We present a pilot validation study performed on 10 European Union (EU) Member States, of a prevalence survey (PPS) of healthcare-associated infections (HAIs) and antimicrobial use in Europe in 2011 involving 29 EU/European Economic Area (EEA) countries and Croatia. A total of 20 acute hospitals and 1,950 patient records were included in the pilot study, which consisted of validation and inter-rater reliability (IRR) testing using an in-hospital observation approach. In the validation, a sensitivity of 83% (95% confidence interval (CI): 79–87%) and a specificity of 98% (95% CI: 98–99%) were found for HAIs. The level of agreement between the primary PPS and validation results were very good for HAIs overall (Cohen’s κappa (κ): 0.81) and across all the types of HAIs (range: 0.83 for bloodstream infections to 1.00 for lower respiratory tract infections). Antimicrobial use had a sensitivity of 94% (95% CI: 93–95%) and specificity of 97% (95% CI: 96–98%) with a very good level of agreement (κ: 0.93). Agreement on other demographic items ranged from moderate to very good (κ: 0.57–0.95): age (κ: 0.95), sex (κ: 0.93), specialty of physician (κ: 0.87) and McCabe score (κ: 0.57). IRR showed a very good level of agreement (κ: 0.92) for both the presence of HAIs and antimicrobial use. This pilot study suggested valid and reliable reporting of HAIs and antimicrobial use in the PPS dataset. The lower level of sensitivity with respect to reporting of HAIs reinforces the importance of training data collectors and including validation studies as part of a PPS in order for the burden of HAIs to be better estimated.

Introduction

In 2011, the European Centre for Disease Prevention and Control (ECDC) initiated the first European point prevalence survey (PPS) of HAIs and antimicrobial use in acute care hospitals [1] involving 29 European Union (EU)/European Economic Area (EEA) countries and Croatia. The objective was to estimate the total burden (prevalence) of HAIs and antimicrobial use in European acute care hospitals.

A pilot validation study was undertaken in the first phase of this PPS in 2011 with two major objectives: (i) to test the sensitivity and specificity of reporting HAIs and antimicrobial use and the level of agreement between primary validation data collectors, whereby this constituted the validation component of the study; (ii) to test the inter-rater reliability (IRR) of hospital data collectors across Europe.

This paper focuses on the aggregated results for several EU Member States of this pilot validation study. Ten EU Member States took part in the validation component (Bulgaria, Finland, Germany, Hungary, Italy, Latvia, Lithuania, Poland, Spain, United Kingdom) and eight of these countries in the IRR component (Bulgaria, Finland, Germany, Italy, Lithuania, Poland, Spain, United Kingdom).

Method

The sample size for the pilot study was calculated to produce validation results overall for the European PPS rather than at individual country specific level. A pilot ECDC PPS had indicated a prevalence of 7.1% and
an average hospital size of 300 beds [2]. The sample size for the pilot validation study was calculated at approximately 2,000 patients for an estimated sensitivity of reporting HAIs of 80% with a precision of +/-5% and a prevalence of 7% [3,4]. These 2,000 patients were specified by the ECDC validation pilot protocol as approximately 200 patients per country within the 10 participating countries, sampled in at least two hospitals per country.

Two approaches, including a validation method by an external validation team (method 1) and an on-site assessment of the IRR of different hospital PPS data collectors (method 2), were taken in order to address the objectives. The methods are summarised here and a full description is available in the ECDC pilot validation protocol [4].

Validation
A standard ECDC protocol was used by all countries [3]. Each country collected data on 100 patient records from each of two hospitals. The hospitals and the patient records were chosen by the national coordinators from each country and not randomly allocated at a country level. A number of approaches were taken including retrospective, simultaneous same day, simultaneous same time, blind and unblind data collection. The approach undertaken by each country was purposively selected dependent on timing of the primary PPS and availability of resources. Countries also had an option of oversampling within the protocol, whereby the number of HAIs in the validation sample was increased on purpose to increase the precision of the specificity estimation, by selecting wards with higher prevalence (e.g. intensive care units) in blind validation or by including all HAI cases detected in the primary PPS in unblind validation.

The validation findings were considered the ‘gold standard’ (true positives and true negatives) as the validation team consisted of at least one trained expert from (and/or acting on behalf of) the national/regional PPS coordinating centre (external to the validated hospital), using the ECDC-PPS protocol and codebook [5] and accompanied by a hospital staff member for the purposes of access and orientation.

Identical data to the primary data collector were collected by the validator using one or more of the approaches outlined above. Patient notes, nursing notes, hospital information systems and clinical ward personal were the data sources used.

From the validation dataset, the positive predictive value (PPV) was calculated as the percentage of patients with true HAIs (or patients receiving antimicrobial as appropriate) among all positive patients in the primary dataset, and the negative predictive value (NPV) as the percentage of true negative cases among all patients identified as negative in the primary sample. The results of the validation were applied to the aggregated primary data by multiplying the number of all positive cases in the primary sample by the PPV to obtain an approximation of the number of true positives to account for potential differences in prevalence due to oversampling. The same procedure was performed for negative cases with the NPV. This allowed determination of the sensitivity and specificity for the primary sample [4,6]. 95% confidence intervals (CI) were calculated using a continuity-corrected version of the Wilson’s score method. They were evaluated as ‘worst case’ instances using a combination method. The effects of omitting these adjustments, in most cases did not result in major differences to the results presented here.

On site assessment of inter-rater reliability
Five HAI-positive and 10 HAI-negative patient records were selected from a single setting, i.e. intensive care unit (ICU), where the prevalence of infection was the highest, or, if access to the ICU was restricted, in a limited number of other wards with expected high HAI prevalence, such as high dependency units.

Between two and five hospital primary PPS data collectors gathered data at an agreed time in the selected ward/setting in turn with the national contact point (validator). A procedure was followed to minimise any potential bias inclusive of the other rater(s) waiting in another room or at a distance where the reproducibility process could not be heard (e.g. use music in the waiting room/area). Data items were collected as detailed in method 1 and agreement between the data collections was analysed using kappa (κ) statistics (0.81–1.00 is very good, 0.61–0.80 is good, 0.41–0.60 is moderate, 0.21–0.40 is fair/marginal, <0.2 is poor; negative values are possible and also denote ‘poor’ agreement [7-9]. κ statistics were also reported for certain variables of the validation approach (method 1), as it can be argued that the external validation team does not truly represent a gold standard for HAIs and variables such as the McCabe score.

Results
Validation
The primary data set that originated from the 20 hospitals in the 10 participating countries comprised 3,958 patient records. Among these, the prevalence of HAIs was 9% (367 patients) and the prevalence of antimicrobial use was 38% (1,504 patients). Validation data were collected from October to December 2011. Of the 3,958 primary patient records, a total of 1,950 were selected for validation in accordance with the calculated sample size. Of those, 1,912 were matched to the primary dataset, since it was not possible to link all patient records due to errors in data entry or missing data. The reported prevalence of HAIs in the matched validation dataset was 12% (233 patients) and the prevalence of antimicrobial use was 46% (878 patients). Due to oversampling in the validation dataset, the prevalence of HAIs in this dataset was significantly higher than in
the ‘primary’ set, $X^2(1) = 7.7$, $p = 0.005$. The prevalence of antimicrobial use was also higher in the validation dataset, $X^2(1) = 27.5$, $p < 0.001$. Four of the ten countries included oversampling.

There was 97% agreement ($\kappa = 0.81$) between primary records and the validation records for the presence of an HAI (Table 1). The level of agreement was very good across all the most important types of HAIs, ranging from $\kappa = 0.83$ for bloodstream infections to $\kappa = 1.00$ for lower respiratory tract infections (Table 2). Specificity of reporting HAIs was 98% (95% CI: 98–99%) with sensitivity comparatively lower at 83% (95% CI: 91–95%) (Table 1). The sensitivity by type of HAI ranged from 83% for bloodstream infections to 100% for lower respiratory tract infections, with specificity values higher than 99% for all types of HAI (Table 2).

Very good results (96% agreement, $k = 0.91$) were achieved with respect to the recording of overall antimicrobial use (Table 3). Sensitivity and specificity were both very high at 94% (95% CI: 92.9–95.3%) and 97% (95% CI: 96.1–97.5%) respectively. Validation of the route of antimicrobial administration oral, parenteral showed that oral antimicrobials were frequently reported as parenteral, resulting in a lower specificity for the parental route and a lower sensitivity for the oral route (Table 4).

At the individual variable level, some variation was noted. Agreement on basic demographic variables was very good: age ($\kappa = 0.95$), sex ($\kappa = 0.93$), specialty of physician ($\kappa = 0.87$). A high level of agreement was also found with respect to the presence of invasive devices, although specificity for the presence of peripheral vascular catheters (93%; 95% CI: 91–95%) was noted to be significantly lower than that of central venous catheters (99%; 95% CI: 98–99%). Variables which required more interpretation such as McCabe score had a moderate score ($\kappa = 0.57$).

### Inter-rater reliability

Eight of ten countries participated in the IRR component of the pilot study with a total of 44 raters across all the participating hospitals, rating 195 patient records. An analysis of IRR by selected variables was undertaken on the dataset. Variables were selected on the basis of their importance and the frequency of reporting in the dataset. Analysis of IRR overall showed a very good level of agreement ($\kappa = 0.92$) for both the presence of HAIs (96%) and antimicrobial use (97%) (Table 5). There was very good IRR ($\kappa = 0.8$) for most of the PPS variables (with the exception of HAI origin) ($\kappa = 0.31$).

### Discussion

Studies on validation of national HAI surveillance are rarely published, and when they are, a variety of approaches are described, according to a recent review in the United States (US) [10]. In that review, of those that included either a validation or an IRR study, the results were varied, underscoring the need for PPS to include validation studies to add confidence to...
Table 2
Validation of the point prevalence survey for healthcare-associated infections (HAIs), by type of HAI, 10 European Union Member States, 2011 (n=1,912 patient records)

<table>
<thead>
<tr>
<th>Types of HAI</th>
<th>N</th>
<th>Sensitivity % (95%CI)</th>
<th>Specificity % (95%CI)</th>
<th>PPV % (95%CI)</th>
<th>NPV % (95%CI)</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>All HAIs</td>
<td>233</td>
<td>83.1 (78.7–86.7)</td>
<td>98.2 (97.2–98.6)</td>
<td>82.8 (77.4–87.4)</td>
<td>98.3 (97.5–98.8)</td>
<td>0.81</td>
</tr>
<tr>
<td>Bloodstream infections</td>
<td>12</td>
<td>83.3 (50.9–97.1)</td>
<td>99.9 (99.6–100)</td>
<td>83.5 (51.6–97.9)</td>
<td>99.9 (99.6–100)</td>
<td>0.83</td>
</tr>
<tr>
<td>Gastrointestinal infections</td>
<td>13</td>
<td>92.9 (64.2–99.6)</td>
<td>100 (99.7–100)</td>
<td>100 (75.3–100)</td>
<td>99.9 (99.7–100)</td>
<td>0.96</td>
</tr>
<tr>
<td>Lower respiratory tract infections</td>
<td>5</td>
<td>100 (46.3–100)</td>
<td>100 (99.7–100)</td>
<td>100 (47.8–100)</td>
<td>100 (99.8–100)</td>
<td>1.00</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>52</td>
<td>95.9 (89.3–98.7)</td>
<td>99.9 (99.7–100)</td>
<td>99.9 (89.9–98.9)</td>
<td>99.9 (99.6–100)</td>
<td>0.96</td>
</tr>
<tr>
<td>Surgical site infections</td>
<td>56</td>
<td>98.2 (89.2–99.9)</td>
<td>99.9 (99.7–100)</td>
<td>98.2 (90.4–100)</td>
<td>99.9 (99.7–100)</td>
<td>0.98</td>
</tr>
<tr>
<td>Urinary tract infections</td>
<td>27</td>
<td>92.6 (74.2–98.7)</td>
<td>99.9 (99.6–100)</td>
<td>92.6 (75.7–99.1)</td>
<td>99.9 (99.6–100)</td>
<td>0.93</td>
</tr>
</tbody>
</table>

CI: confidence interval; NPV: negative predictive value; PPV: positive predictive value.

The 10 European Union Member States that took part in the validation part of the study were Bulgaria, Finland, Germany, Hungary, Italy, Latvia, Lithuania, Poland, Spain, United Kingdom.

1 Number of patient records used for validation, which could be matched to those reported in the primary point prevalence survey data.
2 95% CIs have been adjusted to the overall prevalence among the primary cases (9%).

Table 3
Validation of the point prevalence survey for assessing antimicrobial use, 10 European Union Member States, 2011

A. Validation of the point prevalence survey (n=1,912 patient records)

<table>
<thead>
<tr>
<th>Validation data</th>
<th>Antimicrobial</th>
<th>No antimicrobial</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary data</td>
<td>Antimicrobial</td>
<td>833</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>No antimicrobial</td>
<td>37</td>
<td>997</td>
</tr>
<tr>
<td>Total</td>
<td>870</td>
<td>1,042</td>
<td>1,912</td>
</tr>
</tbody>
</table>

Positive predictive value (PPV): 833/878 = 94.9%; negative predictive value (NPV): 997/1,034 = 96.4%.

B. Results of the validation study applied to the total primary point prevalence survey (n=3,958 patient records)

<table>
<thead>
<tr>
<th>Validation data</th>
<th>Antimicrobial</th>
<th>No antimicrobial</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary data</td>
<td>Antimicrobial</td>
<td>1,427*</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>No antimicrobial</td>
<td>88</td>
<td>2,366b</td>
</tr>
<tr>
<td>Total</td>
<td>1,515</td>
<td>2,443</td>
<td>3,958</td>
</tr>
</tbody>
</table>

Sensitivity: 1,427/1,515*100 = 94.2%; specificity: 2,366/2,443*100 = 96.8%.

The 10 European Union Member States that took part in the validation part of the study were Bulgaria, Finland, Germany, Hungary, Italy, Latvia, Lithuania, Poland, Spain, United Kingdom.

* 1,427 = PPV*1,504.

b 2,366 = NPV*2,454.

Table 4
Validation of the point prevalence survey for antimicrobial use, by administration route, 10 European Union Member States, 2011 (n=1,912 patient records)

<table>
<thead>
<tr>
<th>Antimicrobials administered</th>
<th>N</th>
<th>Sensitivity % (95%CI)</th>
<th>Specificity % (95%CI)</th>
<th>PPV % (95%CI)</th>
<th>NPV % (95%CI)</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients on antimicrobials</td>
<td>878</td>
<td>94.9 (92.9–95.3)</td>
<td>96.8 (96.1–97.5)</td>
<td>94.9 (93.2–96.2)</td>
<td>96.4 (95.1–97.5)</td>
<td>0.91</td>
</tr>
<tr>
<td>Parenteral route</td>
<td>843</td>
<td>97.1 (95–98.3)</td>
<td>88.6 (84.4–91.9)</td>
<td>95.8 (94.3–97.1)</td>
<td>92.5 (88.9–95.3)</td>
<td>0.87</td>
</tr>
<tr>
<td>Oral route</td>
<td>281</td>
<td>88.2 (83.8–91.5)</td>
<td>97.6 (96.3–98.5)</td>
<td>92.9 (89.2–95.6)</td>
<td>95.9 (94.4–97.1)</td>
<td>0.87</td>
</tr>
</tbody>
</table>

CI: confidence interval; NPV: negative predictive value; PPV: positive predictive value.

The 10 European Union Member States that took part in the validation part of the study were Bulgaria, Finland, Germany, Hungary, Italy, Latvia, Lithuania, Poland, Spain, United Kingdom.

* Number of patient records used for validation, which could be matched to those reported in the primary point prevalence survey data.

b 95% CIs have been adjusted to the overall prevalence among the primary cases (38%).
The validation component identified an overall sensitivity of 83% (95% CI: 79–87%) and specificity of 98% (95% CI: 98–99%) for the presence of HAI. The level of agreement between the primary analysis and the validation assessment was very good across all the types of HAI. Previous studies indicated some variation at the level of individual types of HAI. In these studies respiratory tract infections had lower sensitivity, specificity, and inter-rater reliability than other types of HAI [11–13]. However, the results of this pilot study indicated a high level of specificity and a high level of agreement for these types of HAI. It is likely that the training given to support ECDC PPS has had an impact on the good validity results in our pilot study, however, this is difficult to assess and, to our knowledge, no study has been published to date assessing the effect of training on data validity. Moreover, the relatively good sensitivity and specificity results found in our pilot study may have been influenced by the ‘experimental’ conditions (e.g. selection of two hospitals per country willing to participate), which may have resulted in higher sensitivity and specificity than would have been found in validation across a non-selected group of hospitals. Indeed, in four national validation surveys carried out in 2012 during the second phase of the ECDC PPS, the average sensitivity of reporting HAIs was 71.9%, considerably lower in our pilot study [1]. The sensitivity (83%) in our study indicates potential underreporting of HAIs in the ECDC PPS. This underreporting of HAIs may have resulted from difficulties with application of definitions or availability of patient record information at the time of data collection. To the authors’ knowledge, this was the first study which formally validated the reporting of antimicrobial use within a PPS study. Antimicrobial use had a high sensitivity of 94% (95% CI: 93–95%) and specificity of 97% (95% CI: 96–98%) with a very good level of agreement. Validation of the route of antimicrobial showed that oral antimicrobials were frequently reported as parental antimicrobials.

Other variables within the validation dataset were well recorded. A complex patient records review study by Yawn and Wollan (2005) [14] found that demographic data that required copying explicit information (e.g. sex, birth date), ‘free-text’ data that required identifying and copying (e.g. chief complaints and diagnoses), and data that required abstractor judgment in determining what to record (e.g. whether heart disease was considered) differed in terms of rates of agreement. In our study, agreement between the validation and the primary data collectors on more basic demographic variables ranged from moderate for the McCabe score to very good for sex and age. This finding was in line with the scarce literature published to date [11–12,15–19], wherein basic demographic variables such as age and sex tend to have very good levels of agreement compared to those variables where interpretation is required, such as the McCabe score or other markers of co-morbidity. The variables requiring abstractor judgment in this pilot validation study usually involved verification with a clinician present on the ward, which may account for the higher than expected validity.

The IRR component showed that the in-hospital IRR reliability was very good, for HAIs and antimicrobial use. This level of agreement was also found for other variables with the notable exception of HAI origin, which had a fair/marginal kappa. No studies that have looked formally at IRR in more than one country were identified in the literature. One study [17] did examine the difference between teams of data collectors within an Indonesian PPS and indicated that inter-observer variation differed significantly between the teams of data collectors in terms of completeness of data, and most importantly in the number of detected HAIs. Differences of note in this previous study were with respect to surgical site infection, urinary tract infection and septicaemia (p = 0.01) and the reported agreement (k) did not exceed 0.59 for any type of HAI [17]. Their evaluation indicated that ascertainment was affected by underreporting in patient records, and the retrospective nature of data collection for validation purposes.

Table 5
Inter-rater reliability results by selected variables, pilot validation study of a point prevalence survey on healthcare-associated infections and antimicrobial use, eight European Union Member States, 2011 (n=44 raters and 195 patient records)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Number</th>
<th>Agreement rate</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>HAI present</td>
<td>202</td>
<td>96%</td>
<td>0.92</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>133</td>
<td>100%</td>
<td>1.00</td>
</tr>
<tr>
<td>Other lower respiratory infection</td>
<td>133</td>
<td>100%</td>
<td>1.00</td>
</tr>
<tr>
<td>Antimicrobial use</td>
<td>217</td>
<td>97%</td>
<td>0.92</td>
</tr>
<tr>
<td>Fluoroquinolone use</td>
<td>254</td>
<td>97%</td>
<td>1.00</td>
</tr>
<tr>
<td>Oral route</td>
<td>253</td>
<td>99%</td>
<td>0.95</td>
</tr>
<tr>
<td>Parental route</td>
<td>253</td>
<td>99%</td>
<td>0.94</td>
</tr>
<tr>
<td>Surgical prophylaxis</td>
<td>253</td>
<td>99%</td>
<td>0.93</td>
</tr>
<tr>
<td>Device present</td>
<td>93</td>
<td>96%</td>
<td>0.81</td>
</tr>
<tr>
<td>HAI origin</td>
<td>91</td>
<td>96%</td>
<td>0.31</td>
</tr>
</tbody>
</table>

HAI: healthcare-associated infection.

The eight European Union Member States that took part in inter-rater reliability part of the study were Bulgaria, Finland, Germany, Italy, Lithuania, Poland, Spain, United Kingdom.

* Number of variables recorded for all patients; one patient can have more than one HAI.
While the issue of record keeping with respect to device use, infection criteria and antimicrobial use, had been identified as a potential limitation at the outset of this study, the outcomes in the validation aspects of this pilot study were better than expected in this regard. The IRR results overall in this pilot study were better than those published previously in the literature with respect to precision of HAIs and types of HAI.

As with all observation studies of this nature there are a number of potential biases which are acknowledged herein. The first of these is the potential for selection bias as participating countries chose the hospitals and patient records; these were not randomly allocated at a country level. Observer bias potential is also acknowledged as not all the validators were blinded to the primary results although the high levels of IRR indicate minimal risk of this.

In summary, this pilot study suggested that the ECDC PPS dataset of HAIs and antimicrobial use was valid and reliable. Basic demographic data and antimicrobial use data had very good levels of validity and reliability and may not need to be routinely collected in future validation studies. The high specificity and IRR are an indication that the training on the case definitions organised during preparation of the ECDC PPS was effective. The lower sensitivity findings show the potential for underreporting of HAIs in the ECDC PPS and highlight the importance of validation studies for future surveillance activities in order for the burden of HAI to be better estimated.

National Participants in the ECDC pilot validation study
Bulgaria: Rossitsa Vatcheva-Dobrevska, Ivan Ivanov; Finland: Tommi Kärki; Germany: Petra Gastmeier; Hungary: Karolina Böröcz, Ágnes Hajdu; Italy: Silvius Brusafiero, Maria Luisa Moro, Luca Arnoldo; Latvia: Elina Dimina, Uga Dumpis; Lithuania: Jolanta Ašembergienė, Rolanda Valintėlienė; Poland: Aleksander Deptula, Waleria Hryniewicz; Spain: Jose Angel Rodrigo Pendas, Josep Rafart Vaqué.

Conflict of interest
None declared.

Authors’ contribution
JR, LP, JC, SC, SHo, BCoo, WM, GH, OL, BCog, SHa, and CS all contributed to the design of the study and reviewed and commented on the manuscript. In addition JR led the study, LP managed and coordinated the study, JC analysed the data and JR and LP interpreted the results and wrote the manuscript. The following people conducted the study in their respective countries RVD & II (Bulgaria); TK & OL (Finland); PG & SHa (Germany); KB & ÁH (Hungary); SB, MLM & LA (Italy); ED & UD (Latvia); JA & RV (Lithuania); AD & WH (Poland); JARP & JRV (Spain); SHo (United Kingdom).

References