

Supporting Information

Chan et al. 10.1073/pnas.1006219107

SI Methods

Selection of WHO-Confirmed Outbreaks. *Exclusion of outbreaks of seasonal, ongoing, or endemic diseases.* For the creation of the database of distinct WHO-confirmed outbreaks for 1996–2009, outbreaks of diseases considered seasonal, ongoing, or endemic to the region were excluded unless they were described as unprecedented events or unless they were the first cases in a long period. In particular, for meningitis, we excluded reports for regions lying in the “African Meningitis Belt” (according to the 2006 map in the CDC’s Yellow Book for traveler’s health information). Additionally, if a WHO meningitis report is for a country that lies partially in the Meningitis Belt but the report does not specify the exact region, it was excluded as well. For all other diseases, if the report states that the disease is endemic to the region, the outbreak was excluded unless it was also described as a notable or unprecedented event. The rationale for this exclusion criterion was motivated by the fact that in certain regions, seasonal or endemic diseases are not unexpected and therefore usually under continuous surveillance anyway. As for outbreaks of diseases that have an ongoing presence, it is difficult to point to an exact outbreak start date, which makes it difficult to conduct an appropriate assessment and comparison of timing from outbreak start to outbreak discovery and to public communication about the outbreak.

Exclusion of single, isolated cases. Outbreaks consisting of only a single confirmed case were generally excluded unless there were other probable cases in close contacts that could not be confirmed for reasons such as inability to obtain a sample for testing.

Distinction of separate outbreaks. The difficulty of teasing apart cases that occur close together on either a temporal and/or spatial scale into distinct outbreaks is a well-recognized problem. Similarly, when there were WHO reports of ongoing outbreaks with cases that occur regularly but sporadically (as is often the case for H5N1 avian influenza in certain regions), it was difficult to determine whether they constitute separate outbreaks. As much as possible, we attempted to follow the WHO’s lead based on the naming convention for the titles of their reports (i.e., whether the report was titled an update). Spatially, our general guideline was to distinguish outbreaks according to distinct country/disease units, meaning we treated cases arising in new countries as new outbreaks. An exception to this guideline was cases where it was described as a clear spread from one country to another (e.g., cases in towns close to a shared border). Temporally, we generally allowed a 1- to 3-mo gap (this was subjected to a case-by-case basis, depending on the disease and country in question) in between cases occurring in the same country before we treated them as belonging to separate outbreaks.

Database of WHO-Confirmed Outbreaks During 1996–2009. The database of WHO-confirmed outbreaks that was created included the following columns: disease; country; state; city (where it was known, the city where the outbreak started was chosen for outbreaks affecting multiple cities, or if that was unknown, the city that bore the greatest burden in terms of morbidity/mortality); latitude; longitude; GeoNames identification number (<http://www.geonames.org/>); date of the WHO Disease Outbreak News (DON) report (<http://www.who.int/csr/don/en/>); other mentioned earlier WHO date; issue date of the earliest ProMED report (<http://www.promedmail.org/>); issue date of the source cited in the earliest ProMED report; load date of the earliest HealthMap alert (<http://www.healthmap.org/>); publication date of the source cited in the earliest HealthMap alert; identification number of the

earliest HealthMap alert; source of the earliest HealthMap alert; date of receipt of the earliest GPHIN article; publication date of source of the earliest GPHIN article; identification number of the earliest GPHIN article; source of the earliest GPHIN article; date of an announcement made by a local authority figure or medical professional; date of notification to the WHO; date of onset of an associated climate/geological phenomenon; date of a local mass gathering; approximate date of onset of an associated animal or wildlife outbreak; date of outbreak start; approximate date of exposure; date of symptom onset; date of hospitalization or medical visit; date of outbreak detection; start date of a date range (where a case count is given for a specified date range); end date of date range (where a case count is given for a specified date range); date of death; date of laboratory confirmation; date of a preliminary laboratory confirmation; start date of a mass immunization campaign; start date of implementation of vector control measures (e.g., culling); date of declaration of an epidemic or an alert is raised; date of declaration of the end of the epidemic; any earlier mentioned report or announcement date.

Some additional notes about these extracted dates are provided here. For all of these dates, the earliest mentioned date was recorded. Some of the dates of symptom onset, hospitalization or medical visit, and death may have been associated with cases that were probable but for which a final laboratory confirmation was not possible. Where a rough rather than exact date was indicated, an estimated date was noted to maximize the amount of data we could extract. For example, “early March 2005” would be translated to March 1, 2005; “mid-March 2005” to March 15, 2005; and “late March 2005” to March 31, 2005, whereas a statement such as “2 weeks ago” for a report dated October 21, 1999, would be translated to October 7, 1999. If a death or laboratory confirmation was mentioned without specifying a date, the date of the report was taken for those associated columns. For a statement such as, “The Ministry of Health has reported a total of 9,391 cases and 61 deaths from 20 December 2003 to 16 February 2004,” both the column for the start date of a date range and the column for date of death were filled with December 20, 2003. However, for a statement such as, “As of 10 February 2004, the Ministry of Health has reported a total of 236 cases, including six deaths,” February 10, 2004, was placed in both the column for the end date of a date range and the column for date of death.

Date Estimation. We implemented a procedure, described here, for determining a best estimate of “outbreak milestone” dates that we were interested in—(i) outbreak start date, (ii) earliest date of outbreak discovery, and (iii) date of the earliest public communication about the outbreak—but were unknown or not reported. For date of outbreak start, we took the earliest of whichever of the following dates that were available: date of outbreak start, date of symptom onset, and date of hospitalization or medical visit. If none of these dates were available, but a start date of a date range was the earliest date of the available dates, we took the range start date as the estimated outbreak start date. For the date of outbreak discovery, we took the earliest of whichever of the following dates that were available: WHO report (DON) date, other mentioned earlier WHO date, ProMED source date, HealthMap source date, GPHIN source date, date of an announcement made by a local authority figure or medical professional, WHO notification date, hospitalization or medical visit date, outbreak detection date, laboratory confirmation date, preliminary laboratory confirmation date, date of declaration of an epidemic or an alert is raised, and any earlier mentioned an-

nouncement date. For date of public communication about the outbreak, we took the earliest of whichever of the following dates that were available: WHO report (DON) date, other mentioned earlier WHO date, ProMED source date, HealthMap source

date, GPHIN source date, date of an announcement made by a local authority figure or medical professional, date of declaration of an epidemic or an alert is raised, and any earlier mentioned announcement date.