

Università degli Studi di Pisa Dipartimento di Statistica e Matematica Applicata all'Economia

Report n. 230

The pre-vaccination regional landscape of measles in Italy: contact patterns and related amount of needed eradication efforts (and the "EURO" conjecture)

Piero Manfredi, J.R.Williams, E.M.Cleur, S. Salmaso, M. Ciofi

Pisa, settembre 2002

- Stampato in Proprio -

The pre-vaccination regional landscape of measles in Italy: contact patterns and related amount of needed eradication efforts (and the "EURO" conjecture)

Piero Manfredi¹, J.R. Williams², E.M. Cleur¹, S. Salmaso³, M. Ciofi³

¹Dipartimento di Statistica e Matematica Applicata all'Economia, Università di Pisa

Via Ridolfi 10, 56124 Pisa – ITALY;

²Department of Infectious Disease Epidemiology, Division of Primary Care & Population Health Sciences; Faculty of Medicine.

Imperial College of Science, Technology and Medicine

Norfolk Place, London W2 1PG, U.K.

³Laboratorio di Epidemiologia e Biostatistica, Istituto Superiore di Sanità

Viale Regina Elena 261, 00198 Roma

Summary

The pre-vaccination "landscape" of measles in Italy is summarised using standard epidemiological parameters from the basic SEIR age structured model for childhood infectious diseases (forces of infection, WAIFW matrices, basic reproduction rates and corresponding critical vaccination coverages) plus the spectra of incidence profiles. The results suggest clear priorities in the intervention strategy in view of the WHO target for measles. Moreover they offer further insight on the "EURO" conjecture (Edmunds et al. 2001) concerning the structure of forces of infection in Europe, by suggesting that this latter is potentially more complex than suggested by Edmunds et al.

Keywords. Eradication of measles, force of infection, contact patterns, critical vaccination coverages, EURO conjecture.

1. Introduction

This paper deals with the pre-vaccination epidemiology of measles in Italy taking inspiration from two distinct standpoints. The first one is a very practical one, related to the WHO target of elimination of indigenous measles in Europe by 2007. This target represents an historical challenge for Public Health systems. The steps needed to be successfull in this battle are clearly summarised in Gay (2000). Compared to other European countries Italy is still quite far from these achievements because vaccination against measles, being only recommended since the onset of the national immunization programme (1976), has traditionally been characterised by disappointingly low coverage levels, with a national average of only 56% still in 1998, and by a strong heterogeneity in coverages at the regional (and often also within each region) level. As in Italy the offer of vaccination is not planned at the national level, but autonomously carried out by each Region, this means that Regions will play a principal role in the battle for eradication. This makes it necessary to dispose of a clear picture of the regional geography (especially in that it has been often claimed that spatial heterogeneity in transmission rates could exist in view of the large socio-economical differences existing in the Country (Santoro et al. 1984)) of measles in Italy that could be used as a preliminary step for ranking intervention priorities.

In this paper we provide this picture by relying on the fundamental parameters (forces of infection, contact matrices, reproduction rates) of the basic SEIR mathematical model of vaccine preventable diseases. As primarily shown by Anderson & May (1983,1985,1991), such parameters allow, provided it it possible to estimate them from high quality pre-vaccination data, to sharply summarise the natural history (e.g in absence of vaccination) of a given infection in a given country. An example of such a use of mathematical models is the recent paper by Edmunds et al. (2001), who offer a nice summary of the pre-vaccination epidemiology of MMR in the European countries belonging to the ESEN network (QUOTE). Their analysis shows that the pre-vaccination patterns of measles (and mumps) in the different countries considered were broadly similar, suggesting that it may be possible to use parameter values estimated from other countries (e.g. contries with goodquality pre-vaccination data) to model measles (and mumps) in countries with scarce or poor infection data. The case of Italy was however more puzzling. On the one hand they found that the force of infection (FOI since now on) for measles (but also for mumps) computed from Italian national case notifications data differed to such an extent from the rest of available European FOI's, that they propose to dismiss Italian case reports, conjecturing that they suffer of strong selective under-reporting by age. On the other hand they computed the FOI from the elaboration reported by

¹ These are: i) the achievement, as soon as possible, of high levels of routine coverage (>90-95%), ii) their continuation over time during the whole time span until eradication; iii) the use of pulse and other forms of catch-up vaccination of susceptibles as further resources; iv) the improvement of the measles sentinel system.

Santoro et al. (1984) – the unique pre-vaccination measles sero-survey in Italy - finding values of the FOI that, at least for the youngest age groups (0-4), were in agreement with "EURO" values (summarising infection experience in European countries with good infection data, Edmunds et al. 2001). They therefore concluded: "It is tempting to dismiss FOI estimates from Italian case notifications data as the serological data are likely to be more robust".

We believe, however, that this latter conclusion relies more on the generally assumed greater reliability of serological data compared to case reports, rather than on a full demonstration. Indeed, if one looks at the value of the same FOI they estimated from Italian serological data for school children (age 5-10), one finds a value which is just 50% of the corresponding EURO value, and moreover, much smaller compared to the value estimated from case reports. We believe therefore that a prudential attitude should try to obtain further insight on the problem.³

The paper by Edmunds et al. (2001) represents therefore the second source of inspiration for our present work. Our second question here is more theoretical: do we need to assume that the Italian FOI (e.g. the true FOI acting in Italy during the pre-vaccination era) is very likely homologous to the EURO FOI? Or could the two FOI's broadly differ? The implications are relevant especially for modelling purposes: is it advisable to just rely on EURO, or one should consider other possibilities as well? In what follows we will speak about the "EURO conjecture" to denote Edmunds et al. statement that i) the Italian FOI would essentially be homologous to the EURO FOI, ii) Italian case notifications data just camouflage this fact thanks to a broad selective under-reporting.

Since unfortunately we could not access the original serological data, in order to shed more light on both questions raised in this paper, we have more deeply analysed Italian case notifications data, by also looking at patterns at several spatial levels. First we have systematically looked at the structure of the forces of infection (and contact patterns, reproduction rates, etc) for all the Italian Regions and the major Conurbations and Cities. Our feeling was that, since the national datum that Edmunds et al. (2001) used to compute the national FOI, was obtained by pooling the regional data in presence of a strong heterogeneity in spatial under-reporting (as well documented in Williams et al. 2002), a deeper investigation of spatial patterns of infection could reveal the existence of some between regions heterogeneity in age—related transmission rates that could be the spy of a higher force of infection (or in case suggest the presence of selective under reporting by age). Second we have relied on our estimates (Williams et al. 2002) of the overall rates of under-reporting at the regional level in order to provide a corrected national age distribution of cases in order to obtain a corrected estimate of the FOI. Indeed the concurrence of spatial heterogeneity in the overall rates of under-reporting with spatial heterogeneity in age—related transmission rates may be responsible, at

² In Italy vaccination against tetanus, polio, HBV and diphteria are mandatory: all other vaccinations are only recommended.

least theoretically, for selective age-bias in the national age distribution even in absence of selective under-reporting by age at the spatial level. Finally we provided a systematic investigation, by means of time series techniques, of the structure of regional periodicities of measles during the prevaccination era. Our feeling here is that only under exceptional circumstances can selective under reporting by age mask true time patterns, so that again we hoped to find some confirmation in favour or against the EURO conjecture.

The present paper is organised as follows. Section 2 deals methods and data. Section 3 outlines our main results on the pre-vaccination landscape of measles in Italy. The consequences of our results for control of measles in Italy and the EURO conjecture are summarised in the discussion.

2. Materials and methods

The geographic analysis of vaccine preventable diseases is a hard task due to the complex correlations between the various local dynamics. A first useful, though clearly partial, step is the characterization of the forces of infection acting "locally", by treating spatial sub-areas as autonomous epidemiological units (the final discussion also deals with the pittfals intrinsic to such an approach). In order to characterising the pre-vaccination "regional landscape" of measles in Italy we systematically used two types of standard tools:

- 1) summary parameters from the basic SEIR age structured model for vaccine preventable disease (e.g. forces of infection, mixing matrices, reproduction ratios, and related critical coverages, see Anderson & May 1991). In order to ensure comparability with Edmunds et al. (2001) we used the same age groups for the computation of the FOI, e.g. 0-1 years, 2-4 y., 5-10 y., 11-17 y., 18+ y. As regards WAIFW matrices, we have considered standard matrices, e.g. Default, Diagonal, and Proportionate used by Edmunds et al., and also included some Preferred Mixing matrices (Hethcote 1995). The forms of mixing structures adopted in the paper are summarised in the next subsection. An important point is that in many cases neither the Default type, nor the Diagonal one, appears to be suited to represent Italian mixing patterns, as emerging from available data. Basic reproduction ratios (Diekmann et al. 1990), and the related critical vaccination coverages for each region were also computed from each type of "acceptable" mixing matrix.
- characterisation of periodicities of time series of measles incidence (for each Italian region) by means of autocorrelation functions and frequency spectra (firstly used, for characterising oscillatory patterns of childhood diseases by Anderson et al. (1984), (1

³ A further problem of Italian serological data is the lack of any data beyond age 12.

2000), suggesting that, as long as we are concerned with the pre-vaccination era, regional overall under-reporting rates seemed to have remained fairly constant over time, so that the available time series should nonetheless represent a sufficiently reliable picture of the regional dynamics of measles. Standard techniques of data preparation (log transformation of the data in order to stabilize variances; detrending by regressions etc) were used. Periodic components were identified by use of the Fisher's test on the periodogram and spectral analysis. A non-parametric estimate of the spectral density function with the Parzen window was calculated. Various truncation points were used in order to more clearly identify the periodic components.

Analyses 1) and 2) have been carried out for all Italian regions, and for the largest conurbations and cities (with the exception of time series analysis of city data, which were not available on a sufficiently large window of time). Valle d'Aosta, a very small Northern Region with very few reported cases per year was aggregated with the neighbouring Piemonte

Mixing (WAIFW) matrices used in the paper

Mixing matrices (often called WAIFW=Who Acquires the Infection From Whom, Anderson & May 1991) are used, in standard epidemiological modelling, to represent age patterns of contacts between susceptible and infected individuals. Their generic element β_{ij} summarises the risk of acquiring infection for a susceptible individual in age group "i" due to contacts with infective individuals aged "j". As known from basic theory (Anderson & May 1991), a mixing matrix with m age groups can have at most m distinct entries. In the simplest type of mixing, namely homogeneous mixing, the entries of the matrix are independent of age ($\beta_{ij}=\beta$ per ogni i,j). Other types of mixing matrices considered in this paper are listed below (it is useful for the reader to have in mind the adopted age classification: 0-1 y., 2-4,5-10,11-17,18+).

- 2. Fully assortative mixing (matrix DIAG in table 1). This is a boundary case in which individuals of a given age are assumed to mix only with individuals of the same age. The corresponding mixing matrix is diagonal, e.g. it has nonzero entries only on the main diagonal. It is the form of mixing allowing R_0 to achieve, the form of the FOI been given, its upper bound (Greenhalgh and Dietz 1994).
- 3. Realistic assortative mixing (RDIAG). Such a matrix (termed "Diagonal" in Edmunds et al. 2001, Wallinga et al. 2001 use a slightly more flexible form that they term "semi-assortative) has the same diagonal elements as the fully assortative matrix (this ensures preservation of a high degree of assortativeness), but the age groups are opened (non diagonal elements are strictly positive). In the

form actually used by Edmunds et al. (2001) it is assumed that individuals from distinct age groups interact between them at the same rate typical of intra-adult contacts. This choice is actually constrained by the estimated form of the force of infection, as pointed out later on.

- 4. "Default" mixing (matrix DEF in table 1): this type of mixing (discussed in detail in Edmunds et al., 2001) emphasises transmission between school age groups.
- 5. Proportionate mixing (PM). Under PM (Hethcote 1995) contacts occur at random. This implies that in presence of heterogeneity in contact rates of individuals, one will always face a larger probability of meeting more socially active individuals. Under PM the entries of the mixing matrix have the "multiplicative" form $\beta_{ij} = b_i b_j$ j = 1,...m.
- 6. One-parameter preferred mixing (PREF). In this case (Hethcote 1995) the mixing matrix is defined as a one parameter average between the case of proportionate mixing and fully assortative mixing: PREF = (1-h)*DIAG + h*PM where 0 < h < 1. The underlying idea is that everyone's contacts can be split into two components: a selective (e.g. non random) one, basically with individuals from the same age group, and a fully random one. The PREF matrix contains (m+1) distinct entries: the extra parameter h is therefore usually estimated on a "ad hoc" basis or let free to vary in a grid (plausibility considerations may be used to restrict the bounds of acceptable values).

It is to be noticed that matrices 3,4 are defined on a "ad hoc" manner, though they have a behavioural basis, whereas type 1,2,5 are (though in distinct senses) boundary cases. Finally type 6 is an average between two boundary cases. This implies that there is no clear relation between for instance types 3,4 on the one hand and type 6 on the other (though in some cases one can find, by a suitable choice of h, preferred matrices which are quite close to "ad hoc" forms). For this reason in this work we used all forms reported here.

Table. 1. Some types of mixing matrices used in the paper

$$DIAG = \begin{pmatrix} \beta_1 & 0 & 0 & 0 & 0 \\ 0 & \beta_2 & 0 & 0 & 0 \\ 0 & 0 & \beta_3 & 0 & 0 \\ 0 & 0 & 0 & \beta_4 & 0 \\ 0 & 0 & 0 & 0 & \beta_5 \end{pmatrix} ; RDIAG = \begin{pmatrix} \beta_1 & \beta_5 & \beta_5 & \beta_5 & \beta_5 \\ \beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_5 \\ \beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_5 \\ \beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_5 \end{pmatrix} ; DEF = \begin{pmatrix} \beta_1 & \beta_1 & \beta_1 & \beta_5 \\ \beta_1 & \beta_2 & \beta_4 & \beta_4 & \beta_5 \\ \beta_1 & \beta_4 & \beta_3 & \beta_5 & \beta_5 \\ \beta_5 & \beta_5 & \beta_5 & \beta_5 & \beta_5 \end{pmatrix}$$

Sources of data and data preparation

Age structured measles case notifications data were provided by the Istituto Nazionale di Statistica (ISTAT) for the period starting from the first available year, e.g 1971. Monthly measles case notification data were provided by ISTAT for the period from the first available year, 1949, to 1996. Age data for 1975 were unavailable at the regional level and therefore we reconstructed them by interpolation using the few published data. We expect however that this should not cause any serious errore as in all regions the age distribution of cases were quite stable in the pre-vaccination period. Pre-vaccination FOI's were estimated from data on the time window 1971-76. Such 6-years long pre-vaccination window was considered adequate in that about two full three years long epidemic cycles were so considered. We also tried to include in the computation of the FOI's the first few years in the post-vaccination window, when vaccination coverages were known to be extremely small, without getting any significant change in the estimates of the FOI. Previously published data on annual births in Italy were also used.

3. Results: the pre-vaccination landscape of measles in Italy

The structure of the force of infection at the regional level

As a first step we have estimated pre-vaccination forces of infection for all Italian regions and "classical" large clusters (North, Centre, South), and some of the largest conurbations and cities. The age groups considerd were 0-1 y.,2-4y.,5-10y.,11-17y.,18+). We also computed the "corrected" force of infection (denoted as Italy UR in table 2) by removing the bias due to spatial underreporting using the spatial under-reporting rates computed in Williams & et. (2002). We always add for reference the EURO FOI estimated by Edmunds et al. (2001). The results (we only report a sample of them; for details the reader may consult http://www.manfredi.statmat.unipi.it) are as follows:

- a) levels of the FOI's observed in all Italian regions are always systematically lower (in most cases significantly lower) of the "EURO" levels (fig. 1), the unique exception being represented by the very young age groups in Southern Italy.
- b) For what concerns more in detail the geographic patterns of the FOI, some heterogeneity exists, essentially explained by the three classic geographic clusters North, Centre and South (still see fig. 1). For what concerns the two youngest age groups the FOI is substantially higher in the Southern regions compared to the North, and especially compared to the Centre, where the FOI appears to be extremely low. In particular the highest value of the FOI are observed, for the involved age groups, in some of the major Southern cities, where the values are not so different from those reported for EURO and Denmark in the cited study by Edmunds et al. (2001), and broadly agree with the findings of the unique Italian serological study (Santoro et al. 1984). In

the "elementary school" age group (5-10 y.) the Central regions continue to have very small values, compared to the Northern and Central ones, which show quite similar levels. Finally in the two highest age groups there is a sort of recuperation effects: Central and Northern regions have FOI values systematically higher compared to the Southern ones. The average age at infection, which coarsely summarises the FOI, shows similar spatial patterns (this holds for both the overall average age of cases (A), and the restricted average (A_{I5}) over the 0-14 y. age group. For instance A_{I5} is systematically smaller in the South (little above 5 years), is around 7 years in the central regions, and it takes intermediate values in the Northern regions.

c) For what concerns more generally the qualitative shape of the FOI's observed in the Italian regions, these are broadly consistent with the EURO shape (one humped, peaking in the "elementary school" age group 5-10, and with its minimum on the highest age group) in all but one⁴ important point. In fact the aforementioned "recuperation" effect (which is not by itself a necessarily pathological fact, e.g. due to bad reporting) has the consequence that in the Italian regions the force of infection in the highest age group is, with the only exception of the Southerm regions, always higher compared to the FOI in the youngest age group. Provided one can exclude the effects of bad data reporting, this fact (which occurs regardless of how we define the last age group) is potentially suggesting a greater relative importance of contacts between adults, and has interesting effects on our subsequent development.

Fig. 1. Force of infection of measles in Northern, Central, Southern Italy vs EURO FOI

Correcting the force of infection for the effects of overall under-reporting

We noticed in the introduction that the concurrence of spatial heterogeneity in the overall rates of under-reporting with spatial heterogeneity in age—related transmission rates may be responsible, at least theoretically, for selective age-bias in the national age distribution even in absence of selective under-reporting by age. Indeed widely different overall under-reporting rates imply that every regional heterogeneity in epidemic patterns is, when pooling the regional data in the national distribution, weighted by under-reporting. Fig. 1 shows that, however, the extent of the correction is of some relevance only for the first two age groups, and rather small for the others. Though this correction is important for modelling purposes we will not deal with it anymore in this paper as it does not seem to give any further insight on the major questions tackled here.

⁴ Actually there was a second fact worth of consideration, e.g. the systematic presence of a peak around age 20 in all regions. This fact is probably due a very high reporting rate in the male subpopulation at those ages, perhaps due to the military service, compulsory in Italy (we say probably because we could not provide a rigorous argument against alternative assumptions, eg. true measles epidemics in barracks). In these cases we smoothed the age distribution of cases or adopted the sole (much smoother) females profiles, systematically differing from the male ones only in the involved age groups.

Mixing matrices, reprodution rations and amount of effort for measles elimination

From the estimated forces of infection we computed mixing matrices, according to the forms reported in section 2.3, and the related values of the reproduction ratio R_0 (Diekmann et al. 1990), with the corresponding critical vaccination coverage for routine vaccination at birth (given by the known formula $p_c=1-1/R_0$).

Now, it is known that not all forms of mixing will be compatible with a prescribed form of the force of infection. In some cases a mixing matrix will be non feasible for the estimated FOI e.g. it may yield negative values of some of the β coefficients. For such a circumstance Anderson & May (1991) noticed that "the chosen matrix is unappropriate to the observed age dependence in the FOI". H. Hethcote (1995) tried to go a step further by pointing out that not all feasible mixing matrices are "acceptable", in that they could lead to values of the number of contacts between individuals which are outside bounds of practical plausibility. For this reason a careful inspection of the results of mixing computation is always necessary, in order to avoid trivial results. He also proposed a criterion for assessing plausibility, based on a sort of postulate of "preference for assortativeness", which is of simple use for preferred matrices (otherwise, if one is not sure about preference for assortativeness in all age groups, things may be more complex).

Now, a first important problem in dealing with Italian data is that, as previously pointed out, the FOI experienced by adult individuals (e.g. the oldest age groups) is always (with the exception of Southern regions) higher compared to the FOI of the "very young" (0-1 y.).

An immediate, and easily verifiable, consequence of this fact, is that mixing matrices of the type termed "realistic assortative" will never be admissible (they will always have negative coefficients). In other words: such type of mixing patterns is not compatible, apart Southern Regions, with the estimated forms of the force of infection. This is a quite relevant problem because "realistic assortative" mixing is the type of mixing promoting the largest "reasonable" values of the reproduction ratio compatible with an observed EURO-type FOI (Wallinga et al. 2002). The problem could be ruled out in some cases by exchanging the relative role of the youngest and oldest age groups (e.g. by using β_I rather than β_5 as the inter-groups contact rate). This would mean to put more relevance on adult transmission. This latter choice would have, however, a significant consequence: it implies to accept the idea that the Italian forces of infection could differ from the Euro one estimated from reliable data not only in the "levels", but also in the qualitative shape of the FOI profile over ages. For this reason we have chosen a "prudential" starting point, essentially with the purpose to preserve comparability with Edmunds et al. (2001), and have, rather arbitrarily, redefined the value of the force of infection in the oldest age group in order to have, for the FOI's of all the Italian regions (and not only those in the South) the "EURO" shape (e.g. a FOI having its

minimum value over the adult age group). In this manner it was possible to obtain admissible "realistic assortative" mixing matrices for all Italian Regions. The more interesting results are summarised below; we omit for sake of brevity the tedious outputs of mixing computations and rather report the more easily interpretable values of reproduction ratios and critical coverages; results from "fully assortative mixing" leading to trivially high values of R_0 are also omitted. Table 2 reports a full synoptic view of our results.

- 1. Homogeneous mixing. Despite its limited realism, this assumption often represents an excellent starting point for more refined analyses. Values of R_0 (which can be evaluated via the simplified formula $R_0 \cong L/A$ where L is the expectation of life in the population-here taken as 75 years by assuming type 1 mortality, and ignoring regional variability and A the average age at infection in the pre-vaccination era) range between a minimum of about 10 (Toscana) up to a maximum around 13 in Puglia.
- 2. Realistic assortative mixing. Such mixing yields the largest "reasonable" R_0 (29 for the EURO FOI) values compatible with "EURO"-shaped forces of infection. In absence of our previously made assumption on the ultimate age group only Southern Regions yielded to fully admissible mixing rates, with R_0 ranging between 10 and 15, and related critical coverages ranging 92-96%. Our assumption allows admissible solutions for all Italian Regions (regions for which realistic assortative mixing was "recuperated" thanks to our assumption are denoted by an asterisk in tab. 2).
- 3. Default. This type of mixing happens to be never admissible for all Central Regions and some Northern ones. Indeed it is again easily verifiable that Default mixing yields negative transmission rates when the the number of cases in the school age groups is still rather high, compared to the younger age groups (e.g. when the FOI is rather low on the youngest age groups). This occurs for the Central Italy regions because there the FOI at very young ages is quite low, leaving therefore a lot of infection for higher ages. Default mixing is on the contrary admissible in all Southern regions. R_0 and critical coverages happens to be systematically higher in the South. Critical coverages range from 80% in the North up to 88% in the South (reference EURO value being 90%).
- 4. Proporzionate mixing. PM mixing matrices are always admissible by definition. As known (Hethcote 1995, Wallinga et al. 2001) this type of mixing leads to the "practically reasonable" lower bound of R_0 and critical coverages under. Indeed these latter ranged from 74% in the Centre, with intermediate values in the North, up to 84% in the South (EURO=86%). An inspection of the matrices (see Hethcote 1995) reveals for essentially all regions (EURO and Denmark as well) the presence of implausible relations between groups in terms of number of contacts, e.g. a significant level of anti-assortativeness.

5. Preferred mixing. We computed preferred matrices for various values of the "preference for assortativeness" parameter h starting from h=1 (proporzionate mixing), and then, according to Hethcote (1995), progressively decreasing h until achievement of "minimally plausibile" contact rates. In this manner it should be possible to establish more reliable lower bounds on R_0 , compared to the baseline case of Proportionate mixing. For essentially all Italian regions (EURO and Denmark as well) the "pumping" of a rather limited degree (h=0.9) of assortativeness made possible to pass the threshold of "minimal plausibility". This led, compared to the baseline case of proportionate mixing, to increases in the corresponding critical coverages ranging from 1% to 3%. Of course the given values are still to be considered as just refinements of the lower bound.

Which could be the degree of assortativeness h leading to the mixing pattern closest to reality is an open question, as what we have found is just the degree guaranteeing a "minimally plausible" pattern, according to Hethcote's criterion. Certainly one-parameter preferred mixing is a flexible family of mixing patterns, but still too unflexible for real purposes, as it assumes the same degree of preference for assortativeness in all age groups. For istance (as far as we look at schooling as the major source of assortativeness in the young age groups) in the case of pre-vaccination Italy, since boys in the youngest age group were attending cresh schools in a minimal proportion (especially compared to present days), their major source of assortativeness was probably removed. This would therefore suggest that at least two "preference for assortativeness" parameters (h_1, h_2) say) would be neded to describe Italian pre-vaccination infection patterns. This leads to the point of the real gain provided by preferred matrices (just because they are "many") compared to "ad hoc" behavioural matrices. One could ask whether Hethcote's preference for assortativeness criterion is meaningful for mixing matrices which do not conform to the one-parameter "preferred mixing" scheme. Using the criterion of "preference for assortativeness" one would indeed discard, in most cases, "default" mixing. Indeed, under default mixing (apart the too small assortativeness postulated for the youngest age group), most Italian regions (but also Denmark) do not meet the criterion for age groups 2-4 and 5-10. But discarding "default" is of course rash, since default is from other perspectives one of the most "reasonable" among behavioural mixing matrices, which is, moreover, usually characterised by a higher degree, in overall terms, of assortativeness compared to minimally plausible preferred matrices (as showed by higher R_0 values). But, more generally, any answer to such questions is complex, as the general problem of assessing whether a given mixing matrix is better, or simply more plausible, of another mixing matrix, is, as far as we know, rather unsettled. The problem is not an academic one: the predicted critical coverage for measles at age zero under the EURO FOI range between 86% (under proportionate mixing) and 97% under realistic assortative (Edmunds et al. 2001). Using Hethcote criterion one can increase the lower bound up to

87.5% by replacing proportionate mixing (non acceptable under Hethcote criterion) with the corresponding "minimally acceptable" preferred matrix, but the extent of the range still remains significant.

The regional patterns of incidence over time

In order to avoid an overburden of details, we do not report here in detail the results of our time series analysis of incidence data (fully available, with technical details, at www.statmat.unipi.it/cleur); we rather report those summary results at the regional level which appear to be mostrly relelvant for the present paper.

The eye inspection of Italian regional time series of measles incidences suggests much less evident pre-vaccination (1949-76) patterns compared to the sharp 2-years oscillations of England & Wales. The inspection of the frequency spectra of the same data reveals the following features: i) a sharp peak at the annual frequency, a fact well evident from the visual inspection of the data; ii) a second, usually smaller, peak at a lower frequency (and more hardly detectable from visual inspection), denoting a long-term cycle (the only exceptions are Lazio and Puglia, which did not exhibit regular long-term cyclical movements). This longer cycle is characterised by a period varying from essentially 2,5 years in Piemonte and Lombardia, to 4 years in Liguria, Abruzzo, and Sardegna (actually a longer term oscillation with period 2 years is observed in Basilicata, and a 5 years period is observed in Molise, but these are the two smallest regions in the country so that additional caution must be placed on results). It is worth noting that the annual component dominates the long-term cycle in almost all the Italian regions with the ecceptions of Sardegna, Abruzzo, Molise. These results suggest substantial differences compared to the patterns traditionally observed for England & Wales, but they are by no means uncommon (Schenzle 1984, Anderson et al. 1984).

Moreover, a cross-spectral analysis suggests that, whereas the annual components in all the series are very highly correlated (coherencies around 0.9 or higher), there appears to be a much smaller correlation between the long term cycles. In this latter case, coherencies greater than 0.7 were observed, as expected, for a few neighbouring regions such as Emilia-Romagna and Marche, Marche and Umbria, Friuli and Trentino, Lombardia and Piemonte, Liguria and Piemonte.

The inspection of the post-vaccination spectra reveals that, compared to the pre-vaccination period, the role of the longer term oscillation, appears to become more important (less dominance of the seasonal term). Moreover more evident symptoms of increasing synchronization between areas appear. As we are not interested here in the features of post-vaccination dynamics we will not, however, pursue this aspect any further.

The results provided by time series analysis were subsequently grounded against the prediction from the basic homogeneous mixing (e.g. in absence of heterogeneities in contact rates) SEIR

model of vaccine preventable diseases, stating that for diseases in which the sum of the expected lengths of sojour in the latent and infectious states is short compared to the expectation of life of the host (as is most often the case for childhood diseases), the inter-epidemic period of the oscillation around the endemic equilibrium, i.e. the period of the long-term oscillation, is, to an excellent approximation, given by the formula (Anderson and May 1991) $T = 2\pi\sqrt{AK}$, where A is the average age at which infection is acquired, and $K=1/\sigma+1/\nu$ is the sum of the average expected lenghts of the sojours in the latent and infectious states. Tab.3 reports for each region the following informations: i) the average age of cases in the pre-vaccination period (A); ii) the period T_S of the long-term oscillation as estimated from spectral analysis; iii) the period of the long-term oscillation as computed by the formula $T = 2\pi\sqrt{AK}$ (K is taken as 7+7=14 days). With reference to the prevaccination period, tab. 2 shows a substantial agreement between predictions from the SEIR model, and the lengths of the cycles identified by spectral analysis, with a few exceptions, which were Liguria, Basilicata, and Molise (for these two latter regions, however, see the previous caveat; moreover for Lazio and Puglia the absence of an identifiable long term oscillation prevented any comparison). The most important finding is certainly that the period of the long term oscillations computed by time series techniques, appears to be usually around three years (for largest regions) and this is quite well in agreement with the theoretical prediction of the basic SEIR model. This figures are therefore substantially different from the classical 2-years oscillations observed in England & Wales.

4. Discussion

In the present paper an attempt has been made at reconstructing the pre-vaccination landscape of measles in the Italian regions. There are several critical points in analyses as the present one, first of all the fact that, by looking at "local" forces of infection, e.g. treating all spatial units as autonomous epidemiological units, is at best a first step, as it ignores all the complexities related to spatial correlations between the various local dynamics. Unfortunately more refined approaches are not yet common. A second problem is that the present analysis only relies on the available data, e.g. case notifications data. Therefore we are looking at the problem just "one way". Nevertheless several interesting facts emerge.

Amount of eradication effort required at the regional level

The first motivation for this paper was a rather practical one, e.g. the need for a clear picture of the geography of measles in Italy, which could be used for ranking intervention priorities, given the critical role that Regional Public Health Units will have in Italy in the battle for eradication in view of the WHO target of measles elimination by 2007. During the second half of the nineties several

efforts have been undertaken in Italy in order to increase routine coverages, until then disappointingly low. Even if surveillance data suggests that incidence has gone down to its historical minimum during 1999-2001, the national coverage was still below 70% in 2000, but with broad regional heterogeneities. It has a not come as a surprise therefore the big epidemics which has frightened especially Campania (the region with the lowest coverages) during last spring, with 20000 estimated cases between January and May 2002 (Ciofi & Salmaso 2002).

From this standpoint the results of this paper quite sharply suggest that, under a quite broad set of reasonable mixing patterns, the regions which seem to be the more demanding in terms of the amunt of effort needed to eradicate the disease are systematically the Southern ones. This is a rather problematic fact, as Southern regions are those presently characterised by the lowest vaccination coverages (ICONA 1998, Italian Ministery of Health 2001). The absolute amount of the needed eradication effort, as summarised by the critical coverages for routine birth vaccination, is rather variable depending on the chosen type of contact. For instance for Southern regions the critical coverages range from 82%, under "minimally plausible" preferred mixing, up to values around 94-95% under realistic assortative mixing, which represent lower and upper bounds to realistic contact patterns. Intermediate assumptions lead to intermediate values. For instance under Default mixing the critical coverages for the major Southern Regions (Campania, Puglia, Sicilia) range 86-88% which is only two points less than the EURO value of 90% (Edmunds et al. 2001). Keeping in mind that vaccination at birth is just a "theoretical standard", indeed critical coverages quick increase as the age at vaccination is delayed, that we are assuming a 100% effective vaccine, and that, last but not least, the presence of anti-vaccination movements in the Country is rather stable, all these fact denote that a very strong effort is still needed, especially in the Southern part of the Country in order to minimally approach the WHO target.

The EURO conjecture

Our investigations of the pre-vaccination landscape of measles in Italy allowed, as a by product, to add further flesh on a topic which appears to be of critical relevance, e.g. the "EURO" conjecture. The EURO conjecture states that very likely Italian case notifications data suffer a strong bias with age which in turn leads to a gross underestimate of the true force of infection of measles, which, on the contrary should be homologous to the EURO FOI.

Now, certainly, under-reporting is a major problem in Italy. It is widely know (Santoro et al. 1984, Williams et al. 2002) that in the pre-vaccination era the overall reported incidence was an order of magnitude less than true incidence, and that the degree of under-reporting varied widely between regions. The existence of large heterogeneous overall under-reporting rates at the regional level does not necessarily implies, however, a bias in the (national) estimate of the force of infection: this

additionally requires that i) reporting rates exhibit a significant variation with age, or that ii) spatial heterogeneity in overall reporting rates coexist with a marked spatial heterogeneity in infection patterns by age.

We therefore made an extensive onvestigation of forces of infection at several spatial levels (Regions, large Conurbations, large Cities). Our feeling was that by going deeply into the spatial scale it would have been probably possible to see the emergence, in the end, of some footprint (e.g. areas with a "high" FOI, or with significant heterogeneity in FOI's) of the presence of the presumed underlying EURO pattern. Unfortunately we have not been successfull, as documented in section three of the paper. All the "local" forces of infection are systematically significantly smaller compared to the EURO foi, and surprisingly similar in levels and shape: significant heterogeneity exists only in the yougest age groups. This of course raise the point, if the EURO conjecture were true, of which are the social processes and mechanisms, if any, underlying bad age reporting of measles cases, which are capable of so effectively camouflaging the true underlying force of infection. Indeed such processes should, in order to be so stable over space and time, be profoundly radicated in the social attitude of families toward diseases of their children and/or practishioners' behaviour.

We point out that selective under-reporting by age has been documented in Italy (Ciofi et al. 2001) in recent years for varicella, which is still in its pre-vaccination period. Their work, however, shows that reporting rates are rather higher at young ages (0-10 y.) and subsequently tend to decrease. Therefore if it could be reasonable to assume that current under reporting rates for varicella are in some way similar to those for measles during the pre-vaccination period, their result would not carry evidence in favour of a bias leading to an underestimate of the force of infection.

Be that as it may we therefore turned our attention to the investigation of the available time series of measles incidence, comforted by the fact that, as proved in Williams et al. (2002), the reporting rates of measles at the regional level seemed to have been remained fairly constant over time, so that we can reasonably trust that, though unreliable in terms of absolute levels, the available incidence data should be broadly representative of true qualitative patterns. Even in this case we could not find, however, any serious footprint of an underlying "higher" force of infection; rather the results appear to be somewhat consistent with the Italian age patterns observed from case reports. Indeed the observed long-term oscillation of measles between Italy and England & Wales shows (apart other differences) is, for most Italian regions, around three years or more. Such values, which are sharply different from the two years oscillation stably observed in the pre-vaccination incidence series in England & Wales, are therefore consistent, compared to England & Wales, with a much higher value of the average age at infection, and more generally, of the force of infection. Though one must be cautios about these facts, it is also clear, however, that we cannot quickly

dismiss them as the consequence of the presumed existence of age biases in reporting rates: it would mean that patterns of under-reporting are capable of also hiding true time patterns. Now, though this latter possibility can not be excluded "a priori" (one possibility could be that families/practishioners have distinct reporting rates depending on the phase of the epidemic cycle, e.g. an epidemic year vs a non-epidemic year), it appears rather unlikely in most circumstances. Indeed, by respectively denoting as $C_i^*(t)$, C(t) the reported and the true number of cases at time t in area i, and by $\alpha(a,t)$ the reporting rate of cases aged a, it holds $C_i^*(t)$, C(t): $C_i^*(t) = \alpha_i(t)C(t)$, where $\alpha_i(t) = \int_0^\infty \alpha_i(a,t) \frac{C_i(a,t)}{C_i(a,t)} da$ is the

overall reporting rate at time t, which is a weighted average of the age-specific reporting rates. This suggests that, if the percentage age distribution of cases is stationary over time, as we would expect in the pre-vaccination era (a fact that Italian data strongly suggest), then significant (and periodic) time changes in age-specific reporting rates would be needed to effectively mask true incidence patterns.

Future work and conclusive remarks

Which was the true force of infection acting in Italy during the pre-vaccination era ? The high EURO force of infection? Or the low one emerging from case notifications data? Or something in between, e.g. the FOI's estimated from case reports are just lower bounds for the true forces of infection? We feel that reasonably the latter possibility is the most reasonable one. Therefore, from a public health perspective - e.g. a quite prudential attitude - the fact to know that the computed eradication coverages could just represent lower bounds of the true values, which could be much higher, seems to be a sufficient argument for not deviating from the target coverages suggested by WHO. From a theoretical perspective things are different, however. First because for modelling purposes it does make a lot of difference to have huge differences in a fundamental parameter, as the force of infection. An important consequence is that, for instance, certain types of mixing matrices which are considered strongly appropriate, will not be anymore valid. For example "Default" mixing which is considered to be strongly appropriate contact pattern for childhood diseases in the Western world is non admissible for "low" FOI's, as in some cases observed in Italian data (indeed at a more careful glance Default mixing reveals to be appropriate for forces of infection consuming a rather high fraction of susceptible before school ages, as for EURO, but may not be appropriate for "low" forces of infection). One must therefore resort to different types of mixing, probably leading to a very large different in predictions.

Second because this suggests the stimulating possibility that patterns of infection by measles through Europe rather than being essentially homologous, as implicit in the EURO conjecture,

could be (and have been) largely different (as by the way already suggested for rubella by Edmunds et al. 2001). This is a further evidence of the risk of analogical approaches.

As future work on Italy we aim to provide a similar analysis for varicella, which is still in its prevaccination era, and mumps, for which the coexistence of i) a vaccine of low efficacy with, ii) very low routine coverages, should have caused rather little perturbation to pre-vaccination conditions, so partly allowing the use of ESEN serological data.

Acknowledgements. This paper was partly funded by Ministero della Sanità, progetto 1%. We thank Gianni Gallo, Giampaolo Scalia Tomba for useful discussions.

References

A.A.V.V., ICONA: indagine nazionale sulla copertura vaccinale infantile, Rapporto ISTISAN 98/33, 1998.

Anderson R.M., Directly transmitted viral and bacterial infections of man, in Anderson R.M. (ed.), Population biology of infectious diseases, Chapman and Hall, London, 1-39, 1982.

Anderson R.M., May R.M., Vaccination against rubella and measles: quantitative investigation of different policies, J. Hyg. Cambridge, 1983; 90: 259-325.

Anderson R.M., Grenfell B.T., May R.M.(1984), Oscillatory fluctuations in the incidence of infectious diseases and the impact of vaccination: time series analysis, *Journal of Hygiene Cambridge*, 93, 587-698.

Anderson R.M., May R.M., Age-related changes in the rate of disease transmission: implications for the design of vaccination programmes, *J. Hyg. Cambridge*, 1985; 94: 365-436.

Anderson R.M., May R.M.(1991), Infectious diseases of humans: dynamics and control, Oxford University Press.

Gomes M.C., Gomes J.J., Paulo A.C. (2000), Diphteria, pertussis, and measles in Portugal before and after mass vaccination: a time series analysis, preprint.

Benenson A.S., Controlo of Communicable Diseases Manual, American Public Health Ass. Editions, Washington, 1995.

Ciofi degli Atti ML, Salmaso S., Measles increase in Italy, *Eurosurv. weekly*, 2002, 6: 14. Dati disponibili su <u>www.spes.iss.it</u>.

Ciofi degli Atti ML, M. C. Rota, D. Mandolini, et al., Assessment of varicella underreporting in Italy. *Epid. & Infection*, in corso di pubblicazione.

Cleur E.M., Manfredi P., Williams J.R. (2002), The pre- and post-vaccination regional dynamics of measles in Italy: insights from time series analysis, lavoro presentatato al workshop "Modelli statistici per le serie temporali", L'Aquila, 5-6/4/2001.

Diekmann O. Heesterbeek H., Metz J.A.J., On the definition and computation of the basic reproduction ratio, J. Math. Biol., 1990; 28,365-382.

Diekmann O., Heesterbeek H., Mathematical epidemiology of infectious diseases, Springer Verlag, 1999.

DPR 23 luglio 1998: Piano Sanitario Nazionale 1998-2000. Un patto di solidarietà per la salute. G.U. del 10 dicembre 1998.

Edmunds W.J., Van de Heijden O.G., Eerola M., Gay N.J., Modelling rubella in Europe, *Epid. & Infection*, 2000; 125: 617-634.

Edmunds W.J., Gay N.J., Kretzchmar M., Pebody R.G., Wachmann H., The pre-vaccination epidemiology of measles, mumps and rubella in Europe: Implications for modelling studies, *Epid. & Infection*, 2001; 125: 635-650.

Farrington C.P., Modelling force of infection for measles mumps and rubella, Stat. in Medicine, 1990; 9: 953-967.

Farrington C.P., Kanaan M.N., Gay N.J., Estimation of the basic reproduction number for infectious diseases from serological survey data, J. R. Stat. Soc., 2001; C,50:1-33.

Fine P.E.M., Clarkson J.A., Measles in England and Wales: I, an analysis of factors underlying seasonal patterns, *Int. J. Epidemiol.*; 1982; 11:5-14.

Fine P.E.M., Clarkson J., The efficiency of measles and pertussis notification in England & Wales. *Int. J. Epidemiol.*, 1985; 14: 153-168.

Gay N.J., Eliminating measles. No quick fix, Bulletin World Health Organisation, 2000; 78: 8.

Hethcote H.W., Modelling heterogeneous mixing in infectious diseases dynamics, in Isham V., Medley G. (eds), Models for infectious human diseases, Cambridge University Press, 1995, 215-238.

Inaba I., Threshold and stability results for an age structured epidemic model, J. Math. Biol., 1990; 28: 411-433.

ISTAT, Annuari di Statistiche Sanitarie, 1955-1996, ISTAT, Roma.

ISTAT, Annuari di Statistiche Demografiche, 1951-1955, ISTAT, ROMA.

Manfredi P., Williams J.R. (2001), To what extent can inter-regional migration perturb local endemic patterns? Estimating measles incidence and under-reporting rates in the Italian regions, WP 187 Dip. Statistica e Matematica Applicata all'Economia, Università di Pisa, submitted to Epidemiology and Infection.

Manfredi P., Williams J.R. (2002), Complex interaction between low routine coverages and catchup immunization in a changing demographic environment: the case of control of measles in Italy, in preparazione. Salmaso S., Gabutti G., Rota M.C. et al., Patterns of susceptibility to measles in Italy, *Bull. WHO*, 2000;78: 950-955.

Ministero della Sanità. Circolare n. 12 del 13 luglio 1999: Controllo ed eliminazione di morbillo, parotite e rosolia attraverso la vaccinazione.

Santoro R, Ruggeri FM, Battaglia M. et al., Measles Epidemiology in Italy, *Int. J. Epidemiol.*, 1984; 13: 2,201-9.

Walllinga J., Lévy-Bruhl D., Gay N.J., Wachmann C.H., Estimation of measles reproduction ratios for elimination of measles by vaccination in some western european countries, *Epid. & Infection*, 2001; 127: 281-295.

WHO, Operational targets for EPI diseases. Copenhagen; WHO Regional Office for Europe, 1996 [EUR/HFA target 14, EUR/CMDS 01 01 11].

Williams J.R., Manfredi P., Salmaso S., Ciofi M., Heterogeneity in regional notification patterns and its impact on aggregate national case notification data: the example of measles in Italy, WP 216 Dip. Statistica e Matematica Applicata all'Economia, Università di Pisa, 2001, submitted to Epidemiology and Infection.

Type of mixing	Homog	Homogeneous	Def	Default	Realistic a	Realistic assortative	Propor	Proportionate	Preferred	Preferred mixing
pattern							•		(6:0=3)	0.9)
	RO	P _C age 0	RO	P _C age 0	RO	P _c age 0	RO	P _c age 0		P _C age 0
PIEMONTE (N)	10,50	06'0	NA	NA	12,56*	0,92	4,09	9,76	4,30	0,77
LOMBARDIA (N)	12,07	0,92	5,74	0,83	17,27*	0,94	4,80	62'0	5,18	0,81
IKENTINO (N)	11,62	0,91	5,39	0,81	14,59*	0,93	4,50	0,78	4,82	6,79
FRITTIVE (N)	11,67	0,91	5,46	0,82	17,83*	0,94	4,77	6,79	5,13	0,81
LIGURIA (N)	11,10	0,91	4,63	0,78	13,90*	66'0	4,05	0,75	4,32	7.20
EMILIA (N)	9,54	06'0	NA	NA	10,48*	06,0	3,73	0,73	3,95	0,75
,	11,03	0,91	NA	NA	18,37*	0,95	4,41	7,00	4,69	6,79
TOSCANA (C)	6,13	68'0	NA	NA	10,28*	06,0	3,60	0,72	3,79	0,74
UMBRIA (C)	9,23	68'0	NA	NA	9,04*	68,0	3,62	0,72	3,81	0,74
MAKCHE (C)	6,77	06'0	NA	NA	11,14*	0,91	3,83	0,74	4,06	0,75
(C)	10,25	06'0	NA	NA	12,51*	0,92	4,06	0,75	4,27	0,77
ABRUZZO (S)	10,59	0,91	NA	NA	15,20*	0,93	4,64	0,78	5,07	08'0
MOLISE (S)	11,07	0,91	NA	NA	12,45*	0,92	5,68	0,82	6,62	0,85
CAMPANIA (S)	11,82	0,92	7,51	0,87	10,76	0,91	5,91	6,83	68'9	0,85
RASILICATA (S)	12,92	0,92	7,80	0,87	10,33	06'0	6,01	0,83	7,14	98'0
CALABRIA (S)	12,72	0,92	69'9	0,85	15,06	6,03	5,19	0,81	5,86	68'0
SICILIA (S)	12,16	0,92	6,11	0,84	15,71	76,0	5,07	08'0	29'5	0,82
SARDEGNA (S)	12,17	0,92	7,46	0,87	11,38	0,91	5,91	0,83	86'9	98'0
	12,16	0,92	6,28	0,84	15,17	66'0	5,29	0,81	2,93	0,83
NORD	11,18	0,91	4,91	08'0	15,76*	0,94	4,38	11.0	4,66	6,79
CENTRO	9,57	06'0	NA	NA	10,94*	16'0	3,76	0,73	3,96	0,75
SUD	11,98	0,92	85'9	0,85	13,79	6,03	5,45	0,82	61'9	0,84
ITALIA	10,88	16,0	4,86	6,10	14,01*	0,93	4,33	71,0	4,64	0,78
ITALIA CORR UR	11,70	0,91	5,95	0,83	14,31	0,93	5,04	08'0	5,57	0,82
EURO	17,01	0,94	85'6	06'0	29,32	26,0	7,14	98'0	8,95	68'0
DENMARK	18,47	6,95	<i>L</i> 9'6	06'0	66,18	86'0	9,50	68'0	13,16	0,92

Tab. 2. Reproduction ratios and critical routine coverages (age zero) for measles elimination in the Italian regions under distinct assumptions on contact patterns (NA=contact matrix non admissible; *=contact matrix admissible only after FOI redefined on last age group; N=North,C=Centre,S=South)

Table 3. Average age at infection (A) and inter-epidemic period (T) in the Italian regions

	A	Thy Spectral analysis	Thy CFIP model
D; 0 17 77 4			
Fiemonte & Valle Aosta (N)	6,25	2,5	3,08
Lombardia (N)	5,63	2,5	2,92
Trentino (N)	5,71	3	2,94
Veneto (N)	5,80	3-3,5	2,96
Friuli (N)	5,88	3	2,98
Liguria (N)	98'9	45	3,22
Emilia (N)	6,17	3,5	3,06
Toscana (C)	7,08	m	3,27
Umbria (C)	6,79	3,5	3,21
Marche (C)	19'9	3,5	3,16
Lazio (C)	6,28	none	3,08
Abruzzo (S)	6,40	4	3,11
Molise (S)	6,01	5	3,02
Campania (S)	5,56	3	2,90
Puglia (S)	5,15	попе	2,79
Bastlicata (S)	5,27	2	2,82
Calabria (S)	5,56	3-3,5	2,90
Sicilia (5)	5,36	33	2,85
Saraegna (3)	5,63	4	2,92

Fig. 1. Forces of infection by age for measles in Northern, Central & Southern Italy vs "EURO" for

