ELSEVIER

Bayesian Estimation of Asthma Prevalence, and Comparison of Exercise and Questionnaire Diagnostics in the Absence of a Gold Standard

KITAW DEMISSIE, MD, PhD, NEIL WHITE, MD, LAWRENCE JOSEPH, PhD, AND PIERRE ERNST, MD

PURPOSE: To estimate the sensitivities, specificities, and predictive values of exercise challenge and questionnaire, when these tests are used to diagnose asthma in children.
METHODS: Participants were children, predominantly aged 6 to 12 years, selected from three primary school grades among 18 different schools in Montreal. Of 1111 participants, 989 successfully completed a six-minute free running test at school and returned a respiratory questionnaire. A total of 952 children had complete information that could be used for analysis. A history of wheezing in the past year in conjunction with a past diagnosis of asthma defined current asthma by questionnaire. Exercise responsiveness was defined as a $\geqslant$ 10% fall in FEV_1 after a six-minute free run. As there is no perfectly accurate diagnostic test for asthma, we analyzed the data using a previously published Bayesian method that allows for the estimation of test properties when no gold standard test is available.
RESULTS: Current asthma by questionnaire was found to have significantly higher specificity (94.9%, 95% credible interval (CI): 93.2–96.5 versus 82.6%, 95% CI: 79.9–85.1) and positive predictive value (53.8%, 95% CI: 41.0–66.7 versus 19.2%, 95% CI: 12.3–27.8) in comparison to exercise challenge. While there was no statistically significant difference between the two tests with respect to sensitivity and negative predictive values, the estimates were higher for current asthma (64.4%, 95% CI: 50.9–76.6 and 96.7%, 95% CI: 94.6–98.1, respectively) in comparison to exercise challenge (51.3%, 95% CI: 37.8–64.5 and 95.4%, 95% CI: 93.2–97.1, respectively). Agreement between the two diagnostic methods was poor and the combined use of the two tests did not significantly improve the likelihood of correctly identifying children with asthma.
CONCLUSIONS: Our findings support the view that exercise testing adds little to a well designed questionnaire in identifying subjects with asthma in community based studies.
Ann Epidemiol 1998;8:201–208.

KEY WORDS: Sensitivity, Specificity, Asthma, Predictive Values, Bayesian, Exercise, Questionnaire, Gibbs Sampler.

INTRODUCTION

Asthma is a clinical syndrome of reversible airway obstruction with accompanying inflammation. The term asthma is used to embrace a range of airway conditions manifesting in increased airway resistance and increased responsiveness (1). There is no universally accepted definition, nor are there symptoms that are wholly specific for asthma. Therefore, establishing the diagnosis on clinical grounds is sometimes difficult. The recognition that enhanced airway responsiveness may accompany asthma has made provocation testing an important adjunct to the diagnosis of asthma (2, 3). Respiratory questionnaires, provocation tests by pharmacological agents (histamine and methacholine), and physiological stimuli (exercise) have all been used as diagnostic tests and to assess the severity of asthma (4–6).

Large differences between countries in the prevalence of asthma have been reported (7). Comparison of results from different studies is difficult because of the different methods used to define asthma. Recently, the use of both questionnaire responses and airway provocation tests were suggested in defining asthma for epidemiological studies (8).

For population studies, exercise challenge tests are thought to be more readily accepted by parents and children than pharmacological stimulation because of the physiological nature of the provocation which is more readily received by parents and children. A recent study by Ninan and Russel (9) however, suggested that exercise testing enhanced by cold air adds very little to a well designed respiratory questionnaire. Most of the studies which compared the test characteristics (sensitivity, specificity, and predictive val-

From the Respiratory Epidemiology Unit, (K.D., N.W., P.E.), and the Department of Epidemiology and Biostatistics (L.J.), McGill University, Montreal, Quebec, Canada.
Address reprint requests to: Kitaw Demissie, M.D., Ph.D., Department of Family Medicine, UMDNJ—Robert Wood Johnson Medical School, One Robert Wood Johnson Place—CN19, New Brunswick, New Jersey 08903-0019.
Received March 10, 1997; revised August 11, 1997; accepted August 25, 1997.

1047-2797/98/$19.00
PII S1047-2797(97)00168-3

Selected Abbreviations and Acronyms

CI = credible interval
EIB = exercise induced broncoscopy
FEV_1 = forced expiratory volume in 1 second
FVC = forced vital capacity
SES = socioeconomic status
ROC = receiver operating characteristics

ues) of exercise have been limited to hospital or clinic patients who are unlikely to be representative of the general population (6, 10–17). Furthermore, the test characteristics of exercise challenges have often been evaluated against a questionnaire (inquiries or clinical diagnoses) used as the gold standard (18). In fact, one may argue that there is no gold standard for the diagnosis of asthma, and therefore previous estimates that do not recognize this may be biased and therefore invalid. It is nevertheless important for the planning of future studies, and for clinical and public health practice, to have the best possible estimates of asthma prevalence (in any given population) and the parameters (sensitivity, specificity, positive and negative predictive values) for existing diagnostic tests.

There is no study which has compared test characteristics of respiratory questionnaire and exercise challenge on a population based sample without assuming either to be a gold standard. We therefore took the opportunity during a study of the determinants of respiratory health in Montreal school children to compare the sensitivities, specificities, and positive and negative predictive values of questionnaire response, and exercise challenge test in the diagnosis of asthma, as well as to provide a prevalence estimate in this population based on the combined results of the two tests. We used a recently developed Bayesian approach to the analysis of diagnostic test results (19, 20) that permits simultaneous estimation of all test parameters in the case when none is considered to provide perfectly accurate results.

METHODS

Study Population

Eighteen schools were selected on the island of Montreal in order to represent a broad range of socioeconomic status (SES). To achieve this, all schools in the five school boards in central Montreal, were ranked according to neighborhood average house values. Within each school board, schools were selected from upper, middle, and lower ranges of neighborhood average house value. For each of the 18 schools so selected, three classes, one class from each of grades one (6 to 7 years of age), three (8 and 9 years of age), and five (10 to 13 years of age), were selected. A total 1274 children were recruited for the study. Comparison of schools selected with those not selected showed no meaningful difference with respect to neighborhood poverty, income, and educational attainment.

Respiratory Questionnaires

A questionnaire concerning the child's health and home environment (derived from the American Thoracic Society Children's Questionnaire) was completed by parents (in 73% by biological mothers, 24% by biological fathers, and 3% by either a step or grand parent).

Symptoms and diagnosis were defined as follows: "Does the child's chest ever sound wheezy or whistling?" (ever wheeze); "Has the child had wheezing or whistling in the chest at any time in the last 12 months?" (current wheeze); "Does the child's chest usually sound congested or does the child usually cough and bring up mucus?" (cough with mucus); "Do colds usually go to the child's chest?" (chest colds); "Does the child usually cough?" (usual cough); "Has the child been coughing at night or on getting up in the morning during the past month?" (night cough); "Has the child ever been diagnosed as having asthma?" (history of asthma).

For the present analysis, if a child's parent answered "yes" to both of the following questions: "Has a physician ever diagnosed asthma in your child?" and "Has the child had wheezy breathing in the past 12 months?", he/she was defined as having "current asthma."

Lung Function

In the school gymnasium, height and weight were recorded, and inquiries were made concerning respiratory tract infections in all subjects, and current smoking among grade 5 children. Spirometry was carried out sitting and with nose clips using two Collins 10-liter water-sealed spirometers (Warren E. Collins, Braintree, MA) according to current American Thoracic Society guidelines (21) and best FEV_1 and FVC from any flow volume curve used for analysis (22).

After a five minute rest, heart rate was measured using a digital plethysmograph (Heart Rate Inc., Costa Mesa, CA). Children were then asked to run around the gymnasium for six minutes at a pace judged sufficient to attain 90% or more of the predicted maximal heart rate (23). The heart rate was remeasured immediately on completion of exercise. Five and ten minutes after completion of exercise, spirometry was repeated on the same spirometer. Temperature and humidity of the gymnasium were recorded at the time of the test. No specific instructions were given concerning the use of medications including those for asthma. The exercise test was completed successfully in 989 children. For this analysis, a positive exercise test (exercise induced bronchospasm (EIB)) was defined as 10% or more decline in FEV_1 following exercise challenge (24–26). Only 11 children had been using a β-agonists and one theophylline in the days preceding the test.

Statistical Methods

Since neither of the two tests (exercise challenge and asthma by questionnaire) can be considered as a gold standard in defining asthma in epidemiologic studies, there are five unknown parameters about which inference must be made: the population prevalence of asthma, and the sensitivity and specificity of each of the two tests. However, when two tests are applied to a sample of subjects from a given population, there are only three degrees of freedom available for inference from the resulting two-by-two table of data. Therefore, only three of the five parameters may be simultaneously estimated using standard maximum likelihood estimation techniques (27). For example, one test can be considered as a gold standard (so that sensitivity = specificity = 100% is assumed for that test), and the other test parameters, along with the prevalence of the disease, can then be estimated. While this has been used in the past, the approach likely produces biased estimates, since no test is a gold standard for asthma. Furthermore, the confidence intervals around the estimated parameters are too narrow, since they do not account for the uncertainty about the true sensitivity and specificity of the test that is falsely assumed to be perfect.

An alternative is to use a Bayesian approach that simultaneously estimates all five unknown parameters (19, 20). This methods proceeds in two steps: First, a prior distribution is obtained for each unknown parameter. This distribution summarizes the available pre-experimental information about the five parameters. Subsequently, the prior distribution is updated via *Bayes Theorem* to a posterior distribution, using the data and the usual multinomial likelihood function for two-by-two tables. Marginal posterior densities can be derived for each parameter by integration, from which 95% marginal posterior credible intervals (CI, Bayesian analogues of confidence intervals) can be calculated. Since the integration here is analytically intractable, the Gibbs Sampler, a Monte Carlo approach to calculating marginal densities (28), is employed. The above methods allow for simultaneous inferences to be made for all unknown parameters, which takes full advantage of all the information contained in the data, as well as formally incorporating prior information, when available.

The Bayesian approach is, in fact, a generalization of the standard maximum likelihood method, since it includes it as a special case in the following sense: If one chooses prior distributions that concentrate all probability mass on the values of 100% for the sensitivity and specificity of one test, and use flat non-informative prior distributions for the remaining three parameters, then the results coincide numerically with those given by the standard approach. However, this prior distribution is unreasonable, since we do not believe that one of the tests is in fact perfect. Therefore, the Bayesian approach can be seen as more realistic when no perfect test is available.

While the Bayesian approach employs more reasonable assumptions, this advantage comes at a cost of having to specify a prior distribution for each parameter from the information available before the analysis of the data from the current study. Since the choice of prior distribution is not unique, the final posterior distributions are also not unique. Therefore, it is usual to calculate posterior distributions from more than one prior distribution. Here, two analyses were carried out, one for each of the following strategies that were employed to derive prior distributions.

The first involves informative prior distributions for all parameters. A panel of four experts on childhood asthma were asked to provide their best estimate of the sensitivities and specificities of exercise challenge and current asthma by questionnaire, as well as the prevalence of asthma in children of this age group. Prior distributions that summarize the information provided by the experts about each parameter were derived. Table 1 provides the 95% credible intervals from their combined prior information. In our first analysis strategy, we used all five prior distributions given in Table 1. Therefore, the results that arise from this analysis represent our experts' best estimates based on their prior knowledge and the data in the experiment.

As a second strategy, the prior distribution for the prevalence of asthma was used, but nothing was assumed a priori about the sensitivities and specificities of the two tests. Formally, the prior distribution is equal to a constant over the entire range of the parameter (that is, from 0% to 100%), so that a priori, all values are equally likely. This strategy allows the data alone to provide estimates of the test properties, while the prevalence retains the same prior distribution as in strategy (1).

Since the positive and negative predictive values are simply functions of the sensitivity, specificity, and disease prevalence, they are also immediately available once the other parameters are known.

The degree of agreement between the two diagnostic tests (exercise challenge and current asthma by questionnaire) was estimated through a constant predictive probability model (α-parameter) as advocated by Aickin (29). For

TABLE 1. Prior distribution of sensitivities and specificities for exercise and current asthma by questionnaire as well as the population prevalence of asthma in children

	Sensitivity range (%)[a]	Specificity range (%)[a]
Exercise challenge	30–60	80–98
Current asthma by questionnaire	55–80	65–85
Prevalence of asthma in children (%)	3–12	

[a] Range of sensitivities and specificities reported by panel of experts.

comparison, the K-coefficient (30) was also calculated. Although we calculated both *kappa* statistics and the α-parameter to compare the agreement between the two diagnostic tests, we feel the α-parameter to be more valid. This is because in calculating *kappa*, the expected proportion is generally defined in terms of certain marginal probabilities that occur in a model, in which both chance and causal agreement are present. Consequently, the expected proportion tends to include not only the random agreement that is intended to be captured, but in addition, some of the agreement for cause. As pointed out by Aickin (29), Cohen's *kappa*, therefore, penalizes raters who tend to agree because it uses their observed marginal probabilities to correct for chance agreement. This correction term will be larger as the two marginal distributions tend to agree. On the other hand, the constant predictive probability model (α-parameter) obviates this difficulty, since the correlation is made with marginal probabilities of the items that are difficult to classify, and less influenced by excess frequencies in the diagonal cells that represent agreement for cause.

RESULTS

Characteristics of the Study Population

Of the 1274 children selected from 18 schools, the parents of 130 (10.2%) refused participation of their children, while a further 75 (5.9%) children did not return the questionnaire. There were no meaningful differences between participants and non-participants as to age of the child (mean [SD]: 8.8 [1.8] versus 8.0 [1.9]), gender (boys: 50.5% versus 55.4%), race (Caucasians: 78% versus 80.9%), neighborhood SES assessed by the census data (poorest SES quartile: 26.6% versus 22%). Among families who refused permission for their child to participate in testing at school but did return the questionnaire ($n = 99$), mothers were less likely to be currently smoking (18.8% versus 37.9%) but no meaningful differences were seen in terms of respiratory symptoms, type of heating, cooking fuel used, and pets. Spirometry was not performed for a further 23 children because of the child's sickness or absence from school. Of the 1046 children who attempted spirometry, 28 (2.7%) were unable to complete the test successfully and such failure was more common among younger children but was unrelated to reported respiratory illness, symptoms, or SES. The spirometry data of a further 28 (2.7%) children were lost after the test. One child (0.1%) was excluded because of a severe asthmatic attack at the time of the test. After the above exclusions, 989 (77.6%) children remained for analysis. Of these children, 778 (78.7%) were Caucasians, 71 (7.2%) black, 39 (3.9%) Asian, 41 (4.1%) East Indians, and 60 (6.0%) were from other ethnic groups, such as North American Indians.

Table 2 provides descriptive characteristics for responders to the initial questionnaire ($n = 1{,}111$). The prevalence of ever and current wheezing was quite low in comparison to published results in similar populations whereas the prevalence of asthma was comparable (31). This is likely due to the unsatisfactory translation of the term wheezing into French and has been described previously in studies carried out in the province of Quebec (31).

Thirty-nine children had repeat exercise testing on two different days. Agreement as to the presence or absence of exercise induced bronchospasm was 72%.

Comparison of the mean age and atmospheric conditions in those children with and without EIB did not show important differences (Table 3).

Agreement Between the Two Diagnostic Tests (Exercise Challenge and Current Asthma by Questionnaire)

A total of 952 children (85.7%) studied at school had results for both exercise challenge and asthma by questionnaire. Among the children with results for the two diagnostic methods, 215 (22.6%) were classified as positive by at least one of the two tests (Table 4), EIB was present in 184 (19.3%) children, of which only 22 (12.0%) were classified as having current asthma by questionnaire. Of the 53 children who were classified as having current asthma by questionnaire, only 22 (41.5%) had EIB.

When informative priors for all parameters were used, there was a 61.3% probability of having asthma when the child was identified as positive by the two methods (Table 4), followed by 25% when the child was classified as positive by questionnaire, but negative by exercise. The probability

TABLE 2. Descriptive characteristics of responders to initial questionnaire in the cross-sectional survey at schools

		Number	Percentage
Boys $n = 1{,}111$		553	49.9
Symptoms/diagnosis	$n = 1101$		
Usual cough		113	10.3
Night cough		112	10.2
Ever wheeze		117	10.6
History of asthma		133	12.1
Current asthma		51	4.6
Asthma treatment		50	4.5
Parental smoking, current	$n = 1111$		
Neither[a]		538	48.4
Mother only[a]		203	18.3
Father only[a]		180	16.2
Both		190	17.1
Child shares bedroom	$n = 1109$	487	43.9
Child attended daycare	$n = 1102$	373	33.9
Pets in home, current	$n = 1109$		
Any		503	45.4
Cats		209	18.9
Exercise induced bronchospasm (EIB)[b]	$n = 989$	189	19.1

[a] 318 (28.6%) are monoparental.
[b] Fall in $FEV_1 \geq 10\%$.

TABLE 3. Mean age, indicators of effort and atmospheric conditions in exercise challenge tests among children with and without EIB

	Children with EIB (n = 189) Mean (SD)	Children without EIB (n = 800) Mean (SD)
Age in years	8.5 (1.8)	8.8 (1.8)
Pre-exercise heart rate (beats/min)	96 (15)	96 (15)
Post-exercise heart rate (beats/min)	195.6 (11.9)	194.6 (13.5)
Distance run (meters/6 min)	1025 (184)	1052 (187)
Temperature in the room at testing (°C)	19.5 (2.5)	19.2 (2.3)
Humidity in the room at testing	31 (12.3)	30 (11.1)
Percent maximum heart rate[a]	87.8 (5.9)	87.4 (6.3)

[a] Calculation was based on age and gender-specific maximal heart rate by Bruce Treadmill Protocol (reference 32).

of having asthma (predictive value) when the child was classified as positive by exercise but negative by questionnaire was 4.8% (Table 4). The trend was similar when informative prior probabilities for prevalence and uniform distribution for all other parameters were used.

Crude agreement between the two diagnostic tests appeared acceptable (79.7%), largely since most children tested negatively on both tests. The K-coefficient was only 10.9%, however, and the α-parameter was 35.4% (95% CI: 22.1–48.7). Therefore, correcting for chance agreement by either the standard or Aickin's methods showed that agreement for cause is low, almost certainly below 50%.

Prevalence of Asthma

In the sample of children studied, the estimated prevalence was 4.2% (95% CI: 2.4–6.6), using prior information for all parameters (see Table 1), whereas the estimated prevalence of asthma without a prior estimate for sensitivity and specificity (i.e., using non informative priors), but the same prior estimate for asthma prevalence, was 6.8% (95% CI: 3.6–11.9). The latter interval is wider than the former, since more prior information was included in the first estimate.

Comparison of the Test Characteristics of Exercise Challenge and Current Asthma by Questionnaire

Using informative prior distributions for all parameters (Table 1), current asthma by questionnaire was found to have significantly higher specificity and positive predictive value in comparison to exercise challenge (Table 5). While there was no statistically significant difference between the two tests with respect to sensitivity and negative predictive values (see overlapping CI in Table 5), the estimates were higher for current asthma in comparison to exercise challenge. Similarly, with informative prior for prevalence (Table 1) and uniform priors for sensitivity and specificity, the specificity was significantly higher for current asthma by questionnaire than exercise challenge, whereas there was no statistically significant difference between the two tests with respect to sensitivity and positive and negative predictive values (see the overlapping CI in Table 5).

DISCUSSION

We found the estimated prevalence of current asthma to be 4.2% (95% CI: 2.4–6.6) using the combination of an exercise challenge test and a questionnaire in a general population sample of primary school children. Exercise challenge was the less desirable of the two tests in terms of sensitivity, specificity, and positive predictive values, when the purpose was selecting a single test for the identification of subjects with asthma. Use of the two tests did not significantly improve the likelihood of correctly identifying children with asthma. Overall agreement between the two methods was poor.

The test characteristics (sensitivity, specificity, predictive values) of a diagnostic test are usually calculated by compari-

TABLE 4. Distribution of children by their test results

No. of subjects (total: 952)	EIB	Asthma by Questionnaire	% of disease (asthma) Using approach 1[a]	% of disease (asthma) Using approach 2[b]
22	yes	yes	61.3	82.3
162	yes	no	4.8	14.9
31	no	yes	25.0	39.5
737	no	no	1.0	1.7

[a] Informative priors for all parameters (Table 1).
[b] Informative priors for prevalence (Table 1) and uniform priors for sensitivity and specificity.

TABLE 5. Test characteristics of exercise challenge and current asthma by questionnaire with 95% credible intervals in the whole sample

	Exercise	Current asthma by questionnaire
In the sample of children studied at schools, n = 952)		
With informative priors[a]		
Sensitivity, %	51.3 (37.8–64.5)	64.4 (50.9–76.6)
Specificity, %	82.6 (79.9–85.1)	94.9 (93.2–96.5)
Positive predictive value, %	19.2 (12.3–27.8)	53.8 (41.0–66.7)
Negative predictive value, %	95.4 (93.2–97.1)	96.7 (94.6–98.1)
With informative priors for prevalence and uniform priors for sensitivity and specificity[a]		
Sensitivty, %	60.9 (35.0–97.2)	46.8 (17.5–92.9)
Specificity, %	83.7 (80.3–89.1)	97.3 (94.9–99.7)
Positive predictive value, %	21.1 (9.5–49.2)	54.6 (23.9–95.7)
Negative predictive value, %	96.9 (92.3–99.8)	96.2 (89.9–99.7)

[a] Given in Table 1.

son to an established gold standard. In the case of asthma, neither questionnaire nor exercise challenge have proven to be adequate gold standards in epidemiological studies. We have used a Bayesian or non-frequentist approach to compare diagnostic methods in the absence of a gold standard. The basic idea behind the Bayesian approach is to eliminate constraints by first constructing a prior distribution over all unknown quantities (test characteristics and disease prevalence) based on previous experience. The data, through the likelihood function, are then combined with the prior distribution to derive posterior distributions using Bayes Theorem. This allows simultaneous inferences to be made on all parameters. The posterior distributions contain updated beliefs about the values of the model parameter, after taking into account the information obtained from experts or reported in the literature.

A further advantage of the Bayesian approach is that normal distribution approximations are not required. Direct calculations of the posterior distributions can be difficult. The Gibbs Sampler (28, 33, 34), an iterative Markov-chain Monte Carlo technique for approximating analytically intractable posterior densities, was used. This method has been used to estimate parameters in a wide variety of problems in health research (35–37). The basic idea is as follows: Knowing the exact values of the prevalence and all diagnostic test parameters, it is possible to derive posterior distributions of the latent data (the information that is missing when there is no gold standard, that is, the number of true positive test results). Conversely, if the latent data (the number of true positive test results) are known, then deriving posterior distributions of the prevalence and diagnostic test parameters, given the prior distributions, requires only a straightforward application of Bayes Theorem.

A study (38) that compared parental questionnaire responses to in-depth history-taking by experienced physicians found this type of questionnaire to have a sensitivity of 68% and a specificity of 76%. Misclassification of asthmatic children as normal mainly occurred in those with mild symptoms. In a number of community-based studies of children the test characteristics of exercise challenge have been evaluated. Where exercise challenge has been compared to the report of a physician's diagnosis of asthma, the sensitivity has ranged from 46% to 57%, while specificity ranged from 94% to 97% (18, 23, 39). Among a sample of 96 children 8 to 11 year of age, Haby and colleagues (40) found exercise challenge similar to ours to have a sensitivity of 27% and a specificity of 95%. More recently, in a study of 55 children who reported wheeze in the previous 12 months and 54 control subjects, West and colleagues (18) found a sensitivity of 57% and specificity of 94%.

In our study, the sensitivity of exercise challenge test was 51.3% (95% CI: 37.8–64.5), which is not very different from that reported by others. On the other hand, the specificity of 82.6% (95% CI: 79.9–85.1) was much lower than previously reported. This lower specificity, however, is in keeping with the more recent realization that bronchial hyperresponsiveness, in the absence of respiratory symptoms, is unlikely to have clinical relevance (41).

In our study, agreement between the diagnostic methods was poor and exercise challenge, as a single measure of asthma, had a very low predictive value, and when used in addition to questionnaire information, the gain in predicting asthma was also small. This supports the view of Ninan and colleagues that exercise testing adds very little to a well designed respiratory questionnaire (9). On the other hand, our results are in disagreement with the view of Toelle and his group (8) that a combination of both questionnaire and airway responsiveness tests could be used to define asthma epidemiologically. Similarly, the results of our study are applicable to epidemiologic, but not clinical, situations.

We used a 10% or greater fall in FEV_1 to define a positive exercise challenge based upon the receiver operating characteristics (ROC) analysis performed in a sub-sample of children, which found this value to have maximal specificity and sensitivity. This cutoff has been used by others (24–26). While defining a positive response to exercise as 15% or greater fall in FEV_1 may have increased specificity, this would have left insufficient subjects with a positive response to compare with questionnaire responses.

In conclusion, there was poor agreement between exercise challenge and the questionnaire report of asthma. Exercise challenge appeared less useful as compared to a questionnaire in terms of test characteristics, when the purpose was to choose a single test for identifying asthma in community-based studies. The combined use of the two tests did not significantly improve the probability of identifying children with asthma, supporting the view of Ninan and colleagues that exercise adds very little to a well designed questionnaire.

Supported by the Respiratory Health Network of Centers of Excellence and the Medical Research Council of Canada.

REFERENCES

1. Cockroft DE, Hargreave FE. Airway hyperresponsiveness. Relevance of random population data to clinical usefulness. Am Rev Respir Dis. 1990;142:497–500.
2. Woolcock AJ. Test of airway responsiveness in epidemiology. In: Hargreave FE, Woolcock AJ, eds. Airway hyperresponsiveness: Measurement and interpretation. Ontario: Astra Pharmaceuticals Canada Ltd.; 1985.
3. Cockroft DW. Bronchial inhalation tests II. Measurement of allergic (and occupational) bronchial responsiveness (Review). Ann Allergy. 1987;59:89–98.
4. Eggleston PA, Guerrant JL. A standardized method of evaluating exercise-induced asthma. J Allergy Clin Immunol. 1976;58:414–425.
5. Cropp GJA, Schultzer IJ. Grading, time course, and incidence of exercise induced airway obstruction and hyperinflation in asthmatic children. Pediatrics. 1975;56(Suppl):868–879.
6. Bransford RP, McNutt GM, Fink JN. Exercise-induced asthma in adolescent gym class population. Int Arch Allergy Appl Immunol. 1991;94:272–274.
7. Woolcock AJ. Worldwide differences in asthma prevalence and mortality. Chest 1986:90(Suppl):40S–45S.
8. Toelle BG, Peat JK, Salome CM, Mellis CM, Woolcock AJ. Toward a definition of asthma for epidemiology. Am Rev Respir Dis. 1992; 146:633–637.
9. Ninan TK, Russell G. Is exercise testing useful in a community based asthma survey? Thorax. 1993;48:1218–1221.
10. Clough JB, Hutchinson SA, Williams JD, Holgate ST. Airway responsive to exercise and methacholine in children with respiratory symptoms. Arch Dis Child. 1991;16:579–583.
11. Hangaard L, Iversen M, Dahl R. A comparative study of four different bronchial challenge tests. Allergy. 1992;47:138–142.
12. Godfrey S, Springer C, Noviski N, Maayan CH, Avital A. Exercise but not methacholine differentiates asthma from chronic lung disease in children. Thorax. 1991;46:488–492.
13. Bhagat RG, Grunstein MM. Comparison of responsiveness to methacholine, histamine, and exercise in subgroups of asthmatic children. Am Rev Respir Dis. 1984;129:221–224.
14. Kiviloog J. The correlation between exercise-induced bronchoconstriction and bronchial methacholine sensitivity in asthma. Pediatrics. 1975;56(Supp):908–909.
15. Eggleston PA. A comparison of the asthmatic response to methacholine and exercise. J Allergy Clin Immunol. 1979;63:104–110.
16. Foresi A, Mattoli S, Corbo GM, Polidori G, Ciappi G. Comparison of bronchial responses to ultrasonically nebulised distilled water, exercise and methacholine in asthma. Chest. 1986;90:822–826.
17. Chatham M, Bleecker ER, Smith PL, Rosenthal RR, Mason P, Norman PS. A comparison of histamine, methacholine, and exercise airway reactivity in normal and asthmatic subjects. Am Rev Respir Dis. 1982; 126:235–240.
18. West JV, Robertson CF, Roberts R, Olinsky A. Evaluation of bronchial responsiveness to exercise in children as an objective measure of asthma in epidemiological surveys. Thorax. 1996;51:590–595.
19. Joseph L, Gyorkos TW, Coupal L. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. Am J Epidemiol. 1995;141:263–272.
20. Joseph L, Gyorkos T. Inferences for likelihood ratios in the absence of a gold standard. Med Decis Making. 1996;16:412–417.
21. American Thoracic Society. Standardization of spirometry—1987 update. Am Rev Respir Dis. 1987;136:1285–1298.
22. Kanner RE, Schenker MB, Munoz A, Speizer FE. Spirometry in children. Methodology for obtaining optimal results for clinical and epidemiologic studies. Am Rev Respir Dis. 1983;127:720–724.
23. Tsanakas JN, Milner RDG, Bannister OM, Boon WA. Free running asthma screening test. Arch Dis Child. 1988;63:261–265.
24. Eggleston PA, Rosenthal RR, Anderson SA, Anderton R, Bierman CW, Bleecker ER, et al. Guidelines for the methodology of exercise challenge testing of asthmatics. J Allergy Clin Immunol. 1979;64: 642–645.
25. Anderson SD, Seale JP, Ferris L, Schoeffel RE, Lindsay DA. An evaluation of pharmacotherapy for exercise-induced asthma. J Allergy Clin Immunol. 1979;64:612–624.
26. Mellis CM. Kaffan M, Keens TG, Levison H. Comparative study of histamine and exercise challenge in asthmatic children. Am Rev Respir Dis. 1978;117:911–915.
27. Walter SD, Irwig LM. Estimation of test error rates, disease prevalence and relative risk from misclassified data: a review. J Clin Epidemiol. 1988;41:923–937.
28. Gelfand AE, Smith AFM. Sampling-based approaches to calculating marginal densities. J Am Statist Assoc. 1990;85:398–409.
29. Aickin M. Maximum likelihood estimation of agreement in the constant predictive probability model, and its relation to Cohen's kappa. Biometrics. 1990;46:293–302.
30. Streiner DL, Norman GR. Health Measurement Scales: A Practical Guide to Their Development and Use. Oxford: Oxford University Press; 1989;94–95.
31. Osterman JW, Armstrong BG, Ledoux E, Sloan M, Ernst P. Comparison of French and English Versions of the American Thoracic Society Respiratory Questionnaire in a Bilingual Working Population. Int J Epidemiol. 1991;20:138–143.
32. Cumming GR, Everatt D, Hastman L. Bruce Treadmill Test in children: Normal values in a clinic population. Am J Cardiol. 1978; 41:69–75.
33. Gelfand AE, Hills SE, Racine-Poon A, Smith AFM. Illustration of Bayesian inference in normal data using Gibbs sampling. J Am Statist Assoc. 1990;85:972–985.
34. Tanner MA. Tools for Statistical Inference. New York: Springer-Verlag; 1991.
35. Coursaget P, Yvonnet B, Gilks WR, Wang CC, Day NE, Chiron JP. Scheduling of revaccinations against Hepatitis B virus. Lancet. 1991;337:1180–1183.

36. Gilks WR, Clayton DG, Spiegelhalter DJ. Modeling complexity: Application of Gibbs sampling in medicine. J Roy Statist Soc B. 1993; 1:39–52.
37. Richardson S, Gilks WR. A Bayesian approach to measurement error problems in epidemiology using conditional independence models. Am J Epidemiol. 1993;138:430–442.
38. Braback L, Kalvesten L, Sandstrom G. Prevalence of bronchial asthma among school children in a Swedish district. Acta Paediatr Scand. 1988;77:821–825.
39. Terblanche E, Stewart RI. The prevalence of exercise-induced bronchoconstriction in Cape Town school children. S Afr Med J. 1990; 78:744–747.
40. Haby MM, Anderson SD, Peat JK, Mellis CM, Toelle BG, Woolcock AJ. An exercise challenge protocol for epidemiological studies of asthma in children: Comparison with histamine challenge. Eur Respir J. 1994;7:43–49.
41. Kolnaar BGM, Folgering H, van den Hoogen HJM, van Weel C. Asymptomatic bronchial hyperresponsiveness in adolescents and young adults. Eur Respir J. 1997;10:44–50.