Work in progress not for citation or quotation

Paper presented at the HESG conference, Glasgow, June 2004

A comparison of patient and population values for health states in varicose veins patients

Julie Ratcliffe[1], John Brazier[1], Simon Palfreyman[2], Jonathan Michaels[2]

[1] Sheffield Health Economics Group, University of Sheffield

[2] Sheffield Vascular Institute, Northern General Hospital, Sheffield

Corresponding Author:

Julie Ratcliffe

Sheffield Health Economics Group

ScHARR

University of Sheffield

Regent Court

30 Regent St

Sheffield

S1 4DA

Email: j.ratcliffe@sheffield.ac.uk

Tel: 0114 222 0790

Acknowledgements
This work was carried out as part of the REACTIV (randomised and economic assessment of conservative and therapeutic interventions for varicose veins) clinical trial and we would like to acknowledge the contribution of Professor Bruce Campbell, Jackie MacIntrye and Katherine Rigby. The financial support of the Health Technology Assessment Programme is also gratefully acknowledged.

1. Introduction

Within the last two decades economic evaluation has become increasingly important as a tool to aid decision-makers concerning the allocation of scarce resources within health care. As a consequence of its increased prominence in this area, several authorative guidelines for the conduct of economic evaluations in health care have been produced in several countries including the USA [The Washington Panel on Cost Effectiveness, Gold et al, 1996], Canada [Canadian Coordinating Office for Health Technology Assessment, 1997], Australia [Commonwealth of Australia, 1995] and the UK [National Institute for Clinical Excellence 2001 and 2003]. A central component all of the guidelines concerns the methods which should be used to value the benefits of health care and whose values should be incorporated into the measurement of benefits. The guidelines reflect a consensus that the preferred methodology is cost utility analysis (CUA) whereby the benefits of health care interventions are measured according to quality adjusted life years (QALYs). In addition, there is general agreement that a generic preference based measure of health status based upon general population values should be used to calculate QALYs.

The main generic preference based measures of health status include the EQ-5D [Dolan, 1997], the SF-6D [Brazier et al 2002] and the HUI-3 [Feeny et al, 2002]. These instruments are most typically used to elicit responses from patients concerning their health status at defined time intervals. These responses are then converted into values using a pre-determined tariff system which is based upon the values placed upon particular health states by a representative sample of the general population. Such valuation systems have been termed indirect as the value associated with the health state is not obtained directly from the individual/s whose health status is being assessed [Stavem, 1996]. One of the major perceived advantages of such generic measures of health status reflects their ability to provide ‘off the shelf’ values for a wide variety of generic health states. In principle, these values can then be used in any CUA without the requirement for collection of primary data whilst simultaneously providing a mechanism to aid the standardisation of analyses [Gold et al, 1996].

The main argument for using general population (as opposed to patient) values for health states rests upon the premise that in a publically funded health care system the main objective of the health care system is a societal one, namely to maximise health (subject to equity concerns). The social perspective can also be supported on the basis that it is the general public who provide the finance for health care through general taxation [Torrance et al, 1986]. In the context of the theory and practice of willingness to pay, it has been suggested that resource allocation decisions in health care are akin to the decisions about which services should be provided by an insurance package [Gafni, 1991]. Accordingly, public funding for health care can be viewed as a type of public insurance and hence ex ante general population preferences should be used to value health states.

Alternatively, it can be argued that patients themselves should be asked to value their own health state on the basis that, unlike members of the general population, they have first hand experience of the state being valued. Buckingham [1993] summarises the advantages of this approach:

'To ask a person of 20 how s/he will value health at the age of 70 is to ask an enormous amount of their imagination. To ask a 70 year old how important their health is to them is likely to result in far more valuable information'.

There is an increasing amount of evidence to suggest that the choice of whose values to elicit is important as it may influence the resulting values. Patients tend to place higher values on dysfunctional health states than do members of the general population [Sackett and Torrance, 1978; Boyd et al, 1990; Hurst et al, 1994; Nord, 1996; Stavem, 1996] and the extent of this discrepancy tends to be much stronger when patients value their own health state.

Although the empirical link is not proven, it is generally accepted that much of the difference between patients' and the general populations' values for the same health states reflects patients' adaptation to their condition. Patients with chronic illnesses often adapt both physically and psychologically to their health states. Unfortunately, in a resource allocation context, such adaption which may otherwise be considered admirable will result in a deterioration in the influence of the condition and its treatment upon the setting of allocation priorities. A recent article by Ubel and colleagues [2003], outlined a number of possible contributing factors for observed differences between patient and general population values including adaptation effects, a failure to rate the same health state and the use of different measuring rods (or response shifts).

It is very difficult, if not impossible, to construct a complete description of a health state. Most health state descriptive systems contain a limited number of attributes and tend to focus upon the negative aspects of a health state. The difficulties of describing health states inevitably means that individuals tend to bring their own information to the valuation exercise by drawing upon their own personal experiences or stereotypes. Given that the personal experiences of patients and members of the general public are unlikely to be the same it may mean that, in effect, that they are evaluating different health states even when provided with identical descriptions of the health state to be valued. Response shift refers to the possibility that individuals will change their own internal standards for evaluating their own health in response to changes in their health. Response shift may be due to changing expectations (e.g. where an elderly individual rates his/her health according to their expectations of the best possible health at that age rather than best possible health *per se*) or to social comparisons whereby a patient rates his/her health by comparing themselves to other patients rather than to healthy individuals. In either instance this response shift will contribute to discrepancies between patient and general population health state values.

This paper investigates the relationship between patients' valuations of their own health status and general population values of the same health states within the context of a randomised controlled trial to compare alternative treatments for varicose veins.

Varicose veins arise as a consequence of failures in the closing of venous valves and most commonly occur in the legs [Tibbs, 1992]. The characteristic appearance of varicose veins is caused by the extra volume of blood causing the veins to swell and protrude underneath the surface of the skin. Although varicose veins are commonly described as a mild clinical condition, work by Garratt et al [1993] found that varicose veins patients have significantly lower scores than the general population in terms of the physical, emotional, fatigue and pain dimensions of the SF-36. From a cosmetic view-point, the outward appearance of varicose veins may be considered unpleasant and can therefore have an impact upon the emotional and mental well-being of the patient. In addition, physical symptoms can occur including feelings of heaviness or tension in the legs, aching, restless legs, cramps, itching, throbbing, bleeding and ulceration [Bradbury et al, 1999].

Varicose veins treatment receive slow priority in the NHS, reflecting the view that this condition causes cosmetic defects rather than impacting significantly upon health related quality of life. Some health authorities refuse to provide NHS treatment for varicose veins whilst for other health authorities treatment is provided but it is generally limited. Approximately 50,000 operations to treat varicose veins are carried out annually in England and Wales each year costing £400-£600 million to the NHS [Bradbury et al, 1999]. However, there is some debate amongst clinicians concerning which patients should receive treatment and the most appropriate type of treatment.

2. Methods

As part of a randomised clinical trial to compare alternative treatments for varicose veins, patients were surveyed at randomisation and at 12 months following randomisation to assess their own health status using the standard gamble (direct SG) elicitation technique. The SG was administered by a trained interviewer. At the same time points, patients also completed the SF-36 and the EQ-5D health status questionnaires. The randomised controlled trial was designed to compare the clinical effectiveness, impact on quality of life, and cost effectiveness of different varicose vein treatments. Patients with medically uncomplicated varicose veins were randomised to receive either surgical treatment, injection sclerotherapy or conservative treatment according to the severity of their condition classified as either mild (Group 1), moderate (Group 2) or severe (Group 3) varicose veins. Patients who refused randomisation (Group 4) were also invited to participate in clinical and quality of life assessments for observational purposes.

The 36 items of the SF-36 measure eight separate dimensions of HRQOL including physical functioning, role limitations due to physical problems, pain, general health, energy and vitality, social functioning, role limitations due to emotional problems and mental health. Each dimension is scored from 0 to 100 where a higher score represents better health. The EQ-5D is made up of two components. The first component asks respondents to categorise their health status on five dimensions (mobility, self-care, usual activities, pain/discomfort and anxiety/depression), where each dimension has three possible levels of response. The second component involves respondents indicating their own level of health on a visual analogue scale (Direct VAS) bounded by the worst imaginable health state (0) and the best imaginable health state (100).

The EQ-5D VAS algorithm scores were generated using the responses to the second component of the EQ-5D. The responses were converted to a single score based upon general population values (n=2997) for a selection of health states using the VAS [MVH, 1995].[1] The responses to the SF-36 were converted into the SF-6D, a single preference based measure of health using the SG algorithms developed by Brazier and colleagues, version 1 [Brazier et al, 1998] and version 2 [Brazier et al, 2002]. The algorithms are based upon general population values for a selection of health states generated from the SF-36.

SF-6D version 1 algorithm and the direct SG exercise used the same variant of SG - the 'titration' method which was originally developed by Jones- Lee and colleagues at the University of Newcastle in the UK [Jones-Lee et al, 1993]. In the direct SG exercise respondents were asked to choose between living in the health state they were currently experiencing for a period of 10 years or a treatment which may restore them to full health for 10 years or fail and result in immediate death. Version 1 of the SF-6D algorithm was based upon valuations from a convenience sample of health professionals, managers and patients (n=165) who were asked to choose between life in a chronic health state (which was described on a show card) for a period of 10 years or a treatment which may restore them to full health for 10 years or fail and result in immediate death. The titration method presents a range of chances of success from 100% to 10% with a final box for immediate death preferred. Respondents were asked to indicate the chances of success where they would choose the risky treatment. They were then asked to place a cross against the chances of success where they would reject the treatment. Finally they were asked to indicate where it was most difficult to choose.

The SF-6D version 1 VAS algorithm was derived from the same convenience sample of health professionals, managers and patients (n=165) who were asked to place a series of health states, identical to those used for the SG, on a scale bounded by the worst imaginable health state (0) and the best imaginable health state (100) and where the intervals reflect the relative value of the states.[2]

---

[1] To compare the VAS results between respondents, the raw ratings were adjusted relative to each respondent's valuation of the best health state and death. The raw scores were transformed such that the best state had a value of 1.0 and death a value of 0.0 for every respondent. The value of any particular health state could lie within this range, or assume a negative value for states worst than death.

[2] To compare the VAS results between respondents, the raw ratings were adjusted using an identical methodology to that used for transforming the EQ-5D VAS results.

Version 2 of the SF-6D algorithm used a variant of SG using props which was developed by a team at McMaster University in Canada [Furlong et al, 1990]. The algorithm was based upon valuations from a representative sample of the UK population (n=611). As with the titration method, respondents are asked to choose between two alternatives. However the question is framed slightly differently. Respondents are asked to choose between the certain prospect (A) of living in a chronic health state for the rest of their life or the uncertain prospect of two possible outcomes, the best state defined by the SF-6D or the worst (pits). The chances of the best outcome occurring are varied until the respondent is indifferent between the certain and uncertain prospects. The probabilities are displayed on a chance board in both numerical and graphical (pie chart) format. This 'ping-pong with props' version of SG was chosen because of its ease of use for interviewers [Brazier et al, 2002].

The statistical analyses were performed using SPSS version 12.0. All of the statistical tests used in the analysis were parametric. The direct SG scores from patients were compared with general population values based upon the SF-6D version 1 SG algorithm and the SF-6D version 2 algorithm respectively. The statistical significance of any differences found was tested at the 5% level using the paired t-test. Similarly, the direct VAS scores from patients were compared with general population values based upon the SF-6D version 1 VAS algorithm and the EQ-5D VAS algorithm. Given that the SG method used to generate the SF-6D algorithms differed between version 1 and version 2, these comparisons were also able to shed some light on the extent of influence of the variant used in the resulting valuations.

3. Results

A total of 131 patients provided complete responses to the SF-36, EQ-5D VAS and the direct SG at base-line and 12 months time points. The characteristics of the sample are presented in Table 1. In terms of the severity of varicose veins upon clinical diagnosis, 17 patients were classified with mild varicose veins (Group 1); 25 were classified with moderate varicose veins (Group 2); 69 were classified with severe varicose veins (Group 3) and 20 patients agreed to observation but refused to be randomised (Group 4). The mean age was 47 years and there were more females than males.

The baseline data revealed interesting differences in the mean and distribution of the preference based measures. Table 2 summarises the mean scores and differences between patient and general population values for (i) the direct SG and the SF-6D algorithms and (ii) the direct VAS and SF-6D and EQ-5D VAS algorithms. As expected a priori the direct SG produced a higher mean score than either of the SF-6D algorithms (baseline mean (SD) patients' own value: 0.942 (0.116), general population value version 1: 0.913 (0.072), version 2 0.730 (0.108)) and the difference between means for the patient and general population values were statistically significant. For version 1 of the algorithm the difference between patient and general population mean scores was far less pronounced than for version 2. The direct VAS[3] produced a higher mean score than either of the algorithms based upon general population values (baseline mean (SD) patients' own value: 0.762 (0.167), general population value: SF-6D VAS algorithm 0.668 (0.173), EQ-5D VAS algorithm 0.743 (0.173)) with the differences being statistically significant when compared with the SF-6D VAS algorithm but not when compared with the EQ-5D VAS algorithm.

Figure 1 shows a scatter diagram of the relationship between direct SG values and SF-6D version 1 algorithm values. The 45^0 line indicates the line of equality. It can be seen that there is very little agreement between the two methods in terms of the values elicited. An alternative, more informative plot is shown in Figure 2. Here the differences between values (direct SG - SF-6D version 1) have been plotted against the average of the two measurements. The advantages of this plot are that it is possible to see the size of the differences more easily and also their distribution around zero and the relationship between the differences and the size of the measurement [Altman, 1991]. It can be seen that for health state values between 0.80 and 0.90 there is a wide discrepancy between patient and population values with patient values generally being much higher on average. At health state values above 0.90 the level of discrepancy reduces and, as one would expect, patient and population values tend to converge at the highest end of the scale (1.0). Very similar trends are observed when comparing the direct SG values and SF-6D version 2 algorithm values (Figures 3 and 4) with the direct patient valuations being higher on average than general population values.

Table 3 summarises the mean scores and differences between patient and general population values at baseline according to the severity of varicose veins symptoms. With one exception, direct patient based scores were consistently higher than general population algorithm based scores. It can also be seen that, in general, there was concordance between direct SG values and health related quality of life measured according to the SF-36 and EQ-5D in that the magnitude of the differences between

[3] In making comparisons between the direct VAS and VAS algorithm scores it is important to note that the measurement scales are not completely identical since the direct VAS scores represent untransformed (raw) scores where as the SF-6D and EQ-5D VAS algorithms are based solely upon transformed scores.

scores according to the severity of symptoms was similar. Contrary to prior expectations, it was found that patients diagnosed with severe varicose veins consistently reported higher direct SG and VAS scores and rated their health related quality of life higher according to the SF-6D and EQ-5D than did patients diagnosed with moderate varicose veins. However the relatively small numbers of patients in each group means that these results must be interpreted with caution.

Table 4 summarises the mean scores and differences between patient and general population values at the 12 month time-point. For the comparison of direct SG and SG based SF-6D algorithms the trends in differences are very similar to those found at baseline in that the direct SG produced a higher mean score than either of the SF-6D algorithms and for version 1 of the algorithm the difference between patient and general population mean scores was less pronounced than for version 2 (12 months mean (SD) patients' own value: 0.924 (0.115), general population value version 1: 0.912 (0.082), version 2 0.736 (0.108)).The trends in differences between the direct VAS and VAS based algorithms were also very similar to baseline in that the direct VAS produced a higher mean score than either of the algorithms based upon general population values (baseline mean (SD) patients' own value: 0.781 (0.104), general population value: SF-6D VAS algorithm 0.676 (0.191), EQ-5D VAS algorithm 0.771 (0.155)) with the differences being statistically significant when compared with the SF-6D VAS algorithm but not when compared with the EQ-5D VAS algorithm.

Interestingly the mean direct SG score at 12 months is lower than that achieved at baseline and the SF-6D scores remain virtually unchanged whereas the mean direct VAS scores at 12 months show a slight improvement relative to baseline. Upon further inspection it was found that a small number of patients rated their own health very poorly according to the direct SG at 12 months which contributes to a reduction in the mean score overall. A higher proportion of patients rated their own health status at 1.0 according to the direct SG at 12 months (48% compared to 38% at baseline) indicating that for a proportion of patients in this sample treatment does appear to have been successful in improving their health status.

Figures 5 and 6 show scatter diagrams of the relationship between direct SG values and SF-6D version 1 algorithm values at 12 months. Figures 7 and 8 show scatter diagrams of the relationship between direct SG values and SF-6D version 2 algorithm values at 12 months. The pattern is very similar to that exhibited at baseline in that there is generally very little agreement between patient and population values although the values do tend to converge at the highest end of the scale.

Table 5 summarises the mean scores and differences between patient and general population values at 12 months according to the severity of varicose veins symptoms. As with the baseline data, direct patient based scores were consistently higher than general population algorithm based scores. However, in contrast to the baseline data, there was no concordance between direct SG and VAS values and health related quality of life measured according to the SF-6D and EQ-5D in that there was no discernible pattern to the differences between scores across severity groupings.

4. Discussion

This paper adds to the evidence that patients own health valuations tend to be higher than those of a general population sample. This has important implications for economic evaluation. For example, in the context of the results reported here we found that the incremental QALY gain at 12 months for patients receiving treatment for moderate varicose veins would be almost halved if patients own SG values were used to estimate QALYs in contrast to the general population values from the SF-6D version 2.

Varicose veins is often characterised as a relatively mild clinical condition and the results from this study tend to reinforce this view, with patients generally reporting themselves in good health at both baseline and 12 month time points. With the exception of the SF-6D version 2, the observed differences between direct patient and general population values are generally smaller than have been observed in other studies for conditions which may be classified as more severe[4]. There are a number of possible explanations for this finding including a reduction in the adaptation effect, patient selection issues and the lack of impact of the descriptive systems in capturing the ‘experience’ of living with varicose veins.

Firstly, it is likely that, for this sample of patients, the issue of adaptation and coping is less important because varicose veins tend to cause few physical defects. An individual diagnosed with varicose veins is unlikely to lose the ability to perform certain activities and hence the level of physical and emotional adaptation required will be lower than that exhibited by patients with more severe conditions. For the same reasons, the level of

[4] . For example, Boyd et al [1990] found that patients with colostomies assigned a mean health state value of 0.92 to their condition whereas the general population mean health state value for this condition was 0.80.

response shift exhibited (whereby a varicose veins patient rates his/her health according to their expectations of best possible health given their condition rather than to best possible health *per se*) is unlikely to be high.

Secondly, it can be postulated that the observed differences between direct patient and general population values, may be artificially de-flated because this study sample is made up of patients diagnosed with varicose veins who are concerned enough about their condition to present at an out-patients clinic in order to receive treatment. These patients may be reporting lower own SG and VAS scores than would the large number of individuals in the community living with varicose veins who do not actively seek treatment for their condition because they do not consider that it impacts significantly enough upon their overall health.

Thirdly, the SF-36 and the EQ-5D are generic health state descriptive systems which by necessity use a limited number of characteristics or attributes to represent health related quality of life across a broad spectrum of conditions ranging from mild to very severe. It is probable that these instruments will not be sensitive enough to describe varicose veins adequately and therefore they may, to some extent, mis-represent the physical and emotional impact of this condition and its treatment. This finding is important given the recent recommendations by the National Institute of Clinical Excellence that a genetic instrument e.g. EQ-5D should be used to value the benefits of health care interventions [National Institute for Clinical Excellence, 2003] for all clinical conditions. When asked to value their own health state directly, patients inevitably draw upon their own personal experience of their condition and its impact upon their health related quality of life. This may also be a factor when patients are asked to respond to a generic health state descriptive system. Although, in theory, patient and general population values were based upon identical health states, in practice it is highly unlikely that this was the case.

The results from this study also suggest that the variant of elicitation technique used to elicit SG values is likely to more important than the source of those values in revealing discrepancies. We found major differences between SF-6D values for identical health states according to whether these were based upon the version 1 algorithm which used the titration SG method or the version 2 algorithm which used the ping pong with props version of SG, with the titration based method consistently providing higher values. Given that the two algorithms were derived using different general population samples it is not possible to completely exclude respondent bias as a possible cause of the observed differences. However, a recent study which compared the two versions directly within the same sample of respondents also found that the titration method produced higher values for all SF-6D health states valued and the differences were statistically significant [Brazier and Dolan, 2004]. One possible explanation for the discrepancy is an anchoring effect since the titration method elicits preferences from the upper end of the scale initially and works gradually towards the lower end of the scale, whereas the ping pong with props version iterates between the upper and lower ends, thereby reducing any anchor point bias.

In reality, it has been argued that it is difficult to justify the viewpoints of using ***only*** patients or uninformed members of the general population to obtain values for health states to be used within economic evaluation to aid health care decision-making [Fryback 2003; Brazier 2002]. Patient values incorporate aspects of adaptation which may be viewed as unacceptable for public resource allocation decisions e.g. the lowering of expectations. Alternatively, general population values are largely uninformed by what it is like to living in the health state being valued. A number of commentators have recently put forward the view that ultimately general population values are required to assist in resource allocation decisions in a public system. However, these values should be ‘informed’ by providing respondents with more information about what it is like to living in the health states being assessed e.g. by improving health state descriptions to make them more realistic and by providing more information on the size and nature of the adaptation experienced by patients over time [Brazier et al, 2004; Ubel et al 2003; Fryback, 2003]. For the condition of varicose veins, it may be that the main problem is the lack of sensitivity afforded by the health state descriptions generated by the EQ-5D and SF-6D and modification of these systems (e.g. by adding new health state levels or dimensions or changes to the health state descriptions themselves) would help to improve this.

Unfortunately this study was not designed specifically to examine the relative importance of possible sources of the discrepancies between patient and general population values for health states in varicose veins. The need for more comprehensive studies to more fully understand the nature of discrepancies between patient and general population values is clearly a major challenge for future research in this area and represents an essential stepping stone towards the attainment of informed general population values for use within economic evaluation in health care.

References

Boyd NF, Sutherland HJ, Heasman ZK, Cummings BJ [1990]. Whose values for decision making? *Medical Decision Making* 10: 58-67.

Bradbury A, Exam C, Allan P *et al* [1999]. What are the symptoms of varicose veins? The Edinburgh Veins study cross-sectional population survey *British Medical Journal* 318: 1440-44.

Brazier J and Dolan P [2004]. Evidence of preference construction in a comparison of variants of the standard gamble method. Submitted for publication.

Brazier J, Roberts J, Deverill M. The estimation of a preference based single index measure for health from the SF-36 [2002]. *Journal of Health Economics* 21: 271-292.

Brazier J, UsherwoodT, Harper P and Thomas K [1998]. Deriving a preference based single index from the UK SF-36 health survey. *Journal of Clinical Epidemiology* 51: 1115-1128.

Buckingham [1993]. A note on HYE's (healthy year equivalents). *Journal of Health Economics* 12: 301-309.

Canadian Coordinating Office for Health Technology Assessment (CCOHTA) [1997] *Guidelines for economic evaluation of pharmaceuticals: Canada.* 2nd edition. CCOHTA< Canada.

Commonwealth of Australia [1995]. *Guidelines for the pharmaceutical industry on preparation of submissions to the Pharmaceutical Benefits Advisory Committee: including economic analyses.* Department of Health and Community Serrvices. Canberra.

Dolan P [1997]. Modelling valuations for Euroqol health states. *Medical Care* 35: 351-363.

Feeny DH, Furling WJ, Torrance GW *et al* [2002]. Multiattribute and single attribute utility function – the health utility index mark 3 system. *Medical Care* 40: 113-128.

Fryback DG [2003]. Whose quality of life? Or whose decision? *Quality of Life Research* 12: 609-610.

Furlong W, Feeny D, Torrance GW *et al* [1990] *Guide to design and development of health state utility instrumentation.* Centre for Health Economics and Policy Analysis paper 90-9. McMaster University, Hamilton, Ontario.

Gafni A [1991]. Willingness to pay as a measure of benefits: relevant questions in the context of public decision making about health care programmes. *Medical Care* 29: 1246-1252.

Garratt AM, Rutta DA, Abdalla MI [1993]. The SF-36 health survey questionnaire: an outcome measure suitable for routine use within the NHS. *British Medical Journal* 306: 1440-44.

Gold MR, Siegel JE, Russell LB, Weinstein MC [1996]. *Cost effectiveness in health and medicine*. Oxford University Press, Oxford.

Hurst NP, Jopanputra P, Hunter M *et al* [1994]. Validity of EuroQol in patients with rheumatoid arthritis. *British Journal of Rheumatology* 33: 655-662.

Jones-Lee M, Loomes G, O'Reilly D, Phillips P [1993]. *The value of preventing non fatal road injuries: findings of a willingness to pay national sample survey*. Transport Research Laboratory.

MVH Group [1995]. *The measurement and valuation of health: Final report on the modelling of valuation tariffs*. Centre for Health Economics, University of York.

National Institute for Clinical Excellence [2001]. *Guide to the technology appraisal process*, London.

National Institute for Clinical Excellence [2003]. *Guide to the methods of technology appraisal: draft for consultation*, London.

Nord E [1996]. TTO scores in patients with chronic disease. Comparison with the York hypothetical TTO tariff. Paper presented at the EuroQol plenary meeting, Oslo, Norway.

Sackett DL and Torrance GW [1978]. The utility of different health states as perceived by the general public. *Journal of Chronic Diseases* 31: 697-704.

Stavem K [1996]. Valuing theoretical versus own condition: a comparison using TTO and the EuroQol instrument in COPD patients. Paper presented at the EuroQol plenary meeting, Oslo, Norway.
Tibbs DJ [1992]. *Varicose veins and related disorders*. Butterworth-Heineman, Oxford.
Torrance G [1986]. Measurement of health state utilities for economic appraisal: a review. *Journal of Health Economics* 5: 1-30.
Ubel PA, Loewenstein G, Jepson C. Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. *Quality of Life Research* 12: 599-607.

Table 1: Demographic and clinical profile of patients (n=131)

Age (SD)	47.15 (11.41)
Female (%)	92 (70%)
Severity of varicose veins:	
Mild	17 (13%)
Moderate	25 (19.1%)
Severe	69 (52.7%)
Refused randomisation but agreed to observation	20 (15.3%)
*Treatment received for varicose veins:**	
Conservative	51 (38.9%)
Sclerotherapy	19 (14.5%)
Surgery	47 (35.9%)

* missing = 14

Table 2: Summary of patient and general population based scores at baseline (n=131)

Direct SG Mean score (SD)	***SF-6D version 1 Mean score (SD)***	***Difference between means***	***95% confidence interval***
0.942 (0.116)	0.913 (0.072)	0.029 *	0.007 to 0.051
Direct SG Mean score (SD)	***SF-6D version 2 Mean score (SD)***	***Difference between means***	***95% confidence interval***
0.942 (0.116)	0.730 (0.108)	0.212 *	0.187 to 0.238
Direct VAS Mean score (SD)	***SF-6D V1 VAS algorithm Mean score (SD)***	***Difference between means***	***95% confidence interval***
0.762 (0.167)	0.668 (0.173)	0.094 *	0.069 to 0.119
Direct VAS Mean score (SD)	***EQ-5D VAS algorithm Mean score (SD)***	***Difference between means***	***95% confidence interval***
0.762 (0.167)	0.743 (0.173)	0.019	-0.023 to 0.061

* = statistically significant difference between means at 5% level

Figure 1: Scatter plot comparison of patient and general population values (Version 1 SF-6D algorithm)

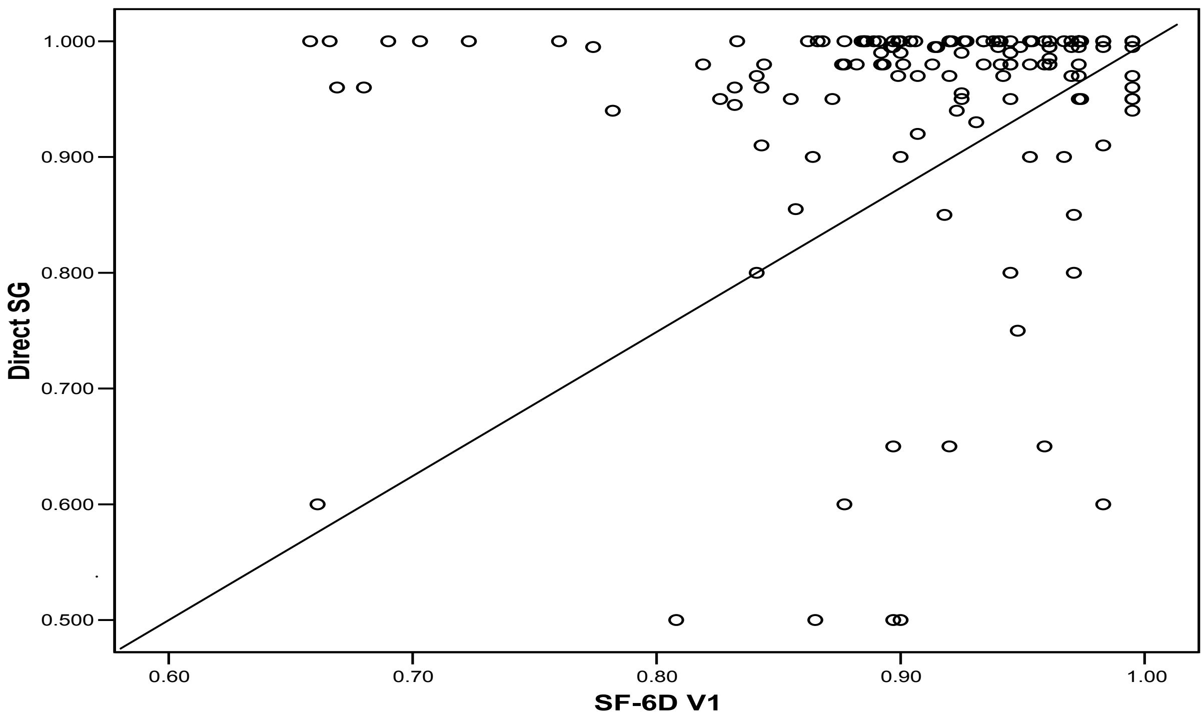

Figure 2: Scatter plot comparison of differences between patient and general population values (Version 1 SF-6D algorithm)

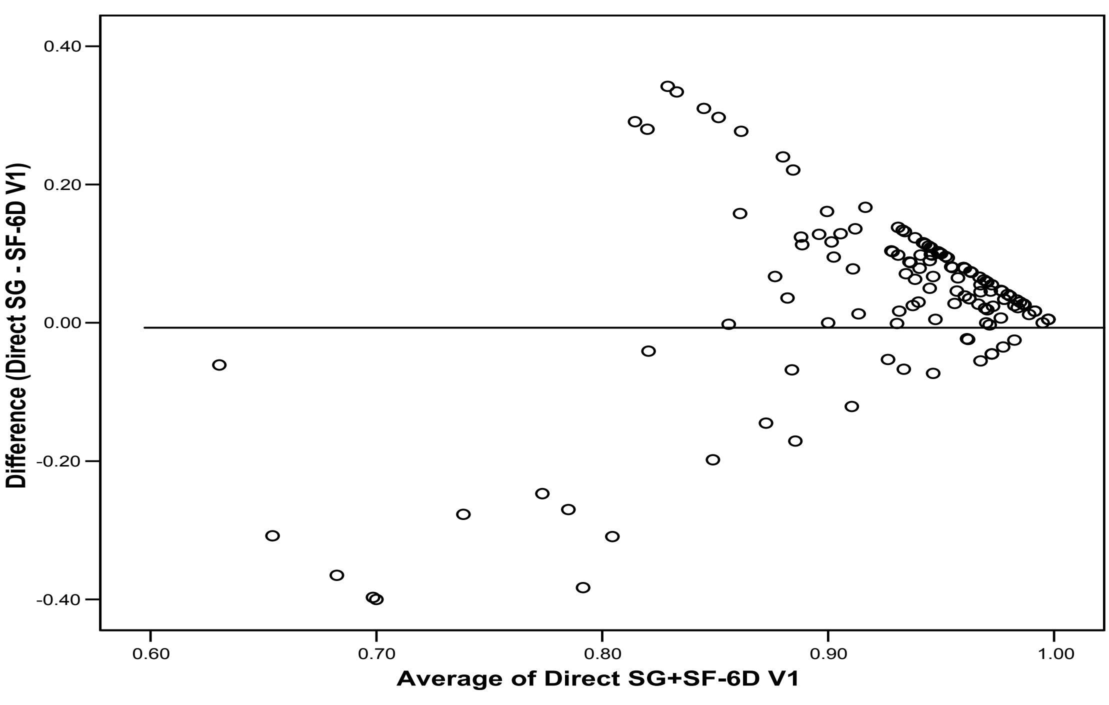

Figure 3: Scatter plot comparison of patient and general population values (Version 2 SF-6D algorithm)

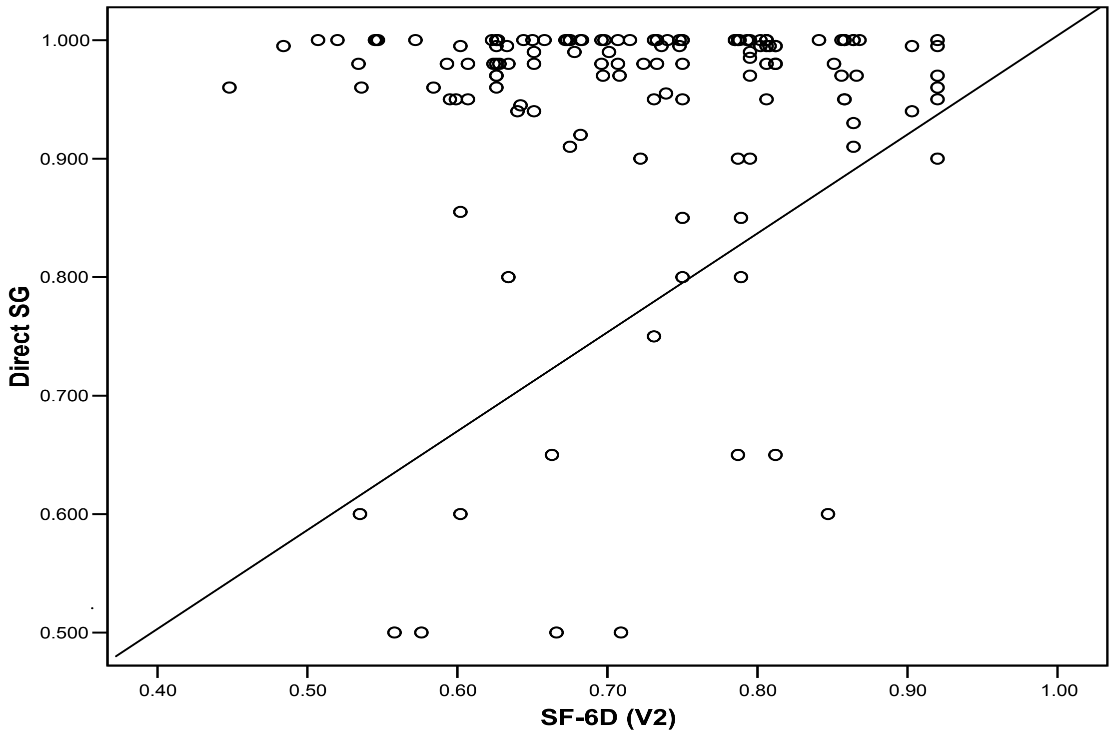

Figure 4: Scatter plot comparison of differences between patient and general population values (Version 2 SF-6D algorithm)

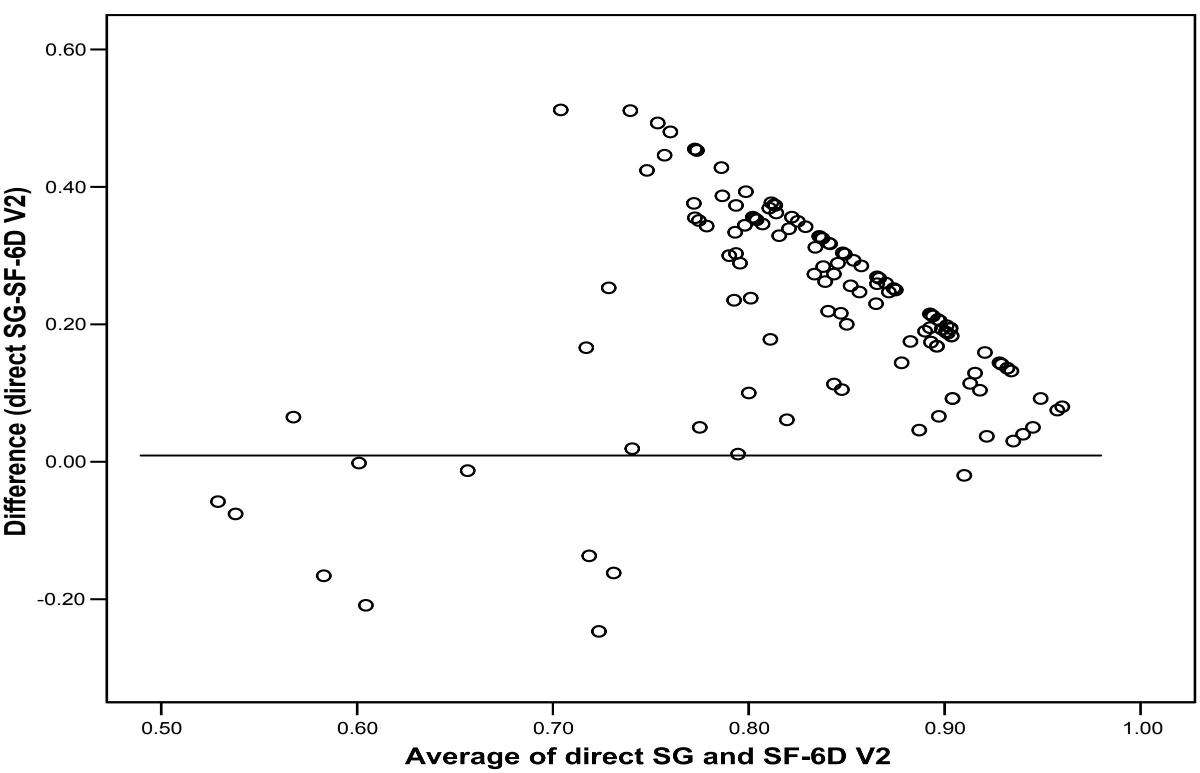

Table 3: Summary of patient and general population based scores at baseline according to clinical condition

Severity of varicose veins[1]	*Direct SG Mean score (SD)*	*SF-6D version 1 Mean score (SD)*	*Difference between means*	*95% confidence interval*
Mild (n=17)	0.956 (0.099)	0.912 (0.075)	0.044	-0.015 to 0.104
Moderate (n=25)	0.899 (0.159)	0.907 (0.052)	-0.009	-0.077 to 0.060
Severe (n=69)	0.941 (0.114)	0.910 (0.084)	0.031*	0.002 to 0.059
[1] Missing = 20				
	Direct SG Mean score (SD)	*SF-6D version 2 Mean score (SD)*	*Difference between means*	*95% confidence interval*
Mild	0.956 (0.099)	0.709 (0.100)	0.247*	0.190 to 0.306
Moderate	0.899 (0.159)	0.691 (0.098)	0.208*	0.147 to 0.283
Severe	0.941 (0.114)	0.744 (0.112)	0.197*	0.169 to 0.234
	Direct VAS Mean score (SD)	*SF-6D V1 VAS algorithm Mean score (SD)*	*Difference between means*	*95% confidence interval*
Mild	0.802 (0.126)	0.669 (0.157)	0.133*	0.076 to 0.190
Moderate	0.737 (0.160)	0.634 (0.148)	0.105*	0.049 to 0.156
Severe	0.767 (0.173	0.682 (0.187)	0.085*	0.047 to 0.122
	Direct VAS Mean score (SD)	*EQ-5D VAS algorithm Mean score (SD)*	*Difference between means*	*95% confidence interval*
Mild	0.802 (0.126)	0.744 (0.157)	0.058*	0.004 to 0.111
Moderate	0.737 (0.160)	0.713 (0.154)	0.024	-0.052 to 0.099
Severe	0.767 (0.173	0.744 (0.195)	0.022	-0.020 to 0.065

* = statistically significant difference between means at 5% level

Table 4: Summary of patient and general population based scores at 12 months (n=131)

Direct SG Mean score (SD)	*SF-6D version 1 Mean score (SD)*	*Difference between means*	*95% confidence interval*
0.924 (0.155)	0.912 (0.082)	0.012	-0.016 to 0.040
Direct SG Mean score (SD)	*SF-6D version 2 Mean score (SD)*	*Difference between means*	*95% confidence interval*
0.924 (0.155)	0.736 (0.108)	0.188 *	0.155 to 0.244
Direct VAS Mean score (SD)	*SF-6D V1 VAS algorithm Mean score (SD)*	*Difference between means*	*95% confidence interval*
0.781 (0.104)	0.676 (0.191)	0.104 *	0.080 to 0.128
Direct VAS Mean score (SD)	*EQ-5D VAS algorithm Mean score (SD)*	*Difference between means*	*95% confidence interval*
0.781 (0.104)	0.771 (0.155)	0.010	-0.007 to 0.050

* = statistically significant difference between means at 5% level

Figure 5: Scatter plot comparison of patient and general population values (Version 1 SF-6D algorithm)

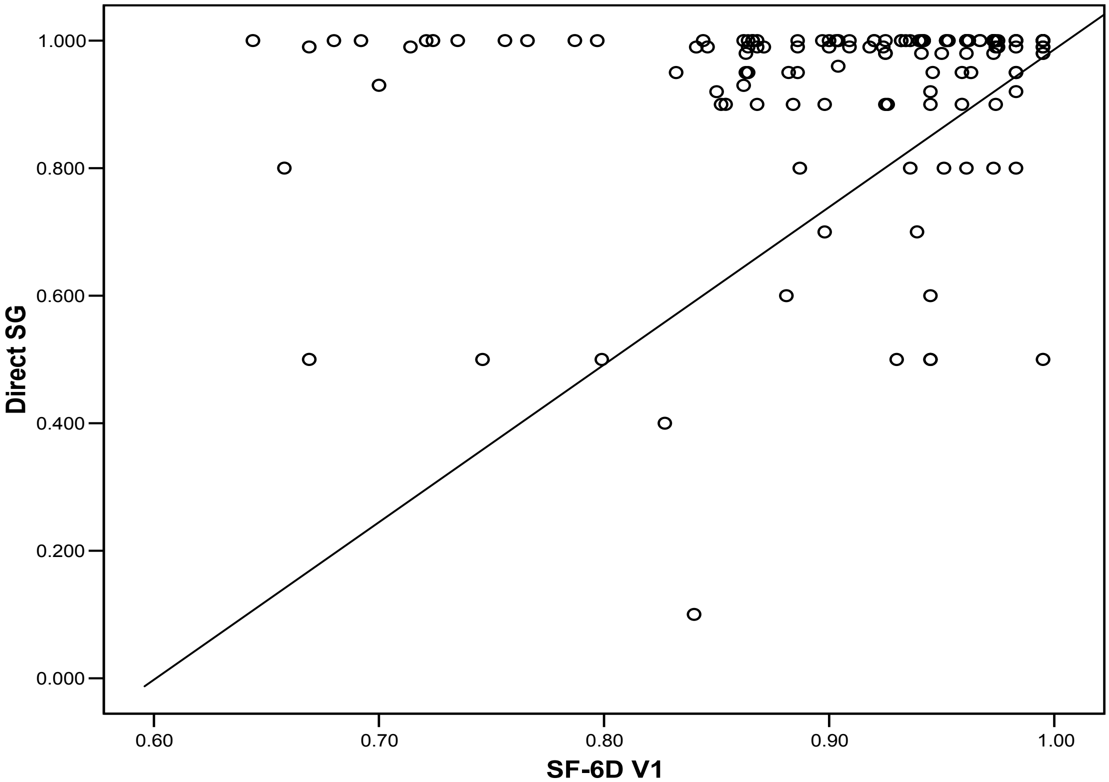

Figure 6: Scatter plot comparison of differences between patient and general population values (Version 1 SF-6D algorithm)

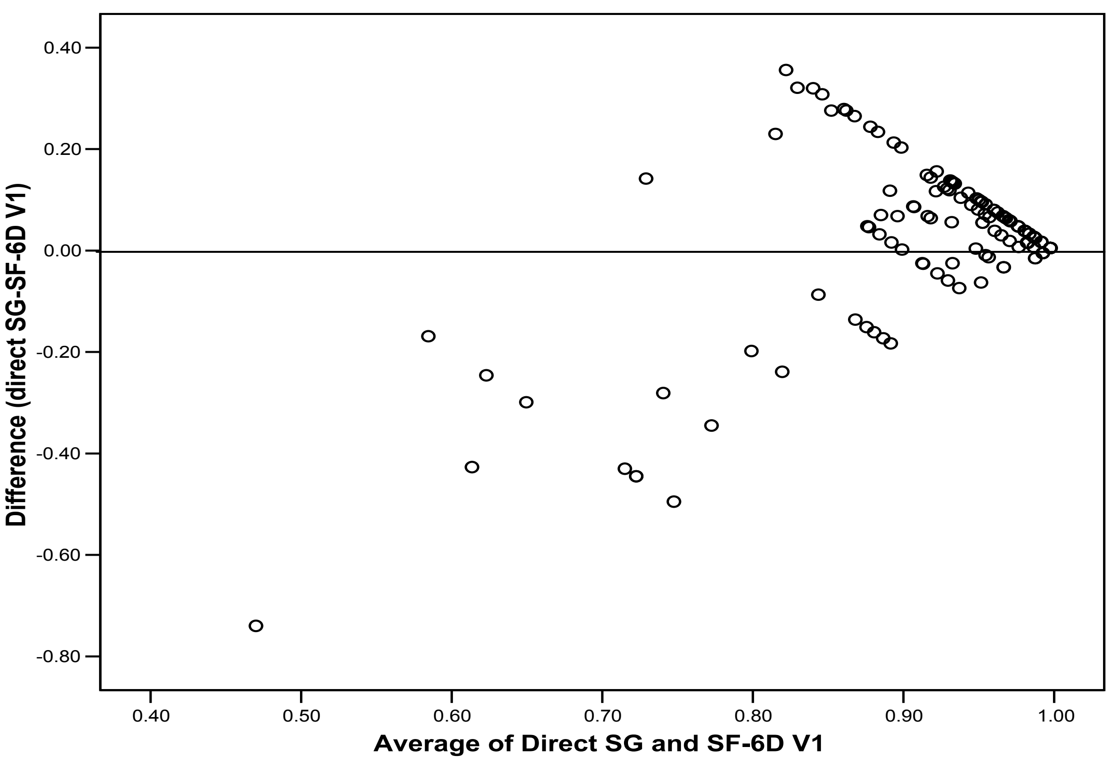

Figure 7: Scatter plot comparison of patient and general population values (Version 2 SF-6D algorithm)

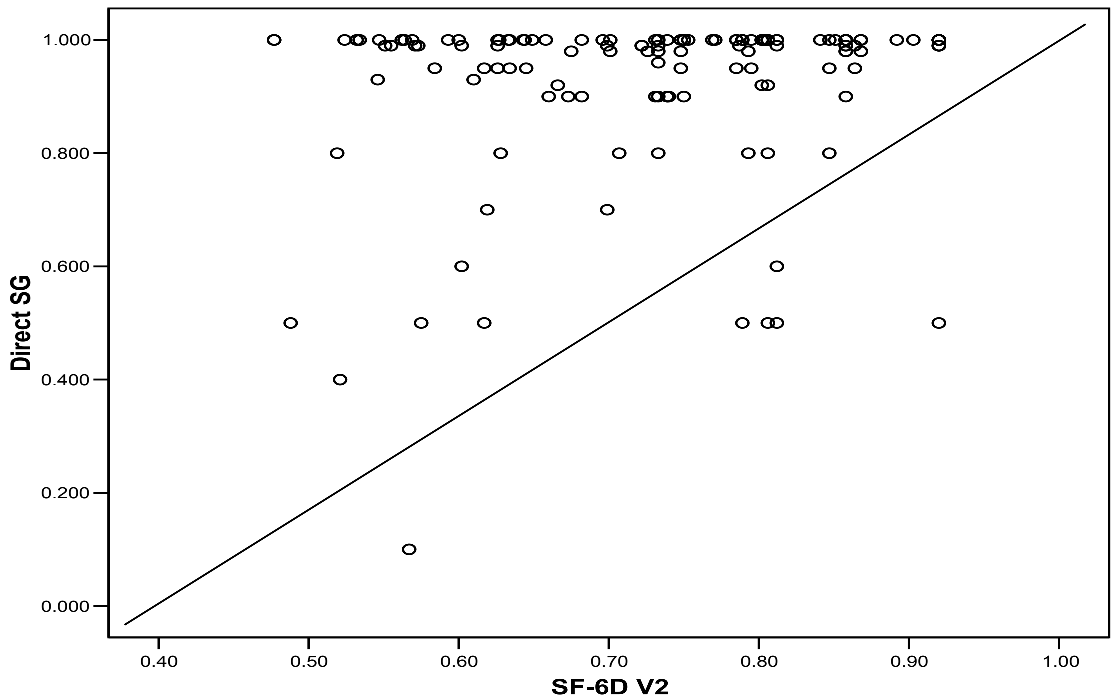

Figure 8: Scatter plot comparison of differences between patient and general population values (Version 2 SF-6D algorithm)

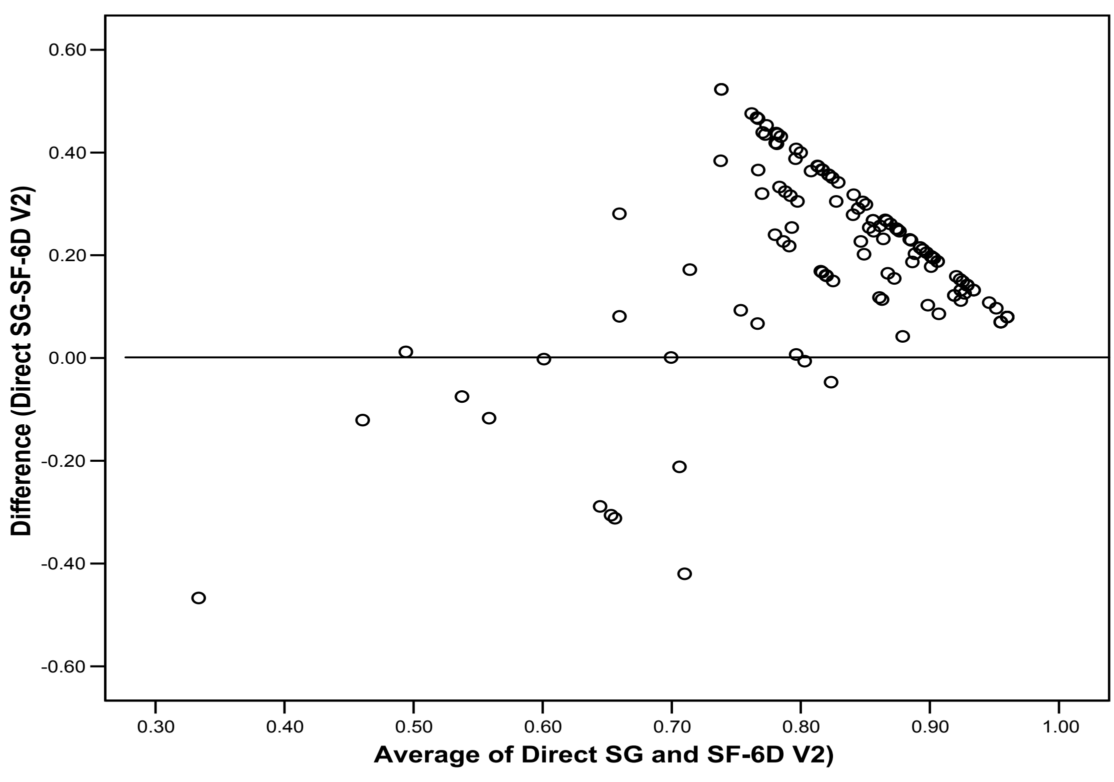

Table 5: Summary of Direct SG and SF-6D scores at 12 months according to clinical condition

Severity of varicose veins[1]	***Direct SG Mean score (SD)***	***SF-6D version 1 Mean score (SD)***	***Difference between means***	***95% confidence interval***
Mild (n=17)	0.955 (0.122)	0.914 (0.087)	0.041	-0.039 to 0.121
Moderate (n=25)	0.948 (0.099)	0.934 (0.069)	0.014	-0.038 to 0.066
Severe (n=69)	0.919 (0.170)	0.910 (0.082)	0.009	-0.031to 0.048
[1] Missing = 20				
	Direct SG Mean score (SD)	***SF-6D version 2 Mean score (SD)***	***Difference between means***	***95% confidence interval***
Mild	0.955 (0.122)	0.732 (0.123)	0.223*	0.128 to 0.319
Moderate	0.948 (0.099)	0.771 (0.105)	0.177*	0.115 to 0.238
Severe	0.919 (0.170)	0.734 (0.114)	0.185*	0.144 to 0.225
	Direct VAS Mean score (SD)	***SF-6D V1 VAS algorithm Mean score (SD)***	***Difference between means***	***95% confidence interval***
Mild	0.792 (0.195)	0.687 (0.199)	0.105*	0.035 to 0.175
Moderate	0.813 (0.163)	0.730 (0.169)	0.083*	0.040 to 0.126
Severe	0.767 (0.173)	0.664 (0.199)	0.104*	0.069 to 0.139
	Direct VAS Mean score (SD)	***EQ-5D VAS algorithm Mean score (SD)***	***Difference between means***	***95% confidence interval***
Mild	0.792 (0.195)	0.735 (0.157)	0.057*	0.006 to 0.115
Moderate	0.813 (0.163)	0.798 (0.149)	0.015	-0.045 to 0.084
Severe	0.767 (0.173)	0.749 (0.191)	0.018	-0.040 to 0.089

* = statistically significant difference between means at 5% level