



Comparison of spirometry criteria for the diagnosis of COPD: results from the BOLD study

W.M. Vollmer*, Þ. Gíslason[#], P. Burney¹, P.L. Enright⁺, A. Gulsvik[§],
A. Kocabas^f and A.S. Buist^{**} for the BOLD Collaborative Research Group

ABSTRACT: Published guidelines recommend spirometry to accurately diagnose chronic obstructive pulmonary disease (COPD). However, even spirometry-based COPD prevalence estimates can vary widely. We compared properties of several spirometry-based COPD definitions using data from the international Burden of Obstructive Lung Disease (BOLD) study.

14 sites recruited population-based samples of adults aged ≥ 40 yrs. Procedures included standardised questionnaires and post-bronchodilator spirometry. 10,001 individuals provided usable data.

Use of the lower limit of normal (LLN) forced expiratory volume in 1 s (FEV₁) to forced vital capacity (FVC) ratio reduced the age-related increases in COPD prevalence that are seen among healthy never-smokers when using the fixed ratio criterion (FEV₁/FVC <0.7) recommended by the Global Initiative for Chronic Obstructive Lung Disease. The added requirement of an FEV₁ either <80% predicted or below the LLN further reduced age-related increases and also led to the least site-to-site variability in prevalence estimates after adjusting for potential confounders. Use of the FEV₁/FEV₆ ratio in place of the FEV₁/FVC yielded similar prevalence estimates.

Use of the FEV₁/FVC < LLN criterion instead of the FEV₁/FVC < 0.7 should minimise known age biases and better reflect clinically significant irreversible airflow limitation. Our study also supports the use of the FEV₁/FEV₆ as a practical substitute for the FEV₁/FVC.

KEYWORDS: Adult, chronic obstructive pulmonary disease, epidemiology

Although chronic obstructive pulmonary disease (COPD) is recognised as a major public health problem worldwide, estimates of its prevalence vary widely [1]. Much of this variation probably reflects differences in the populations studied, spirometry methods and data quality control, and the rules used to define COPD. For example, self-reported physician diagnosis of COPD typically results in estimated prevalences well below those obtained based on spirometry [1, 2].

Although no gold standard definition of COPD exists, published guidelines recommend use of spirometry to define it [3, 4]. However, even spirometry-based COPD prevalence estimates can vary by two-fold or more, depending on the definition used to classify mild disease [5, 6]. The most widely used definition comes from the Global Initiative for Chronic Obstructive Lung

Disease (GOLD), which recommends using a post-bronchodilator forced expiratory volume in 1 s (FEV₁) to forced vital capacity (FVC) ratio <0.7 to define irreversible airflow limitation, and the FEV₁ to stage disease [3]. This “fixed ratio” approach, while easy to apply, appears to overestimate COPD in older individuals [2, 7–10] and to underestimate it in young adults [9, 11]. Alternative definitions that account for normal ageing can alleviate this bias [9, 12] but, in turn, this raises questions about which reference equations are appropriate for which populations. In addition, if pre- (rather than post-) bronchodilator spirometry is used, COPD prevalence may be overestimated by as much as 30% [8, 10, 13, 14].

The Burden of Obstructive Lung Disease (BOLD) study is an international effort to collect population-based estimates of the prevalence and economic burden of COPD using standardised methods

AFFILIATIONS

*Kaiser Permanente, Center for Health Research, Portland, OR
+The University of Arizona, Tucson, AZ, and
**Oregon Health & Science University, Portland, OR, USA.
#University of Iceland, Medical Faculty, Landspítali University Hospital, Reykjavik, Iceland.
¹National Heart and Lung Institute, Imperial College, London, UK.
[§]Institute of Medicine, University of Bergen, Bergen, Norway.
^fCukurova University School of Medicine, Balçalı, Adana, Turkey.

CORRESPONDENCE

W.M. Vollmer
Center for Health Research
3800 N Interstate Ave
Portland
OR
97227-1110
USA
E-mail: william.vollmer@kpchr.org

Received:
Oct 30 2008
Accepted after revision:
April 27 2009
First published online:
May 21 2009

[15, 16]. Using BOLD study data, we examined the impact on prevalence estimates of using the fixed ratio criterion *versus* various other spirometry-based definitions of COPD. We also compare the effects of using central *versus* site-specific prediction equations and of using the FEV₁/FEV₆ ratio in place of FEV₁/FVC.

METHODS AND MATERIALS

The design of the BOLD study is described in detail elsewhere [15, 16] and only summarised here. Participating entities in the BOLD Collaborative Research Group are listed in the online supplement.

Population

Participating sites were expected to recruit population-based samples of ≥ 600 noninstitutionalised adults aged ≥ 40 yrs. We report data from the first 14 BOLD sites (table 1), consisting of 10,001 individuals (93% of all responders) with acceptable post-bronchodilator spirometry. Each site obtained approval from local ethical committees and written informed consent from each participant.

Questionnaires

Questionnaires were administered by trained and certified staff and covered respiratory symptoms, smoking history, respiratory diagnoses and comorbidities. We defined pack-years of cigarette smoking exposure as average number of packs smoked per day (20 cigarettes per pack) multiplied by the number of years smoked. Never-smoking was defined as < 20 packs of cigarettes in a lifetime.

Site-specific prediction equations were developed using never-smokers who had never been told by a healthcare provider that they had emphysema, COPD or tuberculosis, and did not report a current diagnosis of asthma or chronic bronchitis. We were unable to restrict to asymptomatic never-smokers due to the extremely small numbers of (particularly male) never-smokers at some sites.

Height and weight

We measured height (to the nearest centimetre) with the participant standing on a firm, level surface that was perpendicular to the vertical board of the height measurement device (ideally a wall-mounted stadiometer). Participants were

TABLE 1 Summary of Burden of Obstructive Lung Disease (BOLD) study sites included in the analysis

Site	Target population	Sampling design	Respondents [#] n	Response rate [†] %	Cooperation rate [‡] %
Guangzhou, China	Residents of the Number Two Community Center of the Liwan district of the city of Guangzhou	Stratified random sample	602	87	87
Adana, Turkey	Residents of all districts (urban and rural) that make up the Adana province	Stratified cluster sample	875	82	85
Salzburg, Austria	Residents of Salzburg County	Stratified random sample	1349	65	67
Cape Town, South Africa	Residents of the towns of Ravensmead and Uitsig, both located in the general Tygerberg area	Cluster sample	896	63	68
Reykjavik, Iceland	All age-eligible Icelandic citizens living in Reykjavik and surrounding suburbs	Simple random sample	758	81	84
Hanover, Germany	Residents of Hanover city and four of the 20 town councils that make up the rest of Hanover region	Stratified random sample	713	59	61
Krakow, Poland	Residents of the Chrzanów and Proszowice regions within the Malopolska district of Poland	Stratified random sample	603	78	79
Bergen, Norway	Residents of the city of Bergen	Stratified random sample	707	68	71
Vancouver, BC, Canada	Residents of the Vancouver Health Service Delivery Area of Vancouver	Random digit dialling	856	26	51
Lexington, KY, USA	Residents of the fifth congressional district of Kentucky	Random digit dialling	563	14	27
Manila, Philippines	Residents of one of the six districts that make up the city of Manila	Cluster sample	918	58	58
Sydney, Australia	Residents of the federal electorates, Kingsford Smith and Barton, which make up the southern beachside suburbs of Sydney	Stratified random sample	585	25	33
London, UK	The patients served by three general practitioners from West London, which is broadly representative of the Hammersmith and Fulham neighbourhood of West London	Stratified random sample	697	17	37
Uppsala, Sweden	Residents of the city of Uppsala	Stratified random sample	588	61	63

[#]: participants with core questionnaire and any post-bronchodilator spirometry; [†]: denominator includes persons of unknown eligibility status who could not be contacted (only known ineligible participants were excluded); [‡]: denominator includes only participants who were contacted and eligible.

TABLE 2 Definitions of chronic obstructive pulmonary disease (COPD) used in this analysis

Threshold for COPD	Formula
Fixed ratio	FEV ₁ /FVC <0.7 (GOLD stage 1 or higher)
LLN (FEV₁/FVC)	FEV ₁ /FVC <5th percentile (predicted FEV ₁ /FVC minus 1.645 × sd)
GOLD stages 2–4	FEV ₁ /FVC <0.7 and FEV ₁ <80% pred
LLN (FEV₁/FVC) and FEV₁ <80% pred	FEV ₁ /FVC <5th percentile and FEV ₁ <80% pred
LLN (FEV₁/FVC) and LLN (FEV₁)	FEV ₁ /FVC <5th percentile and FEV ₁ <5th percentile
LLN (FEV₁/FEV₆) and FEV₁ <80% pred	FEV ₁ /FEV ₆ <5th percentile and FEV ₁ <80% pred

For all but the fixed ratio, alternative versions may be computed based on choice of prediction equation. This table just describes general formulas. FEV₁: forced expiratory volume in 1 s; FVC: forced vital capacity; GOLD: Global Initiative for Chronic Obstructive Lung Disease; LLN: lower limit of normal; % pred: % predicted; FEV₆: forced expiratory volume in 6 s.

instructed to remove their shoes and stand erect with feet flat on the floor, heels together, and head in the horizontal (Frankfort) plane.

Sites used calibrated scales (preferably balance beam or digital) to measure weight to the nearest 0.1 kg. Participants were instructed to remove shoes, hats, coats, and heavy items in their pockets in order to be weighed in light indoor clothing.

Body mass index (BMI) was computed as weight over height-squared and expressed in units of kg·m⁻².

Spirometry

Lung function data were collected using the ndd EasyOne Spirometer (ndd Medical Technologies, Zurich, Switzerland), which was chosen for its portability and level of accuracy [17]. Lung function was measured before and 15 min after administration of 200 µg of albuterol/salbutamol. Spirometry measures reported here include the FEV₁, FEV₆ and FVC, as well as the FEV₁/FVC and FEV₁/FEV₆ ratios. FEV₁ % predicted, although not reported separately, was used to stage COPD [3].

All spirometry were reviewed by the BOLD Pulmonary Function Reading Center and assigned an overall quality score based on standardised criteria [18]. Local spirometry technicians were trained and certified, and received regular quality control feedback during data collection. Usable spirometry was defined as two or more acceptable blows, with FEV₁ and FVC repeatability within 200 mL. Acceptable manoeuvres were defined as those with a rapid start (back-extrapolated volume <150 mL or <5% of the FVC), lack of a cough during the first second, and a small end-of-test volume (<40 mL during the final second). The calibration of all spirometers was verified to be accurate within 3.0% using a 3.00 L syringe at the beginning of each day of testing. Biological controls were not used.

Definition of COPD

The BOLD study uses the GOLD criteria for defining and staging COPD [3], which are consistent with the 2004 American Thoracic Society (ATS)/European Respiratory Society (ERS) criteria [4] and define COPD as a post-bronchodilator FEV₁/FVC <0.70. The FEV₁ % pred is used to further stage disease (FEV₁ ≥80% pred: stage 1; ≥50 and <80% pred: stage 2; ≥30 and <50% pred: stage 3; <30% pred: stage 4). The BOLD study also uses the prediction equations for Caucasian adult males and females derived from the Third US National Health and

Nutrition Examination Survey (NHANES-III) [19] as its primary reference equations for all participants, although this paper also examined the impact of using equations derived from Norway's Hordaland County Respiratory Health Study [20], as well as site-specific prediction equations, in place of the NHANES-III equations.

In addition, we assessed the impact of restricting COPD to GOLD stages 2 or above, and of using the lower limit of normal (LLN) of the FEV₁/FVC, and the FEV₁ in place of the fixed ratio and the FEV₁ <80% pred criteria, in the GOLD definitions. Finally, we examined the impact of using FEV₁/FEV₆ in place of FEV₁/FVC in our definitions. Table 2 summarises the various definitions of COPD assessed in this manuscript.

Although the text focuses on post-bronchodilator spirometry, the results of comparable analyses based on pre-bronchodilator data are included in the online supplementary material.

Analysis

To provide comparability with earlier reports [16], the site-specific prevalences presented in figure 1 are population-based estimates reflecting sampling designs used at each site. For all other analyses, data are pooled across sites and presented as unweighted prevalences with standard errors accounting only for correlations within the site and, where applicable, for clustering in the sampling plan. Comparisons of the prevalence estimates in figures 1–3 and in table 3 were computed using McNemar's test.

A desired characteristic of any prevalence estimator is that it gives comparable estimates in different populations after adjusting for known confounders. In order to compare the residual site-to-site variability associated with our various prevalence estimators, we report the Wald statistic for the "site" effect, as derived from logistic regression models that adjusted for age (40–49, 50–59, 60–69 and ≥70 yrs), sex, cigarette smoking history (never-smokers, 0–9, 10–19 and ≥20 pack-yrs), BMI (<20, 20–25, 25–30, 30–35 and >35 kg·m⁻²), years worked in a dusty job (0, 1–9 and ≥10 yrs) and interactions of sex with both age and smoking history. We also report Wald statistics for testing the significance of age in selected regression models. Where appropriate, we tested heterogeneity of age effects across strata using appropriate interaction terms. Under the null hypothesis of no effect, the Wald statistic will have an F-

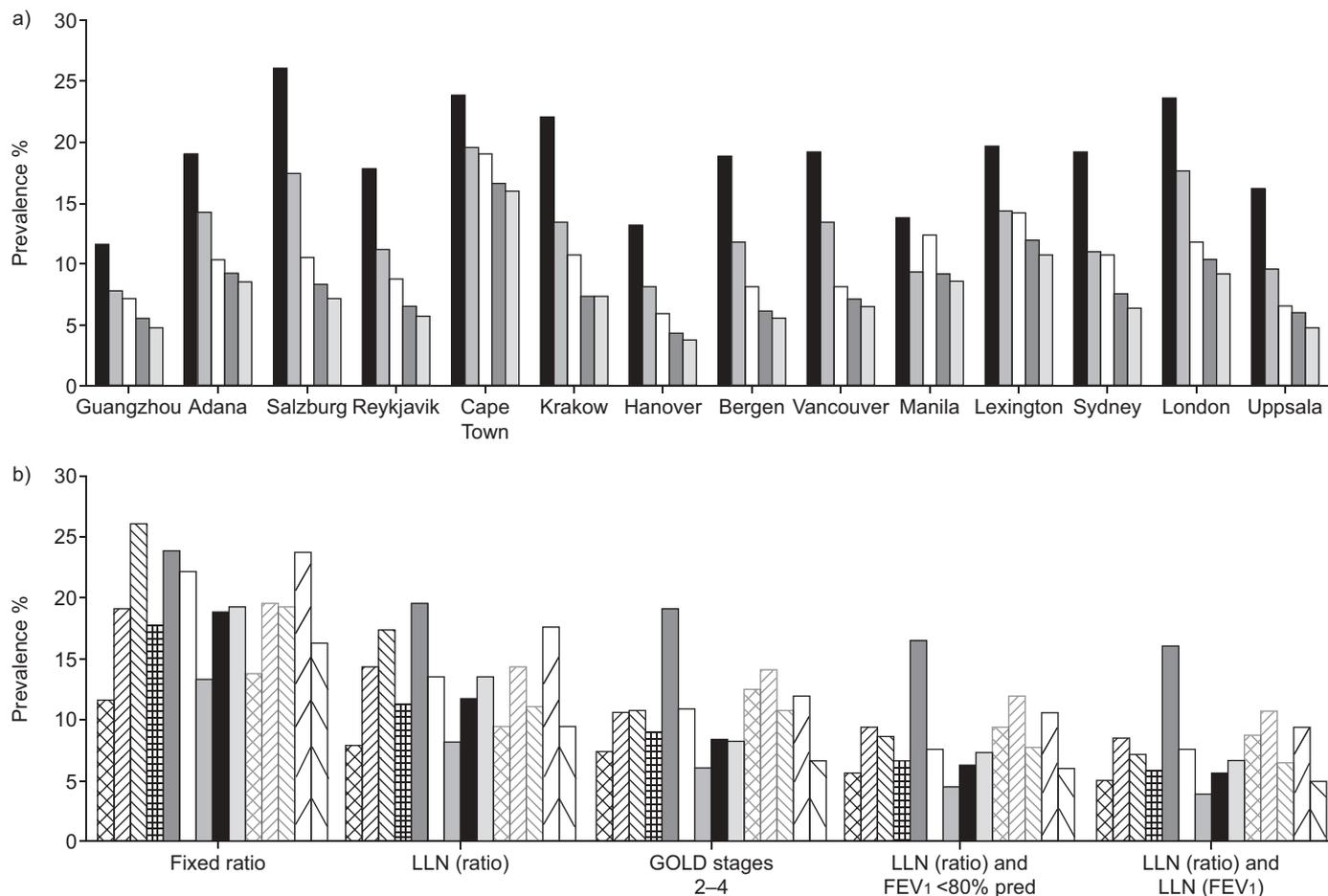


FIGURE 1. Comparison of chronic obstructive pulmonary disease (COPD) prevalence for various alternative definitions of COPD for participants in the Burden of Obstructive Lung Disease (BOLD) study. a) Contrasting prevalences for each definition within site. ■: fixed ratio; ■: lower limit of normal (LLN) (ratio); □: Global Initiative for Chronic Obstructive Lung Disease (GOLD) stages 2–4; ■: LLN (ratio) and forced expiratory volume in 1 s (FEV₁) <80% predicted; ■: LLN (ratio) and LLN (FEV₁). b) Contrasting site differences within each definition. Significant site-to-site variability persisted for each of the estimates even after adjusting for age, sex, cigarette smoking (pack-yrs), body mass index, years worked in a dusty job and interactions of sex with both age and smoking history in logistic regression models (Wald statistics ranging from 7.1 to 8.6, all p-values <0.0001). Prevalences based on the fixed ratio are significantly higher than for all other estimators at each site (all p-values <0.001). ■: Guangzhou, China; ☒: Adana, Turkey; ☒: Salzburg, Austria; ☒: Reykjavik, Iceland; ■: Cape Town, South Africa; □: Krakow, Poland; ■: Hanover, Germany; ■: Bergen, Norway; ■: Vancouver, BC, Canada; ■: Manila, Philippines; ☒: Lexington, KY, USA; ☒: Sydney, Australia; ☒: Uppsala, Sweden.

distribution with an expected value equal to one, and higher values indicate greater heterogeneity across subgroups. All Wald tests are adjusted for clustering in the sampling plan.

All analyses were conducted using Stata version 9.2 (Stata Corp., College Station, TX, USA).

RESULTS

Participants exhibited marked differences in smoking patterns across sites and between sexes within sites (table 4). BOLD sites also differ markedly in prevalences of occupational and other potential COPD risk factors [16].

Use of the fixed ratio criterion (GOLD stage 1 and higher) produced overall population prevalence estimates that, for each site, were significantly greater than those for each of the other estimators (all but one $p < 0.0001$). The fixed ratio estimates were generally 5–11 percentage points higher than those for GOLD stages 2–4 (fig. 1a). The LLN (FEV₁/FVC) criterion produced estimates that tended to be intermediate to

these two GOLD-based definitions, although generally closer to the GOLD stages 2–4 criterion than to the fixed ratio criterion. The added requirement of an FEV₁ <80% pred and an FEV₁/FVC ratio below the LLN resulted in estimates that were 1–3 percentage points lower than estimates for GOLD stages 2–4. Finally, use of FEV₁ <LLN in place of FEV₁ <80% pred in this latter definition further reduced estimates (although generally by less than one percentage point). These patterns were generally consistent across sites.

Regardless of the definition used, we observed sizable site-to-site variation in prevalence estimates (fig. 1b). After adjusting for potential confounders, site-to-site variance in COPD prevalence (as measured by the Wald statistic) ranged from 7.1 to 8.6 and was lowest (7.1 and 7.3, respectively) using the “LLN (FEV₁/FVC) and LLN (FEV₁)” and “LLN (FEV₁/FVC) and FEV₁ <80% pred” criteria, respectively. These Wald statistics all indicated highly statistically significant ($p < 0.0001$) residual site-to-site variability in prevalence estimates.

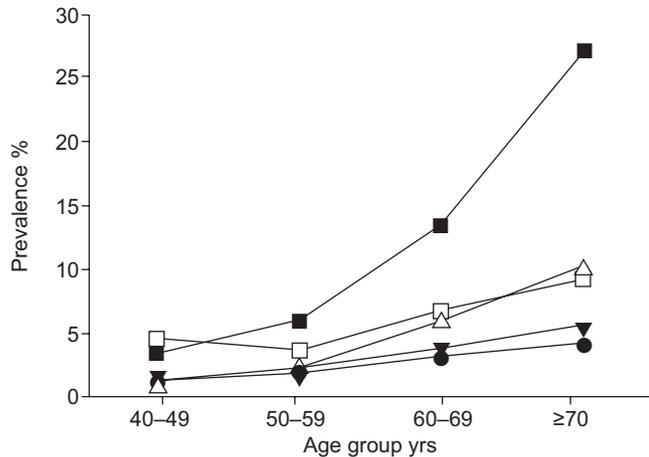


FIGURE 2. Prevalence of “chronic obstructive pulmonary disease” among healthy never-smokers in the Burden of Obstructive Lung Disease (BOLD) study (data from all sites combined). Wald statistics for comparing the four age groups (and adjusted for site, sex, body mass index and years worked in a dusty job) ranged from 62.6 for the fixed ratio criterion, to 24.5 for Global Initiative for Chronic Obstructive Lung Disease (GOLD) stages 2–4, to ~6.6 for the two analogues of these criteria in which the forced expiratory volume in 1 s (FEV₁) to forced vital capacity (FVC) ratio <0.7 is replaced by FEV₁/FVC<LLN (lower limit of normal), to 3.4 for the “LLN (FEV₁/FVC) and LLN (FEV₁)” criteria. All were statistically significant. Significant sex–age interactions were observed only for the fixed ratio criterion. Except for the LLN (ratio) criterion for the lowest age group ($p=0.14$), the fixed ratio prevalence estimates were all significantly greater than those for each of the other estimators ($p<0.0001$). ■: fixed ratio; □: LLN (ratio); △: GOLD stages 2–4; ▼: LLN (ratio) and FEV₁ <80% predicted; ●: LLN (ratio) and LLN (FEV₁).

All prevalences reported in figure 1 were lower than they would have been had we based them on pre-bronchodilator measurements (see online supplementary material). For the fixed ratio criterion, absolute declines between pre- and post-bronchodilator values ranged from 1 to 11 percentage points across centres, while using GOLD stages 2–4 instead of the fixed ratio criterion led to a decline in prevalence ranging from 1 to 6 percentage points across centres. On a relative basis, prevalence estimates declined between 25% to 29% (depending on the definition used) across the five measures in going from pre- to post-bronchodilator measurements.

The prevalence of “COPD” per the fixed ratio criterion increased sharply with age even among healthy never-smokers (fig. 2), a population in which COPD is expected to be rare. By contrast, for the other measures we observed much more muted increases with age and, except for the LLN (ratio) criterion for the lowest age group ($p=0.14$), the fixed ratio prevalence estimates were all significantly greater than those for each of the other estimators ($p<0.0001$). These age-related increases in prevalence were lowest for the “LLN (FEV₁/FVC) and LLN (FEV₁)” and “LLN (FEV₁/FVC) and FEV₁ <80% pred” criteria, for which the age-specific prevalence estimates varied from 2% among 40 yr olds, to 4–5% among those aged ≥ 70 yrs. We observed the same general patterns within each site (data not shown).

The Wald statistic for testing for age effects in figure 2 dropped from a high of 62.6 for the fixed ratio criterion to 24.5 for GOLD stages 2–4, to ~6.6 for the two analogues of these criteria in

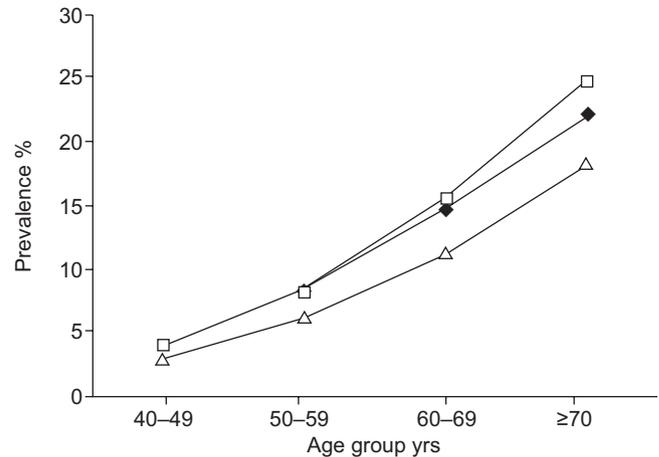


FIGURE 3. Prevalence of Global Initiative for Chronic Obstructive Lung Disease (GOLD) stages 2–4 chronic obstructive pulmonary disease in the Burden of Obstructive Lung Disease (BOLD) study using the Third US National Health and Nutrition Examination Survey (NHANES-III) and site-specific prediction equations (data from all sites combined). Site differences were greater when using the NHANES-III equations (Wald statistic 9.5, $p<0.0001$) or Hordaland county equations (Wald statistic 8.4, $p<0.0001$) than when using the site-specific equations (Wald statistic 4.7, $p<0.0001$). The estimated prevalences were significantly higher ($p<0.0001$ for all age groups) when using common reference equations for all sites (NHANES-III and Hordaland County) than when using local prediction equations. The NHANES-III and Hordaland County prevalence estimates differed significantly overall and for the oldest age group. ◆: NHANES-III; □: Hordaland County; △: local equations.

which FEV₁/FVC <0.7 is replaced by FEV₁/FVC<LLN, and to 3.4 for the “LLN (FEV₁/FVC) and LLN (FEV₁)” criteria. All were statistically significant. We found modest evidence of a statistically significant sex–age interaction using the fixed ratio criterion (Wald statistic 3.1, $p=0.027$) and no evidence of statistically significant sex–age interactions in these healthy never-smokers using any of the other prevalence estimators.

When we assessed site differences in the group of healthy, never-smoking individuals, we observed smaller site differences for the GOLD stages 2–4 criterion (Wald statistic 1.6) than for the LLN (FEV₁/FVC) criterion (Wald statistic 2.9), although once again the smallest site differences were seen for the “LLN (FEV₁/FVC) and LLN (FEV₁)” and “LLN (FEV₁/FVC) and FEV₁ <80% pred” criteria (Wald statistic 0.9–1.1). Indeed, for both of these latter criteria, the site differences did not come close to reaching statistical significance ($p>0.35$), whereas for the other three criteria the p -values were all less than 0.07.

Figure 3 illustrates the impact on prevalence of using a single common prediction equation (the US NHANES-III Caucasian equations) or the Hordaland County Respiratory Health Study equations) versus site-specific prediction equations. For both males and females, the estimated GOLD stage 2–4 prevalences were higher (by 2–3 percentage points overall; $p<0.0001$) when using common reference equations for all sites (NHANES-III and Hordaland County) than when using local prediction equations. The NHANES-III and Hordaland County prevalence estimates were generally similar, although they differed significantly overall and for the oldest age group. The Wald statistic for site differences computed from the site-specific

TABLE 3 Prevalence of chronic obstructive pulmonary disease (COPD) computed using FEV₁/FEV₆ in place of FEV₁/FVC

	Subjects n	COPD criterion		p-value [#]
		LLN (FEV ₁ /FEV ₆) and FEV ₁ <80% pred	LLN (FEV ₁ /FVC) and FEV ₁ <80% pred	
Overall	10001	8.1	8.3	0.10
Age yrs				
40–49	3381	4.1	4.1	0.84
50–59	2968	6.5	7.2	<0.01
60–69	2172	11.2	11.5	0.57
≥70	1480	15.9	15.7	0.73
Smoking history				
Never-smoker	4291	4.0	4.0	1.0
0–10 pack-yrs	1777	5.0	4.8	0.66
10–20 pack-yrs	1270	9.6	9.6	1.0
≥20 pack-yrs	2654	16.3	17.1	0.01
Sex				
Male	4766	9.1	9.1	0.90
Female	5235	7.2	7.6	0.01
Site				
Guangzhou, China	473	6.8	5.5	0.03
Adana, Turkey	806	8.3	9.3	0.04
Salzburg, Austria	1258	7.2	7.5	0.34
Cape Town, South Africa	847	17.1	16.3	0.25
Reykjavik, Iceland	757	5.9	6.5	0.29
Hanover, Germany	683	5.1	5.0	1.0
Krakow, Poland	526	7.6	7.4	1.0
Bergen, Norway	658	5.9	6.8	0.11
Vancouver, BC, Canada	827	5.7	6.4	0.07
Lexington, KY, USA	508	11.2	12.4	0.15
Manila, Philippines	893	9.2	8.4	0.09
Sydney, Australia	541	8.1	7.6	0.38
London, UK	677	9.2	10.2	0.02
Uppsala, Sweden	547	4.9	5.7	0.12

Data are presented as %, unless otherwise stated. Prevalence data are univariate classifications and not adjusted for other terms in the table. FEV₁: forced expiratory volume in 1 s; FEV₆: forced expiratory volume in 6 s; FVC: forced vital capacity; LLN: lower limit of normal; % pred: % predicted. #: two-tailed exact p-values based on McNemar's test for comparing prevalences within each subgroup.

equations (4.7) was less than the Wald statistic for the NHANES (9.5) and Hordaland County (8.4) equations, although all were highly statistically significant ($p < 0.0001$). We observed similar patterns when we replaced the GOLD stage 2–4 criterion with the LLN (ratio) and FEV₁ <80% pred criterion (data not shown), although the Wald statistics were closer (6.5 *versus* 7.3 and 8.9).

Finally, the use of the FEV₁/FEV₆ in place of the FEV₁/FVC when using the “LLN (FEV₁/FVC) and FEV₁ <80% pred” criterion had little clinically relevant impact on prevalence estimates, whether computed overall, by age or smoking history categories, or by site (table 3). When we did observe statistically significant differences, the prevalences were generally smaller for the FEV₁/FEV₆-based criterion.

DISCUSSION

This analysis of data from the BOLD study confirmed previously reported limitations associated with the use of the

fixed ratio criterion to define COPD. Adjusting the FEV₁/FVC for normative ageing effects appears to reduce the rate of false-positive diagnoses that has been reported for older individuals [2, 7–10], and the added requirement of a low FEV₁ further reduced the age-related increases in COPD prevalence seen among healthy never-smokers.

A strength of this analysis is that data were gathered using a standardised approach from a wide range of populations, with close attention paid to spirometry quality control. The qualitative similarity of results across sites (fig. 1a) provides strong evidence for the robustness of our findings. The wide variation in characteristics of BOLD sites enabled us to use site-to-site variation in prevalence (assessed using the Wald statistic) as a convenient metric for comparing alternative measures of COPD prevalence, since a desired characteristic of any prevalence estimate is that it yields comparable estimates in different populations after adjusting for known risk factors.

TABLE 4 Characteristics of Burden of Obstructive Lung Disease study participants included in analysis

	Guangzhou, China	Adana, Turkey	Salzburg, Austria	Reykjavik, Iceland	Cape Town, South Africa	Krakow, Poland	Hanover, Germany	Bergen, Norway	Vancouver, BC, Canada	Manila, Philippines	Lexington, KY, USA	Sydney, Australia	London, UK	Uppsala, Sweden
Subjects n	473	806	1258	757	847	526	683	658	827	893	508	541	677	547
Age yrs														
40–49	43.1	42.1	29.2	35.9	39.6	37.5	26.9	25.1	35.2	47.3	26.6	29.0	28.2	22.3
50–59	26.4	30.6	28.5	29.0	31.1	27.4	26.8	28.1	31.9	31.0	37.4	27.2	26.1	34.2
60–69	20.5	17.5	26.0	18.0	19.7	20.3	29.7	22.0	17.3	14.3	25.6	21.4	27.8	26.3
≥70	9.9	9.8	16.3	17.1	9.7	14.8	16.5	24.8	15.6	7.4	10.4	22.4	17.9	17.2
Males n	236	389	685	403	313	266	349	324	344	378	206	265	323	283
Smoking history														
Never-smoker	18.6	19.5	40.2	38.5	15.6	21.1	30.7	32.1	42.2	16.6	30.1	44.5	32.5	37.5
0–10 pack-yrs	11.9	13.6	12.3	16.1	30.5	11.3	16.6	16.4	18.3	24.3	7.8	18.9	18.6	23.0
10–20 pack-yrs	21.2	11.1	13.1	16.1	24.4	15.8	13.2	19.1	11.0	23.5	6.3	10.2	11.1	14.5
≥20 pack-yrs	48.3	55.8	34.5	29.0	28.9	51.5	39.5	32.4	28.2	35.4	55.8	26.4	37.8	25.1
Height cm	166±6.2	168±6.6	176±6.6	180±6.5	169±7.3	173±6.3	175±7.6	178±7.1	175±7.3	163±6.7	176±7.3	172±7.5	175±7.1	178±7.0
BMI kg·m ⁻²	23.3±3.1	27.7±4.6	26.7±3.5	28.3±4.4	24.7±5.7	27.4±3.9	27.7±4.0	26.9±3.8	27.1±4.0	24.4±4.8	30.3±4.9	28.0±4.4	27.5±4.5	27.2±3.8
Females n	237	417	573	354	532	260	334	334	483	515	302	276	354	264
Smoking history														
Never-smoker	93.7	69.5	55.8	39.7	42.1	56.5	50.0	42.2	51.8	69.5	46.4	53.6	43.2	48.1
0–10 pack-yrs	2.1	15.8	16.8	25.8	27.8	15.0	13.8	19.8	18.2	19.0	11.3	18.1	21.8	20.1
10–20 pack-yrs	1.7	5.8	10.1	13.3	15.8	13.1	12.9	18.9	10.6	5.2	8.3	8.7	9.6	12.5
≥20 pack-yrs	2.5	8.9	17.3	21.2	14.3	14.6	23.4	19.2	19.3	5.2	34.1	19.6	25.4	18.9
Height cm	155±6.5	154±6.0	163±6.1	166±6.3	157±6.7	161±5.8	163±7.0	164±6.4	161±7.4	151±5.9	161±5.8	159±6.6	162±7.3	164±6.2
BMI kg·m ⁻²	23.4±3.4	31.5±5.3	26.1±4.9	27.5±5.5	29.8±7.7	28.0±5.3	26.8±5.1	26.2±4.8	26.4±5.9	25.3±4.6	31.1±7.8	27.9±5.7	26.7±5.3	26.8±4.9

Data are presented as % or mean ± sd, unless otherwise stated. BMI: body mass index.

An obvious limitation of this analysis is the lack of a gold standard against which to assess our putative definitions of COPD (indeed, a more accurate descriptor of what we are measuring may simply be chronic airflow limitation). Nonetheless, it is possible to evaluate how alternative definitions perform in individuals who have a low *a priori* probability of disease. Our results confirm previous reports that the fixed ratio criterion lacks specificity and that, as age increases, increasingly misclassifies apparently healthy never-smokers as having COPD [2, 7–10, 12]. This pattern of (apparent) misclassification with increasing age was greatly muted by adding the requirement that the FEV₁ % pred be below a defined threshold, or by replacing the fixed ratio criterion with a criterion that the FEV₁/FVC be below the LLN (fig. 2). However, only the method requiring both an FEV₁/FVC below the LLN and a low FEV₁ (measured as either FEV₁<LLN or FEV₁ <80% pred) largely eliminated this age-related increase.

The upward trend that still persists in figure 2 even with our “best” definitions of COPD may reflect the fact that our “healthy” never-smokers did include some individuals with symptoms. As noted below, this was a pragmatic decision due to the limited number of never-smokers at some sites. The fact that the NHANES-III prediction equations were fit to a cohort whose upper age limit was 80 yrs also may create an upward bias for very old individuals that helps explain the upward drift in figure 2. However, <4% of the BOLD cohort were aged ≥80 yrs; in addition, the NHANES-III prediction equations for FEV₁ include an age-squared term and so allow for accelerated ageing effects.

Notably, the recent ATS/ERS statement recommends using the LLN of the FEV₁/FVC in place of the fixed ratio criterion to diagnose airflow obstruction [21]; a recent paper by SWANNEY *et al.* [12], albeit using pre-bronchodilator spirometry, also supports this recommendation. Use of both an FEV₁/FVC below the LLN and a low FEV₁ was consistently associated with low site-to-site and age-related variation relative to other measures, after adjusting for known risk factors. Assuming that variability about the prediction equations is stable, using the LLN as a threshold for defining low FEV₁ should produce less misclassification [22], although in practice these two measures performed similarly.

The results of our study also add to the evidence suggesting that, without both a low FEV₁/FVC and a low FEV₁, confidence is low that a true lung function abnormality (or airway disease) exists. The current GOLD stage 1 classification was based solely on expert opinion, not on evidence of airway disease or subsequent rapid loss of lung function. Patients with GOLD stage 1 do not have reduced exercise capacity [23]. Among Lung Health Study participants, a rapid fall in FEV₁ was not seen when baseline FEV₁ was >70% pred [24].

Apart from the fixed ratio criterion, the competing measures we evaluated all require use of prediction equations. One of the purported benefits of the fixed ratio criterion is that it does not rely on such equations. However, as SWANNEY *et al.* [12] note, this easy-to-apply definition is only valid at age ~50 yrs. In addition, the fixed ratio criterion is not necessarily easier to use in practice, since even inexpensive pocket spirometers

include a microprocessor that calculates the appropriate LLN for FEV₁/FVC, FEV₁/FEV₆ and FEV₁. Lastly, even GOLD relies on prediction equations to stage disease, so any advantage of the fixed ratio in terms of its simplicity disappears as soon as one looks at clinically relevant impairment (nominally GOLD stage 2 or higher). The only way to overcome the limitations of the current fixed ratio criterion while still avoiding the need for prediction equations would be to establish a series of separate fixed ratio cut points for different ages.

The question then arises, what is a suitable prediction equation for any given population, and what if normative prediction equations do not exist for that population? While the documented variability in lung function that exists among “healthy” never-smokers in different racial groups may reflect, at least in part, true genetic differences in these populations, it also may represent the cumulative effect of environmental exposures, including childhood factors. For this reason, BOLD chose to use a single set of sex-specific prediction equations for all subjects in all sites. We chose the US NHANES-III equations for Caucasian adults because they were derived from a large study conducted in a diverse population with rigorous attention to quality control. We observed similar prevalence estimates using equations derived from Norway’s Hordaland County Respiratory Health Study [20] in place of the NHANES-III equations.

The PLATINO study (Latin American Project for the Investigation of Obstructive Lung Disease), conducted in five Latin American countries using methods similar to those of the BOLD study, used site-specific prediction equations [25]. In BOLD, the use of local prediction equations led to prevalence estimates 2–3 percentage points lower, on average, than those based on a single, common equation. Whether this means that the BOLD prevalence estimates overestimate the “true” estimate, or the local equations underestimate it, we cannot say, but on balance we prefer to maintain the site-to-site variation and see if it can be explained by other risk factors. Because our local equations were fitted to individuals aged ≥40 yrs, while the NHANES-III equations were fitted to adults aged ≥18 yrs, the former may better describe the accelerated ageing that is known to occur in healthy adults. Also, we included symptomatic individuals in our prediction equations as long as they did not report diagnosed disease, whereas the NHANES-III equations required individuals to be asymptomatic. Since there can be large discrepancies between prediction equations based on individuals with and without major respiratory symptoms [26], this may also help to explain the somewhat lower prevalence estimates between the two approaches. One final consideration relating to the use of site-specific prediction equations, particularly if reliable normative equations for that population do not exist, is that the resulting estimates may be highly variable owing to limited sample sizes. For instance, despite relatively large sample sizes from each site, the number of healthy never-smokers available to build our prediction equations was very limited in some sites due to extremely high rates of ever having smoked.

Considerable attention is now being paid to the use of the FEV₁/FEV₆ as an alternative to the FEV₁/FVC, particularly in older, less healthy populations for whom achievement of a

high quality, reproducible FVC may be problematic [27]. Several studies have shown that the FEV₁/FEV₆, for which reliable reference equations exist [19], is a more reproducible measure than is the FEV₁/FVC [28, 29], and predicts subsequent lung function decline about as well as the FEV₁/FVC [30]. Our results (table 3) show that using the FEV₁/FEV₆ in place of the FEV₁/FVC in our definition of "LLN (FEV₁/FVC) and FEV₁ <80% pred" yields very similar prevalence estimates, thus further supporting the use of this alternative measure in future studies of COPD prevalence. Once an obstructive lung disease has been diagnosed, however, changes in FEV₁ should be used to follow disease progression or treatment responses.

Finally, our observation that use of pre-bronchodilator spirometry results in consistently inflated estimates of chronic airflow obstruction, regardless of the definition used, further emphasises the need for using post-bronchodilator spirometry to classify COPD [14]. Our finding that prevalence estimates dropped, on average, ~25% when using post-bronchodilator spirometry is generally consistent with other reports [8, 10, 13]. Although we recognise that well-assessed, normal pre-bronchodilator spirometry has high negative predictive value even in the absence of post-bronchodilator testing, its use is associated with the more serious risk of increased false-positive diagnoses.

In summary, data from the BOLD study confirm previous reports of misclassification using the fixed ratio criterion to measure COPD. As an alternative, we recommend a definition based on an FEV₁/FVC ratio less than the LLN, and an FEV₁ either <80% pred, or below the LLN. This modification of the current GOLD stage 2 severity threshold appears to better account for known ageing effects in healthy never-smokers. While this new definition will likely miss many individuals with mild COPD, it should capture most individuals with clinically significant disease, while minimising the risk of false-positive diagnoses. Finally, substitution of the FEV₁/FEV₆ in place of the FEV₁/FVC in this definition appears to yield similar prevalence estimates and, based on previous reports, may be a more reproducible and practical measure.

STATEMENT OF INTEREST

Statements of interest for W.M. Vollmer and P.L. Enright can be found at www.erj.ersjournals.com/misc/statements.dtl. The BOLD initiative has been funded in part by unrestricted educational grants to the Operations Center. For full details see the online supplementary material available from www.erj.ersjournals.com

ACKNOWLEDGEMENTS

The BOLD study group wishes to acknowledge the contributions of G. Harmoncourt (ndd Medizintechnik AG, Zurich, Switzerland) and P.L. Enright (The University of Arizona, Tucson, AZ, USA) for their assistance with spirometry training and quality control during the study.

REFERENCES

- Halbert RJ, Natoli JL, Gano A, *et al.* Global burden of COPD: systematic review and meta-analysis. *Eur Respir J* 2006; 28: 523–532.
- Celli BR, Halbert RJ, Isonaka S, *et al.* Population impact of different definitions of airway obstruction. *Eur Respir J* 2003; 22: 268–273.
- Rabe KF, Hurd S, Anzueto A, *et al.* Global strategy for the diagnosis, management, and prevention of chronic obstructive

- pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* 2007; 176: 532–555.
- Celli BR, MacNee W. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. *Eur Respir J* 2004; 23: 932–946.
- Viegi G, Pedreschi M, Pistelli F, *et al.* Prevalence of airways obstruction in a general population: European Respiratory Society versus American Thoracic Society definition. *Chest* 2000; 117: 339S–345S.
- Lindberg A, Eriksson B, Larsson LG, *et al.* Seven-year cumulative incidence of COPD in an age-stratified general population sample. *Chest* 2006; 129: 879–885.
- Hardie JA, Buist AS, Vollmer WM, *et al.* Risk of over-diagnosis of COPD in asymptomatic elderly never-smokers. *Eur Respir J* 2002; 20: 1117–1122.
- Shirtcliffe P, Weatherall M, Marsh S, *et al.* COPD prevalence in a random population survey: a matter of definition. *Eur Respir J* 2007; 30: 232–239.
- Hnizdo E, Glindmeyer HW, Petsonk EL, *et al.* Case definitions for chronic obstructive pulmonary disease. *COPD* 2006; 3: 95–100.
- Perez-Padilla R, Hallal PC, Vazquez-Garcia JC, *et al.* Impact of bronchodilator use on the prevalence of COPD in population-based samples. *COPD* 2007; 4: 113–120.
- Cerveri I, Corsico AG, Accordini S, *et al.* Underestimation of airflow obstruction among young adults using FEV₁/FVC <70% as a fixed cut-off: a longitudinal evaluation of clinical and functional outcomes. *Thorax* 2008; 63: 1040–1045.
- Swanney MP, Ruppel G, Enright PL, *et al.* Using the lower limit of normal for the FEV₁/FVC ratio reduces the misclassification of airway obstruction. *Thorax* 2008; 63: 1046–1051.
- Johannessen A, Omenaas ER, Bakke PS, *et al.* Implications of reversibility testing on prevalence and risk factors for chronic obstructive pulmonary disease: a community study. *Thorax* 2005; 60: 842–847.
- Sterk PJ. Let's not forget: the GOLD criteria for COPD are based on post-bronchodilator FEV₁. *Eur Respir J* 2004; 23: 497–498.
- Buist AS, Vollmer WM, Sullivan S, *et al.* The Burden of Obstructive Lung Disease Initiative (BOLD): rationale and design. *COPD* 2005; 2: 277–283.
- Buist AS, McBurnie MA, Vollmer WM, *et al.* International variation in the prevalence of COPD (the BOLD Study): a population-based prevalence study. *Lancet* 2007; 370: 741–750.
- Perez-Padilla R, Vazquez-Garcia JC, Marquez MN, *et al.* The long-term stability of portable spirometers used in a multinational study of the prevalence of chronic obstructive pulmonary disease. *Respir Care* 2006; 51: 1167–1171.
- Miller MR, Hankinson J, Brusasco V, *et al.* Standardisation of spirometry. *Eur Respir J* 2005; 26: 319–338.
- Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med* 1999; 159: 179–187.
- Johannessen A, Lehmann S, Omenaas ER, *et al.* Post-bronchodilator spirometry reference values in adults and implications for disease management. *Am J Respir Crit Care Med* 2006; 173: 1316–1325.
- Pellegrino R, Brusasco V, Viegi G, *et al.* Definition of COPD: based on evidence or opinion? *Eur Respir J* 2008; 31: 681–682.
- Stanojevic S, Wade A, Stocks J, *et al.* Reference ranges for spirometry across all ages: a new approach. *Am J Respir Crit Care Med* 2008; 177: 253–260.
- Pinto-Plata VM, Celli-Cruz RA, Vassaux C, *et al.* Differences in cardiopulmonary exercise test results by American Thoracic Society/European Respiratory Society – Global Initiative for Chronic Obstructive Lung Disease stage categories and gender. *Chest* 2007; 132: 1204–1211.
- Scanlon PD, Connett JE, Waller LA, *et al.* Smoking cessation and lung function in mild-to-moderate chronic obstructive pulmonary disease. The Lung Health Study. *Am J Respir Crit Care Med* 2000; 161: 381–390.

- 25 Menezes AM, Perez-Padilla R, Jardim JR, *et al.* Chronic obstructive pulmonary disease in five Latin American cities (the PLATINO study): a prevalence study. *Lancet* 2005; 366: 1875–1881.
- 26 Johannessen A, Omenaas ER, Eide GE, *et al.* Feasible and simple exclusion criteria for pulmonary reference populations. *Thorax* 2007; 62: 792–798.
- 27 Pezzoli L, Giardini G, Consonni S, *et al.* Quality of spirometric performance in older people. *Age Ageing* 2003; 32: 43–46.
- 28 Swanney MP, Jensen RL, Crichton DA, *et al.* FEV₆ is an acceptable surrogate for FVC in the spirometric diagnosis of airway obstruction and restriction. *Am J Respir Crit Care Med* 2000; 162: 917–919.
- 29 Rosa FW, Perez-Padilla R, Camelier A, *et al.* Efficacy of the FEV₁/FEV₆ ratio compared to the FEV₁/FVC ratio for the diagnosis of airway obstruction in subjects aged 40 years or over. *Braz J Med Biol Res* 2007; 40: 1615–1621.
- 30 Enright PL, Connett JE, Bailey WC. The FEV₁/FEV₆ predicts lung function decline in adult smokers. *Respiratory Medicine* 2002; 96: 444–449.