

Research article

Open Access

## Heterogeneity in regional notification patterns and its impact on aggregate national case notification data: the example of measles in Italy

John R Williams\*<sup>1,2</sup>, Piero Manfredi<sup>1</sup>, Alisa R Butler<sup>2</sup>, Marta Ciofi degli Atti<sup>3</sup> and Stefania Salmaso<sup>3</sup>

Address: <sup>1</sup>Dipartimento di Statistica e Matematica Applicata all'Economia, Università di Pisa, Via C Ridolfi, 10, 56124 Pisa, Italy, <sup>2</sup>Department of Infectious Disease Epidemiology, Division of Primary Care & Population Health Sciences, Faculty of Medicine, Imperial College London, Norfolk Place, W2 1PG, London, U.K and <sup>3</sup>Istituto Superiore di Sanità, Viale Regina Elena, 98, 00161 Rome, Italy

Email: John R Williams\* - jr.williams@imperial.ac.uk; Piero Manfredi - manfredi@ec.unipi.it; Alisa R Butler - ailsa.butler@wellcome-epidemiology.oxford.ac.uk; Marta Ciofi degli Atti - ciofi@iss.it; Stefania Salmaso - salmaso@iss.it

\* Corresponding author

Published: 18 July 2003

Received: 18 December 2002

BMC Public Health 2003, 3:23

Accepted: 18 July 2003

This article is available from: <http://www.biomedcentral.com/1471-2458/3/23>

© 2003 Williams et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

### Abstract

**Background:** A monthly time series of measles case notifications exists for Italy from 1949 onwards, although its usefulness is seriously undermined by extensive under-reporting which varies strikingly between regions, giving rise to the possibility of significant distortions in epidemic patterns seen in aggregated national data.

**Results:** A corrected national time series is calculated using an algorithm based upon the approximate equality between births and measles cases; under-reporting estimates are presented for each Italian region, and poor levels of reporting in Southern Italy are confirmed.

**Conclusion:** Although an order of magnitude larger, despite great heterogeneity between regions in under-reporting and in epidemic patterns, the shape of the corrected national time series remains close to that of the aggregated uncorrected data. This suggests such aggregate data may be quite robust to great heterogeneity in reporting and epidemic patterns at the regional level. The corrected data set maintains an epidemic pattern distinct from that of England and Wales.

### Background

Highly infectious diseases such as measles necessitate vaccination of a very high proportion of the population in a sustained and effective manner in order to achieve control. Therefore national policy makers must possess data which reflect as closely as possible the true scale and pattern of infection, thus allowing vaccination policy to be tuned to achieve maximum impact.

Optimisation of vaccination programme design can be greatly facilitated through insights gained from the math-

ematical modelling of measles transmission dynamics [1]. Such modelling can shed light on reasons for observed patterns of infection in both pre-vaccination and post-vaccination eras and upon likely effects of the continuation of existing policies. However, modelling work must be based upon sound data if full confidence is to be placed in results and, since epidemiology of specific infectious diseases may differ according to geographic area, the availability of good local data is of great importance. Seroepidemiological surveys provide reliable information about patterns of experience of infection, e.g. Salmaso et al [2],

but when considering historic data, and in the absence of continuing large scale representative programmes of serological surveillance, reliance must be placed upon case notification data for information about patterns of infection over time.

To date, many of the modelling studies have focused on measles with much important analysis having been based on a relatively small number of long data series. These provide a valuable epidemiological resource, but their value relies heavily upon the consistency and the reliability of the relevant systems of case reporting. One European example is measles case notification data from England & Wales, providing raw material for much work in this field [3–6], and widely considered to be of good quality although, even here, concerns have been raised about under-reporting, and misreporting of measles cases remains a danger [7]; also notification efficacy has been found to be more accurate during periods of high incidence than when infection is rare [8]. Elsewhere in Europe, availability and quality of case notification data has varied widely between and within countries [9]. Case based surveillance systems have been employed in the majority of European countries (including Italy and the UK), although case definitions have varied until recently and some countries have also required additional laboratory confirmation or epidemiological linkage; in other European countries sole reliance has been placed upon sentinel systems [9,10]. More reliable data should result from the European surveillance network (EU-VAC-NET) recently constituted with the aim of providing a joint data base with a uniform measles disease classification [9,10]. Additionally, the European Sero-Epidemiology Network (ESEN), was set up in 1996 to gather age-stratified immunity data for measles (and other diseases) in Italy, the UK, Denmark, France, Germany, and the Netherlands, although, being based on residual sera, the data does not form a random cross-sectional survey [9,11].

In Italy measles has been a notifiable disease since the end of the 19<sup>th</sup> century (Figure 1), and in this paper we take Italy as an example of a country possessing a long and potentially valuable time series of measles case notifications, but whose immediate usefulness is potentially undermined by substantial inconsistencies in the reporting process. Data are available by age-group and geographic area from 1949 onwards but it is widely recognised that throughout this period in Italy there has been considerable under-reporting of measles cases [2]. Santoro et al [12], for example, observe that reported incidence may be an order of magnitude less than true incidence. The degree of under-reporting also varies widely between regions; indeed in a few regions notifications are consistently so few as to appear far below the bounds of credibility. This heterogeneous pattern of under-reporting

occurs side by side with the existence of significant regional heterogeneity in epidemic patterns (Figure 2) (e.g. in 1957 epidemic peaks in notifications occur in some regions, such as Marche and Lazio, but only normal seasonal variation in others such as Veneto and Trentino; in 1964 normal seasonal peaks in Lazio and Veneto contrast with epidemic peaks in Abruzzi and Campania; over much of the pre-vaccination period there are marked differences between patterns for Campania and Molise. Close inspection will allow many such contrasts to be identified). With such heterogeneities in under-reporting and epidemic patterns, there has been concern that there may be distortion of the epidemiological patterns observed in aggregate national data with a consequent effect upon the estimates of epidemiological parameters made from such data.

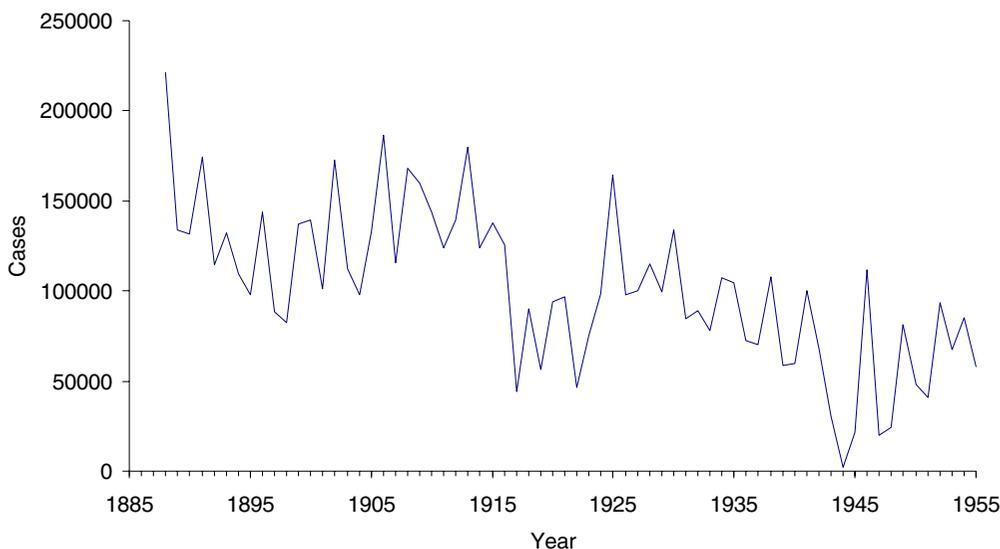
We have here attempted to overcome these inconsistencies by making use of the observed fact that, in the pre-vaccination era, nearly all were infected by measles well before reaching adulthood [3]. Thus, for each region, disregarding migration, there should be an approximate correspondence over time between regional numbers of births and numbers of measles cases. A simple procedure is employed here making use of this correspondence to estimate the degree of under-reporting. These estimates are used to revise regional notifications data in order to provide a corrected aggregate national data set; results are compared with estimates of national case notifications from measles mortality data (for the purposes of further comparison a similar procedure was employed for the national data set of measles cases for England & Wales).

Although in essence this approach to correction of data is not novel, being an extension of that of Clarkson & Fine [8] (see also [12] and [13]), we believe that what is new is its systematic application to regional datasets to provide corrected aggregate national data and that this approach does provide a much sounder basis for the epidemiological analysis of national data sets. Subsequent publications will discuss the time series analysis of this data set, an investigation of the corrected national age distribution of cases and its variation over time.

## Results

### Cases 1000 live births<sup>-1</sup>

For each region, Table 1 shows mean numbers of notifications per thousand live births for the pre-vaccination period 1949–1976, for the first 10 years post-vaccination 1977–1986 (when levels of vaccination cover were believed to be quite low), and for the period 1987–96. To highlight the low numbers of cases shown for some regions, Table 1 also includes national data from England & Wales which demonstrates a 25-fold difference between the pre-vaccination period 1949–1966 and the recent



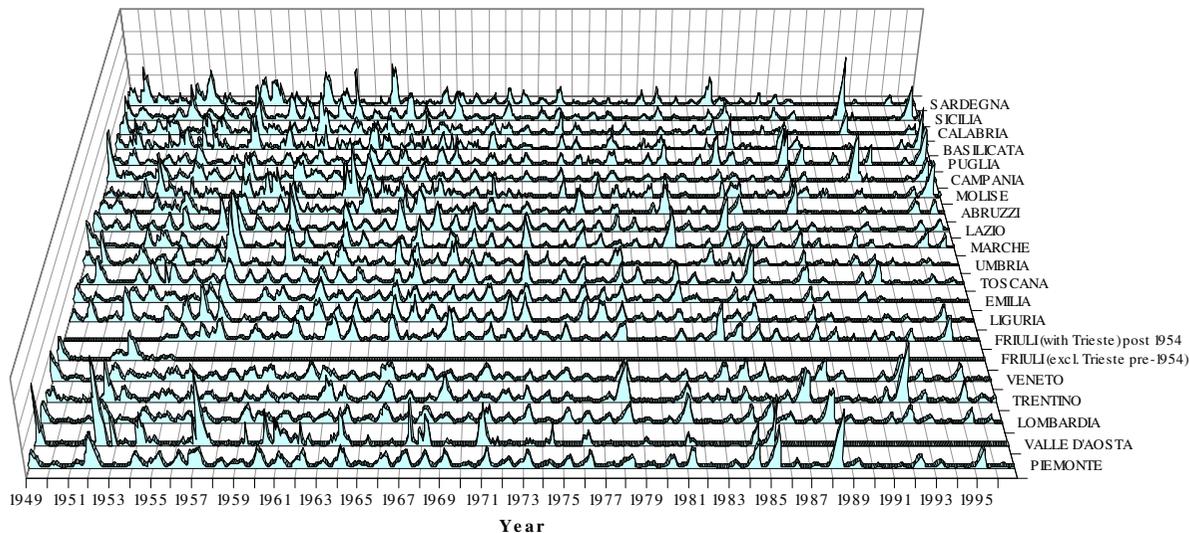
**Figure 1**  
**Time series of Italian measles case notifications from 1888 to 1955.**

post-vaccination period 1989–96. The latter figure of 22.5 notifications 1000 live births<sup>-1</sup>, covers a period when vaccine coverage of the order of 80% had been achieved [14]; in contrast the almost identical figure for the Italian region of Campania during the pre-vaccination period 1949–76 is plainly not credible.

Table 1 makes clear the wide variation in numbers of case notifications between regions in both pre- and post-vaccination eras. At its greatest, during the pre-vaccination period, this reaches an order of magnitude difference between Emilia Romagna in the north (209.3 notifications 1000 live births<sup>-1</sup>) and Campania in the south (22.6 notifications 1000 live births<sup>-1</sup>). If data for the provinces corresponding to the 4 largest cities (population > 10<sup>6</sup>) is taken into account, the difference is even greater with Campania's regional capital of Naples providing only 16.0 notifications 1000 live births<sup>-1</sup>. For half of the

regions there is also clear evidence of a decline in the number of notifications per 1000 births during the pre-vaccination period (not shown). These contrasts are maintained on a wider geographic scale with, for example, figures of 108.9 and 30.4 notifications 1000 live births<sup>-1</sup> in North and South Italy for 1949–76 and, for 1987–96, 76.9 and 34.0 respectively. In contrast, the national notification rate of 572.6 notifications 1000 live births<sup>-1</sup> for England & Wales for the pre-vaccination period 1949–66 is some 8 times greater than that for Italy (73.1) in the comparable period 1949–76.

There is a further strong contrast between Italy and England & Wales in terms of the relationship between recorded measles deaths and notified cases (Figure 3), where data for the majority of Italian regions displays a substantially higher relative number of recorded deaths/case. Although it is possible that this may, in part, reflect



**Figure 2**  
**Monthly measles case notifications for 1949–1996 standardised by regional mean.** Regions are shown in customary geographic order, and data for each region are standardised by their mean. (NB as the city of Trieste was not incorporated into the region of Friuli Venezia Giulia until 1954, when the city's post-war status within Italy was finally resolved, two time series are shown for Friuli V. G.: up to 1954, and 1954 onwards)

real differences in case mortality, it does also lend some support to the argument that fewer cases are reported in Italy.

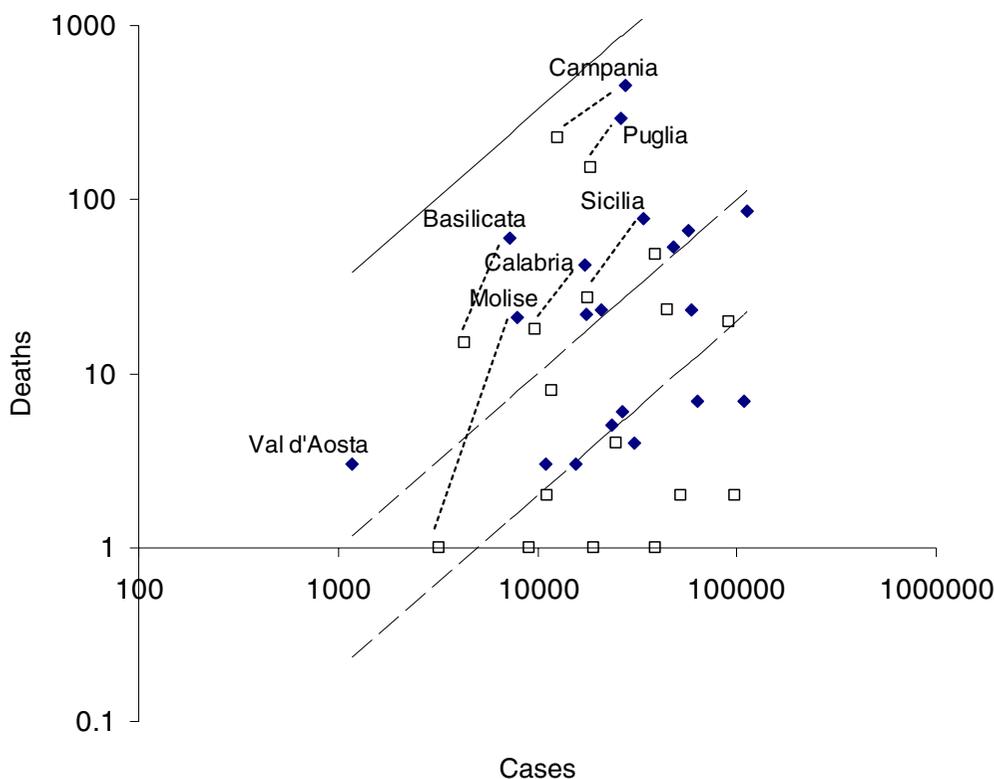
**Smoothing of births and case notifications**

The trends of smoothed case notifications data for the Italian regions (see Additional file: 1 and Additional file: 2) were more varied than those of births, although generally slightly declining towards the start of vaccination in the mid-1970's, in some cases from a broad peak or plateau and in others with occasional sharper peaks or other fluctuations (in contrast smoothing of the national pre-vaccination notifications data for England & Wales showed a more or less constant number of cases). After the start of vaccination in the mid-1970's, some Italian regions saw a

decrease in notifications, but others, strangely, an increase (perhaps occasioned by a heightened awareness of the need for monitoring cases following the introduction of vaccination); yet others showed a decrease followed by increase.

**Under-reporting factors**

For England & Wales, the magnitude of the measure of under-reporting,  $U_t$  (see Methods) suggests about 60% of cases are notified, although with a slowly decreasing trend over time. In Italy regional patterns of  $U_t$  and its mean values are quite variable for the period 1957–76 prior to vaccination (Table 2). When considering the pattern of  $U_t$  values by 5-years intervals, most of the regions show a decline over time; this reaches its maximum in Val



**Figure 3**  
**Relationship between notified measles deaths and notified cases.** Relationship between notified measles deaths and notified cases (unadjusted) for the Italian regions for the periods 1958–1967 (solid markers) and 1968–1977 (open markers) (only selected regions are labelled, with dotted lines joining corresponding data points for earlier and later periods). These are compared with the approximate range in the same relationship for England & Wales for the late 1940's and early 1950's (upper dashed line) and for the late 1950's to early 1960's (lower dashed line). Also shown (solid line) is an example of this relationship from a recent outbreak in a developing country (Peru) [16]

d'Aosta, where  $U_t$  decreased from 15.4% in 1952–1956 (not shown), to 3.6 in 1972–76, but it is particularly evident also in some regions of Southern Italy, such as Campania and Sicilia. Over the whole of the pre-vaccination period the data for Emilia suggests the highest level of reporting, estimated at 21.2% of cases, whereas Campania, the lowest, is estimated to report only 2.0% of cases. Figure 4 makes clear the north-south contrast between regions by providing a summary of these data in map form.

For purposes of further comparison Table 3 shows  $U_t$  for Abruzzo, a region in the middle range of estimates of under-reporting, compared with estimates of under-

reporting where weights of the moving average for births are derived using a catalytic model together with various estimates of the force of infection (see Methods)[15]; a further comparison was made using weighting based upon the 1971–76 age distribution of actual reported cases. The simple moving average,  $U_t$ , compares well with the estimates using the more complex weighting systems estimates. Figure 5 compares the actual national age distribution of case notifications for the years 1971–76 with birth year distributions (i.e. age distributions) suggested by FOI-based estimates and with that of the simple approach used here. It is clear therefore that FOI-based estimates do not necessarily provide a closer correspondence to the age distribution of cases (but see caveat in

**Table 1: Measles notifications for Italy by region.**

	1949-76	1977-86	1987-96
Piemonte	89.7	93.0	93.6
<i>Turin</i>	63.5	81.5	-
<i>Excl. Turin</i>	116.9	107.0	-
Valle D'aosta	94.2	45.8	30.1
Lombardia	86.3	103.2	89.4
<i>Milan</i>	63.5	94.5	-
<i>Excl. Milan</i>	102.0	109.8	-
Trentino A.A.	70.9	94.6	114.7
Veneto	72.3	86.0	89.4
Friuli Venezia G.	168.0	229.0	147.7
Liguria	134.4	134.5	67.7
Emilia Romana	209.3	221.3	92.3
Toscana	116.8	147.1	99.6
Umbria	129.7	99.4	58.9
Marche	113.6	127.9	87.7
Lazio	77.2	56.4	36.0
<i>Rome</i>	82.4	61.9	-
<i>Excl. Rome</i>	65.2	43.6	-
Abruzzo	80.1	61.0	53.3
Molise	83.0	76.4	85.6
Campania	22.6	11.6	22.8
<i>Naples</i>	16.0	7.1	-
<i>Excl. Naples</i>	30.4	17.5	-
Puglia	28.3	33.7	46.3
Basilicata	48.5	45.3	51.3
Calabria	33.7	15.6	14.8
Sicilia	28.3	13.1	27.2
Sardegna	57.3	39.9	23.5
<i>NORTH</i>	108.9	123.1	76.9
<i>CENTRE</i>	88.2	71.9	41.7
<i>SOUTH</i>	30.4	20.2	34.0
<i>ITALY</i>	72.6	69.7	56.0
England & Wales	572.6 (1949-66)		22.5 (1989-96)

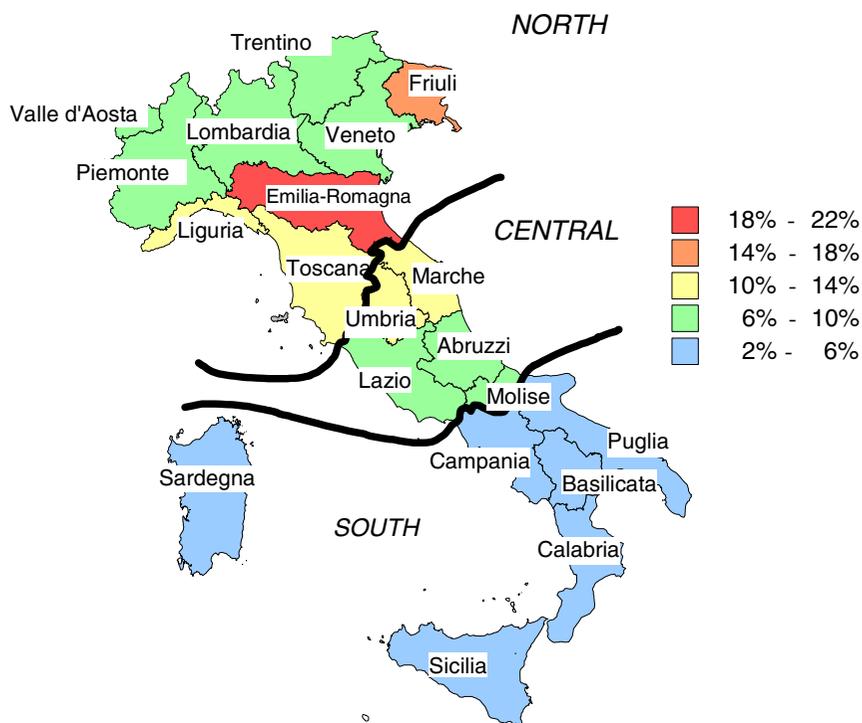
Mean number of notifications of measles cases 1000 live births<sup>-1</sup> for the specified period for Italy by region (from north to south), broad geographic region (North, Central, South (see Fig 4) and for England & Wales. Specific data is not available for the major cities: Milan in Lombardia, Turin in Piemonte, Rome in Lazio, and Naples in Campania; these are represented by their provinces (the major part of whose populations are found in these cities) with, additionally, mean figures for the 4 regions concerned excluding these provinces.

Discussion concerning potential age bias in reporting of cases).

#### **Corrected case notification data for Italy**

Figure 6 shows uncorrected aggregate national notification data in comparison with corrected aggregate notification data where, prior to aggregation, the appropriate mean under-reporting correction is applied to the data for each region. Although there is great variability in magnitude of the correction for the different regions, and the magnitude of the overall correction is very large, the effect of the correction on yearly epidemic patterns is remarkably small with very close correspondence between the shapes of uncorrected and corrected data, and the timing of epidemic peaks is changed by, at the most, one month or so.

Although this adjusted Italian case data is now of similar magnitude to that for England & Wales (see comparison in Figure 7 in which national data for England & Wales is adjusted for under-reporting), the corrected Italian data continues to show a strongly distinct pattern of roughly 3-year periodicity compared with that of 2-years seen for England & Wales. The clear pattern of a single seasonal peak occurring in March or April is also maintained (Figure 8); this occurs a month or so later than that seen in weekly pre-vaccination data for England & Wales [3] which also shows small troughs in January and April, absent, or not clearly apparent, from the monthly Italian data. The results of the present work suggests that such national shapes and trends in notifications data can be quite robust to very substantial variation in under-reporting between regions. It may however be the case that



**Figure 4**  
**Estimated mean percentage of measles cases reported by region for the pre-vaccination period 1953–1976.**  
 The map also shows the conventional partition of Italy into North, Central and South (see Table 1).

detailed time series analysis might reveal differences between corrected and uncorrected data series that are not immediately apparent from visual inspection; this issue will be addressed in a forthcoming paper.

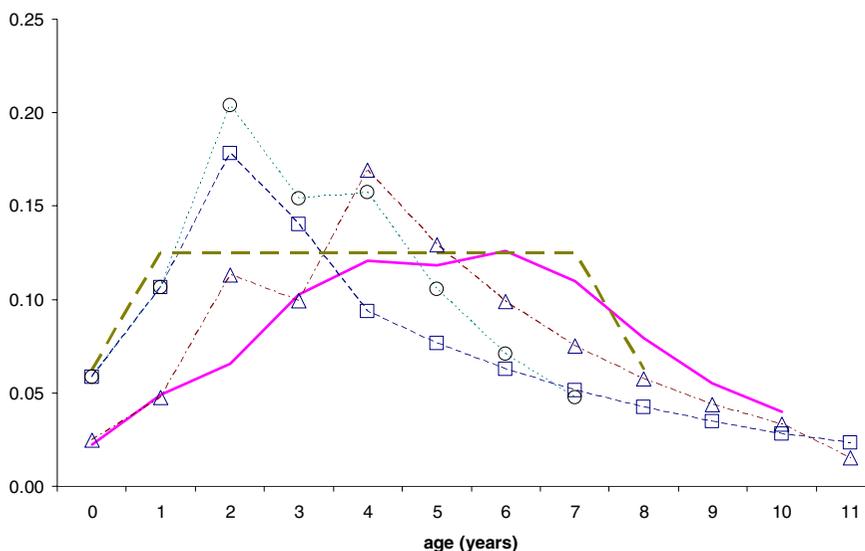
**Discussion**

The degree of under-reporting seen in Italian regional notification data of the pre-vaccination era is in some instances, strikingly large, so that corrected aggregate national data for Italy is an order of magnitude greater than for the uncorrected data. Nevertheless, although they vary widely between regions, shapes, location and trends of adjusted and unadjusted national data are very similar.

**Case reports and measles deaths**

Although the proportion of cases notified may be poor, it seems likely that deaths arising from measles cases would

be more reliably reported, so that a relatively high ratio of deaths to cases might be expected in the circumstances described. Figure 3 confirms this expectation by comparing notified deaths and cases for the period 1958–77 with those observed in England & Wales over similar and earlier timescales [14]. It may be argued that the higher Italian 'deaths to cases' ratios may in part result from poor living conditions or standards of health in certain regions (in this regard an indication of the ratio from one developing country (Peru) is also included for comparison[16]). Some support for this notion is provided by a comparison of infant mortality rates for Italy (52.7 and 35.6 1000<sup>-1</sup> live births in 1951–60 and 1961–70 respectively [17]) with those of England & Wales (29.1 and 17.7 1000<sup>-1</sup> live births in 1950–2 and 1970–2 respectively) [7]and Peru (47.48 1000<sup>-1</sup> live births in 1995 [18]) but it is also likely that the high Italian 'deaths to cases'



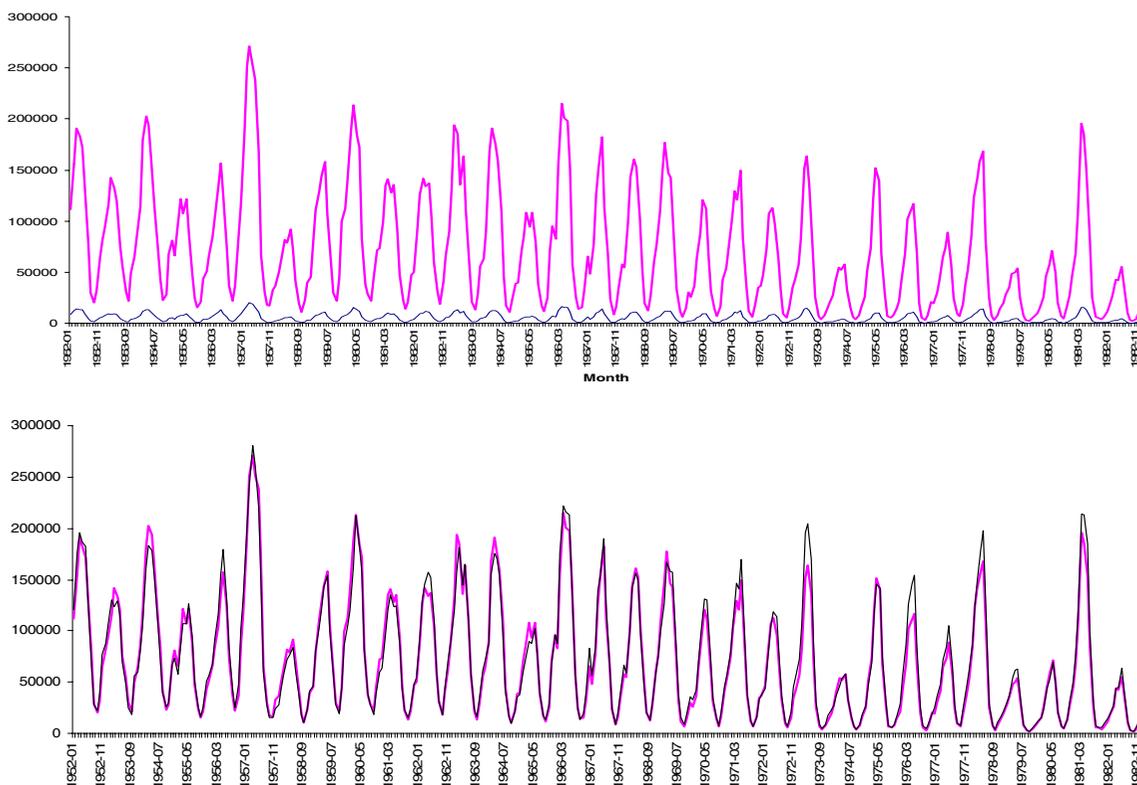
**Figure 5**  
**Age distribution of measles cases in Italy for 1971–76.** Age distribution of measles case reports for Italy for 1971–76 (thick solid line) is compared with age distribution of weights of 9-year moving average (thick dashed line), and those of the age distributions suggested by force of infection (FOI) estimates of Edmunds *et al* [15] made, respectively, from seroprevalence data (squares), case notifications (triangles) and that suggested by the composite 'EURO FOI' of those authors corresponding to a number of European countries (circles)

ratio also reflect the very poor levels of case reporting illustrated above [2] (the term 'case fatality rate' is deliberately avoided because of the low level of case reporting).

**Under-reporting in the vaccination era**

Although the relationship between births and cases clearly begins to break down when vaccination levels increase, it could be argued that levels of under-reporting seen in the pre-vaccination era might continue after the start of a vaccination programme, but this does not take into account changes in reporting regime instituted at the same time nor changes in perception of the value of reporting. Nevertheless, in absence of seroprevalence data or sentinel surveys, correction of post-vaccination data using mean adjustment factors calculated for the whole or part of the

data from pre-vaccination years may prove more satisfactory than reliance upon uncorrected data which there is every reason to suppose substantially under represents true numbers of cases and distorts the contribution of individual regions. Whether it is more appropriate to extrapolate the trend seen in pre-vaccination era correction factors or simply to employ a correction factor calculated for the immediately pre-vaccination years is a matter for debate. At first sight, the former might appear more soundly based but, until underlying social and administrative factors determining trends in the degree of under-reporting are better understood (Figure 9), it would be more prudent to consider only corrections derived from the latter part of the pre-vaccination era. However a number of factors in the post-vaccination period may



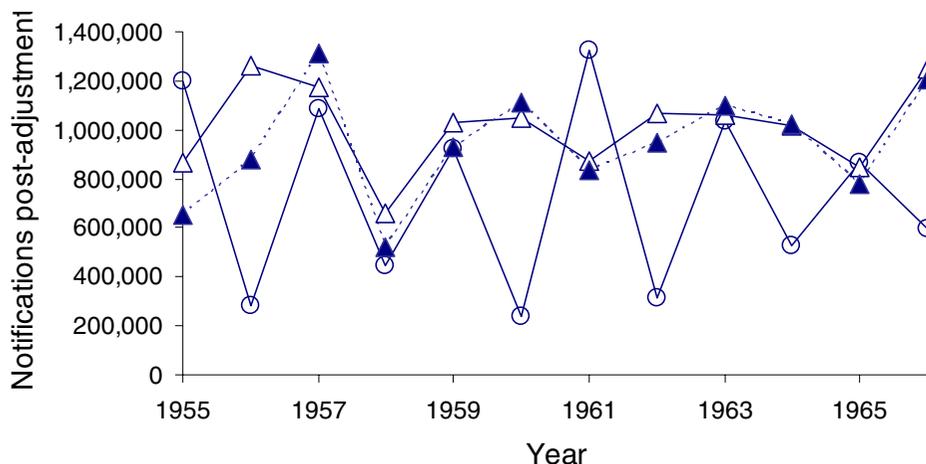
**Figure 6**  
**Adjusted and unadjusted measles notification data 1953–82.** Comparison of (a) unadjusted national Italian measles notification data (thin line) and adjusted (dashed line) nationally aggregated regional notification data for the years 1953–82 and (b) the same data but with the uncorrected national data scaled to be of the same magnitude as that of the aggregate adjusted regional data (using the ratio of the means of the corrected and uncorrected data as the scaling factor).

potentially disturb the magnitude of, or trend in, under-reporting as a result of changing perceptions arising directly from institution of a vaccination programme. For example, in general terms, such a programme, by underlining the public health impact of measles, might encourage increased reporting; alternatively, reporting might decrease as a result of some complacency arising from the very existence of a vaccination programme. Changing the aim of policy from control to elimination, possible concerns over vaccine safety, and the adoption of monetary incentives all may impact upon notification. Also it has been argued that, as measles becomes less common with increasing vaccination, there is an increased tendency to misdiagnose simply as a result of lack of experience in identifying measles cases; as this might lead to

both an increase in under-reporting and an increase in the number of cases of other infections misreported as measles, its overall impact is not clear [7,19].

**Age bias in reporting**

Overall, though corrected age-aggregated notification data for Italy are certainly much closer to true numbers of cases, the issue raised by Edmunds et al [15] of age bias in reporting of measles cases in Italy remains to be considered (this stands as a counter-example to the observation of Fine and Clarkson [3] that the probability of a case of measles being reported appears higher among young children than adolescents and adults). If there is indeed a substantial increase in under-reporting with decreasing age, there would be an overestimation of the average age



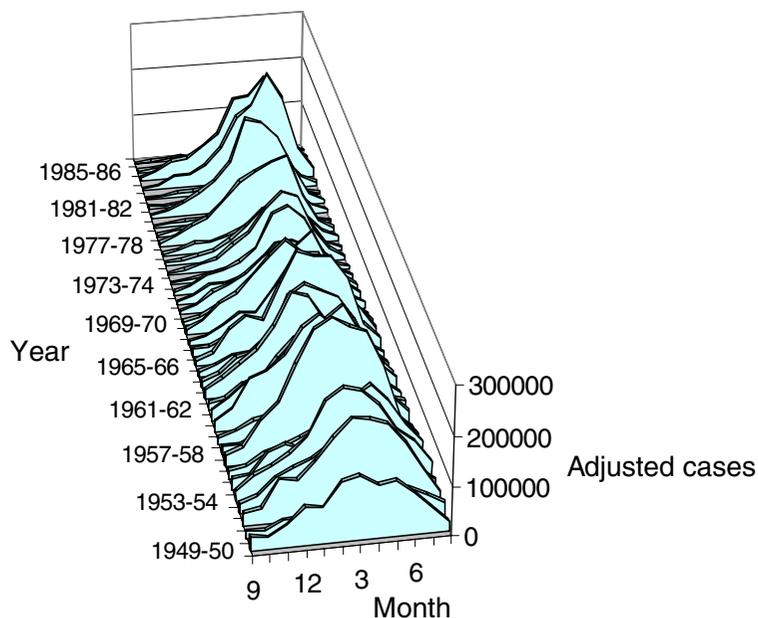
**Figure 7**  
**Post-adjustment comparison of annual notification data for Italy and England & Wales.** Comparison of Italian nationally aggregated regional yearly notification data (open and solid triangles) adjusted for under reporting with national data for England & Wales (aggregated prior to adjustment) (open circles). Adjustments are by moving average of births as described in the text (solid triangles) and by overall means of the moving averages (open triangles and circles).

of infection and underestimation of the force of infection in the youngest age groups. This may in turn impact on health policy both directly and, indirectly, through the effect on parameter estimation for modelling work informing health policy, and, of course, vaccination programmes themselves will influence the average age of infection and so modify the effect of reporting age-bias. It is therefore of great importance that further investigations be undertaken to address this issue of age bias.

Clearly there is a need to improve overall levels of case notification within the regions of Italy. So far, no formal studies to investigate the reasons for under-reporting have been conducted. Nevertheless, the experience of a study conducted with a sentinel network of primary care pedia-

tricians [20] shows that simplicity of reporting and regular feed-back of results can greatly improve case notification. A recent evaluation of varicella under-reporting in Italian children and adolescents, showed also that case notification was more complete in primary-school-aged children compared with that for other age-groups [21], probably because medical certificates for readmission to school are required after certain infectious diseases, including measles and varicella. Such certificates are usually provided by public health officials who may be more aware than general practitioners of the relevance of infectious disease notifications.

Nevertheless, it is clear that further research will need to be undertaken if we are to understand fully the reasons for



**Figure 8**  
**Seasonal pattern for Italy of adjusted measles notification data.** Monthly regional notification data for the period 1953–82 nationally aggregated after adjustment.

such poor levels of reporting and if sound strategies for optimising the impact of vaccination are to be developed. However, it is difficult to foresee a mechanism for overcoming the apparently dramatic extent of under-reporting in some regions that would not depend upon a substantial investment in monitoring systems.

### Conclusions

Despite qualifications outlined above, it is believed that the adoption of the procedure described here provides, by eliminating potential distortions arising from wide variation in regional reporting levels, both i) a national time series of measles data, and ii) a national age distribution of cases, of improved reliability which will prove a useful basic resource to aid further understanding of the epidemiology of measles infection, both within Italy and,

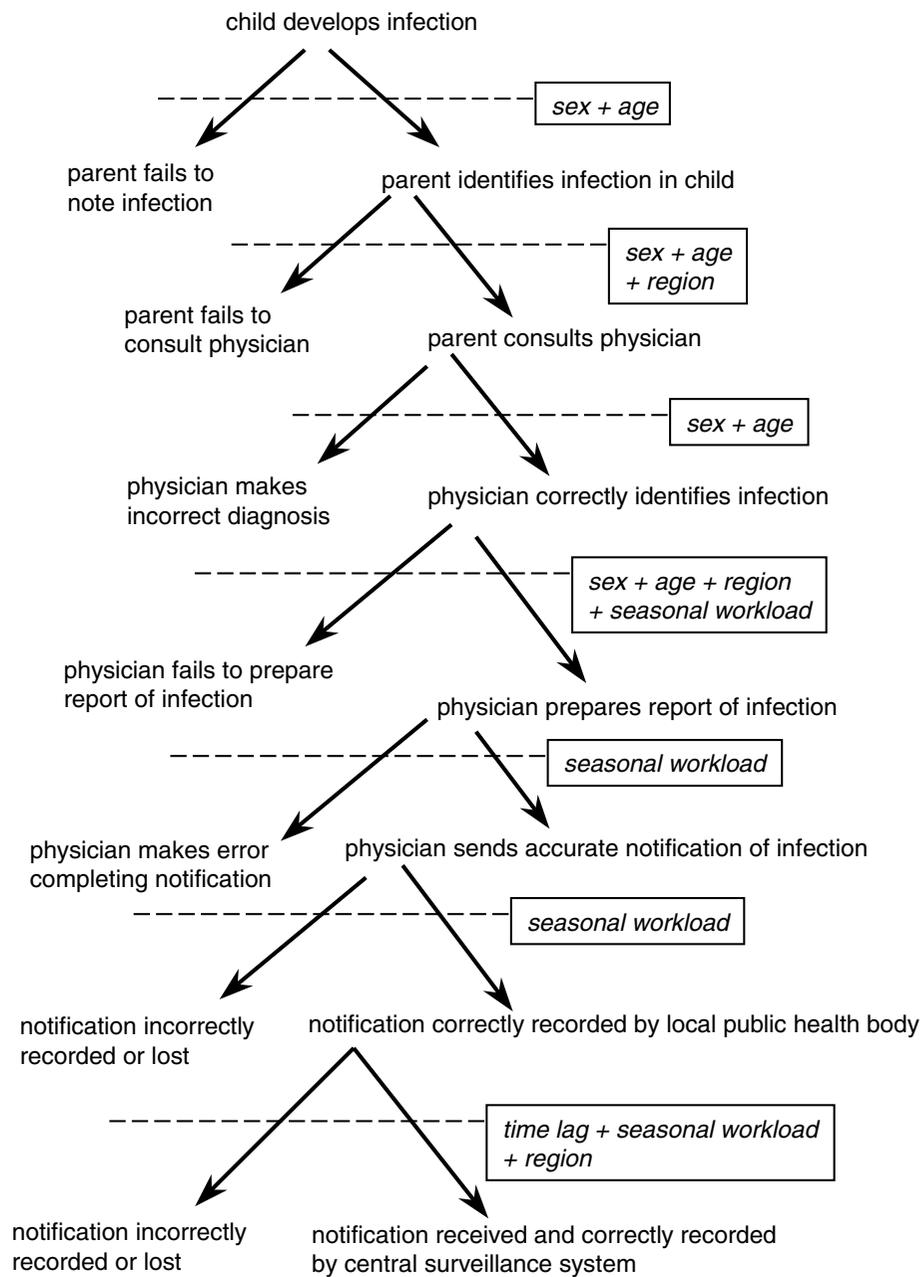
when similar substantial regional heterogeneity in notification levels become apparent, in the wider context.

### Methods

#### Sources of data

Monthly measles case notification data were provided by the Istituto Nazionale di Statistica (ISTAT) for the period from the first available year, 1949, to 1996. These show substantial heterogeneity between regions. Previously published data on annual births in Italy and measles deaths in Italy and England & Wales were also used.

In Italy there is a pattern of mixed private and public sector delivery of vaccination and measles vaccination is classed as 'Recommended' rather than 'Compulsory' (as is also the case for mumps and rubella). Nationally vaccination against measles was first made available towards the



**Figure 9**  
**Illustration of some possible sources of error in case notification and of systematic bias (boxes).** Figure assumes notification system based on case reports only, without laboratory confirmation of infection

**Table 2: Estimates by region of Italy of percentage of measles cases notified for periods specified.**

	1957-61	1962-66	1967-71	1972-76	Pre-vaccination mean
Abruzzo	8.6	8.8	7.4	5.8	7.9
Basilicata	4.9	4.8	4.2	3.2	4.4
Calabria	3.9	3.4	2.7	2.1	3.1
Campania	2.9	2.3	1.5	1.0	2.0
<i>Naples</i>	2.2	1.7	1.1	0.6	1.5
<i>Excl. Naples</i>	3.5	2.9	2.0	1.6	2.6
Emilia Romana	22.8	22.3	19.9	18.4	21.2
Friuli Venezia G.	20.7	20.0	15.8	12.7	17.9
Lazio	9.8	8.3	6.9	4.7	7.8
<i>Rome</i>	11.4	8.7	7.2	4.7	8.4
<i>Excl. Rome</i>	7.1	7.2	6.0	4.7	6.5
Liguria	16.1	14.3	11.1	10.4	13.2
Lombardia	11.0	10.0	7.6	6.3	9.0
<i>Milan</i>	8.9	7.6	5.7	4.7	7.0
<i>Excl. Milan</i>	12.2	11.5	9.1	7.6	10.4
Marche	12.4	11.4	10.6	9.5	11.2
Molise	9.2	11.0	7.7	5.2	8.6
Piemonte	12.8	9.8	7.5	5.7	9.4
<i>Turin</i>	10.2	7.2	5.3	4.3	7.0
<i>Excl. Turin</i>	14.6	12.4	10.0	7.7	11.6
Puglia	3.0	3.4	2.7	2.1	2.9
Sardegna	6.4	5.1	4.3	3.3	5.0
Sicilia	3.4	3.2	2.4	1.6	2.8
Toscana	15.2	14.1	12.4	11.8	13.5
Trentino A.A.	7.1	7.8	6.7	6.2	7.0
Umbria	15.5	13.5	12.0	7.9	12.8
Valle D'aosta	12.3	8.4	6.3	3.6	8.2
Veneto	8.4	8.1	6.5	5.3	7.3

Figures correspond to  $100 U_i$  (estimated percentage of cases reported) with a 4 year lag between moving averages of births and cases.

**Table 3: Effect of different weighting systems on estimates of under-reporting.**

Weighting system	% cases reported
Catalytic model using force of infection (FOI) from Italian seroprevalence data	8.9
Catalytic model using FOI from case notifications data	8.7
Catalytic model using EURO FOI (see text)	8.4
Age distribution of cases 1971-76	8.7
Simple moving average	8.3

Estimated levels of measles case notification 1960-1974 arising from the use of different weighting systems for the moving average of numbers of births using as an example the region of Abruzzo.

end of 1976 and became officially recommended in 1979. There is little data on vaccine uptake but initially it is believed to have been low [2] and, for the first 10 years at least, to have remained at disappointingly low levels [22], perhaps exacerbated by the fact that only a proportion of local health units provide vaccine free of charge.

#### **Estimation of degree of under-reporting**

As note above, prior to widespread commencement of measles vaccination, most individuals were infected by measles before adulthood [3]. Clearly because of the cyclic epidemic pattern, variation in the age at which individuals are infected, and fluctuations and trends in birth numbers, there could not be an exact correspondence between births and cases. Nevertheless, ignoring

migrational flows (see below) and assuming endemic equilibrium, during the pre-vaccination period numbers of measles cases over time would have been approximately equivalent to numbers of births, providing a basis for the estimation of the degree of under-reporting [8].

One approach to such estimation would be to select a suitable value for the force of infection (FOI: the per capita yearly incidence of infection in the susceptible proportion of the population) and, by using a simple catalytic model [1], to estimate the equilibrium age distribution of cases and hence distribution of the birth years of those infected in an average year. Although perhaps theoretically sound, this has two main drawbacks. First, it is quite demanding in terms of births data. Second, it presumes availability of some "suitable" FOI (a reasonable choice used here could be the EURO FOI estimated by Edmunds et al [15] for a wide range of European countries). This assumption is clearly not neutral (nor even tautological), as it implies imposing the structure of the hypothesised FOI upon the very data from which subsequently an estimation of the true FOI would be made. Instead we seek here a correction tool based on minimal assumptions, to minimise possible biases introduced by the correcting algorithm.

In pre-vaccination Europe around half of all measles cases would be expected to occur in the first 4–7 years of life [14], with numbers of cases at older ages becoming more and more widely distributed with increasing age. In these circumstances it was felt that estimation of under reporting using a moving average to smooth yearly birth data would be more than adequate (see below for a theoretical justification). In the event, values obtained by this means proved sufficiently close to those derived using the catalytic modelling approach above to additionally justify its use in this instance on grounds of simplicity, greater transparency and ease of application as a general method. A births curve smoothed by averaging over some appropriately delayed period of years [5], should encapsulate much of the magnitude and trend of pre-vaccination measles cases. Therefore the ratio of the moving averages of births to that of notifications should provide a suitably robust estimate of the degree of under-notification. Here moving averages of both births and cases for each region were calculated as:

$$\gamma_t = \frac{0.5x_{t-n/2} + x_{t-n/2+1} + \dots + x_{t-n/2+(n-1)} + 0.5x_{t-n/2+n}}{0}, \quad n = 8$$

Where  $x_t$  denotes the original time series of births or case notifications, as appropriate,  $n + 1$  the number of terms included in the moving average and  $\gamma_t$  the moving average

series. The estimated level of under-reporting,  $U_t$ , was therefore simply:

$$U_t = \frac{C_t}{B_{t-m}}, \quad t = 1957, 1958, \dots, 1982$$

where  $C_t = \gamma_{t(\text{cases})}$  and  $B_{t-m} = \gamma_{t(\text{births})}$ . The lag,  $m$ , between averages for births and those of cases is intended to take into account the cohorts supplying the greater proportion of cases occurring in any given year [5,15]. As an example, in the present instance where a lag of 4 years was chosen (the same as used by Clarkson & Fine [8]),  $U_{1957}$  is computed from birth numbers for the years 1949–57 and case notifications for the years 1953–1961.

**Using births as estimators for cases: theoretical justification**

We consider the simplest SIR compartmental model [1] describing infection in a stationary homogeneously mixed population in absence of vaccination:

$$\begin{aligned} \dot{X}(t) &= B(t) - (\mu + \lambda(t))X(t) \\ \dot{Y}(t) &= \lambda(t)X(t) - (\mu + \nu)Y(t) \\ \dot{Z}(t) &= \nu Y(t) - \mu Z(t) \end{aligned} \quad (i)$$

where  $X(t)$ ,  $Y(t)$ ,  $Z(t)$  represent numbers of susceptible, infective and recovered/immune individuals at time  $t$ ,  $B(t)$  births per unit time,  $\lambda(t) = \beta Y(t)$  the force of infection (FOI),  $\mu$  the mortality rate (constant for simplicity), and  $\nu$  the (constant) recovery rate. The total population  $N = X + Y + Z$  is assumed stationary with  $B = \mu N$ . At equilibrium numbers of births ( $B$ ) and cases ( $C$ ) respectively are  $B = (\mu + \lambda)X$  and  $C = \lambda X$ , so that:  $C = B - \mu X$ .

A prediction of equations (i) is that the susceptible fraction at equilibrium,  $X/N$ , is the reciprocal of  $R_0$  (the basic reproduction ratio of the infection [1]). Thus:  $C = B - \mu N/R_0 = B(1 - (1/R_0))$ . Therefore for very high values of  $R_0$  such as for pre-vaccination measles, with a sufficiently low mortality rate,  $C/B \approx 1$ . This relation is basically the consequence of the fact that the age window over which the forces of mortality and infection operate are separated. It is worth noting that the basic relation between births and cases holds in more general circumstances. Suppose there is a steady oscillation of (i) around its long term equilibrium, as is typical of periodically forced SIR and SEIR models. In this case the relation  $C/B \approx 1$  remains correct provided we consider average values of cases and births over the appropriate time period. More generally if the model is non-stationary, provided state variables are bounded in the long term, the equilibrium condition (time derivatives of the state variables are zero) may be replaced by the weaker condition that their long term

averages are zero:  $E(\dot{X}(t)) = E(\dot{Y}(t)) = E(\dot{Z}(t)) = 0$ , so that  $E(C) = E(B) - \mu E(X)$  indicating that the basic relation still holds in a broader sense, i.e. if yearly numbers of cases and births are replaced by their averages over a sufficiently long period.

The previous relations are preserved when age-structure is explicitly introduced. Consider a SIR model with chronological age structure [1] in a stationary population with no additional mortality arising from disease

$$\begin{aligned} \frac{\partial X}{\partial a} + \frac{\partial X}{\partial t} &= -(\mu(a) + \lambda(a,t))X(a,t) \quad X(0,t) = B(t) \\ \frac{\partial Y}{\partial a} + \frac{\partial Y}{\partial t} &= \lambda(a,t)X(a,t) - (\mu(a) + \nu)Y(a,t) \quad Y(0,t) = 0 \\ \frac{\partial Z}{\partial a} + \frac{\partial Z}{\partial t} &= \nu Y(a,t) - \mu(a)Z(a,t) \quad Z(0,t) = 0 \end{aligned} \tag{ii}$$

where  $X, Y, Z, \lambda, \mu$  are as in equation (i) with the added dimension of age,  $a$ .

With a stationary population  $B(t) = B$ , and assuming equilibrium, the total number of cases of infection per unit time is:

$$C = \int_0^{\infty} \lambda(a)X(a)da = B \int_0^{\infty} \lambda(a)p(a)\Lambda(a)da \tag{iii}$$

where  $p(a), \Lambda(a)$  denote respectively survival to death and infection functions. Under Type I mortality ( $\mu(a) = 0, a < L, \mu(a) = \infty, a \geq L$ ), commonly used to approximate mortality in industrialised developed countries:

$$C = B \int_0^L \Lambda(a)\lambda(a)da = B(1 - \Lambda(L)) \tag{iv}$$

where  $F(L) = 1 - \Lambda(L)$  is the fraction who have experienced the disease by age  $L$ . Since for measles in the pre-vaccination era  $F(L)$  is usually very close to 100%, again the relation  $C \approx B$  holds. In the special case of homogeneous mixing by age one finds in particular  $C = \lambda X = B(1 - e^{-\lambda L})$ .

Using the approximate relation  $A = 1/\lambda$  (holding under Type I mortality [1]) one gets the further relations  $C \approx B(1 - e^{-L/A}) = B(1 - e^{-R_0})$  showing that yearly numbers of cases at equilibrium are quite close to yearly numbers of births for large  $R_0$  values. Under Type II mortality ( $\mu(a) = \mu$ ) one recovers the relation  $C = B(1 - (1/R_0))$  found for (i).

A possibility offered by (iii) for estimating the degree of under notification of measles cases under more general circumstances appears if one explicitly introduces time,

showing that the total number of cases is a weighted average of past births

$$C(t) = (1 - F(L)) \int_0^L B(t-a)G(a)da \tag{v}$$

the weights being given by the (normalised) density of infection:  $G(a) = \Lambda(a)\lambda(a)/(1 - \Lambda(L))$ ; this is the catalytic approach mentioned in the text. As long as births do not fluctuate wildly, cases can thus be reconstructed from births provided a suitable force of infection is assumed. The further simplified approach used in this paper follows by considering time averages of (v) over a generic interval  $(t_0, t_1)$ :

$$\bar{C}(t_0, t_1) = \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} C(t)dt = \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} \left( \int_0^L B(t-a)G(a)da \right) dt \tag{vi}$$

By the mean value theorem a value  $\zeta$  exists such that

$$\bar{C}(t_0, t_1) = (t_1 - t_0)^{-1} \int_{t_0}^{t_1} B(t-\zeta) \left( \int_0^L G(a)da \right) dt = (t_1 - t_0)^{-1} \int_{t_0}^{t_1} B(t-\zeta)dt = \bar{B}(t_0 - \zeta, t_1 - \zeta) \tag{vii}$$

(vii) suggests that a moving average of births, delayed by a suitable choice of  $\zeta$ , may be used to reconstruct the time series of present (moving averaged) cases. The literature ([15] and references therein) suggests that possible shapes of the age densities of infection for measles in pre-vaccination regimes are sufficiently well behaved so that the choice of  $\zeta$  is easy. Compared to (v), (vii) has the advantage that assumptions about the full shape of the force of infection function are unnecessary.

Finally we point out that the relationship between cases and births is robust to changes in basic model assumptions, extending to more general models such as SEIR's. Moreover, if the population is exponentially increasing rather than stationary (typical of developing countries) it holds in a more general form involving discounted births rather than actual births.

**The effects of migration**

A potential distorting influence on apparent patterns of under-reporting in Italy during the pre-vaccination period is the huge volume of internal migration from 1950 to 1975 (in many regions the major source of population change, e.g. in Turin during 1952-63 net migration was about twice birth numbers) combined with the age distribution of susceptibility amongst these migrant families (we also note that, over the same time window, international migration was, for practical purposes, negligible). We have investigated whether use of birth rates to correct notifications could have over-estimated true case numbers for exporting regions, and the reverse for importing ones. Fortunately, the Italian migration flow during

**Table 4: Force of infection values estimated from Italian case notifications (ICN-UR) and from those from a number of European countries (EURO)**

Age groups	Values of the force of infection (%/year)				
	0-1	2-4	5-10	11-17	18+
ICN-UR	7	15	31	19	6
EURO	12	28	40	20	10

1950-1970 was essentially unidirectional, from Southern "source" regions to Northern "sink" regions. Assuming measles incidence fluctuates around its pre-vaccination equilibrium, partition between "source" and "sink" regions allows straightforward extension of the basic relation between births and cases. Thus, for a source population: Cases  $\equiv$  (Births - Emigrations of susceptibles), and for sink regions (and under some further assumption) the Cases  $\equiv$  (Births + Immigrations of susceptibles). Clearly exit or entry of susceptibles are modulated by the age distributions of the overall migration flow and of infection in the regions of origin. Net migration by age was estimated for each region by standard demographic techniques, using 1951 and 1971 census data, with the reasonable assumption that the force of infection prevailing in pre-vaccination Italy could be bounded between values recently estimated from data from several European countries [15](EURO FOI), and values estimated from Italian case notifications (ICN FOI) (Table 4).

This exercise suggests substantial movements of susceptible individuals: e.g. during 1951-61 for typical industrial "sink" regions, such as Piemonte and Liguria, taking immigration into account expected case numbers were up to 15% larger (under ICN FOI) than those expected by ignoring migration. Despite this, the impact of migration on estimated levels of under-reporting is rather limited. Only in a few regions are there absolute variations ( $|U_{\text{incl.migration}}/U_{\text{excl.migration}}|$ ) of at most 2% (usually much less) with respect to what is predicted by the basic cases-births relation. This broadly confirms the robustness of the simplified approach followed in the paper. Full details of results on estimation of under-reporting in presence of migration are available on <http://statmat.ec.unipi.it/manfredi.html>.

### Competing interests

None declared.

### Authors' contributions

JRW conceived and undertook the data correction procedure and drafted the manuscript. PM provided theoretical justification for use of the data correction procedure with and without migration and the estimates of the migration

section. ARB contributed text on European notification systems. MCdA and SS drew together the data and offered insight on the working of the reporting system in Italy.

### Additional material

#### Additional file 1

Smoothed case notification data

Figures showing smoothed case notification data for each Italian region from the early 1950's to early 1980's.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2458-3-23-S1.doc>]

#### Additional file 2

Smoothed births data

Figures showing smoothed births data for each Italian region from the early 1950's to early 1980's

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2458-3-23-S2.doc>]

### Acknowledgements

This paper was written within the research project Progetto Nazionale "Epidemiologia delle Vaccinazioni in Italia". John Williams was funded under a project grant from the Italian Ministry of University and Scientific Research. We thank Donatella Mandolini, John Edmunds, Alberto Tozzi, and Antonino Bella for their valuable comments and suggestions. Usual disclaimers apply.

### References

1. Anderson RM and May RM: **Infectious diseases of humans: dynamics and control** Oxford, OUP; 1991.
2. Salmaso S, Gabutti G, Rota MC, Giordano C, Penna C, Mandolini D and Crovari P: **Pattern of susceptibility to measles in Italy** *Bulletin of the World Health Organization* 2000, **78**:950-955.
3. Fine PEM and Clarkson JA: **Measles in England and Wales: I, an analysis of factors underlying seasonal patterns** *International Journal of Epidemiology* 1982, **11**:5-14.
4. Anderson RM, Grenfell BT and May RM: **Oscillatory fluctuations in the incidence of infectious disease and the impact of vaccination: Time series analysis** *Journal of Hygiene* 1984, **93**:587-608.
5. Grenfell BT and Anderson RM: **The estimation of age-related rates of infection from case notifications and serological data** *Journal of Hygiene (Cambridge)* 1985, **95**:419-436.
6. Finkenstadt B and Grenfell BT: **Empirical determinants of measles metapopulation dynamics in England and Wales** *Proc. R. Soc. London B* 1998, **265**:211-220.

7. Ramsay M, Brughra R, and Brown D: **Surveillance of measles in England and Wales: Implications of a national saliva testing programme** *Bulletin of the World Health Organization* 1997, **75**:515-521.
8. Clarkson JA and Fine PEM: **The efficiency of measles and pertussis notification in England & Wales** *International Journal of Epidemiology* 1985, **14**:153-168.
9. Butler AR: **Measles, mumps and rubella (MMR) immunisation programmes in Europe: analyses of the impact on the incidence of measles based on mathematical models of viral transmission dynamics** *Zoology Oxford, Oxford*; 2002:402.
10. Glismann S, Rønne T and Schmidt J-E: **The EUVAC-NET survey: national measles surveillance systems in the EU, Switzerland, Norway and Iceland** *Eurosurveillance* 2001, **6**:105-110.
11. Osborne K, Weinberg J and Miller E: **The European Sero-Epidemiology Network** *Eurosurveillance* 1997, **2**:29-31.
12. Santoro R, Ruggeri FM, Battaglia M, Rapicetta M, Grandolfo ME, Annesi I and Cortellessa CM: **Measles Epidemiology in Italy** *International Journal of Epidemiology* 1984, **13**:201-209.
13. Finkenstadt BF and Grenfell BT: **Time series modelling of childhood diseases: a dynamical systems approach** *Appl Statist* 2000, **449**:187-205.
14. Ramsey M, Gay N, Miller E, Rush M, White J, Morgan-Capner P and Brown D: **The epidemiology of measles in England and Wales: rationale for the 1994 national vaccination campaign** *CDR Review* 1994, **4**:R141-R146.
15. Edmunds W John, Gay Nigel J, Kretzchmar M, Pebody RG and Wachmann H: **The pre-vaccination epidemiology of measles, mumps and rubella in Europe: Implications for modelling studies** *Epidemiology & Infection* 2000, **125**:635-650.
16. Sniadack DH, Moscoso B, Aguilar R, Heath J, Bellini W and Chiu MC: **Measles epidemiology and outbreak response immunization in a rural community in Peru** *Bulletin of the World Health Organization* 1999, **77**:545-552.
17. Livi-Bacci M and Breschi B: **Italian fertility - an historical account** *Journal of Family History* 1990, **15**:385.
18. US Census Bureau: **Statistical Abstract of the United States: 1993** 113th edition. Washington DC, US Government Printing Office; 1993.
19. Blackburn N, Schoub B and O'Connell K: **Reliability of the clinical surveillance criteria for measles diagnosis** *Bulletin of the World Health Organization* 2000, **78**:861.
20. Ciofi degli Atti ML, Salmaso S, Bella A, Arigliani R, Gangemi M, Brusoni G, Tozzi AE and Group SPES-Study: **Paediatric sentinel surveillance of vaccine-preventable diseases in Italy** *Pediatric Infectious Disease Journal* 2002, **21**:763-769.
21. Ciofi degli Atti ML, Rota MC, Mandolini D, Bella A, Gabutti G, Crovari P and Salmaso Stefania: **Assessment of varicella underreporting in Italy** *Epidemiology and Infection* 2002, **128**:479-484.
22. ICONA working group: **ICONA: national infant vaccination coverage** *Istituto Superiore di Sanità*; 1998.

### Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2458/3/23/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

