Validity of routine surveillance data: a case study on Swedish notifications of methicillin-resistant Staphylococcus aureus

M Stenhem (mikael.stenhem@smi.se), Å Örtqvist, H Ringberg, L Larsson, B Olsson-Liljequist, S Häggman, M Kalin, K Ekdahl, the Swedish study group on MRSA epidemiology

1. Department of Epidemiology, Swedish Institute for Infectious Disease Control, Karolinska Institutet, Solna, Sweden
2. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Solna, Sweden
3. Department of Medicine, Infectious Diseases Unit, Karolinska Institutet, Solna, Sweden
4. Department of Communicable Diseases Control and Prevention, Stockholm County Council, Stockholm, Sweden
5. Regional Center for Communicable Disease Control and Prevention, Skåne Region, Malmö, Sweden
6. Department of Hospital Hygiene, Sahlgrenska University Hospital, Göteborg, Sweden
7. Department of Bacteriology, Swedish Institute for Infectious Disease Control, Solna, Sweden
8. European Centre for Disease Prevention and Control (ECDC), Stockholm, Sweden
9. The members of the group are listed at the end of the article

Surveillance of communicable diseases is a public health cornerstone. Routine notification data on communicable diseases are used as a basis for public health action as well as for policy making. While there are agreed standards for evaluating the performance of surveillance systems, it is rarely possible to analyse the validity of the data entered into these systems. In this study we compared data on all Swedish cases of methicillin-resistant Staphylococcus aureus (MRSA) routinely notified between 2000 and 2003 with follow-up information collected for each of these cases as part of a public health project. The variables Reason for testing (clinical sample, contact tracing, screening of risk group), Clinical presentation (disease, colonisation), Transmission setting (healthcare-acquired, community-acquired), Country of acquisition (Sweden, abroad) and Risk-occupation (yes, no) were analysed for sensitivity, positive predictive value and completeness of answers. The sensitivity varied between 23% and 83%, the positive predictive values were generally higher (55% to 97%), while missing answers varied from 11% to 59%. The proportion of community-acquired cases was markedly higher when excluding either cases of MRSA colonisation or cases found on routine surveillance figures. Whether or not the case definition includes cases of MRSA colonisation may have a significant impact on population-wide estimates of MRSA occurrence.

Background

The overall aim of disease surveillance is to collect information for public health action. Disease control measures are costly both from a public health and from a healthcare perspective. For the healthcare sector, diseases that spread nosocomially are particularly expensive. Disease control actions become more efficiently focused when based on valid surveillance data. However, it is rarely possible to assess the validity of notification data [1,2], and to our knowledge such an assessment has never been reported for any methicillin-resistant Staphylococcus aureus (MRSA) surveillance system. The epidemiology of diseases, such as MRSA, that can be transmitted both by symptomless carriers and by individuals with clinical infection is complex and their analysis requires a level of detail that can rarely be obtained from routine surveillance data. In contrast to most other countries [3-7], a comparatively lower occurrence of MRSA has hitherto been reported from the Netherlands and the Nordic countries (Denmark, Finland, Iceland, Norway and Sweden) [8]. During the late 1990s, Sweden experienced a large regional outbreak of healthcare-associated MRSA cases, which was brought under control by resolute efforts [8,12,13]. The experience from this outbreak forms the basis for the active strategy against MRSA currently employed in Sweden, with extensive screening of risk groups and contact tracing around known cases (symptomatic cases as well as asymptomatic carriers), aiming at preventing further transmission of MRSA.

The low-endemic MRSA-situation in Sweden and the allocation of resources in the period from 2000 to 2003 to map the epidemiology of MRSA in Sweden in detail, made it possible to collect in-depth data on every case of MRSA notified in the country during that period. This was done in addition to the collection of routine surveillance data. The resulting detailed and unique dataset made it possible to fulfil the two aims of the present study, i.e. to analyse the quality of the data supplied within the routine surveillance system and to show how case finding activities and inclusion or exclusion of MRSA carriers in the case definition influenced the estimated occurrence of MRSA in the population.

Materials and methods

Material

In Sweden, cases of clinical MRSA infection as well as asymptomatic carriage are notified by law to the Swedish Institute for Infectious Disease Control (SMI). The notifications are made
in parallel by the clinicians who diagnosed the patients and the laboratories that identified the pathogens. All MRSA notifications referring to the same individual are merged into one case record at the SMI, using a unique personal identification number. Thus, only new cases of MRSA are counted in the notification system. In this study, we included all cases notified in the years from 2000 to 2003. MRSA isolates from these cases were also sent to the SMI, where the bacteriological diagnosis was confirmed using PCR for the nuc and the mecA genes and epidemiological typing was performed using pulsed-field gel electrophoresis (PFGE).

A prospective, active follow-up on the epidemiological investigation of each notified case was performed in addition to the routine passive surveillance. Once the epidemiological investigation of a case was completed, updated data were collected by MRSA contact persons in each of the 21 counties in Sweden, and entered into a national MRSA study database [8]. These contact persons were infection control and public health officers involved in the local public health work on MRSA, and as such had full access to all information on the MRSA cases.

Definitions for case data evaluation
We analysed a subset of the variables used in the notification forms. The variable Reason for testing defined the reason for taking the first bacteriological sample from which MRSA was isolated from a case, categorised as: a) clinical sample (sample taken for diagnostic purposes), b) contact tracing (sample taken from a contact of a diagnosed MRSA case in order to identify a transmission chain), or c) screening of risk groups (sample taken from a patient with an increased risk of having MRSA, e.g. with healthcare contacts abroad or clinical risk factors such as breakages of the skin barrier or urinary catheter). Clinical presentation was defined as a) disease or b) colonisation. Transmission setting was defined as a) healthcare-acquired (HA), b) community-acquired (CA) or c) unknown. To be considered as HA (including municipal care institutions such as nursing homes), a case would need to have been in contact with a healthcare setting where other MRSA cases with the same PFGE pattern had occurred. If MRSA cases had been in close contact with each other outside any healthcare setting (e.g. family members, child daycare, girl- or boyfriend, work colleagues, sport contacts) and the PFGE pattern did not contradict transmission, or if, in the absence of an epidemiological link, the PFGE pattern was known to occur in the community, the case was considered to be CA. When neither HA nor CA could be ruled out, the transmission setting was considered as unknown. For the purpose of this study, detailed information on Country of acquisition was broadly grouped as a) abroad (acquired outside Sweden), b) domestic (acquired in Sweden) or c) unknown. A notified case was considered as acquired abroad if the patient had been abroad within six months preceding diagnosis and had either an MRSA strain known to have occurred in that part of the world or a strain previously unknown in Sweden and a likely Swedish source could not be found. When neither domestic acquisition nor acquisition abroad could be ruled out, country of acquisition was entered as unknown. Work in healthcare institutions, municipal care facilities and day nurseries was considered a Risk occupation for acquisition of MRSA (answer categories a) yes or b) no).

Data analysis
We compared the information on the routine clinical notification form of each case, with the data in the study database. In case of several clinical notifications on the same case, the first one was used for the analysis. We calculated sensitivity (the percentage of information per variable in the validated study database that was supplied correctly on the clinical notification form) and positive predictive value (PPV, the percentage of information in the first

Table 1
Data from the notifications of MRSA cases in Sweden, 2000-2003 (n=1,733)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Variable category</th>
<th>Number of cases according to study database</th>
<th>Number of cases where notification data were in accordance with study database (sensitivity)</th>
<th>Percentage of cases where notification data were contradictory to study database</th>
<th>Percentage of cases where notification data were missing</th>
<th>Positive predictive value of notification data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk occupation</td>
<td>Yes</td>
<td>140</td>
<td>83% (76–89)</td>
<td>–</td>
<td>17% (11–24)</td>
<td>59% (51–66)</td>
</tr>
<tr>
<td>Country of acquisition</td>
<td>Domestic</td>
<td>1,265</td>
<td>69% (66–72)</td>
<td>7% (6–9)</td>
<td>24% (21–26)</td>
<td>96% (94–97)</td>
</tr>
<tr>
<td></td>
<td>Abroad</td>
<td>444</td>
<td>76% (72–80)</td>
<td>12% (9–16)</td>
<td>11% (8–15)</td>
<td>90% (87–93)</td>
</tr>
<tr>
<td>Clinical presentation</td>
<td>Disease</td>
<td>798</td>
<td>65% (62–68)</td>
<td>19% (16–22)</td>
<td>16% (13–19)</td>
<td>80% (76–83)</td>
</tr>
<tr>
<td></td>
<td>Colonisation</td>
<td>915</td>
<td>66% (63–69)</td>
<td>14% (12–16)</td>
<td>20% (18–23)</td>
<td>79% (76–82)</td>
</tr>
<tr>
<td>Transmission setting</td>
<td>Community-acquired</td>
<td>561</td>
<td>41% (37–45)</td>
<td>40% (36–44)</td>
<td>19% (16–22)</td>
<td>65% (60–70)</td>
</tr>
<tr>
<td></td>
<td>Healthcare-acquired</td>
<td>903</td>
<td>51% (48–54)</td>
<td>34% (31–37)</td>
<td>15% (13–18)</td>
<td>82% (79–85)</td>
</tr>
<tr>
<td>Reason for testing</td>
<td>Clinical sample</td>
<td>203</td>
<td>23% (17–29)</td>
<td>18% (13–24)</td>
<td>59% (52–66)</td>
<td>55% (44–66)</td>
</tr>
<tr>
<td></td>
<td>Contact tracing</td>
<td>437</td>
<td>18% (14–21)</td>
<td>24% (20–29)</td>
<td>35% (30–39)</td>
<td>97% (94–99)</td>
</tr>
<tr>
<td></td>
<td>Screening of risk groups</td>
<td>268</td>
<td>37% (25–37)</td>
<td>32% (27–38)</td>
<td>31% (25–37)</td>
<td>73% (65–80)</td>
</tr>
</tbody>
</table>

MRSA: methicillin-resistant Staphylococcus aureus.
All percentages are presented with 95% confidence intervals. In parentheses.
Clinical notifications were missing for 176 cases. For the variable Reason for testing the analysis was restricted to the 915 cases of MRSA colonisation, since this information was required in the notification form only for those cases.
notification that was in accordance with the information in the study
database), with exact 95% confidence intervals. We also analysed
the completeness of information on the first clinical notification
form. The statistical analyses were performed in Stata version 8.2.

Results
A total of 1,733 MRSA cases were reported during the study
period. Table 1 provides detailed information on each of the
variables in the first clinical notifications compared to the data in
the study database. It shows the sensitivity, the completeness, and
the predictive capacity of the information that public health officers
received in the first clinical notification, i.e., of the information
available for the initiation of public health actions.

Sensitivity of data in original notification
Of 140 cases with Risk occupations according to the study
database, 83% were correctly identified as such in the clinical
notification (Table 1). Sensitivity was also high for the variable
Country of acquisition, with 76% of patients with acquisition
abroad and 69% of patients with acquisition in Sweden correctly
identified in the notification. The sensitivity was low for the variable
Reason for testing, mainly due to missing information in the original
notification forms (see below).

Missing information in original notifications
Missing information for a variable (Table 1) was either due
to missing information for that question or due to the fact that the
clinical notification form was missing altogether. The most
complete variable category was Country of acquisition ‘abroad’: this
information was lacking in only 11% of cases that had acquired
MRSA abroad. Other categories for which the information
given in the first notification to a large extent was present were:
Transmission setting ‘healthcare’ (15% missing information),
Clinical presentation ‘disease’ (16% missing information) and Risk
occupation ‘yes’ (17% missing information). The most incomplete
information was found for the variable Reason for testing.

Positive predictive value of the information in the original
notification
The proportion of accurate information in the original notification
(PPV) was highest for the variable Country of acquisition, with a
PPV of 96% for domestic acquisition and of 90% for acquisition
abroad (Table 1). Least predictive was the information on ‘clinical
sample’ as Reason for testing with only 55% of the cases being
verified. The Transmission setting ‘community-acquired’ also had
a low PPV (65%).

Effect of case definition and method of case finding on estimated
MRSA occurrence
In order to assess the impact of different case definitions on the
distribution of reported MRSA cases, we analysed the variable
Transmission setting within the variables Clinical presentation and
Reason for testing according to the study database (Table 2). Overall,
32% of cases were CA and 52% HA. If only cases with
MRSA-caused disease (and not carriage) had been reported, the
proportion of CA and HA cases would have been 41% and 39%,
respectively. A similar effect on the distribution of cases was seen
when considering only cases diagnosed by cultures that had been
taken on clinical indication: the proportion of HA cases decreased
significantly (43%) and the proportion of CA cases increased (35).

Of the 1,733 cases in the study, 45 were identified through the
isolation of MRSA from blood cultures. Nine of these cases were
CA (20%; 95% CI 9.6-35) and 25 were HA (56%; 95% CI 40-70).
The proportion of CA cases among these was thus significantly lower
than among all clinical MRSA cases (Table 2).

Discussion and conclusion
Far-reaching decisions on public health interventions and policy,
as well as research studies, are based on routine surveillance data.
Surveillance data are also used to compare the disease occurrence
over time and between populations, e.g. when making international
comparisons between countries. When using surveillance data for
such purposes it is essential that the case definitions and measured
variables are valid and comparable. The project with the national
Swedish MRSA-database 2000-2003 provided us with a unique
opportunity to analyse the validity of routine surveillance case-
data in Sweden. There are accepted guidelines for the general
evaluation of public health surveillance systems [14], but such
guidelines do not cover the evaluation of the actual data entered
into the system and their validity – presumably because high quality
reference datasets rarely exist to compare routine surveillance data
against. The validity of notification data has been investigated for
other diagnoses such as tuberculosis and human immunodeficiency
virus infections/acquired immunodeficiency syndrome (HIV/AIDS)
[1,2], but we are not aware of any report on MRSA surveillance
and data validity. The general sensitivity of the Swedish statutory
surveillance system to detect patients diagnosed with a notifiable
disease has recently been analysed and was found to be very high
-well above 90% [15].

<p>| MRSA cases notified in Sweden between 2000 and 2003, according to the validated case information, comparing the proportion of community- and healthcare-acquired cases within the variable categories of Clinical presentation and Reason for testing (n=1,733) |
|-----------------|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>Variable</th>
<th>Variable category</th>
<th>Community acquired</th>
<th>Healthcare acquired</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease</td>
<td>(n=798)</td>
<td>41% (37–44)</td>
<td>39% (35–42)</td>
</tr>
<tr>
<td>Colonization</td>
<td>(n=915)</td>
<td>25% (23–28)</td>
<td>64% (60–67)</td>
</tr>
<tr>
<td>Clinical sample</td>
<td>(n=472)</td>
<td>35% (32–38)</td>
<td>43% (40–46)</td>
</tr>
<tr>
<td>Contact tracing</td>
<td>(n=61)</td>
<td>43% (38–48)</td>
<td>51% (47–56)</td>
</tr>
<tr>
<td>Screening</td>
<td>(n=36)</td>
<td>8% (5–12)</td>
<td>81% (76–85)</td>
</tr>
<tr>
<td>Total</td>
<td>(n=1,733)</td>
<td>32% (30–35)</td>
<td>52% (50–54)</td>
</tr>
</tbody>
</table>

MRSA: methicillin-resistant Staphylococcus aureus
The total number of cases is given for each variable category. Where cell
numbers do not add up to the total of rows or columns, the difference is
due to cases that did not fall under any of the categories.
*Percentages per variable category with exact confidence intervals
Pathogens like MRSA, which are able to colonise individuals as well as cause clinical disease, are particularly challenging for a surveillance system. Patients with clinical disease are more likely to seek healthcare and consequently more likely to be diagnosed and notified. The probability that a colonised individual is diagnosed and notified depends on the vigil in which case finding activities (contact tracing and screening of risk groups) are carried out. The incidence figures presented for different populations would therefore not be comparable if the proportions of colonised individuals identified through case finding activities differed, unless information on clinical presentation and/or reason for testing is specified. It has earlier been noted that differences in reported MRSA incidences can be a result of differences in case finding methods in neighboring health-districts in England [16] as well as between hospital and community populations in an area of Manhattan, New York [17]. Studies of MRSA occurrence often include MRSA carriers [4,8,9,16-18]. To make a comparison valid, investigators need to characterise the cases for the closely interrelated variables Reason for testing and Clinical presentation (disease or colonisation), but this information is often not presented [8,11,18]. Simor et al. suspected an association between screening and colonisation among older MRSA patients in the Canadian Nosocomial Surveillance Program (CNSP) [19], but our study is to our own knowledge the first one to systematically address the effect of case finding on the incidence estimates of MRSA within a complete population on a national level.

The problem presented by an unknown proportion of carriers can be avoided by restricting the case definition to clinical infections only, or even to blood isolates only. In our study, less than two thirds of cases with MRSA colonisation and of cases with MRSA disease were shown to be correctly classified with regards to Clinical presentation. These findings indicate that the MRSA incidence would have been severely biased, if only MRSA disease had been notifiable. If only blood isolates were reported, such misclassification would be less likely. The rationale behind such an approach is that cases found through blood isolates act as a marker for the overall burden of MRSA [20]. Restricting the case definition in this way might however result in a biased estimate of the MRSA occurrence in the general population, as several studies found an association of MRSA bacteraemia with healthcare exposure [21,22]. This is substantiated by our study, in which the proportion of CA cases was significantly lower among those identified by blood culture compared to all cases with MRSA disease. A further advantage of considering all available MRSA cases is the increased statistical power and precision that comes with a larger number of study subjects. In smaller populations, such as single hospitals, this approach may be advantageous even in a high-endemic country like the United Kingdom [23]. Moreover, both MRSA carriers and those infected with MRSA are possible sources for further transmission in the population. From a point of view of MRSA control, a surveillance system should therefore include carriers. Our view is that ideally, all cases of MRSA, colonisation or disease, should be accounted for (provided there is a systematic case finding for colonised cases), along with data on the clinical presentation and/or the reason for testing, so that the analysis and interpretation of the figures can be adjusted accordingly. Public health-initiated case finding is carried out in situations where transmission is known to be high. Not monitoring cases from these settings, which generate a considerable number of new cases, is to neglect an important part of MRSA occurrence. How the surveillance of MRSA and other organisms that both colonise and cause disease is organised also depends on a number of other factors, such as the scope and level of the surveillance (e.g. hospital, district, regional or national), whether it is done in a high-endemic or low-endemic setting, and the available resources.

In conclusion, the present study clearly showed how differences in case definitions can influence the estimated number of MRSA cases categorised as healthcare-acquired or community-acquired, as well as the overall reported MRSA incidence. If carriers are included in the case definition, the overall occurrence and distribution of cases will also depend on the extent of the efforts to control MRSA through contact tracing and screening. We could identify considerable flaws in the quality of case data from routine notifications, e.g. misclassification of cases as colonisation or disease. Consequently, restricting the case definition to clinical cases only, would not be a reliable way to estimate the occurrence and distribution of MRSA. Surveillance systems and population-based epidemiologic studies thus need to specify the proportion of carriers and the reason for testing. This will also increase comparability of figures between countries or regions and between different points in time. Data validity cannot be taken for granted in a surveillance system, but needs to be ensured. For data that ultimately rely on information about transmission chains and results of epidemiological typing, the information should ideally be collected after the completion of the epidemiologic investigation of the cases.

Acknowledgements
The authors wish to thank Sharon Kühlmann-Berenzon for valuable statistical advice.

The Swedish study group on MRSA epidemiology:
A. Andersson, Vastmanland County; IA von Rosen, Jamtland County; B. Ardung, Uppland County; JS Eklumo, Kalmar County; G. Fridh, Värmland County; S. Hedin, Gävleborg County; P. Helgósjo, Västernorrland County; B. Isaksson, Östergötland County; K. Karlman, Norrbotten County; I. Kransberg, Blekinge County; A. Lennel, Örebro County; A. Lindahl, Dalarna County; G. Lindin, Stockholm County; P. Lundholm, Västerbotten County; I. Nilsson, Södermanland County; M. Nilsson, Kronoberg County; P. Nilsson, Gotland County; P. Nilsson, Halland County; UB Thelholm, Stockholm County; PE Åbom, Jönköping County.

References


This article was published on 30 July 2009.