GOLD criteria overestimate airflow limitation in one-third of cases in the general Finnish population

Annette Kainu, Kirsi Timonen, Ari Lindqvist and Päivi Piirilä

**Affiliations**: 1Heart and Lung Center, Peijas Hospital, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. 2Dept of Clinical Physiology, Central Hospital of Central Finland, Jyväskylä, Finland. 3Dept of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital and University of Eastern Finland, Kuopio, Finland. 4Research Unit of Pulmonary Diseases, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. 5Dept of Clinical Physiology and Nuclear Medicine, HUS Medical Imaging Center, University of Helsinki and Helsinki University Hospital, Helsinki, Finland.

**Correspondence**: Annette Kainu, HUCH Heart and Lung Center, Peijas Hospital, University of Helsinki and Helsinki University Hospital, PO Box 900, 00029 HUS, Finland. E-mail: annette.kainu@helsinki.fi

**ABSTRACT**  
The Global Initiative for Chronic Obstructive Lung Disease (GOLD) diagnostic criteria for chronic obstructive pulmonary disease (COPD) use a fixed threshold of forced expiratory volume in 1 s (FEV1)/forced vital capacity (FVC) ratio (<0.70) in post-bronchodilation spirometry to indicate disease, which has been shown to underestimate and overestimate disease prevalence in younger and older adults, respectively, whilst criteria based on reference values have better accuracy. Differences in reference values have limited their use in international studies. However, the new Global Lung Function Initiative reference values (GLI2012) showed FEV1/FVC to be the least dependent on ethnicity. The aim of this study was to assess the prevalence of airflow limitation with GLI2012 and the degree of underdetection or overestimation related to the use of GOLD in the general population.

A Finnish population sample of 1323 subjects (45% male) with post-bronchodilation spirometry was studied.

80 subjects (6.0%) and 55 subjects (4.2%) were identified with airflow limitation with GOLD and GLI2012 criteria, respectively. The proportion of overestimation with GOLD increased with age from 25% of cases in 50-year-olds to 54% in 70-year-olds. Using z-score-based grading resulted in more dispersion in severity grading.

In conclusion, the GOLD criteria cause a marked overestimation already from 50-year-olds and should be replaced with the GLI2012 criteria to improve diagnostic accuracy.

@ERSPublications  
GOLD criteria overestimate COPD, with >30% of cases having normal spirometry using GLI2012 reference values http://ow.ly/IyGr302sZTs

Received: Nov 08 2015 | Accepted after revision: July 15 2016

Support statement: The FinEsS-Helsinki study has received funding from the Special Governmental Subsidy for Health Sciences Research (project codes TYH1235, TYH 2303, TYH 4251 and TYH 2013354). A. Kainu has received a research grant from the Finnish Anti-tuberculosis Foundation and Jalmari and Rauha Ahokas Foundation. Funding information for this article has been deposited with the Open Funder Registry.

Conflict of interest: Disclosures can be found alongside this article at openres.ersjournals.com

Copyright ©ERS 2016. This article is open access and distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0.
Introduction

In the early 1990s, chronic obstructive bronchitis was poorly understood, mostly under-recognised and with limited research focus. In the late 1990s, the international recognition of the need to improve diagnostics and research in the disease spectrum then coined with the name chronic obstructive pulmonary disease (COPD) led to the formation of the Global Initiative for Chronic Obstructive Lung Disease (GOLD) framework with a major effort aiming for earlier recognition and improved research understanding of this condition [1]. Until then, the main focus of national and international efforts in obstructive airways diseases had been in asthma.

The GOLD framework set forth to improve earlier diagnostics by making diagnostic criteria easily accessible and understandable. Whilst it was already understood that criteria for airflow limitation are dependent on sex, height and age (and also to some extent ethnicity), the GOLD guidelines introduced a simplified “rule of thumb” of the so-called fixed-limit criterion of forced expiratory volume in 1 s (FEV1)/forced vital capacity (FVC) <0.70 in post-bronchodilation spirometry indicating airflow limitation in order to make the diagnosis easier [1]. Reference values vary from one population to another and criteria based on reference values were seen to introduce bias into research efforts. As population reference values vary greatly both in models and lower limits of normal (LLNs), their use was considered difficult to advocate and distribute internationally [1].

There is a large body of knowledge that criteria based on reference values are superior to fixed-limit criteria in the diagnosis of airflow limitation consistent with COPD [2–9]. Selected population studies highlighting the differences found between criteria based on fixed limits and reference values are summarised in table 1. In addition to the differences between the reference values used and their representativeness in the study population, the extent of difference of prevalence estimates is affected by age and sex distribution and the prevalence of smoking in the study population. COPD is a multicomponent disease and the use of a simplified criterion of a fixed limit of FEV1/FVC poses a risk of excessive simplification of a clinical entity that is necessarily complex [31]. In a recent review of the clinical relevance between these two criteria, the conclusion was that when these two criteria were in disagreement, an alternative diagnosis should be considered, particularly among older individuals with less severe airflow limitation where the criteria based on reference values performed better [32]. Even the GOLD guidelines included a recommendation in the 2011 revision that criteria relative to reference values should be considered in the elderly, as the fixed-limit criterion causes overdiagnosis in ageing populations [33]. However, it did not specify what age should be considered “old”. In the 2015 update, the GOLD guidelines already recommend the use of reference values whenever available; however, the main document still endorses the fixed-limit criterion as the main diagnostic criterion [34].

Until recently the problem of a lack of uniform reference values has invalidated comparisons between populations and different geographical areas. In 2012, the European Respiratory Society Global Lung Function Initiative introduced the first global reference values for spirometry with the GLI2012 reference equations [10]. These all-age reference values provide models for different ethnic backgrounds using the generalised additive model for location, scale and shape (GAMLSS) model [23, 35]. The method summarises the changing distribution by three curves representing the median (M), coefficient of variation (S) and skewness, the latter expressed as after a Box–Cox power \( \lambda (L) \) transformation [23, 35]. The FEV1/FVC ratio was found to be the least dependent on ethnicity of all evaluated spirometry variables [10]. Although the validity of the global reference equations has been evaluated in only very few population-based samples, these reference values provide for the first time a uniform model that depicts decline of lung function with age, sex and height, and enable the use of LLN criteria for epidemiology and research between different populations [10, 36]. The GLI2012 reference values have been endorsed by all major international respiratory societies [10]. In a recent study among very old adults (≥80 years old) from Belgium, only airflow limitation defined by GLI2012 was independently associated with mortality, whereas subjects fulfilling the GOLD criterion without reduced FEV1/FVC <LLN had no significantly higher risk of mortality or hospitalisation [30].

The aim of our study was to assess the prevalence of airflow limitation with GLI2012, and the degree of underdetection and overestimation of airflow limitation suggestive of COPD related to the use of the GOLD criterion of a fixed limit of FEV1/FVC <0.70 in post-bronchodilation spirometry in the general population. In addition, we aimed to evaluate the GOLD grading of airflow limitation compared with the z-score-based grading of distribution of severity of impairment.

Materials and methods

We used data from the two Finnish centres participating in the FinEsS studies in respiratory epidemiology. The study population was randomly sampled from the National Population Registry in 10-year age cohorts and represents adults aged 20–75 years old. The study protocol has been described in detail previously.
| Study                        | Country      | Subjects (% male) | Age group years | Nonsmoker FEV1/FVC <0.70 | FEV1/FVC < LLN | Comment                                                                 |
|-----------------------------|--------------|-------------------|-----------------|--------------------------|----------------|-------------------------------------------------------------------------|
| Present study               | Finland      | 1323 (45.0)       | 21–74           | 43.2                     | 9.4 3.3        | LLN fifth percentile GLI2012 [10]. Age-dependent increase in difference between fixed limit and LLN from 0.3% (40–49 years) to 10.6% (70–74 years). |
| Celli et al., 2003 [11]     | USA          | 9838 (47.8)       | 30–80           | 43.8                     | 19.9 13.8      | LLN fifth percentile NHANES [12]. Age-dependent increase in difference between fixed limit and LLN from −2.1% (40–44 years) to 28.8% in (70–74 years). |
| Hansen et al., 2007 [4]     | USA          | 9403 (42.9)       | 20–80           | 62.6¶                   | 12.5¶ 8.2¶     | LLN fifth percentile NHANES [12]. Age-dependent increase in difference between fixed limit and LLN from −3.1% and −4.4% (40–49 years) to 19.9% and 12.1% (>80 years) in men and women, respectively. |
| Vamforensen et al., 2012 [13]| The Netherlands | 592 (50.7)        | 40–70           | 34.4                     | 28.5 19.5      | LLN fifth percentile NHANES [12]. |
| Scholes et al., 2014 [14]   | England and Wales | 7879 (42.3)     | 40–95           | 52.5                     | 26.3 18.6      | LLN fifth percentile GLI2012 [10]. The gap in prevalence between GOLD and LLN increased in older age groups. Sex differences in the risk of obstruction were significantly higher in men using GOLD compared with no significant difference using LLN. |
| Hnizdov et al., 2006 [15]   | USA          | 13842             | 20–80           | NA                       | 21.1          | LLN fifth percentile NHANES [12]. Fixed-limit criterion underestimates airflow obstruction by 29% in 20–49-year-olds and overestimates it by 58% in 50–80-year-olds when compared with FEV1/FVC < LLN and FEV1 <100% predicted. |
| Mannino et al., 2007 [16]   | USA          | 4965 (43.4)       | ≥65             | 45.9                     | 42.1          | LLN fifth percentile NHANES [12]. Follow-up of 11 years found subjects with FEV1/FVC <0.70 but >LLN to have increased mortality and COPD-related hospitalisations. |
| Perez-Padilla et al., 2007 [17]| Latin America | 5183             | ≥40             | NA                       | 21.7          | Evaluated data with LLN fifth percentile NHANES [12] and reference values derived from the same study. Data evaluated in groups of high and low risk of COPD based on smoking exposure. |
| Shirtcliffe et al., 2007 [18]| New Zealand  | 749 (54.2)        | 25–70           | 46.2                     | 14.2          | LLN calculated from a concurrent study sample of healthy nonsmokers. Age-adjusted prevalences shown. |
| Roche et al., 2008* [19]    | France       | 4764 (48.0)       | 45–75           | 48.2                     | 8.7           | LLN fifth percentile European Coalition for Steel and Coal [20]. |

Continued
### TABLE 1 Continued

| Study                  | Country                  | Subjects (%) | Age group years | Nonsmoker | FEV\(_1\)/FVC <0.70 | FEV\(_1\)/FVC <LLN | Comment                                                                 |
|------------------------|--------------------------|--------------|-----------------|-----------|----------------------|----------------------|--------------------------------------------------------------------------|
| Van Dijk et al., 2015  | Canada                   | 4882 (43.0)  | 40–80           | 43.0      | 17.0                 | 12.0                 | LLN fifth percentile NHANES [12]. Evaluated the clinical significance of fixed-ratio versus LLN-based criteria. Fixed ratio alone may lead to misdiagnosis. FEV\(_1\)/FVC <LLN and a low FEV\(_1\) is strongly associated with clinical outcomes. Guidelines should be reconsidered to require both spirometric abnormalities to reduce overdiagnosis of COPD. |
| Danielsson et al., 2012 | Sweden                   | 548 (48.2)   | ≥40             | 42.7      | 15.7                 | 9.3                  | LLN fifth percentile NHANES [12]. Evaluated the clinical significance of fixed-ratio versus LLN-based criteria. Fixed ratio alone may lead to misdiagnosis. FEV\(_1\)/FVC <LLN and a low FEV\(_1\) is strongly associated with clinical outcomes. Guidelines should be reconsidered to require both spirometric abnormalities to reduce overdiagnosis of COPD. |
| Waked et al., 2011     | Lebanon                  | 2201 (48.4)  | ≥40             | 33.3      | 9.7                  | 12.5                 | LLN based on LMS method fifth percentile from study sample [23]. The number of never-smokers was small and the LLN-based definition showed a parabolic prevalence estimation by age. |
| Lamprecht et al., 2011 | Austria                  | 1258 [54.5]  | ≥40             | 43.1      | 24.2                 | 15.3                 | LLN fifth percentile NHANES [12]. Discordance increased with age. Discordant cases more often older, male and never-smokers, and had fewer respiratory symptoms and better FEV\(_1\). Discordant cases had heart disease significantly more often. |
| Garcia-Rio et al., 2011| Spain                    | 3802§        | 40–80           | 27.6§     | 10.2                 | 5.5                  | LLN based on Roca et al. [25]. A subsample of 885 subjects (58.2% male) was further evaluated with subjects diagnosed with GOLD but normal with LLN presenting worse self-reported quality of life, but similar exercise, frequency of exacerbations and indices of systemic effects. |
| Vaz Fragoso et al., 2010| USA                      | 3502 [47.8]  | 40–80           | 40.8      | 27.0                 | 13.8                 | LLN based on LMS method LLN fifth percentile from study sample [23]. |
| Swanney et al., 2008   | The Netherlands          | 4557 Dutch   | 17–90           | NA        | 20.5                 | 12.3                 | LLN from Dutch reference values from that substudy, for UK data LLN of Falaschetti et al. [21] and for USA the fifth percentile LLN of NHANES [12]. GOLD guidelines caused false-positive rates of up to 60% when applied to entire populations. |
|                       | UK                       | 24 604 UK    |                 |           |                      |                      |                                                                          |
|                       | USA                      | 6829 USA     |                 |           |                      |                      |                                                                          |

Data are presented as n or %, unless otherwise stated. GLI2012: Global Lung Function Initiative reference values; NHANES: National Health and Nutrition Examination Survey; GOLD: Global Initiative for Chronic Obstructive Lung Disease; COPD: chronic obstructive pulmonary disease; NA: not available. §: where available; ¶: includes only never-smokers and current smokers from the population sample; #: primary care sample of routine preventive visits; ¶: full sample, study analysis of health-related quality of life pertaining to a smaller sample of subjects with no COPD.
The study has been approved by the Helsinki University Central Hospital Board of Ethics (Helsinki, Finland) and Länsi-Pohja Hospital Board of Ethics (Kemi, Finland).

628 adults (41.4% male) and 695 adults (48.3% male) participated in the clinical study in Helsinki and Kemi, respectively, between 1998 and 2003, including post-bronchodilation spirometry fulfilling predefined quality criteria. All participants were of Caucasian ethnic origin. Smoking history was evaluated by an interview questionnaire administered by a nurse and subjects were categorised as nonsmokers, ex-smokers or current smokers. All subjects having smoked any amount of cigarettes, cigars or pipe tobacco during the previous year were classified as current smokers. Ever-smokers were all current smokers and former smokers. Former smokers were required to have quit smoking at least 1 year previously, but to have smoked at least one cigarette per day for 1 year previously. In Kemi, a nonresponder study was conducted in which younger smoking males were found to be less likely to participate in the survey. However, this did not affect the study results significantly [39].

Airflow limitation in the lower airways was defined in this study by spirometry. Spirometry was completed following American Thoracic Society criteria [40] with bronchodilation testing using 400 µg salbutamol aerosol in two 200 µg doses using a spacer (Ventoline E沃haler with Volumatic; GlaxoSmithKline, Brentford, UK). Only post-bronchodilation values of FVC, FEV1 and the FEV1/FVC ratio after bronchodilation were analysed and reported. The descriptive statistics of the study sample are outlined in table 2. The prevalence of smoking in different age categories in this study was at a similar level to other population studies of the time (e.g. the Health 2000 study [41]), thus the sample would seem to be representative of the general population also in terms of exposure levels.

GLI2012 reference values for Caucasians indicated airflow limitation if the FEV1/FVC ratio in the post-bronchodilation spirometry was below the defined fifth percentile LLN of z-score $-1.645$ [10]. GOLD criteria define airflow limitation as post-bronchodilation FEV1/FVC <0.70 [34]. A true-positive case of airflow limitation consistent with COPD was defined as a case with FEV1/FVC <LLN and FEV1/FVC <0.70, a false-negative case was defined as a case with FEV1/FVC <LLN but FEV1/FVC ≥0.70 and a false-positive case was defined as a case with FEV1/FVC ≥LLN but FEV1/FVC <0.70. The false-positive rate is the number of false-positive cases per study sample expressed as a percentage. The proportion of false-positive cases of all cases identified using the GOLD criterion is false positives divided by the sum of true positives and false positives.

The grading of degree of severity of airflow limitation was evaluated with the decreased ventilatory capacity, i.e. lowered FEV1 relative to FEV1 % pred with the GLI2012 reference value as suggested by the GOLD guidelines and relative to the individually calculated z-score as suggested by Quanjer et al. [10, 42]. In the GOLD recommendation, values of FEV1 % pred ≥80 were considered mild COPD, 50–79 as moderate COPD, 30–49 as moderately severe COPD and <30 as severe COPD [34]. In the z-score-based
assessment of FEV1, individual z-score values >−2 were considered mild, between −2 and −2.5 as moderate, between −2.5 and −3 as moderately severe, between −3 and −4 as severe, and <−4 as very severe impairment of the ventilatory capacity [42].

The GLI2012 predicted values and z-scores for each study participant were calculated using R (version 2.15.1; www.cran-R.org) with the macro provided by the Global Lung Function Initiative [43]; all other analyses were conducted using SPSS Statistics for Macintosh version 22.0 (IBM, Armonk, NY, USA). The Chi-squared test was used to compare groups and the Mantel–Haenszel Chi-squared test was used for trends. A p-value of 0.05 was considered significant in all analyses.

Results
The distribution of post-bronchodilation FEV1/FVC relative to GLI2012 reference values versus the absolute FEV1/FVC ratio is shown in figure 1. The GOLD criteria identified overall 6.0% of subjects with airflow limitation; 9.4% of males and 3.3% of females (p<0.001). Correspondingly, the GLI2012 criteria identified 4.2% of subjects with airflow limitation; 6.0% and 2.6% of males and females, respectively (p=0.002). Among subjects with airflow limitation identified by GOLD, 39.3% of cases were false positive in males and 25.0% of cases were false positive in females (p=0.12). The rate of false positives increases with age in both males and females, as shown in figure 2. The rate of false positives increases slightly but not significantly with height: 1.4% for subjects <160 cm, 2.2% for subjects 160–179 cm and 2.5% for subjects ≥180 cm (p=0.713). Of the false-positive subjects (n=28), taller subjects had higher mean z-scores for FEV1/FVC for subjects <160 (−1.57, 95% CI −1.79–−1.35), 160–179 (−1.26, 95% CI −1.39–−1.14) and >180 cm (−1.13, 95% CI −1.36–−0.90), as shown in figure 3. The proportion of false positives of all detected cases was 45.8% in subjects ≥60 years of age (table 3), but false positives were found already in subjects aged 40–49 years. However, the numbers of cases in the younger age categories were limited.

As the GOLD criteria are prone to overdiagnosis of older subjects with normal FEV1/FVC based on LLN of reference values, the excess cases are found mostly in the mild category with normal FEV1, as shown in table 4. These subjects have both normal FEV1/FVC and FEV1 based on LLN of reference values. In the moderate to severe categories, the population prevalences are thus equal. Use of simplified percentage of predicted criteria (GOLD grading) failed to identify the individual variation in the LLN in the older age groups and thus more subjects end up in the moderate COPD category. Figure 4 illustrates the relationship between FEV1 % pred GLI2012 and the individually calculated FEV1 z-score in the whole population sample. Using 80% of predicted overidentifies reduced ventilatory capacity at the population level in 45 subjects (3.4%).

Discussion
The GOLD criteria were originally introduced to improve diagnostics to facilitate earlier identification of subjects with COPD [1]. However, the criteria are known to overdiagnose older individuals [2, 3, 7, 18, 22].

FIGURE 1 Distribution of post-bronchodilation forced expiratory volume in 1 s (FEV1)/forced vital capacity (FVC) relative to Global Lung Function Initiative reference values [GLI2012] [10] versus absolute FEV1/FVC ratio in the study population (n=1323: males n=596; females n=727). Fifth percentile lower limit of normal of z-score −1.645 and the Global Initiative for Chronic Obstructive Lung Disease criteria [34] fixed limit of FEV1/FVC=0.70 are indicated by dotted lines.
The use of more accurate criteria based on reference values has been limited by the variability of local reference values and the lack of a uniform definition of abnormal. The GLI2012 reference values provide a uniform definition for LLN for the first time [10]. We show that in a population-based sample in Finland the spirometric GOLD criteria overestimate COPD prevalence by 35% and up to 54% of cases in the oldest age category of ≥70 years. Overdiagnosis starts already at age 50–59 years where subjects with normal spirometry on LLN criteria would be diagnosed with airflow limitation consistent with COPD according to the GOLD criteria.

Clinical relevance of diagnostic criterion for airflow limitation
COPD is not a single disease but encompasses different diseases with varying prognosis [34]. It indicates a functional disorder characterised by a chronic and progressive reduction in maximum expiratory flow without clarifying the underlying mechanism [23]. There is no doubt that any diagnostic limit will have its limitations as well. Reduced FEV1/FVC is not a dichotomous variable, but instead a continuum of values including elevated, normal, borderline and significantly reduced values. MANNINO et al. [16] have shown that subjects ≥65 years of age with FEV1/FVC <0.70 but above their LLN were at an increased risk of death or COPD-related hospitalisation during an 11-year follow-up. The results were, however, contested after publication, with the rate of diagnosis and events labelled as COPD-related hospitalisations having been overly inclusive [44, 45]. In a recent review by VAN DIJK et al. [32] evaluating the clinical significance of both criteria, the criteria based on reference values worked better in elderly subjects where discordant cases should be further evaluated for other diagnoses. The FEV1/FVC ratio is directly associated with FEV1, which has been shown to be one of the strongest predictors of all-cause morbidity and mortality.

Clinical relevance of diagnostic criterion for airflow limitation
COPD is not a single disease but encompasses different diseases with varying prognosis [34]. It indicates a functional disorder characterised by a chronic and progressive reduction in maximum expiratory flow without clarifying the underlying mechanism [23]. There is no doubt that any diagnostic limit will have its limitations as well. Reduced FEV1/FVC is not a dichotomous variable, but instead a continuum of values including elevated, normal, borderline and significantly reduced values. MANNINO et al. [16] have shown that subjects ≥65 years of age with FEV1/FVC <0.70 but above their LLN were at an increased risk of death or COPD-related hospitalisation during an 11-year follow-up. The results were, however, contested after publication, with the rate of diagnosis and events labelled as COPD-related hospitalisations having been overly inclusive [44, 45]. In a recent review by VAN DIJK et al. [32] evaluating the clinical significance of both criteria, the criteria based on reference values worked better in elderly subjects where discordant cases should be further evaluated for other diagnoses. The FEV1/FVC ratio is directly associated with FEV1, which has been shown to be one of the strongest predictors of all-cause morbidity and mortality.

Clinical relevance of diagnostic criterion for airflow limitation
COPD is not a single disease but encompasses different diseases with varying prognosis [34]. It indicates a functional disorder characterised by a chronic and progressive reduction in maximum expiratory flow without clarifying the underlying mechanism [23]. There is no doubt that any diagnostic limit will have its limitations as well. Reduced FEV1/FVC is not a dichotomous variable, but instead a continuum of values including elevated, normal, borderline and significantly reduced values. MANNINO et al. [16] have shown that subjects ≥65 years of age with FEV1/FVC <0.70 but above their LLN were at an increased risk of death or COPD-related hospitalisation during an 11-year follow-up. The results were, however, contested after publication, with the rate of diagnosis and events labelled as COPD-related hospitalisations having been overly inclusive [44, 45]. In a recent review by VAN DIJK et al. [32] evaluating the clinical significance of both criteria, the criteria based on reference values worked better in elderly subjects where discordant cases should be further evaluated for other diagnoses. The FEV1/FVC ratio is directly associated with FEV1, which has been shown to be one of the strongest predictors of all-cause morbidity and mortality.

Clinical relevance of diagnostic criterion for airflow limitation
COPD is not a single disease but encompasses different diseases with varying prognosis [34]. It indicates a functional disorder characterised by a chronic and progressive reduction in maximum expiratory flow without clarifying the underlying mechanism [23]. There is no doubt that any diagnostic limit will have its limitations as well. Reduced FEV1/FVC is not a dichotomous variable, but instead a continuum of values including elevated, normal, borderline and significantly reduced values. MANNINO et al. [16] have shown that subjects ≥65 years of age with FEV1/FVC <0.70 but above their LLN were at an increased risk of death or COPD-related hospitalisation during an 11-year follow-up. The results were, however, contested after publication, with the rate of diagnosis and events labelled as COPD-related hospitalisations having been overly inclusive [44, 45]. In a recent review by VAN DIJK et al. [32] evaluating the clinical significance of both criteria, the criteria based on reference values worked better in elderly subjects where discordant cases should be further evaluated for other diagnoses. The FEV1/FVC ratio is directly associated with FEV1, which has been shown to be one of the strongest predictors of all-cause morbidity and mortality.

Clinical relevance of diagnostic criterion for airflow limitation
COPD is not a single disease but encompasses different diseases with varying prognosis [34]. It indicates a functional disorder characterised by a chronic and progressive reduction in maximum expiratory flow without clarifying the underlying mechanism [23]. There is no doubt that any diagnostic limit will have its limitations as well. Reduced FEV1/FVC is not a dichotomous variable, but instead a continuum of values including elevated, normal, borderline and significantly reduced values. MANNINO et al. [16] have shown that subjects ≥65 years of age with FEV1/FVC <0.70 but above their LLN were at an increased risk of death or COPD-related hospitalisation during an 11-year follow-up. The results were, however, contested after publication, with the rate of diagnosis and events labelled as COPD-related hospitalisations having been overly inclusive [44, 45]. In a recent review by VAN DIJK et al. [32] evaluating the clinical significance of both criteria, the criteria based on reference values worked better in elderly subjects where discordant cases should be further evaluated for other diagnoses. The FEV1/FVC ratio is directly associated with FEV1, which has been shown to be one of the strongest predictors of all-cause morbidity and mortality.

Clinical relevance of diagnostic criterion for airflow limitation
COPD is not a single disease but encompasses different diseases with varying prognosis [34]. It indicates a functional disorder characterised by a chronic and progressive reduction in maximum expiratory flow without clarifying the underlying mechanism [23]. There is no doubt that any diagnostic limit will have its limitations as well. Reduced FEV1/FVC is not a dichotomous variable, but instead a continuum of values including elevated, normal, borderline and significantly reduced values. MANNINO et al. [16] have shown that subjects ≥65 years of age with FEV1/FVC <0.70 but above their LLN were at an increased risk of death or COPD-related hospitalisation during an 11-year follow-up. The results were, however, contested after publication, with the rate of diagnosis and events labelled as COPD-related hospitalisations having been overly inclusive [44, 45]. In a recent review by VAN DIJK et al. [32] evaluating the clinical significance of both criteria, the criteria based on reference values worked better in elderly subjects where discordant cases should be further evaluated for other diagnoses. The FEV1/FVC ratio is directly associated with FEV1, which has been shown to be one of the strongest predictors of all-cause morbidity and mortality.

Clinical relevance of diagnostic criterion for airflow limitation
COPD is not a single disease but encompasses different diseases with varying prognosis [34]. It indicates a functional disorder characterised by a chronic and progressive reduction in maximum expiratory flow without clarifying the underlying mechanism [23]. There is no doubt that any diagnostic limit will have its limitations as well. Reduced FEV1/FVC is not a dichotomous variable, but instead a continuum of values including elevated, normal, borderline and significantly reduced values. MANNINO et al. [16] have shown that subjects ≥65 years of age with FEV1/FVC <0.70 but above their LLN were at an increased risk of death or COPD-related hospitalisation during an 11-year follow-up. The results were, however, contested after publication, with the rate of diagnosis and events labelled as COPD-related hospitalisations having been overly inclusive [44, 45]. In a recent review by VAN DIJK et al. [32] evaluating the clinical significance of both criteria, the criteria based on reference values worked better in elderly subjects where discordant cases should be further evaluated for other diagnoses. The FEV1/FVC ratio is directly associated with FEV1, which has been shown to be one of the strongest predictors of all-cause morbidity and mortality.
even in the absence of lung disease [46–49]. Subjects with a low but still normal FEV1/FVC ratio might be less healthy, and they do have lower lung function than their age and gender peers with higher FEV1/FVC ratio and normal FEV1, but this does not imply causality with the diagnostic criteria for COPD. A reduced FEV1/FVC ratio should be understood as a signal, but the diagnosis of COPD should be limited to truly abnormal values. The use of assessment based on reference values also facilitates better understanding of the grey areas around diagnostic criteria, i.e. a continuum of values around the LLN could improve the understanding of general practitioners of patients at risk. The predicted FEV1/FVC ratio is also an estimate with an associated degree of prediction error. Transition to the use of z-scores instead of percentage of predicted with the implementation of the GLI2012 reference values will improve the accuracy of the LLN especially in the elderly. Further research is needed to identify the significance of borderline categories of z-score combined with exposure and symptoms to possibly facilitate earlier identification of subjects at increased risk in primary care. However, airflow limitation is just one of the functional components of COPD. The addition of FEV1 % pred and ratio of residual volume to total lung capacity has been found to improve the identification of clinically relevant COPD [50].

**Potential effect of height on degree of overdiagnosis of airflow limitation**

The continued use of FEV1/FVC fixed-limit criterion by GOLD guidelines has been advocated with similarities with the fixed limits agreed upon for other physiological measures such as elevated blood pressure [31]. There are, however, fundamental physiological differences in these variables. One of the most evident is

| TABLE 4 Grading of degrees of airflow limitation in the study population (n=1323) according to the Global Initiative for Chronic Obstructive Lung Disease (GOLD) [34] and to QUANJER et al. [42] in subjects with lower limit of normal criteria airflow limitation (forced expiratory volume in 1 s [FEV1]/forced vital capacity (FVC) post-bronchodilation z-score < −1.645 on Global Lung Function Initiative reference values [GLI2012]) or with GOLD fixed-limit criterion [FEV1/FVC post-bronchodilation <0.70] |
|---|---|---|---|---|---|---|---|
| FEV1/FVC <0.70 | FEV1/FVC <LLN |
| Subjects | 80 | 55 |
| GOLD grading | | | | |
| Mild (>80% predicted) | 51.2 | 40.0 |
| Moderate (50–80% predicted) | 42.5 | 50.9 |
| Moderately severe (30–50% predicted) | 6.3 | 9.1 |
| Severe (<30% predicted) | 0.0 | 0.0 |
| z-score-based grading | | | | |
| Mild [z-score > −2] | 70.0 | 60.0 |
| Moderate [z-score −2 to −2.5] | 8.8 | 9.1 |
| Moderately severe [z-score −2.5 to −3] | 10.0 | 14.5 |
| Severe [z-score −3 to −4] | 11.3 | 16.4 |
| Very severe [z-score < −4] | 0.0 | 0.0 |

Data are presented as n or % of subjects with airflow limitation.
the dependence of the FEV1/FVC ratio on a person’s height. Using a fixed-limit criterion can cause an increasing level of overdiagnosis of airflow limitation with increasing height. Shorter subjects have normally higher ratios and taller subjects have physiologically lower ratios [10]. In our study the rate of false positives was not significantly higher in taller cohorts; however, of the false-positive cases, taller subjects had a significantly higher FEV1/FVC ratio relative to reference values (figure 3). Unfortunately the absolute number of obstructive individuals in our sample was too small for further analyses. Geographically, when comparing populations, taller populations could thus have higher rates of false positives. Additionally, the younger age cohorts in many countries are taller than their parents and grandparents. In Finland, the average height of the population increased by 5 cm in both males and females from the 1970s to the 2010s [51]. Thus, the level of overdiagnosis of airflow limitation using the FEV1/FVC fixed-limit criterion might have an increasing trend as taller generations are reaching middle age.

Impact of overestimation with GOLD criteria

The GOLD criteria were introduced to facilitate the interpretation of spirometry, which was considered difficult [1]. Generalisations can be useful, but they should not hinder correct diagnostics and treatment choices. In moderately to severely reduced FEV1, the GOLD criterion functions well and correct conclusions are reached even with this simplified “rule of thumb” [34]. These are also the patients that should be considered for COPD medication and among whom the results from pharmaceutical trials, using the inclusion criteria based on the GOLD criterion, are likely to be applicable to judge the costs and benefits of treatment. It is, however, the potential for harm that should drive pulmonary professionals towards change. All subjects benefit from the main treatment alternatives: quitting smoking, exercise, vaccinations and eating a healthy diet to maintain normal weight. The opportunity for harm is twofold: when pharmaceutical treatment is initiated to treat “reduced lung function” that is in fact normal and, moreover, when other reasons for shortness of breath, dyspnoea and other respiratory symptoms are not sought, when the patient is mislabelled as having COPD. There are no published studies that would have proven any benefit of treating subjects with normal lung function (without asthma) with inhaled steroids or long-acting bronchodilators, which on the contrary may have significant side-effects for the patients. The level of overestimation of airflow limitation, determined here as 35% of the population, can result in hugely excessive costs in terms of unnecessary pharmaceutical treatments.

Grading of degree of airflow limitation

We evaluated the grading system proposed by GOLD for applicability relative to z-score-based severity grading. The GOLD limits of 80%, 50% and 30% are not based on actual variability of FEV1 in healthy nonsmokers or subjects with COPD, but were originally chosen based on expert opinion as convenient limits that are easy to remember [1, 34]. The GLI2012 reference values introduce individual z-scores, with increasing variation of lung function with age taken into consideration with a separately modelled coefficient of variation (S) [10]. Traditionally, a fixed limit of the fifth percentile, commonly 80% of predicted, has been considered as the LLN for all ages [10]. The true fifth percentile varies with age and what commonly is 80% of predicted in those 20–30 years old is more likely 65–75% in the older age categories [10]. Thus, the use of z-scores to incorporate this variability gives a more accurate estimate of LLN, but also the degrees of reduced lung function, as seen in figure 3 and table 3. The GOLD limits fail to identify differences in FEV1 levels. Use of more accurate z-score-based grading of COPD could improve the grading of COPD and assessment of treatment options, even though airflow limitation is just one of the phenotypes of COPD [52].
Limitations of the study

There is no golden standard for the true prevalence of airflow limitation in the general population. Among healthy nonsmokers the recommended fifth percentile LLN should by definition identify 5% of subjects with reduced values. In the absence of a gold standard, comparisons between different diagnostic limits are relative. The population sample has a limited number of subjects with pathological airflow limitation consistent with COPD to assess, which limits further analysis. However, we have shown that the GLI2012 values differentiate between subjects with airflow limitation consistent with COPD and can potentially give severity grading of reduced ventilatory capacity more substance. Further studies on patient populations are needed to evaluate the grading system more thoroughly. The GLI2012 limit of $z$-score $\leq -3$ would seem to correspond to the previous limit of 50% of predicted fairly well (figure 4). The low prevalence of airflow limitation found in the general population in comparison with other studies (table 1) is notable especially regarding various selection biases. However, compared with other Finnish studies from the same timespan (e.g. the Health 2000 study [41]), the population prevalences of GOLD criteria airflow limitation and smoking are similar. In addition, a nonresponder study was performed in Kemi, where the nonresponders were found to be more often smoking young men, which did not significantly affect survey results [39]. Airflow limitation consistent with COPD is prevalent in subjects $\geq 40$ years old and there was no difference in these age groups in terms of participation.

The GLI2012 reference values have been shown to underestimate lung volumes by 3–5% in the Finnish population, thus further study using severity grading from locally derived reference equations [39] is needed. Also in this study, the GLI2012 reference values would seem to underestimate lung volumes (FEV1), thus underestimating the degree of COPD by spirometric criteria, especially in females (table 3 and figure 4). The new Finnish reference values have been calculated with GAMLSS modelling similar to the GLI2012 model [53]. Local data can bridge the gap between GLI2012 and the clinical patient evaluation. However, this does not invalidate the use of GLI2012 reference values in international comparisons. Instead, the local data should be evaluated with models compatible with the modelling approaches taken with the Global Lung Function Initiative. The Global Lung Function Initiative has made software tools available to calculate correction factors for divergent population samples and it is foreseen that the Global Lung Function Initiative models will be further developed to incorporate new data in the future [10].

Conclusions

The GLI2012 reference values provide a uniform standard for the diagnosis of airflow limitation and should be used in lieu of the current GOLD criteria. Clinicians should be aware of the overestimation of airflow limitation and COPD resulting from the current widespread use of the fixed-limit GOLD criterion. One-third of cases identified with airflow limitation using the GOLD criterion had normal spirometry with GLI2012 reference values. Given the age distribution at the population level, this could result in potentially over half of cases identified being false positives. Overestimation of airflow limitation consistent with COPD in the middle aged and elderly can result in excessive pharmaceutical costs, side-effects of drugs used to treat COPD, and under-recognition of other causes of dyspnoea, cough and sputum production in the ageing population. We recommend the speedy recognition of GLI2012 as a uniform definition of airflow limitation in epidemiological research to substitute the spirometric GOLD criterion in the diagnosis of post-bronchodilation airflow limitation consistent with COPD.

Acknowledgements

We warmly thank the Research Unit of Pulmonary Diseases in Helsinki University Hospital for high-quality measurements and undertaking the FinEsS-Helsinki study, and Dr Jyrki-Tapani Kotaniemi (Päijät-Häme Central Hospital, Lahti, Finland) for overseeing the FinEsS-Kemi study.

References

1. Global Initiative for Chronic Obstructive Lung Disease. Global Strategy for the Diagnosis, Management and Prevention of Chronic Obstructive Pulmonary Disease. NHLBI/WHO Workshop Report. Bethesda, National Heart, Lung and Blood Institute, 2001.
2. 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.
3. Roberts SD, Farber MO, Knox KS, et al. FEV1/FVC ratio misclassifies patients with obstruction at the extremes of age. *Chest* 2006; 130: 200–206.
4. Hansen JE, Sun X-G, Wasserman K. Spirometric criteria for airway obstruction. Use percentage of FEV1/FVC ratio below the fifth percentile, not <70%. *Chest* 2007; 131: 349–355.
5. 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.
6. Schermer TR, Smele U, Thoonen BP, et al. Current clinical guideline definitions for airflow obstruction leads to substantial overdiagnosis of COPD in primary care. *Eur Respir J* 2008; 32: 945–952.
7. Swanney MP, Ruppel G, Enright PL, et al. Using the lower limit of normal for the FEV1/FVC ratio reduces the misclassification of airflow obstruction. *Thorax* 2008; 63: 1046–1051.
Mannino DM, Buist AS, Vollmer WM. Authors’ reply to two letters to the editor. Thorax 2007; 62: 1108–1109.

Quanjer PH, Furberg CD. GOLD COPD stage I is not associated with increased risk of death. Respir Med 2012; 106: 153.

Ryan G, Knuiman MW, Divitini ML, et al. Decline in lung function and mortality: the Busselton Health Study. J Epidemiol Community Health 1999; 53: 230–234.

Schünemann HJ, Dorn J, Grant BJB, et al. Pulmonary Function is a long-term predictor of mortality in the general population: 29-year follow-up of the Buffalo Health Study. Chest 2000; 118: 656–664.

Staenv K, Aaser E, Sandvik L, et al. Lung function, smoking and mortality in a 26-year follow-up of healthy middle-aged males. Eur Respir J 2005; 25: 618–625.

Sin DD, Wu LL, Man SFP. The relationship between reduced lung function and cardiovascular mortality. A population-based study and a systematic review of the literature. Chest 2005; 127: 1952–1959.

Güder G, Brenner S, Angermann CE, et al. GOLD or lower limit of normal definition? A comparison with expert-based diagnosis of chronic obstructive pulmonary disease in a prospective cohort-study. Respir Res 2012; 13: 13.

Helldán A, Helakorpi S. Health Behaviour and Health among the Finnish Adult Population, Spring 2014. Helsinki, National Institute for Health and Welfare, 2015.

Vaz Fragoso CA, Concato J, McAvay G, et al. Staging the severity of chronic obstructive pulmonary disease in older persons based on spirometric Z-scores. J Am Geriatr Soc 2011; 59: 1847–1854.

Kainu A, Timonen KL, Toikka J, et al. Reference values of spirometry for Finnish adults. Clin Phys Funct Imaging 2016; 36: 246–358.