Completeness of Notifiable Infectious Disease Reporting in the United States: An Analytical Literature Review

Timothy J. Doyle, M. Kathleen Glynn, and Samuel L. Groseclose

Despite state and local laws requiring medical providers to report notifiable infectious diseases to public health authorities, reporting is believed to be incomplete. Through means of an analytical literature review, the authors synthesize current knowledge on the completeness of disease reporting and identify factors associated with reporting completeness. The review was limited to published studies, conducted in the United States between 1970 and 1999, that quantitatively assessed infectious disease reporting completeness. Thirty-three studies met the inclusion criteria. Reporting completeness, expressed between 0% and 100%, was treated as the dependent outcome variable in statistical analysis; disease, study location, time period, study design, and study size were treated as independent variables. Fifty-six distinct measures of reporting completeness were identified for 21 diseases. Reporting completeness varied from 9% to 99% and was most strongly associated with the disease being reported. The mean reporting completeness for acquired immunodeficiency syndrome, sexually transmitted diseases, and tuberculosis as a group was significantly higher (79%) than for all other diseases combined (49%) (p < 0.01). Am J Epidemiol 2002;155:866–74.

Communicable disease control; disease notification; population surveillance; review literature

Surveillance for infectious diseases is a critical element in providing effective public health disease control and prevention services. In the United States, disease reporting is mandated by state and local laws. These laws require medical providers and laboratories to notify state or local public health authorities of patients diagnosed with reportable conditions (1). A reportable condition is one for which “…regular, frequent, and timely information regarding individual cases is considered necessary for the prevention and control of the disease” (2, p. v). On the basis of the unique disease control priorities within each state, the state determines which conditions are reportable within its jurisdiction. States then voluntarily report diseases notifiable at the national level to the federal Centers for Disease Control and Prevention, and a summary of notifiable disease activity in the United States is published annually (2). Although this national system of disease surveillance probably results in incomplete case enumeration, it is less clear what proportion of notifiable disease events are captured by this confluence of local, state, and national public health surveillance activities.

Periodic evaluation of disease surveillance activities may include quantitative measurements of the sensitivity of systems to detect conditions under surveillance (3). The sensitivity of surveillance systems includes both a case detection and diagnosis component and a disease-reporting component. Disease-reporting completeness refers to the proportion of those diagnosed with a notifiable condition that are reported to the appropriate public health authorities. Disease reporting, if it is representative and consistent over time, need not be complete to successfully monitor demographic, spatial, and temporal trends (4). Completeness becomes more important for infrequently occurring diseases, however, and some measure of reporting completeness is necessary to accurately interpret disease incidence or to make national and international comparisons among public health jurisdictions (5).

Several studies have evaluated the completeness of reporting for a particular disease, in a particular jurisdiction, and over a particular period. We review here the published literature on the completeness of notifiable infectious disease reporting in the United States. Through means of an analytical literature review, we attempt to synthesize current knowledge on the completeness of disease reporting and to identify factors associated with reporting completeness.

MATERIALS AND METHODS

We reviewed all studies of notifiable infectious disease-reporting completeness that were published between 1970 and 1999 in the peer-reviewed biomedical and public health
methods for measuring reporting completeness (10, 11). Study size was expressed as a continuous variable by the denominator of the fraction used to calculate reporting completeness. Only results reported for discrete disease entities (rather than disease categories such as viral hepatitis), with a minimum of 10 cases reported in the denominator, were included in the statistical analysis. Statistical analyses were performed using SAS JMP version 3.1 software for Windows (SAS Institute, Inc., Cary, North Carolina).

Finally, we divided the measures of reporting completeness observed in this review into two broad disease categories for comparison. The first category included the observations of reporting completeness for tuberculosis, acquired immunodeficiency syndrome (AIDS), and sexually transmitted diseases (i.e., chlamydia, gonorrhea, and syphilis). We grouped these conditions because surveillance and prevention activities for tuberculosis, AIDS, and sexually transmitted diseases are often grouped programmatically, since the populations affected by these conditions and the surveillance methods and data sources are similar. The second category included observations of reporting completeness for all other disease entities for which data were available.

RESULTS

A total of 33 studies were identified in the published literature that met the stated inclusion criteria. Among the published studies, we observed two distinct methods for calculating reporting completeness. In 22 studies (67 percent), researchers measured reporting completeness by dividing the number of cases reported to public health authorities by the total number of cases identified through active case detection and the use of supplemental data sources (table 1) (8, 12–32). This method does not account for the number of cases left undetected by all data sources and is referred to as the uncorrected method. In the remaining 11 studies, researchers measured reporting completeness by dividing the number of cases reported to public health authorities by the total number of cases estimated through the use of capture-recapture methods for comparing two or more data sources (table 2) (33–43). This method attempts to account for the number of cases undetected by all available data sources and is referred to as the underascertainment corrected method (44). When determining the accuracy of this method, one should consider the validity of the data from each source, the dependency relation between data sources, and the criteria used to match persons between data sources. Specific observations related to these factors are noted for each study in the limitations column of table 2.

From the 33 published studies, 56 measures of reporting completeness were recorded, involving 21 different infectious diseases reportable in the jurisdictions where the studies were conducted. Reporting completeness differed significantly among the 21 distinct disease categories (Kruskal-Wallis test, \( p < 0.05 \)). No temporal trends in disease-reporting completeness were noted. Reporting completeness for studies conducted during the 1990s was similar to reporting completeness during the 1980s. Reporting completeness was generally
<table>
<thead>
<tr>
<th>Author(s) (reference no.)</th>
<th>Location</th>
<th>Time period</th>
<th>Disease</th>
<th>Supplemental data sources</th>
<th>Observed reporting completeness*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marier (12)</td>
<td>Washington, DC</td>
<td>1971</td>
<td>Various</td>
<td>Hospital discharge records</td>
<td>Total†: 35% (200 of 570)</td>
</tr>
<tr>
<td>Murray et al. (13)</td>
<td>Nassau County, NY</td>
<td>1972</td>
<td>Tuberculosis</td>
<td>Laboratory data</td>
<td>Haemophilus influenzae: 32% (7 of 22)</td>
</tr>
<tr>
<td>Eisenberg and Wiesner (8)</td>
<td>Alaska</td>
<td>1973–1974</td>
<td>Gonorrhea</td>
<td>Medical records</td>
<td>Salmonellosis: 42% (11 of 26)</td>
</tr>
<tr>
<td>Kimball et al. (14)</td>
<td>Washington, DC</td>
<td>1978</td>
<td>Shigellosis</td>
<td>Medical records</td>
<td>Shigellosis: 62% (21 of 34)</td>
</tr>
<tr>
<td>Davis and Bohn (15)</td>
<td>Wisconsin</td>
<td>1980–1982</td>
<td>Meningococcal disease</td>
<td>Laboratory data</td>
<td>Tuberculosis: 63% (127 of 200)</td>
</tr>
<tr>
<td>Vogt et al. (16)</td>
<td>Vermont</td>
<td>1982–1983</td>
<td>Various</td>
<td>Hospital discharge records</td>
<td></td>
</tr>
<tr>
<td>Chamberland et al. (17)</td>
<td>New York, NY</td>
<td>1982–1983</td>
<td>AIDS‡</td>
<td>Medical records and laboratory data combined</td>
<td>65% (32 of 49)</td>
</tr>
<tr>
<td>Harkess et al. (18)</td>
<td>Oklahoma</td>
<td>1985</td>
<td>Shigellosis</td>
<td>Laboratory data</td>
<td>42% (76 of 183)</td>
</tr>
<tr>
<td>Lindan et al. (19)</td>
<td>San Francisco, CA</td>
<td>1985–1986</td>
<td>AIDS</td>
<td>Medical records</td>
<td>30% (43 of 136)</td>
</tr>
<tr>
<td>Conway et al. (20)</td>
<td>South Carolina</td>
<td>1986–1987</td>
<td>AIDS</td>
<td>Medical records</td>
<td>23% (28 of 120)</td>
</tr>
<tr>
<td>Modesitt et al. (21)</td>
<td>Oregon</td>
<td>1986–1987</td>
<td>AIDS</td>
<td>Medical records</td>
<td>23% (28 of 120)</td>
</tr>
<tr>
<td>Campos-Outcalt et al. (22)</td>
<td>Pima County, AZ</td>
<td>1986–1988</td>
<td>Various</td>
<td>Hospital discharge records and death certificates combined</td>
<td>96% (409 of 425)</td>
</tr>
<tr>
<td>Fife et al. (23)</td>
<td>Philadelphia, PA</td>
<td>1986–1991</td>
<td>AIDS</td>
<td>Provider-based case registry</td>
<td>91% (267 of 295)</td>
</tr>
<tr>
<td>Kirsch et al. (24)</td>
<td>Washington, DC</td>
<td>1989</td>
<td>Gonorrhea</td>
<td>Emergency department medical records and laboratory data combined</td>
<td>91% (204 of 223)</td>
</tr>
<tr>
<td>Study</td>
<td>Location</td>
<td>Period</td>
<td>Disease/Condition</td>
<td>Data Sources</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>-------------------</td>
<td>---------------</td>
<td>--------------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Elcock et al. (25)</td>
<td>San Mateo County, CA</td>
<td>1989–1991</td>
<td>AIDS</td>
<td>Medical records</td>
<td></td>
</tr>
<tr>
<td>Standaert et al. (26)</td>
<td>Tennessee</td>
<td>1989–1992</td>
<td>Invasive bacterial infections</td>
<td>Laboratory data</td>
<td></td>
</tr>
<tr>
<td>Ewert et al. (27)</td>
<td>Los Angeles, CA</td>
<td>1990–1991</td>
<td>Measles</td>
<td>Community survey</td>
<td></td>
</tr>
<tr>
<td>Driver et al. (28)</td>
<td>Puerto Rico</td>
<td>1992</td>
<td>Tuberculosis</td>
<td>Hospital discharge records, pharmacy records, laboratory data, and case registries combined</td>
<td></td>
</tr>
<tr>
<td>Yokoe et al. (29)</td>
<td>Massachusetts</td>
<td>1992–1996</td>
<td>Tuberculosis</td>
<td>Automated clinical, laboratory, and pharmacy data from health maintenance organization</td>
<td></td>
</tr>
<tr>
<td>Smucker and Thomas (30)</td>
<td>Rural North Carolina</td>
<td>1993</td>
<td>Sexually transmitted diseases</td>
<td>Laboratory data compared with case reports during both active and passive surveillance periods</td>
<td></td>
</tr>
<tr>
<td>Dembek et al. (31)</td>
<td>Connecticut</td>
<td>1994–1996</td>
<td>Vancomycin-resistant enterococci</td>
<td>Laboratory data</td>
<td></td>
</tr>
<tr>
<td>Trepka et al. (32)</td>
<td>Wisconsin</td>
<td>1995</td>
<td>Tuberculosis</td>
<td>Hospital discharge records and laboratory data</td>
<td></td>
</tr>
</tbody>
</table>

*Observed reporting completeness is the number of routine surveillance case reports divided by the observed number of cases in routine and supplemental data sources combined.
† Includes other notifiable conditions.
‡ AIDS, acquired immunodeficiency syndrome.
<table>
<thead>
<tr>
<th>Author(s) (reference no.)</th>
<th>Location</th>
<th>Time period</th>
<th>Disease</th>
<th>Supplemental data sources</th>
<th>Limitations</th>
<th>Estimated reporting completeness*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cochi et al. (33)</td>
<td>United States</td>
<td>1970–1985</td>
<td>Congenital rubella syndrome</td>
<td>Birth defects registry</td>
<td>Personal identifiers not used for matching</td>
<td>22% (263 of 1,186)</td>
</tr>
<tr>
<td>Sutter et al. (34)</td>
<td>United States</td>
<td>1979–1984</td>
<td>Tetanus mortality</td>
<td>Death certificates</td>
<td>Outcome (alive/dead) variable in case reports not validated; personal identifiers not used for matching</td>
<td>40% (129 of 326)</td>
</tr>
<tr>
<td>Prevots et al. (35)</td>
<td>United States</td>
<td>1980–1991</td>
<td>Vaccine-associated paralytic polio</td>
<td>Claims filed with national vaccine injury compensation program</td>
<td>Matching criteria unspecified</td>
<td>81% (92 of 114)</td>
</tr>
<tr>
<td>Hardy et al. (36)</td>
<td>Boston, MA; Chicago, IL; Washington, DC; New York, NY</td>
<td>1985</td>
<td>AIDS†</td>
<td>Death certificates</td>
<td>Inclusion and search criteria varied across study sites</td>
<td>89% (487 of 548)</td>
</tr>
<tr>
<td>Sutter and Cochi (37)</td>
<td>United States</td>
<td>1985–1988</td>
<td>Pertussis</td>
<td>Hospital discharge records and death certificates</td>
<td>Hospitalization and outcome data from case reports not validated; discharge diagnoses and death certificates not validated; personal identifiers not used for matching</td>
<td>32% (4,404 of 13,557 hospitalized)</td>
</tr>
<tr>
<td>Rosenblum et al. (38)</td>
<td>Various sites, United States</td>
<td>1988</td>
<td>AIDS</td>
<td>Hospital discharge and Medicaid data combined</td>
<td>Matching criteria unspecified</td>
<td>92% (4,157 of 4,500)</td>
</tr>
<tr>
<td>Davis et al. (39)</td>
<td>New York, NY</td>
<td>1991</td>
<td>Measles</td>
<td>Medical records</td>
<td>No validation of physician diagnosis</td>
<td>45% (664 of 1,487)</td>
</tr>
<tr>
<td>Ikeda et al. (40)</td>
<td>New York State, excluding New York City</td>
<td>1991</td>
<td>Perinatal hepatitis B exposure</td>
<td>Mother/infant pairs matched from maternal prenatal, maternal perinatal, and newborn infant screening program data</td>
<td>Matching criteria unspecified</td>
<td>96% (363 of 378)</td>
</tr>
<tr>
<td>Ackman et al. (41)</td>
<td>New York State, excluding New York City</td>
<td>1991</td>
<td>Meningococcal disease</td>
<td>Hospital discharge records</td>
<td></td>
<td>93% (100 of 107)</td>
</tr>
<tr>
<td>Barat et al. (42)</td>
<td>Southwestern United States</td>
<td>1995</td>
<td>Malaria</td>
<td>Laboratory data</td>
<td>Matching criteria unspecified</td>
<td>69% (43 of 62)</td>
</tr>
<tr>
<td>Effer et al. (43)</td>
<td>Hawaii</td>
<td>1998</td>
<td>Various</td>
<td>Automated laboratory data</td>
<td>Matching criteria unspecified</td>
<td>Total‡: 44% (156 of 357) Salmonellosis: 50% (102 of 205) Shigellosis: 54% (16 of 30) Giardiasis: 33% (26 of 79) VRE†: 22% (7 of 32) Invasive Streptococcus pneumoniae: 9% (5 of 55)</td>
</tr>
</tbody>
</table>

* Estimated reporting completeness is the number of routine surveillance case reports divided by the estimated number of cases, after correcting for underascertainment using capture-recapture methods.
† AIDS, acquired immunodeficiency syndrome; VRE, vancomycin-resistant enterococci.
‡ Includes other notifiable conditions.
lower for studies conducted during the 1970s, but these earlier studies preceded the identification of human immunodeficiency virus and mandatory reporting of AIDS. After stratification by disease, reporting completeness during the 1970s was similar to completeness during subsequent decades. No other factor (i.e., geographic location, study design, or study size), either individually or after stratification by disease, was significantly associated with reporting completeness.

When we compared the individual measures of reporting completeness for tuberculosis, AIDS, and sexually transmitted diseases ($n = 25$) with those for all other diseases ($n = 31$), the mean of the individual measures of reporting completeness was significantly higher for tuberculosis, AIDS, and sexually transmitted diseases (79 percent) than for all other diseases (49 percent; $t$ test, $p < 0.01$).

**DISCUSSION**

The published literature reveals variation in infectious disease-reporting completeness in the United States, ranging from 9 to 99 percent. Though highly variable, reporting completeness appears to be most strongly related to the disease or condition being reported. Other factors, such as geographic location, appear to be less important in determining disease-reporting completeness. No temporal trends were noted and, based on the studies we reviewed, infectious disease-reporting completeness seems neither to have improved nor to have deteriorated appreciably in the United States during the last three decades. Reporting completeness was significantly greater for tuberculosis, AIDS, and sexually transmitted diseases as a group than for all other notifiable infectious diseases combined. The reasons for this are not clear but may be related to the perceived seriousness of these diseases or to the greater financial and human resources devoted to treating and preventing them, often involving contact tracing by case workers in the community. Historically, the strong reliance of tuberculosis and sexually transmitted disease patients on the public health system for case management and treatment has also augmented the surveillance for these conditions.

Studies conducted in other industrialized countries have found patterns of reporting completeness similar to those found in the United States (45–52). Both in the United States and abroad, AIDS surveillance has generally been more frequently evaluated and AIDS cases more completely reported than for other diseases. The only other published reviews on the subject of infectious disease-reporting completeness have focused exclusively on AIDS reporting (53, 54). Buehler et al. (53) reviewed several published and unpublished AIDS surveillance evaluation activities in the United States, whereas the review by Gertig et al. (54) provides an international comparison of AIDS-reporting completeness. Ours is the first known review of reporting completeness for infectious diseases in general, and it suggests important differences depending on the condition under surveillance.

Studies in this review provide general support for the commonly held belief that active disease surveillance results in more complete case enumeration than passive surveillance methods. This observation has been frequently noted and further suggested by studies that were unable to control for true temporal or spatial differences in disease incidence (55–58). Evaluation activities that supplement routine data sources to derive a measure of reporting completeness are, albeit temporary, a form of active surveillance. Such evaluation activities, whether published or not, undoubtedly contribute to more complete national disease reporting (53). However, because national infectious disease incidence data are derived from a mixture of active and passive surveillance methods, the observation that active surveillance results in increased reporting completeness does not readily provide a clearer interpretation of national data. Thus, as a means to both better understand surveillance data and contribute to increased reporting completeness, ongoing surveillance evaluation of the kind represented by studies in this review should be encouraged.

Previous reports have delineated the specific stages in the disease identification and reporting process for conditions of public health importance. Using shigellosis as an example, Rosenberg et al. (59) estimated that, for every 100 persons infected with *Shigella*, 76 became symptomatic, 28 consulted a physician, nine submitted stool cultures, seven had positive culture results, six were reported to the local health department, and five were reported nationally to the Centers for Disease Control and Prevention. Thus, they proposed a multiplication factor of 20 to estimate the overall number of *Shigella* infections based on national shigellosis case reports. Similar strategies have been used to estimate the number of *Salmonella* (60) and *Chlamydia* (61) infections in the United States. Our review of the literature specifically addressed the later stage in this process between clinical or laboratory diagnosis and reporting to local, state, or national health authorities. Early components of surveillance sensitivity related to underdiagnosis, such as the proportion of those infected who become symptomatic, seek treatment, submit diagnostic specimens, and obtain laboratory confirmation of infection, were not addressed by this review but are also important to consider when interpreting surveillance data.

Several authors have described reasons for failure of health-care providers and laboratories to report notifiable diseases (9, 62–66). The reasons cited include a lack of awareness of the legal requirement to report, a lack of knowledge of which diseases are reportable, a lack of understanding of how or to whom to report, an assumption that someone else will report the case, intentional failure to report to protect patient privacy, and insufficient reward for reporting or penalty for not reporting. Interventions aimed at reducing these barriers have had limited success at improving provider and laboratory reporting behavior (9, 64–66).

Advances in provider- and laboratory-based information management hold promise for automated disease reporting and surveillance. The use of electronic health information systems should help to reduce dependence on individual provider behavior for routine disease surveillance, yet still allow patient privacy and confidentiality to be maintained. One study in our review demonstrated the added completeness of disease reporting achievable through the use of automated systems for capturing electronic laboratory data (43).
When compared with traditional paper-based morbidity reporting, automated reporting of electronic laboratory data was found by Effler et al. to have resulted in a 2.3-fold increase in case reports. Furthermore, electronic reports arrived an average of 3.8 days earlier than conventional reports and were more likely to provide complete patient-, specimen-, and provider-related information. Another study demonstrated the usefulness of automated health maintenance organization data for detecting unreported cases of tuberculosis (29). Yokoe et al. found that health maintenance organization pharmacy-dispensing data, suggesting the use of two or more antituberculosis drugs, were a useful marker for identifying tuberculosis patients, particularly those without positive microbiologic results. As further advances are made in information technology and clinical and laboratory data standards, coupled with further consolidation of healthcare delivery in the marketplace, automated electronic provider- and laboratory-based reporting will likely become a more prominent component of routine disease surveillance in the United States (67).

One third of the studies included in our review used capture-recapture methods to estimate the number of cases undetected by all available data sources. Such methods are based on the key assumption that the data sources are independent; that is, the probability of being present or “captured” in one data source is unrelated to the probability of being present in the other data sources (68). These methods can result in inaccurate estimates when patients are misdiagnosed or when patient diagnoses are inaccurately recorded in health information systems (69, 70). Therefore, data validation is an important aspect of accurately applying capture-recapture methods. Estimates using capture-recapture methods are also influenced by record linkage errors (71). False negative matches will underestimate reporting completeness, whereas false positive matches will inflate estimates of reporting completeness. Therefore, for these methods to be successfully applied, appropriate criteria must be defined for accurately matching persons between data sources.

However, for any given study, applying underascertainment corrected methods will always result in an equal or lower estimation of reporting completeness than would result from using uncorrected methods. This is evidenced by Prevots et al. (35) when evaluating completeness of reporting for vaccine-associated paralytic polio. Their conclusion of 81 percent (92 of 114) reporting completeness using capture-recapture methods is lower than the 94 percent (92 of 98) completeness that could be confirmed using uncotted methods. Whether underascertainment corrected methods result in a more accurate estimate of reporting completeness depends on how these methods are applied and the individual characteristics of the data sources being used (72, 73). Despite the effect of different methods on any given study in our review, study design was not significantly associated with reporting completeness. Therefore, studies using the two different methods were combined for our analyses.

In conclusion, surveillance systems require ongoing maintenance and evaluation if the data that result from them are to be accurately interpreted. We have presented here a review of quantitative evaluations of infectious disease surveillance and case-reporting completeness conducted in the United States during the last three decades. As health-care services and information technology continue to evolve, the possibility exists for numerous changes in conducting routine public health surveillance. Ongoing evaluation of disease-reporting completeness will continue to be a necessary part of public health surveillance, enabling more accurate interpretation of surveillance information for disease control and prevention.

ACKNOWLEDGMENTS

The authors thank Lori Hutwagner for statistical advice and Dr. G. David Williamson for critical review of the manuscript.

REFERENCES


