Epidemiol. Infect. (2011), **139**, 505–515. © Cambridge University Press 2010
doi:10.1017/S0950268810001664

Heterogeneity in vaccination coverage explains the size and occurrence of measles epidemics in German surveillance data

S. A. HERZOG[1]*†, M. PAUL[2]† AND L. HELD[2]

[1] *Institute of Social and Preventive Medicine, University of Bern, Switzerland*
[2] *Biostatistics Unit, Institute of Social and Preventive Medicine, University of Zurich, Switzerland*

(*Accepted 15 June 2010; first published online 12 July 2010*)

SUMMARY

The objective of this study was to characterize empirically the association between vaccination coverage and the size and occurrence of measles epidemics in Germany. In order to achieve this we analysed data routinely collected by the Robert Koch Institute, which comprise the weekly number of reported measles cases at all ages as well as estimates of vaccination coverage at the average age of entry into the school system. Coverage levels within each federal state of Germany are incorporated into a multivariate time-series model for infectious disease counts, which captures occasional outbreaks by means of an autoregressive component. The observed incidence pattern of measles for all ages is best described by using the log proportion of unvaccinated school starters in the autoregressive component of the model.

Key words: Infectious disease epidemiology, measles (rubeola), MMR vaccination, modelling.

INTRODUCTION

Measles is a highly contagious disease and still an important health concern [1]. Numerous efforts such as routine childhood vaccination programmes or the WHO measles elimination plan have significantly reduced the incidence of measles in Europe. The epidemic pattern has changed from a roughly biennial cycle to an irregular sequence of outbreaks [2]. However, disease has not been eradicated. The incidence of measles varies widely, with large outbreaks in Romania, Germany, UK, Switzerland and Italy in 2006 and 2007, whereas in other countries such as Finland, Slovakia and Hungary almost no cases were reported [1]. Since most measles cases were unvaccinated or incompletely vaccinated, the differences in incidence are likely to be due to differences in the success of national vaccination programmes [1, 2]. For instance, there have been several outbreaks in some of the 16 federal states of Germany in recent years [3–6]. Detailed investigations of selected outbreaks showed that most cases occurred in unvaccinated individuals [4].

National surveillance systems such as that at the Robert Koch Institute (RKI), Germany, typically provide weekly time-series of counts stratified by for example, region, age or sex. Accordingly, statistical methods for the analysis of multivariate time-series of counts are needed. It is of public health interest to investigate empirically the relationship between vaccination coverage and the occurrence and size of measles epidemics using such data.

* Author for correspondence: Ms. S. Herzog, Institute of Social and Preventive Medicine, University of Bern, Finkenhubelweg 11, 3012 Bern, Switzerland.
(Email: sherzog@ispm.unibe.ch)
† These authors contributed equally to this work.

source: https://doi.org/10.7892/boris.1160 | downloaded: 12.10.2024

Cummings *et al.* [7] used a linear model to analyse the sum of measles cases over 5 years in several provinces of Cameroon, including vaccination coverage among other covariates. However, the time-series aspect was not considered. Multivariate time-series methods for counts of infectious diseases have only recently been developed and applied to epidemiological data. However, these models are not able to cope with occasional large outbreaks. For instance, Frank *et al.* [8] investigated the association between human infection with Shiga toxin-producing *Escherichia coli* (STEC) and cattle density based on German notification data. A Bayesian Poisson regression model was used to analyse the weekly number of cases in each age group and district of Germany. The model accounted for temporal and seasonal trends, spatial variation and cattle density as explanatory factors. No large STEC gastroenteritis outbreaks occurred in the time period considered. Hens *et al.* [9] modelled the yearly, age-stratified incidence of hepatitis B in Bulgaria using a log-additive Poisson model, where age and time were modelled as non-parametric functions. The impact of vaccination was taken into account by including indicators for various immunization programmes as covariates. The log-additive Poisson model chosen was justified since the data contained no outbreaks.

If there are outbreaks in the data, a more realistic formulation for (multivariate) time-series of infectious disease counts has been suggested by Held *et al.* [10]. The model decomposes the disease incidence into two additive components. One component represents an autoregression on past counts which allows for temporal dependence beyond regular patterns, i.e. epidemic behaviour. The other component accounts for regular, endemic behaviour. However, this method did not consider the inclusion of covariates.

The aim of this paper is to investigate the association between vaccination coverage and the size and occurrence of measles epidemics. We first describe the data about measles incidence [11] and vaccination coverage [12] in Germany obtained from the RKI. The approach of Held *et al.* [10] is extended to allow for the inclusion of covariates and applied to the measles data using vaccination coverage as an explanatory variable. Different formulations of the proposed model are compared based on Akaike's Information Criterion (AIC [13]). A simulation study is performed in order to further investigate the ability of AIC to identify the underlying true model.

DATA

Measles incidence

In Germany, introduction of the measles vaccine had reduced the incidence of measles to a historical low of 0·2 cases/100 000 inhabitants in 2004 [3], before the disease re-emerged due to outbreaks in a few regions. We used measles surveillance data from Germany for the years 2005–2007, which contain weekly counts of cases for all ages in all 16 federal states reported to the RKI [11]. Figure 1 shows the notified measles cases in the years 2005–2007 for six selected federal states to illustrate the different incidence patterns. Large outbreaks occurred in Hesse and Bavaria in 2005 [3], in North Rhine-Westphalia in 2006 [4] and in North Rhine-Westphalia and Bavaria in 2007 [5]. The majority of cases ($\sim 80\%$) occurred in children and adolescents. About 12% occurred in infants aged <2 years. This pattern was very similar in all three years considered. A brief summary of the number of reported cases in each state is shown in Table 1 together with population numbers at 31 December 2006 obtained from the Federal Statistical Office of Germany [14].

Measles-mumps-rubella (MMR) vaccination

Coverage levels of the combined MMR vaccine were derived from vaccination cards presented at medical examinations, which are conducted by local health authorities at school entry [12]. Records include information about receipt of the first and second doses of MMR, but no information about dates or age of the child at vaccination. Age at school entry ranges between states from 4 to 7 years [15], therefore the information collected typically refers to vaccinations received 3–5 years previously [16].

The estimated coverage data do not include any information from children who did not present a vaccination card on the day of the medical examination (5–13% of children attending the school entry examination in different states). This is likely to overestimate true coverage, because the vaccination status of children with vaccination cards is generally more complete than in those without a card [4, 17]. However, there are no national data about the degree of overestimation. We made an assumption, which was used in a previous German study [18], that for each dose, the percentage of children without a vaccination card, 'non-card holders' was half that of 'card holders'. We applied this adjustment to all

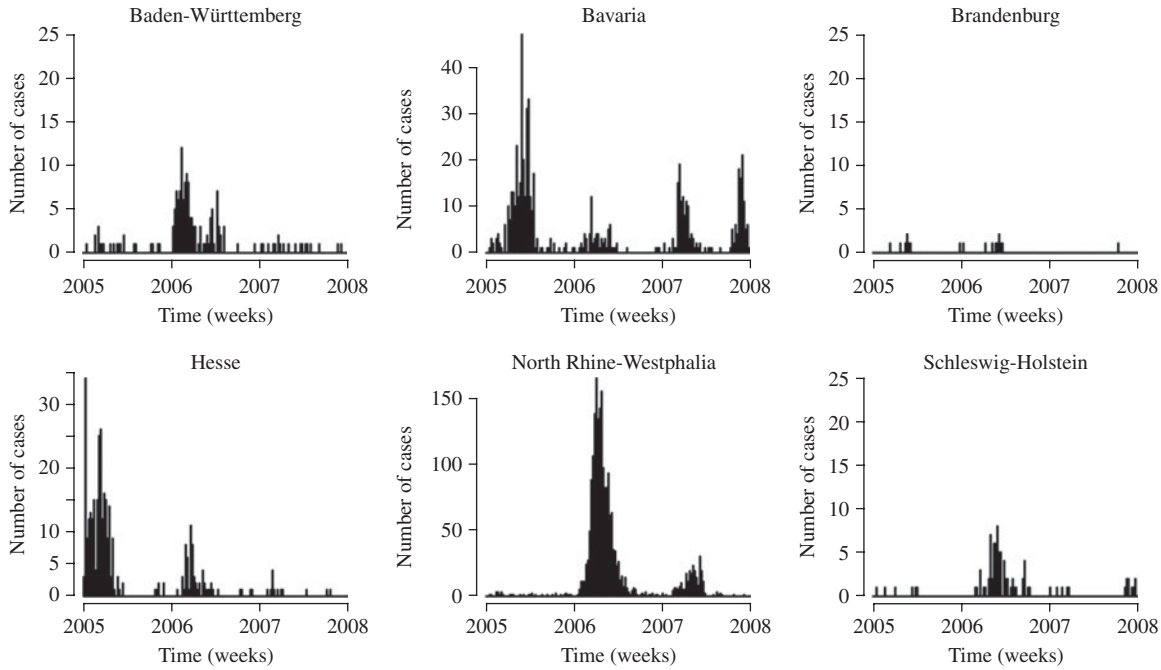

Fig. 1. Number of weekly measles cases in selected German federal states for the years 2005–2007. Note that the y-axis is not the same for all states.

analyses and conducted a sensitivity analysis to examine the robustness of the assumption.

Coverage levels for both the first and the second dose were higher in the new, re-established states in East Germany (Brandenburg, Mecklenburg-Western Pomerania, Saxony, Saxony-Anhalt, Thuringia) than in West Germany (Table 1). This might reflect continuing adherence to different childhood vaccination policies before re-unification [15, 19]. Immunization is voluntary in Germany now, but it was mandatory in the former German Democratic Republic.

METHODS

To investigate a possible association between the occurrence of measles epidemics and MMR vaccination coverage, we first examined the correlation between the number of observed cases in a region and region-specific vaccination coverage. One possibility is to apply the variance-stabilizing transformation for Poisson counts [20], i.e. taking the square root of cases, before estimating the empirical correlation coefficient which might improve the goodness of the corresponding confidence intervals. An alternative approach, based on a Poisson regression model [21, 22], assumes that the sum of cases in region i, aggregated over all three years, has mean

$$\mu_i = \exp(\alpha + \beta x_i), \tag{1}$$

where x_i denotes the coverage in state i. For example, to adjust for regionally varying population numbers, the right hand side of equation (1) can be multiplied by an offset n_i. Conclusions about the effect β of the covariate x_i in equation (1) remain the same when considering the weekly number of cases instead of the sum of cases, assuming that the weekly counts are independent. However, a multivariate time-series analysis of counts is able to incorporate autocorrelation and provides many more possibilities compared to the analysis of temporally aggregated data.

In the following, $y_{i,t}$ denotes the number of cases of a specific disease in a defined geographical region $i = 1, \ldots, I$ at time $t = 1, \ldots, T$. A fundamental assumption of a Poisson regression model is that the response variables $y_{i,t}$ are independent given the covariates. Thus the above model is not suited for the analysis of the measles data as the weekly counts are clearly dependent. Regular temporal dependence can easily be accounted for by including covariates for long-term or seasonal trends in the model. For instance, seasonal variation can be modelled parametrically using a superposition of harmonic waves [10, 23] or non-parametrically [9, 24]. However, such

Table 1. *Measles cases and estimated vaccination coverage in the 16 federal states of Germany*

State	Population	Measles cases		Coverage (%)		Presented cards (%)
		Max.	Sum	1st dose	2nd dose	
Baden-Württemberg (BW)	10 738 753	12	162	93·7	78·7	92·1
Bavaria (BY)	12 492 658	47	606	91·7	75·7	93·4
Bremen (HB)	663 979	1	4	94·6	76·9	86·9
Hamburg (HH)	1 754 182	3	29	93·9	84·0	91·7
Hesse (HE)	6 075 359	34	336	94·8	81·2	92·4
Lower Saxony (NI)	7 982 685	12	144	95·4	81·6	91·2
North Rhine-Westphalia (NW)	18 028 745	165	2036	95·2	81·6	88·5
Rhineland-Palatinate (RP)	4 052 860	9	85	94·9	80·8	91·4
Saarland (SL)	1 043 167	0	0	95·2	85·6	91·1
Schleswig-Holstein (SH)	2 834 254	8	89	94·7	83·6	89·8
Berlin (BE)	3 404 037	8	104	93·8	83·6	91·9
Brandenburg (BB)	2 547 772	2	18	97·1	89·8	93·5
Mecklenburg-Western Pomerania (MV)	1 693 754	1	4	97·5	91·6	92·1
Saxony (SN)	4 249 774	2	18	97·3	85·0	93·9
Saxony-Anhalt (ST)	2 441 787	2	12	97·7	89·8	92·6
Thuringia (TH)	2 311 140	3	8	97·4	88·3	94·6

Population estimated at 31 December 2006; maximum and total number of weekly measles cases from week 1, 2005 to week 52, 2007; coverage at school entry for the first and second dose of MMR vaccine in 2006 estimated from children presenting vaccination cards at school entry examinations; percentage of children with a vaccination card.

a model may still not adequately capture occasional outbreaks typical for infectious diseases.

A natural way to incorporate temporal dependence beyond seasonal variation is to consider the number of past cases as additional explanatory variables in the model. Held *et al.* [10] suggest a Poisson regression model with an identity link, where the (conditional) mean $\mu_{i,t}$ of $y_{i,t}$ is additively decomposed into two parts

$$\mu_{i,t} = \lambda y_{i,t-1} + \nu_{i,t}. \tag{2}$$

The first part with conditional rate $\lambda y_{i,t-1}$ is called the 'epidemic' component and the second part with rate $\nu_{i,t}$ the 'endemic' component. The former component captures occasional (epidemic) outbreaks whereas the latter describes regular (endemic) patterns.

To include region-specific covariate information, we allow the autoregressive parameter λ in equation (2) to vary across regions, i.e. we switch notation from λ to λ_i and model λ_i as a function of these covariates. Furthermore, covariates can also be considered in the other component $\nu_{i,t}$. Note that the conditional mean $\mu_{i,t}$ needs to be non-negative. This can be ensured by modelling both λ_i and $\nu_{i,t}$ on a log-scale.

Our first model (type A) assumes that the coverage levels in all states, x_i, enter into the epidemic component and the model is given by

$$\log(\lambda_i) = \beta_0 + \beta_1 x_i, \tag{3}$$

$$\log(\nu_{i,t}) = \alpha_0 + \{\gamma \sin(2\pi t/f) + \delta \cos(2\pi t/f)\} + \log(n_i), \tag{4}$$

where β_0 is an intercept and β_1 quantifies the influence of vaccination coverage. The parameter α_0 denotes the intercept of the endemic component and the offset $\log(n_i)$ represents population fractions, computed from Table 1. The terms in curly brackets in equation (4) are used to model seasonal variation. The number of data points per season is denoted by f. For instance, for a season of 1 year and weekly data $f = 52$. For ease of interpretation, the seasonal terms can be written equivalently as a sine wave with amplitude A describing the magnitude, and phase difference φ describing the onset of the seasonal pattern [23]. In the second model, the term $\beta_1 x_i$ is omitted in equation (3) and the coverage levels x_i are included instead in the endemic component with coefficient α_1. Altogether, the model (type B) is given by

$$\log(\lambda_i) = \beta_0, \tag{5}$$

$$\log(\nu_{i,t}) = \alpha_0 + \alpha_1 x_i + \{\gamma \sin(2\pi t/f) + \delta \cos(2\pi t/f)\} + \log(n_i). \tag{6}$$

Table 2. *Estimated Pearson's correlation coefficient, r, with 95% confidence intervals*

	Adjusted vaccination coverage			
	1st dose		2nd dose	
	r	95% CI	*r*	95% CI
Sum of cases	−0·34	−0·71 to 0·19	−0·34	−0·72 to 0·19
Square root of sum of cases	−0·44	−0·77 to 0·07	−0·48	−0·79 to 0·02

To investigate the impact of the explanatory variable, we also consider a model of type C, given by equations (4) and (5), where no covariate is included. Additionally, a standard log-linear Poisson regression model without the autoregressive component is fitted (model D).

For the model of type A we use the log proportion of unvaccinated school starters as explanatory variable x_i in equation (3) in accordance with the mass action principle [25]. This principle assumes that the rate of disease spread is proportional to the product of the density of susceptibles (unvaccinated school starters) multiplied by the density of infected individuals (reported cases). Taking the logarithm of the proportion of unvaccinated school starters produces the multiplicative relation (model A_0). Similarly, the log proportion of all school starters who received at most one dose of MMR vaccine is used as an explanatory variable. We used the same covariates in the model of type B.

Maximum likelihood (ML) estimates of parameters and standard errors (S.E.) are obtained by numerically maximizing the respective Poisson log-likelihood. Standard software for linear Poisson regression cannot be used because of the nonlinearity of the parameters. Therefore, the quasi-Newton BFGS method implemented in the R [26] function optim is used for optimization. The fitting procedure and the measles data are integrated in the R package surveillance ([27]; http://surveillance.r-forge.r-project.org). Note that models involving more than one covariate, time-varying covariates or additional seasonal terms at higher frequencies [28] can also be fitted with this function in surveillance.

The models investigated in the Results section are compared based on the model choice AIC criterion. We were particularly interested in the ability of AIC to distinguish between the model types A and B. In order to investigate this we conducted a simulation study (see Appendix).

RESULTS

The sum of cases over the years 2005–2007 in each state is negatively correlated with coverage for both the first and second dose of MMR vaccine (Table 2). Absolute correlation increases slightly when taking the square root of cases. However, the statistical evidence for correlation is weak, since the upper 95% confidence limits are always positive.

We describe here an analysis of the multivariate time-series of counts to further investigate the measles incidence patterns. The generation time [25] for measles, i.e. the average time between the onset of symptoms in one case and the onset of symptoms in a second case directly infected by the first, is about 10 days [25, 29]. We therefore aggregate measles cases in successive bi-weekly periods to better reflect this characteristic time-scale [30, 31]. AIC is used as a model choice criterion. The simulation study, discussed in detail in the Appendix, showed that this criterion is suitable for the comparison of the different model formulations.

The results of the analysis of the bi-weekly aggregated measles data are summarized in Table 3. All considered models contain an overall intercept α_0, a seasonal term and population fractions n_i as offset. The last two models in the table contain no covariates. When including only an intercept in the epidemic component (model C), the fit improves substantially compared to a model without autoregression (model D). The ML estimate of $\lambda = \exp(\beta_0)$ is quite high, $\hat{\lambda} = 0{\cdot}85$ (S.E. $= 0{\cdot}02$), which indicates a strong dependence on the number of counts at the previous time point after adjustment for seasonal effects. Consequently, the use of a Poisson regression model (without autoregression) seems inappropriate for these data. Indeed, the series of deviance residuals obtained from model D showed considerable autocorrelation compared with model C, which showed almost no autocorrelation.

Table 3. *Analysis of bi-weekly aggregated measles data*

				Epidemic component		Endemic component			
Model	log(L)	p	AIC	β_0 (S.E.)	β_1 (S.E.)	α_0 (S.E.)	α_1 (S.E.)	A (S.E.)	φ (S.E.)
Log proportion of unvaccinated school starters									
A_0	−1778·1	5	3566·1	3·01 (0·52)	1·38 (0·23)	1·78 (0·06)	—	0·66 (0·08)	−0·10 (0·12)
B_0	−1783·4	5	3576·8	−0·17 (0·02)	—	5·43 (0·69)	1·52 (0·29)	0·73 (0·09)	−0·10 (0·39)
Log proportion of school starters who received at most 1 dose of MMR vaccine									
A_1	−1787·1	5	3584·1	1·34 (0·31)	1·02 (0·21)	1·76 (0·06)	—	0·65 (0·08)	−0·08 (0·13)
B_1	−1790·7	5	3591·4	−0·17 (0·02)	—	3·59 (0·45)	1·17 (0·29)	0·71 (0·09)	−0·09 (0·41)
No covariates									
C	−1799·4	4	3606·8	−0·16 (0·02)	—	1·76 (0·06)	—	0·66 (0·08)	−0·06 (0·12)
D	−5213·9	3	10433·8	—	—	3·25 (0·03)	—	1·65 (0·04)	−0·52 (0·02)

The log-likelihood is denoted by log(L); p is the number of parameters and Akaike's Information Criterion (AIC) $= -2\log(L) + 2p$; lower AIC values indicate better fit. The parameters β_0 and α_0 denote intercepts; β_1 and α_1 denote the effect of the covariate; A and φ denote the amplitude and onset of the seasonal pattern. The standard error is denoted by S.E.

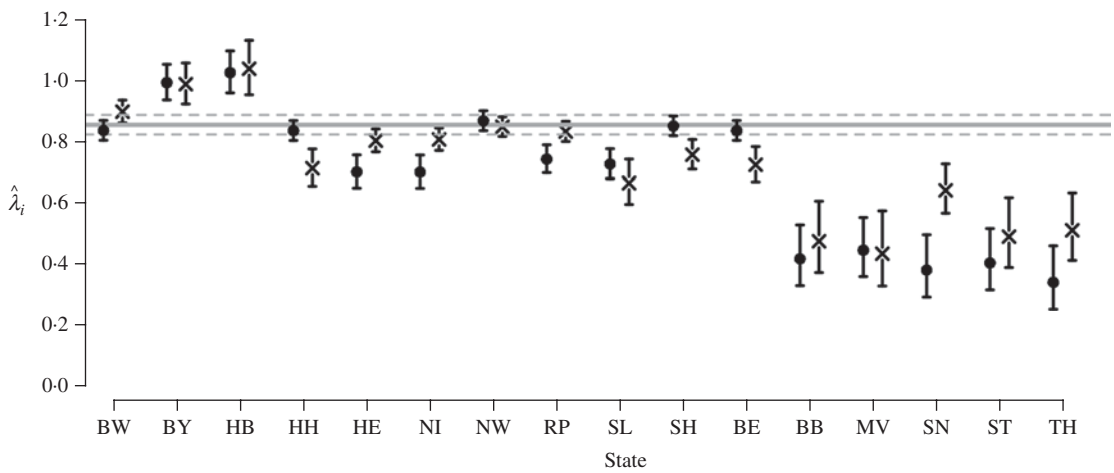

Fig. 2. Estimated autoregressive parameters $\hat{\lambda}_i$ and corresponding 95% confidence intervals for models A_0 (●) and A_1 (×). For comparison, the horizontal line denotes the estimated parameter $\hat{\lambda}$ for model C without covariates with the dashed lines representing the corresponding 95% confidence intervals. For definition of state abbreviations see Table 1.

In the next step, we investigated the impact of the inclusion of vaccination coverage in either the epidemic or endemic component compared to model C. Inclusion of the log proportion of unvaccinated school starters in the epidemic component (model A_0) leads to a considerably better fit.

The effect of the covariate β_1 in model A_0 is clearly significant ($P < 0{\cdot}0001$). Note that the estimated coefficients in the endemic component remain similar as in model C while the autoregressive parameter now varies across states. Inclusion of the covariate into the endemic component (model B_0) also improves the fit compared to model C but is worse compared to model A_0 according to AIC.

The above conclusions also hold when including the log proportion of school starters with at most one dose of MMR vaccine (models A_1, B_1). However, the model fit is considerably worse in terms of AIC. All results in Table 3 are based on the assumption that the coverage levels of the non-card holders are half those of card holders (adjustment factor 0·5). We tried several adjustment factors to investigate the robustness of our results. The ranking of the models according to AIC does not change for an adjustment factor $< 0{\cdot}6$. With regard to AIC an adjustment factor of 0·2 yields the best fit.

Figure 2 shows the estimated parameters λ_i and corresponding 95% confidence intervals for models

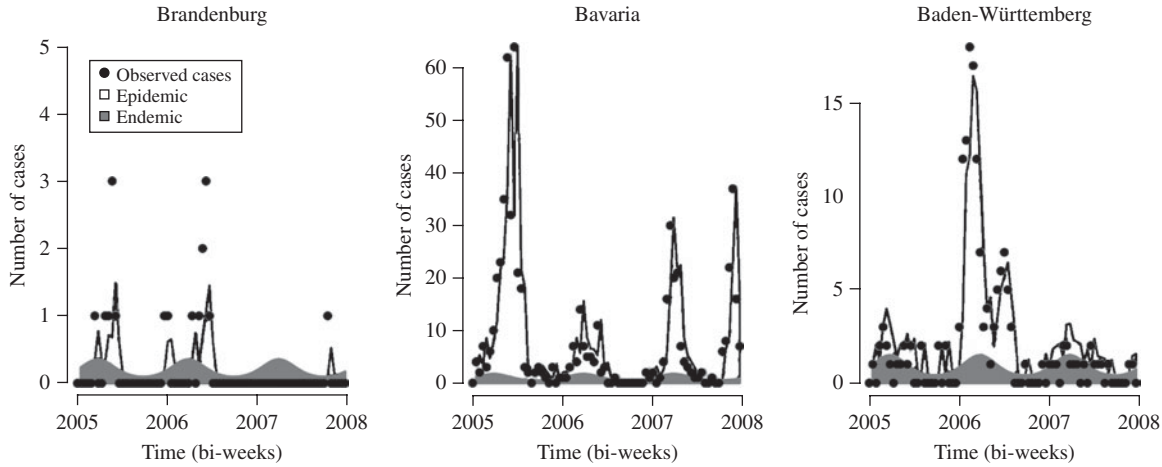

Fig. 3. Fitted mean for model A_0, which includes the log proportion of unvaccinated school starters as covariate in the epidemic component, in selected states.

A_0 and A_1 for each state. There is considerable heterogeneity across states. The ML estimates for the five states in East Germany are markedly lower than estimates for the remaining states. Vaccination coverage is considerably higher in these states. Note that model A_0 which includes the log proportion of unvaccinated school starters in the epidemic component performs better in terms of AIC than a model with the original (untransformed) proportion.

The analysis of the multivariate time-series of measles surveillance counts showed that there is an association between vaccination coverage and the occurrence and size of measles epidemics within states, with model A_0 fitting best. Figure 3 shows the fitted number of cases, decomposed into endemic and epidemic components, for this model in three of the states shown in Figure 1 for illustrative purposes. The estimated mean is clearly dominated by the epidemic component.

DISCUSSION

We observed a significant association between estimated vaccination coverage at school entry and the overall incidence of measles in the federal states of Germany (Table 3). The inclusion of the log proportion of unvaccinated school starters in the epidemic component of the model is the most suitable formulation to describe the occurrence and size of measles epidemics. This is plausible since the proportion of unvaccinated school starters acts as a proxy for the population of susceptibles, and the number of cases at a future time point depends on the number of infectious cases in the present as well as on the number of individuals susceptible to infection.

A strength of the proposed model is the decomposition of the disease incidence into an endemic and an epidemic component. Compared to a standard log-linear Poisson regression model our formulation is able to account for occasional outbreaks by including an autoregressive component. This is particularly important for the analysis of highly infectious diseases such as measles. In addition, information about vaccination coverage was included to cope with regional heterogeneity.

There are some limitations to this study. The RKI also provides estimates of vaccination coverage at school entry for children aged 4–7 for the years 2005 and 2007. However, the measles data comprise cases of all ages. Thus, changes in age-specific vaccination coverage may lead to shifts in the age distribution of the number of cases, but it will be impossible to discern such shifts from age-aggregated surveillance data. In addition, there is uncertainty about the true vaccination status, when obtained from school entry examinations. Hence small changes in coverage levels in successive years are not expected to be particularly meaningful. Therefore, we used only data for 2006 as an approximate measure of the overall immunization status in each state in all age groups.

We were aware that vaccination coverage was probably overestimated because vaccination uptake in school starters who presented vaccination cards is assumed to be higher [12]. Roughly 10% of school starters did not present vaccination cards and coverage for them is unknown. To assess the sensitivity of

the assumed coverage for those without cards (0·5 times that of card holders) we considered values ranging from the same coverage as children who presented cards (corresponding to 1) to all children who did not present cards being unvaccinated (corresponding to 0). In terms of AIC, model B where the covariate is included in the endemic component is not very sensitive with regard to the assumed coverage. In contrast, the AIC for model A where the covariate is included in the epidemic component changes considerably. When coverage for non-card holders is $>0{\cdot}6$ times that of card holders, model B is preferred.

Wichmann *et al.* [4] investigated a local outbreak in a school in Duisburg (North Rhine-Westphalia) in 2006. They estimated that receipt of one dose of MMR in the 22% without cards was 75% (significantly lower than the coverage of 95% in students with vaccination cards). This corresponds to a coverage level for non-card holders around 0·8 times that of card holders. However, this investigation involved only one school and no information about uncertainty around the estimated 75% coverage was given. The results are probably not generalizable to data at state level in this study. According to AIC, the measles data in our study are best described assuming coverage in non-card holders of 0·2 times that of card holders and using a model in which the proportion of unvaccinated school starters is incorporated in the epidemic component of the model.

To investigate the ability of AIC to identify the correct type of the model, we conducted a simulation study (Appendix). We used a simple model, comparable to the model of type A, where vaccination coverage influences the epidemic component. The simulation study showed that AIC identifies the true underlying model as long as the influence of vaccination coverage is strong or non-existent.

The proposed model approach allows us to consider infectious disease counts with several time-varying covariates. If quarterly, age-specific vaccination coverage was available, it could also be investigated whether vaccination-related trends in age-specific incidence [32] are observable using such notification data. Another interesting aspect would be to investigate the behaviour of the model where vaccination coverage is simultaneously included as an explanatory variable in both components. In this case, attention should be paid to potential issues related to multicollinearity or identifiability of parameter estimates.

Table 4. *Population sizes (N_i) and corresponding vaccination coverage levels (x_i) used in the simulation study*

Region	State	N_i	n_i	x_i	$\log(1-x_i)$
1	Bavaria	12 492 658	0·44	0·90	−2·30
2	Lower Saxony	7 982 685	0·28	0·85	−1·90
3	Saxony	4 249 774	0·15	0·85	−1·90
4	Berlin	3 404 037	0·12	0·80	−1·61

The states used in the simulation study were selected at random. The population fraction is denoted by n_i.

Table 5. *Models for the simulation analysis*

Model	Epidemic component $\log(\lambda_i)=\beta_0+$	Endemic component $\log(\nu_i)=\log(n_i)+\alpha_0+$
A	$\beta_1 \log(1-x_i)$	—
B	—	$\alpha_1 \log(1-x_i)$
C	—	—

In order to apply the proposed model to data at a finer spatial resolution we would need more detailed information about vaccination coverage because there are great regional and local differences leading to immunization gaps [6, 15]. For example, coverage levels for one dose of MMR vaccine ranged from 77·5% to 98% in the 77 health districts of Bavaria at school entry examinations 2005/2006 [33]. At a finer spatial resolution, it might also be necessary to account for spatio-temporal dependence, e.g. due to commuting. This could be done by including the previous number of cases in adjacent regions in the epidemic component [10, 23].

Although the data on measles incidence and vaccination coverage have some limitations, clear associations were observed. The pattern observed in the reported measles cases for all ages is best described by including the log proportion of unvaccinated school starters as an explanatory variable in the autoregressive (epidemic) component of the model.

APPENDIX: Simulation study

We investigated whether AIC identifies the correct structure of the model with a simulation study. Multivariate time-series of length $T=156$ (3 years of weekly data) were simulated based on a model where the number of cases $y_{i,t}$ in region i at time t is influenced by vaccination coverage as a covariate. Each

Table 6. *Results for the simulation study*

	Fixed parameters			Average number of cases				AIC % of model			
Sim	c	$\bar{\lambda}$	β_1	Reg 1	Reg 2	Reg 3	Reg 4	A	B	C	True model
1	10^{-4}	0·5	0	3813	2296	1201	1061	10·1	8·8	81·1	C
2	10^{-4}	0·5	0·1	3587	2613	1332	1014	26·1	19·8	54·1	A
3	10^{-4}	0·5	0·5	3232	2294	1259	1246	81·9	18·1	0·0	A
4	10^{-4}	0·8	0	4004	2396	1318	660	10·9	10·6	78·5	C
5	10^{-4}	0·8	0·1	3139	2377	1080	997	48·3	22·0	29·7	A
6	10^{-4}	0·8	0·5	2111	2247	1214	2162	99·3	0·7	0·0	A
7	10^{-5}	0·5	0	364	223	129	96	12·6	12·4	75·0	C
8	10^{-5}	0·5	0·1	376	249	141	98	14·1	11·6	74·3	A
9	10^{-5}	0·5	0·5	305	295	127	132	62·7	12·6	24·7	A
10	10^{-5}	0·8	0	333	256	206	107	13·7	13·8	72·5	C
11	10^{-5}	0·8	0·1	413	258	183	76	20·9	13·6	65·5	A
12	10^{-5}	0·8	0·5	234	146	146	359	87·1	3·9	9·0	A

AIC, Akaike's Information Criterion.
Parameter values are shown for the simulations (Sim), the mean number of cases for each region (Reg), and how often each model has the lowest AIC value (AIC % of model).

simulated dataset is analysed with different models and AIC is calculated.

We assumed that vaccination coverage influences the epidemic component, which also contains an intercept. The endemic component contains no seasonal terms, an overall intercept α_0 and population fractions n_i as offset. Four randomly selected regions are used where the population sizes N_i are selected from population data of Germany in 2006 and artificial vaccination coverage levels x_i are attached (see Table 4). The coverage levels x_i differ between the regions and have been transformed with $\log(1-x_i)$ as in the measles analysis. The simulation model corresponds to model A in Table 5 and is similar to model A_0 for the measles data (Table 3).

We chose different values for the yearly incidence c (10^{-4}, 10^{-5}) and the basic level of the epidemic component not influenced by covariates, $\bar{\lambda}$ (0·5, 0·8). Furthermore, we assumed that vaccination coverage has either no ($\beta_1=0$), a small ($\beta_1=0{\cdot}1$), or a strong ($\beta_1=0{\cdot}5$) influence. All combinations of these values give 12 different simulation scenarios. For each of these scenarios, 1000 datasets have been simulated. The incidence c and the population size N_i are used to calculate the mean number of cases for the first week $\mu_{i,1}$ for each region with $\mu_{i,1}=cN_i/52$. The parameter $\bar{\lambda}$ is used to calculate the intercept β_0 as a basic level

$$\beta_0=\log(\bar{\lambda})-\text{mean}(\beta_1\log(1-x_i)).$$

Next, the epidemic component λ_i is calculated as in model A (Table 5) and used for the simulation. The endemic component ν is calculated with the stationary mean equation [10]

$$\nu_i=\mu_{i,t}\frac{(1-\bar{\lambda})}{n_i}=\frac{cN_i}{52}\frac{(1-\bar{\lambda})\sum_i N_i}{N_i}$$
$$=\frac{c}{52}(1-\bar{\lambda})\sum_i N_i$$

and is the same for all regions. The cases $y_{i,t}$ are simulated for each region i and point in time t as follows:

$$y_{i,t}\sim Po\left(\frac{n_i\nu}{1-\lambda_i}\right) \qquad (t=1),$$
$$y_{i,t}\sim Po(\lambda_i y_{i,t-1}+n_i\nu) \quad (t=2,\ldots,T).$$

For the analysis of each simulated dataset three different models, listed in Table 5, have been considered. The models differ with regard to the influence of vaccination coverage: in the epidemic component, in the endemic component, or none. Note that the values of the covariates used in the analysis are the same as in the simulation.

The results of the analysis are shown in Table 6. In all simulations where there was no influence of vaccination coverage the true underlying model C resulted most frequently in the lowest AIC value (i.e. highest AIC %). When there was a small influence of vaccination coverage in the epidemic component, AIC in general preferred model C with no influence, followed by model A with influence in the epidemic component. When there was a strong influence, model A is clearly preferred. In summary, AIC identifies the true

underlying model as long as the influence of vaccination coverage is strong or non-existent.

ACKNOWLEDGEMENTS

We thank J. C. M. Heijne and N. Low for valuable comments and suggestions. Financial support by the Swiss National Science Foundation is gratefully acknowledged.

DECLARATION OF INTEREST

None.

REFERENCES

1. **Muscat M, *et al.*** Measles in Europe: an epidemiological assessment. *Lancet* 2009; **373**: 383–389.
2. **Wallinga J, Heijne JCM, Kretzschmar M.** A measles epidemic threshold in a highly vaccinated population. *PLoS Medicine* 2005; **2**: e316.
3. **Siedler A, *et al.*** Two outbreaks of measles in Germany 2005. *Eurosurveillance* 2006; **11**: 131–134.
4. **Wichmann O, *et al.*** Large measles outbreak at a German public school, 2006. *Pediatric Infectious Disease Journal* 2007; **26**: 782–786.
5. **Bernard H, *et al.*** An outbreak of measles in Lower Bavaria, Germany, January–June 2007. *Eurosurveillance* 2007; **12**: pii = 3278.
6. **Wichmann O, *et al.*** Further efforts needed to achieve measles elimination in Germany: results of an outbreak investigation. *Bulletin of the World Health Organization* 2009; **87**: 108–115.
7. **Cummings DAT, *et al.*** Improved measles surveillance in Cameroon reveals two major dynamic patterns of incidence. *International Journal of Infectious Diseases* 2006; **10**: 148–155.
8. **Frank C, *et al.*** Cattle density and Shiga toxin-producing Escherichia coli infection in Germany: increased risk for most but not all serogroups. *Vector-Borne and Zoonotic Diseases* 2008; **8**: 635–643.
9. **Hens N, *et al.*** Estimating the impact of vaccination using age-time-dependent incidence rates of hepatitis B. *Epidemiology and Infection* 2008; **136**: 341–351.
10. **Held L, Höhle M, Hofmann M.** A statistical framework for the analysis of multivariate infectious disease surveillance counts. *Statistical Modelling* 2005; **5**: 187–199.
11. **Robert Koch Institute.** SurvStat (http://www3.rki.de/SurvStat). Accessed 14 October 2009.
12. **Robert Koch Institute**. On vaccination coverage estimated at school entry examinations in Germany, 2006 [in German]. *Epidemiologisches Bulletin* 2008; **7**: 55–57.
13. **Lindsey JK, Jones B.** Choosing among generalized linear models applied to medical data. *Statistics in Medicine* 1998; **17**: 59–68.
14. **Statistisches Bundesamt.** 12411-0009 Projections of the resident population for German states on reference date (https://www-genesis.destatis.de/genesis/online/logon). Accessed 21 January 2009.
15. **Kalies H, *et al.*** Immunisation status of children in Germany: temporal trends and regional differences. *European Journal of Pediatrics* 2006; **165**: 30–36.
16. **Reiter S, Poethko-Müller C.** Current vaccination coverage and immunization gaps of children and adolescents in Germany [in German]. *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz* 2009; **52**: 1037–1044.
17. **Poethko-Müller C, *et al.*** Vaccination coverage against measles in German-born and foreign-born children and identification of unvaccinated subgroups in Germany. *Vaccine* 2009; **27**: 2563–2569.
18. **Tischer A, Siedler A, Rasch G.** Surveillance of measles in Germany [in German]. *Gesundheitswesen* 2001; **63**: 703–709.
19. **Hellenbrand W, *et al.*** Progress toward measles elimination in Germany. *Journal of Infectious Diseases* 2003; **187** (Suppl. 1): S208–S216.
20. **Palmgren J.** Poisson distribution. In: Armitage P, Colton T, eds. *Encyclopaedia of Biostatistics*, 2nd edn. West Sussex: John Wiley and Sons, 2005, pp. 4109–4113.
21. **Kirkwood B, Sterne J.** *Essential Medical Statistics*, 2nd edn, 2003. Malden: Wiley.
22. **Kuhn L, Davidson LL, Durkin MS.** Use of Poisson regression and time series analysis for detecting changes over time in rates of child injury following a prevention program. *American Journal of Epidemiology* 1994; **140**: 943–955.
23. **Paul M, Held L, Toschke M.** Multivariate modelling of infectious disease surveillance data. *Statistics in Medicine* 2008; **27**: 6250–6267.
24. **Knorr-Held L, Richardson S.** A hierarchical model for space-time surveillance data on meningococcal disease incidence. *Journal of the Royal Statistical Society Series C* 2003; **52**: 169–183.
25. **Anderson RM, May RM.** *Infectious Diseases of Humans: Dynamics and Control*. Oxford: Oxford University Press, 1991.
26. **R Development Core Team.** *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2009.
27. **Höhle M.** Surveillance: an R package for the monitoring of infectious diseases. *Computational Statistics* 2007; **22**: 571–582.
28. **Diggle PJ.** *Time Series. A Biostatistical Introduction*. New York: Oxford University Press, 1990.
29. **Fine PEM, Clarkson JA.** Measles in England and Wales – I: An analysis of factors underlying seasonal patterns. *International Journal of Epidemiology* 1982; **11**: 5–14.
30. **Cliff AD, Haggett P.** Statistical modelling of measles and influenza outbreaks. *Statistical Methods in Medical Research* 1993; **2**: 43–73.
31. **Finkenstädt BF, Grenfell BT.** Time series modelling of childhood diseases: A dynamical systems approach.

Journal of the Royal Statistical Society: Series C 2000; **49**: 187–205.

32. **Fine PEM, Clarkson JA.** Measles in England and Wales – II: The impact of the measles vaccination program on the distribution of immunity in the population. *International Journal of Epidemiology* 1982; **11**: 15–25.
33. **Bayerisches Landesamt für Gesundheit und Lebensmittelsicherheit.** MMR vaccination coverage of children starting school in Bavaria in 2006/07 at a regional level (http://www.lgl.bayern.de/gesundheit/gesundheitsindikatoren/themenfeld07/indikator0714.htm). Accessed 10 June 2009.