Under-reporting and case fatality estimates for emerging epidemics

Katherine E Atkins,^{1,2} Natasha S Wenzel,² Martial Ndeffo-Mbah,² Frederick L Altice,³ Jeffrey P Townsend,⁴ Alison P Galvani²

¹Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, UK

²Center for Infectious Disease Modeling and Analysis, Yale School of Public Health, New Haven, CT, USA

³Department of Medicine, Section of Infectious Diseases, New Haven, CT, USA

⁴Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA

Correspondence to: K E Atkins
katherine.atkins@lshtm.ac.uk
Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/BMJ.h1115>)

Cite this as: *BMJ* 2015;350:h1115
doi: 10.1136/bmj.h1115

Accepted: 29 December 2014

Using the example of Ebola, this article highlights the problems of using population level data to estimate case fatality rates in epidemics

The case fatality risk (CFR) is the probability that an infection results in death. For an emerging infectious disease, CFR is a vital metric to assess clinical severity. In this paper, we highlight methodological considerations fundamental to calculations of CFR during an ongoing epidemic. Ignoring information biases inherent to population level data may lead to inaccurate estimates of CFR, especially during the nascent phase of an epidemic. The epidemic of Ebola virus disease in west Africa provides an ideal case study both because the high death rate is a pressing international concern and because many CFR estimates have neglected to account for these biases, leading to inaccurate quantification.

Population level and individual outcome data

During the early phase of an emerging epidemic, population level cumulative case and death counts may be the only data available from which to estimate epidemiological statistics. Population level data provide information on the total number of confirmed infections and total number of confirmed deaths, but not on individual progression of disease—that is, the data do not link individuals reported as infected with their death or recovery. By contrast, individual level data on disease progression follow new cases through time and ascertain each person's clinical outcome. Throughout this article, we refer to data as population level when individual outcomes are unknown, and as individual level when individuals have been followed through to the definitive clinical outcome of their infections.

Correct calculation of case fatality risk

CFRs calculated from individual outcome data are likely to be more reliable than estimates calculated from population level data. Although over-representation of

more severe or symptomatic cases in the ascertainment of individual outcome data may mean the result is an upper bound of the true CFR, use of individual outcome data to calculate CFR nonetheless quantifies the likelihood that an ascertained case leads to death. To accurately estimate CFR from population level data, the same fraction of confirmed cases and deaths must be reported, or unrealistically, the under-reporting of both cases and deaths must be known. Moreover, CFR estimates from population level data must include the lag time between reporting cases and reporting deaths in order to account for reported cases for whom the disease outcome is yet unknown. Below, we detail the potential biases that can arise in estimating CFR.

Biases

Calculating the CFR from population level data can introduce several biases (see supplementary information):

Underestimation of deaths—Calculating the CFR as the ratio of deaths to cases at a single point in time inaccurately assumes that all people infected at the time of ascertainment will survive, thereby underestimating the CFR.¹ This underestimation is particularly pronounced during the exponentially increasing phase of an epidemic, when the number of new, unresolved infections is high relative to the number of cumulative cases.

Separate reporting systems may be used to record confirmed cases and deaths, leading to differential reporting of either outcome and thereby inaccurately estimates of CFR in either direction.

Country reporting methods—Countries have different systems for reporting confirmed cases and deaths.² Dividing the total reported deaths by the total reported cases over multiple countries neglects such variability across countries, generating confidence intervals of the CFR estimate that are inaccurately narrow. It is unlikely that the under-reporting of cases and the under-reporting of deaths in one country will be exactly balanced by those of another country because the reporting of cases and deaths is unlikely to be symmetrically distributed across countries. A pooled CFR calculated across multinational data is an average based on data biased by inconsistent reporting, and likewise produces a biased estimate.

Ascertainment—The use of either population level data or individual patient outcome data can also introduce a fourth bias. Preferential reporting of severe cases during either disease surveillance or cohort studies neglects mild or asymptomatic infections less likely to be fatal. This bias leads to an overestimation of CFR.

Below we show how these biases have led to discrepancies in CFR estimates in the Ebola epidemic.

SUMMARY POINTS

Accurate calculations of case fatality risk (CFR) from population level data must consider both the time lag between reporting of cases and deaths and country specific differences in under-reporting of cases and deaths

Best estimates are achieved by using outcome data from individual patients

Individual estimates can also be inaccurate if mild or asymptomatic cases are not reported

For the current Ebola epidemic, our assessment of case fatality estimates indicates significant differences in reporting of cases and deaths among countries

Inconsistency of case fatality risk estimates for Ebola

The WHO response team estimated the average CFR for the current Ebola epidemic across west Africa as 71% (95% confidence interval 69% to 73%).³ This was calculated from individual outcome data as the percentage of fatal cases among reported cases with a definitive clinical outcome. We term this metric the individual outcome CFR measure and regard it as the best available metric for CFR, although still prone to bias if reporting is not independent of case severity. The individual outcome CFR measure is consistent across countries at 71% (67% to 74%), 72% (69% to 75%) and 69 (65% to 73% for Guinea, Liberia, and Sierra Leone, respectively.³ The country specific CFRs were calculated from patient data collected between 30 December 2013 and 14 September 2014 and therefore represent averages across this period, with a higher weighting for patient data toward the end of the period when most cases occurred.

In general, CFR may be contingent on the degree of treatment or care that patients receive, such that CFR could vary both between countries and over time. For instance, CFR can decrease over the course of an epidemic because new treatment is introduced or it can increase if the health system becomes overwhelmed at the height of an epidemic. The consistency of the country CFRs despite Ebola arriving at different times suggests that no temporal variability has manifested during the current Ebola epidemic, probably because there is no effective drug treatment.

Bias in CFR measures from individual outcome data emerges if the infected cohort is not representative of the severity profile in the general population. In reality, CFR is likely to be calculated from individual outcome data that oversample severe cases and do not follow asymptomatic cases, thereby leading to overestimation. Nevertheless, the individual outcome CFR measure is likely to be best available estimate. A CFR estimate from population level data that has appropriately accounted for under-reporting and reporting delays should align with the individual outcome CFR measure. The figure compares individual CFR measures with those calculated from population level data as it was reported (naïve), accounting for the lag in reporting of cases and deaths (delayed), and including an estimate of asymptomatic cases.

Naïve estimates

We calculated naïve Ebola CFR estimates for each country separately by dividing the cumulative number of WHO reported confirmed deaths by the cumulative number of WHO reported confirmed cases on specific dates when the data were publicly released in situation reports.^{1 4 5} These estimates vary both temporally and between countries and are highly inconsistent with the individual outcome estimates (figure).

Adjusting for reporting delays

In contrast to naïve estimates, delayed estimates account for the lag between country specific reporting of cases and deaths. We used WHO data on the average time from onset of symptoms to WHO notification and

Figure | Effect of using different methods to estimate case fatality risk of Ebola in Guinea, Liberia, and Sierra Leone: naïve—number of deaths by the number of cases at time t ; delayed—adjusted for delay between WHO reporting of the case and the death; individual outcome—known outcome (recovery or death) of a group of patients over entire sampling period 30 December 2013 to 14 September 2014 for the three countries (95% CI); and asymptomatic—estimated as the country specific individual outcome measure multiplied by 50% (assuming that half of cases are asymptomatic)

the average time from WHO notification to death³ to calculate the average time lag for each country between WHO notification of an Ebola case and WHO notification of the death.³ We found that the lags were 3.3 days, 2.3 days, and 5.7 days for Guinea, Liberia, and Sierra Leone, respectively. These lag times are shorter than the time from symptom onset to death, which is typically over a week,^{3 6} suggesting that deaths are confirmed and reported more swiftly than other cases. Consequently, the delay from symptom onset to case report is longer than the delay from death to fatality report. To calculate the delayed CFR estimate on day t we divided the number of cumulative deaths that have been recorded up until day $t + d$ by the cumulative case count that have been recorded up until day t , where d is the country specific lag time between case and fatality

reporting. This method requires fatality count data for d days after the date on which the CFR is calculated. We assumed that confirmed deaths are uniformly distributed across the 2–4 day intervals between WHO reports. Calculations based on the delayed CFR estimates for Guinea, Liberia, and Sierra Leone are higher than the naive estimates. Nonetheless, the delayed estimates are highly inconsistent both between countries and with the individual outcome measures (fig 1). The between country discrepancies suggest that there are widespread differences in the relative under-reporting of cases and deaths across countries.

Adjusting for country specific under-reporting

A common assumption in CFR calculations is that cases and deaths are reported to the same degree. Even if there is substantial under-reporting, as has been suggested for the current Ebola outbreak,⁷ it is possible to estimate CFR accurately provided that the extent of under-reporting is the same for cases as it is for deaths. However, equal reporting of cases and deaths is unlikely in an emergency epidemic situation, and even less likely to be consistent across three different countries. The resulting bias can be corrected by calculating the relative under-reporting of cases versus deaths. For each country, we divided the individual outcome CFR measures by the delayed CFR estimate, providing country specific estimates of the relative under-reporting of cases versus deaths (under the assumption that reporting does not vary temporally).

Our calculation suggests that CFRs based on population level data will not be consistent with individual outcome measures unless country specific reporting of confirmed deaths versus cases is incorporated. Given that the delayed CFR estimate was significantly lower than the individual outcome CFR in Sierra Leone and Guinea, we can infer that confirmed Ebola cases are 72–95% more likely to be reported than confirmed deaths in Sierra Leone and 7–20% more likely to be reported in Guinea, assuming constant under-reporting since the beginning of the epidemic. Conversely in Liberia, the delayed CFR estimate was considerably higher than the individual outcome CFR, signifying that confirmed Ebola deaths are 55–64% more likely to be reported than confirmed cases. Consequently, the over-reporting of confirmed deaths relative to cases in Liberia led to a logically impossible estimate of a CFR that exceeds 100% (fig 1). These estimates point toward vulnerabilities in each country's disease surveillance and laboratory testing infrastructure.

Complications with pooling multicountry data

Our results highlight the sensitivity of CFR estimates to under-reporting of cases versus deaths. When under-reporting is higher for either deaths or cases, it is inaccurate to calculate the average CFR as the ratio of cumulative deaths by cumulative cases across all countries, because variable under-reporting biases the cumulative counts. As the epidemic spreads at different rates through countries, and country specific surveillance varies over time, the CFR estimates from pooled

counts fluctuate substantially. For example, the delayed CFR estimates from pooled data across three countries range from 55% on 23 September to 85% on 1 October, inconsistent with the 95% confidence interval of 69% to 73% measured by the individual outcome study.³ Strikingly, the delayed CFR estimate calculated from the pooled data falls within the confidence interval for the individual outcome CFR only on 5 July, 30 August, and 5 September.

Individual outcome data and asymptomatic infection

Cohort data collected by the WHO Response Team on individual outcomes includes only people who present for medical care or are considered suspected cases from contact tracing efforts.³ Data on previous Ebola outbreaks suggest that some infected people will not show symptoms, although how many is uncertain.^{8–10} Unrecorded mild or asymptomatic infections can have a large effect on estimates of CFR. If we assume that 50%⁸ of infections are undetected, the ratio of reported cases to reported fatalities would also be halved. Individual outcome CFRs thus establish an upper bound for the true CFR. Monitoring asymptomatic infections in an outbreak situation may not be a priority, but a retrospective study that assessed the serostatus of all close contacts of patients irrespective of symptoms would help to determine the proportion of asymptomatic and mild infections.

Conclusion

Estimating case fatality risks from population level data requires accounting for both the delay between the report of the case and death and country specific under-reporting of cases relative to deaths; this information may not be known. Furthermore, if the under-reporting of deaths or cases varies significantly between different geographical regions, it is problematic to pool data from multiple regions. We suggest that CFRs estimated from individual outcome data are the best available. Nonetheless, even these estimates can be biased if reporting varies according to the severity of a case. In such circumstances, the individual outcome CFR measure will constitute an upper bound of the true CFR.

Our analysis of data from the 2014 Ebola epidemic suggests that the differential under-reporting of cases and deaths in each country contributes more substantially to inaccurate CFR estimation than delays in reporting. Without knowledge of the relative reporting of cases to deaths, estimates of CFR calculated from population level data provide are likely to deviate substantially from the true CFR.

We thank Tolbert Nyenswah, Luke Bawo, and Mosaka Fallah from the Liberian Ministry of Health and Social Welfare, and Ryan Boyko for their helpful discussion.

Competing interests: All authors have completed the unified competing interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare the work was supported by the National Institutes of Health (NIH U01 GM087719, U01 GM105627, and K24 DA017072), and the National Science Foundation (NSF RAPID 1514673). This research was partly funded by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Immunisation at the London School of

Hygiene and Tropical Medicine in partnership with Public Health England (PHE). The views expressed are those of the authors and not necessarily those of the NIH, NSF, NHS, NIHR, the Department of Health or Public Health England.

Provenance and peer review: Not commissioned; externally peer reviewed.

- 1 Ghani AC, Donnelly CA, Cox DR, Griffin JT, Fraser C, Lam TH, et al. Methods for estimating the case fatality ratio for a novel, emerging infectious disease. *Am J Epidemiol* 2005;162:479–86.
- 2 World Health Organization. WHO global alert and response, 26 November 2014. www.who.int/csr/en/.
- 3 WHO Ebola Response Team. Ebola virus disease in west Africa—the first 9 months of the epidemic and forward projections. *N Engl J Med* 2014;371:1481–95.
- 4 Centers for Disease Control. Outbreaks chronology: Ebola hemorrhagic fever. 2014. www.cdc.gov/vhf/ebola/resources/outbreak-table.html.
- 5 Kucharski AJ, Edmunds WJ. Case fatality rate for Ebola virus disease in west Africa. *Lancet* 2014;384:1260.
- 6 Chertow DS, Kleine C, Edwards JK, Scaini R, Giuliani R, Sprecher A. Ebola virus disease in west Africa—clinical manifestations and management. *N Engl J Med* 2014;371:2054–7.
- 7 World Health Organization. Why the Ebola outbreak has been underestimated. 2014. www.who.int/mediacentre/news/ebola/22-august-2014/en/.
- 8 Bellan SE, Pulliam JRC, Dushoff J, Meyers LA. Ebola control: effect of asymptomatic infection and acquired immunity. *Lancet* 2014;384:1499–500.
- 9 Leroy E, Baize S, Volchkov V, Fisher-Hoch S, Georges-Courbot M-C, Lansoud-Soukate J, et al. Human asymptomatic Ebola infection and strong inflammatory response. *Lancet* 2000;355:2210–5.
- 10 Heffernan JM, Smith RJ, Wahl LM. Perspectives on the basic reproductive ratio. *J R Soc Interface* 2005;2:281–93.

© BMJ Publishing Group Ltd 2015