



# Swedish obese subjects (SOS) – an intervention study of obesity. Two-year follow-up of health-related quality of life (HRQL) and eating behavior after gastric surgery for severe obesity

J Karlsson<sup>1</sup>, L Sjöström<sup>2</sup> and M Sullivan<sup>1</sup>

<sup>1</sup>Health Care Research Unit and <sup>2</sup>Institute of Internal Medicine, Sahlgrenska University Hospital, University of Göteborg, Göteborg, Sweden

**OBJECTIVE:** To examine the effects of weight loss on health-related quality of life (HRQL) in subjects with severe obesity.

**DESIGN:** Controlled clinical trial of the outcomes of surgical vs conventional weight reduction treatment.

**SUBJECTS:** The first 487 surgical cases and their conventionally treated, matched controls were followed for two years in the Swedish Obese Subjects (SOS) intervention study.

**MEASUREMENTS:** A battery of generic and study-specific self-assessment instruments or subscales was used to characterize HRQL in the severely obese (BMI  $\geq 34$  kg/m<sup>2</sup> for males and BMI  $\geq 38$  kg/m<sup>2</sup> for females). Measures of general health perceptions (general health rating index; current health), mental well-being (mood adjective check list; pleasantness, activation and calmness), mood disorders (hospital anxiety and depression scale; anxiety and depression) and social interaction (sickness impact profile), were supplemented by obesity-specific modules on obesity-related psychosocial problems and eating behavior (three-factor eating questionnaire; restrained eating, disinhibition and perceived hunger). Assessments were conducted prior to treatment and repeated after 6, 12 and 24 months.

**RESULTS:** Poor HRQL before intervention was dramatically improved after gastric restriction surgery, while only minor fluctuations in HRQL scores were observed in the conventionally treated controls. Peak values were observed in the surgical group at 6 or 12 months after intervention with a slight to moderate decrease at the two-year follow-up. The positive changes in HRQL after two years were related to the magnitude of weight loss, that is, the greater the weight reduction, the greater the HRQL improvements. Eating behavior improved accordingly.

**CONCLUSION:** Quality of life in the severely obese is improved by substantial weight loss. Most patients benefit from weight reduction surgery, while HRQL in surgical patients with minor reduction in overweight is less positive. Further research is needed to determine outcome predictors of the surgical management of severe obesity and to ensure that HRQL improvements are maintained.

**Keywords:** quality of life; eating behavior; health assessment; severe obesity; gastric surgery; controlled clinical trial

## Introduction

The prevalence of negative health impacts of obesity has made its prevention and treatment a public health priority. Severe obesity is associated with a number of serious medical conditions such as hypertension, lipid disturbances, and diabetes mellitus, all of which are known risk factors for coronary heart disease.<sup>1</sup> Several large epidemiological studies have also shown that the mortality rate among the severely obese is elevated.<sup>1</sup> Moreover, obesity is a stigmatized condition in western culture and the discrimination of obese people in several areas of social life has been documented.<sup>2–4</sup>

Available data on quality of life indicate poor well-being and serious psychosocial dysfunction among the extremely obese. Results from the first 1743 subjects registered in the Swedish Obese Subjects (SOS) study showed a markedly lower mood state and more anxiety and depression symptoms than in healthy reference subjects.<sup>5</sup> High levels of obesity-related psychosocial problems in everyday life and dysfunction in social interaction were also observed. Furthermore, health-related quality of life (HRQL) was worse in the severely obese than in several other groups of patients with chronic conditions.

Poor HRQL is assumed to be improved by weight reduction surgery and substantial weight loss.<sup>6</sup> Since conventional approaches (for example, dieting) have proven unsuccessful for the long-term management of severe obesity, attention has turned to surgical techniques. The radical weight loss and psychosocial improvements reported in earlier studies on the effects of intestinal bypass operations were recognized more than two decades ago among both surgeons and

Correspondence: Marianne Sullivan, Health Care Research Unit, Institute of Internal Medicine, Sahlgrenska University Hospital, S-413 45 Göteborg, Sweden.

Received 14 April 1997; revised 22 September 1997; accepted 30 September 1997

E-mail: healthcare.research@medicine.gu.se

psychiatric-behavioral authorities.<sup>7</sup> Because many severe post-operative complications resulted from intestinal bypass surgery, it was replaced by more benign gastric restriction methods.<sup>8</sup> Outcome studies of gastric surgery have further confirmed the positive effects on well-being and functioning in most patients.<sup>9-11</sup> However, there are also reports of adverse effects in subgroups of patients and controversy exists concerning overall outcome on life quality.<sup>12-13</sup> Reliable information on the long-term effects of weight reduction treatment is still lacking.<sup>14</sup> Few studies have followed patients for longer periods and very few report on quality of life using well-established instruments. Such results from controlled longitudinal studies on the management of severe obesity are still not available.

Quality of life has been widely accepted as an important concept in medical research and a growing number of clinical trials include measures of functional status and well-being in outcome assessment. Although no single measure or strategy in quality of life research is apt to be adopted as the gold standard, greater understanding of major health dimensions and refined measurement techniques have given rise to more efficient operational definitions and standardization of methods. There is now agreement on a set of core dimensions to be included in the assessment of HRQL.<sup>15,16</sup> The study of Swedish obese subjects (SOS) was designed in accordance with these guidelines.

The SOS intervention study is an ongoing controlled clinical trial on the effects of long-term weight loss on health and well-being.<sup>17</sup> Subjects who lose weight by surgical means are observed and compared with conventionally treated, matched controls over a ten-year period. The aim of the current paper is to report on HRQL in subjects followed for two years. HRQL in the first 487 surgical patients and their controls was assessed prior to treatment and measurements were repeated after 6, 12 and 24 months.

The following specific questions were addressed:

- Does weight reduction improve HRQL in the severely obese?
- Is weight reduction surgery effective in the two-year perspective?
- Is there a correspondence between the magnitude of HRQL changes and weight loss?
- Does self-assessed eating behavior (TFEQ) improve after surgical treatment?

## Methods

### SOS study

SOS is an ongoing nationwide, multi-center project which comprises a registry study and an intervention trial. Since its start in October 1987 about 7500

(August 1997) severely obese people have been accepted in the registry study. Subjects are recruited by advertisements in the media. Inclusion criteria are age at accrual (37-57 y) and body mass index (BMI)  $\geq 34$  kg/m<sup>2</sup> for males and BMI  $\geq 38$  kg/m<sup>2</sup> for females. All subjects complete a series of extensive questionnaires and thereafter take part in a health examination. A detailed description of the SOS study design, recruitment and assessment procedures is reported elsewhere.<sup>17</sup> A baseline description of HRQL in the first 1743 severely obese included in the registry study, is presented in Sullivan *et al.*<sup>5</sup>

*SOS intervention study.* Subjects are recruited from the registry study to take part in an intervention trial designed to test if the negative effects of severe obesity on mortality, morbidity and quality of life are reduced during long-term weight reduction. Inclusion criteria are age (37-60 y) and BMI  $\geq 34$  kg/m<sup>2</sup> for males and  $\geq 38$  kg/m<sup>2</sup> for females. Two groups of patients are followed for ten years, one surgically treated (gastric banding, vertical banded gastroplasty and gastric bypass) and one conventionally treated. The surgical cases are also recruited from existing waiting lists at surgical departments. All patients choosing surgery, undergo a comprehensive pre-operative evaluation. Exclusion criteria for the intervention study have previously been described.<sup>17</sup>

A non-randomized, matched-group design was chosen for the intervention study, that is, a control case is selected from the registry (according to the method of sequential assignment by Pocock and Simon<sup>18</sup>) to match each included surgical patient. In order to obtain two groups that do not differ systematically with respect to prognosis prior to treatment, a computerized matching procedure selects an optimal control case taking into account 18 variables, six of which are psychosocial.<sup>17</sup>

The SOS intervention study is projected to comprise 2000 surgical cases and 2000 matched controls. Obesity surgery and follow-up examinations are performed at county or university hospitals. All surgical patients are given instructions on nutrition. The control subjects are offered treatment at primary health care centres (more than 450 centres are currently participating in the SOS study) and the treatment regimen varies according to the local routines. All surgical and non-surgical patients return regularly for complete medical check-ups during the ten-year observation period.

Assessments of HRQL are conducted before intervention and repeated after six months and one, two, three, four, six, eight and ten years after intervention. Main quality of life aspects of the intervention study concern treatment effects in surgical cases vs controls on functional status, mental well-being and general health perceptions. The SOS study has been approved by the local ethics committees of all Swedish universities.

### Study sample characteristics

The study sample at the present stage of analysis includes the first 487 surgical patients followed for two years and their conventionally treated, matched controls. There were twice as many females ( $n = 327$ ) as males ( $n = 160$ ) in each group. Mean age on the operation day was 46.6 (95% confidence interval (CI) 46.1–47.1) y in the surgical group and 47.7 (CI: 47.2–48.3) y in the control group. Differences in overweight characteristics at baseline were observed. As seen in Table 1, the surgical subjects were significantly more obese than controls. In both groups, the average body weight was higher in males than females, while BMI was lower in males.

**Surgical techniques.** Vertical banded gastroplasty was performed in 315 (65%) of the operations, gastric banding techniques were used in 136 (28%) and gastric bypass in 36 (7%) cases.

**Weight change.** Rapid weight loss was observed during the first six months after surgery. Weight loss was higher in males than females; 29.1 (CI: 27.2–31.0) kg vs 26.0 (CI: 25.0–27.0) kg on average. A further average loss of 3.5 (CI: 2.9–4.0) kg was observed during the second half of the first year, while an average weight regain of 2.5 (CI: 1.9–3.2) kg was registered during the second year. The average reduction of BMI, two years after surgery was 9.3 (CI: 8.5–10.1) kg/m<sup>2</sup> in males and 9.8 (CI: 9.2–10.3) kg/m<sup>2</sup> in females. Average weight loss in controls was minor; 0.8 (CI: 0.3–1.3) kg after six months, 1.2 (CI: 0.4–1.9) kg after one year and 0.7 (CI: –0.2–1.6) kg after two years.

**Compliance rate.** The compliance rate was high in both treatment groups. In the surgical group, the

HRQL battery was fully completed by 95% at 6-month, 98% at 12-month and 98% at 24-month follow-ups. Corresponding rates in the control group were 90, 90 and 84%.

**Compliers vs non-compliers in the control group.** At the two year follow-up, data on HRQL was missing for 24 (15%) male and 56 (17%) female controls. Comparison of compliers and non-compliers showed a tendency towards somewhat poorer ratings on all HRQL measures at baseline in non-compliers. Female drop-outs had a slightly higher mean BMI at baseline, 41.7 (CI: 40.0–43.5) kg/m<sup>2</sup> in non-compliers vs 40.4 (CI: 39.9–41.0) kg/m<sup>2</sup> in compliers.

### Weight parameters

Body weight was measured in light clothing without shoes to the nearest 0.1 kg using calibrated balances or electronic scales. Height was measured in a standing position without shoes to the nearest 0.01 m. BMI was calculated as weight divided by height squared (kg/m<sup>2</sup>).

### HRQL assessment instruments

A comprehensive battery of generic and study-specific self-assessment measures was used to characterize HRQL in the severely obese.<sup>5</sup>

**Health perception.** Subjective health was measured by the current health scale (CH) selected from the general health rating index (GHRI).<sup>19</sup> The CH scale includes nine general statements on perceived current health status. The original five point response format was modified in the SOS study to a four-point scale with two acceptance and two rejection alternatives. Psychometric properties of the scale in SOS with

**Table 1** Body weight and body mass index (BMI) in surgical cases ( $n = 487$ ) and controls ( $n = 487$ ), by gender, before intervention and at two-year follow-up

	Mean (95% confidence interval)	
	Surgical cases*	Control cases*
Body weight		
Before intervention		
Males	130.8 (128.1–133.4)	124.4 (121.8–127.0)
Females	115.1 (113.6–116.5)	109.5 (107.8–111.3)
Two-year follow-up**		
Males	100.6 (98.0–103.2)	123.6 (120.7–126.5)
Females	88.4 (86.7–90.1)	108.6 (106.8–110.5)
BMI		
Before intervention		
Males	40.8 (40.1–41.5)	38.7 (38.0–39.5)
Females	42.3 (41.9–42.7)	40.7 (40.1–41.2)
Two-year follow-up**		
Males	31.5 (30.7–32.3)	38.5 (37.6–39.3)
Females	32.5 (32.0–33.0)	40.3 (39.7–40.9)

\*All comparisons between surgical cases and controls were significant ( $P < 0.0001$ , Fisher's nonparametric permutation test).

\*\*Within-group comparisons (Fisher's nonparametric permutation test) of change between baseline and two-year follow-up were significant in the surgical group ( $P < 0.0001$ ), while comparisons in the control group were nonsignificant.

original and modified response formats have been examined.<sup>20</sup>

**Mental well-being/mood disorders.** Mental well-being was measured by the mood adjective check list (MACL).<sup>21</sup> The short version of the MACL applied in the SOS study contains 38 adjectives on a four-point response scale with two acceptance and two rejection categories. Items are grouped into three bipolar dimensions; pleasantness/unpleasantness (for example, satisfied, optimistic/depressed, resigned), activation/deactivation (for example, alert, active/passive, apathetic) and calmness/tension (for example, relaxed/tense, distressed). An overall mood score is also calculated. The MACL has been widely applied in evaluation of mental health status in patients with chronic conditions.<sup>5</sup> The instrument has proven valid in detecting mood changes during weight loss and relapse in dieting obese women.<sup>22</sup>

The hospital anxiety and depression scale (HAD)<sup>23</sup> was used as a complement to the MACL. The HAD instrument is frequently used in screening for psychiatric morbidity in somatically ill patients. It contains 14 items on a four-point response scale, which are summed to separate scores on anxiety and depression. Each individual is also grouped according to a clinically tested classification of psychiatric morbidity, that is, a scale score < 8 is in the normal range, a score 8–10 indicates a possible case and a score > 10 indicates a probable mood disorder. The validity of the Swedish version has been documented in several studies of HRQL in patients with chronic disease/injury.<sup>5</sup>

**Psychosocial functioning.** Social interaction (SI), the main psychosocial category from the sickness impact profile (SIP),<sup>24</sup> was chosen to assess health-related dysfunction in social life. The SIP/SI category contains 20 statements on quality and quantity of social interaction within the family, among friends and in the community. Respondents are asked to indicate only those statements that describe a limitation related to their health status 'today'. Items agreed to are summed according to predetermined weights, divided by the sum of weighted scores for all the 20 items and multiplied by 100. Thus, a scale score is obtained expressing the percentage of maximum dysfunction.

A study-specific module was used to assess obesity-related psychosocial problems (OP) in everyday life. The OP-module comprises eight items on a four-point response scale. Subjects are asked to rate how bothered they are by their obesity as regards private gatherings, going to restaurants or community activities, holidays away from home, trying on and buying clothes, bathing in public places and intimate relations with partner. Item content, psychometric properties and descriptive statistics have previously been presented.<sup>5</sup>

**Self-assessment of eating behavior.** Eating behavior was assessed by the three-factor eating questionnaire (TFEQ), which includes 36 items with an agree/disagree response format, 14 items on a four-point response scale and one vertical rating with a 0–10-point scale.<sup>25</sup> All items are dichotomized and aggregated by summated ratings into three main factors: restrained eating, disinhibition and hunger. The restrained eating scale (21 items) measures the amount of intentional restriction of food intake, the disinhibition scale (16 items) concerns the inability to resist food stimuli or social/emotional eating cues, and the hunger scale (14 items) measures subjective feelings of hunger. The reproducibility of the Swedish version was examined under varied conditions with satisfactory results. Clinical relevance of the Swedish version was tested in a two-year follow-up study on the effects of dieting on functioning and well-being in moderately obese women.<sup>22</sup>

### Statistical methods

Differences between groups were tested by nonparametric methods: two-group comparisons were performed by Fisher's permutation test<sup>26</sup> and for comparison of three or more groups the Kruskal-Wallis' ANOVA of mean ranks was used.<sup>27</sup> Tukey's range test was used for *post hoc* testing of differences between mean values.<sup>28</sup> Correlations were tested for significance by Pitman's nonparametric permutation test.<sup>26</sup> The limit for significance was set at the 5% level. Pearson's correlation coefficient and 95% confidence intervals for the mean were calculated for descriptive purposes.<sup>29</sup> Analysis of longitudinal treatment effects on HRQL in surgical cases *vs* controls was performed using Fisher's nonparametric test. Mantel's pooling technique<sup>30</sup> applied to Fisher's nonparametric test was used to adjust for baseline differences in the longitudinal analysis.

## Results

### HRQL in surgical cases and controls prior to and after intervention

Group comparisons showed generally lower levels of HRQL at baseline in surgical than control cases. After intervention, a consistent pattern of change across all HRQL measures was observed. We found radical positive effects in the surgical group, with peak values on average at 6 or 12 months, followed by a slight to moderate reduction after 24 months. Only minor fluctuations in scores during the observation period were noted in the conventionally treated subjects. Small initial improvements were mostly lost after two years, except for a slight decrease in levels of OP (Table 2) and HAD (Table 3). Analysis of treatment effects from baseline to the 24-months follow-up in surgical cases *vs* controls, showed

**Table 2** Psychosocial functioning in surgical cases ( $n=487$ ) and controls ( $n=487$ ), by gender, before intervention and at two-year follow-up

Psychosocial functioning scales	Mean score <sup>a</sup> (95% confidence interval)				P*
	Surgical cases <sup>b</sup>		Control cases <sup>b</sup>		
<b>OP-scale<sup>c</sup></b>					
Before intervention					
Males	1.60	(1.47–1.73)	0.99	(0.87–1.12)	0.0001
Females	1.94	(1.85–2.03)	1.45	(1.36–1.54)	0.0001
Two-year follow-up					
Males	0.60	(0.48–0.72)	0.92	(0.78–1.05)	0.001
Females	0.84	(0.75–0.93)	1.28	(1.18–1.39)	0.0001
Two-year change**					
Males	–1.01	(–1.14––0.87)	–0.07	(–0.17–0.03)	0.0001
Females	–1.10	(–1.19––1.00)	–0.16	(–0.22––0.09)	0.0001
<b>SIP/SI category<sup>d</sup></b>					
Before intervention					
Males	10.4	(8.7–12.0)	8.2	(6.5–10.0)	ns
Females	11.3	(10.2–12.5)	7.4	(6.4–8.4)	0.0001
Two-year follow-up					
Males	7.0	(4.9–9.1)	9.8	(7.7–11.9)	ns
Females	6.2	(5.1–7.3)	8.2	(6.8–9.7)	0.05
Two-year change**					
Males	–3.3	(–5.0 ––1.5)	1.5	(0.2–3.2)	0.001
Females	–5.2	(–6.5 ––4.0)	1.2	(0.2–2.2)	0.0001

<sup>a</sup>High scores represent dysfunction.

<sup>b</sup>There were 160 males and 327 females in each group at baseline. At two year follow-up, 154 male and 323 female surgical cases and 126 male and 271 female control cases remained in the study.

<sup>c</sup>Obesity-related psychosocial problems (scale range 0–3).

<sup>d</sup>Sickness impact profile; Social interaction (scale range 0–100).

\*P values refer to surgical and control group comparison (Fisher's nonparametric permutation test). Ns = not significant.

\*\*Within-group comparisons (Fisher's nonparametric permutation test) of change between baseline and two-year follow-up were significant in the surgical group; OP ( $P < 0.0001$ ) and SIP/SI (females;  $P < 0.0001$ , males;  $P < 0.001$ ). In the control group, two-year change was significant in females (OP;  $P < 0.0001$ , SIP/SI;  $P < 0.05$ ), while changes in male controls were not significant.

highly significant improvements in the surgical group on all HRQL measures ( $P < 0.0001$ ).

**Current health perception.** The surgical candidates rated their health status at baseline substantially worse than controls ( $P < 0.0001$ ). Mean (95% CI) GHRI/CH scale score was 26.9 (26.1–27.7) in the surgical and 29.4 (28.5–30.2) in the control group. The two groups' levels of self-rated health at follow-ups are shown in Figure 1a. At the 6-month follow-up, the current health perception was strongly improved in the surgical group. The positive alteration in subjective health was maintained after one year, while a decrease of about 20% of the initial change score was registered at the two-year follow-up. In the control group, a slight but significant rise in health outlook after 6 months deteriorated after two years. The average (95% CI) GHRI/CH scale score after two years was markedly higher in the surgical cases; 34.3 (33.4–35.1) in surgical cases vs 30.2 (29.4–31.1) in controls ( $P < 0.0001$ ).

**Psychosocial functioning.** Levels of OP in everyday life are presented in Table 2. Baseline scores varied greatly between treatment groups and by gender. The surgical cases reported markedly more psychosocial

dysfunction due to their obesity than controls and the highest level was reported by female surgical subjects. Also, the average score for male surgical candidates was worse than for male controls. As shown in Figure 1b, dramatic improvements in functioning were reported post-surgically. OP-scale scores two years after the operation were reduced on average by 63% in males and 57% in females. Major change was observed in all areas covered by the construct. In the control group, a slight average reduction in scale scores at two-year follow-up (7% in males and 12% in females) was related to weight loss ( $P < 0.0001$ ). Levels of OP after two years were significantly lower in surgical compared to control cases and, in both groups, lower in males than females (Table 2).

Health-related impacts on social relations were measured by the SIP/SI category (Table 2). High levels of dysfunction were noted in both surgery and control subjects at baseline. Female surgical cases reported more limitations than female controls. Change patterns in the two groups during the two-year follow-up are shown in Figure 1c. A substantial decrease in dysfunction six months after surgery, greater in females than males, was followed by a moderate increase. Two years after surgery, SIP/SI scores decreased by one-third in males and almost one-half in females. The SIP/SI score in controls

**Table 3** Mental well-being (MACL, HAD) in surgical cases ( $n=487$ ) and controls ( $n=487$ ) before intervention and at two-year follow-up

Mental well-being scales	Mean score <sup>a</sup> (95% confidence interval)		$P^*$
	Surgical cases <sup>b</sup>	Control cases <sup>b</sup>	
<b>MACL<sup>c</sup></b>			
Before intervention			
Pleasantness/Unpleasantness	2.96 (2.90–3.02)	3.04 (2.98–3.10)	ns
Activation/Deactivation	2.86 (2.81–2.92)	3.01 (2.95–3.07)	0.001
Calmness/Tension	2.90 (2.84–2.96)	2.98 (2.92–3.04)	ns
Two-year follow-up			
Pleasantness/Unpleasantness	3.17 (3.11–3.22)	3.02 (2.95–3.08)	0.001
Activation/Deactivation	3.18 (3.13–3.23)	3.02 (2.96–3.08)	0.0001
Calmness/Tension	3.11 (3.05–3.16)	2.97 (2.91–3.04)	0.01
Two-year change**			
Pleasantness/Unpleasantness	0.21 (0.16–0.26)	–0.04 (–0.09–0.01)	0.0001
Activation/Deactivation	0.32 (0.27–0.37)	0.00 (–0.04–0.05)	0.0001
Calmness/Tension	0.20 (0.15–0.26)	–0.01 (–0.06–0.04)	0.0001
<b>HAD<sup>d</sup></b>			
Before intervention			
Anxiety	6.3 (5.9–6.7)	5.7 (5.3–6.1)	ns
Depression	5.2 (4.9–5.5)	4.5 (4.2–4.8)	0.01
Two-year follow-up			
Anxiety	4.6 (4.2–5.0)	5.1 (4.6–5.5)	ns
Depression	3.0 (2.7–3.3)	4.1 (3.7–4.4)	0.0001
Two-year change**			
Anxiety	–1.7 (–2.0–1.4)	–0.6 (–0.9–0.2)	0.0001
Depression	–2.2 (–2.5–1.9)	–0.4 (–0.6–0.1)	0.0001

<sup>a</sup>High scores represent well-being on mood adjective check list (MACL) and symptoms on hospital anxiety and depression scale (HAD).

<sup>b</sup>At two-year follow-up, 477 and 407 compliers were registered in the surgical and control groups, respectively.

<sup>c</sup>Mood adjective check list (scale range 1–4).

<sup>d</sup>Hospital anxiety and depression scale (range 0–21).

\* $P$  values refer to surgical and control group comparison (Fisher's nonparametric permutation test).

Ns = not significant.

\*\*Within-group comparisons (Fisher's nonparametric permutation test) of change between baseline and two-year follow-up were significant in the surgical group ( $P < 0.0001$ ). In the control group, changes in HAD anxiety ( $P < 0.001$ ) and depression ( $P < 0.01$ ) scores were significant, while comparisons of MACL scores were not significant.

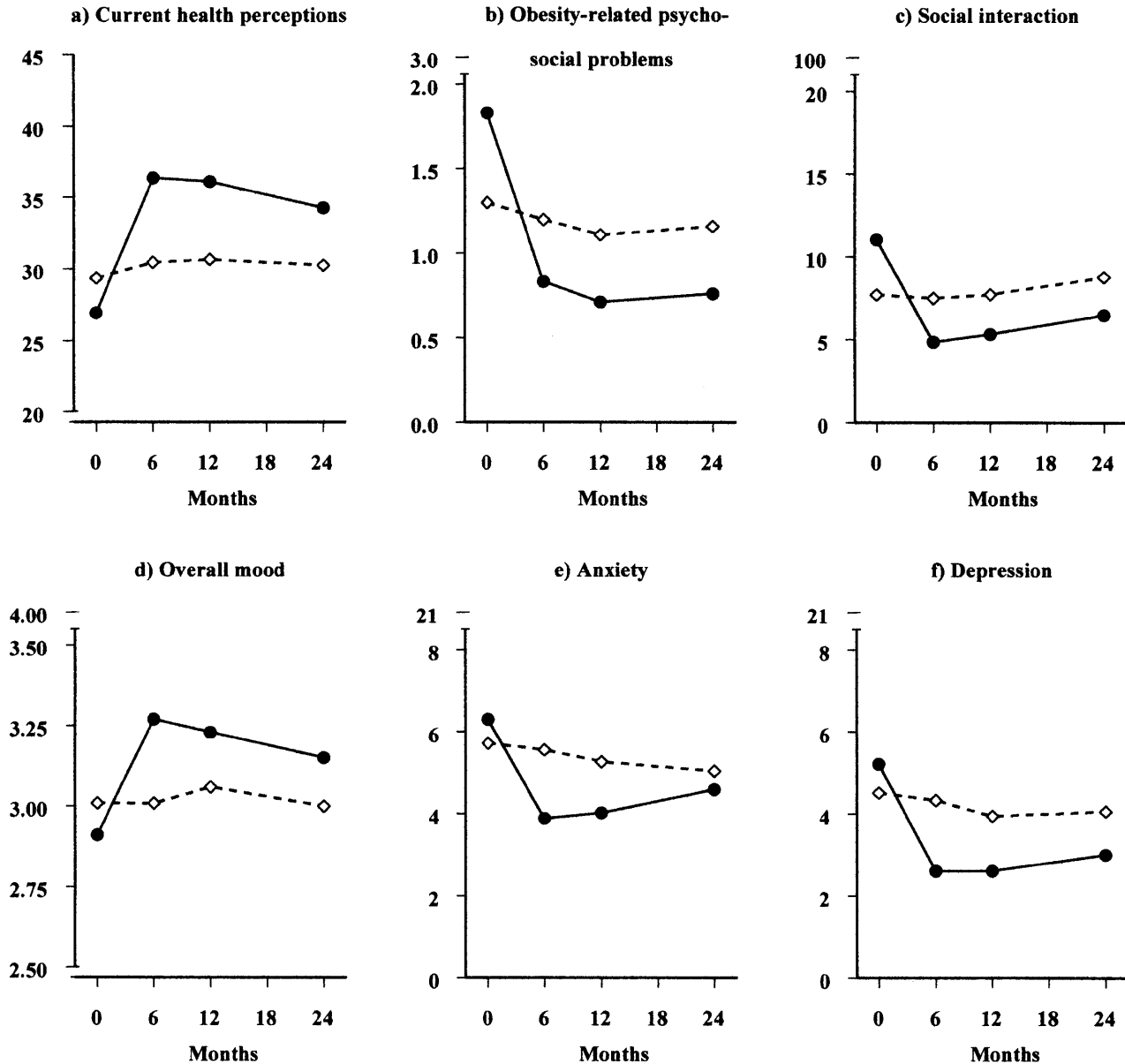
remained stable during the first year, while at the two-year follow-up a minor increase was noted. Health-related dysfunction in social interaction two years after intervention was lower in surgical than in control subjects ( $P < 0.01$ ). However, group comparison by gender showed significantly lower scores only in the female surgical cases (Table 2).

**Mental well-being.** Poor mood ratings were observed in both groups prior to treatment (MACL, Table 3). The activation/deactivation dimension score was significantly worse in the surgical group. The pattern of mood change in surgical and control cases during the two-year follow-up period is shown in Figure 1d. A dramatic rise in emotional well-being, more pronounced in females, was seen in the surgical group at the 6-month follow-up. From 6 to 24 months, the initial improvements decreased by 19% in males and 37% in females. However, the positive change in mental well-being two years after surgery was still substantial and all three mood dimensions, pleasantness/unpleasantness, activation/deactivation and calmness/tension, were significantly higher in the surgical than control group. Only minor changes in mood ratings were noted in controls, that is, a small

rise in overall mood scores at 12 months returned to baseline levels after two years.

High average scores of anxiety and depression were noted at baseline in both surgical and control cases (HAD, Table 3). Males in the surgical group reported somewhat more anxiety symptoms than male controls ( $P < 0.05$ , data not shown) and female surgical cases scored higher than controls on the depression scale ( $P < 0.01$ , data not shown). A substantial reduction of mental distress was observed 6 months after surgery (Figure 1e and f). The decrease in depression symptoms was more pronounced in females. A slight deterioration on both HAD scales was seen in females from 6 to 24 months, while male scores remained stable. Two years after intervention, the surgical cases scored significantly lower than controls on the depression scale, while scores on anxiety were nonsignificantly lower. The decrease in depressive symptoms was more pronounced in females ( $P < 0.05$ , data not shown). Although minor in absolute terms, a reduction in HAD anxiety ( $P < 0.001$ ) and depression ( $P < 0.01$ ) scores was observed from baseline to the 24-month follow-up in the control group. The positive change on the depression scale was related to weight loss ( $P < 0.01$ ), while no such association to the change on the anxiety scale was found.

● Surgical cases  
-◇- Control cases



**Figure 1** Mean scale scores on (a) current health perceptions (GHRI/CH; scale range 9–45), (b) obesity-related psychosocial problems (OP) in everyday life (scale range 0–3), (c) social interaction (SIP/SI; scale range 0–100), (d) overall mood (MACL; scale range 1–4), (e) anxiety and (f) depression (HAD; scale range 0–21) in surgical cases and controls at baseline and at 6, 12 and 24 months after intervention. High scores represent well-being on GHRI/CH and MACL scales, symptoms on HAD scales and dysfunction on OP and SIP/SI scales. GHRI = general health rating index; CH = current health; SIP/SI = social interaction category of sickness impact profile; MACL = mood adjective check list; HAD = hospital anxiety and depression scale.

The prevalence of psychiatric morbidity according to HAD classifications was analyzed. High rates of probable psychiatric morbidity before intervention in the surgical group were decreased by 50% post-operatively. At baseline, 21% of the surgical cases exceeded the cut-off level (scale score > 10) for probable morbid anxiety and at the two-year follow-up this was reduced to 11%. The percentage of surgical cases with probable morbid depression (scale score > 10) was reduced from 9% to 4%. Somewhat lower rates of probable psychiatric morbidity prior to treatment in the control group (17% and

6%, respectively, exceeded the levels for severe anxiety and depression) did not significantly change after two years.

#### HRQL improvements in relation to weight loss after surgical treatment

Relationships between weight loss and HRQL change scores at 6-, 12- and 24-month follow-ups were tested for significance. Six months after surgery, weight reduction was associated only to improvements in OP and SIP/SI scale scores (Table 4). However, the

**Table 4** Correlations between quality of life improvements (HRQL change scores) and weight loss at 6 and 24 month follow-ups after gastric weight reduction surgery

HRQL scales <sup>b</sup>	Weight loss (kg)			
	6 month (n = 456) <sup>a</sup>		24 month (n = 477) <sup>a</sup>	
	r	P	r	P
General health perception				
Δ Current health scale (GHRI/CH) <sup>b</sup>	−0.05	0.2548	−0.18	0.0001
Psychosocial functioning				
Δ Obesity-related problems (OP)	0.21	0.0001	0.30	0.0001
Δ Social interaction (SIP/SI)	0.11	0.0165	0.16	0.0005
Mental well-being				
Δ Overall mood (MACL) <sup>b</sup>	−0.02	0.7188	−0.12	0.0109
Δ Anxiety (HAD)	0.02	0.7031	0.15	0.0012
Δ Depression (HAD)	0.05	0.2847	0.11	0.0178

<sup>a</sup>Pearson's correlation.

<sup>b</sup>Correlations are negative due to inverse scoring on GHRI/CH and MACL scales.

GHRI = general health rating index; CH = current health; MACL = mood adjective check list; OP = obesity related psychosocial problems; SIP/SI = social interaction category of sickness impact profile; HAD = hospital anxiety and depression scale.

associations were stronger after one year and after two years the changes in all HRQL measures were significantly related to the magnitude of weight loss, that is, the greater the weight loss, the greater the quality of life improvements. The decrease of obesity-specific psychosocial problems (OP) in every day life and improvements of current health perceptions (GHRI/CH) were most strongly related to weight loss (Table 4).

The variation in HRQL scores two years after surgery was significantly associated to reduction in overweight. In Figure 2, the surgically treated subjects are grouped by amount of weight loss (kg) two years after surgery and the mean OP-scale scores are plotted for each measurement time point. There were no significant differences between groups at baseline. After six months, levels of psychosocial problems were substantially reduced in all groups, with a more positive trend seen in subjects with major long-term weight reduction. At 12 and 24 months a

distinct pattern of change among groups can be observed, namely, subjects with more favorable long-term weight reduction reported significantly lower levels of obesity-related psychosocial problems. A tendency for scores to return to baseline levels can be seen in groups with the smallest weight loss (less than 20 kg).

In Table 5, two-year follow-up scores (mean, 95% CI) on general health perception, psychosocial functioning and mental well-being are presented for the surgical subjects grouped into two categories of weight loss (< 20 kg and ≥ 20 kg). As shown in Table 5, ratings on all HRQL measures were significantly more positive in subjects with greater weight loss.

#### Self-assessment of eating behavior before and after intervention

Group comparisons of baseline scores on the TFEQ showed less intentional control of food intake

**Table 5** Comparison of health-related quality of life (HRQL) two years after gastric weight reduction surgery between two weight loss categories

HRQL scales <sup>a</sup>	Mean score <sup>a</sup> (95% confidence interval)		P*
	Weight loss < 20 kg (n = 152)	Weight loss ≥ 20 kg (n = 325)	
General health perception			
Current health scale (GHRI/CH)	32.1 (30.7–33.6)	35.2 (34.2–36.2)	0.001
Psychosocial functioning			
Obesity-related problems (OP)	1.16 (1.03–1.29)	0.58 (0.50–0.66)	0.0001
Social interaction (SIP/SI)	8.2 (6.1–10.2)	5.7 (4.5–6.8)	0.05
Mental well-being			
Overall mood (MACL)	3.03 (2.94–3.12)	3.21 (3.15–3.27)	0.001
Anxiety (HAD)	5.5 (4.8–6.3)	4.1 (3.7–4.6)	0.01
Depression (HAD)	3.9 (3.4–4.5)	2.6 (2.3–2.9)	0.0001

<sup>a</sup>High scores represent well-being on GHRI/CH (scale range 9–45) and MACL (scale range 1–4) scales, dysfunction on OP (scale range 0–3) and SIP/SI (scale range 0–100) scales, and symptoms on HAD (scale range 0–21) scales.

\*P values refer to group comparisons (Fisher's nonparametric permutation test).

GHRI = general health rating index; CH = current health; OP = obesity related psychosocial problems; SIP/SI = social interaction category of sickness impact profile; MACL = mood adjective check list; HAD = hospital anxiety and depression scale.

**Table 6** Self-assessment of eating behavior (TFEQ) in surgical cases ( $n=487$ ) and controls ( $n=487$ ) before intervention and at two-year follow-up

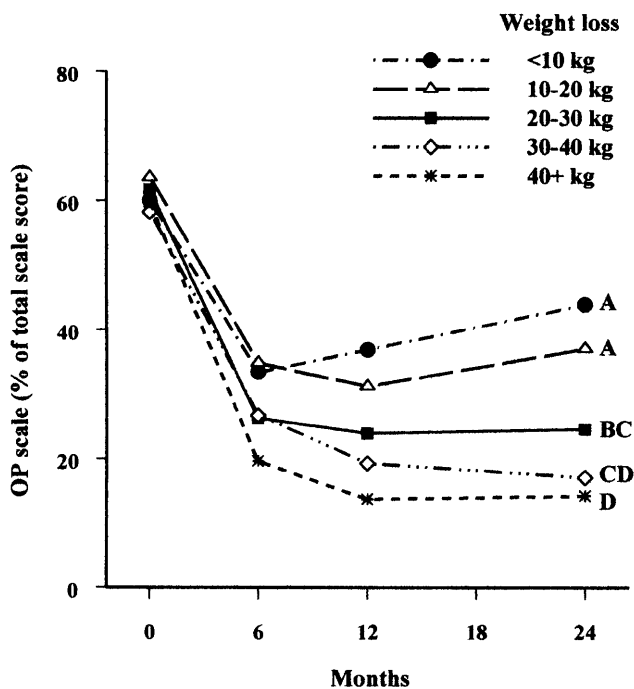
TFEQ <sup>a</sup>	Self-assessment of eating behavior		
	Mean score <sup>a</sup> (95% confidence interval)		P*
	Surgical cases <sup>b</sup>	Control cases <sup>b</sup>	
Before intervention			
Restrained eating	8.8 (8.4–9.1)	9.9 (9.5–10.4)	0.0001
Disinhibition	9.4 (9.1–9.8)	8.5 (8.2–8.8)	0.0001
Hunger	7.2 (6.9–7.5)	6.2 (5.9–6.5)	0.0001
Two-year follow-up			
Restrained eating	12.8 (12.4–13.2)	10.7 (10.2–11.2)	0.0001
Disinhibition	4.9 (4.6–5.1)	8.1 (7.8–8.5)	0.0001
Hunger	3.4 (3.1–3.6)	5.7 (5.4–6.1)	0.0001
Two-year change**			
Restrained eating	4.0 (3.6–4.4)	0.8 (0.4–1.2)	0.0001
Disinhibition	-4.6 (-4.9–-4.3)	-0.5 (-0.8–-0.2)	0.0001
Hunger	-3.9 (-4.2–-3.6)	-0.5 (-0.8–-0.3)	0.0001

<sup>a</sup>High scores represent more restrained eating (scale range 0–21), disinhibition (scale range 0–16) and hunger (scale range 0–14).

<sup>b</sup>At two-year follow-up, 477 and 407 compliers were registered in the surgical and control groups, respectively.

\*P values refer to surgical and control group comparison (Fisher's nonparametric permutation test).

\*\*Within-group comparisons (Fisher's non-parametric permutation test) of change between baseline and two-year follow-up were significant in the surgical group ( $P < 0.0001$ ) and in the control group; restrained eating ( $P < 0.001$ ), disinhibition ( $P < 0.001$ ) and hunger ( $P < 0.0001$ ). TFEQ = three-factor eating questionnaire.



**Figure 2** Obesity-related psychosocial problems (OP) in everyday life before and after gastric restriction surgery, by weight reduction two years after intervention. Mean level at baseline and at 6, 12 and 24 months after treatment are plotted for five groups of patients differing in the amount of weight loss: < 10 kg ( $n=38$ ), 10–20 kg ( $n=114$ ), 20–30 kg ( $n=136$ ), 30–40 kg ( $n=88$ ) and > 40 kg ( $n=101$ ). OP scale scores are transformed into a 0–100 scale, that is, raw scores are divided by the highest possible score and multiplied by 100. High scores represent dysfunction. Groups with different letters are significantly different at two-year follow-up ( $P < 0.05$ ; Tukey's range test).

(restrained eating), more difficulties in resisting eating cues (disinhibition) and more hunger feelings (hunger) in the surgical group (Table 6).

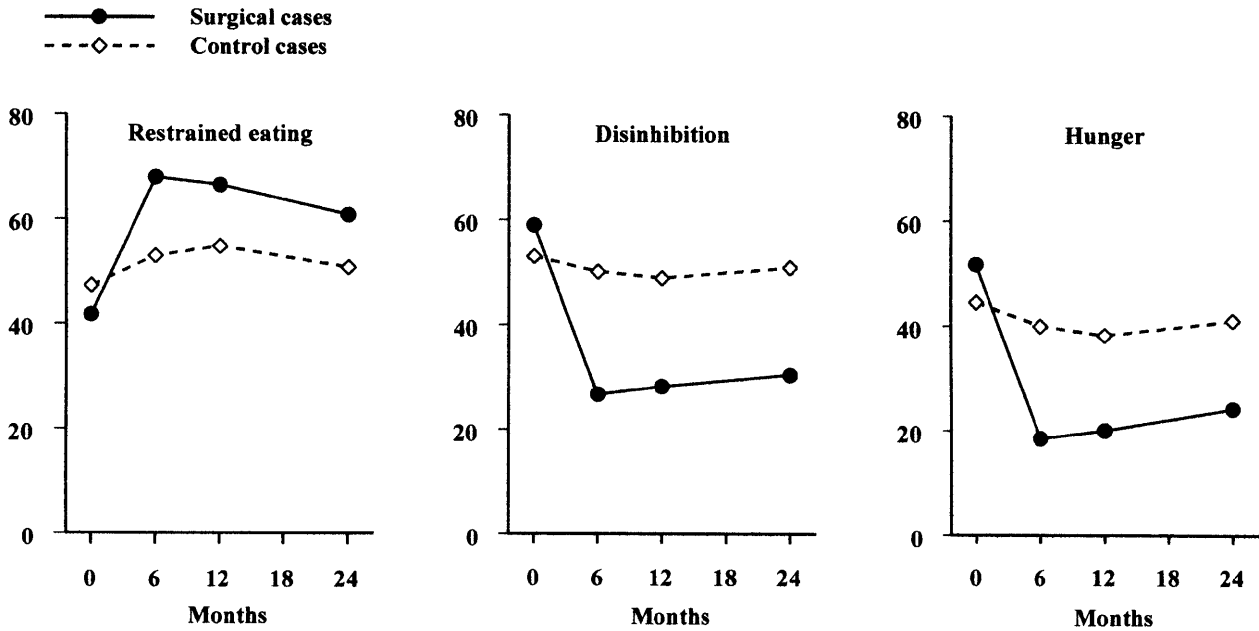
Eating patterns in surgical cases and controls prior to and after treatment are shown in Figure 3. Changes

in eating behavior after intervention followed the hypothesized pattern, with restrained eating increasing and disinhibition and hunger decreasing. Dramatic changes were seen in the surgical group six months after the operation. The restrained eating scale score substantially increased and scores on the disinhibition and hunger scales decreased by almost two-thirds. The long-term changes were consistent with the pattern found after six months. However, a slight return to baseline scores, that is, less restrained eating and somewhat more disinhibition and hunger, were observed at the 24-month follow-up. As shown in Figure 3, only minor changes in self-assessed eating behavior were registered in the control group.

Weight loss in the surgical group was strongly related to the variation in disinhibition and hunger scores at the two-year follow-up, that is, more problems in resisting social/emotional eating stimulus and stronger feelings of hunger were associated with less weight reduction ( $P < 0.0001$ ). Restrained eating scores after two years did not vary with weight loss. However, the two-year change in scale score was significant ( $P < 0.01$ ).

## Discussion

This paper presents data on HRQL before and after treatment in 487 surgical cases and their controls followed for two years in the SOS intervention study. Poor HRQL before intervention was dramatically improved after gastric restriction surgery, while only minor fluctuations in HRQL scores were observed at follow-ups in the conventionally treated controls. Radical long-term improvements in self-



**Figure 3** Self-assessed eating behavior (three-factor eating questionnaire (TFEQ)) at baseline and at 6, 12 and 24 months after intervention in surgical and control cases. Scores on the restrained eating, disinhibition and hunger scales are transformed into a 0–100 scale, that is, raw scores are divided by the highest possible score on each scale and multiplied by 100. High scores represent more restrained eating, disinhibition and hunger.

assessed eating behavior (TFEQ) were also registered in the surgical group. The marked positive effect on HRQL found in the present study is in line with earlier reports on psychosocial change after weight reduction surgery.<sup>4,6,7</sup> Most patients benefit from the operation and report substantial overall improvements in quality of life. This finding supports the assumption that poor well-being and psychosocial dysfunction among the severely obese are in most cases secondary to the obese state and, consequently, can be reversed by long-term weight reduction.

However, long-term improvements after surgery were related to the magnitude of weight loss and subgroups of surgical patients did not manage to lose substantial amounts of overweight; about 8% of the surgical cases had lost < 10 kg and in another 24% weight loss was moderate (10–20 kg). As shown in the present paper, HRQL ratings after two years were significantly more positive in patients with greater weight loss ( $\geq 20$  kg) and a tendency for scores to return to baseline levels was seen in the group with the smaller weight loss (< 20 kg). Patients must adjust to radically new eating patterns after gastric restriction surgery if outcomes are to be successful and coping problems have been reported.<sup>11</sup> It therefore seems important to identify characteristics of patients with high and low weight loss and to define outcome predictors. This issue will be further addressed within the SOS intervention study.

Although substantial effects in the surgical group were observed, it is still important to distinguish between true treatment effects and placebo or general participation effects, since patients will report improvements due to both. The distinct pattern of change in HRQL scores at 6-, 12- and 24-month

follow-ups in the surgical group is noteworthy. Powerful short-term improvements were followed by a slight to moderate long-term decline in average group scores. This pattern of change was most obvious in ratings of emotional well-being. Very low mood scores (MACL) prior to treatment were dramatically improved 6 months after surgery, especially among females, to levels clearly above age-matched general population norms (Persson L-O, unpublished data). This was followed by a decrease and after two years, a substantial part of the initial positive change had diminished. Similar short-term effects on self-assessed mood have been observed in diet-induced weight reduction programs during early weight loss, but when weight is regained the initial effects are likely to vanish.<sup>22,31</sup> In the present study, the dramatic early rise in HRQL scores after surgery is probably influenced by the patients' general positive beliefs in treatment effects. This assumption is supported by the weak relation between HRQL change and clinical change (that is, weight loss) at the 6-month follow-up. At the two-year follow-up, however, the association was much stronger and part of the decline in average group scores could be attributed to decreases in HRQL scores among surgical patients with low weight loss. It is suggested that short-term effects on HRQL indicators in weight loss studies should be interpreted with caution. Repeated post-treatment measurements of HRQL, including long-term follow-ups, are necessary to confirm effects of obesity interventions.

As concluded by Stunkard and Wadden,<sup>4</sup> there is very little and incomplete information on the psychological aspects of severe obesity. Most data on psychosocial status before and after treatment has been

obtained from heavily selected surgical candidates in studies without adequate controls, and thus such data are probably not representative of the general obese population. In SOS, group comparison at baseline showed generally poorer HRQL in surgical than control cases; less positive ratings of subjective health (GHRI/CH), clearly more fatigue (MACL; activation/deactivation), depressive symptoms (HAD) and dysfunction in social interaction (SIP/SI), and markedly more obesity-related psychosocial problems (OP). Baseline scores on self-assessed eating style (TFEQ) showed higher levels of disinhibition and perceived hunger in the surgical cases, indicating more difficulties in controlling eating behavior. These findings suggest that obese persons who choose surgical treatment suffer from a heavier psychosocial burden of the disease and that improvements in life quality are among the main motives for their choice of treatment. It should also be noted that the surgical cases had a higher probability of improvement than controls; however, when these differences were controlled, the beneficial effects on HRQL among the surgical cases were still in evidence.

In contrast to most weight loss studies, compliance rates in the SOS intervention study were high. In the present analysis 99% of the surgical cases and 84% of the control group were still registered after two years. Furthermore, the proportion of males participating in weight loss studies is generally small and gender differences in baseline characteristics or in response to treatment are sparsely reported in the literature. In SOS, the proportion of male surgical candidates (33%) is greater than in earlier corresponding studies. Males in both treatment groups reported markedly fewer psychosocial problems (OP) in everyday life due to their obesity at baseline. In the surgical group, no significant gender differences were found for general health perception (GHRI/CH), mental well-being (MACL, HAD) and dysfunction in social interaction (SIP/SI), while in controls, lower levels of anxiety symptoms (HAD) were reported by males. Short-term responses to surgical treatment were more positive in females; however, the two-year changes were largely the same.

Severe obesity is a chronic condition found to be associated with a wide range of medical complications as well as poor psychosocial functioning. Although life quality is assumed to be poor among the severely obese, few attempts have been made in obesity research to assess HRQL using standardized instruments.<sup>6,32</sup> Outcomes in treatment studies have focused on weight parameters, ignoring important long-term effects of interventions on mortality, morbidity and quality of life. This has now been clearly recognized and several authors have argued for a broadening of outcome measurement.<sup>33,34</sup> New standards proposed for evaluating the success of obesity interventions include quality of life assessments.<sup>14</sup> However, several methodological issues should be considered when health status measures are included in clinical research. Quality of life has been concep-

tualized and measured in a variety of ways in clinical research and particular attention should be paid to measurement strategy and instrument properties. Both conceptualization and measurement techniques are rapidly evolving and important advances in methods for assessing patient perspectives about functional status, well-being and general health perceptions have been made during the past decade.<sup>35</sup> In recent years, the use of standardized, generic self-assessment instruments in quality of life research has increased dramatically. As these measures are designed to be applicable across diseases and treatments as well as demographic and cultural subgroups, they facilitate comparisons among groups. On the other hand, a wide variety of disease-specific instruments have been developed for numerous medical conditions in order to assess particular clinically discernible impacts on health caused by the specific disease. Because of the restricted and more specialized content of disease-specific measures they are considered more sensitive to clinical change than are generic measures. However, unexpected effects on health status are more likely to be detected by the latter.<sup>36</sup> According to Ware,<sup>35</sup> assessments of health status should include standardized, generic measures as well as measures of symptoms and problems unique to diseases and treatments. Recent achievements in the field of health status assessment have not been sufficiently acknowledged in clinical obesity research, however, much could be gained by using modern HRQL methodology in outcomes evaluation. A few promising attempts have recently been made to validate HRQL measures in obese populations and to assess the impact of obesity on quality of life.<sup>37-39</sup> In the SOS study, we applied a multidimensional approach to characterize the quality of life in severely obese people.<sup>5</sup> The HRQL battery is intended to tap a broad range of health impacts from obesity, and generic instruments or subscales on functioning, well-being and subjective health are supplemented by obesity-specific modules.

The capacity of a measure to detect change in health status, clinically significant effects and the relative efficiency of measures, is an important consideration when analyzing and interpreting scores from serial applications of HRQL. To be useful in evaluative studies, a health status measure must demonstrate both reproducibility and responsiveness. Evidence of reproducibility is obtained when repeated administrations of a measure, yield stable scores in subjects with no change in health status, while responsiveness concerns the capacity of an instrument to reflect true clinical change. The sensitivity of instruments to detect change in health over time has been studied far less than other aspects of validity.<sup>40</sup> A controlled study, such as the SOS, provides a unique opportunity for interpreting the reliability of observed changes since both within-subject and between-subject change can be analyzed. If the direction and magnitude of improvements or

deteriorations for groups hypothesized to differ in outcome are as expected, this can be deduced as evidence of validity.

In the present two-year analysis, the reproducibility of HRQL measures was generally supported by stable ratings at repeated points of measurement in the control group. Long-term weight reduction is not expected in controls and quality of life indicators are hypothesized to be more or less stable longitudinally. No long-term changes in scores on general health (GHRI/CH) and mood (MACL) scales were observed in controls, while a slight increase in health-related impacts on social interactions (SIP/SI) was registered. There is, however, individual variability in response to conventional treatment and some control cases will lose weight. Minor long-term reductions in obesity-related psychosocial problems (OP) and depression symptoms (HAD) were related to weight loss in controls; however, the decrease in anxiety (HAD) scores was not associated with weight reduction.

Evidence of responsiveness of measures was obtained by the powerful long-term effects in the surgical group. There were, however, differences in the magnitude of change indicated by HRQL measures after surgical treatment. Several methods have been suggested for assessing the relative efficiency of health status measures to reflect clinical change. A common method is to compare longitudinal change in scores with established criteria of clinical change. When analyzing treatment effects of weight reduction on quality of life, correlations to weight loss may be used. In this study, the disease-specific OP scale was the most responsive. Ratings of subjective health (GHRI/CH) were also responsive to reduction in overweight, which is in line with research findings on self-rated health obtained in clinical and public health areas.<sup>41</sup> Measures of mental well-being/mood disorders (MACL, HAD) and social interaction (SIP/SI) were somewhat less responsive to weight loss. Our results indicate a major advantage of a multidimensional approach, where physical, mental and psychosocial domains are clearly defined, in assessing treatment effects on HRQL in obesity research.

Although the analyses of treatment effects after two years, in surgical cases *vs* controls, showed highly statistically significant improvements in the surgical group, it may still be difficult to interpret the clinical relevance of the change in each of the specific HRQL domains. This is especially true in a study like the SOS where the large number of subjects studied increases statistical power and, thus, even minor group differences are likely to be detected. The actual meaning of a difference in instrument scores and the interpretation of HRQL change in longitudinal studies is much debated<sup>40,42-44</sup> and this important issue will certainly receive increased attention in the years to come. The problem of defining clinically meaningful effect sizes is not unique to HRQL research, but is partly a reflection of the newness of the research area and our inexperience with the con-

cepts and measures.<sup>42</sup> HRQL instruments have arbitrary measurement units which have no direct clinical or commonly-understood meaning. The interpretability of HRQL data is further hampered by the great variation in scaling techniques among instruments.

Several solutions to the problem of interpreting differences in HRQL scores have been discussed in the literature. First, the interpretability of scale scores could be simplified by transforming the raw scale scores to a 0-100 scale.<sup>45</sup> Transformed scale scores represent the percentage of the total possible score. A 5-point difference on a standardized 100-point health status scale has been proposed as a clinically relevant effect size.<sup>45</sup> In the present study, transformed change scores after two years were all of this magnitude or greater in the surgical group, while all changes in the control group were less than 5 points. The average reduction of obesity-related psychosocial problems (OP), two years after surgery corresponds to a 36-point change and the improvements in self-rated health (GHRI/CH) correspond to a 21-point change on a 100-point scale. Transformed change scores on the mental health scales (MACL, HAD) were between 7 and 11 points on average. This comparison indicates major positive outcomes two years after obesity surgery on psychosocial functioning and general health perceptions, while the effect on mental well-being is moderate. However, treatment effects on HRQL were more pronounced in the group of surgical subjects with greater weight loss ( $\geq 20$  kg).

Another commonly used interpretation strategy is to relate scores in the study sample to the distribution of scores derived in other populations (known groups). Comparative HRQL data for a large sample of severely obese males and females in the SOS registry study have earlier been presented together with score levels for healthy reference subjects and chronically diseased or injured patient groups.<sup>5</sup> Available comparisons (HAD, MACL, SIP/SI) between the surgical cases in the present study and the healthy reference subjects indicate markedly improved mental health status after intervention in the surgical subgroup with substantial weight loss ( $\geq 20$  kg). Average two-year follow-up scores on HAD anxiety and depression in this subgroup are in the range of the reference group, while the overall mood (MACL) score is somewhat lower. However, the average level of dysfunction in social interaction (SIP/SI) at the two-year follow-up are higher in the surgical cases than in healthy reference subjects. A third strategy used in the interpretation of HRQL data is to compare the study group with general population norms. Norm-based interpretation demands well-defined and representative population data and the SOS study is currently being supplemented by a general population study.

Factors not taken into account in the present analysis, but which are important in the overall evaluation of life quality after weight reduction surgery, include physical functioning, post-operative complications

and morbidity rates. It should also be noted that results presented in this paper are based on data from subjects aged 37–60 and baseline levels as well as longitudinal change may be different in younger obese persons.

We conclude that poor quality of life in the severely obese can be substantially improved by long-term weight reduction. However, results are based on patient reports two years after surgery, which should be regarded as an intermediate period for evaluating treatment effects. It is of vital importance to ensure that improvements in health-related quality of life after gastric restriction surgery are permanent. Forthcoming analyses of HRQL data in the SOS intervention study will address this issue.

### Acknowledgments

We are grateful for the statistical advice provided by Nils-Gunnar Pehrsson and for helpful comments on the manuscript from Charles Taft. This project has been supported by the Swedish Council for Social Research (project 97-0355:1b), the Swedish Foundation for Health Care Sciences and Allergy Research (project V96 065), the Swedish Medical Research Council and the Faculty of Medicine, Göteborg.

### References

- 1 Sjöström L. Impacts of body weight, body composition, and adipose tissue distribution on morbidity and mortality. In: Stunkard AJ, Wadden TA (eds). *Obesity: Theory and Therapy* (2nd edn). Raven Press: New York, 1993, pp.13–41.
- 2 Rand CSW, MacGregor AMC. Morbidly obese patients' perceptions of social discrimination before and after surgery for obesity. *South Med J* 1990; **83**: 1390–1395.
- 3 Sobal J. Obesity and nutritional sociology: a model for coping with the stigma of obesity. *Clin Soc Rev* 1991; **9**: 125–141.
- 4 Stunkard AJ, Wadden TA. Psychological aspects of severe obesity. *Am J Clin Nutr* 1992; **55**: 524S–532S.
- 5 Sullivan M, Karlsson J, Sjöström L, Backman L, Bengtsson C, Bouchard C, Dahlgren S, Jonsson E, Larsson B, Lindstedt S, Nöslund I, Olbe L, Wedel H. Swedish Obese Subjects (SOS)—an intervention study of obesity. Baseline evaluation of health and psychosocial functioning in the first 1743 subjects examined. *Int J Obes* 1993; **17**: 503–512.
- 6 Kral JG, Sjöström L, Sullivan MBE. Assessment of quality of life before and after surgery for severe obesity. National Institutes of Health Consensus Development Conference on Gastro-Intestinal Surgery for Severe Obesity, March 25–27, 1991. *Am J Clin Nutr* 1992; **55**: 611S–614S.
- 7 Stunkard AJ, Stinnett JL, Smoller JW. Psychological and social aspects of the surgical treatment of obesity. *Am J Psych* 1986; **143**: 417–429.
- 8 Mason EE. Vertical banded gastroplasty for obesity. *Arch Surg* 1982; **117**: 701–706.
- 9 Hawke A, O'Brien P, Watts JM, Hall J, Dunstan RE, Walsh JF, Slavotinek AH, Elmslie RG. Psychosocial and physical activity changes after gastric restrictive procedures for morbid obesity. *Aust NZ J Surg* 1990; **60**: 755–758.
- 10 Larsen F. Psychosocial function before and after gastric banding surgery for morbid obesity. A prospective psychiatric study. *Acta Psychiatr Scand* 1990; **82** (Suppl): 1–57.
- 11 Hafner RJ, Watts JM, Rogers J. Quality of life after gastric bypass for morbid obesity. *Int J Obes* 1991; **15**: 555–560.
- 12 Bull RH, Legorreta G. Outcome of gastric surgery for morbid obesity: weight changes and personality traits. *Psychother Psychosom* 1991; **56**: 146–156.

- 13 Wölfel R, Günther K, Rümenapf G, Koerfgen P, Husemann B. Weight reduction after gastric bypass and horizontal gastroplasty for morbid obesity. Results after 10 years. *Eur J Surg* 1994; **160**: 219–225.
- 14 Atkinson RL. Proposed standards for judging the success of the treatment of obesity. *Ann Intern Med* 1993; **119**: 677–680.
- 15 Ware JE. Standards for validating health measures: definition and content. *J Chronic Dis* 1987; **40**: 473–480.
- 16 Sullivan M. Quality of life assessment in medicine: concept, definitions, purposes and basic tools. *Nord J Psychiatry* 1992; **46**: 79–83.
- 17 Sjöström L, Larsson B, Backman L, Bengtsson C, Bouchard C, Dahlgren S, Hallgren P, Jonsson E, Karlsson J, Lapidus L, Lindroos A-K, Lindstedt S, Lissner L, Narbro K, Näslund I, Olbe L, Sullivan M, Sylvan A, Wedel H, Ågren G. Swedish Obese Subjects (SOS). Recruitment for an intervention study and a selected description of the obese state. *Int J Obes* 1992; **16**: 465–479.
- 18 Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics* 1975; **31**: 103–115.
- 19 Davies AR, Sherbourne CR, Peterson JR, Ware Jr JE. Scoring manual: Adult health status and patient satisfaction measures used in RAND's health insurance experiment. The RAND Corporation: Santa Monica, CA, (No. N-2190-HHS), 1988.
- 20 Karlsson J, Sjöström L, Sullivan M. Swedish Obese Subjects (SOS)—an intervention study of obesity. Measuring psychosocial factors and health by means of short-form questionnaires. Results from a method study. *J Clin Epidemiol* 1995; **48**: 817–823.
- 21 Sjöberg L, Svensson E, Persson LO. The measurement of mood. *Scand J Psychol* 1979; **20**: 1–18.
- 22 Karlsson J, Hallgren P, Kral J, Lindroos AK, Sjöström L, Sullivan M. Predictors and effects of long-term dieting on mental well-being and weight loss in obese women. *Appetite* 1994; **23**: 15–26.
- 23 Zigmund AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983; **67**: 361–370.
- 24 Bergner M, Bobitt RA, Carter WB, Gilson BS. The sickness impact profile: development and final revision of a health status measure. *Med Care* 1981; **19**: 787–805.
- 25 Stunkard AJ, Messick S. The three-factor eating questionnaire to measure dietary restraint, disinhibition and hunger. *J Psychosom Res* 1985; **29**: 71–83.
- 26 Bradley JW. *Distribution-free statistical tests*. Prentice-Hall: London, 1968, pp 68–86.
- 27 Siegel S. *Nonparametric statistics for the behavioral sciences*. McGraw-Hill: New York, 1956, pp 184–194.
- 28 Winer BJ. *Statistical principles in experimental design*. MacGraw-Hill: London, 1970, pp 46–104.
- 29 Gardner MJ, Aldman DG (eds). *Statistics with confidence*. British Medical Association: London, 1989.
- 30 Mantel N. Chi square test with one degree of freedom: extensions of the Mantel-Haenszel procedure. *J Am Stat Assoc* 1963; **58**: 690–700.
- 31 Wing RR, Epstein LH, Marcus MD, Kupper DJ. Mood changes in behavioral weight loss programs. *J Psychosom Res* 1984; **28**: 189–196.
- 32 Sullivan MBE, Sullivan LGM, Kral JG. Quality of life assessment in obesity: physical, psychological and social function. *Gastroenterol Clin North Am* 1987; **16**: 433–442.
- 33 Brownell KD, Wadden TA. Etiology and treatment of obesity: understanding a serious, prevalent, and refractory disorder. *J Consult Clin Psychol* 1992; **60**: 505–517.
- 34 Wardle J. The assessment of obesity: Theoretical background and practical advice. *Behav Res Ther* 1995; **33**: 107–117.
- 35 Ware JE. The status of health assessment 1994. *Ann Rev Pub Health* 1995; **16**: 327–354.
- 36 Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. *Med Care* 1989; **27**: Supplement, S217–S232.

- 37 Kolotkin R, Head S, Hamilton M, Tse C. Assessing impact of weight on quality of life. *Obes Res* 1995; **3**: 49–56.
- 38 Fontaine K, Cheskin L, Barofsky I. Health-related quality of life in obese persons seeking treatment. *J Fam Practice* 1996; **43**: 265–270.
- 39 Mathias S, Williamson C, Colwell H, Cisternas M, Pasta D, Stolshek B, Patrick D. Assessing health-related quality-of-life and health state preference in persons with obesity: a validation study. *Qual Life Res* 1997; **6**: 311–322.
- 40 Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Strategies for evaluation. *Contr Clin Trials* 1991; **12**: 142S–158S.
- 41 Björner JB, Sondergaard Kristensen T, Orth-Gomér K, Tibblin G, Sullivan M, Westerholm P. *Self-rated health—a useful concept in research, prevention and clinical medicine*. Swedish council for planning and coordination of research: Report 96:9, 1996.
- 42 Lydick EG, Epstein RS. Clinical significance of quality of life data. In: Spilker B (ed). *Quality of life and pharmacoeconomics in clinical trials* (2nd edn). Lippincott-Raven Publishers: Philadelphia, 1996, pp 461–465.
- 43 Editorial. Quality of life and clinical trials. *Lancet* 1995; **346**: 1–2.
- 44 Staquet M, Berzon R, Osoba D, Machin D. Guidelines for reporting results of quality of life assessments in clinical trials. *Qual Life Res* 1996; **5**: 496–322.
- 45 Ware JE, Snow KK, Kosinski M, Gandek B. *SF-36 health survey. Manual and interpretation guide*. Health Institute, New England Medical Center: Boston, MA, 1993.