Key points

- Arbitrary differences in the way lung function is expressed and interpreted may result in mismanagement of patients as well as hindering our understanding of the global burden of lung disease.

- Currently, international and regional boundaries, together with individual preferences, may have as much impact on estimates of disease prevalence and treatment decisions as does the true pathophysiological heterogeneity of disease.

- Use of the all-age (3–95 years), multiethnic Global Lung Function Initiative (GLI) spirometry equations, which provide well defined lower limits of normal, will allow global standardisation of how spirometry results are interpreted. This will also avoid errors that have occurred in the past due to overdependence on fixed thresholds to diagnose lung disease or extrapolation of prediction equations in either very young or elderly patients

Image: NASA

Sanja Stanojevic[1],
Philip Quanjer[2],
Martin R. Miller[3],
Janet Stocks[4]

[1]Division of Respiratory Medicine, Hospital for Sick Children, Toronto, Canada
[2]Department of Pulmonary Diseases and Department of Paediatrics, Erasmus Medical Centre, Erasmus University, Rotterdam, The Netherlands
[3]Institute of Occupational and Environmental Medicine, University of Birmingham, Birmingham
[4]Portex Respiratory Unit, Institute of Child Health, University College London, London, United Kingdom

Sanja Stanojevic:
Respiratory Medicine,
Hospital for Sick
Children, Toronto,
Canada

Sanja.stanojevic@sickkids.ca

The Global Lung Function Initiative: dispelling some myths of lung function test interpretation

Educational Aims

- To summarise limitations and implications associated with using outdated spirometry reference equations to interpret lung function.
- To describe the Quanjer *et al.*, 2012 "Global Lung Function Initiative" (GLI) spirometry equations and the advantages of using these in both clinical practice and research studies.
- To discuss the necessary steps and challenges when switching to the GLI, including adjustment for ethnicity, re-calculation of previous results for accurate trend reports and education of both patients and professionals.

Summary

Lung function results can help with establishing a diagnosis, with assessment of treatment effects and with making a prognosis. However, arbitrary differences in the way lung function is expressed and interpreted may result in mismanagement of patients as well as hindering our understanding of the global burden of lung disease. In this article, we summarise the Global Lung Function Initiative spirometry reference equations and dispel some common myths related to the use and interpretation of spirometry results.

HERMES syllabus link module: D.1

Introduction

Every patient with, or suspected of having, a respiratory condition should routinely have pulmonary function tests (PFTs) performed to assess the type and severity of the disease and/or to monitor disease progress and treatment effectiveness. Patients with respiratory disease only account for ~60% of referrals to most PFT laboratories [1]; patients receiving general anaesthetics, transplants and treatment for cancer or other chronic conditions also have lung function measured routinely. PFTs are also used to estimate the global burden of respiratory disease [2]. With such a high volume and demand for this physiological measurement, why is it that the same patient seen in a different country, a different hospital in the same country or even by different physicians in the same hospital on the same occasion, can have markedly different interpretation of lung function results?

Correct interpretation of lung function depends on knowing how big a patient's lungs should be. However, arbitrary differences in the way in which lung function is currently expressed and interpreted [3, 4] not only results in potential confusion and mismanagement of patients, but also hinder our understanding of the global burden of lung disease. Recently, the Global Lung Function Initiative developed a unified and global approach for the interpretation of spirometry results, such that, for the first

time in history, there is the potential to standardise how lung function is interpreted around the world [5].

Global Lung Function Initiative Reference Equations for Spirometry

The objective of the Global Lung Function Initiative (GLI) Task Force (www.lungfunction.org) was to derive ''all-age'' reference equations for spirometry from preschool children to the elderly that covered as many ethnic groups as possible. As a result of unprecedented international cooperation, tens of thousands of spirometric measurements from healthy, non-smoking males and females have been made available by some 70 centres and organisations worldwide. These data were collated and analysed with modern statistical techniques (fig. 1), and have led to the derivation of the Quanjer ''GLI-2012'' prediction equations [5].

The GLI-2012 equations have been endorsed by all major respiratory societies and, for the first time, provide a unified approach to the interpretation of PFTs that has the potential to streamline presentation and interpretation across laboratories, cities and countries worldwide. Within less than a year of publication, the GLI equations have already been validated in several populations [6–8], widely cited in the literature [9–17], and implemented into numerous lung function devices (see www.lungfunction.org/manufacturers.html). While the GLI represent a huge step forward, the lack of awareness and logistical roadblocks for their appropriate implementation can hamper widespread clinical use. In this article we will describe and dispel some common myths related to the presentation and interpretation of lung function results and demonstrate how a unified global approach to interpret spirometry results can benefit patients worldwide.

Myths

Myth 1: our current PFT reports are reliable and appropriate for our local needs

Several jurisdictions (e.g. ATS/ERS, European Drivers' License, ARTP) [18, 19] have produced
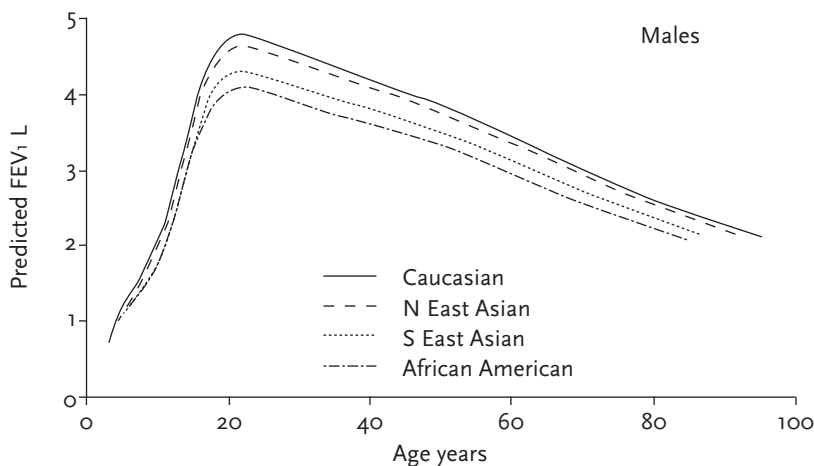

Figure 1

Predicted values for FEV1 in males for the four ethnic groups considered within the GLI. A further equation (''other/mixed'') has also been derived for those of mixed ethnicity or who are not represented by the groups shown above.

evidence-based recommendations to facilitate accurate data collection, quality control and interpretation of spirometry results. The choice of how PFTs are presented and interpreted is, however, largely at the discretion of the individual laboratory or physician. Consequently, in many centres, there remains an enormous gap between these evidence-based guidelines and the actual clinical interpretation of pulmonary function test results [3]. Furthermore, there is generally a naïve trust placed in reports automatically produced by software, with little (if any) thought given as to how tests were performed or how the numbers were derived.

There are several aspects of the PFT report that ought to be questioned.

Are the patient details recorded correctly?

The major determinants of spirometric lung function are height, age, sex and ethnicity [5]. Despite the dependence of predicted values on height, it is not uncommon (particularly in adult centres) for the patient's height to be self-reported rather than measured. Men tend to over-report their actual height [20], and adults tend to ''shrink'' as they age, but often report their height as the maximum achieved during adulthood. Since height is a major determinant of expected lung function, discrepancies in height measurements, including those resulting from a poorly calibrated stadiometer, can lead to misinterpretation of results [11].

> It is recommended that height is measured at each visit to one decimal point using an accurate and regularly calibrated stadiometer.

Age is also an important determinant of lung function throughout the life span, such that accurate documentation (in years to one decimal point) is essential, particularly during childhood when growth and development are so rapid [11]. Current practice in many commercial devices of either truncating or ''rounding'' age to the nearest year, or dependence on self-reported age (rather than that based on difference between date of test and date of birth) can also lead to serious misinterpretation [11].

> It is recommended that age be calculated accurately in years to one decimal point using the patient's date of birth and the date of test.

Finally, PFT reports should also consider a patient's ethnicity. Although it is well established that there are ethnic differences in lung volumes after correcting for age, height and sex [21–24], PFT results in non-Caucasians are typically compared to healthy Caucasian subjects of white European descent, which can lead to significant under-estimation of lung function in such subjects.

> In addition to clearly stating exactly which reference equations have been used, PFT reports should display a patient's ethnic group, as well as the ethnic group of the reference population used to derive the predicted values.

Myth 2: it doesn't really matter which reference equation is used

There are more than 300 published reference equations for spirometry, not to mention the numerous unpublished equations available on PFT equipment. It is important to appreciate that not all reference equations are created equal. Within each spirometer, the user has the option to select a reference equation that they believe is appropriate for the local population. Oftentimes the default equations provided by the manufacturer are never changed and, even if these are changed at the request of the lab manager at time of equipment delivery, it is not uncommon that automatic re-booting of computer systems or installation of new software restores default reference equations without the knowledge of the user.

It is also common practice for reference equations to be stitched together such that a wider age range can be tested without the need for the user to manually switch between equations when testing different patients. Often these prediction sets or modules are not evidence-based but simply derived for convenience, or to meet demands of users who do not fully appreciate the consequences

of developing such prediction modules. As a result, out-dated or inappropriate reference equations are often used to interpret spirometry results, with arbitrary break points between specific age groups (*e.g.* preschool to school-age children, adolescents to adults), which can lead to serious misinterpretation of results [25].

Equally concerning is the lack of transparency to the general user as to which reference equations have been used in any selected module. Commercial devices will sometimes allow extrapolation of prediction equations beyond the age range they were derived for (*e.g.* interpretation of lung function in a 4-year-old from equations derived from children aged above 6– 8 years, extrapolation of European Community for Steel and Coal (ECSC) and the National Health and Nutrition Examination Survey (NHANES) to those aged above 70 and 80 years respectively). Furthermore, within any age range, different spirometric outcomes may be interpreted using entirely different equations, with potentially serious impact on interpretation of the relative sensitivities of different variables.

The potential misinterpretation of spirometry results is greatest during the transition
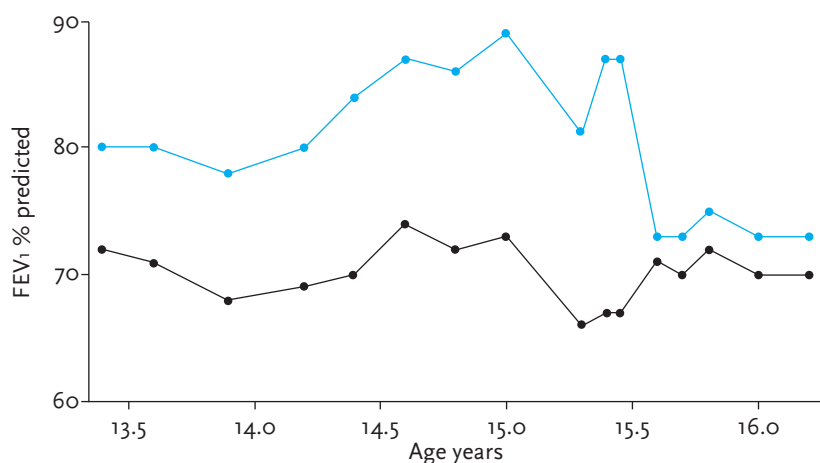
from paediatric to adult care in patients with chronic respiratory conditions [25]. For example, until recently in the UK, it has been recommended that the Rosenthal equations [26] be used during childhood, and the ECSC equations [27] during adulthood. Many laboratories use a prediction module which joins these equations at 18 years of age. In figures 2 and 3 we highlight a few examples where interpretation of results is influenced by the inconsistencies both within and between reference equations from the same ''prediction module''.

Even if the investigator/clinician is aware of exactly which reference equation has been selected, there are numerous factors, including the inclusion/exclusion criteria used to select the reference population, how well-nourished the reference population was, and the age span which will inevitably affect what the predicted value will be at any given age and height. Equations derived from 100 individuals will be far less representative than equations derived from 10,000 individuals [30]. Furthermore, equations that are derived for the entire age range, will be far more robust than separate equations that are artificially stitched together to cover the entire age range.

Awareness of the limitations of outdated reference equations has led to an increasing number of attempts to construct population-specific or even centre-specific reference equations. These are often derived for a population over a limited age range and using a limited sample of patients such that comparison with larger, more rigorously derived equations will inevitably identify differences, all of which may largely be attributed to the phenomenon of sampling variability [11]. Recent evidence suggests that validation of reference equations for local use requires a minimum of 300 subjects (150 males and 150 females) [11].

Figure 2

16 serial measurements between 13 and 16 years of age in a boy with Cystic Fibrosis with height measurements at the 2nd centile [28]. % predicted values according to Rosenthal equations (blue line) [26] are much higher than those calculated from GLI (black line). This is likely due to the fact that the Rosenthal equations do not take age into account, and thus result in marked over-estimation of % predicted in growth restricted children. Furthermore, there was a sudden fall in % predicted results by 14% at around 15.5 years of age when expressed according to Rosenthal even though there had been no change in clinical status and the absolute FEV$_1$ had continued to increase as expected with growth. This was caused by the fact that, when using the Rosenthal equations, this adolescent's height reached a critical break point between two consecutive visits resulting in automatic transfer by the software from the male child's equations (where predicted FEV$_1$ for a height of 162.4 cm was 2.77 L) to the male ''post-pubertal'' reference equation for which the predicted FEV$_1$ for a height of 163.9 cm was 3.43 L [29].

PFT technologists, respiratory clinicians and referring physicians should be aware of which reference equations are used to interpret PFTs, and whether these are appropriate for their patient population, paying particular attention to whether prediction modules are used and the potential impact of any break-points on interpretation of results.

Myth 3: results below 80% predicted are "abnormal"

Clinicians and patients both want to know whether the PFT result falls within the normal or expected range. The interpretation of PFTs thus hinges around knowing what "normal" is. Results from clinical chemistry, haemoglobin, lipids and so forth are compared with a reference range, which summarises measurements made in a group of healthy individuals in the absence of disease. The reference range is derived from the upper and lower values that contain 95% of the healthy reference population.

Similarly, pulmonary function measurements are compared with a group of healthy individuals, although traditionally interpretation has been slightly different from that observed in other disciplines since lung size is strongly associated with body size, dimensions of the thoracic cavity, sex and age. During childhood and adolescence, growth is particularly rapid with lung function increasing 20-fold during the first 10 years of life [31]. Furthermore, during childhood, forced vital capacity (FVC) outgrows forced expiratory volume in 1 second (FEV1), leading to falls in FEV1/FVC; these trends are reversed during adolescence [32]. The correct description of the range of normal values requires that these physiological factors are taken into account. This has rarely been possible when analysing relatively small datasets over limited age ranges or when using simple linear regression models and is responsible for many of the discrepancies seen when comparing results from children and adolescents that have been interpreted according to GLI-2012 with those obtained from previously published equations (see figs 2 and 3).

Respiratory clinicians, technicians and their patients are accustomed to the practice of expressing and interpreting PFT results as % predicted without realising the inherent errors in this methodology. The % predicted is calculated by taking the observed measurement (absolute values of FEV1 and FVC in L) and dividing it by a predicted value multiplied by 100 (% predicted = (observed/predicted) × 100). The predicted values are obtained from a group of healthy individuals, such that 100% predicted reflects the average value expected in a healthy individual of any given size, sex and age. Although 80% predicted is commonly used as the cut-off for identifying abnormal results, the use of this
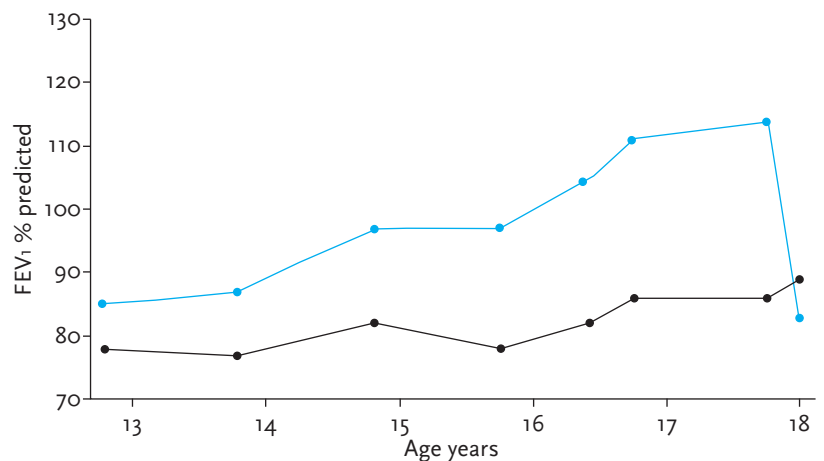

Figure 3

Eight serial measurements are presented from a white Caucasian male below the 2nd centile for height [28] who was tested between his 12th and 18th birthday. There is increasing discrepancy over time between the GLI (black line) and Rosenthal equations (blue line) [26]. The difference between the two equations is as large as 25% at 17 years. At 18 years, the software automatically switches from Rosenthal to the ECSC equations and the % predicted value plummets from 114% to 83%. By contrast when the GLI equations are used, the patient's results tracked seamlessly across the pubertal period.

fixed cut-off comes from unsubstantiated claims in the 1960s that suggested this was a good "rule of thumb" [33]. The major limitation of using % predicted (and 80% predicted as a fixed threshold) is that it does not take into account the fact that the natural variability of spirometry outcomes in health is highly age and outcome dependent. The practical implication being that the "normal range" for FVC or FEV1 is considerably wider than the frequently quoted "80–120% predicted" both for young children and for subjects older than 30 years [5, 34]. This leads to a high percentage of false-positive findings particularly in the elderly (fig. 4a).

The use of the fixed cut-off for the FEV1/FVC ratio has similar consequences. Several large population studies have shown that the FEV1/FVC has a strong negative agedependency, such that the frequently used fixed threshold of 0.7 for FEV1/FVC isnot attained until about 50 years of age inmen and later in women. Consequently, airway obstruction in younger subjects is missed [5, 34–38], whereas the 0.7 cut-off falsely identifies a large number of olderhealthy subjects as having lung disease (fig. 4b) [10, 39–41].

Myth 4: I can't understand Z-scores or SRS or explain them to my patient

For many years, both the ERS and ATS have recommended that results are presented and
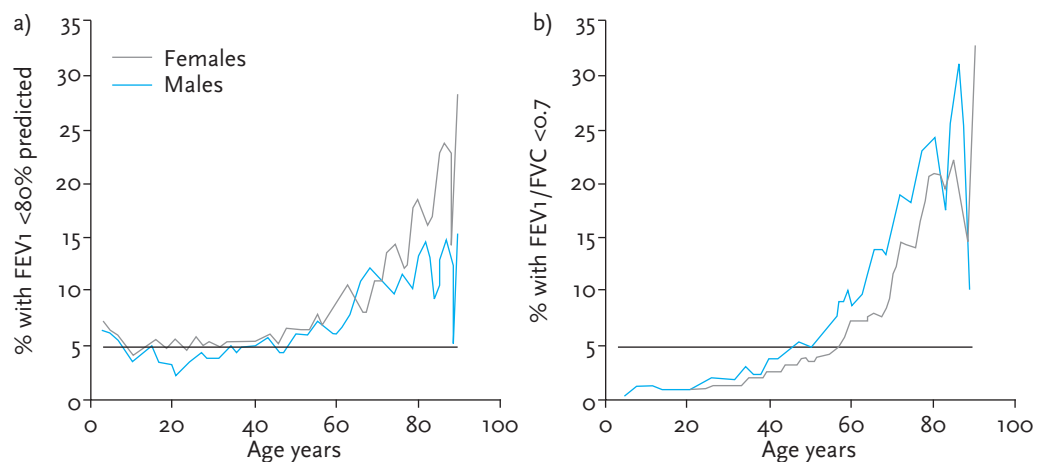
Figure 4
The percentage of healthy subjects in whom FEV1 is less than 80% predicted (left), or in whom the FEV1/FVC ratio is less than 0.7 (right) [5]. As can be seen, use of 80% as a fixed threshold for FEV1 leads to a high percentage of false positives in the elderly, while use of <0.7 as the threshold for abnormal FEV1/FVC will lead to under-diagnosis of airway obstruction in the young and over-estimation in the elderly.

interpreted in the context of the normal range wherein the lower limit of normal (LLN) is age specific. The normal range can be represented as a pictogram (fig. 5), or as Z-scores (otherwise known as standardised residual scores or SRS) in which 95% of healthy subjects will have Z-scores (or SRS) values within ± 2 z-scores, and 90% within ± 1.64 z-scores. The Z-score indicates how many standard deviations a measured value is from predicted ((observed-predicted)/standard deviation). Although it is customary to classify the severity of lung function impairment using FEV1 as a percentage of predicted (ATS/ERS 2005) [18], the use of Z-scores removes the age-related bias [15]. Further work is necessary to better understand what Z-score values represent clinically meaningful outcomes, as was done for % predicted FEV1 in the past.

Myth 5: the Z-score cannot replace the clinically established % predicted

The use of % predicted is associated with age- and height-related bias and should therefore be abandoned. A valid alternative method of reporting lung function is to express results as Z-scores. Use of Z-scores solves many potential problems by taking into account age, height, sex and ethnic group, as well as the age-dependent reference range. Unlike % predicted, it is therefore free of any bias. An obvious advantage is that any given Z-score indicates comparable lung function between individuals, irrespective of their sex, height, age or ethnicity. The Z-score also facilitates bias-free interpretation of serial measurements within a person during growth and ageing, and direct comparison between different lung function outcomes.

Because lung function tests are not applied indiscriminately to the population, when patients have symptoms or known risk factors (*e.g.* smoking history), it is usual practice to use the -1.64 Z-scores cut-off to identify subjects outside the normal range. If lung function tests are used for untargeted screening, then -1.96 Z-scores should be used. Both these cut-offs indicate the probability of a false positive (5% and 2.5% respectively for -1.64 and -1.96 Z-scores), or the proportion of completely normal subjects with values below these cut-offs.

When interpreting results, it is important to remember that there will always be a degree of within-person variability, so that by chance a measurement may be just outside the normal range on one occasion, but just within it on the next. It is also essential to take other clinical information into account, and to weigh the consequences of an erroneous false positive against that of a missed diagnosis. Particular caution is required when interpreting results which lie close to the somewhat arbitrary cut offs between health and suspected disease, especially when results are limited to a single test occasion.

However, both % predicted and Z-scores have limitations when classifying the severity of obstructive lung disease or defining prognosis. This can be illustrated as follows. Both indices indicate how far a measurement is from an average value in a healthy subject; however, mortality relates more directly to how close a measured value is from the minimum that is compatible with life. For males and females aged >50 years this is an FEV_1 of ~500 and 400 mL, respectively [42, 43]. At age 50 years, for a man of average height this represents 13.2% predicted (Z-score -5.77), for a woman (165 cm) 14.0% (Z-score -5.92). At age 85 years, corresponding findings are 18.9% (Z-score -3.73) and 21.4% (Z-score -3.88), respectively. Hence, in clinical assessment both the Z-score and remaining ventilatory reserves needs to be taken into account. Further investigations to define disease severity and how this translates to Z-scores using the GLI are necessary.

A lung function test must never be used in isolation to define disease severity and prognosis; a number of factors, including quality of life, are likely to contribute, and the ideal approach remains to be determined. Neither % predicted nor Z-scores used in isolation can answer those fundamental questions.

Figure 5
Illustration of the normal distribution and corresponding Z-scores and percentiles. The pictogram (horizontal bars) demonstrates the normal range (white region) with arrows indicating how far from the normal range an observation is (Z-scores) [5]. The 50th centile (0 Z-scores) is equivalent to 100% predicted.

Myth 6: I don't need to worry about reference equations as I only look at the absolute values

Physicians that care for patients with chronic respiratory conditions are particularly interested in identifying changes and deteriorations over time that are outside the normal variability for an individual patient. When serial measures of absolute values of lung function (*i.e.* FEV_1 and FVC in L) are used to track progress in young and middle-aged adults, these can be referred internally to the subject's own "best" values and interpreted without the need for any reference equation. This approach cannot, however, be used when attempting to compare results between patients or centres. Even within an individual, interpretation of change becomes much more complicated during periods of growth or ageing. Under such circumstances, it will be necessary to express results in relation to some accepted reference.
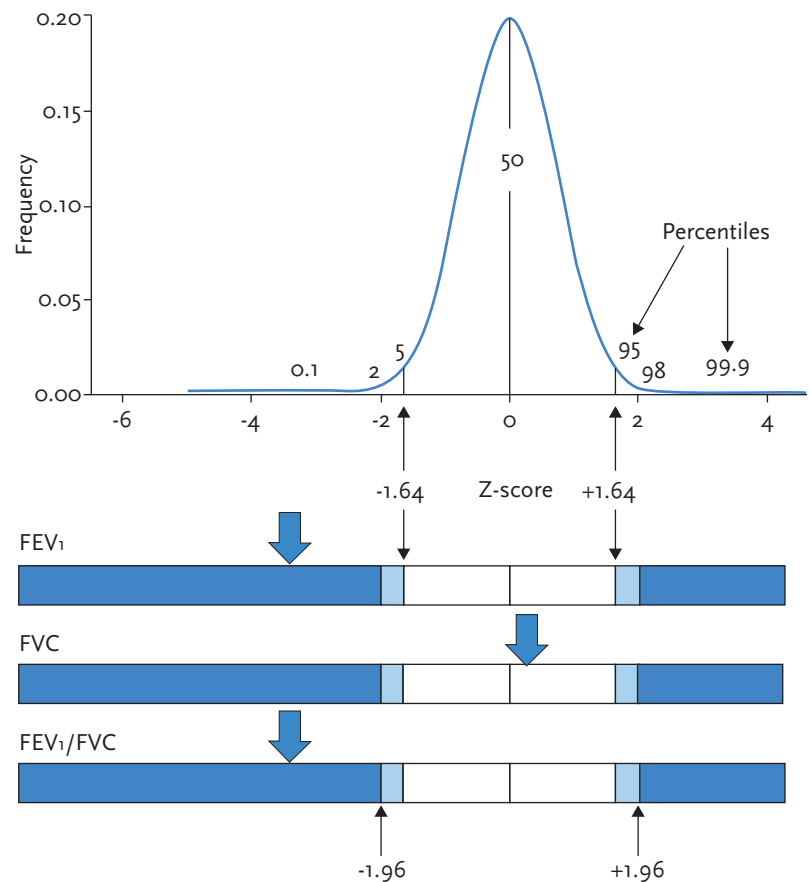
Myth 7: if I follow patients over time, it doesn't matter which reference equation is used, as long as it is the same one

Accurate interpretation of how much change can be attributed to disease progression or response to therapy needs to be made in context of how much change is likely to occur as a result of within-subject, between-test variability, as well as in relation to the changes attributed to the natural process of growth, development and ageing.

In contrast to popular belief, interpretation of changes in lung function depends markedly on which reference equation is selected. In contrast to the GLI, older reference equations were often derived using simple linear regression based on relatively few subjects over a limited age range, with age not being included as a determinant in many of the paediatric equations still in common use. Consequently, the predicted value at any given age can vary markedly
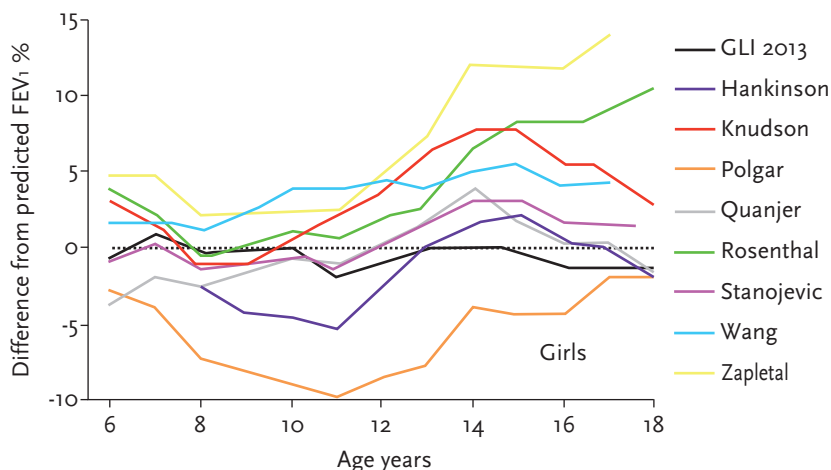
Figure 6
Differences from GLI in predicted values compared to observed values for 9 different reference equations used for children [11].

between and within equations, with marked subsequent impact on the way in which a subject's results will be interpreted either on a single occasion or over time (see Myth 2 and figure 6).

In practice, the interpretation of serial measurements is further complicated by the fact that, with the exception of the GLI-2012 equations, no spirometric reference equation encompasses the entire age span. Consequently when following a patient over time, results are interpreted either by extrapolating results beyond the age they were derived for (a practice that should be discouraged strongly due to the major errors it can introduce) or by stitching reference equations derived from different populations together as part of "prediction modules" (see Myth 2). With increasing use of spirometry in preschool children as young as 3 years of age [44] these issues are now of greater relevance and can also lead to serious misinterpretation particularly when patients transition from paediatric to adult care (fig. 3) [25].

> Most commercial PFT devices facilitate the re-calculation of trend reports based on a single reference. Therefore the transition to GLI equations should be accompanied with a re-calculation of trend reports to facilitate interpretation of results in context of previous results.

It is also not uncommon for patients to acquire disease and suffer from a significant reduction in lung function, even though their lung function values remain within the normal range. Understanding the subject's baseline lung function is critical for correct interpretation of the changes observed due to disease. When tracking a patient over time or monitoring treatments, it is important to consider the patient's "best lung function" as a reference point and how much change has been observed. A major limitation of any reference equation is that it is based on a cross-sectional snap-shot of a population comprised of individuals at different ages, which does not reflect how individual patients change over time.

Myth 8: the GLI might make patients' results that were normal abnormal and *vice versa*. How can we explain this to the patient/their doctor?

For reasons discussed above, there will inevitably be some changes in interpretation of results when they are re-analysed using GLI 2012, and this will need to be dealt with sensitively when discussing with the patient and their family [10]. When examined in clinical populations, there were minor, clinically unimportant changes in the prevalence rate of airways obstruction upon adopting the GLI equations, whereas the prevalence rate of a "spirometric restrictive pattern" (low FVC but no obstruction) was higher [6, 9, 10]. As a matter of principle, the fact that new improved methods of interpreting results have been introduced should be discussed *prior* to any patient receiving a copy of results based on the new equations to allay any anxiety, accompanied by simple explanatory leaflets. Similar communication will be necessary with all relevant health professionals in contact with the patient. Appropriate educational material is being developed currently and will be made widely available free of charge in the near future.

In the meantime, it is important to remember that no reference equation is capable of determining whether a patient has normal or abnormal lung function in isolation (see Myth 5). Reference equations (and the lower limit of normal) give an indication of how an individual patient's result compares to that of a group of healthy individuals. The further away a result lies from the normal range, the more certain we can be that the results are not compatible with health. Thus results from a patient whose results fall on or near the lower limit

of normal and who has evidence of respiratory symptoms and other clinical indications of disease should be interpreted quite differently from an otherwise healthy individual with similar predicted values but no apparent signs or symptoms of disease. By definition, when using a LLN of -1.64 Z-scores or (the 5th centile) 5% of healthy individuals (one in 20 PFTs) will have results below the lower limit of normal (fig. 5); therefore simply relying on a number to indicate whether lung function is normal or abnormal will inevitably lead to a misdiagnosis. One should treat the patient, not the numbers.

> When new improved methods of interpreting results are introduced, these should be discussed prior to any patient receiving a copy of results based on the new equations to allay any anxiety, accompanied by simple explanatory leaflets.

Myth 9: It is better to wait until the GLI equations have included ALL ethnic groups before we switch.

The GLI represent a huge step forward by providing a unified approach to interpreting lung function in several ethnic groups. Importantly the GLI analyses were able to show that a wide range of countries and ethnicities have lung function that is consistent with a white "Caucasian", minimising the need for multiple reference equations where the differences are not physiologically or clinically meaningful [45]. Second, the ethnic differences identified were proportional such that the general growth and decline of lung function was systematic across ethnic groups.

While the GLI do include several ethnic groups, they are by no means comprehensive of all ethnic groups. Notably missing from the GLI are data from the African continent, South Asia (Indian sub-continent) and Latin America. Several on-going efforts are underway to collect data from these regions for the next instalment of the GLI in the future [7]. Since ethnic differences are proportional, interim ethnic-specific correction factors can be derived for new ethnic groups currently not represented within the GLI.

However, even these efforts pale in comparison to the breadth of ethnic diversity represented in the world. The growing number of migrants and bi-racial children pose two significant limitations to any attempt to develop reference equations for every ethnic group. Migration (both within and between countries), adoption of Western lifestyles and secular changes in anthropometric characteristics in resource poor nations will all need to be investigated in the future.

Myth 10: it is better to wait until GLI equations are available for all pulmonary function tests before we switch

Whenever reference equations are developed, the outcomes included and age range encompassed is usually at the discretion of the investigators conducting the study. No set of reference equation contains all possible spirometric outcomes, and very few measure multiple outcomes (e.g. from spirometry, plethysmography and gas transfer) on the same subjects. It is therefore unrealistic to hope that a single reference population for multiple pulmonary function tests, over the entire age range and for multiple ethnic groups will ever be a reality. Consequently, reports which present predicted results for a range of PFTs generally include predicted values from multiple prediction equations derived from different healthy populations. Furthermore, many PFT devices allow for prediction modules to be developed, such that the user only needs to select one reference, although in fact this represents multiple reference equations arbitrarily stitched together. Of course, if the predicted spirometric FVC is derived from one reference population and the plethysmographic VC from a different population, then it is highly unlikely that these two values will match. Rather than producing all-inclusive prediction sets that are, at best, arbitrary, efforts should be put towards investing in appropriate studies that collect high quality data that will further help to improve the interpretation of PFTs. For this reason, a GLI working party is currently trying to establish predicted values for transfer factor.

Currently, results from each PFT need to be interpreted independently, and in the context of the patient. While report of a FVC value within the normal range and VC value outside the normal range does present a conundrum, reference equations and presentation of results in relation to what is expected in a subject of that age, height, sex and ethnicity is only one indication of a patient's health status. These values should

be interpreted in the context of the patient's other clinical findings. Until a suitable alternative to these discrepancies becomes available, the interpretation of discordant results, such as these, requires careful clinical judgement, rather than inappropriate application of out-dated reference equations.

When switching to GLI-2012 equations for spirometry, laboratories should investigate which reference equations are being used for other PFTs and decide whether these remain the most appropriate for their patient population at the present time. There are obvious limitations to using reference equations for different PFTs derived from different populations, and these should be considered when interpreting discordant findings from different PFTs.

Myth 11: we cannot use GLI because it does not include our favourite outcomes

While it is possible to extract hundreds of outcomes from a spirometry test, and almost all commercial devices will allow you to display your favourite selection of these on reports, the clinical utility of these additional outcomes is highly questionable [46]. Similarly, the reference populations used to derive the predicted values for many of these outcomes should be examined more closely before they are used as a basis for any clinical decisions.

The GLI spirometry equations were developed to include physiologically relevant outcomes for which there were sufficient data to develop robust reference equations. The GLI focussed on outcomes recommended by the ATS/ERS guidelines (*i.e.* FEV_1, FVC and FEV_1/FVC) [47], and thus do not include all possible outcomes, although the facility to interpret $FEV_{0.75}$ in younger subjects has been included for white children (data on this outcome not being available for other ethnic groups) together with forced expiratory flow between 25 and 75% of FVC (FEF_{25-75}) and at 75% of FVC (FEF_{75}) for various ethnic groups.

It is inadvisable to include predicted values for spirometry outcomes not included in the GLI from external reference equations in the form of prediction sets.

Myth 12: it is impossible to interpret results easily using GLI because it does not present a typical reference flow–volume loop with which to compare my patient's results

Many physicians are familiar with the presentation of an ''expected'' flow–volume loop superimposed on top of a patient's flow–volume loop commonly presented on reports. To create this it is necessary to have accurate recordings of peak expiratory flow and other flow outcomes; data that were not available in many of the spirometric datasets included in GLI. Furthermore, these ''typical curves'' on which clinicians place so much reliance when interpreting their patient's results, are fundamentally flawed. Despite popular belief, flow–volume curves do not actually contain any information that is not available from FEV_1, FVC and FEV_1/FVC. The myth that flows detect small airways disease which goes undetected when using conventional spirometric indices arose because flows were expressed as % predicted, using inappropriate prediction equations and disregarding the true LLN. The ''expected'' flow–volume loop is drawn from predicted values from a mixture of sources, which do not always correspond with the predicted values and % predicted values presented on the report. In addition the typical curves are inappropriate for children. Generally, whenever FEV_1, FVC and FEV_1/FVC are within the normal range, so are $FEF_{25-75\%}$ and $FEF_{75\%}$, making flow indices redundant [46].

Understanding whether a patient's results are within the normal range, and the nature and severity of any abnormality is much better assessed using a pictogram (fig. 5), which clearly delineates the range of values compatible with normal and how an individual's results compare to this range. Assessment of the shape of the flow–volume loop should be done independently of any predicted diagram and ideally in the context of pre–post bronchodilator results.

Myth 13: changing to GLI will just confuse everyone

The reality of the current situation is that most people are confused when it comes to using reference equations, though few are aware of the magnitude of error that can occur by misusing such equations. Bearing in mind that a large proportion of pulmonary function laboratories [3, 4, 48], let alone general practice

offices, are not aware of which reference equations are being used to interpret the results, ignorance or fear of the effects of making the change to GLI-2012 is not an acceptable reason to continue the misinterpretation of PFT results. Patients deserve more from the respiratory community. As highlighted above, the key to successful transition lies in appropriate education and ensuring that, once the switch is made, all previous results from each patient are re-analysed using the same standards, so that reliable trend reports are obtained.

Conclusions

The use of inappropriate reference equations, and misinterpretation even when using potentially appropriate equations, can lead to serious errors in both under- and over-diagnosis, with the associated burden in terms of financial and human costs. It is no longer acceptable to continue applying blind faith in results produced by equipment when interpreting lung function tests. Any algorithm can produce a result, but that does not mean it is the correct result. Having endorsed these equations, national and international respiratory societies now need to follow the example of the Association of Respiratory Technology and Physiology (ARTP) and formally recommend usage of the GLI equations to all their members. Similarly, researchers, clinicians and technicians are all equally responsible for enforcing pressure to request the latest ATS/ERS recommendations are available for the interpretation of patient results. Finally, manufacturers of lung function testing equipment need to be proactive and actively participate in facilitating the switch to GLI.

Suggested reading:

- A review of how reference equations are derived and used: Stanojevic S, Wade A, Stocks J. Reference values for lung function: past, present and future. *Eur Respir J* 2010; 36: 12–19.
- A detailed description of how the GLI-2012 were derived: 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.
- A comparison of several commonly used reference equations, and how interpretation changes when the GLI-2012 are applied: Quanjer PH, Brazzale DJ, Boros PW, *et al*. Implications of adopting the Global Lungs Initiative 2012 all-age reference equations for spirometry. *Eur Respir J* 2013; 42: 1046–1054.
- Evidence supporting the misuse of fix cut-offs to interpret spirometery results: Miller MR, Quanjer PH, Swanney MP, *et al*. Interpreting lung function data using 80% predicted and fixed thresholds misclassifies more than 20% of patients. *Chest* 2011; 139: 52–59.
- A summary of the clinical benefits of adopting the GLI-2012: Swanney MP, Miller MR. Adopting universal lung function reference equations. *Eur Respir J* 2013; 42: 901–903.

References

1. Pretto JJ, Brazzale DJ, Guy PA, *et al*. Reasons for referral for pulmonary function testing: an audit of 4 adult lung function laboratories. *Respir Care* 2013; 58: 507–510.
2. Buist AS, McBurnie MA, Vollmer WM, *et al*. International variation in the prevalence of COPD (the BOLD Study): a population-based prevalence study. *Lancet* 2007; 370: 741–750.
3. Mohanka MR, McCarthy K, Xu M, *et al*. A survey of practices of pulmonary function interpretation in laboratories in Northeast Ohio. *Chest* 2012; 141: 1040–1046.
4. Pattishall EN. Pulmonary function testing reference values and interpretations in pediatric training programs. *Pediatrics* 1990; 85: 768–773.
5. 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.
6. Quanjer PH, Weiner DJ. Interpretative consequences of adopting the global lungs 2012 reference equations for spirometry for children and adolescents. *Pediatric Pulmonology* 2013; [In press; doi: 10.1002/ppul.2287].
7. Bonner R, Lum S, Stocks J, *et al*. Applicability of the global lung function spirometry equations in contemporary multiethnic children. *Am J Respir Crit Care Med* 2013; 188: 515–516.
8. Hall GL, Thompson BR, Stanojevic S, *et al*. The Global Lung Initiative 2012 reference values reflect

contemporary Australasian spirometry. *Respirology* 2012; 17: 1150–1151.

9. Brazzale DJ, Hall GL, Pretto JJ. Effects of Adopting the New Global Lung Function Initiative 2012 Reference Equations on the Interpretation of Spirometry. *Respiration* 2013; 86: 183–189.

10. Quanjer PH, Brazzale DJ, Boros PW, *et al.* Implications of adopting the Global Lungs Initiative 2012 all-age reference equations for spirometry. *Eur Respir J* 2013; 42: 1046–1054.

11. Quanjer PH, Hall GL, Stanojevic S, *et al.* Age- and height-based prediction bias in spirometry reference equations. *Eur Respir J* 2012; 40: 190–197.

12. Vaz Fragoso CA, McAvay G, Gill TM, *et al.* Ethnic differences in respiratory impairment. *Thorax* 2013; [In press; doi: 10.1136/thoraxjnl-2013-203631].

13. Schultz ES, Gruzieva O, Bellander T, *et al.* Traffic-related air pollution and lung function in children at 8 years of age: a birth cohort study. *Am J Respir Crit Care Med* 2012; 186: 1286–1291.

14. Coates AL, Graham BL, McFadden RG, *et al.* Spirometry in primary care. *Can Respir J* 2013; 20: 13–21.

15. Quanjer PH, Pretto JJ, Brazzale DJ, *et al.* Grading the severity of airways obstruction: new wine in new bottles. *Eur Respir J* 2013; [In press; doi: 10.1183/09031936.00086313].

16. Vaz Fragoso CA, Gill TM, McAvay G, *et al.* Respiratory impairment in older persons: when less means more. *Am J Med* 2013; 126: 49–57.

17. Vandevanter DR, Pasta DJ. Evidence of diminished FEV and FVC in 6-year-olds followed in the European cystic fibrosis patient registry, 2007–2009. *J Cyst Fibros* 2013; [In press; doi: 10.1016/j.jcf.2013.04.005].

18. Pellegrino R, Viegi G, Brusasco V, *et al.* Interpretative strategies for lung function tests. *Eur Respir J* 2005; 26: 948–968.

19. Steenbruggen I, Mitchell S, Severin T, *et al.* Harmonising spirometry education with HERMES: training a new generation of qualified spirometry practitioners across Europe. *Eur Respir J* 2011; 37: 479–481.

20. Merrill RM, Richardson JS. Validity of self-reported height, weight, and body mass index: findings from the National Health and Nutrition Examination Survey, 2001–2006. *Prev Chronic Dis* 2009; 6: A121.

21. Greenough A, Hird MF, Everett L, *et al.* Importance of using lung function regression equations appropriate for ethnic origin. *Pediatr Pulmonol* 1991; 11: 207–211.

22. Harik-Khan RI, Muller DC, Wise RA. Racial difference in lung function in African-American and White children: effect of anthropometric, socioeconomic, nutritional, and environmental factors. *Am J Epidemiol* 2004; 160: 893–900.

23. Ip MS, Karlberg EM, Karlberg JP, *et al.* Lung function reference values in Chinese children and adolescents in Hong Kong. I. Spirometric values and comparison with other populations. *Am J Respir Crit Care Med* 2000; 162: 424–429.

24. Wang X, Dockery DW, Wypij D, *et al.* Pulmonary function between 6 and 18 years of age. *Pediatr Pulmonol* 1993; 15: 75–88.

25. Kirkby J, Aurora P, Spencer H, *et al.* Stitching and switching: the impact of discontinuous lung function reference equations. *Eur Respir J* 2012; 39: 1256–1257.

26. Rosenthal M, Bain SH, Cramer D, *et al.* Lung function in white children aged 4 to 19 years: I--Spirometry. *Thorax* 1993; 48: 794–802.

27. 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.

28. Cole TJ, Freeman JV, Preece MA. British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. *Stat Med* 1998; 17: 407–429.

29. Rees S, Kirkby J, Fettes E, *et al.* Appropriate interpretation of longitudinal spirometry assessments. *Eur Respir J* 2013; Suppl. 348s – 349s.

30. Quanjer PH, Stocks J, Cole TJ, *et al.* Influence of secular trends and sample size on reference equations for lung function tests. *Eur Respir J* 2010; 37: 658–664.

31. Quanjer PH, Stocks J, Polgar G, *et al.* Compilation of reference values for lung function measurements in children. *Eur Respir J* 1989; 4: Suppl. 184S–261S.

32. Quanjer PH, Stanojevic S, Stocks J, *et al.* Changes in the FEV/FVC ratio during childhood and adolescence: an intercontinental study. *Eur Respir J* 2010; 36: 1391–1399.

33. Bates DV, Christie RV, eds. Respiratory Function in Disease. Philadelphia, WB Saunders, 1964.

34. 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.

35. Brändli O, Schindler C, Künzli N, *et al.* Lung function in healthy never smoking adults: reference values and lower limits of normal of a Swiss population. *Thorax* 1996; 51: 277–283.

36. Falaschetti E, Laiho J, Primatesta P, *et al.* Prediction equations for normal and low lung function from the Health Survey for England. *Eur Respir J* 2004; 23: 456–463.

37. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med* 1999; 159: 179–187.

38. Kuster SP, Kuster D, Schindler C, *et al.* Reference equations for lung function screening of healthy never-smoking adults aged 18-80 years. *Eur Respir J* 2008; 31: 860–868.

39. Miller MR, Quanjer PH, Swanney MP, *et al.* Interpreting lung function data using 80% predicted and fixed thresholds misclassifies more than 20% of patients. *Chest* 2011; 139: 52–59.

40. Quanjer PH, Enright PL, Miller MR, *et al.* The need to change the method for defining mild airway obstruction. *Eur Respir J* 2011; 37: 720–722.

41. Swanney MP, Miller MR. Adopting universal lung function reference equations. *Eur Respir J* 2013; 42: 901–903.

42. Miller MR, Pedersen OF. New concepts for expressing forced expiratory volume in 1 s arising from survival analysis. *Eur Respir J* 2010; 35: 873–882.

43. Ward H, Cooper B, Miller MR. Validation of lung function prediction equations from patient survival data. *Eur Respir J* 2012; 39: 1181–1187.

44. Beydon N, Davis SD, Lombardi E, *et al.* An official American Thoracic Society/European Respiratory Society statement: pulmonary function testing in preschool children. *Am J Respir Crit Care Med* 2007; 175: 1304–1345.

45. Kiefer EM, Hankinson JL, Barr RG. Similar relation of age and height to lung function among Whites, African Americans, and Hispanics. *Am J Epidemiol* 2011; 173: 376–387.

46. Quanjer PH, Weiner DJ, Pretto JJ, *et al.* Measurement of FEF25–75% and FEF75% does not contribute to clinical decision making. *Eur Respir J* 2013; [In press; DOI: 10.1183/09031936.00128113].

47. Miller MR, Hankinson J, Brusasco V, *et al.* Standardisation of spirometry. *Eur Respir J* 2005; 26: 319–338.

48. Basson E, Stewart RI. The standard of spirometry in the RSA. *S Afr Med J* 1991; 79: 361–363.