# An Assessment of Electronically Captured Data in the Patient Care Enhancement System (PACES) for Syndromic Surveillance

BCH Ang,[1] MIC Chen,[2] *MBBS, M Med (Public Health), MSc (Epidemiology)*, TLH Goh,[3] *MBBS*, YY Ng,[4] *MBBS, MRCS (A&E)*, SW Fan,[3] *MBBS, Dipl Aviation Med, M Med (Occup Med)*

## Abstract

**Introduction: A common approach to the surveillance of emerging infectious diseases and agents of bioterrorism is to analyse electronically captured data for disease syndromes. The Patient Care Enhancement System (PACES) is a form of electronic medical records presently in service in the Singapore Armed Forces (SAF). We assess the feasibility of PACES data for surveillance, describe time-trends, and identify methods of sub-analysis which could improve performance. Materials and Methods: Medical consults from July 2000 to June 2003 were extracted. Diagnosis codes were mapped to 7 infectious disease syndromes according to the categorisation in the Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE): gastrointestinal (GI), fever (FEVER), respiratory, (RESP), coma (COMA), neurological (NEURO), dermatologic-haemorrhagic (DERMHEM) and dermatologic-infectious (DERM-INF). Results: A total of 732,233 episodes of care were analysed. Weekly periodicity was observed, with decreased weekend consults; there were no obvious seasonal trends in any of the syndromes. RESP, FEVER and GI syndromes were common events. Sub-analyses, either by restricting to cases with a repeated consultation, or grouping the data by medical centres, could dramatically lower thresholds used to flag outbreaks. Conclusion: In spite of the level of background noise inherent in a system consisting mainly of primary care consults, sub-analysis by medical centre, or restriction to cases with repeated consults were able to yield sensitive thresholds for outbreak detection.**

**Ann Acad Med Singapore 2005;34:539-44**

**Key words: Computerised medical records systems, Disease outbreaks, Primary healthcare, Statistics and numerical data**

## Introduction

There is growing awareness of the need for efficient, real-time identification of infectious disease outbreaks. The bioterrorism attacks involving anthrax bacilli,[1] and the sudden emergence of the Severe Acute Respiratory Syndrome (SARS),[2] a previously unknown infectious disease, have emphasised the reality of both intentional and naturally occurring outbreaks. Early detection of an outbreak is important for both outbreak control and the clinical management of cases.[3] Disease notifications, which rely on physicians to notify suspicious cases, often occur after the definitive diagnosis, and hence may be unable to provide early warning for novel infectious agents.

One suggested means of overcoming the deficiencies in traditional disease notification is through the surveillance of acute disease syndromes. Commonly, "syndromic surveillance systems" utilise data obtained from Emergency Departments (ED) by forms[4] or logs[5] of diagnoses and chief complaints; there are also systems utilising the automated extraction of data from electronic medical records (EMR).[6]

[1] Yong Loo Lin School of Medicine
National University of Singapore, Singapore
[2] Department of Clinical Epidemiology
Tan Tock Seng Hospital, Singapore
[3] Chemical, Biological and Radiological Defence Branch
Headquarters Medical Corps, Singapore Armed Forces, Singapore
[4] Preventive Medicine Branch
Headquarters Medical Corps, Singapore Armed Forces, Singapore
Address for Reprints: Mr Bryan Ang Chin Hou, 21-A, Ah Soo Garden, Singapore 539983.
Email: tootle@singnet.com.sg

The Patient Care Enhancement System (PACES) is a form of EMR that has served as the universal medical records system for the Singapore Armed Forces (SAF) in recent years. Health information from all primary care consults of servicemen is keyed into PACES directly by medical officers throughout all SAF medical centres and transmitted to a central data centre.

## Materials and Methods

All medical consults from the period July 2000 to June 2003 were included following their extraction from the PACES computer system. We opted for the most recent 3 years of data, as this was a manageable time period which would allow us to avoid any surveillance artifacts from a longer time-series, and yet was long enough to allow seasonal patterns for any of the disease syndromes to be apparent. Fields extracted included age, date of consult, medical centre and ICD-9-CM (International Classification of Diseases Clinical Modification)[7] diagnosis code.

The ESSENCE (Electronic Surveillance System for the Early Notification of Community-based Epidemics) classification was used to map ICD codes to disease syndromes. ESSENCE was developed by the Department of Defence-Global Emerging Infections Surveillance and Response System (DoD-GEISaRS) in the United States,[8] and includes the following syndromic categories: gastrointestinal (GI), fever (FEVER), respiratory (RESP), coma (COMA), neurological (NEURO), dermatologic-haemorrhagic (DERMHEM) and dermatologic-infectious (DERMINF) syndromes. In a minor modification to ESSENCE-I, several diagnoses seen in the local context were mapped to the following syndromes:

- FEVER: Infectious Mononucleosis (ICD 075), Other Viral Disease (ICD 078), Dengue Fever (ICD 061)
- RESP: Meliodosis (ICD 025), Primary Tuberculosis (ICD 010), Pulmonary
- Tuberculosis (ICD 011)
- NEURO: Tuberculosis Meningitis (ICD 013)
- Diagnoses not categorised into acute disease syndromes are henceforth referred to as “OTHER” consults.

In assessing weekly periodicity, mean consults per day were computed for the days of the week. In other analyses, 7-day moving aggregates were used. Repeat consults were defined as consults wherein a patient had an episode of care for any of the acute disease syndromes in the preceding week. This definition allowed for undifferentiated illnesses, such as a febrile disease that eventually results in a rash, to be counted as a repeat consult at the later visit.

In the sub-analysis by the medical centre, the centres were grouped by the mean number of consults a week, with median, mean, and standard deviation computed for each group. Medical centres with fewer than 100, 100 to 200, and more than 200 mean number of consults per week were arbitrarily classified as small, medium and large respectively. For the purpose of outbreak surveillance, threshold levels were arbitrarily set at 2 standard deviations above the mean number of consults for that disease syndrome. All statistical analyses were performed using SPSS 10.0.

## Results

A total of 732,233 episodes of care were studied (Table 1). The 20-24 age group accounted for the majority of the episodes. The percentage of episodes that each age group contributed remained fairly consistent through the years, varying within 2% to 3%.

Figure 1 presents the weekly pattern of consults for the FEVER, GI, RESP syndromes and OTHER consults. The mean episodes for NEURO, DERMHEM, DERMINF were 0.0694, 0.0511 and 0.695 respectively, and there were only 2 episodes for COMA; these are hence not shown in Figure 1.

Figures 2a, 2b, 2c and 2d give the plots of 7-day moving aggregates for RESP, FEVER, GI and OTHER consults respectively, while Figure 2e shows the less common disease syndromes. There is no clear seasonal trend over the 3 years. However, there are peaks that indicate sporadic outbreaks in specific diseases. For example, the peak in the DERMINF graph (Fig. 2e) arising around the week of 11/11/2000 corresponds to an outbreak of coxsackie viral disease (ICD-9 code 074). As this could possibly present as Hand-Foot-Mouth-Disease (HFMD), we had mapped this as a DERMINF syndrome. However, it could alternatively have denoted an outbreak of conjunctivitis, which is also

Table 1. Distribution of Consults Analysed by Age and Time Period (n = 732,233)

| Age group | Number of consults (%) | | | All time periods |
|---|---|---|---|---|
| | 1 Jul 2000 to 30 Jun 2001 | 1 Jul 2001 to 30 Jun 2002 | 1 Jul 2002 to 30 Jun 2003 | |
| 15 to 19 | 47,966 (17.8%) | 40,445 (17.6%) | 32,808 (14.1%) | 121,219 (16.6%) |
| 20 to 24 | 155,478 (57.8%) | 134,902 (58.6%) | 135,799 (58.3%) | 426,179 (58.2%) |
| 25 to 29 | 24,725 (9.2%) | 22,714 (9.9%) | 27,088 (11.6%) | 74,527 (10.2%) |
| 30 and above | 40,762 (15.2%) | 32,306 (14%) | 37,240 (16%) | 110,308 (15.1%) |
| Total | 268,931 | 230,367 | 232,935 | 732,233 |

Table 2. Repeat Consults for RESP, GI and FEVER Syndromes (Based on 7-day Moving Aggregate for 1095 days)

| | RESP | | GI | | FEVER | |
|---|---|---|---|---|---|---|
| | All | Repeat | All | Repeat | All | Repeat |
| ***Median*** | 1390 | 128 | 354 | 31 | 38 | 3 |
| Mean | 1400.2 | 132.4 | 354.4 | 33.8 | 39.8 | 3.8 |
| 2SD upper limit | 2052.7 | 224.8 | 514.3 | 61.9 | 63.5 | 8.8 |
| Minimum | 596 | 22 | 151 | 4 | 10 | 0 |
| Maximum | 2514 | 300 | 626 | 110 | 89 | 15 |
| Ratio of means | 0.093 | | 0.093 | | 0.094 | |
| Ratio of 2SD upper limit | 0.130 | | 0.142 | | 0.198 | |

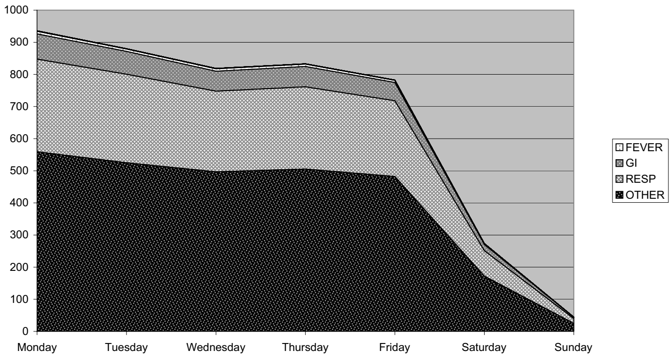


Fig. 1. Mean episodes for each syndrome by day of the week.

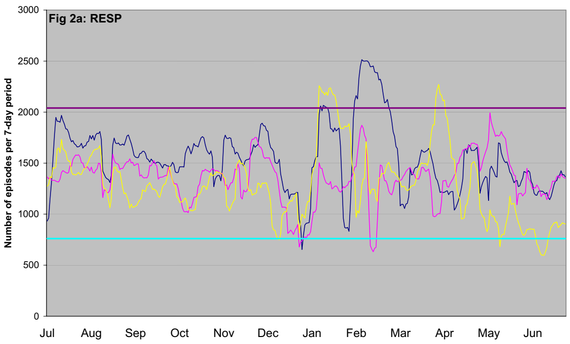


Fig. 2a.

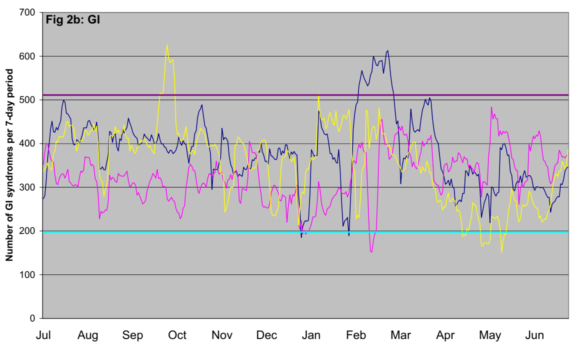


Fig. 2b.

caused by the Coxsackie virus. The intermittent peaks from January to March 2003 were caused by clusters of chickenpox cases (ICD-9 code 052). The increase in "RESP" and "GI" syndromes in February to March 2001 coincided with an increase in "OTHER" consults, and is likely due to an increase in the total population serviced by the medical centres; unfortunately, we are unable to verify this, as detailed historical records of total population served by time period were not available for this study.

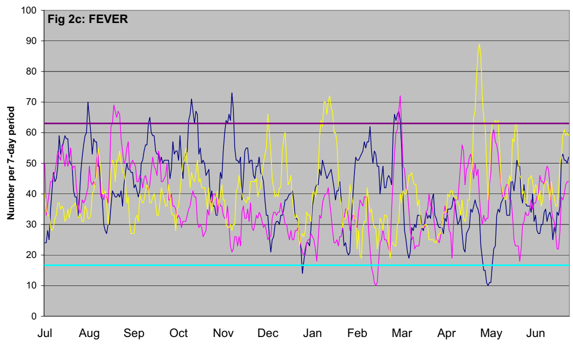


Fig. 2c.

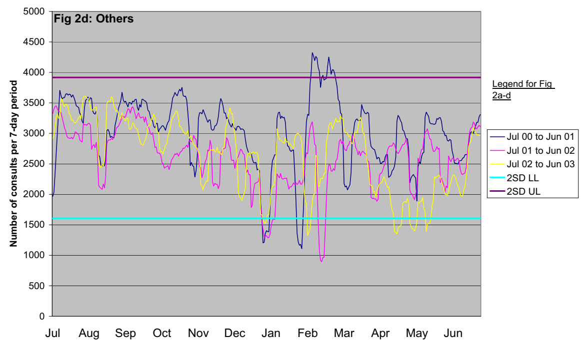


Fig. 2d.

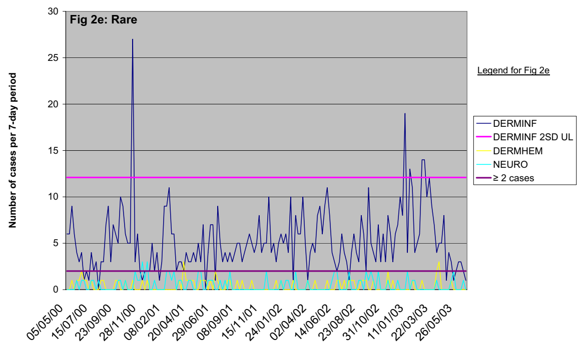


Fig. 2e.

Fig. 2. Seven-day moving aggregate of episodes for syndromic groups.

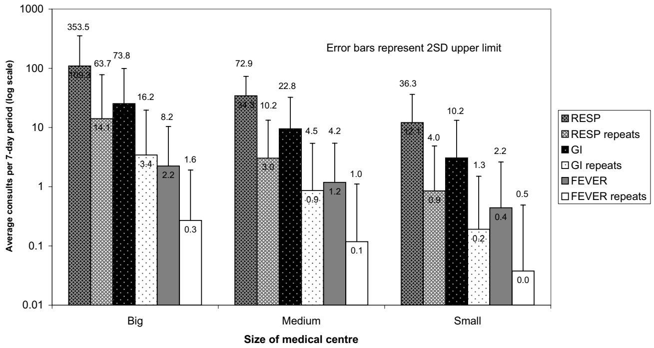


Fig. 3. Syndromes by medical centre size.

An analysis of repeat consults is presented in Table 2 for RESP, GI and FEVER syndromes. If the respective 2SD upper limits for repeat consults are used as thresholds, the system potentially detects far smaller outbreaks. For example, if an outbreak is superimposed on the 7-day period with the minimum number of RESP consults, it would have taken close to 1500 extra episodes to generate an alert; in contrast, if repeat consults were to be used, only about 200 episodes would have been required.

Figure 3 presents the number of consults sub-analysed by the medical centre. In the smaller medical centres, 2SD as an upper limit is sensitive to outbreaks of about 10 cases or less for all but the RESP syndromes. For diseases presenting with RESP syndrome at the repeat consult, thresholds can be set at 10 cases even for medium-sized medical centres.

## Discussion

This study is one of the largest collections of primary care consults known, consisting of over 730,000 episodes of care. In keeping with a primary care setting, weekly periodicity was observed, with increased caseloads on Mondays, declining to a minimum on Saturdays (half-day of work) and Sundays (emergency cases only). The use of a 7-day moving aggregate was a convenient method for smoothening out variations by day of week. However, smaller time-windows may perform better in detecting a surge of cases occurring after a simultaneous exposure. Therefore, other methods of adjusting for decreased caseloads on weekends have to be explored.

Three years was a reasonable span of time within which to observe for seasonal variations. Although seasonal peaks have been observed for respiratory diseases in temperate climates,[9] no such pattern was detected in this tropical country. However, fluctuations in the size of the populations served by the medical centres are to be expected; such changes will have to be accounted for when setting surveillance thresholds.

Sensitive thresholds can be set with present numbers for coma, neurological, dermatologic-haemorrhagic and dermatologic-infectious syndromes. However, some form of sub-analyses are required to improve the signal-to-noise ratio for respiratory, gastrointestinal and fever syndromes, as conditions which present in this fashion are common in primary care. Unpublished data from Tan Tock Seng Hospital showed that repeat medical consults occurred in healthcare workers prior to admission in the earlier part of the SARS outbreak. Hence, restricting the analyses to repeat consults should improve the performance of the surveillance system. The main drawback would be the delay in awaiting repeat consults to present. Other less time-sensitive indicators of severity, such as vital signs, are being explored.

By resolving to medical centre level, performance in smaller medical centres is impressive. Thresholds could be set at fewer than 40 consults-a-week for RESP syndrome; in contrast, the paper by Reis and Mandl[9] estimates that 30-visits-per-day outbreaks were required for the ED under study, with equivalent specificity.

There are several limitations to the study. Firstly, detailed data on the denominator population served over different time periods were not available for this study, and the best available option was to use the number of episodes from the category of "Other" consults as a proxy indicator for the size of the denominator population served. It would be better, in any prospective system, to use actual denominators, so that rates of various syndromes can be computed, hence reducing surveillance artifacts arising from changes in the size of the population served. Secondly, information on key symptoms is not available in the present PACES system. Assumptions were thus made as to which syndrome each diagnosis code should be mapped to. These assumptions may not always be valid. This appeared to have been the case in the outbreak caused by Coxsackie viral disease. Coxsackie viruses can present with various clinical syndromes, ranging from herpangina to conjunctivitis to HFMD. Unfortunately, we were unable to retrieve the record of what the outbreak actually was. HFMD is a notifiable disease within the SAF. As no outbreaks of HFMD were reported in the SAF at the time, we think it is more likely that the cluster of cases represented an outbreak of acute viral conjunctivitis, but we are unable to verify this. Another example on the ambiguity posed by using ICD-9 codes is the case of dengue fever. This disease will present in the early stages as a "FEVER" syndrome, but may progress to a DERMHEM syndrome should haemorrhagic tendencies manifest; however, it was assumed here that dengue always presents as FEVER, regardless of disease stage.

One other limitation of a system based on PACES is incomplete data capture. Soldiers can report sick to civilian medical centres when they are off-duty, but there is a mechanism for such medical leave to be notified to their respective medical centres. However, this study was unable to assess the impact of off-duty medical leave on any potential surveillance system based on PACES data. The actual extent of incomplete data capture can be assessed, and the impact of this can be estimated using a sensitivity analysis, to see if anticipated outbreak sizes are able to give rise to a surveillance signal even when a proportion of cases do not immediately report their illness to the medical centre – these can be the subjects of separate studies.

A related limitation is that analysis by unit sub-groupings assumes transmission within SAF premises, and not in the general community. To detect outbreaks where the exposures were community-based, data from SAF centres will have to be pooled with information from cases presenting to civilian institutions, as part of a national surveillance system.

Our descriptive study also did not attempt to estimate measures of performance for any prospective surveillance system based on the PACES data. Possible approaches would be to compare surveillance signals based on the data from PACES with historical data from reported outbreaks, as well as to perform simulations using artificially generated outbreak data, such as that attempted by Reis and Mandl.[9] These can be the subject of future studies, and are beyond the scope of our present evaluation.

Other areas for future work include comparing the trends detected in this system with those detected by national level syndromic surveillance systems based on primary care acute respiratory infection consults and diarrhoeal disease, as well as non-syndromic systems such as the national level laboratory-based surveillance system for influenza. It would be interesting, in particular, to examine whether heightened influenza activity, as indicated by increased proportions of laboratory isolates testing positive for influenza, is reflected in any of the peaks observed in the numbers of consults for the RESP syndrome. The same dataset from PACES can also be used to examine how other outbreaks with specific aetiologies would present, such as epidemics of viral conjunctivitis, for which national data from both primary care consults and laboratory surveillance are also available. Such studies are potential areas for collaborative work between the SAF and the Ministry of Health, and consists of worthwhile follow-up work.

The ESSENCE syndromic surveillance system has been in use by the US military since 1999, and has shown itself capable of picking up outbreaks.[10] Another community-based system using retail data for over-the-counter medications could detect increased seasonal activity of paediatric gastrointestinal and respiratory illnesses in children.[11] With some calibration, a system based on PACES data is likely to be able to do the same, although this will have to be validated, either through historical outbreak data or through prospective data collection.

While the above assessment adds several insights into the possible use of electronic data for syndromic surveillance, it should be noted that simple thresholds are already in use at the level of medical centres, and have thus far proved useful in detecting naturally occurring outbreaks. Sensitive detection limits, whether based on simple thresholds of case-counts, or more sophisticated analysis, can be effective in the armed forces as it is made up of a well-circumscribed population of young, healthy adults, unlike in the general community. As such, it may be inappropriate to extrapolate the above models for use outside of the military context.

## Conclusions

Most syndromic surveillance systems utilise information from emergency departments.[7,8,11] This paper is amongst the first to propose a viable syndromic surveillance system in a primary care setting. It shows that utilising repeat consults has potential in improving the signal-to-noise ratio, and that resolving to the level of subunits at risk can detect outbreak sizes difficult to achieve with community-based systems. It remains to be seen if syndromic surveillance will add any value in the SAF, but the analysis does support the present practice of using simple thresholds of case counts at the medical centre level to detect outbreaks.

## Acknowledgements

The authors acknowledge the Chief of Medical Corps, BG (Dr) Wong Yue Sie, and Senior Medical Officer Healthcare, Col (Dr) John Wong, for the use of PACES data for this analysis, and for their support of the project.

REFERENCES

1. Greene CM, Reefhuis J, Tan C, Fiore AE, Goldstein S, Beach MJ, et al; CDC New Jersey Anthrax Investigation Team. Centers for Disease Control and Prevention. Epidemiologic investigations of bioterrorism-related anthrax, New Jersey, 2001. Emerg Infect Dis 2002;8:1048-55.
2. Centers for Disease Control and Prevention. Update: Outbreak of severe acute respiratory syndrome—worldwide, 2003. MMWR Morb Mortal Wkly Rep 2003;52:241-8. Erratum in: MMWR Morb Mortal Wkly Rep 2003;52:284.
3. Centers for Disease Control and Prevention. Biological and chemical terrorism: strategic plan for preparedness and response. MMWR Recomm Rep 2000;49(RR-4):1-14.
4. Das D, Weiss D, Mostashari F, Treadwell T, McQuiston J, Hutwagner L, et al. Enhanced drop-in syndromic surveillance in New York City following September 11, 2001. J Urban Health 2003;80:i76-88.

5. Begier EM, Sockwell D, Branch LM, Davies-Cole JO, Jones LH, Edwards L, et al. The National Capitol Region's Emergency Department syndromic surveillance system: do chief complaint and discharge diagnosis yield different results? Emerg Infect Dis 2003;9:393-6.
6. Lazarus R, Kleinman KP, Dashevsky I, DeMaria A, Platt R. Using automated medical records for rapid identification of illness syndromes (syndromic surveillance): the example of lower respiratory infection. BMC Public Health 2001;1:9.
7. ICD-9-CM international classification of diseases, ninth revision, clinical modification, 6th edition. National Center for Health Statistics (US). Washington D.C.: US Dept of Health and Human Services, Centers for Disease Control and Prevention, Centers for Medicare and Medicaid Services, 2002.
8. US Department of Defense GEISaRS. ESSENCE: Syndromic surveillance to track emerging infectious diseases. Available at: http://www.geis.ha.osd.mil/GEIS/SurveillanceActivities/ESSENCE/ESSENCE.asp. Accessed June 2003.
9. Reis BY, Mandl KD. Time series modeling for syndromic surveillance. BMC Med Inform Decis Mak 2003;3:2.
10. Pavlin JA. Investigation of disease outbreaks detected by "syndromic" surveillance systems. J Urban Health 2003;80:i107-14.
11. Hogan WR, Tsui FC, Ivanov O, Gesteland PH, Grannis S, Overhage JM, et al; Indiana-Pennsylvania-Utah Collaboration. Detection of pediatric respiratory and diarrheal outbreaks from sales of over-the-counter electrolyte products. J Am Med Inform Assoc 2003;10:555-62.