Infectious Diseases

Surveillance for Rare Infectious Diseases: is one passive data source enough for Haemophilus influenzae?

Astrid Milde-Busch¹, Helen Kalies¹, Simon Rückinger¹, Anette Siedler², Joachim Rosenbauer³, Rüdiger von Kries¹

Background: The completeness of a compulsory reporting system of systemic Haemophilus influenzae infections in children in Germany is studied by means of cross-linking registry data from three sources and applying capture-recapture methods. Methods: Cases were collected for the years 2001–05 by three national data sources: a passive administration registry (SurvNet@RKI), an active hospital surveillance system and an active laboratory surveillance system. The case definition required cultural detection of H. influenzae in blood or cerebrospinal fluid. Linkage was carried out by month and year of birth, sex, geographical region and date of disease onset. Capture-recapture models were used to estimate the incidence of invasive H. influenzae infections. Results: SurvNet@RKI reported 113 H. influenzae and 38 H. influenzae type b (Hib) cases, compared to a total of 231 and 68 cases, respectively, reported by all three sources combined. Best-fitting 3-source capture-recapture estimations amounts to 258 (95% confidence interval: 247–276) H. influenzae and 71 (69–74) Hib cases. SurvNet@RKI data depicted a similar decrease in annual H. influenzae cases as the capture-recapture estimates but failed to detect the underlying decrease in Hib cases which was observed in the capture-recapture estimates due to a considerable annual variability of ascertainment of serotyped cases in SurvNet@RKI ranging from 14% to 69%. Conclusions: Because of small variability of ascertainment, the compulsory passive reporting system depicted trends in H. influenzae incidence, although less than half of the cases were ascertained. However, time trend in Hib cases could not be depicted, because of highly variable serotyping proportions.

Keywords: completeness, Haemophilus influenzae, Hib, passive surveillance system

Introduction

Vaccines represent one of the most powerful weapons of public health and are routinely used since many years with dramatic impact: Smallpox has been eradicated; poliomyelitis is likely to be eliminated in few years.¹² For other diseases, vaccination has to be maintained in order to keep the attack rates low.

However, even with vaccination programs in place, infections may re-emerge because of failures of the vaccine or the vaccination program. Although the introduction of conjugate vaccines against Haemophilus influenzae type b (Hib) in the 1990s prompted dramatic decreases in the incidences of Hib-related infections,³–⁸ a recurrence in vaccinated children and adolescents has been observed in Great Britain and The Netherlands.⁹–¹³ These examples underline the necessity for a continuous surveillance on systemic H. influenzae infections which have become a rare disease with <50 invasive Hib cases in infants younger than 16 years annually in Germany.¹⁴

For rare diseases, however, even small changes in absolute case numbers may indicate relevant changes in the infectious pattern. Case ascertainment therefore has to be as complete as possible to detect critical trends in the incidence. In most countries, notifiable disease surveillance systems rely on mandatory, mostly passive, case reporting by physicians and laboratories.

With the Protection against Infection Act,¹⁵ invasive H. influenzae infections became notifiable in Germany in 2001. Laboratories have to report invasive H. influenzae cases isolated from blood or cerebrospinal fluid (CSF) to the local health departments. An electronic reporting system for surveillance of notifiable infectious diseases (SurvNet@RKI) allows for information flow between local, state and federal institutions.¹⁶,¹⁷

How complete is this passive mandatory reporting system? Is case ascertainment constant over time? Does case ascertainment with passive reporting allow for depicting secular trends?

Methods

Case definition

Definition of invasive H. influenzae cases required hospitalization for a systemic infectious disease (meningitis, pneumonia, epiglottitis, septicemia, cellulitis, arthritis) in a paediatric hospital in Germany and isolation of H. influenzae from a normally sterile body site (blood or CSF).

Data sources

Cases of invasive H. influenzae infections in children younger than 16 years were collected from three sources: mandatory notification data provided by SurvNet@RKI, a hospital-based surveillance system (Hospital-ESPED) and a laboratory-based surveillance system (Laboratory-ESPED) over the period 2001–05. While the first is a passive system, the latter two
require active reporting: the absence of cases has to be reported as well and reminders for missing reports are included on a monthly basis.

The Robert-Koch-Institute (RKI), the central national agency for infectious disease epidemiology in Germany, provided data on notified H. influenzae cases in SurvNet@RKI up to 5 December 2006. Laboratories are obliged by law (Protection against Infection Act) to report all cases of invasive H. influenzae, isolated from blood or CSF, to one of the local health departments, according to site of resident of the patient. The local health department asks for additional information on the patient, such as clinical symptoms and vaccination information, and traces contacts for possible post-exposure prophylaxis. If the report meets the case definition anonymised case-based data are electronically submitted to the respective state health department. From there, data are submitted to the RKI by the data system SurvNet@RKI for national processing. The notification data at RKI contain: a unique identification case number, month and year of birth, sex, federal state, county, date of disease onset (not always given), date of notification, sampling material (blood or CSF), bacteriologic serotype (if known) and clinical symptoms. Furthermore, information on vaccination is reported if available (e.g. number of doses, date of last vaccination and vaccine type).

In Hospital-ESPED, physicians from all paediatric hospitals in Germany voluntarily report incident cases of a number of rare conditions on a monthly basis since 1992 to a central study office. Reporting on H. influenzae was included since 1998. The reports provide information on: an identification number for Hospital-ESPED, location of the reporting hospital, date of birth, age at disease onset, sex, date of disease onset, bacteriologic serotype, detection material, anamnestic and clinical data, i.e. underlying disease and outcome as well as vaccination information.

For isolation of respective pathogens paediatric hospitals send specimen of body tissue to microbiological laboratories. From 1998 to 2006, all laboratories throughout Germany serving the paediatric hospitals were asked to report voluntarily detection of H. influenzae in physiologically sterile materials in children to Laboratory-ESPED, a study group at RKI that was working in parallel to the Hospital-ESPED system. The data include comparable data to Hospital-ESPED, location of the reporting hospital, date of birth, age at disease onset, sex, date of disease onset, bacteriologic serotype, detection material, anamnestic and clinical data, i.e. underlying disease and outcome as well as vaccination information.

In single cases, variations of even more days have been accepted (e.g. when disease onset occurred around holidays). Matching on all four criteria was required or, in case of only three matching criteria, at least one out of detection material and disease or symptoms, respectively, was matching.

In case of non-corresponding serotypes reported, data from Laboratory- and Hospital-ESPED was considered as the gold-standard.

**Capture–recapture estimates**

The idea of the capture–recapture (CRC) method is to ‘...equate the unknown proportion of a population recorded in a list (unknown because the population size is unknown) to the known proportion of one or more subpopulations also recorded in that list...’. Thereby, the overlap between reports of different sources is used to calculate the number of cases missed by any of the sources and allows for calculating the total number of expected cases. However, the application of the CRC method depends on the following assumptions: (i) a closed study population, (ii) independence of sources (in case of 2-source CRC) and (iv) homogeneity within sources (for a single source, each case has the same probability of ascertainment). While the first two criteria are mostly fulfilled in epidemiological studies, problems may arise with the last two assumptions. Source independency means that the probability of notification in one source is independent of notification in the other sources. Violations in homogeneity result, for example, from differences in case-finding strategies (e.g. source 1 registers primarily individuals from one special sub-population, while source 2 registers primarily individuals from another sub-population). With two sources, the number of cases only reported by source 1, the number of cases only reported by source 2 and the number of cases reported by both sources can be identified. With these data, the number of missed cases can be estimated under the assumption of independence of the two sources.

Independence of sources is, however, not required for 3-source CRC. In particular, pair-wise dependencies of sources can explicitly be modelled within a log-linear model approach for CRC; separate models have to be calculated for each type of supposed source interaction. Different assumptions of source dependencies are modelled based on 3-source CRC systems. These models can be compared by the likelihood-based Akaike Information Criterion (AIC). The best fitting model is indicated by the lowest value of AIC.

Capture–recapture estimates have been calculated separately for each year as well as taken together for the total study period, for all H. influenzae and Hib cases, respectively.

**Further statistical analyses**

Completeness of each source is estimated by dividing the number of reported cases in each source by the estimate of the 3-source CRC procedure of the total population.

In order to assess time trends, analysis of the incident cases as a function year of reporting data was performed.

Statistical analyses have been performed with the SAS software package (version 9.1, SAS Institute Inc. Cary, NC, USA).

**Results**

**Number of case reports**

Table 1 presents a summary of the number of reported cases stratified by sources. In SurvNet@RKI, 113 H. influenzae cases have been reported from 2001 to 2005. The total number of reports in either source was 231 cases. Forty-five of the SurvNet@RKI cases (39.82%) were reported by all three
Table 1 Distribution of all reported invasive H. influenzae cases and Hib cases (proportion of Hib cases of all reported H. influenzae cases) in the three surveillance systems in Germany between 2001 and 2005

<table>
<thead>
<tr>
<th>Cases reported by</th>
<th>H. influenzae cases</th>
<th>Hib cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>SurvNet@RKI</td>
<td>Hospital-ESPED</td>
<td>Laboratory-ESPED</td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Total number of cases</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AIC, Akaike Information Criterion; NCRC, capture-recapture estimate of the total number of cases, with 95% CI; S-H, assumed dependency between SurvNet@RKI and Hospital-ESPED; S-L, assumed dependency between SurvNet@RKI and Laboratory-ESPED; H-L, assumed dependency between Hospital-ESPED and Laboratory-ESPED.

Table 2 Log-linear models fitted to three sources of the number of H. influenzae and Hib infections

<table>
<thead>
<tr>
<th>Model</th>
<th>Interactions</th>
<th>AIC</th>
<th>NCRC (95% CI)</th>
<th>Completeness of SurvNet@RKI (95% CI)</th>
<th>Completeness of Hospital-ESPED (95% CI)</th>
<th>Completeness of Laboratory-ESPED (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No</td>
<td>–0.6922</td>
<td>250 (244–259)</td>
<td>45% (44–46%)</td>
<td>49% (47–50%)</td>
<td>73% (70–75%)</td>
</tr>
<tr>
<td>2</td>
<td>S-H</td>
<td>1.3483</td>
<td>250 (243–260)</td>
<td>45% (43–47%)</td>
<td>49% (47–50%)</td>
<td>73% (70–75%)</td>
</tr>
<tr>
<td>3</td>
<td>S-L</td>
<td>1.3683</td>
<td>251 (243–265)</td>
<td>45% (43–47%)</td>
<td>49% (46–50%)</td>
<td>73% (69–75%)</td>
</tr>
<tr>
<td>4</td>
<td>H-L</td>
<td>–1.7641</td>
<td>258 (247–276)</td>
<td>44% (41–46%)</td>
<td>47% (45–49%)</td>
<td>71% (66–74%)</td>
</tr>
<tr>
<td>5</td>
<td>S-H, S-L</td>
<td>3.3067</td>
<td>250 (242–266)</td>
<td>45% (42–47%)</td>
<td>49% (46–50%)</td>
<td>73% (68–75%)</td>
</tr>
<tr>
<td>6</td>
<td>S-H, H-L</td>
<td>0.2117</td>
<td>259 (246–281)</td>
<td>44% (40–46%)</td>
<td>47% (43–50%)</td>
<td>70% (65–74%)</td>
</tr>
<tr>
<td>7</td>
<td>S-L, H-L</td>
<td>–1.5967</td>
<td>275 (249–340)</td>
<td>41% (33–45%)</td>
<td>44% (36–49%)</td>
<td>66% (54–73%)</td>
</tr>
<tr>
<td>8</td>
<td>S-H, S-L, H-L</td>
<td>0</td>
<td>285 (249–388)</td>
<td>40% (29–45%)</td>
<td>43% (31–49%)</td>
<td>64% (47–73%)</td>
</tr>
</tbody>
</table>

In SurvNet@RKI, 26 of the 68 Hib cases have been reported as estimated from the best-fitting 3-source CRC model for the years 2001–05. These data indicate that the annual number of reported and estimated H. influenzae cases remains at a very low level.

**Time trends in H. influenzae infections**

Reported and estimated case numbers have been analyzed as a function of year of disease onset. Figure 1 presents the annual number of H. influenzae cases reported by SurvNet@RKI and as estimated from the best-fitting 3-source CRC model for the years 2001–05. These data indicate that the annual number of reported and estimated H. influenzae cases remains at a very low level. However, some variability can be observed, with a trend for decreasing case numbers over time detectable for the CRC estimations and, equally, for the SurvNet@RKI reports. The proportion of cases reported in SurvNet@RKI compared to the CRC estimation was roughly constant over time, varying between 33% and 49% between years.

**Capture–recapture estimation on incidence of Hib infections**

Similarly as for the overall H. influenzae cases, incidence of Hib was estimated by 3-source CRC, based on all reported Hib sources. In either of the other sources, 47 of the SurvNet@RKI cases (41.59%) have been reported. Twenty-one cases (18.58%) have only been reported by SurvNet@RKI.

For 133 of the 231 reported H. influenzae cases (57.58%), bacteriologic serotype was assessed, with 68 proven Hib cases (29.44% of all reported cases; 51.13% of all typed cases) and 65 none-type b cases (28.14% of all reported cases; 48.87% of all typed cases). For four further individuals (1.73%), serotype could not be determined, and for 94 cases (40.69%), no serotype data were available. See table 1 right column for the distribution of Hib reports in the various sources.

In SurvNet@RKI, 26 of the 68 Hib cases have been reported by 3-source CRC, based on all reported Hib cases. 27 cases [95% confidence interval (CI): 16–45] of invasive H. influenzae infections have not been registered in any of the three sources in the period 2001–05. The total number of cases would, therefore, amount to 258 (95% CI: 247–276); of these cases, 44% [95% CI: 41–46%] were reported in SurvNet@RKI.
cases, even if serotype information was missing from one or two sources. Surprisingly, the best-fitting model was not the same as for the overall number of *H. influenzae* cases, but rather Model 1, the one assuming no source dependencies (see Table 2, lower part). Based on this model, a total of 71 (95% CI: 69–74) Hib cases would have been expected, which corresponds to a reporting deficit of three (95% CI: 1–6) individuals. Calculations basing on Model 1 yield source completeness of 54% (95% CI: 51–55%) for SurvNet@RKI.

**Development of Hib infections over time**

Figure 2 presents trends of the Hib incidence, separately for serotyped SurvNet@RKI cases, all SurvNet@RKI cases and the results of Model 1. It shows a clear decreasing trend for CRC estimations over time, but rather unsystematic variation for the SurvNet@RKI reports. Completeness of SurvNet@RKI for serotype Hib cases varied between 20% and 56% for Hib cases between years, improving over time.

**Discussion**

The first major finding of the present investigation is that neither of the surveillance systems implemented in Germany to assess *H. influenzae* infections is complete. Application of the CRC method indicates that only 44% of the estimated cases of all invasive *H. influenzae* infections and 54% of estimated Hib cases have been caught by mandatory surveillance (SurvNet@RKI) between 2001 and 2005.

The availability of data from three sources allows adjusting CRC calculations for underlying pair-wise dependency between reports, providing estimations for all possible potential source dependencies. The three sources used in this article appeared to be widely independent for ascertainment of Hib cases. Although the best-fitting model for *H. influenzae* was a model assuming interdependency between Hospital- and Laboratory-ESPED, the incidence estimate for this model was only marginally different from the other models.

Secondly, it was shown that SurvNet@RKI reports for all invasive *H. influenzae* cases and the respective CRC estimates follow the same temporal trends and that completeness of SurvNet@RKI was fairly robust between years, with a supposed outlier in 2002 with only 33% coverage. Thus, even though coverage is not satisfactory, SurvNet@RKI seems to be sufficient to map secular trends in infectious pattern, as long as completeness remains stable. Therefore, SurvNet@RKI reporting can be regarded as an indicator for the temporal development of *H. influenzae* incidence.

However, the third major finding suggests that SurvNet@RKI reports are not sufficient to map development of Hib incidence correctly over time. Two potential reasons can be pointed out: (i) completeness of SurvNet@RKI for Hib cases varies between 20% and 56% between years and (ii) the proportion of cases with serotype information is generally low for SurvNet@RKI (48 of 113 cases: 42%) and varied between 14% and 69% between years (data not shown). Therefore, SurvNet@RKI reporting for Hib is unlikely to suffice for monitoring of Hib incidence.

**Strength and limitations**

A major strength of this study is the availability of three sources, each with nationwide coverage, to judge interdependencies between surveillance systems and to estimate the incidence of *H. influenzae* and Hib infections over a five year time period. This allowed for a valid evaluation of the ‘routine’ passive reporting system. Such single passive surveillance systems constitute the only basis for monitoring *H. influenzae* and Hib in many other European countries, mostly due to economical reasons. However, such passive surveillance systems are less complete than active systems, a finding that could also be replicated with the present study with 54%, 63% and 76% coverage for Hib cases for SurvNet@RKI, Hospital-ESPED and Laboratory-ESPED, respectively.

To our knowledge there are no other studies elaborating the limitations of 1-source passive surveillance of Hib in a similar manner. Although it is not unexpected that inconsistent serotyping and reporting rates will account for problems in monitoring, this has not yet been demonstrated for specific 1-source surveillance systems. Regarding Hib, almost complete or at least time-constant case ascertainment is essential, because (i) failures of the vaccination program have occurred...
in the past10–13 and (ii) small changes in absolute numbers may already indicate a significant hint for problems in the vaccination program, since the absolute numbers for Hib in vaccinated populations are low—even in a population of considerable size such as Germany with some 80 million inhabitants.

A limitation of our data lies in small numbers of cases, which is an inherent problem of surveillance of rare diseases. Therefore, we could not apply formal test statistics to confirm the differences in variability of reporting between the *H. influenzae* and Hib cases.

**Conclusions**

Although a single passive mandatory surveillance system appeared to be sufficient to monitor secular trends in overall invasive *H. influenzae* infections pattern, this was not the case for Hib infections. Reasons for this difference are likely to be related to variability in completeness of assessment and, to a greater extent, to variable serotyping rates. Since Hib is the *H. influenzae* species in focus of surveillance to monitor the effect of Hib vaccination programs, it is crucial that passive systems, which are likely to be fraught with underreporting, maintain not only constant ascertainment rates, but also constant serotyping rates.

**Acknowledgements**

The authors would like to thank all reporting paediatricians and laboratories for their time and efforts. The help of Beate Heinrich in keeping the surveillance program running was invaluable. We thank Britta Gründahl who delivered ESPED serotype data and Hermann Claus who collected SurvNet@RKI data.

**Conflicts of interest:** This study is part of a project evaluating vaccine effectiveness against *H. influenzae* type b which is financially supported by Sanofi Pasteur MSD and GlaxoSmithKline Biologicals (A.M.-B., H.K.).

**Key points**

- A single passive reporting system appears to be sufficient to monitor secular trends of *H. influenzae* incidence.
- To monitor trends in Hib incidence, a combination of multiple independent data sources is necessary.
- Completeness and consistence of reporting systems and, especially, the serotyping rate should be increased to improve reliability of surveillance systems.

**References**


Received 12 November 2007, accepted 13 March 2008