Can economic indicators predict infectious disease spread? A cross-country panel analysis of 13 EU countries

Paul R Hunter[1], Felipe J Colón-González[2], Julii Brainard[1]*, Batsirai Majuru[1], Debora Pedrazzoli[3], Ibrahim Abubakar[4], Girmaye Dinsa[5], Marc Suhrcke[6], David Stuckler[7], Tek-Ang Lim[8] and Jan C. Semenza[9]

[1] Norwich Medical School, University of East Anglia NR4 7TJ, UK

[2] School of Environmental Sciences, University of East Anglia, UK

[3] Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK

[4] Institute for Global Health, University College London, UK

[5] T.H. Chan School of Public Health, Harvard University MA USA

[6] Centre for Health Economics, University of York, UK

[7] Dondena Research Centre, University of Bocconi, Milan, Italy

[8] Science and International Office, French Public Health Agency, France

[9] European Centre for Disease Prevention and Control (ECDC), Stockholm, Sweden

* Corresponding Author: j.brainard@uea.ac.uk. Tel. 01603-591151, Orcid No. = 0000-0002-5272-7995

Abstract

Aims: It is unclear how economic factors impact on the epidemiology of infectious disease. We evaluated the relationship between incidence of selected infectious diseases and economic factors including recession in 13 European countries between 1970 and 2010. *Methods*: Data were obtained from national communicable disease surveillance centres. Negative binomial forms of Generalised Additive and Generalised Linear Models (GAM and GLM) were tested to see which best reflected transmission dynamics of: diphtheria, pertussis, measles, meningococcal disease, hepatitis B, gonorrhoea, syphilis, hepatitis A and salmonella. Economic indicators were gross domestic product per capita (GDPpc), unemployment rates, and recession. *Results*: GAM models produced the best goodness of fit results. The relationship between GDPpc and disease incidence was often nonlinear. Strength and directions of association between population age, tertiary education levels, GDPpc and unemployment were disease dependent. Overdispersion for almost all diseases validated the assumption of a negative binomial relationship. Recession was not independently linked to disease incidence. ***Conclusions*: Social and economic factors can be correlated with many infections. However, the trend is not always in the same direction and these associations are often non-linear. Recession as an indicator of increased disease risk may be better replaced by GDPpc or unemployment measures.**

Keywords: Gonorrhoea; Hepatitis B; Measles; meningococcal disease; Pertussis; Salmonella

WORD COUNT = 3600

Background

Lessons from historical literature and research suggest that the impacts of an economic downturn can be far reaching [1]. These include: shifts in trends in health risk, such as immunisation levels and utilisation of health services; differential impact on vulnerable population groups; and shifts in demand for health services from private sector to public sector [2].

In high-income countries such as most of Europe, successes in infectious disease prevention and control have been high. However, 9% of the disease burden in Europe in Disability Adjusted Life Years (DALYs) is still attributable to infectious diseases [3]. The real-world impacts from financial crisis on infectious disease prevention and control in modern European conditions are unclear [4,5]. It is clear that austerity budgets often lead to reduced service provision. For instance, diminished screening, treatment and case management services were documented over the period 2008-2009 for American services in sexual health [6]. A 2009-2010 survey among European infectious disease experts and policy-makers also found considerable concern about budget cuts specifically in relation to infectious disease prevention services and in particular those programmes targeted at vulnerable subgroups such as drug abusers and sex workers [7].

There is much literature evaluating the relationships between health outcomes and economic status. Concerns that infection control measures would specifically be reduced because of austerity budgets have been voiced [4]. However, research on how the recent economic recession influenced epidemiological patterns is limited. Generally, incidence of communicable diseases in high income countries seems to have fallen in the last decade [5,8]. but some increases in observed communicable diseases were observed following the 2007/08 financial

crisis, notably for influenza, West Nile virus, HIV and indigenous malaria in Greece [9,10]. However, most studies to review incidence of poor health linked to the financial crisis have tended to dwell on chronic diseases (including impacts on mental health) and overall mortality rates, rather than communicable diseases [5,8,11]. Lack of research on changes in infectious diseases is partly due the paucity of observed and comparable data across countries and regions and differences in models and indicators used [5,12].

The analysis in this study seeks to improve understanding of the relationship between macroeconomic factors and infectious disease incidence in the European Union (EU) through a quantitative assessment across 13 EU countries. Such an understanding is valuable, as it would provide guidance on the future impact of macroeconomic downturns for the spread of infectious diseases. The ensuing appropriate and effective policy responses and strategies would not only mitigate potential negative health impact, but also maximise the efficiency of health systems during periods of macroeconomic downturn.

Methods

Epidemiological data

We approached all national infectious diseases surveillance centres within the European Union for access to their infectious disease surveillance data from the years 1970 onwards. Some data were already available in the public domain, and some were generated specifically for the researchers. In choosing which diseases to study we wanted to ensure that we had representative selections of vaccine preventable diseases, infections spread by sexual contact and also food-borne and person-to-person spread enteric infections. In addition, we were only interested

in infections that were likely to be included in most national disease surveillance systems over most of the study period. We excluded diseases based on these criteria:

• Diseases where widespread use of effective diagnostic methods was only recent or was slow to be adopted across Europe (such as Hepatitis C, Chlamydia, Cryptosporidiosis and Campylobacter).

• Notifiable diseases where the recent incidence in most countries was very low/non-existent (such as cholera).

• Infectious diseases where diagnosis is most often delayed (such as HIV).

• Diseases with substantial year on year variation because of pathogen specific factors (i.e. influenza).

• Tuberculosis because of long latency.

These decision criteria led to selection for detailed analysis of four vaccine preventable diseases (diphtheria, measles, pertussis and meningococcal infection), three sexually transmitted or blood-borne disease (gonorrhoea, syphilis and hepatitis B) and two food borne diseases (salmonella and hepatitis A).

Economic data

We acquired data on unemployment, GDP per capita (measured in USD), tertiary education enrolment and demographic characteristics from the World Bank's World Development Indicators [13]. Data were obtained for Finland, Germany, Hungary, Ireland, Italy,

the Netherlands, Norway, Poland, Portugal, Spain, Romania, Sweden and the United Kingdom. Data are available in supplementary file.

Analytical approach

Our outcome variable was the disease total annual count of disease reports per country. Our principal macroeconomic predictor variables were unemployment and Gross Domestic Product per capita (GDPpc). We also incorporated a Boolean variable representing recession years determined by whether or not GDP had declined over the previous year as a proxy for macroeconomic downturn. Hence, a "recession" in the context of this study was defined as negative growth in GDPpc from one year to the next. We also adjusted the models for the proportion of people who were enrolled in tertiary education (gross enrolment ratio from World Bank data).

We initially explored the effects of the most recent recession period (i.e. 2008–2010) upon the different disease reports. For comparative purposes, we also explored the effects of other recession periods (1970–2010) using an Organisation of Economic Development (OECD) Composite Leading Indicators data set [14]. Boolean variables were created to indicate recession periods on both the 2008–2010 and the 1970–2010 data sets.

Statistical model

The expected number of cases for each disease E(Yit) for country i = 1,. . . ,I at year t = 1,. . . ,T was modelled using a generalized additive model (GAM) approach to account for potential nonlinear associations between variables. GAMs are semi-parametric extensions of the

widely used generalized linear model where the linear predictor is replaced by the sum of smooth functions of the covariates [15]. The optimal degree of non-linearity between the outcome and the predictors is estimated using Generalized Cross-Validation [16]. To account for possible over-dispersion in the data, we fitted Negative Binomial models. The general algebraical definition of the models is given by:

$$\text{Log}(\mu_{it}) = \eta_{it} \qquad (1)$$

$$\eta_{it} = \alpha + Log(\xi_{it}) + Log(Y_{it-1}) + t' + \sum_{p=1}^{P} f(x_{it,l}) + \sum_{q=1}^{Q} \beta(z_{it}) + d_i \qquad (2)$$

where η_{it} is a natural logarithmic link function of the expectation $E(Y_{it} \equiv \mu_{it})$, with Y_{it} as the time series of annual disease reports. The term α corresponds to the intercept; $\text{Log}(\xi_{it})$ denotes the natural logarithm of the population at risk for country *i* and year *t* included as an offset to adjust the epidemiological data by population. $\text{Log}(Y_{it} - 1)$, is the natural logarithm of the outcome disease counts lagged one year to account for potential auto-correlation in the data [17]. Here, t' is an index variable of year to control for possible long-term trends. The term $f(x_{it}, l)$ corresponds to smoothed relationships between the socioeconomic predictors lagged zero to three years and disease incidence defined by cubic regression splines. Lagged socioeconomic variables were computed using a four-year moving average for the year variable to account for the lagged effects of the socioeconomic predictors on disease incidence. The demographic predictor term z_{it} with regression coefficients β enter the model linearly. Country-specific fixed effects (d_i) were included to account for unknown or unobserved variables in the models [18].

Model selection

Some arguments were constant to all models. First, the logarithm of the population was incorporated as an offset to estimate relations on the crude incidence rate rather than on the total number of cases. Second, the logarithm of the outcome variable lagged one year was included in all models because epidemiological observations near in time are likely to be more similar than those distant in time [17]. Third, an index variable for the year of the observations was incorporated to control for potential long-term trends that may be due to factors other than socioeconomic development [19]. Finally, we incorporated country-specific fixed effects to account for unknown or unobserved variables in the model such as diagnostic performance variability, and interventions [18]. All models were fitted using the "mgcv" package for R [20].

A series of models were then fitted using all socioeconomic predictors (GDP per capita and unemployment and tertiary education) lagged 0:3 years, and all demographic predictors (population 15-65 years, and population over 65 years) in isolation, as well as in all their possible combinations. Thus, we successively fitted all possible models containing one socioeconomic or demographic predictor at a time, then two predictors at a time, and so on, until all predictors were included altogether in a single model. We measured the goodness of fit of each model using the Akaike information criterion (AIC) [21]. The model with the lowest AIC value was selected.

Comparative analyses

To ensure the robustness of our estimates, we compared the results of our Negative Binomial GAM against those obtained using a Negative Binomial generalised linear model (GLM). The model specification of these models was as in the GAM models except for the smooth predictors f(xit,l) that entered the model linearly instead.

Results

Overall, disease incidences had a downward trend from 1970 to 2010. Table I presents the summary statistics for each disease and economic variable. Table II indicates the level of missing data by disease and country. The panel data set contained country-specific observations for a 41 year period for a total of 513 country-year data points. The average sample per disease amounts to 296 country-year observations due to missing data.

Table I: Summary statistics across all countries and years

Outcome	Mean	Median	SD	25th percentile	75th percentile	N	Missing
Diphtheria	19.3	0.0	73.3	0.0	3.0	331	182
Gonorrhoea	6012.0	981.0	9993.7	284.0	8275.0	305	208
Hepatitis A	6956.0	489.0	15576.5	166.0	2416.0	261	252
Hepatitis B	2160.5	387.5	3709.3	166.2	2084.5	256	257
Measles	18185.0	415.0	41272.5	28.0	9024.0	363	150
Meningococcal	498.7	233.0	827.7	96.5	500.5	303	210
Pertussis	3482.0	587.0	7447.4	107.0	3092.0	342	171
Salmonella	17393.0	6653.0	27037.5	2304.0	23097.0	239	274
Syphilis	1566.0	545.0	2402.9	190.0	1822.0	263	250
Predictor	**Mean**	**Median**	**SD**	**25th percentile**	**75th percentile**	**N**	
GDPpc	21929.0	21077.0	8924.7	15340.0	27934.0	438	
Unemployment	8.5	7.7	4.5	5.3	10.5	310	
Tertiary	36.8	31.5	21.1	18.9	54.2	453	

Notes: Data from sources described in text. Units are, Diseases: counts per country. GDPpc: USD. Unemployment: %. Tertiary education: gross enrolment ratio.

Table II. Percentage of completion in disease reporting per EU country.

Country	Diphtheria	Gonorrhoea	Hepatitis A	Hepatitis B	Measles	Meningococcal	Pertussis	Salmonella	Syphilis
Finland	39.0	39.0	39.0	39.0	39.0	39.0	39.0	39.0	39.0
Germany	0.0	0.0	51.2	51.2	24.4	51.2	24.4	51.2	24.4
Hungary	97.6	97.6	36.6	36.6	97.6	97.6	97.6	97.6	97.6
Ireland	100.0	34.1	70.7	70.7	100.0	70.7	100.0	70.7	22.0
Italy	97.6	46.3	29.3	56.1	97.6	39.0	97.6	29.3	46.3
Netherlands	51.2	80.5	51.2	51.2	51.2	51.2	51.2	0.0	80.5
Norway	22.0	58.5	97.6	85.4	78.0	97.6	87.8	0.0	61.0
Poland	95.1	95.1	95.1	73.2	95.1	95.1	95.1	95.1	36.6
Portugal	51.2	68.3	51.2	46.3	48.8	0.0	51.2	31.7	68.3
Romania	95.1	95.1	73.2	73.2	95.1	95.1	95.1	70.7	95.1
Spain	95.1	65.9	29.3	29.3	95.1	95.1	65.9	48.8	65.9
Sweden	29.3	29.3	29.3	29.3	29.3	29.3	29.3	29.3	29.3
UK	82.9	82.9	31.7	31.7	82.9	26.8	48.8	68.3	24.4

We tried 31 different model specifications for each disease in the data set. The model specification with the lowest AIC estimate was selected. The selection based on the lowest AIC lead to different model parameters being used for each disease. Overall, generalised additive models (GAM) resulted in lower AIC values than generalised linear models. Table III presents the results for the final models selected for each disease. The values in bold font indicate that the estimate was significant at the 0.05 level. Note that a great deal of the deviance is explained by

the selected models (explained deviance range: 85%–99%). With the exception of the model for Diphtheria that suggests that data were under-dispersed (dispersion = 0.478), the dispersion statistic of all models was close to one suggesting that the Negative Binomial specification of the model was adequate to account for potential over-dispersion in the data.

Table III. Comparing the relative quality of generalised additive and generalised linear models.

Disease	GAM AIC	GLM AIC
Diphtheria	495.5	494.9
Gonorrhoea	3233.0	3262.8
Hepatitis A	3493.9	3495.4
Hepatitis B	2295.6	2294.7
Measles	2980.1	2979.1
Meningococcal	2316.0	2348.7
Pertussis	2823.4	2827.5
Salmonella	2670.0	2695.2
Syphilis	2855.1	2879.3

The Boolean variable used here to specify the 2008–2010 recession period had p-values > 0.05 for all diseases after accounting for the effects of autocorrelation, long-term trends, and the demographic and macroeconomic predictors in the model. Such observation suggests that the 2008–2010 recession did not play an important role on the occurrence of the diseases under scope. Similar results were obtained when accounted for all recession years (i.e. 1970, 1971, 1973, 1975, 1980–1982, 1986, 1990–1993, 1995, 1996, 1998, 2000–2003, and 2008–2010) based on the OECD data.

Table IV. GAM estimated relationships between disease incidence and each predictor incorporated in the final models

Term	Diptheria	Gonorrhoea	Hepatitis A	HepatitisB	Measles	Men'1 Disease	Pertussis	Salmonella	Syphillis
Log(Yit−1)	-0.027	**0.748**	**0.624**	**0.697**	**0.282**	**0.742**	**0.575**	**0.512**	**0.453**
Year	**0.526**	-0.030	0.045	-0.001	0.024	-0.008	**0.089**	-0.001	0.010
Recession	-0.656	0.084	0.083	-0.045	0.583	-0.063	-0.194	-0.030	0.019
Tertiary	-	**0.008**	>0.001	0.002	**-0.089**	0.001	0.006	**-0.012**	**0.027**
Population 15-65 yrs	**-0.554**	**-0.092**	-0.069	0.049	-0.202	0.052	**-0.261**	-0.053	-0.055
Population Age 65+	**-2.287**	**-0.096**	**-0.434**	-0.175	**-0.315**	-	**-0.469**	-	-
Country fixed effects	Included	Included	Included	Included	Included	Included	Included	Included	Included
Smoothers (edf)									
GDPpc	**2.932**	1.001	3.733	**7.210**	**6.794**	**1.001**	4.317	**5.836**	**7.174**
Unemploy-ment	1.000	1.000	2.436	2.491	1.006	1.000	1.002	**5.369**	**4.615**
Model statistics									
Dispersion	0.478	1.099	1.506	1.296	1.412	1.202	1.437	1.098	1.229
Deviance explained	0.958	0.975	0.940	0.974	0.846	0.967	0.901	0.990	0.975

Notes: Values in bold font were significant at ≤ 0.05 level. *edf* = estimated degrees of freedom. The dashes indicate that the parameter was not included in the model.

A positive relationship was observed in Table IV between the increasing access to tertiary education and the occurrence of gonorrhoea and syphilis suggesting that as access to tertiary education increases, so does the incidence of these two sexually transmitted diseases. Conversely, a negative relationship was observed between the increasing access to tertiary education and the occurrence of measles and salmonella suggesting a possible protective effect of education against these diseases.

The proportion of the population with 65 years of age or more showed a negative effect on disease occurrence in all models where it was included, at $p \leq 0.05$, except for Hepatitis B (Table IV). Similarly, we estimate a negative relationship between the proportion of the population between 15 and 65 years of age and the occurrence of diphtheria, gonorrhoea and pertussis. The estimated degrees of freedom (edf) of the smoothed macroeconomic predictors suggest that the relationship between GDPpc and disease occurrence is highly nonlinear (edf > 1) for all diseases except for gonorrhoea and meningococcal infection where we estimate a log-linear relationship (edf ≈ 1). The relationship between disease occurrence and unemployment was also highly nonlinear for Hepatitis A, Hepatitis B, Salmonella and Syphilis, and approximately log-linear for all the other diseases.

Figure 1 depicts the functional form of the estimated effects of GDPpc and unemployment on disease occurrence. The X axis indicates the values of the macroeconomic predictors, and the Y axis the estimated response of the disease outcome variable in a logarithmic scale. Diphtheria data are described in text but not shown in Figs. 1-2 due to a greater than two factor of two difference in average incidence compared to all the other diseases, as indicated by summary statistics in Table I.

It is important to note that the direction of the association varies between the different

diseases. Overall, we estimate a protective effect (decline in disease) with rising GDPpc for diphtheria, hepatitis A, hepatitis B, meningococcal infection and syphilis. Conversely, a positive effect was estimated for gonorrhoea and measles. GDPpc showed a quadratic-like relationship with Salmonella with a gradual rise in log-incidence up to about USD 25,000 after which the rate of increase in incidence levels off and then decays at about USD 35,000. The relationship between GDPpc and syphilis is interesting as there is a marked decrease in log-incidence for values below USD 28,000 after which the effect flattens. The estimated effect of unemployment was positive for measles, pertussis, and salmonella and negative or ambiguous for the other diseases. Similar results were estimated for the two macroeconomic predictors when accounting for the effects of all recession periods in the model.

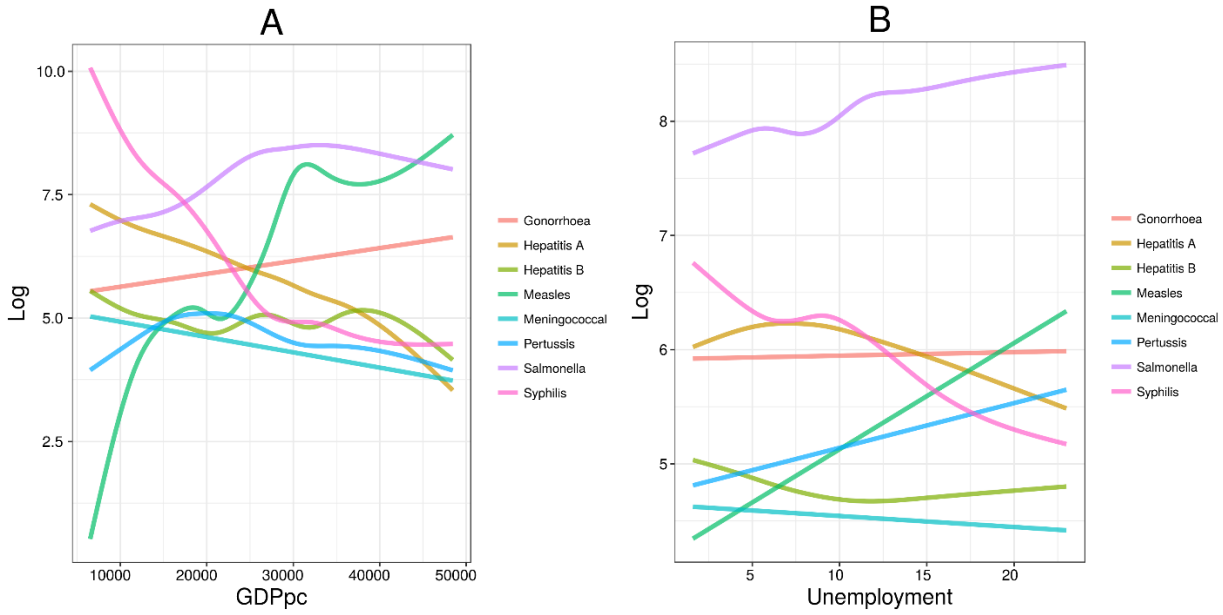

Figure 1: GAM-estimated smooth relationships between disease occurrence and (A) GDP per capita, and (B) unemployment. Diptheria not shown on from the figures due to scale incompatibility.

The direction of the GLM-estimated relationships between disease incidence, GDPpc and unemployment was similar for most GAMs and GLM models (results of GLM models shown in Figure 2). This was not the case, however, for the relationships between GDPpc, Hepatitis B, Salmonella and Syphilis.

When we compared the results obtained with our Negative Binomial GAM against those from the Negative Binomial GLM, we observed an increase in the dispersion parameter and a decrease in the explained deviance (Table V). The increase in the dispersion parameter suggests that some of the covariates in the Negative Binomial GAM may indeed have a nonlinear effect. Replacing the smoothed macroeconomic predictors with linear predictors resulted in relationships that were mainly not significant at the 0.05 level both for GDPpc and unemployment.

Table V. GLM estimated coefficients and model statistics

Term	Diptheria	Gonorrhoea	Hepatitis A	HepatitisB	Measles	Men'1 Disease	Pertussis	Salmonella	Syphillis
Log(Yit−1)	0.045	**0.749**	**0.674**	**0.785**	**0.334**	**0.743**	**0.613**	**0.771**	**0.690**
Year	**0.608**	**-0.030**	0.028	-0.031	0.056	-0.009	0.066	0.011	-0.018
Recession	-1.122	0.084	0.108	-0.076	0.449	-0.064	-0.084	-0.068	0.046
Tertiary-education	-	**0.008**	-0.001	-0.004	**-0.056**	0.001	0.016	-0.006	0.006
Population 15-65 yrs	**-0.520**	**-0.092**	-0.033	0.020	**-0.082**	**0.052**	**-0.150**	-0.001	**-0.102**
Population age 65+	**-2.089**	**-0.096**	**-0.310**	-0.023	**-0.311**	-	**-0.412**	-	-
Country fixed effects	Included	Included	Included	Included	Included	Included	Included	Included	Included
GDPpc	**-0.001**	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Unemployment	-0.184	0.003	**-0.036**	-0.017	-0.003	-0.010	**0.045**	0.009	-0.021
Model statistics									
Dispersion	0.559	1.116	1.590	1.450	1.831	1.205	1.486	1.177	1.305
Deviance-explained	0.931	0.970	0.900	0.939	0.778	0.956	0.880	0.969	0.948

Note: Values in bold font were significant at the 0.05 level. edf = estimated degrees of freedom.

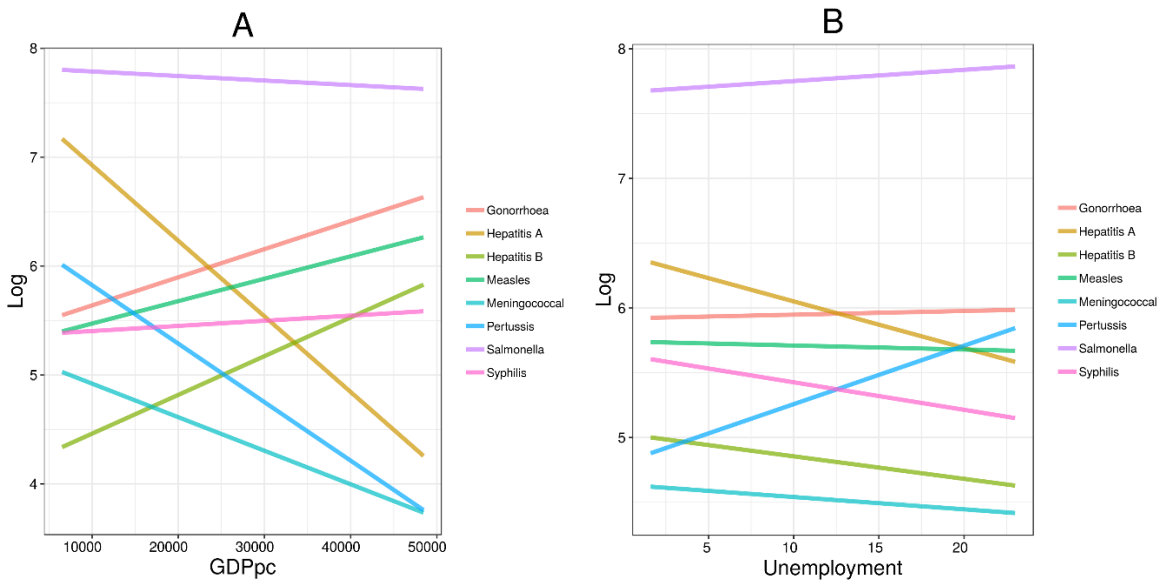

Figure 2: GLM-estimated smooth relationships between disease occurrence and (A) GDP per capita, and (B) unemployment. Diptheria not shown on from the figures due to scale incompatibility.

Discussion

For this study we brought together one of the most comprehensive annualised datasets on infectious disease reporting across different European Union member states to date. We then used this dataset to model the impact of standard macroeconomic variables on disease incidence. We demonstrate that the relationships between these economic variables and disease incidence are highly non-linear.

If, as we suspect, non-linearity between economic variables and outcome variables is common and consistent, then using linear models to predict changes in incidence of public health problems [22-25] may be inappropriate, as it will likely under-estimate the impact of macroeconomic drivers on disease incidence.

Our discussion focuses on the superior GAM models (Figure 1), which analysis found strong but non-linear relationships between the incidence of many of the infectious diseases and economic variables. We found that generally, rising per capita GDP had a protective effects against diphtheria, hepatitis A and B and meningcoccal disease. In contrast, rising unemployment (an indicator of economic downturn) also had a protective effect against most diseases, except gonorrhoea, measles pertussis and salmonella. While it is interesting to speculate on the mechanisms underlying the associations between disease incidence and economic variables identified in this study, it is important to remark that any suggestions for such mechanisms do not themselves come out of the data and cannot be confirmed or refuted from the available data. It should also be remembered that the non-linear relationship between many of the infectious diseases and economic variables means that a change in an economic variable, either within the same country over time or between countries will depend on the starting value.

Economic theory suggests that unemployment and GDP should have opposite relationships with the same indicators (because unemployment falls as GDP rises, see Okun's Law [26]). However, collinearity means that it is reasonable to view GDP as the primary predictor in these models; and thus where the relationship between disease incidence and unemployment runs counter to intuition, this may reflect collinearity with GDP in the same model, rather than an underlying distinct economic relationship. Therefore the below discussion will focus on possible explanations for the apparent relationship between GDP (only) and disease incidence.

An increase in gonorrhoea with rising GDP may reflect increasingly relaxed social attitudes about sexual activity, accompanied by widespread subclinical and hence untreated

infection. Increased incidence of sexual activities and consequential sexually transmitted disease has been linked to higher rates of tertiary education [27, 28] .These observations may seem to contradict the fall of cases of syphilis (fell with rising GDP in our dataset). However, the clinical course of syphilis is very different from gonorrhoea. Syphilis is noted for multiple phases of symptoms, sometimes including long dormant periods followed by quite severe health problems. This contrasts with gonorrhoea which may never cause symptoms at any stage of infection.

The fluctuating relationships (rises and falls) of some diseases (hepatitis B, meningoccocal) with GDP probably reflects multiple unconsidered factors, such as different immunisation programmes in different countries for these diseases. Hepatitis A, whose main risk factor is sanitation and hygiene practices [29], shows a marked decline with increasing GDPpc. As the disease has been more common in poorer regions such as Eastern and Central Europe [30], the improvement in sanitation associated with increased GDP in these regions could likely be a factor in the reduction of the disease. There was a marked decline in hepatitis A incidence in Hungary, Poland and Romania (data not shown). While the incidence of vaccine preventable infections generally shows a negative association with increasing GDPpc, measles is a striking exception. Increasingly large outbreaks of the disease continue to occur in Europe, as a result sub-optimal vaccination uptake [31-34]. Higher rates of measles in our data probably links indirectly to less tertiary education for both socially marginalised groups and religious minorities, who may also be less likely to pursue tertiary education. For instance, Roma and Travellers often lack both formal education and consistent access to vaccination programmes ([35]). Some other specific religious and philosophical minorities who have been especially affected in European measles outbreaks are known to prefer to not vaccinate [35,36] and may also be less likely to pursue tertiary education than their socioeconomic counterparts in the

general population [37-39].

We have been transparent about the limitations in the work presented here. Because our dataset is based on annual data, our definition of recession is limited to year on year decline in GDP, which differs somewhat from the more commonly used definition of two consecutive quarters of declining GDP [40]. Thus, our reliance on annual changes in GDP prevents us from detecting within-year relationships in the variables. Also, although our study presents what is perhaps the most comprehensive dataset of notifiable infectious diseases across a number of European countries, this dataset is not complete for all countries for all years. This could have masked the full extent of relationships between macroeconomic factors and infectious diseases. Overall differences in the sensitivity of surveillance between countries are accounted for in the panel structure of our analyses. However, the analyses cannot account for potential temporal changes in sensitivity of surveillance within countries. Our modelling strategy was narrow, we did not consider other modelling approaches, such as segmented regression which can also overcome some of the shortcomings associated with using simple linear regression to describe the possible relationship between disease incidence and economic indicators [41-43].

One issue that deserves further comment is the inclusion of lagged log incidence in order to control for autocorrelation in disease incidence within country. It has been pointed out that the use of lagged variables in Generalised Least Squares or Ordinary Least Squares can "squash" the apparent effect of other predictor variables [44]. Achen showed that this could be artefactual and due to a combination of high serial correlation and heavy trending in the exogenous variables [44]. However, as pointed out more recently, the use of lagged dependent variables is often appropriate for dynamic models [45]. Given that most models of infectious disease epidemiology are highly dynamic the inclusion of a lagged dependent variable is appropriate. Furthermore, all

of the evidence against the use of lagged variables is based on linear regression models and may not apply to the GAM used here.

Conclusions

The prevailing macroeconomic climate and its impact on disease outcomes remains an important concern. Evidence on both the nature and the likely outcomes of these relationships is key to decision-making. Here we have reported on the relationships between macroeconomic factors and infectious disease outcome across 13 European countries and over a period of 40 years (1970–2010). Most notably, compared to linear models, our application of Generalised Additive Models proved to give a valid and best fit. A key finding from this study are highly non-linear relationships between macroeconomic indicators and infectious disease incidence. We found limited evidence of the effect of recession on infectious disease independent of any effect of GDPpc, but other macroeconomic factors may be important drivers of the disease trends observed.

Acknowledgements

This study was supported by the European Centre for Disease Control and Prevention. Paul Hunter, Julii Brainard and Felipe J Colón-González are affiliated to the National Institute for Health Research (NIHR) Health Protection Research Unit in Emergency Preparedness and Response led by Kings College London in partnership with Public Health England (PHE). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health, or PHE.

Conflict of interest: The authors have no financial conflicts.

Contributions: PRH, IA, MS, JCS, DS and TAL conceived of the study. Database assembly by PRH, DP and IA. Analysis by FCG, PRH, JB, GD, MS, DS and BM. FCG undertook GLM and

GAM modelling. PRH wrote the first draft of the manuscript. PRH, FCG and JB assembled revisions and the final article. All authors approve of this version of the manuscript for submission.

References

1. Suhrcke M, Stuckler D, Suk JE, Desai M, Senek M, McKee M, et al. The impact of economic crises on communicable disease transmission and control: a systematic review of the evidence. *PloS One* 2011; 6(6): e20724.

2. Phua K-L. Can we learn from history? Policy responses and strategies to meet health care needs in times of severe economic crisis. *Open Public Health J* 2011; 4: 1-5.

3. Mladovsky P, Allin S, Masseria C, Hernández-Quevedo C, McDaid D, Mossialos E. Health in the European Union: Trends and analysis. European Observatory on Health Systems and Policies, 2009.

4. Semenza JC, Tsolova S, Lim TA. Economic crisis and infectious disease control: a public health predicament. *Eur J Public Health* 2012; 22(1): 5-6.

5. Karanikolos M, Mladovsky P, Cylus J, Thomson S, Basu S, Stuckler D, et al. Financial crisis, austerity, and health in Europe. *Lancet* 2013; 381: 1323-1331.

6. National Coalition of STD Directors. STD programs facing severe cutbacks, affecting public health infrastructure: Economic crisis impacting health department capacity to address rising STD rates. 2009. http://www.ncsddc.org/sites/default/files/docs/releases/ncsd_member_survey_and_surveillance_press_release_11.pdf (accessed 20 January 2012).

7. Rechel B, Suhrcke M, Tsolova S, Suk JE, Desai M, McKee M, et al. Economic crisis and communicable disease control in Europe: a scoping study among national experts. *Health Policy* 2011; 103(2-3): 168-75.

8. Baumbach A, Gulis G. Impact of financial crisis on selected health outcomes in Europe. *Eur J. of Public Health* 2014; 24; 399-403.

9. Quaglio G, Karapiperis T, van Woensel L, Arnold E, McDaid D. Austerity and health in Europe. *Health Policy* 2013; 113; 13-19.

10. Bonovas S, Nikolopoulos G. High-burden epidemics in Greece in the era of economic crisis. Early signs of a public health tragedy. *Journal Preventative Medicine and Hygiene* 2012; 53; 169-171.

11. Toffolutti V, Suhrcke M. Assessing the short term health impact of the Great Recession in the European Union: A cross-country panel analysis. *Preventative Medicine* 2014; 64; 54-62.

12. Or Z. Determinants of health outcomes in industrialised countries: A pooled, cross-country, time-series analysis. Organisation for Economic Co-operation and Development (OECD), 2000.

13. World Bank World Development Indicators Online (WDI) database.
http://data.worldbank.org/indicator, World Bank, 2010.

14. FRED (Federal Reserve Bank of St Louis) OECD based Recession Indicators for Euro Area from the Period following the Peak through the Trough (EUROREC) https://fred.stlouisfed.org/series/EUROREC, FRED, 2016.

15. Hastie T, Tibshirani R. Generalized Additive Models. Stat Sci 1986; 1(3): 297-310.

16. Wood, SN. Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC. 2006

17. Imai C, Armstrong B, Chalabi Z, Mangtani P, Hashizume M. Time series regression model for infectious disease and weather. *Environ Res* 2015; 142:319–327.

18. Colón-González FJ, Fezzi C, Lake IR, Hunter PR. The Effects of Weather and Climate Change on Dengue. *PLoS Neglected Tropical Diseases* 2013;7(11):e2503.

19. Bhaskaran K, Gasparrini A, Hajat S, Smeeth L, Armstrong B. Time series regression studies in environmental epidemiology. *International Journal of Epidemiology* 2013;42(4):1187-1195.

20. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2010.

21. Akaike H. Akaike's information criterion. In International Encyclopedia of Statistical Science, pp 25, Berlin Heidelberg: Springer; 2011.

22. Gerdtham UG, Ruhm CJ. Deaths rise in good economic times: Evidence from OECD. *Economics and Human Biology* 2006; 4; 298-316.

23. Stuckler D, Meissner CM, King LP. Can a bank crisis break your heart? *Globalization and Health* 2008; 4.

24. Nikolopoulos GK, Fotiou A, Kanavou E, Richardson C, Detsis M, Pharris A, Suk JE, Semenza JC, Costa-Storti C, Paraskevis D and Sypsa V. National income inequality and declining GDP growth rates are associated with increases in HIV diagnoses among people who inject drugs in Europe: a panel data analysis. *PLoS One* 2015; 10; 4; e0122367.

25. Odone A, Signorelli C, Rodrigues LC. Tuberculosis and the economic crisis: an old threat for the new European agenda. *Scandanavian Journal of Public Health* 2014; 42; 834-835.

26. Lee J. The robustness of Okun's law: Evidence from OECD countries. *Journal of Macroeconomics* 2000; 22(2): 331-356.

27. Wolfinger NH. Promiscuous America: Smart, Secular, and Somewhat Less Happy. 2018. https://ifstudies.org/blog/promiscuous-america-smart-secular-and-somewhat-less-happy..

28. Martí-Pastor M, De Olalla PG, Barberá MJ, Manzardo C, Ocaña I, Knobel H, Gurguí M, Humet V, Vall M, Ribera E, Villar J. Epidemiology of infections by HIV, syphilis, gonorrhea and lymphogranuloma venereum in Barcelona City: a population-based incidence study. *BMC Public Health*. 2015 Dec;15(1):1015.

29. FitzSimons D, Hendrickx G, Vorsters A, Van Damme P. Hepatitis A and E: Update on prevention and epidemiology. *Vaccine* 2010; 28(3): 583-8.

30. Mogilevsky R. The impact of the global financial crisis on public health expenditures in the economies of the former Soviet Union: Center for Social and Economic Research, 2011.

31. Cottrell S, Roberts RJ. Measles outbreak in Europe. *BMJ* 2011; 342: d3724

32. Muscat M. Who Gets Measles in Europe? *J Infect Dis* 2011; 204(suppl 1): S353-S65.

33. Lim TA, Marinova L, Kojouharova M, Tsolova S, Semenza JC. Measles outbreak in Bulgaria: poor maternal educational attainment as a risk factor for medical complications. *Eur J Public Health* 2013; 23(4): 663-9.

34. European Centre for Disease Prevention and Control. Epidemiological update: Measles - monitoring European outbreaks, 8 September 2017. Epidemiological Update 2017; https://ecdc.europa.eu/en/news-events/epidemiological-update-measles-monitoring-european-outbreaks-8-september-2017.

35. Fournet N, Mollema L, Ruijs WL, Harmsen IA, Keck F, Durand JY, Cunha MP, Wamsiedel M, Reis R, French J, Smit EG. Under-vaccinated groups in Europe and their beliefs, attitudes and reasons for non-vaccination; two systematic reviews. *BMC Public Health* 2018 Dec;18(1):196.

36. Mollema L, Wijers N, Hahné SJ, van der Klis FR, Boshuizen HC, de Melker HE. Participation in and attitude towards the national immunization program in the Netherlands: data from population-based questionnaires. *BMC Public Health*. 2012 Dec;12(1):57.

37. Anonymous. Steiner Schools: An Alternative Education System. http://www.independentschoolparent.com/school/steiner-schools-an-alternative-educational-system/. Accessed 10 Sept 2018.

38. Matthews D. Europe's Forgotten Minority. *Times Higher Education*. 3 Jun 2016. https://www.insidehighered.com/news/2016/06/03/only-tiny-minority-roma-enter-higher-education-europe.

39. Brown, M. Inside the private world of London's ultra-Orthodox Jews. *The Daily Telegraph*.

5 Feb 2011.

40. Lee K, Shields KK. Decision-making in hard times: What is a recession, why do we care and how do we know when we are in one? *N Am Journal Econ Finance* 2011; 22(1): 43-60.

41. Chang SS, Gunnell D, Sterne JAC, Lu TH, Cheng ATA. Was the economic crisis 1997–1998 responsible for rising suicide rates in East/Southeast Asia? A time–trend analysis for Japan, Hong Kong, South Korea, Taiwan, Singapore and Thailand. *Social Science and Medicine* 2009; 68(7): 1322-1331.

42. Regidor E, Barrio G, Bravo MJ, de la Fuente L. Has health in Spain been declining since the economic crisis? *Journal of Epidemiology and Community Health* 2014; 68: 280-282.

43. Funk S, Bogich TL, Jones KE, Kilpatrick AM, Daszak P. Quantifying Trends in Disease Impact to Produce a Consistent and Reproducible Definition of an Emerging Infectious Disease. *PloS One* 2013; 8(8): e69951.

44. Achen CH. Why lagged dependent variables can suppress the explanatory power of other independent variables. Presented at the Annual Meeting of the Political Methodology, Los Angeles. 2000 http://www.polmeth.wustl.edu/media/Paper/achen00.pdf

45. Keele L, Kelly N J. Dynamic models for dynamic theories: The ins and outs of lagged dependent variables. *Political Analysis* 2006; 14(2): 186-205.