

## Original Articles

# A reliable method to retrieve Accident & Emergency data stored on a free-text basis

U. N. PREMARATNE\*, G. B. MARKS\*, E. J. AUSTIN<sup>†</sup> AND P. G. J. BURNEY<sup>†</sup>

\*Department of Public Health Medicine, UMDS, St Thomas' Hospital, London SE1 7EH, U.K.

<sup>†</sup>Information Department, Greenwich Healthcare NHS Trust, The Memorial Hospital, London SE18 3RZ, U.K.

Accident & Emergency (A & E) data on asthma-related attendances are useful for studies on the effectiveness of asthma interventions, and to determine the relationship of environmental factors to asthma and asthma epidemics. The final diagnoses made in the A & E departments are not usually coded when entered into hospital databases in the U.K., although the 'presenting complaint' can be retrieved from the computerized Hospital Information & Support Systems (HISS), from a free-text attendance diagnosis field entered by the reception clerk when the patient arrives at the A & E department. The validity of this as an indication of the final diagnosis is unevaluated. The aim of this study was to measure the validity of the string 'asth' in the A & E attendance diagnosis field for identifying patients attending the A & E departments of two hospitals for asthma-related conditions. A list of patients who attended the A & E department of two hospitals was retrieved from the HISS along with the attendance diagnosis field. If the attendance diagnosis field contained the text string 'asth', mentioned wheeze or breathing problems, or the patients were referred by their GP without any diagnostic information entered on HISS, the records were selected for evaluation. The remaining attendances, which were not evaluated further, were attributed to another diagnosis based on the evidence of the recorded attendance diagnosis. The results indicated that the string 'asth' in the attendance diagnosis field had a sensitivity of 80.3% (95% CI 75.1-85.5%) and a specificity of 96.7% (95% CI 95.6-97.8%) for a final diagnosis of asthma. It is concluded that free-text attendance diagnosis fields in hospital databases can be searched with suitable strings to obtain reliable data on attendance with asthma. As part of another investigation, the present authors attempted to retrieve a list of the attendances with asthma at the same two A & E departments at a time that was reportedly associated with an epidemic of asthma following a thunderstorm. On this occasion, the string 'asth' proved to be significantly less sensitive. The possible reasons for this and the implications for using this method for identifying cases are discussed.

RESPIR. MED. (1997) 91, 61-66

## Introduction

There is a substantial, and probably increasing, burden of ill-health attributable to asthma in the U.K. (1). The Department of Health believes that further research and development is required

before national targets can be set for asthma; hence, although a strong candidate, it was not identified as a Key Area in the Health of the Nation statement (2), which defines the government's targets for prevention. Evidence of the effectiveness of interventions targeted across the spectrum of asthma care is poor. However, there have been encouraging reports that tackling severe asthma can reduce hospital admissions (3). Asthma is also an important acute respiratory problem where patients commonly present at Accident & Emergency (A & E) departments. A

Received 26 August 1995 and accepted in revised form 10 March 1996.

Correspondence should be addressed to: U. N. Premaratne, Department of Public Health Medicine, UMDS, South Wing Block 8, Basement, St Thomas' Hospital, Lambeth Palace Road, London SE1 7EH, U.K.

recent survey showed that 51% of subjects with asthma had sought urgent medical help from their general practitioner (GP) or an A & E department (Marks, pers. comm.). Therefore, A & E attendance data could be used as an important outcome indicator when investigating improved asthma management in the community.

Visits to A & E departments have also been used in the past to determine the relationship of environmental factors to asthma (4) and to detect asthma epidemics (5–8). Although a wealth of information is available in many A & E units, attendance diagnoses stored in free-text format are difficult and time-consuming to retrieve accurately. It is important to estimate the validity and reliability of the information that is already stored if it is to be retrieved and used for research purposes (9).

In 1991, the first national pilot site for a fully computerized integrated Hospital Information Support System (HISS) was established at Greenwich, covering the full range of hospital services: inpatient, A & E, pathology, radiology and pharmacy, as well as the services of professions allied to medicine. This original system makes use of free-text descriptions when diagnoses are entered. Since that time, a number of new HISS developments have emerged across the U.K., some of which are now more advanced than the early modules developed at Greenwich. The new systems rely on full and accurate coding of diagnoses and procedures in A & E, making the analysis of data both easier and more reliable.

It is difficult to retrieve patient information from A & E clinical records, and the accuracy of the data collected is in question. What is noted as the attendance diagnosis in most cases is what the patient tells the admissions clerk on arrival at the A & E unit. This diagnosis depends on the patient's perception of the illness and what the admissions clerk finally records, and it is this unvalidated account which often remains as the final diagnosis on the HISS at Greenwich. This may or may not equate with the final medical diagnosis. However, these computerized data are the only data available, short of retrieving and abstracting all the written clinical records, and it is therefore important to know whether the data retrieved are accurate enough to be used for research purposes.

The value of these free-text data for identifying patients attending the A & E department for

asthma or an asthma-related condition were assessed. The aim of this study was to measure the sensitivity and specificity of the term 'asth' in the 'attendance diagnoses' field in the data stored in the HISS as a marker for a diagnosis of asthma, so that the original system which makes use of free-text descriptions can still be used as a good source of data on A & E attendances for asthma or related conditions.

A thunderstorm-associated asthma epidemic was reported in June 1994. There was a significant increase in the number of patients with asthma attending A & E departments across London and other parts of England following the thunderstorm (10,11). This epidemic was studied using this marker in order to test its usefulness in a research context.

## Methods

Records of patients visiting the A & E departments at Greenwich District Hospital and Brook Hospital with conditions related to asthma were studied over 3 months from January to March 1994. In these hospitals, data on each A & E attendance are stored in a computerized HISS. The attendance diagnoses were entered into the information system on a free-text basis by the admissions clerk prior to the patient being seen by a medical officer.

Records of all patients who attended the A & E departments of Greenwich District Hospital and Brook Hospital during January–March 1994 were scrutinized. Initially, a complete list of patients who attended the A & E departments during the 3-month period was retrieved from the HISS, together with their attendance diagnoses. After going through their free-text 'attendance diagnoses', those who attended the A & E department for conditions which clearly were unlikely to be asthma were identified and removed. The rest of the patients identified on free-text diagnosis fell into four broad categories: those referred by their general practitioner (GP) without further diagnostic information in the HISS; those for whom asthma or some cognate term was used; those with breathing problems; and those with a wheeze. From the stated 'attendance diagnoses' itself, it was difficult to decide whether these patients in fact attended the A & E department because of a

problem related to asthma. Therefore, the records of these patients which gave more details of their visits to the A & E departments were studied to identify the final diagnosis. Some of the records, however, did not indicate a final diagnosis. In such instances, on the basis of the signs and symptoms that were noted, the treatment that was given and whether or not the patient's symptoms improved with treatment, a diagnosis was assigned by the observer. When notes of a casualty medical officer were not available, the reason for attending the A & E department was identified by examining other relevant information such as notes of the admissions clerk and the casualty nurse, and/or letters of GPs that were sent with patients. The records that were either so incomplete that a diagnosis could not be assigned to them or were missing and could not be traced, were not included in the validity and reliability calculations.

A diagnosis of asthma was made if any of the following criteria were met: (1) A final diagnosis of asthma made by the medical officer; (2) Presenting symptoms of wheeze, shortness of breath, difficulty in breathing, tightness of chest or an asthma attack and, either (i) past history of asthma, or (ii) given bronchodilators at the A & E department and improvement in symptoms or peak flow, and no other main cause of the symptoms established; (3) Known asthmatic and presented (i) because of symptoms as in (2) in the recent past but which have now been resolved, or (ii) to obtain medication for asthma.

An 'attendance diagnosis' of asthma was excluded if: (1) a clear alternative diagnosis was noted by the medical officer; or (2) if there was sufficient other evidence (notes of A & E nurse/GP letter) to exclude asthma.

To estimate the reliability of this assessment of final diagnosis, a sample of records were studied by two independent observers, and the kappa coefficient was calculated to assess the inter-observer variability.

On the night of 24 June 1994 there was a thunderstorm. For several A & E departments in Southern England, a 10-fold increase in the number of asthmatic patients attending the A & E departments was reported (10,11). To observe whether the attendance diagnoses noted by the admissions clerks depend on the patient's perception of the illness, the attendance data of the

A & E departments of Greenwich District Hospital and Brook Hospital entered onto the HISS during this month and the following month were retrieved. The data were first retrieved using the string 'asth'. A second set of data were retrieved using several additional strings such as 'breath', including breathing difficulty, breathless, breathlessness etc.; 'wheez', including wheezing, wheezy etc.; 'DIB' or 'D.I.B.', common abbreviations for difficulty in breathing; and 'S.O.B.' and 'SOB', common abbreviations for shortness of breath. Using these two data sets, the number of attendances per day at the A & E departments was plotted for June and July to see whether 'epidemic' asthma days could be detected.

The case notes of the patients retrieved by the above strings were subsequently retrieved and abstracted, and the diagnoses were updated according to the medical diagnoses that were made after treatment. This was used to validate the data retrieved with different strings from the HISS.

## Results

A total of 1185 records were studied and 209 of these were not used for the calculations, either because the records were not available at the A & E departments or because they did not have sufficient data to allocate a diagnosis (Table 1). It was, however, observed that the percentage of those who had a diagnosis of asthma on the free-text admission diagnosis field was 22% in the group that was excluded, and 20.6% in the group that was further analysed in whom a final diagnosis was available.

With the results available, the specificity and the sensitivity of the stem 'asth' to retrieve patients that attended the A & E department for asthma-related conditions was calculated (Table 2). The text string 'asth' in the 'attendance diagnosis' field had a sensitivity of 80.3% (95% CI 75.1–85.5%) and a specificity of 96.7% (95% CI 95.6–97.8%) for a final diagnosis of asthma.

To assess the inter-observer agreement on arriving at a final diagnosis, a sample of 261 records were studied. The kappa coefficient was calculated ( $K=0.76$ , 95% CI 0.68–0.85), confirming that there was excellent agreement between the diagnoses of the two observers (12).

The number of patients attending the A & E departments with asthma-related conditions

TABLE 1. Details of patients (under 60 years of age) who attended the Casualty Departments of Greenwich District Hospital and Brook Hospital during January–March 1994 which could have been related to ‘asthma’

Contents of attendance diagnosis field	Final diagnosis		Records either not traced or containing no diagnosis
	Asthma	Non-asthma	
‘Asth’ <i>n</i> =247	183 (74%)	18 (7%)	46 (19%)
‘Difficulty in breathing’ related <i>n</i> =145	20 (14%)	96 (66%)	29 (20%)
‘Wheeze related’ <i>n</i> =21	5 (24%)	9 (43%)	7 (33%)
Referred by general practitioner <i>n</i> =772	20 (3%)	625 (81%)	127 (16%)

TABLE 2. Validity of diagnosis of asthma using a common string ‘asth’, from recorded casualty admission data, entered on a ‘free-text’ basis in the Hospital Information Support System (HISS)

Presentation	Diagnosis		Total
	Asthma	Non-asthma	
Asthma	183	18	201
Other	45	730	775
Total	228	748	976

Sensitivity, 80.3%; specificity, 97.6%

immediately following the thunderstorm on 24 June 1994 was increased only about two-fold (from 5 to 11) during the thunderstorm, when the data were retrieved using the string ‘asth’ (Fig. 1). However, when additional strings were used, a four-fold increase in the number of patients with acute asthma-related conditions was detected (from 12.4 to 50) compared to the average number of similar patients attending the A & E departments on other days (Fig. 2).

## Discussion

This study suggests that carefully selected strings can be used to retrieve data on patients attending A & E departments with asthma-related problems, from the data stored on a ‘free-text’ basis in the HISS.

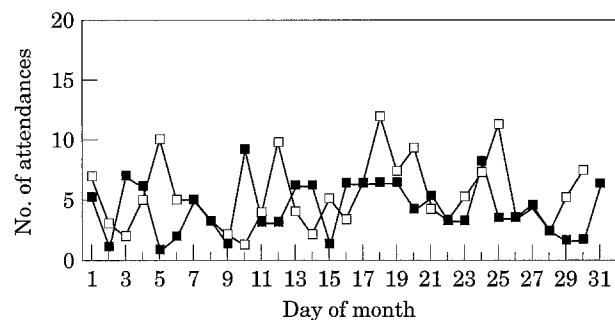


FIG. 1. Accident & Emergency department attendances during June (□) and July (■) 1994 for asthma.

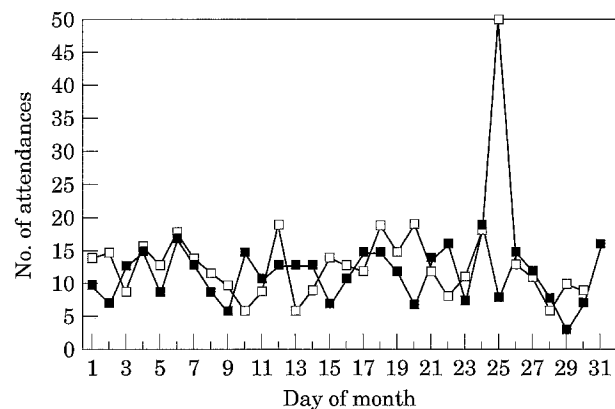


FIG. 2. Accident & Emergency department attendances during June (□) and July (■) 1994 for asthma and related reasons.

As the diagnoses are entered in free-text, care should be taken when selecting a string. In this study, it was observed that the term ‘asthma’ was sometimes misspelled (asthama, asthm), and that only the first four letters were consistent. If a

string which went beyond the first four letters were used, several of the asthma cases would not have been retrieved.

The noted admission diagnosis relied heavily on the patient's perception of whether or not they had this illness. This was confirmed when data on asthma attendances during the thunderstorm-associated asthma epidemic in June 1994 were retrieved. In an earlier study where an asthma outbreak has been reported which has been associated with exceptional weather conditions (6), the data had been collected in the usual manner of going through casualty department registers, identifying patients presenting with symptoms of respiratory disease and subsequently validating them by going through individual case records. When admission data during the June thunderstorm was retrieved from the HISS using the string 'asth', this increase in number of asthmatics was not detected (Fig. 1). However, when additional strings (breath . . . , wheez . . . , DIB, D.I.B., S.O.B., SOB) were included, an increase in the number of patients attending the A & E departments with respiratory problems was identified (Fig. 2). The explanation that could be offered is that with a string such as 'asth', those who are being detected are the patients with asthma who have already been diagnosed, who have apparently been on treatment over a long period and who are aware of their diagnosis. These patients apparently were able to inform the admission clerk that they had come with a probable asthma attack which would prompt the admissions clerk to enter 'asth' within the attendance diagnosis (e.g. 'asth'matic, 'asth'ma attack, 'asth'ma problem). The other group of 'asthmatics' consist of a group of individuals who have experienced their first-ever acute asthma attack, in this case triggered by exposure to high levels of aeroallergens during the storm (13,14). They would not yet have been labelled or diagnosed as 'asthmatics'. Thus, 'asth' would not have been entered on the free-text diagnosis and their attendances were retrieved only when additional stems were included in retrieving the data. The fact that there was an increased number of asthmatics seen during this day at Greenwich District Hospital and Brook Hospital was confirmed by going through all the admission records, as had been done in a previous study (6). Many who did

not get the string 'asth' into their admission diagnosis, but who came in with difficulty in breathing, were diagnosed as having an acute asthma attack after medical assessment.

Monitoring emergency room admissions has been found to be very useful in studying asthma epidemics in Barcelona (9), and assessing the effects of air pollution on health (15,16). However, care must be taken when retrieving data from the HISS with the method described because of the problems that have been highlighted, and one must be aware of the limitations of the methodology. The usefulness of this method is that it enables one to retrieve from the computerized databases, all possible asthmatic patients who would have presenting symptoms categorized earlier, from the total number of patients who attended the A & E departments. However, to validate the data retrieved from the present databases during unusual periods such as thunderstorms, one will have to go through the clinic records of the selected patients. On the other hand, when A & E asthma data during other periods are retrieved by this method, the results obtained are not significantly different from results obtained by manual retrieval of the same data. Even in the former situation, as going through A & E records is a laborious task, the time saved on this type of initial retrieval will be of immense value for research studies.

The way A & E admissions are entered into the HISS is changing and an attempt is being made, through the introduction of a new national A & E Minimum Dataset, to code the attendance diagnoses as confirmed by the medical officer. This change will improve the quality of data available in the future. The aim of this study was to identify ways in which the already available data could be used for epidemiological and other studies on asthma. Similar methodologies could be developed for other conditions.

## References

1. Burney PGJ. Strategy for asthma. *BMJ* 1991; **303**: 571-573.
2. Secretary of State for Health. *The Health of the Nation*. London: HMSO, 1991, Cm1523.
3. Mayo PH, Richman J, Harris HW. Results of a programme to reduce admissions for adult asthma. *Ann Intern Med* 1990; **112**: 864-871.

4. Toepfritz R, Goren A, Goldsmith JR, Donaghi A. Epidemiological monitoring: methods for analyzing routinely collected data. *Sci Total Environ* 1984; **32**: 233–246.
5. Derrick EH, Thatcher RH, Trappett LG. The seasonal distribution of hospital admissions for asthma in Brisbane. *Australas Ann Med* 1960; **9**: 180–187.
6. Packe GE, Ayres JG. Asthma outbreak during a thunderstorm. *Lancet* 1985; **2**: 199–204.
7. Salvaggio J, Seabury J, Schoenhardt A. New Orleans asthma V. Relationship between Charity Hospital asthma admission rates, semiquantitative pollen and fungal spore counts, and total particulate aerometric sampling data. *J Allergy* 1971; **48**: 96–114.
8. Goldstein IF, Cuzick J. Daily patterns of asthma in New York City and New Orleans: an epidemiological investigation. *Environ Res* 1983; **30**: 211.
9. Martinez F, Sunyer J, Anto JM. Reliability of a monitoring system for respiratory emergency room admissions. *Eur Resp J* 1993; **6**: 337–341.
10. Murray V, Venables K, Laing-Morton T, Partridge M, Thurston J, Williams D. Epidemic of asthma possibly related to thunderstorms. *BMJ* 1994; **309**: 131–132.
11. Higham JH. Thunderstorm peak in Luton. *BMJ* 1994; **309**: 604.
12. Beard CM, Yunginger JW, Reed CE, O'Connell EJ, Silverstein MD. Interobserver variability in medical record review: an epidemiological study of asthma. *J Clin Epidemiol* 1992; **45**: 1013–1020.
13. Bellomo R, Gigliotti P, Treloar A *et al*. Two consecutive thunderstorm associated epidemics of asthma in the city of Melbourne. The possible role of rye grass pollen. *Med J Aust* 1992; **156**: 834–837.
14. Knox RB. Grass pollen, thunderstorms and asthma. *Clin Exp Allergy* 1993; **23**: 354–359.
15. Bates DV, Baker-Anderson M, Sizto R. Asthma attack periodicity: a study of hospital emergency visits in Vancouver. *Environ Res* 1989; **51**: 51–70.
16. Sunyer J, Anto JM, Murillo C, Saez M. Effects on urban air pollution on emergency room admissions for chronic obstructive pulmonary disease. *Am J Epidemiol* 1991; **134**: 277–286.