Evaluation of a hierarchical ascendant clustering process implemented in a veterinary syndromic surveillance system

Isabelle Behaeghel [a,*], Anouk Veldhuis [b], Libo Ren [c], Estelle Méroc [a],
Frank Koenen [a], Pierre Kerkhofs [a], Yves Van der Stede [a,d], Jacques Barnouin [c],
Marc Dispas [a]

[a] Veterinary and Agrochemical Research Centre (CODA-CERVA), Brussels, Belgium
[b] GD Animal Health, Deventer, The Netherlands
[c] National Institute for Agricultural Research, Clermont-Ferrand, France
[d] Department of Veterinary Virology, Parasitology and Immunology, Ghent University, Belgium

ARTICLE INFO

ABSTRACT

Syndromic surveillance is considered as one of the surveillance components for early warning of health-related events, as it allows detection of aberrations in health indicators before laboratory confirmation. "MoSS-Emergences 2" (MoSS-E2), a tool for veterinary syndromic surveillance, aggregates groups of similar clinical observations by hierarchical ascendant classification (HAC).

In the present study, this HAC clustering process was evaluated using a reference set of data that, for the purpose of this evaluation, was a priori divided and defined as Bluetongue (BTV) positive cases (PC) on the one hand and BTV negative cases (NC) on the other hand. By comparing the clustering result of MoSS-E2 with the expected outcome, the sensitivity (the ability to cluster PC together) and specificity (the ability to exclude NC from PC) of the clustering process were determined for this set of data. The stability of the classes obtained with the clustering algorithm was evaluated by comparing the MoSS-E2 generated dendrogram (applying complete linkage) with dendrograms of STATA® software applying average and single linkage methods. To assess the systems' robustness, the parameters of the distance measure were adjusted according to different scenarios and obtained outcomes were compared to the expected outcome based on the a priori known labels. Rand indexes were calculated to measure similarity between clustering outcomes. The clustering algorithm in its default settings successfully segregated the reference BTV cases from the non-BTV cases, resulting in a sensitivity of 100.0% (95% CI: 89.0–100.0) and a specificity of 100.0% (95% CI: 80.0–100.0) for this set of data. The different linkage methods showed similar clustering results indicating stability of the classes (Rand indexes of respectively 0.77 for average and 0.75 for single linkage). The system proved to be robust when changing the parameters as the BTV cases remained together in meaningful clusters (Rand indexes between 0.72 and 1). The configurable MoSS-E2 system demonstrated its suitability to identify meaningful clusters of clinical syndromes.

* Corresponding author at: CODA-CERVA, Groeselenberg, 99 B-1180 Brussels, Belgium. Tel.: +32 02379 06 28; fax: +32 02379 06 64.
E-mail address: Isabelle.Behaeghel@coda-cerva.be (I. Behaeghel).

1. Introduction

Because of the potential major social and economic impact of emerging diseases, it is imperative to timely detect changes in the behaviour of the animal population (Dórea et al., 2011). Classical surveillance, based on reporting of and screening for diseases, is an effective tool to control diseases. Though in classical surveillance, information on health indicators like clinical signs, milk production and fertility is usually a priori not known whereupon disease detection is often based on laboratory analysis. In case of emerging diseases, detecting and diagnosing the event can take several weeks to months (Elbers et al., 2008). Therefore, the systematic collection and analysis of pre-diagnostic data has become a prominent tool in surveillance systems to support risk managers in taking decisions to prevent and/or control animal diseases. Syndromic surveillance (SyS), defined by Triple S (2011) as "the real-time (or near real-time) collection, analysis, interpretation and dissemination of health-related data to enable the early identification of the impact (or absence of impact) of potential human or veterinary public-health threats which require effective public health action", can be an added value to timely detect an outbreak. SyS does use pre-diagnostic non-specific health indicators, often gathered for other purposes than surveillance (Doherr and Audige, 2001; Katz et al., 2011; Triple S, 2011).

To detect health-related events, cases that show the same clinical signs should preferably be grouped into a same cluster (Warns-Petit et al., 2010). Cluster analysis is the organization of objects or observations into clusters, based on similarity (Jain et al., 1999). SyS tools generally use spatio-temporal analysis to define clusters of potential outbreaks: Hyder et al. (2011) performed this type of analysis on laboratory test results for which a final diagnosis could not be reached whereas Odoi et al. (2009) focused on clustering abortions in mares in time and space. Hierarchical ascendant classification (HAC) is nowadays frequently used to identify clusters of similar observations among descriptions of undiagnosed cases. Two recent European veterinary SyS systems included HAC in their clustering method: (i) for the analysis of necropsy descriptions obtained from wildlife, in order to identify clusters of pathological profiles (Warns-Petit et al., 2010) and (ii) to detect meaningful syndromes in the descriptions of animals that were condemned in French slaughterhouses, using the available demographic data of these animals and the reason for condemnation of their carcasses (Dupuy et al., 2013).

A third system, "MoSS-Emergences2" (MoSS-E2), aims to cluster atypical cases using the clinical descriptions of the observed cases as reported by veterinarians and spatio-temporal information, as soon as possible after the occurrence of health events (Herr et al., 2009) in livestock. Hence, clinical data can be collected and analyzed before obtaining laboratory test results or observing mortality. As this system allows the clustering, using a HAC algorithm without any a priori information about the final diagnosis, it can be defined as an 'unsupervised clustering system'. In this type of unsupervised clustering, a syndrome can be defined as a combination of clinical signs that repeatedly occurs in different observations, indicating a possible presence of disease (Fricker et al., 2008).

In the present study, the unsupervised clustering process of MoSS-E2 was evaluated by using a reference set of 60 real observed cases, which were a priori defined as Bluetongue (BTV) positive (case) or BTV negative (control). Performance of the clustering algorithm was assessed by calculating the (i) sensitivity (Se: the ability to include all BTV positive cases in a cluster) and specificity (Sp: the ability to exclude non-BTV cases from BTV clusters) for this set of data, (ii) the stability (consistency between different linkage methods) of the classes and (iii) the robustness (the degree of similarity between the MoSS-E2 outcomes and the expected outcome based on known labels) of the clustering procedure.

2. Material and methods

2.1. Short description of MoSS-E2

Briefly, MoSS-E2 is an online tool for the voluntary registration of atypical cases by veterinarians and a communication platform to support rapid diagnosis under the leadership of voluntary veterinary experts, based on sharing information and opinions on specific forum pages. A combination of clinical signs is considered as atypical when this combination is unknown or is unusual in the affected species or when the affected animals do not respond to the classical (or most commonly used) treatments. Each MoSS-E2 record contains information on clinical signs and affected body parts, animal species and spatio-temporal localization of the observation. A short epidemiological inquiry is also included to gather information on breed, age, breeding type, duration of disease, laboratory results, mortality and morbidity. The HAC procedure of MoSS-E2 is embedded in the system and continuously applied to all registered cases. The clustering outcome is daily monitored by the MoSS-E2 administrator, a veterinarian with epidemiological background. When the administrator distinguishes a meaningful cluster in the clustering outcome (first validation, based on expert opinion of the administrator using a descriptive summary of the characteristics of the cluster), he launches a cluster investigation (expertise process). At this point, the system cross-matches the specifications of the identified cluster with the profiles from all veterinary experts registered in MoSS-E2. Veterinary experts were recruited amongst the second line veterinarians of the regional veterinary associations and academics working at the veterinary universities or in the national reference laboratories in fields like internal medicine, pathology, epidemiology, virology and bacteriology. Based on the veterinary specialization of the experts, the content of the cases and the native language of the users, the best fit expert is automatically selected and invited to lead the expertise process. When the selected expert considers the cluster as relevant (second validation step), an online forum is automatically created in MoSS-E2, in which notifying practitioners and veterinary experts are invited to discuss the respective cases. Using this structured communication tool, the system aims to identify the causative agent of the detected syndrome as soon as possible. Expert advice is the