




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

Annette Kainu <sup>1</sup>, Kirsi Timonen<sup>2,3</sup>, Ari Lindqvist<sup>4</sup> and Päivi Piirilä<sup>5</sup>

**Affiliations:** <sup>1</sup>Heart and Lung Center, Peijas Hospital, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. <sup>2</sup>Dept of Clinical Physiology, Central Hospital of Central Finland, Jyväskylä, Finland. <sup>3</sup>Dept of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital and University of Eastern Finland, Kuopio, Finland. <sup>4</sup>Research Unit of Pulmonary Diseases, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. <sup>5</sup>Dept 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](mailto: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 (FEV<sub>1</sub>)/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 FEV<sub>1</sub>/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](http://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 (FEV<sub>1</sub>)/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 FEV<sub>1</sub>/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 FEV<sub>1</sub>/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 ( $\geq 80$  years old) from Belgium, only airflow limitation defined by GLI2012 was independently associated with mortality, whereas subjects fulfilling the GOLD criterion without reduced FEV<sub>1</sub>/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 FEV<sub>1</sub>/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

TABLE 1 General population studies comparing the diagnostic criteria of forced expiratory volume in 1 s (FEV<sub>1</sub>)/forced vital capacity (FVC) <0.70 and FEV<sub>1</sub>/FVC < lower limit of normal (LLN)

Study	Country	Subjects (% male <sup>#</sup> )	Age group years	Nonsmoker	FEV <sub>1</sub> /FVC <0.70		FEV <sub>1</sub> /FVC <LLN		Comment
					Male	Female	Male	Female	
<b>Present study</b>	Finland	1323 (45.0)	21–74	43.2	9.4	3.3	6.0	2.6	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).
<b>CELLI <i>et al.</i>, 2003 [11]</b>	USA	9838 (47.8)	30–80	43.8	19.9	13.8	15.0	13.5	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).
<b>HANSEN <i>et al.</i>, 2007 [4]</b>	USA	9403 <sup>¶</sup> (42.9)	20–80	62.6 <sup>¶</sup>	12.5 <sup>¶</sup>	8.2 <sup>¶</sup>	11.4 <sup>¶</sup>	9.4 <sup>¶</sup>	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.
<b>VANFLETEREN <i>et al.</i>, 2012 [13]</b>	The Netherlands	592 (50.7)	40–≥70	34.4	28.5	19.5	19.7	17.8	LLN fifth percentile NHANES [12].
<b>SCHOLES <i>et al.</i>, 2014 [14]</b>	England and Wales	7879 (42.3)	40–95	52.5	26.3	18.6	15.0	11.3	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.
<b>HNIZDO <i>et al.</i>, 2006 [15]</b>	USA	13 842	20–80	NA	21.1		18.5		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 FEV <sub>1</sub> /FVC <LLN and FEV <sub>1</sub> <100% predicted.
<b>MANNINO <i>et al.</i>, 2007 [16]</b>	USA	4965 (43.4)	≥65	45.9	42.1		19.3		LLN fifth percentile NHANES [12]. Follow-up of 11 years found subjects with FEV <sub>1</sub> /FVC <0.70 but >LLN to have increased mortality and COPD-related hospitalisations.
<b>PEREZ-PADILLA <i>et al.</i>, 2007 [17]</b>	Latin America	5183	≥40	NA	21.7		11.7		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. Age-dependent increase in prevalence in both groups.
<b>SHIRTCLIFFE <i>et al.</i>, 2007 [18]</b>	New Zealand	749 (54.2)	25–≥70	46.2	14.2		9.0		LLN calculated from a concurrent study sample of healthy nonsmokers. Age-adjusted prevalences shown.
<b>ROCHE <i>et al.</i>, 2008* [19]</b>	France	4764 (48.0)	45–≥75	48.2	8.7		6.4		LLN fifth percentile European Coalition for Steel and Coal [20].

Continued

TABLE 1 Continued

Study	Country	Subjects (% male <sup>#</sup> )	Age group years	Nonsmoker	FEV <sub>1</sub> /FVC <0.70		FEV <sub>1</sub> /FVC <LLN		Comment
					Male	Female	Male	Female	
SWANNEY <i>et al.</i> , 2008 [7]	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 <i>et al.</i> [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			13.4		14.1		
	USA	6829 USA			21.8		15.5		
Vaz FRAGOSO <i>et al.</i> , 2010 [22]	USA	3502 (47.8)	40–80	40.8	27.0		13.8	LLN based on LMS method LLN fifth percentile from study sample [23].	
GARCIA-RIO <i>et al.</i> , 2011 [24]	Spain	3802 <sup>§</sup>	40–80	27.6 <sup>§</sup>	10.2		5.5	LLN based on ROCA <i>et al.</i> [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.	
LAMPRECHT <i>et al.</i> , 2011 [26]	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 <sub>1</sub> . Discordant cases had heart disease significantly more often.	
WAKED <i>et al.</i> , 2011 [27]	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.	
DANIELSSON <i>et al.</i> , 2012 [28]	Sweden	548 (48.2)	≥40	42.7	15.7		9.3	LLN fifth percentile NHANES [12].	
VAN DIJK <i>et al.</i> , 2015 [29]	Canada	4882 (43.0)	40–≥80	43.0	17.0		11.2	LLN fifth percentile NHANES [12]. Evaluated the clinical significance of fixed-ratio- versus LLN-based criteria. Fixed ratio alone may lead to misdiagnosis. FEV <sub>1</sub> /FVC <LLN and a low FEV <sub>1</sub> is strongly associated with clinical outcomes. Guidelines should be reconsidered to require both spirometric abnormalities to reduce overdiagnosis of COPD.	
TURKESHI <i>et al.</i> , 2015 [30]	Belgium	411 (37.0)	≥80	69.6	27.0		9.2	LLN fifth percentile GLI2012 [10]. Only airflow limitation by GLI2012 was independently associated with mortality. Subjects with FEV <sub>1</sub> /FVC <0.70 but above GLI2012 LLN had no significantly higher risk of mortality or hospitalisation.	

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. <sup>#</sup>: where available; <sup>¶</sup>: includes only never-smokers and current smokers from the population sample; <sup>\*</sup>: primary care sample of routine preventive visits; <sup>§</sup>: full sample, study analysis of health-related quality of life pertaining to a smaller subsample of subjects with no COPD.

[37–39]. 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 Evohaler with Volumatic; GlaxoSmithKline, Brentford, UK). Only post-bronchodilation values of FVC, FEV<sub>1</sub> and the FEV<sub>1</sub>/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 FEV<sub>1</sub>/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 FEV<sub>1</sub>/FVC  $<0.70$  [34]. A true-positive case of airflow limitation consistent with COPD was defined as a case with FEV<sub>1</sub>/FVC  $<LLN$  and FEV<sub>1</sub>/FVC  $<0.70$ , a false-negative case was defined as a case with FEV<sub>1</sub>/FVC  $<LLN$  but FEV<sub>1</sub>/FVC  $\geq 0.70$  and a false-positive case was defined as a case with FEV<sub>1</sub>/FVC  $\geq LLN$  but FEV<sub>1</sub>/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 FEV<sub>1</sub> relative to FEV<sub>1</sub> % 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 FEV<sub>1</sub> % pred  $\geq 80$  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

TABLE 2 Characteristics of the study population

	Male	Female
<b>Subjects</b>	596	727
<b>Age years</b>	48.8±13.2	48.2±13.0
<b>Height cm</b>	176.5±6.6	163.0±6.1
<b>BMI kg·m<sup>-2</sup></b>	26.4±3.9	25.9±4.9
<b>FEV<sub>1</sub> L</b>	3.9±0.9	2.9±0.6
<b>FVC L</b>	4.9±0.9	3.5±0.6
<b>FEV<sub>1</sub>/FVC</b>	0.796±0.082	0.822±0.063
<b>FEV<sub>1</sub> z-score GLI2012 [10]</b>	0.04±1.07	0.11±0.94
<b>FVC z-score GLI2012 [10]</b>	-0.02±0.90	-0.03±0.84
<b>FEV<sub>1</sub>/FVC z-score GLI2012 [10]</b>	0.09±1.07	0.19±0.86
<b>Smoking status</b>		
Nonsmoker	31.4	52.8
Ex-smoker	36.4	22.3
Current smoker	32.2	24.9
<b>Physician-diagnosed asthma</b>	6.4	7.0
<b>Physician-diagnosed COPD</b>	2.9	2.8

Data are presented as n, mean±SD or %. BMI: body mass index; FEV<sub>1</sub>: forced expiratory volume in 1 s; FVC: forced vital capacity; GLI2012: Global Lung Function Initiative reference values; COPD: chronic obstructive pulmonary disease.

assessment of FEV<sub>1</sub>, 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.

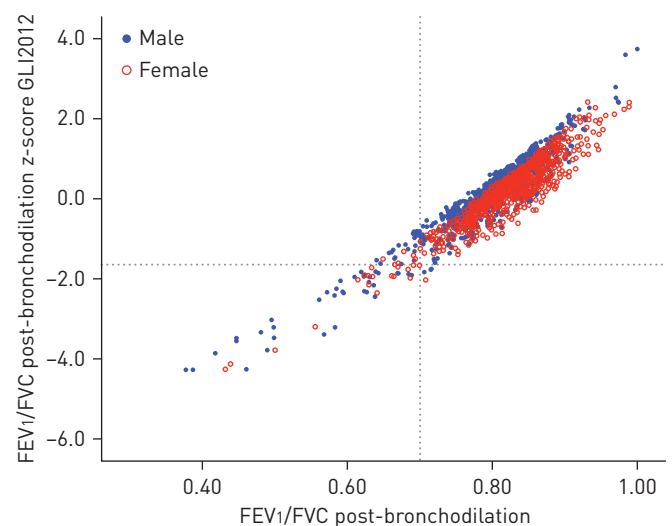
## Results

The distribution of post-bronchodilation FEV<sub>1</sub>/FVC relative to GLI2012 reference values *versus* the absolute FEV<sub>1</sub>/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 FEV<sub>1</sub>/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.

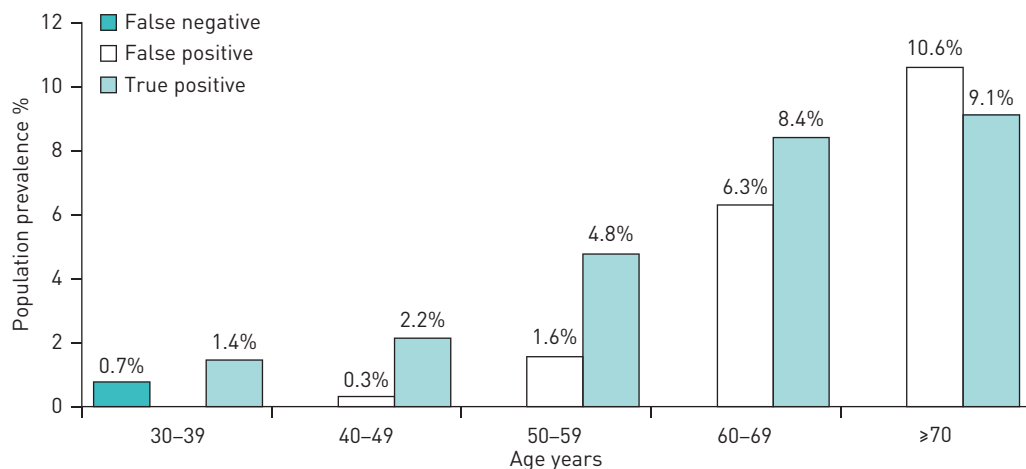
As the GOLD criteria are prone to overdiagnosis of older subjects with normal FEV<sub>1</sub>/FVC based on LLN of reference values, the excess cases are found mostly in the mild category with normal FEV<sub>1</sub>, as shown in table 4. These subjects have both normal FEV<sub>1</sub>/FVC and FEV<sub>1</sub> 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 FEV<sub>1</sub> % pred GLI2012 and the individually calculated FEV<sub>1</sub> 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 (FEV<sub>1</sub>)/forced vital capacity (FVC) relative to Global Lung Function Initiative reference values (GLI2012) [10] *versus* absolute FEV<sub>1</sub>/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 FEV<sub>1</sub>/FVC=0.70 are indicated by dotted lines.

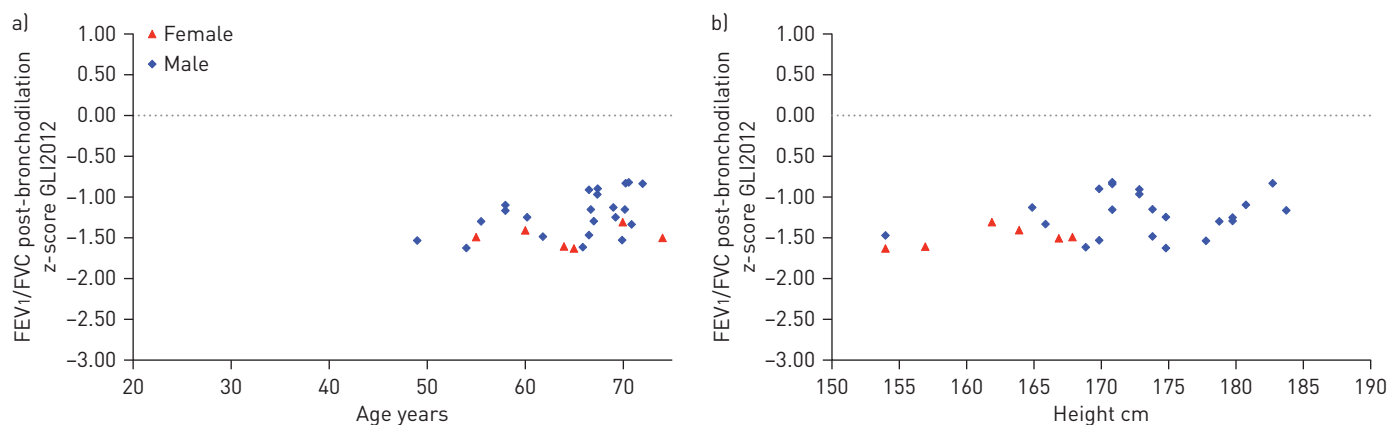


**FIGURE 2** Prevalence of true-positive, false-negative and false-positive cases of chronic obstructive pulmonary disease identified with the fixed-limit Global Initiative for Chronic Obstructive Lung Disease criterion [34] using the Global Lung Function Initiative reference values (GLI2012) [10] as gold standard.

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  $\geq 70$  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 FEV<sub>1</sub>/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  $\geq 65$  years of age with FEV<sub>1</sub>/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 FEV<sub>1</sub>/FVC ratio is directly associated with FEV<sub>1</sub>, which has been shown to be one of the strongest predictors of all-cause morbidity and mortality



**FIGURE 3** Level of forced expiratory volume in 1 s (FEV<sub>1</sub>)/forced vital capacity (FVC) post-bronchodilation as a function of a) age and b) height among subjects (n=28: males n=6; females n=22) with normal FEV<sub>1</sub>/FVC relative to Global Lung Function Initiative reference values (GLI2012) [10] (FEV<sub>1</sub>/FVC  $>$  lower limit of normal) but identified with airflow limitation on Global Initiative for Chronic Obstructive Lung Disease criterion [34] (FEV<sub>1</sub>/FVC  $<$  0.70) in post-bronchodilation spirometry in the study population.

TABLE 3 True-positive, false-negative and false-positive rates in age decades stratified by sex in the study population (n=1323)

Age years	Male				Female			
	True-positive rate	False-negative rate	False-positive rate	Proportion of cases false positive <sup>#</sup>	True-positive rate	False-negative rate	False-positive rate	Proportion of cases false positive <sup>#</sup>
<30	0.0	2.3	0.0	0.0	0.0	0.0	0.0	0.0
30–39	3.3	0.8	0.0	0.0	0.0	0.7	0.0	0.0
40–49	2.0	0.0	0.7	25.0	2.4	0.0	0.0	0.0
50–59	5.2	0.0	3.0	36.4	4.5	0.0	0.6	11.1
60–69	15.7	0.0	11.1	41.4	2.3	0.0	2.3	50.0
≥70	9.1	0.0	15.2	62.5	9.1	0.0	6.1	40.0
All	5.7	0.3	3.7	39.3	2.5	0.1	0.8	25.0

Data are presented as %. <sup>#</sup>: false positive/(true positive+false positive).

even in the absence of lung disease [46–49]. Subjects with a low but still normal FEV<sub>1</sub>/FVC ratio might be less healthy, and they do have lower lung function than their age and gender peers with higher FEV<sub>1</sub>/FVC ratio and normal FEV<sub>1</sub>, but this does not imply causality with the diagnostic criteria for COPD. A reduced FEV<sub>1</sub>/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 FEV<sub>1</sub>/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 FEV<sub>1</sub> % 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

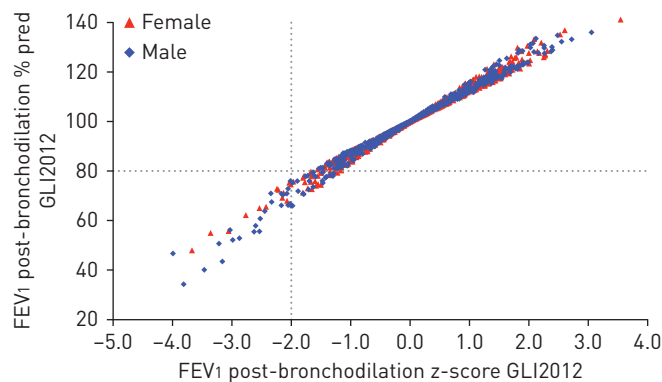
The continued use of FEV<sub>1</sub>/FVC fixed-limit criterion by GOLD guidelines has been advocated with similarities with the fixed limits agreed upon for other physiological measures such as elevated blood pressure [31]. There are, however, fundamental physiological differences in these variables. 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 (FEV<sub>1</sub>)/forced vital capacity (FVC) post-bronchodilation z-score < -1.645 on Global Lung Function Initiative reference values (GLI2012)) or with GOLD fixed-limit criterion (FEV<sub>1</sub>/FVC post-bronchodilation <0.70)

	FEV <sub>1</sub> /FVC <0.70	FEV <sub>1</sub> /FVC <LLN
<b>Subjects</b>	80	55
<b>GOLD grading</b>		
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
<b>z-score-based grading</b>		
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.





**FIGURE 4** Distribution of forced expiratory volume in 1 s (FEV<sub>1</sub>) relative to Global Lung Function Initiative reference values [GLI2012] [10] expressed as FEV<sub>1</sub> % pred versus FEV<sub>1</sub> z-score stratified by sex (males n=596; females n=727). Reference values of 80% of predicted and z-score -2.0 are indicated by dotted lines.

the dependence of the FEV<sub>1</sub>/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 FEV<sub>1</sub>/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 FEV<sub>1</sub>/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 FEV<sub>1</sub>, 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 FEV<sub>1</sub> 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 FEV<sub>1</sub> 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  $-3$  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 (FEV<sub>1</sub>), 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.* FEV<sub>1</sub>/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 FEV<sub>1</sub>/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 FEV<sub>1</sub>/FVC  $<70\%$  as a fixed cut-off: a longitudinal evaluation of clinical and functional outcomes. *Thorax* 2008; 63: 1040–1045.
- 6 Schermer TR, Smeele 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 FEV<sub>1</sub>/FVC ratio reduces the misclassification of airway obstruction. *Thorax* 2008; 63: 1046–1051.

- 8 Miller MR, Quanjer PH, Swanney MP, *et al.* Interpreting lung function data using 80 percent of predicted and fixed thresholds misclassifies over 20% of patients. *Chest* 2011; 139: 52–59.
- 9 Vollmer WM, Gislason P, Burney P, *et al.* Comparison of spirometry criteria for the diagnosis of COPD: results from the BOLD study. *Eur Respir J* 2009; 34: 588–597.
- 10 Quanjer PH, Stanojevic S, Cole TJ, *et al.* Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. *Eur Respir J* 2012; 40: 1324–1343.
- 11 Celli BR, Halbert RJ, Isonaka S, *et al.* Population impact of different definitions of airway obstruction. *Eur Respir J* 2003; 22: 268–273.
- 12 Hankinson JL, Oldenclantz JR, Fedan KB. Spirometric reference values from a sample of the general US population. *Am J Respir Crit Care Med* 1999; 159: 179–187.
- 13 Vanfleteren LEGW, Franssen FME, Wesseling G, *et al.* The prevalence of chronic obstructive pulmonary disease in Maastricht, the Netherlands. *Respir Med* 2012; 106: 871–874.
- 14 Scholes S, Moody A, Mindell JS. Estimating population prevalence of potential airflow obstruction using different spirometric criteria: a pooled cross-sectional analysis of persons aged 40–95 years in England and Wales. *BMJ Open* 2014; 4: e005685.
- 15 Hnizdo E, Glindmeyer HW, Petsonk EL, *et al.* Case definitions for chronic obstructive pulmonary disease. *COPD* 2006; 3: 95–100.
- 16 Mannino DM, Buist AS, Vollmer WM. Chronic obstructive pulmonary disease in the older adult: what defines abnormal lung function? *Thorax* 2007; 62: 237–241.
- 17 Perez-Padilla R, Hallal PC, Vásquez-García JC, *et al.* Impact of bronchodilator use on the prevalence of COPD in population-based samples. *COPD* 2007; 4: 113–120.
- 18 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.
- 19 Roche N, Dalmau F, Perez T, *et al.* FEV<sub>1</sub>/FVC and FEV<sub>1</sub> for the assessment of chronic airflow obstruction in prevalence studies: do prediction equations need revision? *Respir Med* 2008; 102: 1568–1574.
- 20 Quanjer PH, Tammeling GJ, Cotes JE, *et al.* Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. *Eur Respir J* 1993; 6: Suppl. 16, 5–40.
- 21 Falaschetti E, Laiho J, Primates P, *et al.* Prediction equations for normal and low lung function from the Health Survey for England. *Eur Respir J* 2004; 23: 456–463.
- 22 Vaz Fragoso CA, Concato J, McAvay G, *et al.* Chronic obstructive pulmonary disease in older persons: a comparison of two spirometric definitions. *Respir Med* 2010; 104: 1189–1196.
- 23 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.
- 24 Garcia-Río F, Soriano JB, Miravittles M, *et al.* Overdiagnosing subjects with COPD using the 0.7 fixed ratio: correlation with a poor health-related quality of life. *Chest* 2011; 139: 1072–1080.
- 25 Roca J, Burgos F, Sunyer J, *et al.* Reference values for forced spirometry. Group of the European Community Respiratory Health Survey. *Eur Respir J* 1998; 11: 1354–1362.
- 26 Lamprecht B, Schirrhofer L, Kaiser B, *et al.* Subjects with discordant airways obstruction: lost between spirometric definitions of COPD. *Pulm Med* 2011; 2011: 780215.
- 27 Waked M, Khayat G, Salameh P. Chronic obstructive pulmonary disease prevalence in Lebanon: a cross-sectional descriptive study. *Clin Epidemiol* 2011; 3: 315–323.
- 28 Danielsson P, Olafsdóttir IS, Benediktsdóttir B, *et al.* The prevalence of chronic obstructive pulmonary disease in Uppsala, Sweden – the Burden of Obstructive Lung Disease (BOLD) study: cross-sectional population-based study. *Clin Respir J* 2012; 6: 120–127.
- 29 Van Dijk W, Tan W, Li P, *et al.* Clinical relevance of fixed ratio vs lower limit of normal of FEV<sub>1</sub>/FVC in COPD: patient-reported outcomes from the CanCOLD cohort. *Ann Fam Med* 2015; 13: 41–48.
- 30 Turkeshi E, Vaes B, Andreeva E, *et al.* Airflow limitation by the Global Lungs Initiative equations in a cohort of very old adults. *Eur Respir J* 2015; 46: 123–132.
- 31 Di Marco F, Tantucci C, Pellegrino G, *et al.* Chronic obstructive pulmonary disease diagnosis: the simpler the better? Not always. *Eur J Intern Med* 2013; 24: 199–202.
- 32 Van Dijk WD, Gupta N, Tan WC, *et al.* Clinical relevance of diagnosing COPD by fixed ratio or lower limit of normal: a systematic review. *COPD* 2014; 11: 113–120.
- 33 Vestbo J, Hurd SS, Agustí AG, *et al.* Global strategy for the diagnosis, management and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* 2013; 187: 347–365.
- 34 Global Initiative for Chronic Obstructive Lung Disease. Global Strategy for the Diagnosis, Management and Prevention of COPD – 2016. [www.goldcopd.org](http://www.goldcopd.org) Date last accessed: May 31, 2016.
- 35 Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat Med* 1992; 11: 1305–1319.
- 36 Backman H, Lindberg A, Sovijärvi A, *et al.* Evaluation of the global lung function initiative 2012 reference values for spirometry in a Swedish population sample. *BMC Pulm Med* 2015; 15: 26.
- 37 Kotaniemi J-T, Sovijärvi A, Lundbäck B. Chronic obstructive pulmonary disease in Finland: prevalence and risk factors. *COPD* 2005; 3: 331–339.
- 38 Kainu A, Lindqvist A, Sarna S, *et al.* FEV<sub>1</sub> response to bronchodilation in an adult urban population and in healthy adults. *Chest* 2008; 134: 387–393.
- 39 Kotaniemi J-T, Hassi J, Kataja M, *et al.* Does non-responder bias have a significant effect on the results of a postal questionnaire study? *Eur J Epidemiol* 2001; 17: 809–817.
- 40 American Thoracic Society. Standardization of spirometry. 1994 update. *Am J Respir Crit Care Med* 1995; 152: 1107–1136.
- 41 Kanervisto M, Vasankari T, Laitinen T, *et al.* Low socioeconomic status is associated with chronic obstructive airway diseases. *Respir Med* 2011; 105: 1140–1146.
- 42 Quanjer PH, Pretto JJ, Drazzale DJ, *et al.* Grading the severity of airways obstruction: new wine in new bottles. *Eur Respir J* 2014; 43: 505–512.
- 43 GLI Team. Global Lung Function Initiative. [www.ers-education.org/guidelines/global-lung-function-initiative.aspx](http://www.ers-education.org/guidelines/global-lung-function-initiative.aspx) Date last accessed: May 31, 2016.

- 44 Mannino DM, Buist AS, Vollmer WM. Authors' reply to two letters to the editor. *Thorax* 2007; 62: 1108–1109.
- 45 Quanjer PH, Furberg CD. GOLD COPD stage I is not associated with increased risk of death. *Respir Med* 2012; 106: 153.
- 46 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.
- 47 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.
- 48 Stavem 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.
- 49 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.
- 50 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.
- 51 Helldán A, Helakorpi S. Health Behaviour and Health among the Finnish Adult Population, Spring 2014. Helsinki, National Institute for Health and Welfare, 2015.
- 52 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.
- 53 Kainu A, Timonen KL, Toikka J, *et al.* Reference values of spirometry for Finnish adults. *Clin Phys Funct Imaging* 2016; 36: 246–358.