time. In general, G^2 and X^2 cannot be used to study goodness-of-fit of regression models with (nearly) continuous predictors. Indeed, each possible value of the predictor would be used to create a contingency table, and this would result in many sparse cells. Hence, G^2 and X^2 may not have approximate chi-square distributions. One way of dealing with this problem consists of grouping the data artificially by combinations of certain values of the predictor.

A second solution to study goodness-of-fit for logistic regression with continuous predictors is provided by Hosmer and Lemeshow (2000). These authors propose an alternative way of grouping the data using a partitioning based on estimated probabilities. For example, with 10 groups of equal size, the first pair of observed counts and corresponding fitted counts refers to the n/10 observations having the highest estimated probabilities; the next pair refers to the n/10 observations having the second decile of estimated probabilities, and so forth (Agresti, 2007). The Hosmer-Lemeshow test then uses X^2 to compare observed and fitted counts for this partition. While it does not exactly have a limiting chi-square distribution anymore, the authors note that when the number of distinct patterns of covariate values is close to the sample size, the null is approximately chi-square with (number of groups – 2) degrees of freedom, and significant result on the test would indicate lack-of-fit for the model under consideration.

Simulation Study

Alternatively, another possible approach to evaluate the appropriateness of the above proposed models for the aggregated-level data would consist of a *simulation study* where data are generated based on characteristics observed in the data. With this approach we can study how sensitive the analyses are to several aspects of the design. For example, how sensitive are the model parameters to non-linear trends in the data, how many time points are needed to obtain accurate estimates and predictions, etc.

1.1.4. Evaluating the Likelihood of Achievement of *Salmonella* Reduction Targets

Once the appropriate model(s) have been agreed upon, the likelihood of achievement of the *Salmonella* reduction targets can be assessed by extrapolating the model to *estimate future prevalence* values. It could be of interest to combine this estimate with a measure of precision or a measure of how close the prediction is to the pre-specified target. For instance, a *prediction interval* could be determined which, with a pre-specified coverage probability, would contain a future observation from a population. The methodology is well established in the context of linear regression (see for example Kutner *et al.*, 2004). Denoting by $Y_{h(new)}$ a future observation for which a prediction interval needs to be constructed, then the (1–α) prediction limits, corresponding to X_h are:

$$Y_{h(new)} \pm t(1-\alpha/2; n-p)\sqrt{Var(pred)},$$

where

$$Var(pred) = Var(\mu) + Var(\hat{Y}_{h(new)}).$$

Very little information, however, is available on how to construct such prediction intervals for non-normal responses. This approach could be worth exploring in the context of this project.
A second very interesting approach consists of using the estimated prevalence to construct the underlying binomial distribution and derive from this distribution the *likelihood* to observe, for

example, 1% or less (depending on the reduction target) of infected flocks. For instance, let $\hat{\pi}_{2009}$ represent the predicted prevalence by the end of 2009 and let n_{2009} represent the number of expected flocks to be bred in 2009. Then, $B(n_{2009}, \hat{\pi}_{2009})$ describes the distribution of the (expected) number of infected flocks by the end of 2009.

1.1.5. Investigating the Impact of Different Monitoring Schemes

A notable feature of the aggregated data that should be considered is the fact that important changes were implemented in the sampling scheme. For breeding hen data, starting from 2007, a new Regulation (EC) No 1003/2005[11], which was more intensive than the requirements set out in the former Directive 92/117/EC[12], obliged MSs to run control programmes in breeding flocks for *S.* Enteritidis and *S.* Typhimurium. For laying hen data, changes were implemented in 2008.

Change-Point Models

Since the aggregated-level data covers the years 2004-2009, within which these changes were implemented, it would be of interest to investigate whether such changes in the monitoring scheme affect resulting predictions on the proposed models. A possible approach to look into such an impact would be to consider *change-point models*. Under the latter, different structural forms can be considered for the periods before and after the change, and differences in trends in the two periods can then be tested to assess the impact of the change.
Within, for example, the logistic regression models, an indicator type variable distinguishing observations made before and after the monitoring scheme changed can be included. If there is reason to believe that the trend remains the same in the two periods but with a slight upward or downward shift after the change, then one might start out by contemplating a change-point logistic regression model with period-specific intercepts and a common slope. The assumption is somewhat simplistic, but can be meaningful in some situations. If such an assumption of similar trends over the two periods is not reasonable, a more flexible set of assumptions might be considered. For instance, different slopes, as well as intercepts, might be used for the periods before and after the monitoring changes were implemented. The latter approach essentially fits different logistic regression models for the two periods, and as such, would require more than two data points for the period before the change and also after the change. Moreover, if only few data points are available for each of the periods, the results can be quite unstable.

For the available aggregated-level data, assuming that there are no missing responses for a country from 2004 through 2009, for breeding hen data, the series of prevalence values could be split into the period before the monitoring scheme changed, 2004-2006, and the period after, from 2007-2009. This provides only 3 data points per period. For laying hen data, prevalence values for 2004-2007 comprise the first period, while the second period will consist of prevalence values for 2008 and 2009 only. With such limitations – only two or three data points in the period after the change – this approach will most probably not yield very reliable results. However, with longer time sequences, the approach can be fully explored and may be meaningful.

[11] Regulation (EC) No 1003/2005 implementing Regulation (EC) No 2160/2003 as regards a Community target for the reduction of the prevalence of certain *Salmonella* serotypes in breeding flocks of *Gallus gallus* and amending Regulation (EC) No 2160/2003. OJ L 170, 1.7.2005, p. 12.

[12] Directive 92/117/EC of 17 December 1992 concerning measures for protection against specified zoonoses and specified zoonotic agents in animals and products of animal origin in order to prevent outbreaks of food-borne infections and intoxications, OJ L 62, 15.3.1993, p. 38–48.

Weighting

Another possibility to assess the impact of the change in the monitoring scheme would be to incorporate some form of ***weighting scheme*** within the prescribed models. A larger weight might be taken for the period after the changes in the sampling scheme were implemented, since in this period, monitoring was more intensive. Actual quantification of the weights to be used, however, is often less than straightforward and can pose additional complexity to the modelling exercise.

Adjusted Prevalence Estimates

Finally, one could explore the possibility of incorporating the information regarding the change in monitoring scheme to ***adjust resulting prevalence estimates*** from the proposed models. This approach is described briefly here.

Typically, the tests used to classify whether a flock is infected or not, are imperfect. Since interest is in the true prevalence, π, reflecting the true disease status, and not in the apparent prevalence p (i.e., the prevalence of a positive test), it is important to correct for the misclassification of the test. When sensitivity and specificity of the diagnostic test is known, it is possible to derive the true prevalence from the apparent prevalence, using the following equation (Rogan and Gladen, 1978):

$$p = Se\,\pi + (1 - Sp)(1 - \pi),$$

where *Se* and *Sp* respectively denote the sensitivity and specificity of the testing scheme. Note that, to be a valid estimate, the Rogan-Gladen estimate requires: (1) $Se > 1\text{-}Sp$, i.e., that the probability of a test-positive result of an infected flock is larger than that of a non-infected one, (2) $Se \geq p$, i.e., that the probability of a test-positive result of an infected flock is larger than the probability of a positive test, and (3) $Sp \geq 1\text{-}p$, i.e., that the probability of a test-negative result of a non-infected flock is larger than the probability of a negative test (Rogan and Gladen, 1978).

This correction could be applied to the observed apparent prevalence to obtain an estimate for the true prevalence, π, which could then be used in Model (1)-(2) to evaluate the likelihood of achievement of the *Salmonella* reduction targets.

Often the sensitivity and specificity are not known fixed values (Bollaerts *et al*., 2010). Instead, confidence bounds for the sensitivity and specificity are known from literature. It is not always clear how to use such confidence bounds and how to account for the uncertainty of the sensitivity and specificity in the analysis. In such a setting, the strength of a Bayesian modelling framework might be used to define a prior distribution for the sensitivity and specificity, instead of assuming a fixed value for the diagnostic test characteristics.

A possible way to quantify the prior knowledge on the sensitivity (specificity) of the test is through a beta distribution Beta(a,b) (Faes *et al*., 2010). Parameters of the beta-distribution can then be selected to best represent the bounds of sensitivity (specificity) and the most probably value of sensitivity (specificity). The parameters a and b are estimated using the minimal (m_1), maximal (m_3) and most probable (m_2) value of the sensitivity (specificity), using the following equations (three-point estimation; Grubbs, 1962):

$$a = \frac{\mu^2(1-\mu) - \sigma^2\mu}{\sigma^2} \qquad \text{and} \qquad b = a\frac{1-\mu}{\mu},$$

where the mean μ and standard error σ of the beta-binomial distributed are approximated with

$$\mu = \frac{m_1 + 4m_2 + m_3}{6} \qquad \text{and} \qquad \sigma = \frac{m_3 - m_1}{6}.$$

Note that other approaches are possible as well to obtain a prior distribution describing the knowledge on the sensitivity and specificity. In the context of the aggregated-level data, the information regarding the change in monitoring scheme might be useful in specifying prior distributions.

1.2. Sample-level data

1.2.1. Data

Six MSs provided pilot non-aggregated, detailed sample-level data on a voluntary, confidential basis to EFSA, as well as additional information on the date of setting up of the flock, the date of outcome, the date of slaughtering of negative flocks, and some historical data.

Numbers of sampled holdings

An overview of the reported data on the yearly number of holdings sampled from breeding and laying hens, for each of these MSs, is provided anonymously in Table3. Countries A, E, P and V provided data on *Salmonella* monitoring in both flocks of breeding and laying hens, while countries N and I provided data only on *Salmonella* monitoring in flocks of laying hens.

Since the data reported by countries N and A lacked a unique holding identifier, the trend analysis of the *Salmonella* prevalence data for these two countries was limited to simple models which ignore the natural clustering of flocks within holdings. Note that, unless otherwise indicated, succeeding tables, figures and analyses on the sample-level data are based on the years of availability.

A breakdown of the number of holdings presented in Table 3 in terms of the number of unique flocks sampled is further provided in Table 4. For instance, for country V, of the 69 holdings of breeding hens, 63 holdings had between one and five unique flocks sampled, and in three holdings between six and 10 unique flocks were sampled. From this table it can be seen that some holdings contribute a substantial amount of flocks to the database. For laying hens, it can be observed that most holdings contribute information on between one and five flocks. Country E differs from the other countries, in the sense that in some holdings more flocks were considered (18 holdings with more than 20 sampled flocks).

Table 3. Number of sampled holdings, by country, by production type and by year, sample-level data, six MSs, 2007-2009.

Breeding hens					
Country	**Overall number of holdings**	**Comment**	**2007**	**2008**	**2009**
A	1	No unique holding identifier, 209 samples			1
E	88		78	80	81
N	-				
P	16			15	15
I	-				
V	69				69
Laying hens					
Country	**Overall number of holdings**	**Comment**	**2007**	**2008**	**2009**
A	14	1 unit without unique holding identifier, 13 holdings with unique holding identifier with only 1 sample		13	1
E	83		77	76	70
N	1	No unique holding identifier, 25 samples		1	
P	107			99	105
I	783			301	693
V	1,171				1,171

Table 4. Number of unique flocks tested, by country, by production type, and by holding, sample-level data, six MSs, 2007-2009.

Breeding hens								
Country	**Number of holdings with *X* number of unique flocks tested**							**Total number of holdings**
	1-5	**6-10**	**11-15**	**16-20**	**21-25**	**26-30**	**>30**	
A	-	-	-	-	-	-	-	NA♣
E	21	24	13	10	6	7	7	88
P	2	1	3	3	2	1	4	16
V	63	6	-	-	-	-	-	69
Laying hens								
Country	**Number of holdings with *X* number of unique flocks tested**							**Total number of holdings**
	1-5	**6-10**	**11-15**	**16-20**	**21-25**	**26-30**	**>30**	
A	13	-	-	1+	-	-	-	14
E	39	12	5	9	7	4	7	83
N	-	-	-	-	-	-	-	NA
P	99	4	2	1	-	1	-	107
I	777	4	2	-	-	-	-	783
V	1,099	59	11	1	-	1	-	1,171

+17 flocks without unique holding identifier

♣ Unique flock identifiers missing

A more detailed graphical display of these results, by year, is provided by the density histograms in Figure 5 (for breeding hens) and Figure 6 (for laying hen flocks). In the density histograms, the area covered by each bar (obtained as width x height) represents the proportion of holdings with the number of unique flocks sampled on the X-axis. Since the number of flocks varied widely among MSs, proportions rather than counts are used in this display. Adding up the areas of all bars leads to 1. For any specific value on the X-axis, the height of the bar denotes the proportion of holdings with the number of unique flocks sampled equal to that specific X-value.

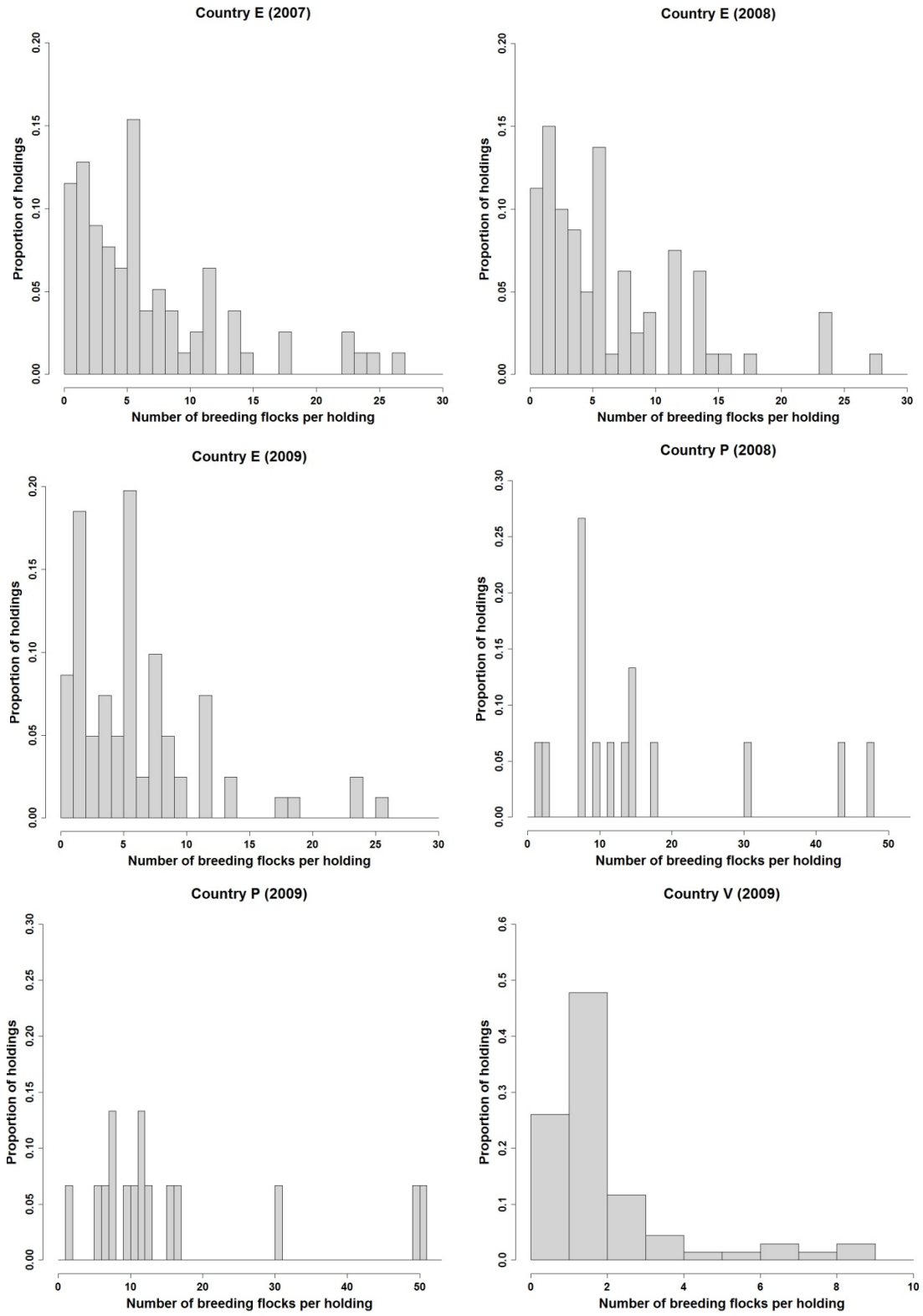
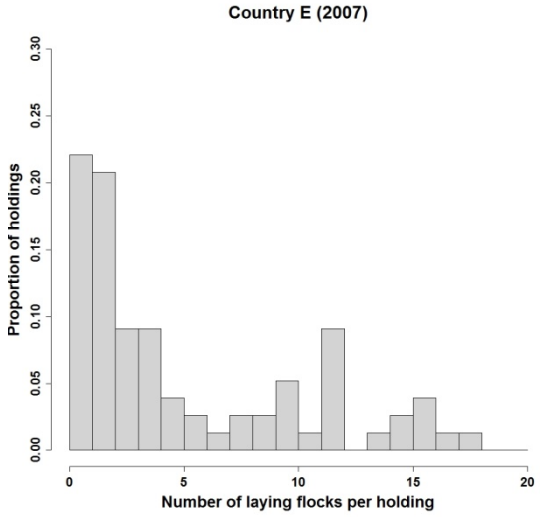
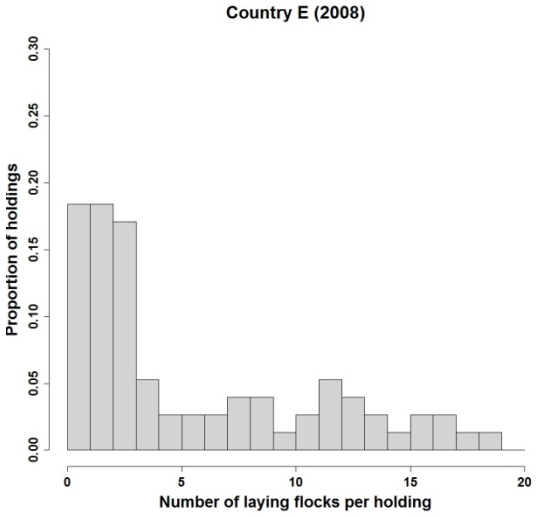
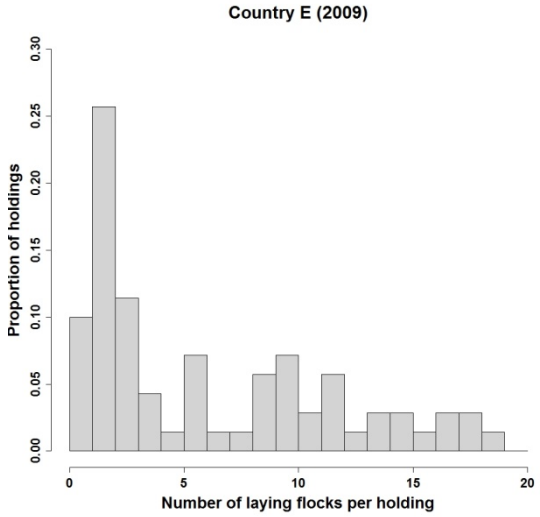
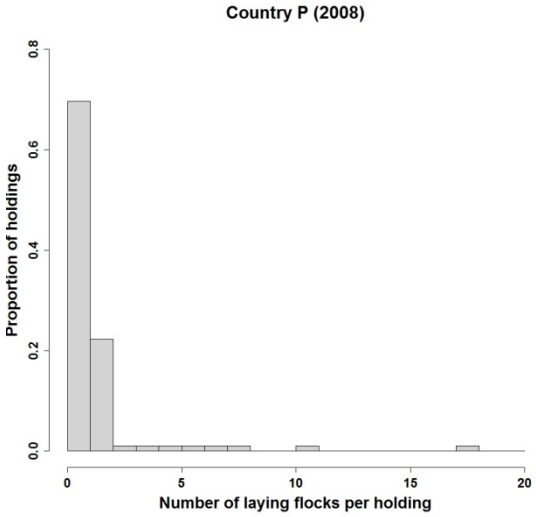


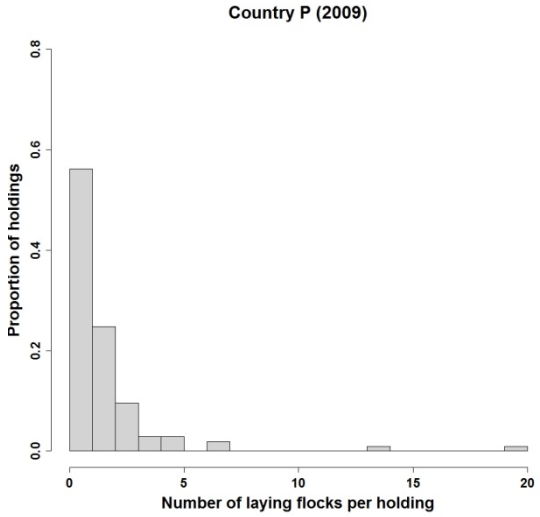
Figure 5. Density histograms for the number of unique breeding flocks per breeding holding, by year, by country, sample-level data, four MSs, 2007-2009.

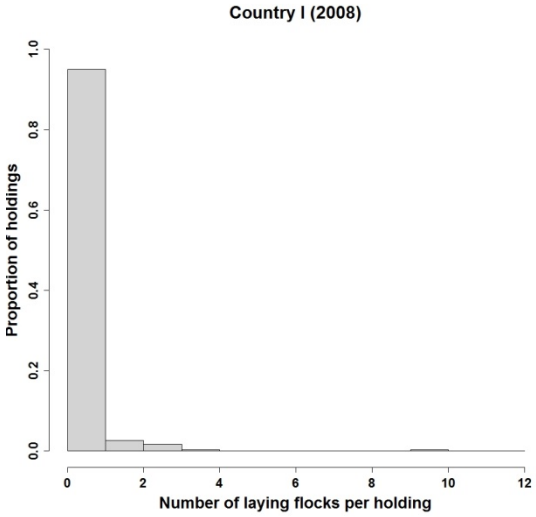












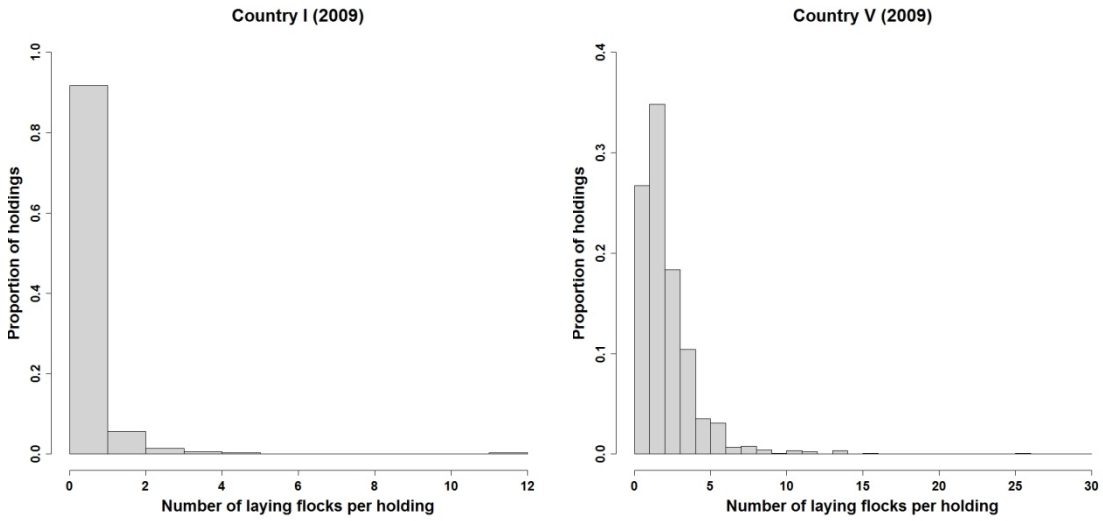


Figure 6. Density histograms for the number of unique flocks of laying hens per laying hen holding, by year, by country, sample-level data, four MSs, 2007-2009.

Numbers of sampled flocks

The number of unique[13] flocks sampled over all holdings, by year, by production type and by country is provided in Table 5. For breeding hens, country A reported on 48 flocks (without providing the holding ID), country E reported on 1,225 flocks, country P on 444 flocks and country V on 171 flocks. For laying hens, countries E, I and V contribute a considerable amount of information, on 949, 1,125 and 3,044 flocks, respectively. It can be noted that yearly values do not necessarily sum to the overall value since flocks can be resampled over several years.

Table 5. Number of unique flocks tested, by year, by country, sample-level data, 2007-2009.

Flocks of breeding hens				
Country	**Overall**	**2007**	**2008**	**2009**
A	48+			48
E	1,225	558	557	534
P	444		244	254
V	171			171
Flocks of laying hens				
Country	**Overall**	**2007**	**2008**	**2009**
A	30		13	17
E	949	428	449	461
N	14+		14	
P	334		175	216
I	1,125		331	794
V	3,044			3,044

+ No unique holding identifier available

[13] A unique flock is reported positive (or negative) only once, per year, irrespective of how many times and how many samples were received/analyzed for those years.

In the monitoring and surveillance programmes, flocks are sampled several times and tested until found positive. Table 6 (first panel) and Figure 7 present a summary of the distribution of the number of samples per flock of breeding hens. Information was available on a total of 22,775 samples. A major part is contributed by country E, for which the majority of flocks have been sampled at least six times.

A similar overview is provided for flocks of laying hens in the second panel of Table 6 and in Figure 8. In total, information on 14,913 samples was available in this subset of the database. Countries V, E and I contributed mostly to the data, followed closely by country P. Most flocks were sampled in between 1 to 5 times on different measurement occasions. Few flocks have been sampled more than 5 times. It can be noted that laying flocks are sampled less frequently than breeding flocks.

Table 6. Number of unique flocks tested and number of samples taken per unique flock, by country, sample-level data, 2007-2009.

Flocks of breeding hens						
Country	**Number of Flocks with *X* Number of Samples**					**Total Number of Samples**
	1-5	**6-10**	**11-15**	**16-20**	**>20**	
A	34	13	1	-	-	210
E	198	102	236	623	66	17,433
P	247	27	49	47	74	3,983
V	94	43	19	6	9	1,149

Flocks of laying hens											
Country	**Number of Flocks with *X* Number of Samples**										**Total Number of Samples**
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**≥10**	
A	14	8	3	1	1	-	-	-	1	2	88
E	223	185	280	189	51	9	10	1	-	1	2,588
N	6	5	3	-	-	-	-	-	-	-	25
P	64	45	57	66	61	26	8	2	4	1	1,168
I	36	1,015	34	4	36	-	-	-	-	-	2,364
V	610	930	550	473	333	90	32	15	4	7	8,680

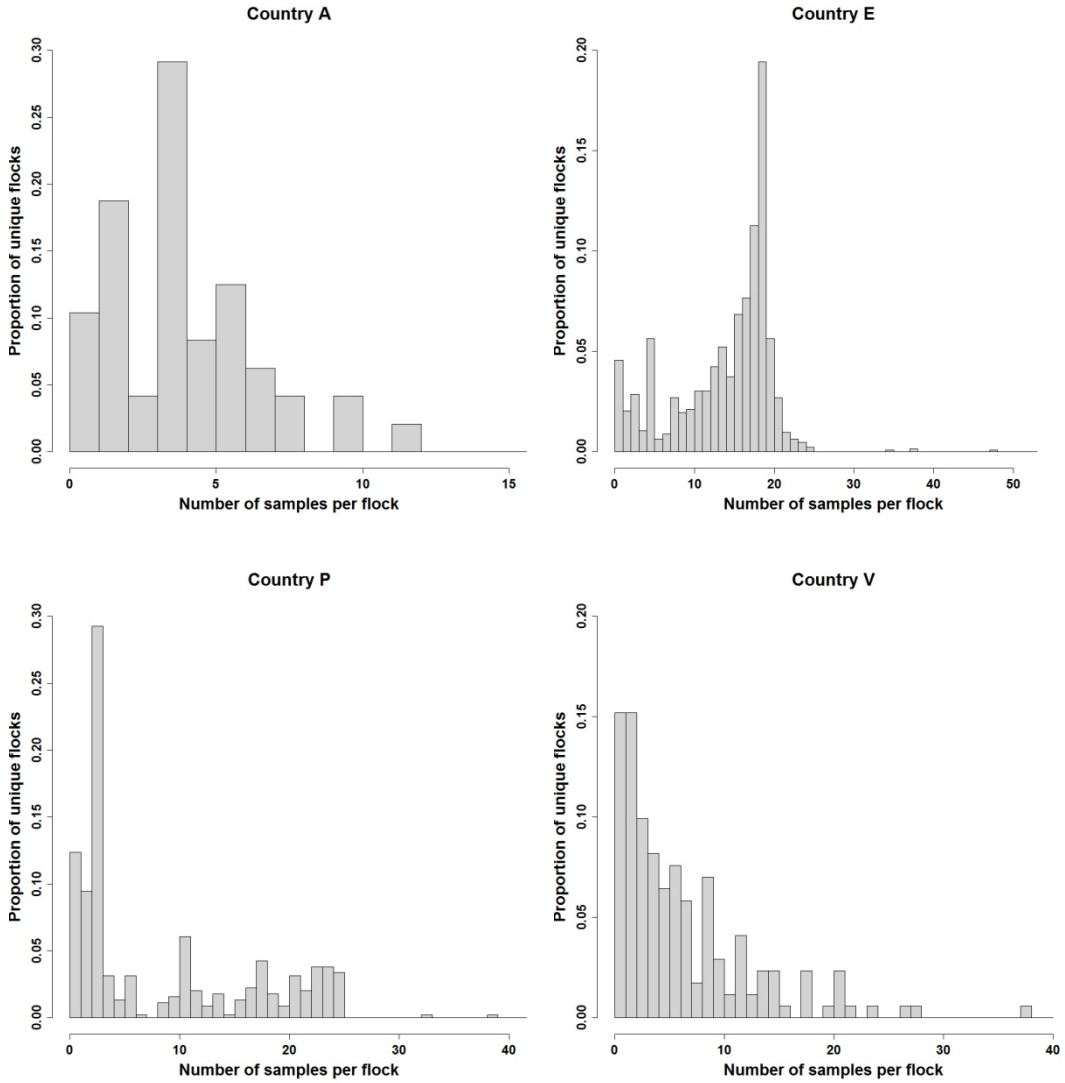


Figure 7. Density histograms for the number of samples taken per unique breeding flock, by country, sample-level data, four MSs, 2007-2009.

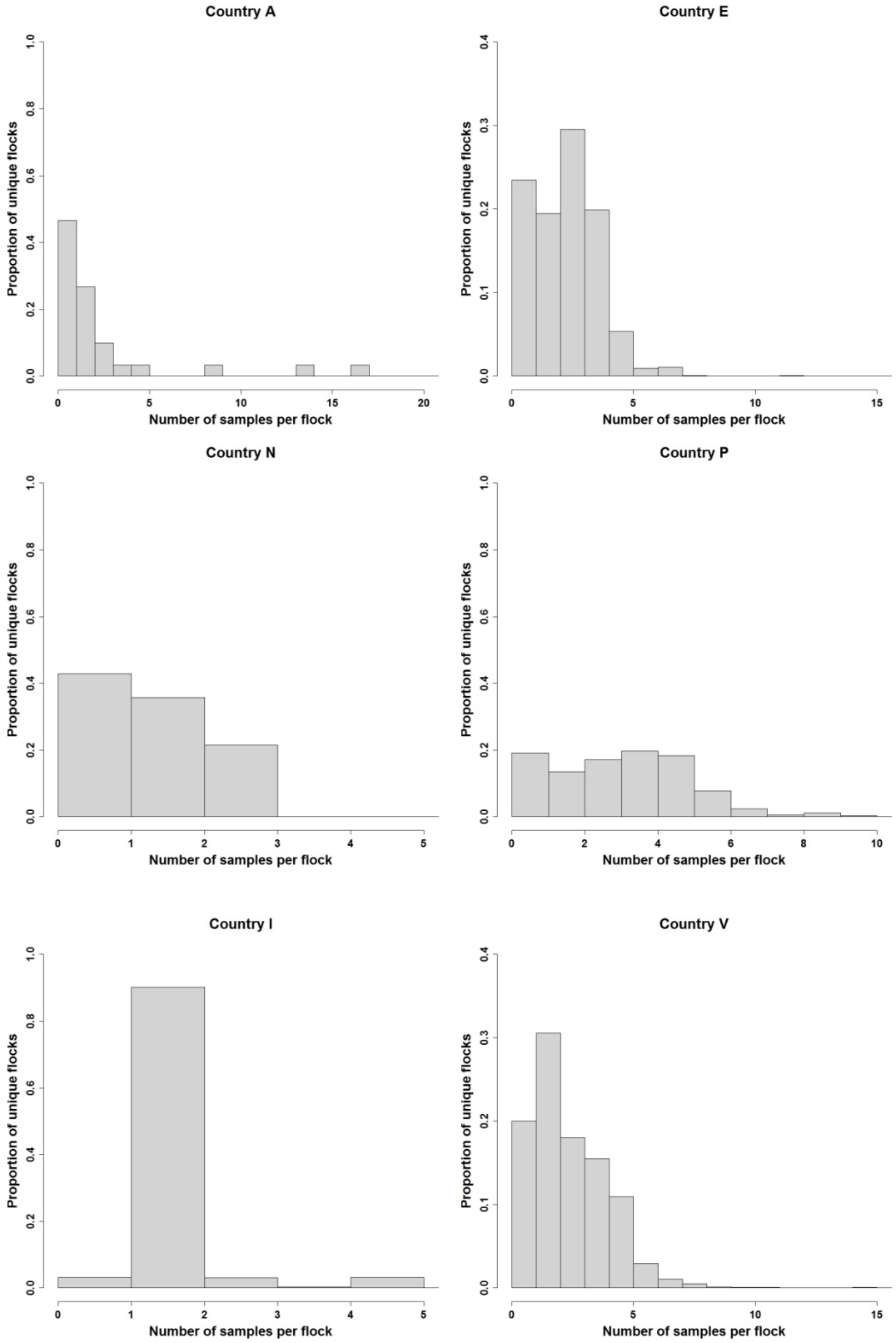


Figure 8. Density histograms for the number of samples taken per unique laying flock, by country, sample-level data, six MSs, 2007-2009.

Country-specific sample-level data

To study how the presence of *Salmonella* evolves over time in flocks, one can compare over time the number of positive flocks over the total number of tested flocks.

The month was considered as the time unit of interest. Since a flock could be sampled more than once per month, it could contribute several test results to a particular month and hence induce bias in the results. To avoid this issue, only one sample result per flock per month was retained in the analysis database. This was done as follows:

- if all sample results of a flock within a month were negative, then the last (in terms of sampling date) negative result was retained,
- if all samples for a flock within a month were positive, then the first (earliest) positive result was retained,
- if some sample results are positive and some negative, then the retained result would be the first instance of a positive result.

Thus, a flock can contribute only one result per month, but it can contribute information for several months.

An overview of the resulting proportion of positive flocks is provided from Table 7 to Table 11. Note that a possibly important limitation of the current data is the uncertainty about how positive flocks are dealt with after testing positive. While a positive breeding flock is destroyed after detection, the procedure for a positive laying flock is less clear. The way of dealing with such flocks depends on MSs-specific legislation, which may not require the removal of a positive flock. Since at this moment there is no way of knowing how the positive flocks were dealt with, and since flocks are tested until positive only, this may introduce bias in the sample-level data analysis. This may also influence a comparison of the sample-level data with the aggregated data.

Country A

Table 7 summarizes for country A the distribution of the number of flocks tested and the test results on a monthly basis. Data on breeding flocks were only available from 2009. In contrast, information on laying flocks is available for 2008 and 2009; however, in 2008 only few (monthly) samples were taken. More information is available on flocks sampled in 2009. Still, there were only very few positive tests (to target *Salmonella* serovars), both in breeding and in laying flocks. Hence, there is not a sufficient amount of non-negative data available to perform a meaningful and appropriate trend analysis for country A. Thus, no illustration was produced to graphically display these results.

Table 7. Number of unique flocks tested† and proportion (number) of flocks positive, country A, by month, by production type, sample-level data, 2008-2009.

	Flocks of breeding hens			Flocks of laying hens		
Month	Number of unique flocks tested	Number of positive flocks	% Positives	Number of unique flocks tested	Number of positive flocks	% Positives
Jan/08				1		
Feb/08				1		
Mar/08				1		
Apr/08				1		
May/08				1		
Jun/08				2		
Jul/08				1		
Aug/08				1		
Sep/08						
Oct/08				1		
Nov/08				2		
Dec/08				1		
Jan/09						
Feb/09	3	1	33.3			
Mar/09	3			1		
Apr/09				11		
May/09	11			7		
Jun/09	11			4	2	50.0
Jul/09	12			1		
Aug/09	10					
Sep/09	13			2		
Oct/09	11					
Nov/09	18	1	5.6	2		
Dec/09	23					

† A flock can contribute only one result per month, but can contribute over several months.

Country E

Table 8 illustrates for country E the distribution of the number of unique flocks tested and the test result by month of sampling for the years 2007-2009. Few breeding flocks tested in 2008 and 2009 are positive. Although census sampling of laying hen flocks only came into force in 2008 by EU law, the number of flocks sampled in 2007 does not appear to differ much from the number of flocks sampled in the following years. Therefore remedial measures for a change in the sampling scheme may not be necessary. Further, it seems that a sufficient number of tests were performed during each month of sampling, with sufficient numbers of positive flocks. This will facilitate a month-based time trend analysis.

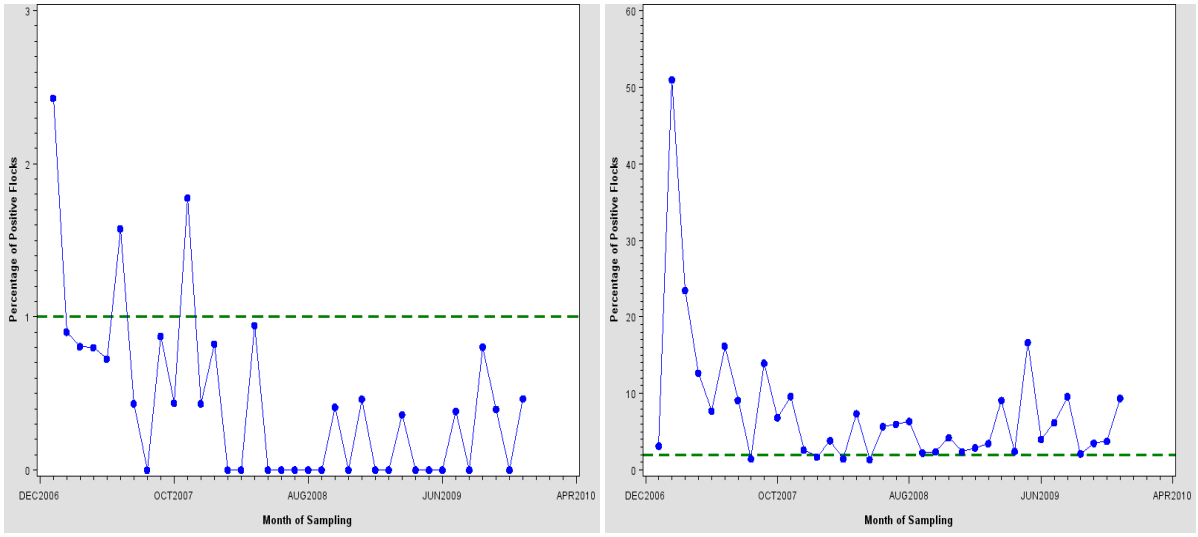
Table 8. N**umber of unique flocks tested**† **and proportion (number) of flocks positive, country E, by month, by production type, sample-level data, 2007-2009.**

	Flocks of breeding hens			Flocks of laying hens		
Month	Number of unique flocks tested	Number of positive flocks	% Positives	Number of unique flocks tested	Number of positive flocks	% Positives
Jan/07	247	6	2.4	32	1	3.1
Feb/07	222	2	0.9	51	26	51.0
Mar/07	248	2	0.8	81	19	23.5
Apr/07	250	2	0.8	79	10	12.7
May/07	276	2	0.7	52	4	7.7
Jun/07	254	4	1.6	68	11	16.2
Jul/07	231	1	0.4	77	7	9.1
Aug/07	235			69	1	1.4
Sep/07	229	2	0.9	43	6	14.0
Oct/07	229	1	0.4	73	5	6.8
Nov/07	225	4	1.8	104	10	9.6
Dec/07	231	1	0.4	77	2	2.6
Jan/08	243	2	0.8	59	1	1.7
Feb/08	208			78	3	3.8
Mar/08	202			69	1	1.4
Apr/08	212	2	0.9	95	7	7.4
May/08	255			74	1	1.4
Jun/08	236			88	5	5.7
Jul/08	215			67	4	6.0
Aug/08	224			63	4	6.3
Sep/08	236			45	1	2.2
Oct/08	244	1	0.4	84	2	2.4
Nov/08	215			71	3	4.2
Dec/08	216	1	0.5	84	2	2.4
Jan/09	226			69	2	2.9
Feb/09	238			58	2	3.4
Mar/09	277	1	0.4	88	8	9.1
Apr/09	279			82	2	2.4
May/09	265			54	9	16. 7
Jun/09	244			75	3	4.0
Jul/09	261	1	0.4	81	5	6.2
Aug/09	243			73	7	9.6
Sep/09	249	2	0.8	47	1	2.1
Oct/09	252	1	0.4	86	3	3.5
Nov/09	244			80	3	3.8
Dec/09	215	1	0.5	64	6	9.4

† A flock can contribute only one result per month, but can be sampled over several months.

The observed (monthly) prevalence in breeding flocks is already much lower than the (annual) reduction target of 1% indicated by the horizontal dashed line (see Figure 9a). Nevertheless, a slight increase is observed towards the end of the study period. Also in laying hens a general downward trend can be observed (see Figure 9b). However, although for this production type *Salmonella*

positivity seems to stabilize as of January 2008, there seems to be a slightly increase again towards the end of the study period, away from the (annual) reduction target of 2% (represented by the green line).



(a) Breeding Flocks (b) Laying Flocks

Figure 9. Monthly proportion of target *Salmonella* serovars-positive flocks, country E, sample-level data, 2007-2009, with *Salmonella* EU reduction target (dashed green line).

Country N

For country N, data pertained only to 2008 and were about laying hen flocks. No information on holding identifiers was available and no information on the time of sampling. The proportion of *Salmonella*–positive laying hen flocks was 52% (13 positive flocks out of 25). These data are less appropriate to perform a time trend analysis.

Country P

Table 9 illustrates for country P the distribution of the number of unique flocks sampled and the test result by month of sampling and production type in 2008 and 2009. In country P only 1 flock of breeding hens tested positive for at least one of the five target *Salmonella* serovars. Hence, while the prevalence observed in this MS for breeding flocks is safely below the EU reduction target, a time trend analysis will not provide any meaningful conclusions.

Table 9. N**umber of unique flocks tested**† **and proportion (number) of flocks positive, country P, by month, by production type, sample-level data, 2008-2009.**

Month	Flocks of breeding hens			Flocks of laying hens		
	N. of unique flocks tested	N. of positive flocks	% Positives	N. of unique flocks tested	N. of positive flocks	% Positives
Jan/08	76			56		
Feb/08	75			23	1	4.3
Mar/08	64			43	3	7.0
Apr/08	85			49	1	2.0
May/08	107			27	6	22.2
Jun/08	103			32	2	6.2
Jul/08	88			53	1	1.9
Aug/08	69			26		
Sep/08	80			56		
Oct/08	94			43	3	7.0
Nov/08	91	1	1.1	28	3	10.7
Dec/08	83			43	2	4.7
Jan/09	81			39		
Feb/09	77			18		
Mar/09	80			41	1	2.4
Apr/09	75			65	2	3.1
May/09	89			26	1	3.8
Jun/09	130			53		
Jul/09	94			21	1	4.8
Aug/09	86			39		
Sep/09	79	1	1.3	55		
Oct/09	101			37	2	5.4
Nov/09	74			58		
Dec/09	102			33	2	6.1

† A flock can contribute only one result per month, but can contribute over several months.

The evolution of *Salmonella*-positivity in laying flocks is graphically illustrated in Figure 10. The (monthly) prevalence seems to vary around 3%, which is slightly above the (annual) EU reduction target (of 2%). Formal procedures are needed to study the presence of a time trend and the evolution of the prevalence.

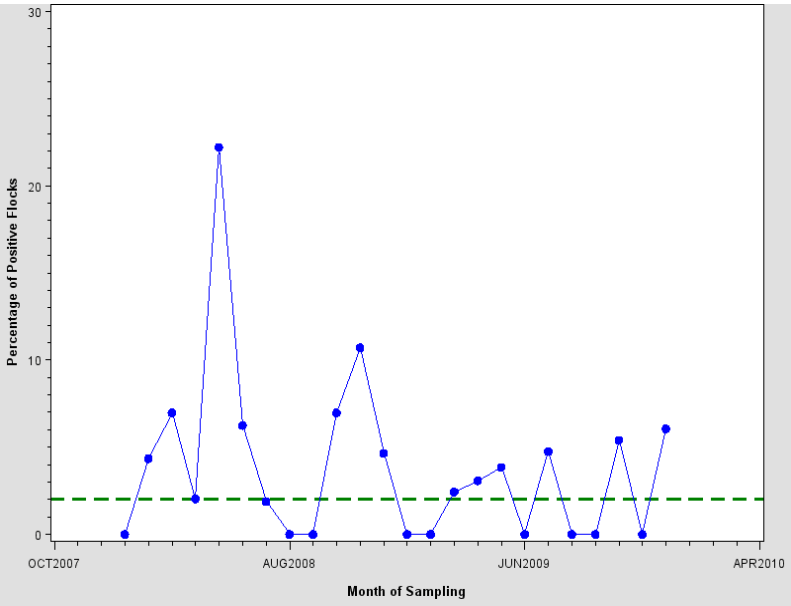


Figure 10. Monthly proportion of two target *Salmonella* serovars-positive laying hen flocks, country P, sample-level data, 2008-2009, with *Salmonella* EU reduction target (dashed green line).

Country I

Table 10 provides an overview of the distribution by month of the tested laying flocks in country I, sampled between January 2008 and December 2009. A general decreasing trend of *Salmonella* positivity can be seen in Figure 11. However, the observed line remains considerably above the (annual) EU reduction target.

Table 10. N**umber of unique flocks of laying hens tested**† **and proportion (number) of flocks positive, country I, by month, by production type, sample-level data, 2008-2009.**

Month	N. of unique flocks tested	N. of positive flocks	% Positives
Jan/08	15	4	26.7
Feb/08	21	3	14.3
Mar/08	23	6	26.1
Apr/08	25	9	36.0
May/08	35	2	5.7
Jun/08	28	5	17.9
Jul/08	23	7	30.4
Aug/08	14	3	21.4
Sep/08	21	2	9.5
Oct/08	45	10	22.2
Nov/08	55	10	18.2
Dec/08	26	4	15.4
Jan/09	57	19	33.3
Feb/09	44	1	2.3
Mar/09	53	6	11.3
Apr/09	41	5	12.2
May/09	97	14	14.4
Jun/09	73	8	11.0
Jul/09	65	4	6.2
Aug/09	43	7	16.3
Sep/09	80	12	15.0
Oct/09	101	7	6.9
Nov/09	87	9	10.3
Dec/09	53	6	11.3

† A flock can contribute only one result per month, but can contribute over several months.

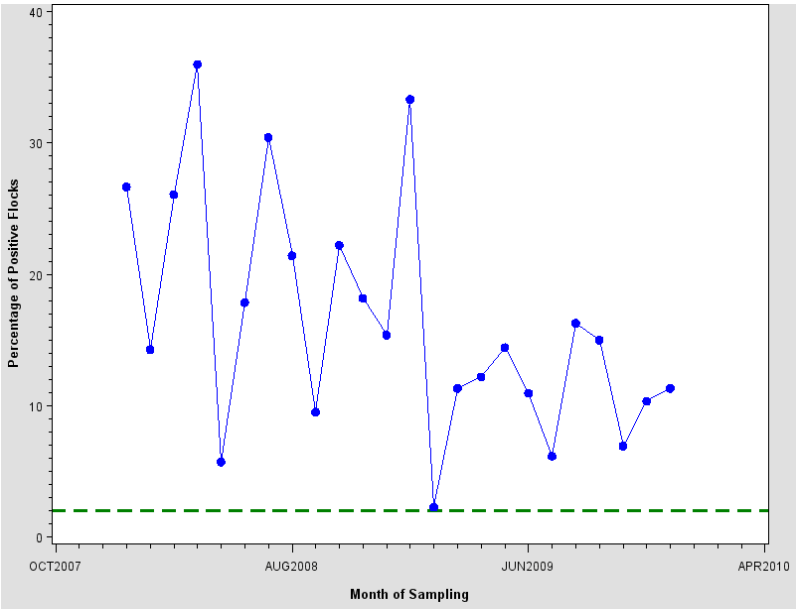


Figure 11. Monthly proportion of two target *Salmonella* serovars-positive laying hen flocks, country I, sample-level data, 2008-2009, with *Salmonella* EU reduction target (dashed green line).

Country V

For country V, data pertained only to 2009. Only two breeding flocks tested positive, one in January and another in February. Hence, there is not a sufficient amount of non-negative findings available to perform a meaningful and appropriate trend analysis for this production type in country V.

In laying hen flocks, a considerable number of flocks were tested each month. However, few laying flocks tested positive. In Figure 12 it can be seen that the observed (monthly) prevalence is already much lower than the (annual) EU reduction target (of 2%). Further, the trend seems to be relatively stable. Nevertheless, few positive outcomes may complicate the computational aspect of the analysis.

Table 11. Number of unique flocks tested† and proportion (number) of flocks positive, country V, by month, by production type, sample-level data, 2009.

Month	Flocks of breeding hens			Flocks of laying hens		
	N. of flocks tested	N. of positive flocks	% Positives	N. of flocks tested	N. of positive flocks	% Positives
Jan/09	61	1	1.6	563	6	1.1
Feb/09	52	1	1.9	349	5	1.4
Mar/09	51			667	3	0.4
Apr/09	51			506	3	0.6
May/09	58			414		
Jun/09	66			639	5	0.8
Jul/09	62			575	9	1.6
Aug/09	61			481	6	1.2
Sep/09	59			650	11	1.7
Oct/09	61			585	13	2.2
Nov/09	62			541	4	0.7
Dec/09	50			708	8	1.1

† A flock can contribute only one result per month, but can contribute over several months.

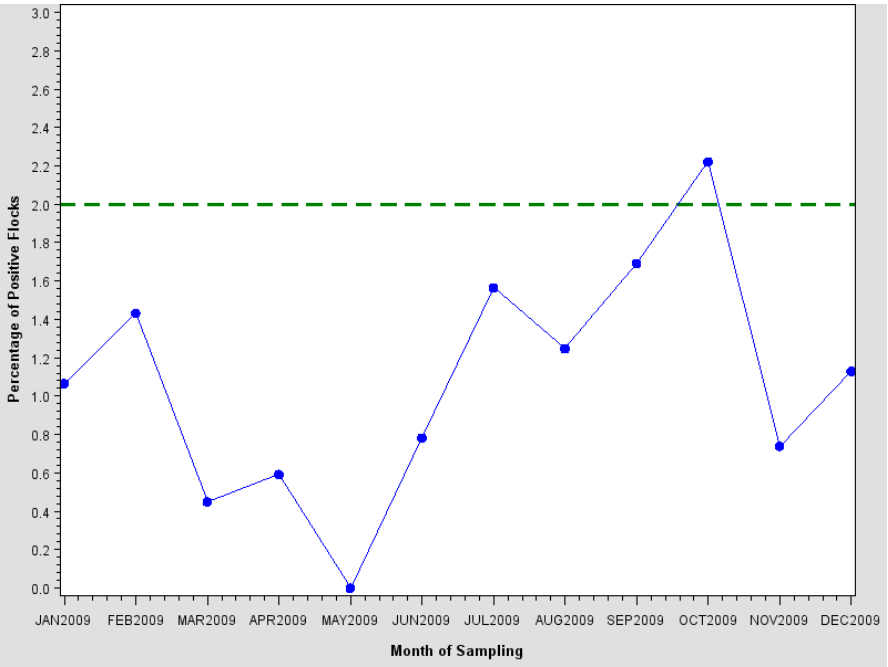


Figure 12. Monthly proportion of the two target *Salmonella* serovars-positive laying hen flocks, country V, sample-level data, 2009, with *Salmonella* EU reduction target (dashed green line).

Further considerations on the sample-level data

Possible issues which make it difficult to obtain meaningful results from the modelling exercise of sample-level data, include:

- the few positive tests results for breeding flocks in most countries providing sample-level data
- the lack of information on the time of sampling for country N
- the lack of a unique identifier of the holding from which samples were taken for countries A and N
- only few monthly samples on laying flocks from country A were provided, especially in 2008. Additionally, the target *Salmonella* serovars were detected in only three out of the 88 samples tested, and all three were detected in June 2009

The sample-level findings presented in this report are based on tests performed by the competent authorities and the industry. An overview of the number of tests executed by competent authorities and industry in each country is provided in Table 12.

Table 12. Distribution of the number of samples by type of check, by country and production type, sample-level data, six MSs, 2007-2009.

Country	Flocks of breeding hens		Flocks of laying hens	
	Industry	Authority	Industry*	Authority
A		210 (100%)		88 (100%)
E	14,632 (84%)	2,801 (16%)	1,685 (65%)	903 (35%)
N				25 (100%)
P	3,336 (84%)	647 (16%)	797 (68%)	369 (32%)+
I				2,364 (100%)
V	1,036 (90%)	113 (10%)	5,336 (61%)	3,344 (39%)

* The label 'industry' covers own checks as well as check by food business operators.

+ 2 samples with missing information

While ignoring this aspect of the data may impose bias on the analysis due to heterogeneity between both testing procedures, as well as different underlying sampling probabilities, it was decided to use all samples in the analysis. Indeed, using only information from competent authorities could result in a considerable reduction of the data, which in some MSs was sparse already.

To conclude this section, an overview of the available information on the production size of a holding is provided. This includes the holding size, the number of birds in a holding, the number of flocks and the average number of hens. In Table 13**,** the symbol X indicates which type of information is available from the participating MSs. When marked as 'X/M' (available/missing) this indicates that the information is not available for all samples. Note that such information can be useful to construct weights which reflect the sampling probability of each flock.

Table 13. Available information at time of sampling for weight construction, sample-level data, six MSs, 2007-2009.

Country	Flocks of breeding hens				Flocks of laying hens			
	Holding size	Number of hens in holding	Number of flocks	Number of hens in flock	Holding size	Number of hens in holding	Number of flocks	Number of hens in flock
A	-	-	-	X	X/M	X/M	X/M	X/M
E	-	X	-	-	-	X	-	-
N					X		X	X
P	X				X		X/M	
I					X	X	X/M	X
V	X	X	X	X	X	X	X	X

X: Available, -: Missing, X/M: incomplete information

1.2.2. Methodology

General Models for Statistical Trends

Starting point for the model building on the sample-level data is again the logistic regression model specified in equations (1) and (2). Model refinement might be achieved by means of the use of exact inferences. This is due to the fact that sample sizes for the sample-level data are on a month-to-month basis and are thus much smaller in comparison to those on a yearly basis (aggregated-level data). Hence, the use of exact inference for the logistic regression model is quite appealing in this setting.

Also, in contrast to the previous setting, information is now available on the holding to which each of the flocks belong. Therefore, sufficient information is available to be able to *account for correlation*. This correlation can be treated as a nuisance characteristic and can be corrected for by means of computing a so-called *design effect*. Roughly, the design effect is a factor comparing the precision under simple random sampling with the precision of the actual design. Standard errors, computed as if the design had been simple random sampling, can then be inflated using the design effect.

In contrast to the previous viewpoint, one can have a genuine scientific interest in the correlation itself. The intraclass correlation should then be addressed in order to obtain valid statistical inference. In this case specialized methods that model the correlation should be used. There are two important families of models which can be used for this purpose: random-effects models and marginal models.

In a *marginal* or *population-averaged model*, marginal distributions are used to describe the outcome vector ***Y***, given a set *X* of predictor variables. A marginal model can be used to evaluate the overall (or population-averaged) trend as a function of covariates. For binary data, one possible approach is to fit a logistic regression model, while correcting the estimated standard errors for clustering. The association structure is typically captured using a set of association parameters, such as correlations or odds ratios. Often, *generalized estimating equations* (GEE) (Zeger and Liang, 1986 and Liang and Zeger, 1986) are used to account for the clustering of outcomes. In this approach, instead of specifying the full distribution for the correlated binary response, we make assumptions about the mean, variance and correlation. For example, it can be assumed that the number of positive flocks y_{ijt} in holding *j* at time point *t* in country *i* has mean and variance specified by:

$$E[y_{ijt}] = n_{ijt}\pi_{it} \qquad \text{and} \qquad Var[y_{ijt}] = n_{ijt}\pi_{it}(1-\pi_{it}).$$

A variety of possible working correlation structures can be considered. Some of the more popular choices are:

- Independence: The simplest choice is the working independence model, i.e.,
$$\mathrm{Corr}(Y_{ij}, Y_{ik}) = 0.$$

- Exchangeable: When there is no logical ordering for the observations within a cluster, an exchangeable correlation structure (or equicorrelated structure) may be most appropriate:
$$\mathrm{Corr}(Y_{ij}, Y_{ik}) = \alpha.$$

- Autoregressive: When repeated samples are taken at the same holding, an autoregressive correlation structure might be of interest, assuming that the correlation between samples depends on the time lag between samples:
$$\mathrm{Corr}(Y_{ij}, Y_{ik}) = \alpha^{|t_j - t_k|}.$$

- Unstructured: A totally unspecified correlation matrix given by:
$$\mathrm{Corr}(Y_{ij}, Y_{ik}) = \alpha_{jk}.$$

Any of these choices is justified, since the GEE method is robust against misspecification of the working correlation structure. However, misspecification of the correlation structure comes at the cost of efficiency of the parameter estimates.

Alternatively, in a *random-effects* model, also called cluster-specific or multilevel model or generalized linear mixed model (GLMM), the predictor variables *X* are supplemented with a vector ***b*** of random effects (specific to the cluster/holding), conditional upon which the components of ***Y*** are usually assumed to be independent. Thus, cluster-specific models are differentiated from population-averaged models by the inclusion of parameters that are specific to the cluster/holding. In random-effects models, the intracluster correlation is assumed to arise from natural heterogeneity in the parameters across clusters (holdings). There are two routes to introduce randomness into the model parameters. The first approach introduces the random effects on the probability scale, such as the beta-binomial model (Skellam, 1948). The second approach introduces the random effects in the linear predictor, yielding the classical mixed-effects models (Stiratelli, Laird and Ware, 1984). A random effects logistic regression model is an example of the second approach, where it is assumed that the number of positive flocks y_{ijt} in holding *j* at time point *t* in country *i* follow a binomial distribution:

$$y_{ijt} \sim \mathrm{B}(n_{ijt}, \pi_{ijt}),$$

with mean modelled through a linear predictor containing fixed regression parameters β_i and holding-specific parameters b_{ij}:

$$\mathrm{logit}(p_{ijt}) = \beta_{0i} + \beta_{1i} x_t + b_{ij}.$$

It is assumed that the holding-specific effects are normally distributed with mean zero and some variance σ_i^2, i.e., $b_{ij} \sim N(0, \sigma_i^2)$.

The above model can be interpreted as a logistic regression model for each holding, where some of the regression parameters are specific (random effects), while others are not (fixed effects). The random effects b_{ij} express how unit-specific trends deviate from the population-averaged trends. In the case of repeated samples, the above model can be generalized by inclusion of a holding-specific time trend (random effect). This is often called a random-slopes model.

Unlike for correlated Gaussian outcomes, the parameters of the cluster-specific and population-averaged models for correlated binary data describe different types of effects of the covariates on the response probabilities (Neuhaus, 1992). The choice between population-averaged (i.e., marginal models) and cluster-specific (i.e., mixed models) strategies may heavily depend on the scientific goals. Population-averaged models evaluate the overall trend in the population. With the cluster-specific approach, the response rates are modelled as a function of time, specific to a holding. In such models, the interpretation of time-related parameters is conditional on a constant level of the holding-specific parameter (e.g., random effect). Population-averaged comparisons, on the other hand, make no use of within-holding comparisons for holding-varying covariates and substantially underestimate within-holding risks. Diggle, Liang and Zeger (1994) and Diggle *et al*. (2002) recommend the random-effects model for inferences about individual responses and the marginal model for inferences about margins, that is, the objectives (or the types of inferences) in a study should determine which suitable statistical model to use. For more details, see e.g., Aerts *et al*. (2002) and Molenberghs and Verbeke (2005).

Since in this report the objective is to study the likelihood of a MS to achieve the reduction target, we are mainly interested in inferences on the level of the population, in this case the MS. Hence, a population-averaged approach using GEE was adopted to account for clustering.

Semi-Parametric Models

When long sequences of test results in holdings over time are present, such as in the sample-level data, the evolution of prevalence may not be constant or linear over time. Additional quadratic or cubic effects of time can be included. However, such trends may lead to improper extrapolations when one is interested in future predictions. In order to make reliable predictions about the progress of MS to achieve the *Salmonella* reduction target, it is therefore necessary to properly capture the form of the progress over time. A more flexible approach to modelling the time trend is through penalized splines (Eilers and Marx, 1996; Ruppert *et al*., 2003).

Consider the observed pairs (y_t, x_t) for each MS. Rather than model $g(\pi_t)$ using a linear relation with time, in nonparametric regression one considers $g(\pi_t) = f(x_t)$, where f is a smooth function determined by the data. The unknown smooth function f can be modelled as a piecewise linear smoother (see also Friedman and Silverman, 1989)

$$f(x_t) = \beta_0 + \beta_1 x_t + \sum_{k=1}^{K} u_k (x_t - K_k)_+$$

where K_k is the location of the *k*th knot $(k = 1, \ldots, K)$ and

$$(x_t - K_k)_+ = \begin{cases} 0, & x_t \leq K_k \\ x_t - K_k, & x_t > K_k \end{cases}.$$

The basis functions represent broken lines with knots K_k as joint points. Following Ruppert (2002), the knots are selected using the quantile spacing approach. Note that it is always possible to choose f sufficiently complicated that it perfectly fits the data. Therefore, one needs to find a balance in the number of knots – a large enough number to ensure the desired flexibility while avoiding over-fitting

of the data. This can be done through the implementation of a penalty, to ensure a smooth fit free from random fluctuations.

Note that the described model can also easily be extended to account for correlation between observations from the same holding. An example in the context of GLMM is provided by Faes *et al*. (2006), whereas Welsh *et al*. (2002) illustrates the use of splines in a GEE framework.

Accounting for Changes in the Sampling Scheme

Another interesting characteristic of the data is the fact that for *laying hens*, census sampling of flocks only came into force (by EU law) since 2008. Hence, data observed prior to 2008 may necessitate correction since they represent only a sample of the total population of flocks.
A key concept in the analysis of such complex survey data is *weighting*. Weighting arises naturally in a variety of contexts:

a) with stratification: different strata have different selection probabilities (e.g. oversampling of some subgroups);
b) with clustering: weights differ within and between clusters; and,
c) in general: units are given probabilities of selection, e.g., proportional to their size.

Assigning weights to the observations is one possible approach to correct for the differences between the complex survey design and simple random sampling. In general, by using weights, we try to "reconstruct the total population" in order to avoid certain strata or subpopulations being over- or under-represented. Many procedures and functions in SAS have a WEIGHT statement or option in order to include weights in the analysis. There exists a vast amount of literature on the use of weights in the analysis of survey data, including several text books such as Barnett (2002), Chambers and Skinner (2003), Knottnerus (2003), and Skinner, Holt and Smith (1989).

Note that defining an appropriate weight is not an exact science. Different choices can be made by different analysts in different contexts, and such a choice may impact the results of the analysis in different ways. Therefore, if a sampling frame was set up to collect data for a particular study, it is recommended to follow the design as closely as possible.

Time-to-Event Methodology

In the detailed database, flocks are repeatedly sampled until they are found positive (for the five – in flocks with breeding hens - or two – in flocks with laying hens - target serovars). Such an approach results in interval-censored data, in the sense that for some flocks the event of interest has:

a) occurred before the first measurement was taken,
b) not occurred at the time the data are analyzed, or
c) occurred during two sampling moments, respectively.

These data are typically analyzed using models for survival data or failure-time data. In this approach, the dependent variable or response is the waiting time until the occurrence of a well-defined event. These are analyzed using predictors or explanatory variables whose effect on the waiting time we wish to assess or control.

In survival analysis, the situation is somewhat different as compared to the case of normally distributed data. Here, we usually do not model the response directly. Instead, we often focus on the hazard, which can be linked to the survival function. This model can be extended relatively easy to

account for clustering by including random effects, which results in the so-called frailty model (see Clayton and Cuzick, 1985).

Suppose Y_{ijk}, the failure time of interest for flock *k* in holding *j* for country *i*, is known only to have taken place in an observed time interval. We will use the method develop by Farrington (1996), and further covered by Collet (2003), which assumes proportional hazards, based on a nonlinear model for binary data to model such problems. The baseline survivor function will be modelled as a step function, where the steps occur at *a* sampling moments $(t_{(1)}, t_{(2)}, \cdots, t_{(a)})$. Thus the baseline survivor function at any time, t_{ijk}, is given by

$$S_0(t_{ijk}) = \exp\left(\sum_{\ell=1}^{a} \theta_\ell d_{ijk\ell}\right), \quad \text{where } d_{ijk\ell} = \begin{cases} 1, & \text{if } t_{ij} \leq t_{(\ell)} \\ 0, & \text{if } t_{ij} > t_{(\ell)} \end{cases}.$$

Thus the response probability can be expressed as follows

$$p_{ijk} = 1 - \exp\left\{-\exp(\beta' x_{ijk}) \sum_{\ell=1}^{a} \theta_\ell d_{ijk\ell}\right\},$$

where x_{ijk} is the vector of values of the *p* explanatory variables for flock *k* in holding *j* in country *i*. These could reflect characteristics of the flock, such as the time of setting up the flock, conditions of the bird population, or holding-level information. These covariates can then be used to study the difference of median survival times between holdings and countries.
In our particular case, having observations from several flocks belonging to the same holding, introduces another level of complexity into the model, in general, called cluster effect. In order to take into account the dependency between flocks belonging to the same holding, a random effect (b_{ij}, also called unobserved frailty) associated to the holding *j* in country *i* can be introduced. As in the previous mixed model methodology, this is an effect that is shared by all animals in the holding, such that, conditional on b_{ij}, observations from holding *j* in country *i* are independent. Thus the probability becomes now:

$$p_{ijk} = 1 - \exp\left\{-\exp(\beta' x_{ijk} + b_{ij}) \sum_{\ell=1}^{a} \theta_\ell d_{ijk\ell}\right\}.$$

Then the likelihood contribution for the *j*th holding can be expressed as the product of differences of the (conditional) survivorship functions evaluated at the observed lower (L) and upper (U) time point:

$$L_{ij}(\beta, \theta) = \int_{b_{ij}} \prod_{k=1}^{n_j} [S_{ijk}(L_{ijk}|b_{ij}) - S_{ijk}(U_{ijk}|b_{ij})] f(b_{ij}) db_{ij},$$

where $f(b_{ij})$ is the assumed density function for the unobserved frailties. Often normally distributed frailties are considered, however the methodology allows the use of other distributions as well. This model can also accommodate both left and right censoring as special cases of interval censoring. Indeed, for left censored the time interval would be $(0, U_{ijk})$ whereas (L_{ijk}, ∞) can be used for right censored observations.

For the sample-level data, it would be possible to consider the above approach to investigate the average time till a flock tests positive. The method, however, would necessitate some quantity by which the time to positivity can be measured. Namely, a 'start date' as well as an 'end date' would be

needed. In the current database, however, only the date of sampling is available for the six MS that provided data. Though many of these MS do have a start date or a set-up date, which could be used as a starting point in calculating the time to positivity, these variables contain large proportions (some even 100%) of missing values. If analysis would be restricted to only those flocks with both 'start' and 'end' dates, the base dataset for analysis would be significantly reduced. Thus, this approach will not be considered further in this report.

1.2.3. Evaluation of the Proposed Methodology

As mentioned in Section 1.1.3, the appropriateness of logistic regression models can be done by means of various goodness-of-fit measures (e.g., Deviance and Pearson statistics) or, for the case of continuous covariates, by means of a Hosmer-Lemeshow test.

Goodness-of-Fit for Logistic Regression with Sparse Data

The standard goodness-of-fit tests for logistic regression based on the Deviance and Pearson statistics tend to behave unsatisfactorily when data are sparse (Kuss, 2002). The latter tests are based on the assumption of large cell counts and may breakdown when many of these cell counts are small (e.g., less than 5). Several alternative measures to assess goodness-of-fit can be found in the literature and have been documented in Kuss (2002). These include proposals by Osius and Rojek (1992), McCullagh (1985), Farrington (1996), White (1982), Orme (1988) and Copas (1989). Details are provided in Kuss (2002) and in the original cited references. These alternative goodness-of-fit measures have been implemented in a SAS macro (%goflogit).

For the sample-level data, for some MS, sample sizes and the observed counts of positive flocks are quite small (see Tables 7-11), with many instances of zero counts for positivity. The standard goodness-of-fit tests using the Deviance and Pearson statistic may thus be somewhat unreliable in assessing the fit of the models applied on these data. Hence, the alternative measures for goodness-of-fit shall also be considered in evaluating the appropriateness of the logistic regression models fitted on these data.

Goodness-of-Fit for GEE

In the presence of clustering, the way to proceed is less clear. An easy extension of the above goodness-of-fit statistics is not available. Several proposals have been suggested over the last few years (Evans and Hosmer, 2004; Evans and Li, 2005). A SAS macro (%goflgee.mac) for implementation is also available from the latter authors. The methods proposed therein, however, consider the case of equally-sized clusters. In the context of the flock database considered in this report, holdings almost always differ in terms of the number of flocks within, and as such, would not be suitable for application of the methods cited in these references.

Cross-Validation

In addition to the standard goodness-of-fit measures and their variations, a second approach that could be used to assess the appropriateness of the fitted models consists of a so-called *cross-validation* technique. This method is often used in many statistical contexts and can likewise be applied in the current setting. The rationale is to fit the model of interest using a portion of the data and reserve the remainder of the cases for a subsequent cross-validation.
For illustration, consider the case of country E where monthly data are available from 2007 to 2009. To do the cross-validation, a model was fitted (e.g., logistic regression or GEE model) using data only from January 2007 up to June 2009. The observations for the last 6 months of 2009 are excluded from

the model fitting stage. Once the model is built, predictions for the last 6 months of 2009 can be obtained. These predictions can then be compared to the actual observed values in order to assess how well the model can predict future or new cases.

A cross-validation approach would be meaningful when there is a reasonable amount of observations to begin with. If there are only a few cases (e.g., 1 year or 12 observations), the procedure might not be entirely worthwhile. Splitting the data will reduce the small number of cases even further and a model built on such a limited set would probably be inadequate. Hence, the cross-validation will only be considered for the models for the country E. The sample-level data for the other MS consist of monthly data for only 1 or 2 years, in which case the cross-validation would probably not be very informative.

1.2.4. Evaluating the Likelihood of Achievement of *Salmonella* Reduction Targets

In section 1.1.4, predictions intervals were described for the case of normal responses, and extension to the non-normal case are limited. In practice, a prediction interval is constructed for the binomial random variable, where the data consist of X out of n trials from a $B(n,\pi)$ distribution, with π representing the probability of obtaining a 'success'. The most commonly used prediction interval for such a variable was constructed by Nelson (1982). Let Y be the future number of successes out of m trials from a $B(m,\pi)$ distribution, then a large sample $(1-\alpha)$ prediction interval is given by

$$\hat{Y} \pm z_{1-\alpha/2}(m\hat{\pi}(1-\hat{\pi})(m+n)/n)^{1/2},$$

where $\hat{Y} = m\hat{\pi}$ and $\hat{\pi}$ can be obtained from an appropriate model. Note that this prediction interval is based on normal approximations, hence it will not perform well when either one of X, $(n-X)$, Y or $(m-Y)$ is small or when π is close to 0 or 1. Other approximations could be considered as well. For example, Nelson (1982) proposed a Poisson approximation when n is large and Y is small. However, in the context of the sample-level time trend analysis, it is not clear to what n should correspond to (e.g., number of tested flocks at last observation, average number of tested flocks, etc). Additionally, one needs a reliable estimate for the future number of trials/tests m to be able to construct such interval. Finally, note that such an interval would provide a prediction interval for the number of positive flocks out of a number of tested flocks; it does not provide a result in terms of probabilities.

The models discussed in section 1.2.2 allow the user to make predictions for future observations. The logistic models which are considered, are formulated on the scale of $\tilde{\pi}$. Hence, these models can provide a prediction for π as well as a $(1-\alpha)$ confidence interval for future unobserved time points. This prediction provides an estimate for the probability of obtaining a positive test, and hence also prevalence of *Salmonella*–positive flocks in a MS. Observe that this approach is different from the ones explained in this section. Indeed, the methods here aim to provide a prediction interval for one particular observation, whether in the context of a continuous outcome or a binomial variable. In our setting, dealing with binary outcomes, this would imply constructing a prediction interval for a new test which can only take two outcome values, 0 or 1. Such an interval would not be very meaningful, unlike a prediction for the probability of observing a positive test. Nevertheless, one should keep in mind that there is a conceptual difference between predicting one particular observation at a new time point and an expected value of the response at the same new time point (Kutner *et al.*, 2004). CIs are much narrower for the mean response. Additionally, they do not take into account the uncertainty of making predictions for new observations.

All analysis were performed and graphs were produced by the software SAS, version 9.2 and R version 2.10.

2. RESULTS

2.1. Aggregated-level data

2.1.1. Simple logistic regression model with linear effect of time

Flocks of breeding hens

The first model considered is a simple logistic regression with a linear effect of time. Note that in this model, time is an indicator for the year of sampling defined such that 2004 corresponds to 1, 2005 to 2, etc. Observe that due to this definition, the intercept can be interpreted in terms of the baseline prevalence at time 0, in this case 2003.

The parameter estimates and odds ratios are shown Table 14. Significant *P*-values for the linear effect of time, shown in grey-shaded cells, are observed for countries 4, 5, 6, 8, 10, 18 and 25. This, along with the algebraic sign of the estimate for the time effect, indicates a significant decreasing linear trend of *Salmonella* prevalence in breeding flocks over time for most of these MSs, with the exception of country 6. For the latter, the estimate of the time effect was positive and significant, implying an increasing trend over time.

These observations and results are further depicted in Figure 13, in which the model fit for each country is plotted. Countries with comparable response values are grouped together for better visualization. In this display, the blue line indicates the observed data, where dots are used to represent the available information, and a break in the line indicates a missing response value. The red line represents the fitted curve obtained from the logistic regression and connects the predictions obtained for each year (even if this information is missing in the database). The grey lines represent the corresponding 95% CIs. Note that the linear trend was modelled on the logit scale of the response, and, as a result, the fit may not necessarily be presented as a straight line. The dashed green line represents the EU reduction target.

In most cases, the predicted curve seems to follow the general trend in the observed values. The countries for which a significant decreasing time effect was obtained (countries 4, 5, 8, 10, 18 and 25) do indeed exhibit decreasing observed percentages over time. Further, from this display it is quite clear that for country 6, observed prevalence values for 2006, 2007 and 2008 were below the reduction target, but in 2009 an increased value that exceeded the reduction target was observed. The significance of this trend might be attributed to the fairly large number of breeding flocks tested for this country (see Figure 3).

Table 14. Parameter estimates and odds ratios from logistic regression of prevalence of *Salmonella*–positive flocks with breeding hens, aggregated data, EUSR 2004-2009.

Country	Parameter	Estimate	standard error	*P*-value	odds ratios	95% CI	
1	Intercept	-5.784	1.223	0			
	time	0.215	0.291	0.461	1.239	0.701	2.192
4	Intercept	-3.597	0.359	0			
	time	-0.182	0.083	0.0295	0.834	0.708	0.982
5	Intercept	-2.099	0.711	0.0032			
	time	-0.719	0.262	0.0061	0.487	0.292	0.814
6	Intercept	-12.424	0.468	0			
	time	1.515	0.086	0	4.55	3.841	5.39
8	Intercept	-0.812	0.111	0			
	time	-0.471	0.032	0	0.624	0.586	0.665
9	Intercept	-2.941	0.239	0			
	time	0.049	0.062	0.4305	1.05	0.93	1.186
10	Intercept	-2.861	0.433	0			
	time	-0.225	0.1	0.0248	0.798	0.656	0.972
14	Intercept	-3.493	0.924	0.0002			
	time	-0.296	0.281	0.2916	0.744	0.429	1.289
16	Intercept	-5.023	0.505	0			
	time	-0.021	0.13	0.8724	0.979	0.759	1.263
17	Intercept	-5.26	0.415	0			
	time	-0.054	0.106	0.6098	0.948	0.77	1.165
18	Intercept	-2.527	0.133	0			
	time	-0.115	0.036	0.0013	0.891	0.831	0.956
22	Intercept	-4.59	0.518	0			
	time	-0.104	0.125	0.4065	0.901	0.705	1.152
25	Intercept	0.861	0.238	0.0003			
	time	-0.78	0.074	0	0.458	0.396	0.53
28	Intercept	-4.631	0.803	0			
	time	-0.01	0.187	0.9581	0.99	0.686	1.429
29	Intercept	-3.94	0.799	0			
	time	-0.121	0.162	0.4561	0.886	0.645	1.218

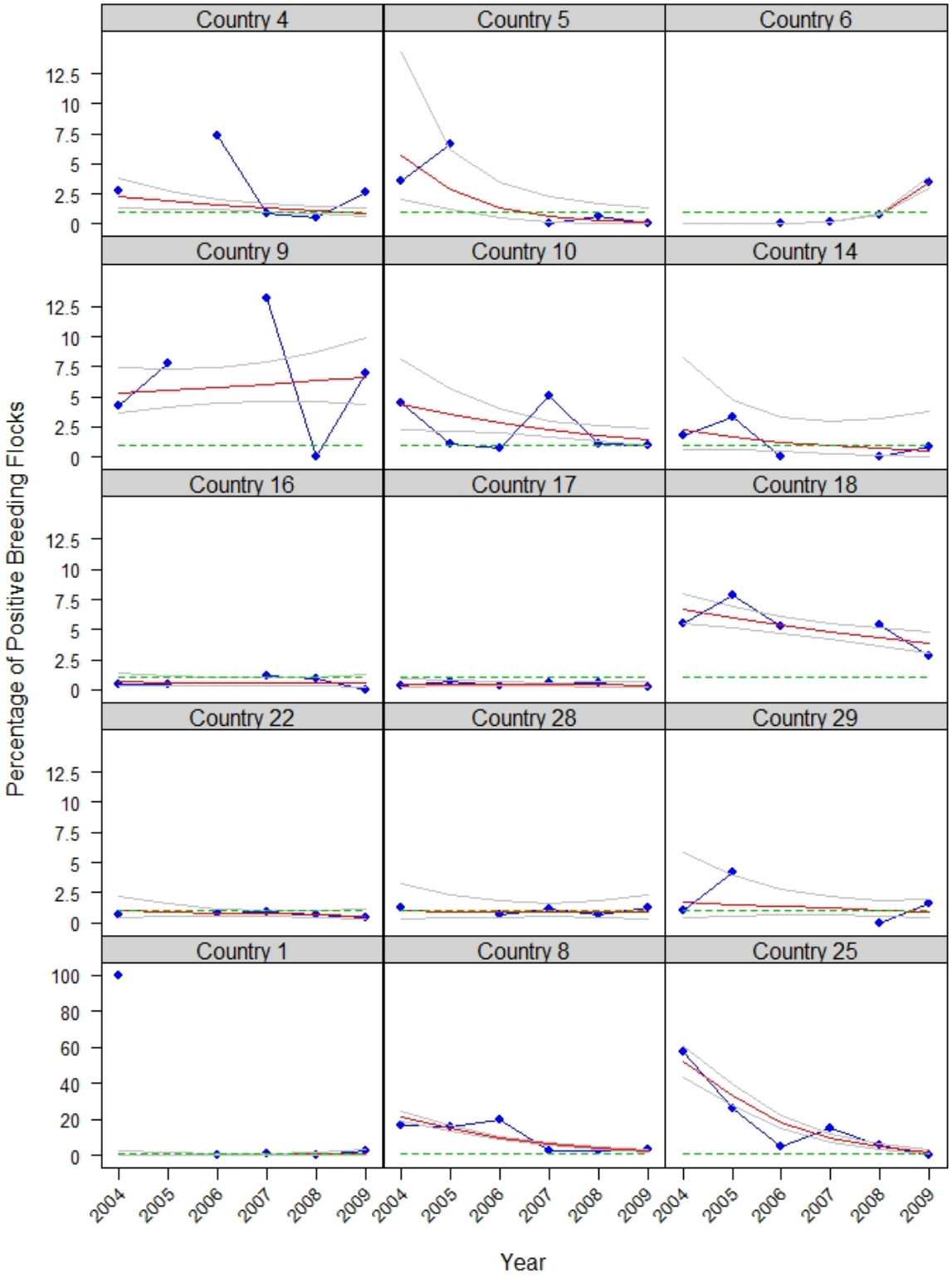


Figure 13 Observed (blue line) and estimated (red line) prevalence of *Salmonella*–positive flocks with breeding hens with 95% Confidence Intervals (grey lines) from logistic regression, aggregated data, EUSR 2004-2009. The dashed green line represents EU reduction target.

With respect to the achievement of the EU reduction target of 1%, among the countries with significant decreasing time trend, countries 4, 5, 8, 10 and 25 have predicted values which are already within or quite close to the target, while country 18 may still need a few years before the target can be achieved. If the current increasing trend for country 6 continues, it will proceed to move further away from the target in the coming years. For the other countries, no significant trend was observed and achievement of the target seems unlikely in the coming years. Else, more information might be required to validate these trends in these countries.

Flocks of laying hens

A similar model was fitted to the data on laying flocks. The results of this analysis are shown in Table 15 and Figure 14. Countries 7, 8 and 22 registered a significant decreasing trend over time, with the 2009 projection for country 22 already quite close to the EU reduction target of 2%. For countries 7 and 8, if the current estimated trend continues, reduction targets may be reached within the next year or two. Countries 6, 13, 14 and 18 were observed to have a significant increasing trend. If such a trend prevails, prevalence values for these countries cannot be expected to attain the EU reduction targets, and are in fact moving further from such.

Table 15. Parameter estimates and odds ratios from logistic regression of prevalence of *Salmonella*–positive flocks with laying hens, aggregated data, EUSR 2004-2009.

Country	Parameter	Estimate	standard error	*P*-value	odds ratios	95% CI	
1	Intercept	-2.839	0.297	0.0000			
	time	-0.147	0.088	0.0961	0.863	0.726	1.027
3	Intercept	-6.202	0.714	0.0000			
	time	0.025	0.178	0.8861	1.026	0.724	1.453
5	Intercept	-3.910	0.483	0.0000			
	time	0.187	0.105	0.0756	1.206	0.981	1.482
6	Intercept	-5.172	0.129	0.0000			
	time	0.332	0.028	0.0000	1.394	1.321	1.472
7	Intercept	1.421	0.630	0.0240			
	time	-0.642	0.144	0.0000	0.526	0.397	0.697
8	Intercept	0.257	0.120	0.0327			
	time	-0.484	0.030	0.0000	0.616	0.580	0.654
10	Intercept	-1.520	0.288	0.0000			
	time	-0.076	0.059	0.2005	0.927	0.826	1.041
13	Intercept	-5.390	0.554	0.0000			
	time	0.462	0.126	0.0002	1.588	1.240	2.032
14	Intercept	-4.449	0.150	0.0000			
	time	0.140	0.036	0.0001	1.150	1.072	1.234
17	Intercept	-3.279	0.104	0.0000			
	time	-0.044	0.025	0.0825	0.957	0.910	1.006
18	Intercept	-4.080	0.129	0.0000			
	time	0.252	0.030	0.0000	1.287	1.214	1.364
19	Intercept	-5.784	0.853	0.0000			
	time	0.068	0.278	0.8069	1.070	0.621	1.844
21	Intercept	-1.055	2.636	0.6890			
	time	-0.707	0.641	0.2701	0.493	0.141	1.537
22	Intercept	-2.583	0.147	0.0000			
	time	-0.172	0.037	0.0000	0.842	0.784	0.905
24	Intercept	-2.973	1.236	0.0162			
	time	-0.741	0.411	0.0711	0.477	0.213	1.066
26	Intercept	-8.091	1.445	0.0000			
	time	0.317	0.299	0.2883	1.373	0.765	2.466
28	Intercept	-5.609	0.509	0.0000			
	time	0.192	0.122	0.1151	1.211	0.954	1.537
29	Intercept	-3.575	0.227	0.0000			
	time	0.115	0.050	0.0207	1.122	1.018	1.238

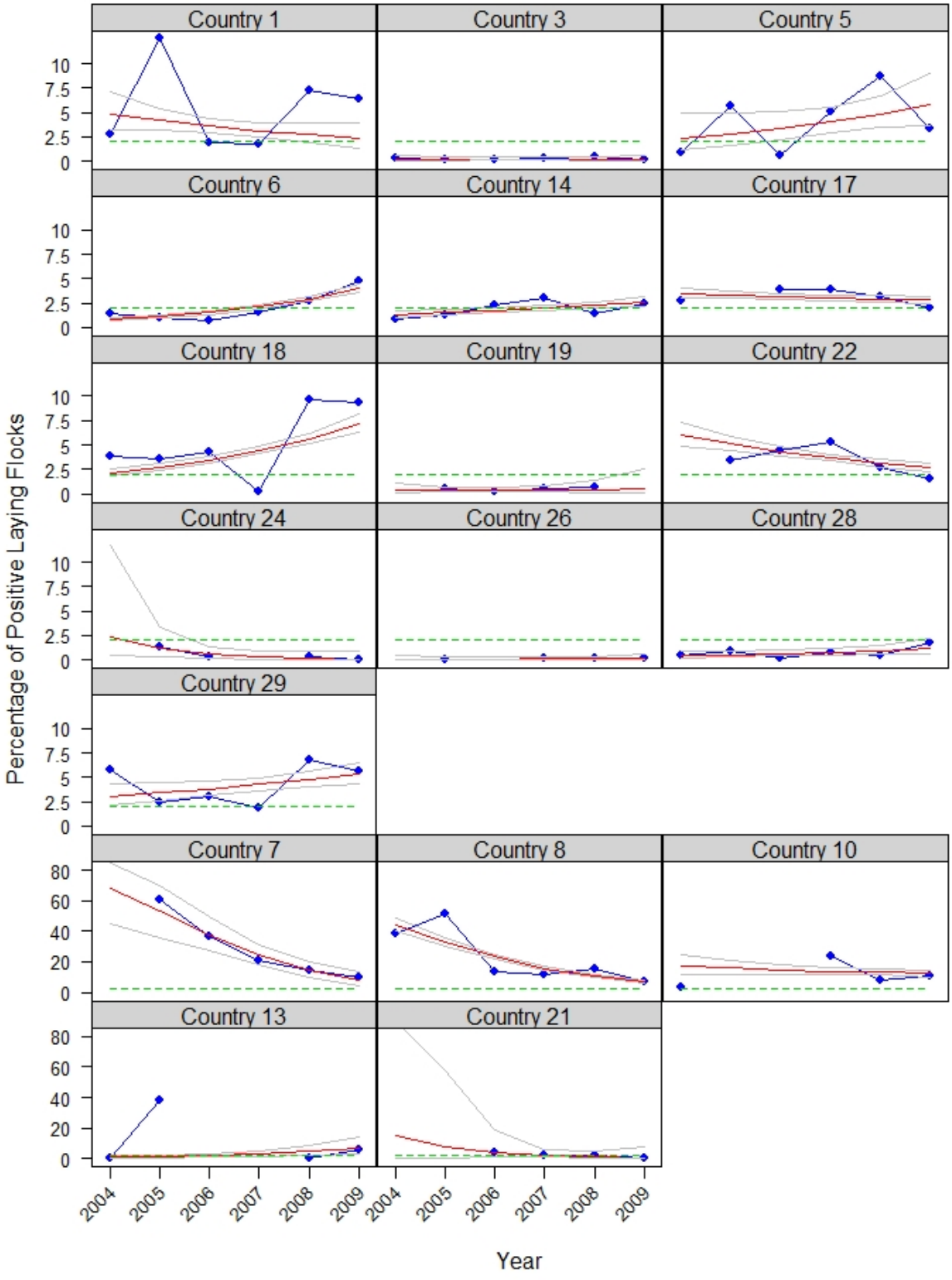


Figure 14. Observed (blue line) and estimated (red line) prevalence of *Salmonella*–positive flocks with laying hens with 95% Confidence Intervals (grey lines) from logistic regression, aggregated data, EUSR 2004-2009. The dashed green line represents EU reduction target.

Evaluation of the fitted logistic regression models with linear term for time

For each of the fitted logistic regression models for *Salmonella* prevalence in breeding and in laying flocks, goodness-of-fit tests based on the Deviance and Pearson statistics were computed. Note that for the case of the aggregated-level data, the primary covariate of interest is time, with a maximum of 6 time points. As such, time, in this setting, is a discrete type of covariate, rather than a continuous one. Hence, the Hosmer-Lemeshow test for goodness-of-fit, which is applicable for models with continuous covariates, is not appropriate here. The corresponding *P*-values for the Deviance and Pearson goodness-of-fit tests for each of the MS are summarized in Table 16. Models with poor fit, i.e., models under which the null hypothesis of a good fit are rejected, are indicated in grey-shaded cells.

Table 16. *P*-Values for Goodness-of-Fit Tests† for the Logistic Regression Models of *Salmonella* Prevalence by Country and Production Type.

Country	Flocks with breeding hens		Flocks with laying hens	
	Deviance	Pearson	Deviance	Pearson
1	0.0003	0	0	0
3			0.6221	0.6177
4	0	0		
5	0.1886	0.2202	0.0018	0.0027
6	0.9492	0.9504	0	0
7			0.7415	0.7429
8	0	0	0	0
9	0.0019	0.0098		
10	0	0	0	0
13			0	0
14	0.346	0.4807	0	0
16	0.0216	0.0595		
17	0.2894	0.3058	0	0
18	0.0055	0.0054	0	0
19			0.1204	0.1464
21			0.646	0.6966
22	0.7516	0.7446	0	0
24			0.4391	0.4515
25	0.0041	0.0078		
26			0.8675	0.8654
28	0.8081	0.8128	0.1061	0.1413
29	0.0011	0.0051	0	0

†The null hypothesis for these tests is good model fit; rejection of the null, *P*-value<0.05, implies poor model fit.

A large number of the models considered indicate evidence of a poor fit. This result is somewhat expected and might be attributed to a number of factors. First, data is available for only 4, 5 or 6 years. With such limited information, positivity trends tend to be poorly estimated by the logistic regression models. Additional data for other years may significantly improve the fit of these models. The poor fit might also be attributed to the fact that the models considered are quite simplistic, including only an intercept and a linear time effect. Addition of other covariates that may be related to *Salmonella* positivity or consideration of a more complex structure might also improve the fit of these models. The latter approach, however, would still be somewhat limited by the number of data points available. Moreover, specific choices would have to be made for each MS depending on the particular shapes/trends in the observed prevalence.

2.1.2. Extended logistic regression model with quadratic term for time

In an attempt to address the poor fit of the above-identified models, other forms of statistical models that might better fit the available data were explored. For a number of MSs, the observed trends in Figure 13 and Figure 14 indicated some amount of curvature, which may not be captured by a purely linear model. A logistic regression model containing a quadratic term for time was thus considered for the countries which indicated a poor fit for the logistic model that was linear in time. The full results are provided in Appendix A. For breeding hens, no improvement was observed in the lack-of-fit statistics. In contrast, for laying hens, the logistic regression with a quadratic term in time improved the fit for countries 6, 17, and 22.

Evaluation of the fitted logistic regression models with quadratic term for time

The resulting parameter estimates and goodness-of-fit test *P*-values are shown in Table 17.

For all cases, the model terms are all highly significant and the goodness-of-fit tests indicate adequate model fit. This improvement is also quite evident in Figure 15, which plots the observed (blue) and predicted (red) prevalence, along with 95% CIs (grey), for the extended logistic model. The estimated curves now follow the observed trends much more closely for these countries than the previous logistic model with only a linear effect of time.

Table 17. Parameter estimates from logistic regression models with quadratic time effect for prevalence of *Salmonella*–positive flocks with laying hens, aggregated data, EUSR 2004-2009.

Country	Model Parameters				Goodness-of-Fit Test† *P*-Values	
	Parameter	Estimate	standard error	*P*-value	Deviance	Pearson
6	Intercept	-3.968	0.216	0.0000	0.0529	0.0650
	time	-0.491	0.133	0.0002		
	time^2	0.110	0.018	0.0000		
17	Intercept	-4.050	0.193	0.0000	0.8361	0.8357
	time	0.558	0.121	0.0000		
	time^2	-0.088	0.017	0.0000		
22	Intercept	-5.527	0.474	0.0000	0.0660	0.0697
	time	1.511	0.254	0.0000		
	time^2	-0.218	0.033	0.0000		

†The null hypothesis for these tests is good model fit; rejection of the null, *P*-value, implies poor model fit.

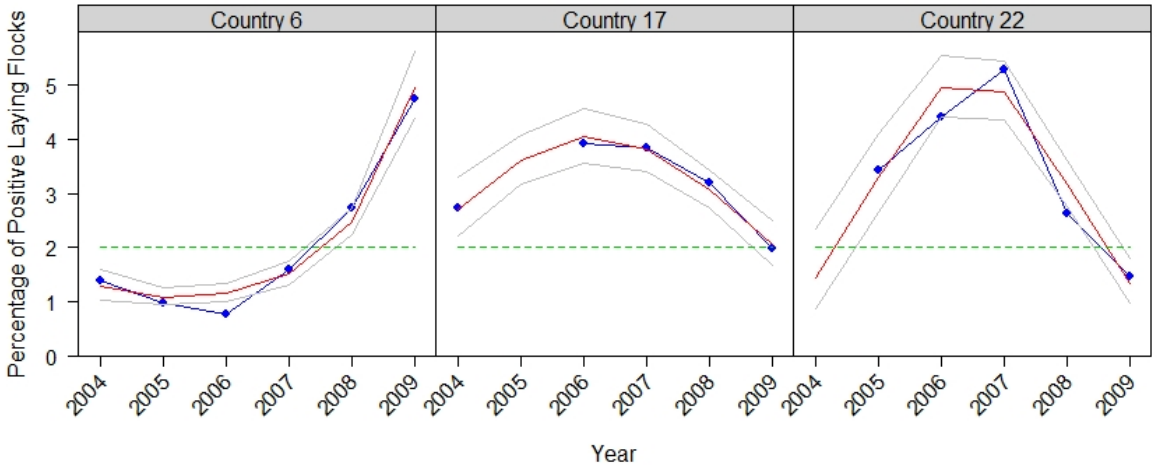


Figure 15. Observed (blue line) and estimated (red line) prevalence of *Salmonella*–positive flocks with laying hens with 95% Confidence Intervals (dashed lines) from logistic regression model with quadratic time effect, aggregated data, EUSR 2004-2009. The dashed green line represents EU reduction target.

For the remainder countries, fully non-linear models were further considered. However, resulting estimates were quite unstable, as might be expected when fitting a complex model on just a few observations. In most cases, convergence was not attained, and in others, resulting estimates could not be deemed reliable.

2.2. Sample-level data

In line with the structure of the methodology to investigate the evolution of the prevalence of *Salmonella*-positive flocks, this section starts with an overview of the general models used to evaluate statistical trends, followed by evaluation of the likelihood of achievement of *Salmonella* reduction targets. In each case, results are presented and discussed separately by production type.

2.2.1. Investigating trends in flocks with breeding hens (countries A and E)

Logistic regression

The first method applied to study the statistical trend in the prevalence of *Salmonella*-positive flocks consists of a simple logistic regression model with a fixed intercept and a time (month) effect. The parameters and odds ratio estimates resulting from the logistic regression (for breeding flocks tested in countries A and E are shown in Table 18 and Table19. Time was defined as the number of months since January 2007, where January 2007 was fixed as 1. This date corresponds to the first available sample-level information in the database (recorded for country E). Its definition was kept fixed for all MSs. As a result the intercept can be interpreted in terms of the prevalence of *Salmonella*-positive flocks at time zero, in this case corresponding to December 2006. For example, for country E, -4.3 is the estimate of the intercept on the logit scale. Using the transformation given in equation (5) in section 1.1.2, at time zero, i.e. $t_j = 0$, or December 2006, the probability of observing a positive flock is estimated as 1.3% for this MS. For country A, a similar interpretation can be obtained. Nevertheless, one should be careful not to put too much emphasis on the estimate for this parameter, as for country A it presents a rather extreme extrapolation. Indeed, only information from 2009 was available to fit the model for country A. Therefore it is very difficult to make extrapolations to a period of time for which no observations are available. Still, this fixed time point for all analyses was kept, as the interest of the assignment is mainly on studying the evolution of the prevalence of *Salmonella*-positive flocks over time.

Population-averaged modelling (GEE)

Samples taken in one holding are expected to be more alike than samples from a different holding, due to conditions specific to each holding. To account for this in the analysis, the clustering of samples within a holding needs to be taken into account. As explained in Section 2.2.1, the implemented approach is a population-averaged one. Indeed, since the main interest for each MS is in its average evolution, rather than a holding-specific approach, mixed effects models are less suitable. Generalized Estimating Equations were fitted to the observed data, assuming an exchangeable working correlation between flocks within a holding. This implies an assumed constant correlation between any two flocks within a holding. The results of this analysis are shown in the second panel of Table 18 and Table 19. The analysis is limited to the data provided by country E, as no holding identifier was available for the flocks of country A.

The results in Table 18 indicate that under both the logistic regression and GEE models, a significant (decreasing) time effect can be observed for country E, but not for A. The former implies that the prevalence of *Salmonella*-positive breeding flocks in country E can be expected to decline further over time. The same, however, cannot be said for country A and this result might be attributed to the fact that only very few samples, as well as a limited period of sampling, are available for this MS.

Exact logistic regression

Since few samples of breeding flocks tested positive, an exact logistic regression was also executed. However, parameter estimates, standard errors and resulting conclusions were very similar to the ones obtained from the logistic regression. Hence they are not reported here but are provided in Appendix B.

Table 18. Parameter estimates, standard errors and corresponding *P*-values from logistic regression and GEE on prevalence of *Salmonella*–positive flocks with breeding hens, sample-level data, countries A and E, 2007-2009.

Country	Parameter	Logistic Regression			GEE		
		Est.	S.E.	*P*-value	Est.	S.E.†	*P*-value
A	Intercept	4.557	7.653	0.5516	-*	-	-
	Time	-0.272	0.249	0.2762	-	-	-
E	Intercept	-4.322	0.260	<.0001	-4.305	0.381	<.0001
	Time	-0.072	0.018	<.0001	-0.068	0.028	0.0169

†Empirically corrected standard errors

* No holding identifier was available for the flocks of country A

The strong similarity between the parameter estimates obtained from the logistic regression and the GEE approach is not unexpected given that the parameter estimators under both models are consistent even when the correlation structure is misspecified. Nevertheless, we can observe some difference between the efficiency of the parameter estimates provided by the two approaches. As expected, larger standard errors are associated with the GEE model(Table 18) and a slightly wider CI for the month effect odds ratio is observed (Table 19). In Table 19, the odds ratio estimates for the effect of time for both the logistic regression and GEE models are presented. For the country E, for breeding flocks, the odds of *Salmonella* in a particular month are only about 93% the odds of *Salmonella* in the previous month, i.e., there is a reduction of about 7% in the odds of *Salmonella* over a month. For country A, the odds ratio is 0.762 and the corresponding CI under logistic regression indicates that this odds ratio is not significantly different from 1, further implying no change in the odds of *Salmonella* from 1 month to the next.

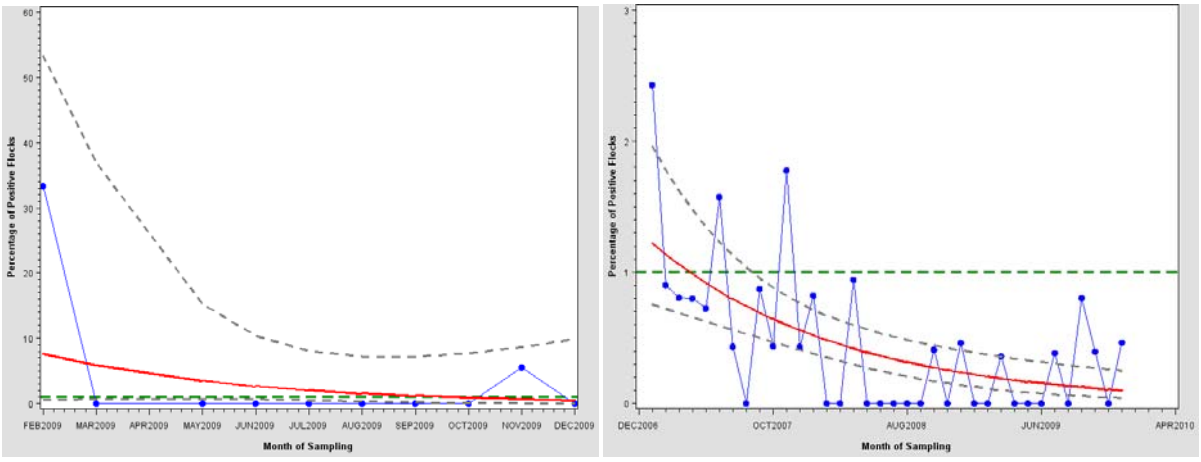
Using the fit of the logistic regression, an expected curve with a 95% CI (grey lines) for the prevalence of *Salmonella* can be produced. In Figure 16, this curve (the red line) is compared to the observed percentage of positive flocks (blue line) for countries E and A.

Table 19. Odds ratio estimates with 95% Confidence Intervals for the time effect from logistic regression and GEE on prevalence of *Salmonella*–positive flocks with breeding hens, sample-level data, countries A and E, 2007-2009.

Country	Logistic Regression			GEE		
	Estimate	95% CI		Estimate	95% CI†	
A	0.762	0.468	1.243	-*	-	-
E	0.931	0.899	0.964	0.935	0.884	0.988

†Based on empirically corrected standard errors.

* No holding identifier was available for the flocks of country A



(a) Country A (b) Country E

Figure 16. Observed (blue line) and estimated (red line) prevalence of *Salmonella*–positive flocks with breeding hens with 95% Confidence Intervals (grey dashed lines) from logistic regression, sample-level data, countries A and E, 2007-2009. The dashed green line represents EU reduction target.

Observe that even though a linear time trend was considered, the red line of fitted values in Figure 16 is not straight. This can be explained by the fact that the linear time trend was modelled on the logit scale (see also Equation 3), whereas the result of the model fit is displayed on the probability scale, after applying the transformation in Equation 3.

Note that in both figures above, the dashed green line represents the reduction target for *Salmonella* in breeding flocks. It can be seen that the observed values for country A in Figure 16 evolve primarily below the target, while the fitted curve stays mostly above the target until the last 2 or 3 months of 2009. This fit, however, is based only on two non-negative observations, which makes it very difficult to model and interpret this trend. For instance, while from these observed data it appears that *Salmonella* is not an issue anymore for the breeding flocks of country A apart from a sporadic 'outbreak', the corresponding CI is wide and contains the target as well as values (considerably) above the target.

A similar graph, illustrating the observed and estimated prevalence for breeding flocks in country E, along with corresponding 95% CIs, as obtained from a GEE model, is provided in Figure 17.

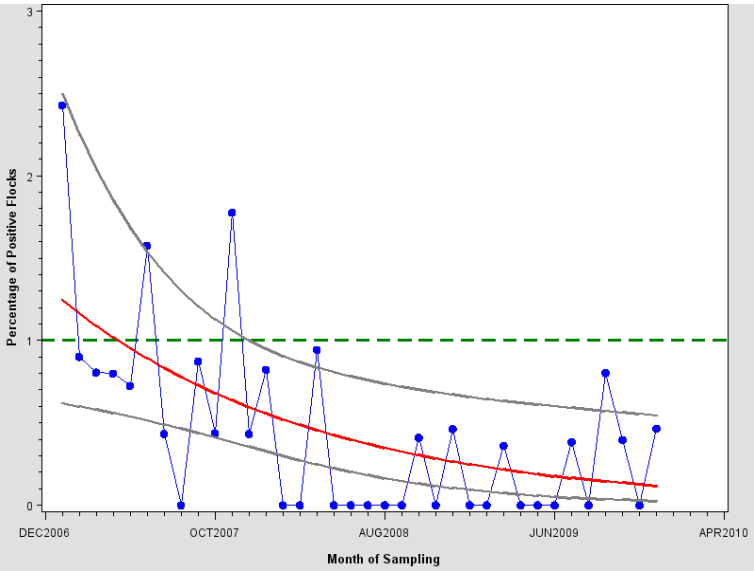
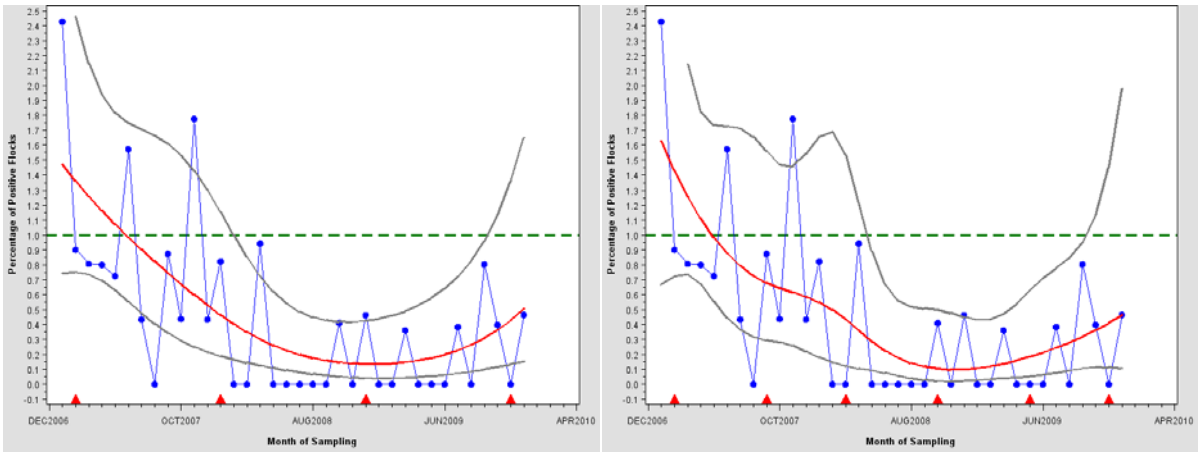


Figure 17. Observed (blue line) and estimated (red line) prevalence of *Salmonella*–positive flocks with breeding hens with 95% Confidence Intervals (grey lines) from GEE, sample-level data, country E, 2007-2009. The dashed green line represents EU reduction target.

The significant decreasing time trends from the results of the logistic regression and GEE analyses, which are also observed in Figure 16 (b) and Figure 17, indicate that there should be no immediate problems anticipated for country E regarding the reduction targets. Indeed, in both figures the estimated prevalence remains safely below the green line, with the exception of a few cases in the earlier periods.

Semi-parametric modelling

For the case of country A, a slight widening of the CIs in Figure 16 (a) can be observed towards the end of the study period. This suggests that a linear trend alone may not be sufficiently suitable for the analysis of these data. In this case a more flexible trend could be considered using splines (see Section on Semi-Parametric Models). Even though such an approach is not available in the standard SAS procedures, (unvalidated) macros can be found on the internet to fit such models. These macros can work with user-specified knots, but can also determine the optimal knots for the procedure. In Figure 18, this approach is illustrated for the data from country E, using GEE, and using different numbers of knots to capture a time interval of roughly a year (4 knots) and approximately a semester (6 knots). The red triangles on the horizontal axis indicate the location of the aforementioned knots.



(a) 4 knots (b) 6 knots

Figure 18. Observed (blue line) and estimated (red line) prevalence of *Salmonella*–positive flocks with breeding hens with 95% Confidence Intervals (grey lines) from a semi-parametric GEE model, sample-level data, country E, 2007-2009. The dashed green line represents EU reduction target.

In this figure the time evolution is modelled allowing different linear trends (on the logit scale) which are connected at each knot. Clearly, the larger the number of knots, the more flexible the fitted curve becomes. Additionally, observe that the distance between two knots is not constant, but depends on the amount of available information. Indeed, knots are selected using a quantile-based spacing method.

Both displays in Figure 18 seem to describe well the trend observed in the data. However, in both settings, the estimated curve and the upper limit of the CI seem to increase towards the end of the study, indicating an increased prevalence of *Salmonella* positive flocks. To select an appropriate model, the QIC (Quasi-likelihood Information Criterion; Pan, 2001) associated with the GEE and the semi-parametric models were compared. This criterion is an informal model selection tool equivalent to the AIC but valid when dealing with non-likelihood based methods such as GEE models. It is based on an estimate of the quasi-likelihood but penalizes for the number of parameters in the model; the smaller the value, the better the model fit. According to this criterion, for the breeding flocks in country E, the GEE model extended with splines based on 4 knots (QIC=488.4) might be preferred over the simple GEE model (QIC=491.1) and the GEE model with splines based on 6 knots (QIC=492.3).

2.2.2. Investigating trends in flocks with laying hens (countries E, P, I and V)

Logistic regression and Population-averaged modelling (GEE)

The results of implementing logistic regression and GEE on sample-level data from laying flocks are summarized in Table 20 (parameter estimates) and Table 21 (odds ratio estimates).

Table 20. Parameter estimates, standard errors and corresponding *P*-values from logistic regression and GEE on prevalence of *Salmonella*–positive flocks with laying hens, sample-level data, countries E, P, I and V, 2007-2009.

Country	Parameter	Logistic Regression			GEE		
		Est.	S.E.	*P*-value	Est.	S.E.†	*P*-value
E	Intercept	-1.722	0.141	<.0001	-1.751	0.206	<.0001
	Time	-0.048	0.008	<.0001	-0.046	0.014	0.0011
P	Intercept	-2.349	0.627	0.0002	-2.442	0.694	0.0004
	Time	-0.045	0.027	0.0924	-0.041	0.026	0.1195
I	Intercept	-0.446	0.348	0.2007	-0.785	0.387	0.0423
	Time	-0.050	0.013	0.0001	-0.040	0.014	0.0042
V	Intercept	-6.420	1.096	<.0001	-7.472	1.304	<.0001
	Time	0.062	0.035	0.0755	0.090	0.041	0.0277

†Empirically corrected standard errors.

A significant downward effect of time is estimated for countries E and I, using both logistic regression and GEE (grey shaded cells). The decreasing trend for country P was not found to be statistically significant. Oddly, in country V a slight (non-significant) upward trend was observed under logistic regression, which, under GEE was significant. Similar conclusions are indicated by the odds ratio estimates and corresponding CIs (Table 21).

Table 21. Odds Ratio Estimates with 95% Confidence Intervals for the Time Effect from logistic regression and GEE on prevalence of *Salmonella*–positive flocks with laying hens, sample-level data, countries E, P, I and V, 2007-2009.

Country	Logistic Regression			GEE		
	Estimate	95% CI		Estimate	95% CI†	
E	0.953	0.938	0.968	0.9546	0.9285	0.9815
P	0.956	0.908	1.007	0.9599	0.9117	1.0107
I	0.952	0.928	0.976	0.9607	0.9346	0.9874
V	1.063	0.994	1.138	1.0942	1.0099	1.1856

†Based on empirically corrected standard errors.

Figure 19 graphically represents the monthly prevalence estimates with 95% CIs for these flocks with laying hens, obtained from a logistic regression (left panels) and GEE model (right panels). Both models provide similar fits to the data.

Country E

In Figure 19 (a) both the prevalence estimate and the CI slightly overestimate the observed numbers for country E. This could be driven by the model trying to accommodate the peak observed in February 2007.

Country P

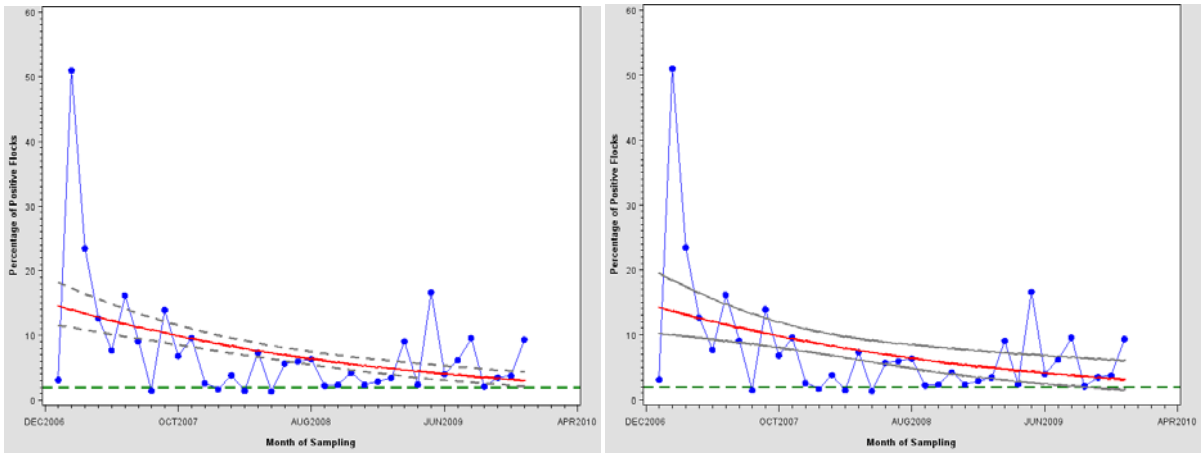
The model fit for country P in Figure 19 (b) suggests a slightly downward trend. Nevertheless, as shown in Table 20, this downward trend is not significant. Therefore, given the current data and evolution, country P may not reach the *Salmonella* reduction target set for laying flocks.

Country I

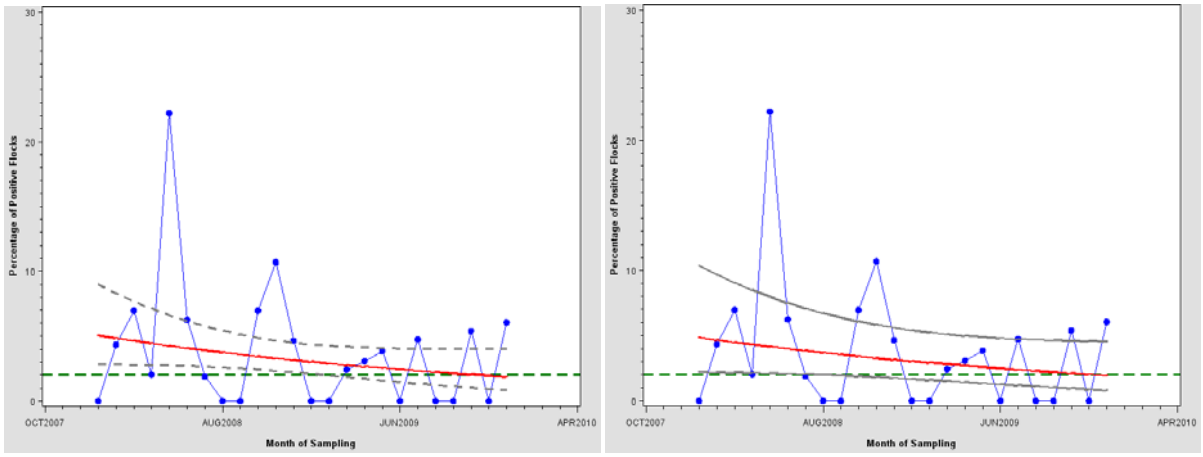
As shown in Figure 19 (c), while still considerably above the reduction target, the significant downward trend observed in country I, suggests that this MS can reach the 2% limit for *Salmonella* positivity within a few more months.

Country V

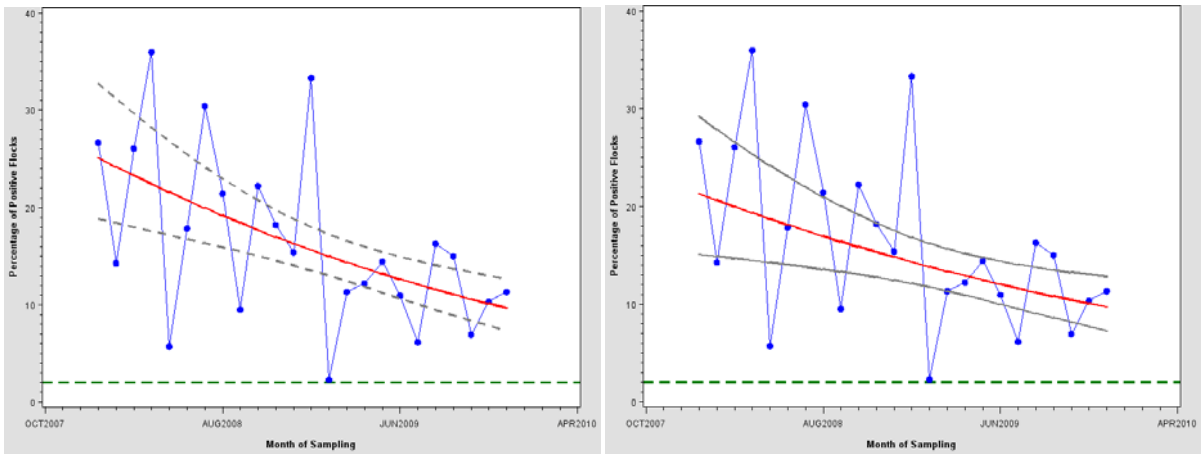
Figure 19 (d) illustrates the slight upward trend estimated for the prevalence of *Salmonella*–positive flocks with laying hens in country V. Nevertheless, the prevalence remains below the target reduction of 2%.



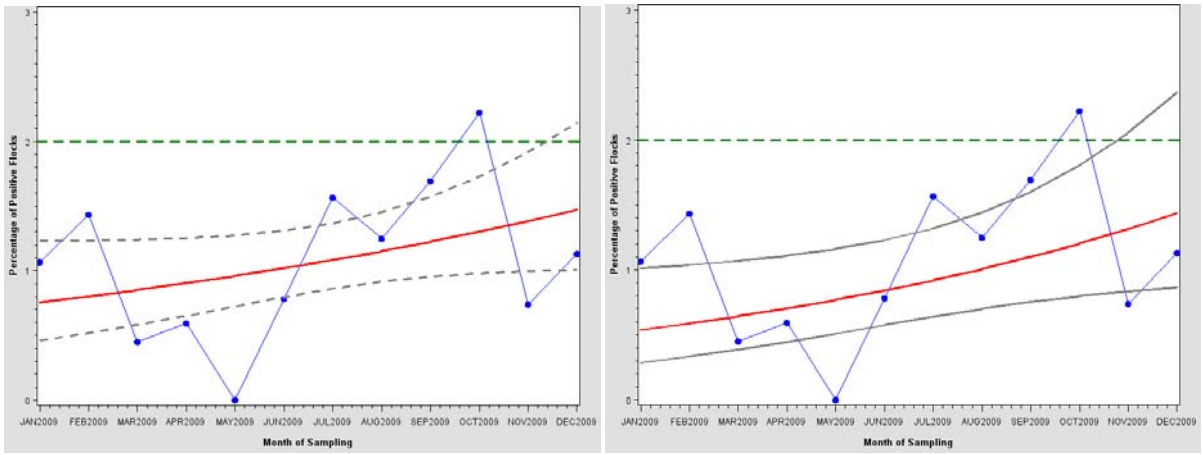
(a) Country E



(b) Country P



(c) Country I



(d) Country V

Figure 19. Observed (blue line) and estimated (red line) prevalence of *Salmonella*–positive flocks with laying hens with 95% Confidence Intervals from logistic regression (grey dashed lines, left figures) and GEE (grey lines, right figures), sample-level data, countries E, P, I and V, 2007-2009. The dashed green line represents EU reduction target.

Semi-parametric modelling

In a next step, one can also consider the semi-parametric approach, based on splines to allow a more flexible modelling of the time trend. This approach is graphically illustrated in Figure 20.

Country E

As in the analysis of breeding flocks in country E, 4 and 6 knots were considered. The same was used for the analysis of laying flocks. The results of this analysis are shown in Figure 20 (a). In this setting, the difference between the two settings is small. In comparison with Figure 19 (a), the model is now better able to capture the peak at the start of the study.

Country P

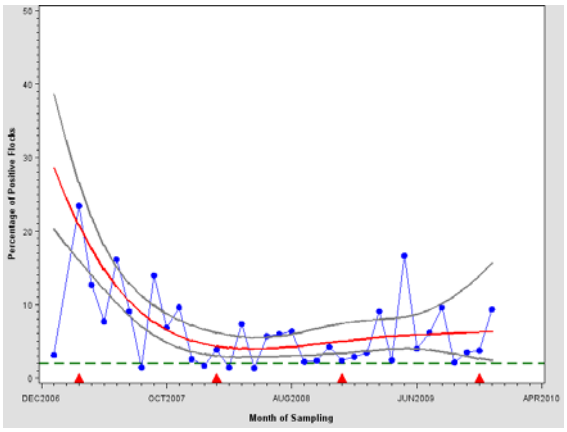
Since for country P only 2 years of information is available, the number of knots was fixed to 3 and 5, resulting in time intervals between the knots of roughly 1 year and 6 months. The result is shown in Figure 20 (b). Here, the difference between the two settings is quite pronounced. While the setting of 3 knots shows a general downward trend, for 5 knots, a slight upward trend is observed in the beginning, followed by a decreasing trend. These observations, however, should not be over interpreted, since the logistic regression and GEE models indicated no significant trends in *Salmonella*-positivity over time.

Country I

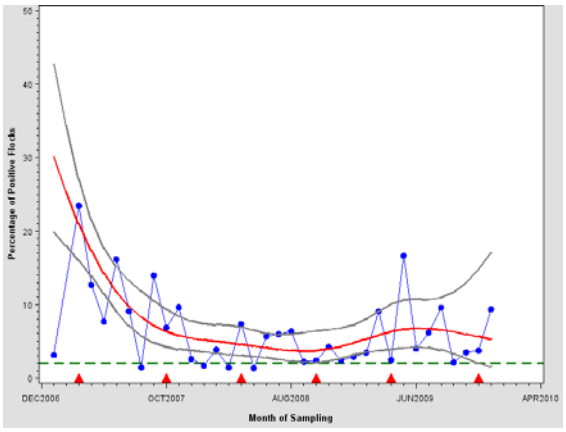
The same numbers of knots were also considered for country I, for which two years of information were also available. The results of this analysis are shown in Figure 20 (c). As in the logistic regression with a linear time trend, a downward trend is observed. This suggests that a more complicated model may not be necessary for this MS.

Country V

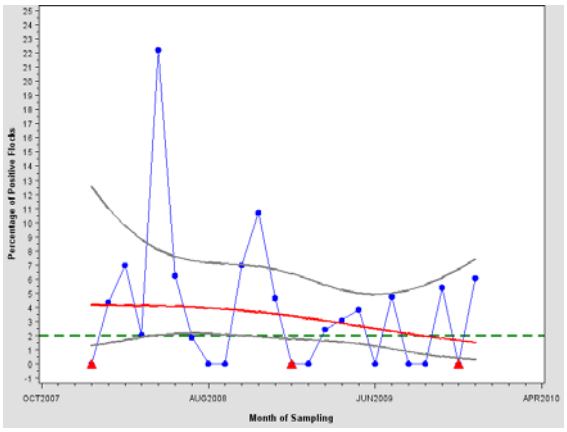
Since for country V only one year of data is available, the number of knots was limited to 3 and 4, resulting in time intervals between the knots of roughly 6 and 4 months. In Figure 20 (d), the fitted curves for 3 and 4 knots are generally similar, with slightly more curvature in the case of the latter. As mentioned previously, increasing the number of knots to the maximum brings about a perfect fit to the data. While such a model might be able to describe well the data at hand, it may not be very appropriate to obtain future predictions.



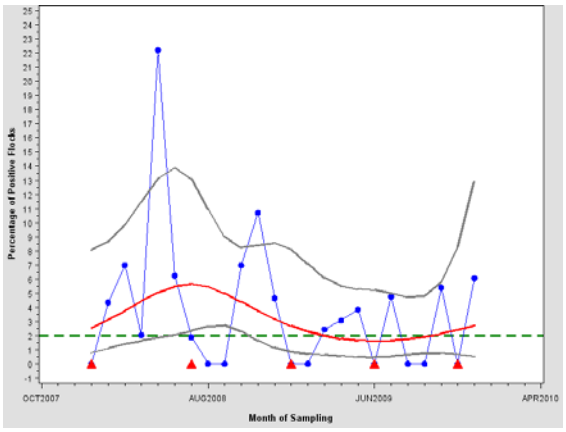
(a) Country E: 4 knots



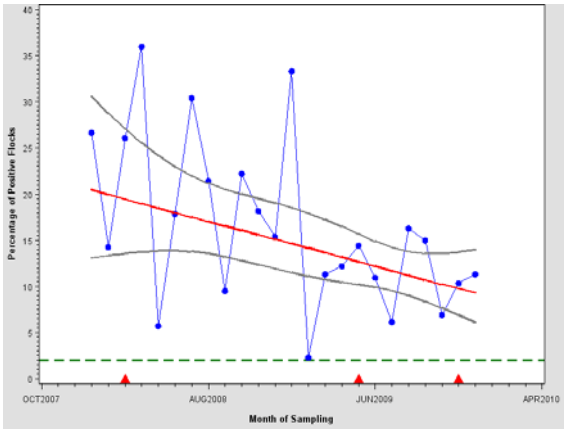
(a) Country E: 6 knots



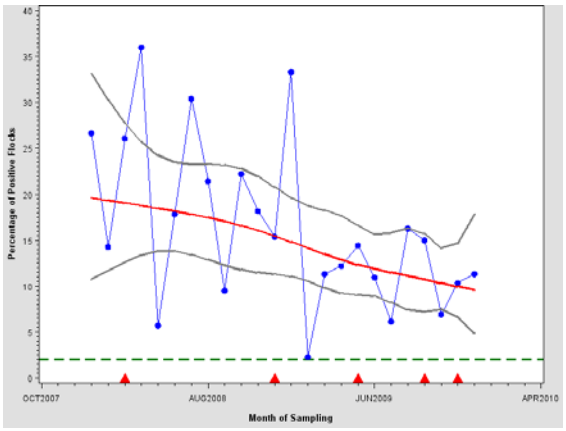
(b) Country P: 3 knots



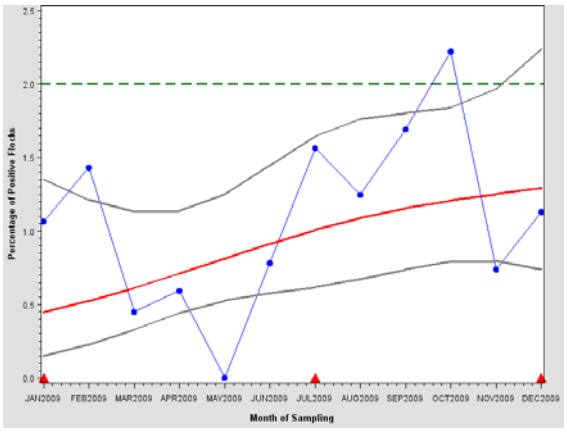
(b) Country P: 5 knots



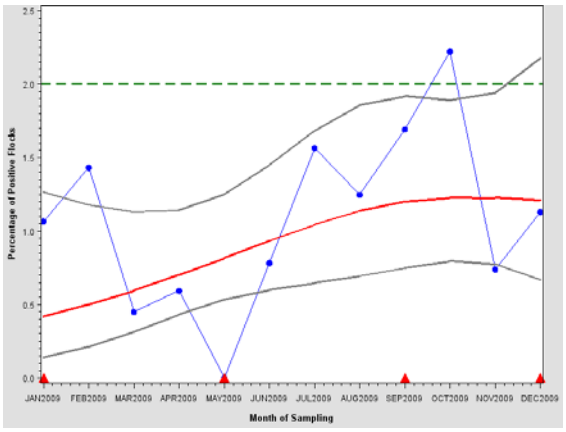
(c) Country I: 3 knots



(c) Country I: 5 knots



(d) Country V: 3 knots



(d) Country V: 4 knots

Figure 20. Observed (blue line) and estimated (red line) prevalence of *Salmonella*–positive flocks with laying hens with 95% Confidence Intervals (grey lines) from a Semi-Parametric GEE Model, sample-level data, countries E, P, I and V, 2007-2009. The dashed green line represents EU reduction target.

A comparison of the different models via QIC (Quasi-likelihood Information Criterion) is provided in Table22. With the exception of country E, the simpler GEE model seems to be preferable over the extended GEE model. This might have already been expected from the observations made on the fitted curves depicted in Figure 20. For country E, however, the criterion selects an extended GEE model based on 4 knots, which shows significant improvement in QIC over the simple GEE model. Note that this extended GEE model with 4 knots was also the choice for breeding flocks in this country.

Table 22. Overview of Quasi-likelihood Information Criterion Values for the Different Models Considered, prevalence of *Salmonella*–positive flocks with laying hens, sample-level data, countries E, P, I and V, 2007-2009.

Country	Model	QIC
E	GEE	1307.9
	GEE + splines (4 knots)	**1270.3**
	GEE + splines (6 knots)	1274.6
P	GEE	**278.8**
	GEE + splines (3 knots)	282.1
	GEE + splines (5 knots)	284.3
I	GEE	**922.6**
	GEE + splines (3 knots)	924.5
	GEE + splines (5 knots)	928.7
V	GEE	**810.1**
	GEE + splines (3 knots)	813.4
	GEE + splines (4 knots)	813.6

2.2.3. Evaluation of the Proposed Methodology

In the second and third columns for breeding flocks of Table 23, the Deviance and Pearson statistics for the logistic models in Table 18 are presented. These statistics were obtained by grouping the observation by month of sampling during the study period. For both countries, both test statistics fail to reject the null hypothesis of lack of fit. Hence, the logistic regression models considered for the breeding flocks of countries A and E seem to fit fairly well.

These results, however, have to be interpreted with caution. Even when the information is aggregated by month of sampling, few positive breeding flocks were observed in these participating MSs. As noted at the end of Section 1.2.1, when expected counts below 5 occur, the limiting distribution may not be valid. Therefore, the Hosmer-Lemeshow test was also applied to the above mentioned models and the results of which are further given in Table 23. Although with slightly different *P*-values, similar conclusions hold, implying no evidence of lack-of-fit for the logistic models considered for the breeding flocks of these two countries.

For laying flocks, the results from the Hosmer-Lemeshow test seem to indicate that the logistic regression model for country E may have some lack-of-fit problems. Nevertheless, recall that country E contributes a considerable amount of information, which makes it easier to detect possible issues with the model.

Because the models being evaluated here were fitted to sparse data, the goodness-of-fit tests variants under such, in particular, White's (1982) information matrix (IM) test was also considered and the resulting *P*-values are shown in the last column of Table 23. For most cases, the results led to the same conclusions as those for the Hosmer-Lemeshow test. For a few cases, however, the SAS macro failed to provide the test results.

Table 23. *P*-Values for Goodness-of-Fit Test† Results for the Logistic Regression Models on prevalence of *Salmonella*-positive flocks with breeding or laying hens, sample-level data, countries A, E, P, I and V, 2007-2009.

Flocks of breeding hens				
Country	**Deviance, G^2**	**Pearson, X^2**	**Hosmer-Lemeshow**	**White's IM Test**
A	0.492	0.227	0.303	0.102
E	0.218	0.227	0.113	no result
Flocks of laying hens				
Country	**Deviance, G^2**	**Pearson, X^2**	**Hosmer-Lemeshow**	**White's IM Test**
E	<.0001	<.0001	<.0001	<.0001
P	0.002	0.001	0.051	0.098
I	0.010	0.016	0.284	0.707
V	0.024	0.070	0.070	no result

†The null hypothesis for these tests is good model fit; rejection of the null, *P*-value<0.05, implies poor model fit.

As mentioned in Section 1.2.3, extensions of these statistics for GEE models are limited. Current implementations are restricted to cases where clusters (e.g., holdings) are of equal size. For the particular data setting being investigated here, it is more often the case that the clusters (holdings) do not have the same number of subjects (flocks).

To further evaluate the proposed methodology, a cross-validation approach was considered for country E. Under this method, models are built from a portion of the available data. Predictions are then obtained for the excluded data and these are compared with the actual observations that were reserved for the cross-validation. In this setting, since country E had data available for 3 years (2007, 2008 and 2009), logistic and GEE models were fitted using data from 2007 up to mid-2009. The data for the last 6 months of 2009 are reserved for the cross-validation. Predictions for these last 6 months are obtained from the fitted models and subsequently compared to the actual observed data. The results are shown in Table 24.

Table 24. Cross-validation results for the logistic regression and GEE models on *Salmonella*–positive flocks, in country E, sample-level data, 2007-2009.

	Flocks of breeding hens				Flocks of laying hens			
	Obs.	**Pred. Prev.**			**Obs.**	**Pred. Prev.**		
Month	**Prev.**	**Logistic**	**GEE**	**Splines**	**Prev.**	**Logistic**	**GEE**	**Splines**
Jul/09	0.383	0.054	0.063	0.067	6.173	2.446	2.444	8.133
Aug/09	0.000	0.048	0.057	0.063	9.589	2.284	2.283	8.955
Sep/09	0.803	0.043	0.051	0.059	2.128	2.133	2.133	9.869
Oct/09	0.397	0.039	0.046	0.056	3.488	1.991	1.993	10.879
Nov/09	0.000	0.035	0.041	0.053	3.750	1.859	1.862	11.985
Dec/09	0.465	0.031	0.037	0.050	9.375	1.735	1.739	13.188

It can be observed that the predictions of the prevalence of *Salmonella*–positive breeding flocks are quite different from the observed prevalence under all 3 models considered. In addition, for the cases where the observed prevalence is zero, the predictions are somewhat less distant from the observed values.

For laying hens, as was already noted earlier, the logistic regression and GEE models give comparable results. For September, October and November 2009, the latter 2 methods yield better predictions than the GEE approach with splines. But for July, August and December, it is the GEE with splines approach that seems to do better in terms of predictions of prevalence of *Salmonella*–positive laying hen flocks.

Further considerations

Countries with few positive events observed

For two MS providing information on breeding flocks, the data were not considered in the modeling exercise due to an extremely small number of observed positive flocks. While the available information would allow a model fit which is seemingly fine, it's result would be very difficult to interpret. To illustrate this, both a logistic regression and GEE models have been fitted to the data of country P. The parameter estimates and odds ratios are provided in Table 25. Even though the parameter estimates obtained from both models are again very similar, there is quite some discrepancy between the estimated standard errors. Oddly, a smaller CI is obtained from fitting the GEE model. In this case, this could be a symptom of a problematic model fit, even though no warning was produced by the SAS software. Indeed, it is very difficult to account for correlation between flocks in a holding when only one flock tested positive (see Table 9). Even when assuming an independent working correlation, similar results are obtained. When displaying the results graphically (Figure 21), though observed and fitted probabilities of observing a positive flock are fairly comparable for most time points, the CIs for the fitted probabilities are quite huge, which could result in extremely variable predicted values. As such, it is clear that the obtained model fit is not very meaningful. This illustrates that statistical models should not to be used as automated, black-box procedures. Clearly, a model fit may be inappropriate even though the software does not produce a warning.

Table 25. Parameter estimates, standard errors, corresponding *P*-values and odds ratio estimates with 95% Confidence Intervals from logistic regression and GEE on prevalence of *Salmonella*–positive flocks with breeding hens, sample-level data, country P, 2007-2009.

	Logistic Regression			GEE		
Parameter	**Estimate**	**S.E.**	***P*-value**	**Estimate**	**S.E.†**	***P*-value**
Intercept	-8.834	3.190	0.0056	-8.718	2.422	0.0003
Time	0.071	0.111	0.5207	0.070	0.082	0.3963
Odds Ratio	**Estimate**	**95% CI**		**Estimate**	**95% CI†**	
Time	1.074	0.864	1.335	1.072	0.913	1.260

†Based on empirically corrected standard errors.

Weights

Another important aspect not considered in the current analysis is the use of *weights*. Indeed, since for different MS different information is available for the construction of weights, this makes it very difficult to explore a general approach for constructing weights. These would have to be constructed on a MS-by-MS basis. For example, for breeding flocks, information is available on the number of hens in the holdings in country E, whereas in country P (not considered for analysis) information on the holding size is provided. On the other hand, for country A, only the number of hens in the flock is provided. Since most information is available on the level of the birds rather than on the flocks, this further complicates the construction of appropriate weights.

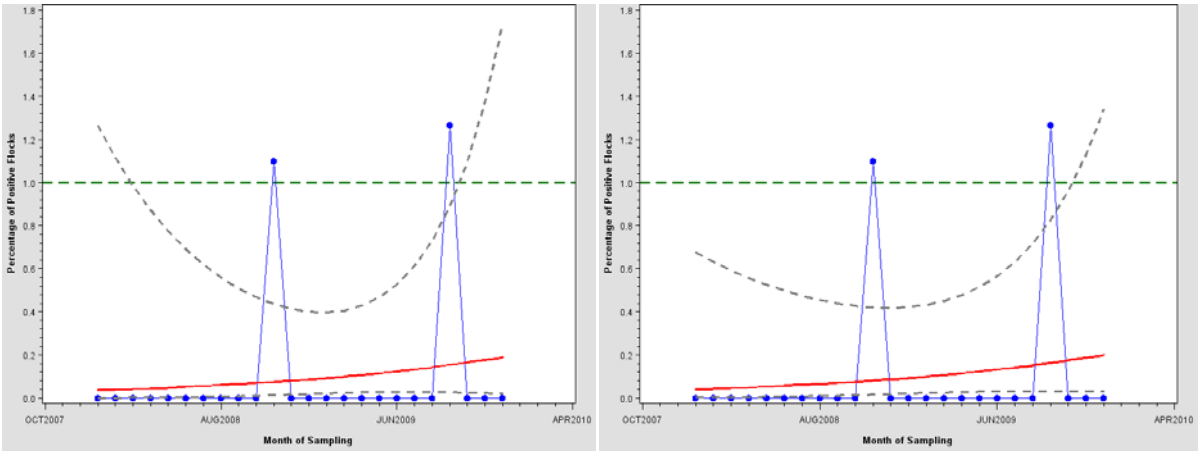


Figure 21. Observed (blue line) and estimated (red line) prevalence of *Salmonella*–positive flocks with breeding hens with 95% Confidence Intervals (grey dashed lines) from logistic regression (left figure) and GEE (right figure), sample-level data, country P, 2007-2009. The dashed green line represents EU reduction target.

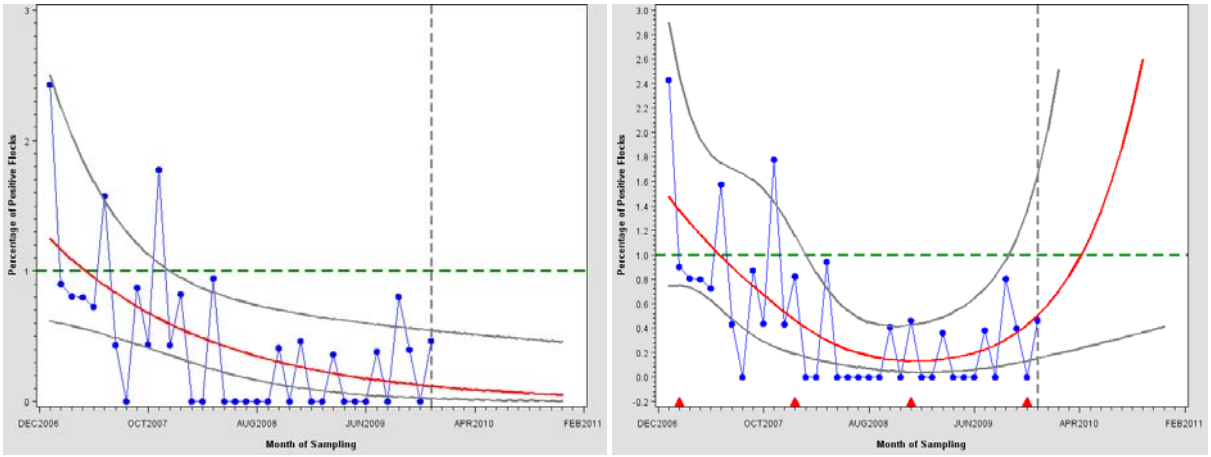
2.2.4. Evaluating the Likelihood of Achievement of *Salmonella* Reduction Targets

In a next step, the models obtained in the previous sections can be used to evaluate the likelihood of achieving the *Salmonella* reduction targets. An important constraint of the current analysis is that the proportion of positive tests is estimated on a monthly basis. While the general trend can provide some insight into the evolution over time of the prevalence of *Salmonella*–positive flocks, one has to be careful with extrapolating the results of this analysis to the targets set on a yearly basis. In this section, focus shall first be on the data on breeding flocks in country E to illustrate the methodology. Since the GEE model takes into account an important aspect of the data, only this model will be considered, with its extension using splines. Next an overview of the results from similar analyses for the other MS is given.

Flocks with breeding hens (countries E and A)

Monthly predictions of the probability of observing a positive breeding flock for 2010, based on the GEE model and the GEE model with splines based on 4 knots, are shown in Figure 22. Though the latter model was highlighted by the QIC as best fitting the observed data, its (long-term) predictions

do not appear very realistic. Moreover, the increased width displayed by the CI for the predictions in 2010 illustrates the increased uncertainty towards future observations under this model. In contrast, the prediction for the prevalence in Figure 22 (a) under the ordinary GEE model follows the observed downward trend, with quite stable confidence limits.



(a) GEE Model (b) GEE with Splines (4 knots)

Figure 22. Predictions of the prevalence of *Salmonella*–positive breeding flocks with 95% Confidence Intervals, for 2010 (one year of future observations), obtained from a GEE model and a GEE model with splines using 4 knots, sample-level data, country E, 2007-2009. The vertical line marks the end of the observed information. The dashed green line represents EU reduction target.

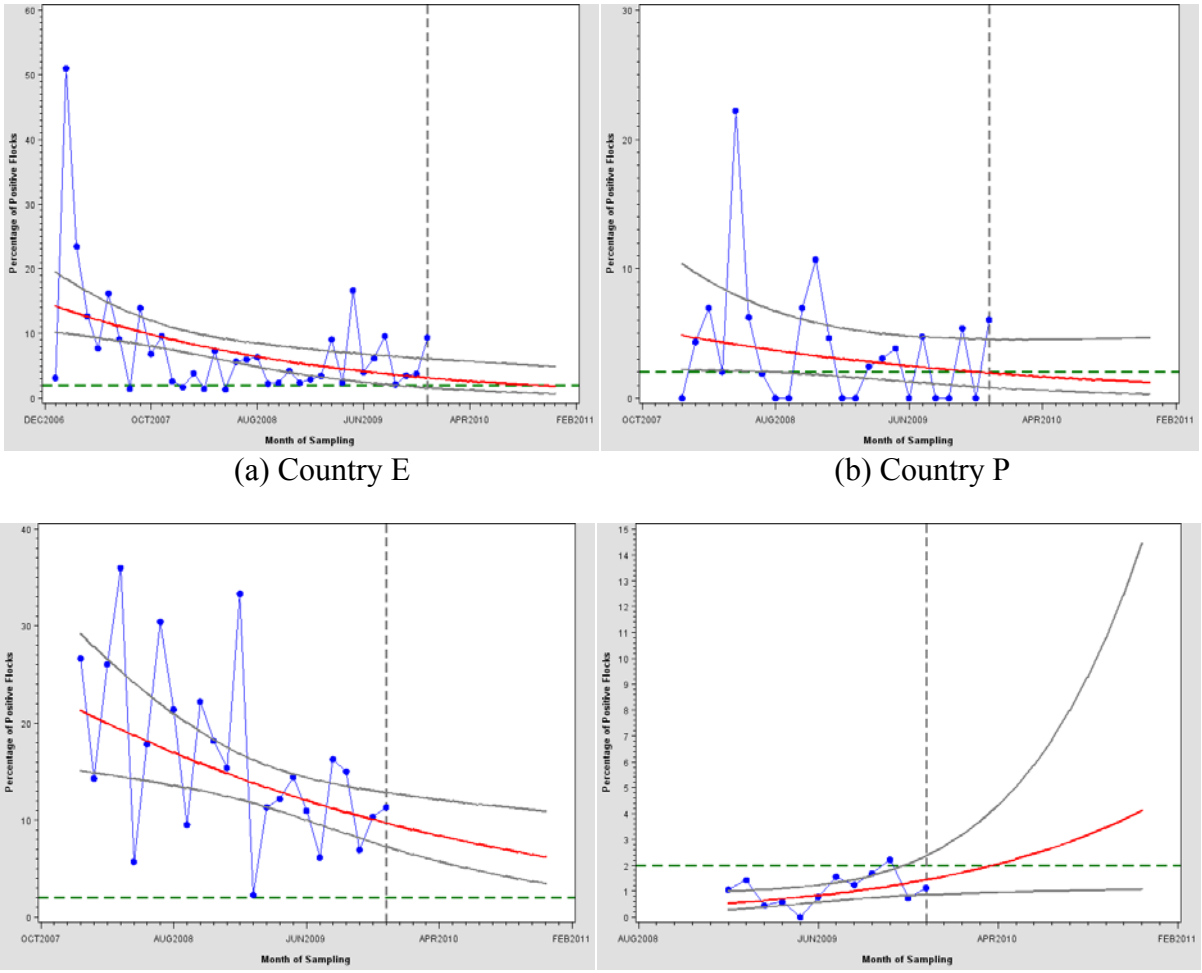
Flocks with laying hens (countries I, E, P, and V)

Predictions from the GEE models for the data on laying flocks are graphically illustrated in Figure 23. The predictions follow the slightly upward trend estimated for the data from country V. If the current trend continues, the monthly prevalence is expected to exceed the reduction target by May 2010. Nevertheless, the observed increase was not statistically significant. Should more data (historical or future) become available, it is well possible that a (downward) correction of the slope of the line will be observed. For the other MSs, a downward trend is observed, which suggests that, if the current trend continues, the monthly prevalence may reach the reduction target by around August 2010 for country E and by early 2010 for country P. While approaching the target line, country I may not be able to reach the target before the end of 2010.

Recall that for country E, the QIC identified an extension of the GEE model with splines based on 4 knots as being more appropriate for the data at hand. Prevalence predictions based on this model are displayed in Figure 24. It can be observed that though the prevalence predictions remain fairly stable after 2009, this is accompanied by a very wide CI, indicating a considerable level of uncertainty associated with the prediction. Moreover, the prediction is not expected to reach the target within 2010.

Clearly, there is quite some difference between predictions obtained from a simple GEE model with a linear trend over time. While such a trend may be unrealistic in practice, it seems to perform quite well for this study. On the other hand, more flexibility for the time trend in the model does not necessarily

guarantee a realistic prediction. This illustrates that an appropriate model for describing a trend and one for predicting a trend may be totally different.



(a) Country E (b) Country P

(c) Country I (d) Country V

Figure 23. Predictions of the prevalence of *Salmonella*–positive laying flocks with 95% Confidence Intervals, for 2010, obtained from a GEE model, sample-level data, countries E, P, I and V, 2007-2009. The vertical line marks the end of the observed information. The dashed green line represents EU reduction target.

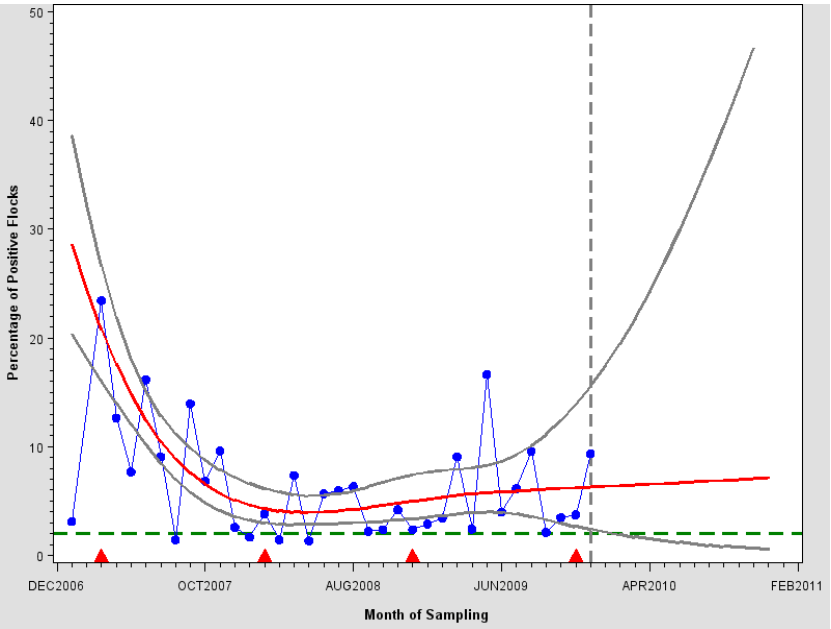


Figure 24. Predictions of the prevalence of *Salmonella*–positive laying flocks with 95% Confidence Intervals, for 2010, obtained from a GEE model with splines using 4 knots, sample-level data, country E, 2007-2009. The vertical line marks the end of the observed information. The dashed green line represents EU reduction target.

Discussion and Conclusions

The present investigation was undertaken to investigate the appropriateness of different statistical methodologies in evaluating the progress made by MSs towards, or the achievement of, the EU *Salmonella* reduction targets in breeding and laying hens of *Gallus gallus*. This was done based on two databases of prevalence data – one consisting of aggregated-level data submitted by MSs in their annual zoonoses reports and the other consisting of more detailed sample-level data provided voluntarily by six MSs.

Aggregated-level data

For the aggregated-level data, due to the very limited amount of information available (e.g., 4 to 6 years of data), the models that were considered were kept relatively simplistic. It is important to point out that though there is an abundance of methods in the literature that are appropriate for modelling complex data structures, these often require a reasonable amount of data. The choices here, however, were largely restricted by the limitations of the current data setting. Nevertheless, the results for a few MSs were reasonably informative considering the given limitations. However, the extremely short time sequence on which these models were based can hardly be sufficient to establish a reliable trend analysis. It is therefore wise to take caution in assessing and interpreting the results of the foregoing analyses on the aggregated-level data.

Regarding the specific statistical methodologies proposed for the evaluation of the progress made by the MS towards, or the achievement of, the EU *Salmonella* reduction targets, based on the aggregated-level prevalence data submitted by the different MSs to EFSA, several points can be made.

- Simple modelling strategies can be considered for the aggregated-level prevalence data. In this report a simple logistic regression for *Salmonella* prevalence with a linear time effect was considered. In a number of situations, the latter model was sufficient and, assuming the observed trends persist in these cases, could provide reasonable predictions for future prevalence values, which will in turn be useful in evaluating the progress made towards the *Salmonella* reduction targets.

- Extensions of the previous type of models can also be considered. When observed trends exhibit some amount of non-linearity in time, additional terms for the time effect (e.g., quadratic) could be considered. This was true for some countries and the extension of the linear logistic regression model to a quadratic one significantly improved the fit of the model to the observed data.

- When improvement of the model fit cannot be achieved by means of additional terms for the time effect, as would probably be the case for very irregular observed trends, one might also further consider fully nonlinear types of models. Though these allow much more flexible structures that can capture irregularities in observed trends, they can also be numerically difficult, especially when the available data are quite limited. For instance, one can almost always expect to run into convergence problems when considering a model with 3 parameters when only 4 data points are available. Moreover, one needs to find a balance between capturing the observed trends and avoiding overfitting. A model should adequately capture observed trends, but at the same time should not be entirely dependent on or driven by the data at hand.

- Finally, though the abovementioned analyses on the aggregated-level prevalence data largely focussed on modelling positivity trends in time, and as such, considered only time as the

primary predictor, additional covariates which could be known to have some impact on *Salmonella* prevalence can also be included within the proposed models. These, if indeed meaningfully related to positivity, could provide further improvement of the model fit.

Another main issue regarding the analysis of the aggregated-level data is the fact that harmonization of the monitoring scheme only started in breeding flocks in 2007 and in laying flocks in 2008. Due to possible underreporting in the period prior to harmonization, bias can be introduced in the estimate of the time trend. Possible approaches to deal with this issue include the following:

- modelling the harmonization as an additional parameter in the model. This would allow a different time trend before and after harmonization. While it takes into account all information, the information prior to harmonization will have very little impact on the future predictions. This approach would therefore be similar to one where all information prior to harmonization is ignored.

- using the information prior to harmonization as a prior in a Bayesian analysis. Also here one has to be careful with the possible underreporting associated with this information. If few data points posterior to the harmonization are available, the prior may have a considerable impact on the analysis. A possible solution in this case would be to focus on countries that are known to have control programs even before harmonization of the monitoring scheme.

- using the time effect estimated in the data prior to harmonization as a prior for the time effect post-harmonization. This could be an interesting approach assuming that the evolution of *Salmonella* positivity is not changed by the harmonization.

Sample-level data

In contrast with the aggregated-level data, the sample-level data provided a much richer basis for exploration. In addition to sufficiently lengthy time sequences on which *Salmonella* prevalence trends could be modelled, information on the holding and flock identification was also available. These enabled the application of statistical models for clustered data. With the increased flexibility accorded by the enriched information, the range of choices for modelling strategies was considerably broadened in comparison with what was done for the aggregated-level data. For most of the MSs, well-fitting models were obtained and predictions for future prevalence values were reasonably stable. One major drawback, however, for the sample-level data was the fact that for a few MSs, though prevalence values were observed monthly over a sufficiently long period of time, only 1 or 2 months registered a positive result. While it is indeed true that one can obtain a seemingly stable model in such a case, the meaningfulness of such a model can be somewhat questionable, and, more importantly, predictions can be extremely volatile. It is thus important to be cautious in the use and application of statistical models when this type of situation is present.

With respect to the various statistical methodologies proposed for the evaluation of the progress made by the MSs towards, or the achievement of, the EU *Salmonella* reduction targets, based on the non-aggregated, sample-level prevalence data submitted by a number of MSs to EFSA, the following remarks can be made.

- The clustered nature of the sample-level data was initially ignored as a starting point for analysis and a logistic regression of *Salmonella* prevalence on a linear time effect was considered. Given the clustered structure of the data, this approach might be considered somewhat 'naïve', since the clustering within the data can have some effect on the variability of the resulting parameter estimates. Though the estimates themselves are quite comparable to those obtained using clustered approaches, the precision of these estimates are underestimated

when ignoring clustering, and as such, could impact the conclusions about the significance of these parameters.

- To appropriately account for the clustering within the data, a cluster-based statistical model, namely, the generalized estimating equations (GEE) approach, was employed. Indeed, the results indicated increased variability in the parameter estimates as compared to the results for ordinary logistic regression, indicating the effects of correlations among flocks within the same holding. For most of the MSs analyzed, the GEE method proved to be preferable over more complex types of models.

- A semi-parametric approach combining GEE with splines was also considered to try to capture even more closely distinct patterns in the observed trends. While this technique was indeed able to follow observed trends quite closely, predictions for the latter time periods exhibited fairly large variability, which would not be advisable when using such predictions in evaluating the progress made by the MS towards, or the achievement of, the *Salmonella* reduction targets. Moreover, this method was only found to be preferable for one MS (e.g., country E), which might have been expected since the observed trends for this MS were somewhat more irregular than those observed for the other MSs. Finally, it is again important to emphasize that a balance needs to be attained between a sufficiently well-fitting model that does not tend to over-fit the observed values.

In this report, various statistical methodologies were considered and assessed with regards to their appropriateness in evaluating the progress made by MSs towards the achievement of the EU *Salmonella* reduction targets in breeding and laying hens. Several modelling strategies were applied on the aggregated- as well as the sample-level data and these were primarily dictated by the nature of the available information within the said databases. The differences in the structure of the aggregated- and the sample-level data naturally led to different choices of approaches. Whereas the aggregate-level data was quite limited, the sample-level data provided a richer basis on which models could be built. For the former, many of the strategies considered yielded less than satisfactory results, which is not surprising given the limitations in the amount of available information. As such, conclusions derived from these approaches should be taken with extreme caution. In contrast, for the sample-level data, a number of modelling approaches proved meaningful and stable enough to provide insight into the progress made by MSs towards the achievement of the *Salmonella* reduction targets in breeding and laying hens.

Recommendations

One of the main difficulties encountered in the foregoing modelling exercise was the lack of sufficient information, particularly for the aggregated-level data. Firstly, the time sequence under investigation consisted of at most 6 time points and this significantly limited the types of models that could be considered for the analyses. Change-point models, for instance, could not be further entertained due to the extremely small amount of data points for the periods before and after the monitoring scheme was changed. Also, more complex model structures (e.g., fully non-linear models), which allow greater flexibility in capturing more complicated trends, were not feasible because of the data limitations. In contrast, the sample-level data consisted of more time periods (e.g., monthly data), thereby allowing the use of much more flexible models by which trends in *Salmonella* prevalence could be more adequately captured. If MSs can provide prevalence data for monthly or even quarterly periods, rather than yearly values, then perhaps the resulting models for the aggregated-level data could prove much more informative.

An additional limitation in the aggregated-level data is the unavailability of information identifying the particular holding from which the flock data are taken. Because flocks of the same holding might be somewhat correlated, models accounting for such should be considered. This was possible for some of the MSs considered in the sample-level data analyses. There, it was clear that clustering effects due to holding could have an impact on the precision of the resulting estimates. Since no such information was available for the aggregated-level data, any clustering effects were altogether ignored. It is therefore recommended that MSs try to include, within their monitoring scheme, some form of identification by which flocks of the same holding can be distinguished.

In the previous analyses, for both the aggregated- and sample-level data, only time was considered as the primary covariate of interest. Though, in general, this was indeed the main concern, i.e., to estimate trends of *Salmonella* prevalence over time, results can be further refined by the inclusion of other important covariates that may be believed to have an impact on the said trends. Naturally, this would imply that information on such covariates should also be recorded for possible inclusion in the proposed models.

The time-to-event methodology could not also be further investigated here owing to the lack of information on the start and end dates of sampling. This could be a worthwhile approach to look into, particularly if the time to positivity is of interest. If there is a clear record of the starting and ending period of monitoring, then this approach could be investigated further.

Finally, weighted analyses could not be explored because of the incompleteness in the information that could possibly be used for weight construction. Had such information been available for all MSs, then weighted modelling approaches could be considered.

References

Aerts M, Geys H, Molenberghs G and Ryan L, 2002. Topics in Modelling of Clustered Data. London: Chapman & Hall.

Agresti A, 2007. An introduction to Categorical Data Analysis (2nd ed). New York: Wiley.

Barnett V, 2002. Sample Survey: Principles and Methods (3rd ed). London: Arnold.

Bollaerts K, Aerts M, Hens N, Shkedy Z, Faes C, Van der Stede Y and Beutels P, 2010. Estimating epidemiological parameters by using mixture-models for serological data. Submitted.

Chambers RL and Skinner CJ, 2003. Analysis of Survey Data. New York: Wiley

Copas JB, 1989. Unweighted sum of squares test for proportions. Applied Statistics 38, 71-80.

Dempster AP, Selwyn MR and Weeks BJ, 1983. Combining historical and randomized controls for assessing trends in proportions. Journal of the American Statistical Association 78, 221-227.

Diggle PJ, Liang KY and Zeger SL, 1994. Analysis of Longitudinal Data. Oxford: Oxford University Press.

Diggle PJ, Heagerty PJ, Liang KY and Zeger SL, 2002. Analysis of Longitudinal Data (2nd ed). Oxford Science Publications. Oxford: Clarendon Press.

EFSA (European Food Safety Authority), 2011. The European Union Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-borne Outbreaks in 2009. EFSA Journal 2011; 9(3):2090.

Eilers PHC and Marx BD, 1996. Flexible smoothing with B-splines and penalties. Statistical Science 11, 89-121.

Evans S and Hosmer DW, 2004. Goodness of fit tests for the logistic GEE models: simulation results. Communications in Statistics: Simulation and Computation 33, 247-258.

Evans S and Li L, 2005. A comparison of goodness of fit tests for the logistic GEE model. Statistics in Medicine 24, 1245-1261.

Faes C, Aerts M, Geys H, Bijnens L, Ver Donck L and Lammers W, 2006. GLMM approach to study the spatial and temporal evolution of spikes in the small intestine. Statistical Modelling 6, 300-320.

Faes C, Aerts M, Litiere S, Meroc E, Van der Stede Y and Mintiens K, 2010. Estimating the herd-prevalence based on aggregate testing of animals. Journal of the Royal Statistical Society A 00, 000-000.

Farrington CP, 1996. On assessing goodness of fit of generalized linear models to sparse data. Journal of the Royal Statistical Society B 58, 349-360.

Friedman JH and Silverman BW, 1989. Flexible parsimonious smoothing and additive modelling. Technometrics 31, 3-39.

Gelman A, Carlin JB, Stern HS and Rubin DB, 2004. Bayesian Data Analysis (2nd ed). London: Chapman & Hall.

Grubbs FE, 1962. Attempts to validate certain PERT statistics or picking on PERT. Operations Research 10, 912-915.

Hosmer DW and Lemeshow S, 2000. Applied Logistic Regression (2nd ed). Wiley: New York, 2000.

Knottnerus P, 2003. Sample Survey Theory. New York: Springer.

Kuss O, 2002. Global goodness-of-fit tests in logistic regression with sparse data. Statistics in Medicine 21, 3789-3801.

Kutner MH, Nachtsheim CJ, Neter J and Li W, 2004. Applied Linear Statistical Models (5th ed). McGraw-Hill/Irwin.

Liang KY and Zeger SL, 1986. Longitudinal data analysis using generalized linear models. Biometrika 73, 13-22.

McCullagh P, 1985. On the asymptotic distribution of Pearson´s statistic in linear exponential-family models. International Statistical Review 53, 61-67.

Maringwa JT, Geys H, Shkedy Z, Faes C, Molenberghs G, Aerts M, Van Ammel K, Teisman A and Bijnens L, 2006. Analysis of a cardiovascular safety experiment with longitudinal data using penalized splines. In: Proceedings of the 21st Internation Workshop on Statistical Modelling. Hinde, J., Einbeck, J. and Newell, J. (Eds) Galway Ireland, 346-353.

Molenberghs G and Verbeke G, 2005. Models for Discrete Longitudinal Data. New York: Springer-Verlag.

Nelson W, 1982. Applied Life Data Analysis. New York: Wiley.

Neuhaus JM, 1992. Statistical methods for longitudinal and clustered designs with binary responses. Statistical Methods in Medical Research 1, 249-273.

Orme C, 1988. The calculation of the information matrix test for binary data models. The Manchester School 54, 370-376.

Osius G and Rojek D, 1992. Normal goodness-of-fit tests for multinomial models with large degrees of freedom. Journal of the American Statistical Association 87, 1145-1152.

Pan W, 2001. Akaike's information criterion in generalized estimating equations. Biometrics 57, 120-125.

Parise H, Wand MP, Ruppert D and Ryan L, 2001. Incorporation of historical controls using semi-parametric mixed models. Applied Statistics 50, 31-42.

Rogan WMJ and Gladen B, 1978. Estimating prevalence from the results of a screening test. American Journal of Epidemiology 107, 71-76.

Robert CP, 2007. The Bayesian Choice. Springer.

Ruppert D, 2002. Selecting the number of knots for penalized splines. Journal of Computational and Graphical Statistics 11, 735-757.

Ruppert D, Wand MP and Carroll RJ, 2003. Semiparametric Regression. Cambridge University Press.

Skellam JG, 1948. A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials. Journal of the Royal Statististical Society B 10, 257-261.

Skinner CJ, Holt D and Smith TMF, 1989. Analysis of Complex Surveys. New York: Wiley.

Stiratelli R, Laird N and Ware JH, 1984. Random effects models for serial observations with binary responses. Biometrics 40, 961-971.

Welsh AH, Lin X and Carroll RJ, 2002. Marginal longitudinal nonparametric regression: locality and efficiency of spline and kernel methods. Journal of the American Statistical Association 97, 482-493.

White H, 1982. Maximum likelihood estimation of misspecified models. Econometrica 50, 1-25.

Zeger SL and Liang KY, 1986. Longitudinal data analysis for discrete and continuous outcomes. Biometrics, 42, 121-130.

Appendices

APPENDIX A

Table A1. Parameter estimates from logistic regression models with Quadratic Time Effect of prevalence of *Salmonella*–positive flocks with breeding hens, aggregated data, EUSR 2004-2009.

Country	Model Parameters				Goodness-of-Fit Test† *P*-Values	
	Parameter	Estimate	S.E.	*P*-value	Deviance	Pearson
1	Intercept	4.040	3.039	0.1837	0.0035	0.0031
	time	-4.683	1.513	0.0020		
	time^2	0.565	0.175	0.0013		
4	Intercept	-2.301	0.480	0.0000	0.0000	0.0000
	time	-1.284	0.333	0.0001		
	time^2	0.168	0.049	0.0006		
8	Intercept	-1.359	0.219	0.0000	0.0000	0.0000
	time	-0.020	0.156	0.8989		
	time^2	-0.069	0.024	0.0033		
9	Intercept	-3.530	0.555	0.0000	0.0012	0.0158
	time	0.590	0.454	0.1929		
	time^2	-0.077	0.064	0.2276		
10	Intercept	-4.503	1.080	0.0000	0.0000	0.0001
	time	0.782	0.573	0.1718		
	time^2	-0.134	0.073	0.0676		
18	Intercept	-3.086	0.275	0.0000	0.0331	0.0319
	time	0.330	0.189	0.0814		
	time^2	-0.064	0.027	0.0165		
25	Intercept	0.670	0.494	0.1745	0.0017	0.0034
	time	-0.616	0.380	0.1044		
	time^2	-0.025	0.058	0.6611		
29	Intercept	-2.060	1.411	0.1443	0.0008	0.0003
	time	-1.840	1.190	0.1221		
	time^2	0.243	0.168	0.1469		

†The null hypothesis for these tests is good model fit; rejection of the null, $p<0.05$, implies poor model fit.

Table A2. Parameter estimates from logistic regression models with Quadratic Time Effect of prevalence of *Salmonella*–positive flocks with laying hens, aggregated data, EUSR 2004-2009.

Country	Model Parameters				Goodness-of-Fit Test† *P*-Values	
	Parameter	Estimate	S.E.	*P*-value	Deviance	Pearson
1	Intercept	-1.446	0.443	0.0011	0.0000	0.0000
	time	-1.160	0.280	0.0000		
	$time^2$	0.156	0.041	0.0001		
5	Intercept	-5.404	1.215	0.0000	0.0019	0.0032
	time	1.088	0.637	0.0879		
	$time^2$	-0.115	0.079	0.1430		
6	Intercept	-3.968	0.216	0.0000	0.0529	0.0650
	time	-0.491	0.133	0.0002		
	$time^2$	0.110	0.018	0.0000		
8	Intercept	2.806	0.360	0.0000	0.0000	0.0000
	time	-1.946	0.198	0.0000		
	$time^2$	0.183	0.024	0.0000		
10	Intercept	-3.993	0.722	0.0000	0.0000	0.0000
	time	1.378	0.334	0.0000		
	$time^2$	-0.186	0.039	0.0000		
13	Intercept	-9.909	1.326	0.0000	0.0000	0.0000
	time	5.271	1.149	0.0000		
	$time^2$	-0.696	0.166	0.0000		
14	Intercept	-5.401	0.356	0.0000	0.0007	0.0016
	time	0.764	0.206	0.0002		
	$time^2$	-0.083	0.027	0.0019		
17	Intercept	-4.050	0.193	0.0000	0.8361	0.8357
	time	0.558	0.121	0.0000		
	$time^2$	-0.088	0.017	0.0000		
18	Intercept	-2.227	0.199	0.0000	0.0000	0.0000
	time	-1.027	0.127	0.0000		
	$time^2$	0.174	0.017	0.0000		
22	Intercept	-5.527	0.474	0.0000	0.0660	0.0697
	time	1.511	0.254	0.0000		
	$time^2$	-0.218	0.033	0.0000		
29	Intercept	-2.351	0.376	0.0000	0.0000	0.0000
	time	-0.732	0.234	0.0018		
	$time^2$	0.114	0.031	0.0003		

†The null hypothesis for these tests is good model fit; rejection of the null, $p<0.05$, implies poor model fit.

APPENDIX B

Table B1. Parameter estimates, standard errors and corresponding *P*-values from exact logistic regression models of prevalence of *Salmonella*–positive flocks, sample-level data, countries A, E, P, I and V, 2007-2009

		Breeding Hens			Laying Hens		
Country	**Parameter**	**Estimate**	**S.E.**	***P*-value**	**Estimate**	**S.E.**	***P*-value**
A	Intercept	*					
	Time	-0.268	0.247	0.3301			
E	Intercept	-4.331	0.260	0.0000	-1.724	0.141	0.0000
	Time	-0.072	0.018	0.0000	-0.048	0.008	0.0000
P	Intercept				-2.365	0.627	0.0003
	Time				-0.045	0.027	0.0923
I	Intercept				-0.449	0.348	0.2601
	Time				-0.050	0.013	0.0001
V	Intercept				-6.426	1.096	0.0000
	Time				0.061	0.035	0.0763

*Degenerate estimate.

Table B2. Odds Ratio estimates with 95% Confidence Intervals for the Time Effect from exact logistic regression models of prevalence of *Salmonella*–positive flocks, sample-level data, countries A, E, P, I and V, 2007-2009.

	Breeding Hens			Laying Hens		
Country	**Estimate**	**95% CI**		**Estimate**	**95% CI**	
A	0.765	0.436	1.287			
E	0.931	0.898	0.963	0.953	0.938	0.968
P				0.956	0.907	1.007
I				0.952	0.928	0.976
V				1.063	0.994	1.140

Glossary / Abbreviations

EC	European Commission
EFSA	European Food Safety Authority
EU	European Union
EUSR	European Union Summary Report
Est	Estimate
GEE	Generalized Estimating Equations
MS	Member State
OR	Odds Ratio
QIC	Quasi-likelihood Information Criterion
S.E.	Standard Error