

# Systematic review and meta-analysis of controlled and prospective cohort efficacy studies of endoscopic radiofrequency for treatment of gastroesophageal reflux disease

Ronnie Fass<sup>1</sup> · Frederick Cahn<sup>2</sup> · Dennis J. Scotti<sup>3</sup> · David A. Gregory<sup>4</sup>

Received: 9 October 2016 / Accepted: 20 January 2017 / Published online: 23 February 2017  
© Springer Science+Business Media New York 2017

## Abstract

**Background** The endoscopic radiofrequency procedure (Stretta) has been used for more than a decade to treat patients with gastroesophageal reflux disease (GERD). However, the efficacy of the procedure in improving objective and subjective clinical endpoints needs to be further established.

**Aim** To determine the efficacy of the Stretta procedure in treating patients with GERD, using a systematic review and meta-analysis of controlled and cohort studies.

**Methods** We conducted a systematic search of the PubMed and Cochrane databases for English language clinical studies of the Stretta procedure, published from inception until May 2016. Randomized controlled trials (RCTs) and cohort studies that included the use of the Stretta procedure in GERD patients were included. A generalized inverse weighting was used for all outcomes. Results were calculated by both fixed effects and random effects model.

**Results** Twenty-eight studies (4 RCTs, 23 cohort studies, and 1 registry) representing 2468 unique Stretta patients were included in the meta-analysis. The (unweighted) mean follow-up time for the 28 studies was 25.4 [14.0,

36.7] months. The pooled results showed that the Stretta reduced (improved) the health-related quality of life score by  $-14.6$  [ $-16.48$ ,  $-12.73$ ] ( $P < 0.001$ ). Stretta also reduced (improved) the pooled heartburn standardized score by  $-1.53$  [ $-1.97$ ,  $-1.09$ ] ( $P < 0.001$ ). After Stretta treatment, only 49% of the patients using proton pump inhibitors (PPIs) at baseline required PPIs at follow-up ( $P < 0.001$ ). The Stretta treatment reduced the incidence of erosive esophagitis by 24% ( $P < 0.001$ ) and reduced esophageal acid exposure by a mean of  $-3.01$  [ $-3.72$ ,  $-2.30$ ] ( $P < 0.001$ ). Lower esophageal sphincter (LES) basal pressure was increased post Stretta therapy by a mean of  $1.73$  [ $-0.29$ ,  $3.74$ ] mmHg ( $P = \text{NS}$ ).

**Conclusions** The Stretta procedure significantly improves subjective and objective clinical endpoints, except LES basal pressure, and therefore should be considered as a viable alternative in managing GERD.

**Keywords** GERD · Heartburn · PPI · Endoscopy · Radiofrequency ablation · pH test

Gastroesophageal reflux disease (GERD) is very common affecting 20% of the adult population in the US weekly and 7% daily [1, 2]. The prevalence estimates of GERD are 18.1–27.8% in North America and 8.8–25.9% in Europe [3]. The incidence of GERD per 1000 person-years is about 5.0 in the UK and US populations [3]. Overall, in the US, GERD is the most common outpatient diagnosis in gastroenterology and is associated with a significant burden on the healthcare system.

Presently, there are medical, endoscopic and surgical therapeutic modalities for GERD. While medical therapy remains the most popular therapeutic approach for GERD, recent years have seen a marked shift in the development

✉ Ronnie Fass  
ronnie.fass@gmail.com

<sup>1</sup> The Esophageal and Swallowing Center, Division of Gastroenterology and Hepatology, Department of Medicine, The MetroHealth Medical Center and Case Western Reserve University, 2500 MetroHealth Drive, Cleveland, OH 44109, USA

<sup>2</sup> BioMedical Strategies, La Jolla, CA, USA

<sup>3</sup> Alfred E. Driscoll Professor of Healthcare & Life Sciences Management, Fairleigh Dickinson University, Teaneck, NJ, USA

<sup>4</sup> Baker Tilly Virchow Krause, LLC, New York, NY, USA

of therapeutic modalities for GERD, with a focus on non-medical techniques [1]. Those primarily include novel endoscopic and surgical procedures [4]. The Stretta procedure, a radiofrequency (RF) application to the lower esophageal sphincter (LES), was introduced about 15 years ago as an alternative to chronic medical therapy or surgical intervention for GERD [5]. Since the initial introduction of the Stretta RF system, several improvements have been implemented that ensure ease of use and proper application of the technique.

The Stretta procedure appears to result in thickening of the LES, decreased transient LES relaxation rate and reduced esophageal acid exposure [5]. Concerns about adverse events (AEs), such as esophageal stricture or neurolysis, have been refuted over time [6]. Two recent meta-analyses evaluated the impact of the Stretta procedure on objective and subjective clinical endpoints with conflicting results [6, 7]. Perry et al. performed a meta-analysis of randomized controlled trials (RCTs) and cohort studies that included a total of 1441 patients from 18 studies [6]. The authors demonstrated that the Stretta procedure significantly improved heartburn scores, GERD–health-related quality of life (HRQL), esophageal acid exposure and DeMeester score. Lipka et al. also performed a meta-analysis of only RCTs of the Stretta procedure [7]. The analysis included 165 patients from four trials [3 Stretta vs. sham and 1 Stretta vs. proton pump inhibitor (PPI) treatment]. The pooled results of this study revealed no difference between the clinical outcome of the Stretta procedure and sham or PPI treatment, although the authors cautioned that the quality of the evidence was poor. There was no significant improvement in % time pH less than 4, LES basal pressure, ability to stop PPI or HRQL [7]. The authors concluded that the Stretta procedure does not produce significant clinical or physiological changes as compared with sham therapy.

Due to the variable designs and inconsistent results of the aforementioned meta-analyses, the aim of our study was to determine the efficacy of the endoscopic RF procedure (Stretta), using a systematic review and meta-analysis of all currently available controlled and cohort studies.

## Methods

### Data sources and search strategy

We performed a literature search to identify all controlled trials and cohort studies of Stretta therapy for the relief of symptoms associated with GERD. Electronic databases (PubMed and Cochrane Central Registry of Controlled Trials) were queried from inception of the Stretta procedure (year 2000) to 18 May 2016 using the key words:

(1) “Stretta” and “endoscopic radiofrequency” and (2) the combination of (“Stretta” OR “radiofrequency”) AND (“GERD” OR “gastroesophageal reflux”) present in the title, abstract, or any fields. The reference list of the studies identified was then hand-searched to identify additional studies that may have been missed in the initial search.

### Selection criteria

All potentially relevant articles were examined to determine their eligibility using the following inclusion criteria: (1) at least 3 months follow-up, (2) study design was controlled trial or cohort study and (3) sufficient data for at least one of the six selected outcome variables (defined below). Exclusion criteria included patients from special populations (e.g. obese, paediatric or gastroparesis patients), (2) patients undergoing combined treatment modalities, (3) letters, editorials, review articles and animal studies and (4) non-English publications. Overall, 28 studies met the selection criteria, and each was crosschecked with studies included in previous Stretta meta-analyses to ensure that all relevant studies had been captured. The present systematic review and meta-analysis were conducted using a protocol consistent with the guidelines described in the Cochrane Handbook for Systematic Reviews of Interventions [8] and the PRISMA statement [9].

### Study outcomes

The outcomes of interest were the relief of symptoms associated with GERD. Because the included studies reported GERD relief using several different variables and multiple scoring systems, we have identified three self-reported symptom variables and three physiological markers that appear with sufficient frequency in the studies to enable meta-analysis: (1) PPI use, (2) GERD–HRQL, (3) heartburn score, (4) presence of erosive esophagitis, (5) esophageal acid exposure and (6) LES basal pressure.

PPI use was by far the most commonly reported outcome measure for which data were available in 23 studies. Data from the HRQL instrument—a validated scale for GERD symptom relief that ranges from 0 (asymptomatic) to 50 (incapacitating symptoms) [10, 11]—were reported in 11 studies. Heartburn score, perhaps a more specific measure of GERD symptoms, was reported in 13 studies. Because these studies used several different scales, we standardized the heartburn scores [12] before testing for a statistically significant effect of Stretta treatment. Erosive esophagitis data were extracted from 12 studies that performed upper endoscopy at baseline and follow-up. Studies either used the Los Angeles or the Savary–Miller classification; therefore, only the presence of any of these mucosal abnormalities was considered. Esophageal acid exposure

was reported in either percentage total time  $\text{pH} < 4.0$  over a 24-h period or DeMeester score. Eleven studies reported esophageal acid exposure time and eight studies reported DeMeester score. Finally, nine studies reported LES basal pressure (mmHg).

### Assessment of risk of bias

Risk of methodological bias (quality) of the RCT studies was assessed using the criteria from the Cochrane Handbook [13]: randomization risk, allocation concealment risk, performance (blinding) risk, detection bias risk, attrition bias risk, and reporting bias risk.

We used a modified Newcastle–Ottawa quality assessment scale [14, 15] to evaluate the quality of the 23 prospective cohort studies and 1 registry that are included in our analysis. There are some important distinctions between the cohort studies we included in this analysis and the type of study for which the Newcastle–Ottawa tool is optimally designed. Similar to the RCT studies, our cohort studies are not retrospective observational studies with comparative groups, but prospective interventional studies with a pre/post-experimental design; “with exposure” being defined as Stretta treatment and the non-exposed cohort being the baseline measurement before treatment. The specific criteria and scoring scheme are detailed in Appendix 1.

Funnel plots were used to visualize asymmetric patterns suggestive of publication bias or other small study effects.

### Data extraction and calculations

For continuous data, each outcome measure was expressed as the mean change from baseline to the longest follow-up. The variation in the mean change was expressed as standard error of the mean (SE) and, for studies not reporting SE, we converted them to SE [16]: when standard deviation ( $s$ ) was reported, we divided  $s$  by  $\sqrt{N}$  to obtain the SE. For the data expressed as mean and confidence interval, we calculated the SE as the confidence interval divided by twice the inverse  $t$ -distribution at a significance of 0.05 and with  $n$  degrees of freedom. (For large  $N$ , the divisor is approximately 3.92.) For data reported as median and range, the mean was estimated as the mean of the median and the limits of the range while the standard deviation was estimated using the approximation that  $s = \text{range}/4$ . Finally, for data reported as median and interquartile range, the mean and  $s$  were estimated by the method of Hozo et al. [17].

When the mean changes in outcome measures from baseline to follow-up were not directly reported, we calculated them by subtraction and calculated the SE of the difference using a general propagation of errors method [13, 16]: the variance of the mean of each baseline and follow-up measurement was calculated by squaring its SE, and

the variances added to estimate the variance of the mean change from baseline. This variance was adjusted for the expected correlation between baseline and follow-up measurements by multiplying by  $(1 - \text{correlation coefficient})$ , using an imputed value of 0.5 for the correlation coefficients since they were unknown [16, 18]. The square root of the adjusted variance was then taken to obtain the SE of the mean change.

Because the heartburn data were derived from different measurement scales, we standardized the mean change from baseline using Review Manager (RevMan) software. Erosive esophagitis and PPI use were quantified as discrete data and expressed as patient counts in the studies we analysed. We extracted data on the total number of patients and the number of those patients with erosive esophagitis at baseline, and the total number of patients and the number of those patients with erosive esophagitis at follow-up. We extracted these data from 12 studies, including 4 studies that compared sham treatment to Stretta treatment and 8 cohort studies. We chose to treat PPI use as a binary variable (use or no use) because of the inconsistency in reporting among studies. We extracted data on the total number of patients and the number of those patients using PPI at baseline, and the total number of patients and the number of those patients that were using PPI at follow-up or were lost to follow-up (intent-to-treat analysis). Many studies reported the number of patients not taking PPI, and we obtained the number of patients taking PPI by subtraction. We extracted PPI data from 23 studies.

We analysed dichotomous data as risk ratios using the equation provided in the Cochrane Handbook [13]. The risk ratio is equal to the quotient of the fraction of patients with symptoms of esophagitis or using PPI at follow-up and the fraction at baseline (before treatment).

### Subgroup and sensitivity analyses

Prespecified subgroup analyses of RCTs and cohort studies were conducted. Subgroups comprised Stretta in cohort or laparoscopic fundoplication (LF) comparative studies, Stretta in RCT studies, control (sham or PPI treatment) in RCT studies and LF in LF comparative studies. Mean change or risk ratio data (“treatment effects”) were pooled for each subgroup by fixed effects or random effects models (as specified). Where data permitted, we tested the hypothesis that the pooled treatment effect of Stretta was significantly different from the pooled treatment effect of control (sham or PPI treatment) or that the pooled treatment effect was significantly different between Stretta treatment subgroups (cohort vs. RCT). We did not test LF treatment with either Stretta or control treatments.

Heterogeneity was assessed for the pooled estimates. Sensitivity of results to study quality was planned for

analysis where treatment effect findings are equivocal. We also performed meta-regression to explore the potential influence of duration of follow-up on the various outcomes' treatment effects.

### Statistical analysis

All conversions of data to means and SEs were performed with IBM SPSS version 22 for OS X (10.11). The studies were weighted by the inverse variance of the mean difference data such that  $w = \left[ \frac{1}{(SE^2)} \right]$ .

RevMan [for OS X], Ver. 5.3. (Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014) was used to calculate the meta-analyses employing the random effects model, with initial weights derived by the inverse variance method. RevMan 5.3 was also used to generate funnel plots (using the fixed effects model). For hypothesis testing, the primary hypothesis was that the overall mean treatment effect of Stretta on each of the six variables was different than zero for continuous data and not equal to one for dichotomous data.

Results are presented as forest plots with the mean difference from baseline and 95% CIs, and as weighted means (random/fixed effects model) with 95% CIs in brackets. All statistical tests were evaluated at an a priori significance level of 0.05.

## Results

### Study selection and characteristics

Our search strategy yielded a total of 56 publications suitable for full-text screening (Fig. 1). After evaluating each publication for eligibility, 28 studies representing 2468 unique Stretta patients were retained and included in the present meta-analysis. These studies included 4 RCTs (1 multi-centre US study with 8 clinical sites and 3 single-centre RCTs from Belgium, France and the USA) and 23 cohort studies (1 multi-centre US study with 13 clinical sites and 22 single-centre studies from Australia, Belgium, China, France, Germany, Italy, Japan, Puerto Rico and the USA). Additionally, we included an international registry with 33 sites. The characteristics of the included studies are presented in Table 1. Notably, the (unweighted) mean follow-up time for the 28 studies was 25.4 [14.0, 36.7] months.

### Risk of bias

Using the Cochrane Collaboration criteria [13], we evaluated the overall risk of bias for the three RCT studies that

used sham control as low (Table 2). The study with PPI control [19] is subject to higher risk of bias because it could not be blinded. We considered the two studies that did not measure PPI use to be of unclear risk of reporting bias. The overall quality of the cohort studies according to the modified Newcastle–Ottawa scale was relatively high (Table 3). Most studies were representative of the population experiencing GERD and all had direct ascertainment of GERD diagnosis by study investigators. Since these were interventional studies, ascertainment of exposure was always known. Because of the pre/post-treatment design, the non-exposed cohort (i.e. the pretreatment measurement) was also known and comparable with respect to comorbidities.

All 28 studies were assessed against each element of the “PICOTS” framework [20]. We used the framework to assess the feasibility of combining the RCTs and cohort studies for analysis (Appendix 2).

We used funnel plots to evaluate the risk of publication bias for our six outcomes (see Fig. 2). The outcomes do not show substantial evidence of publication bias. Furthermore, because of the large treatment effects we observed, we do not believe that our results would be sensitive to biases, including publication bias from small studies.

### Heterogeneity

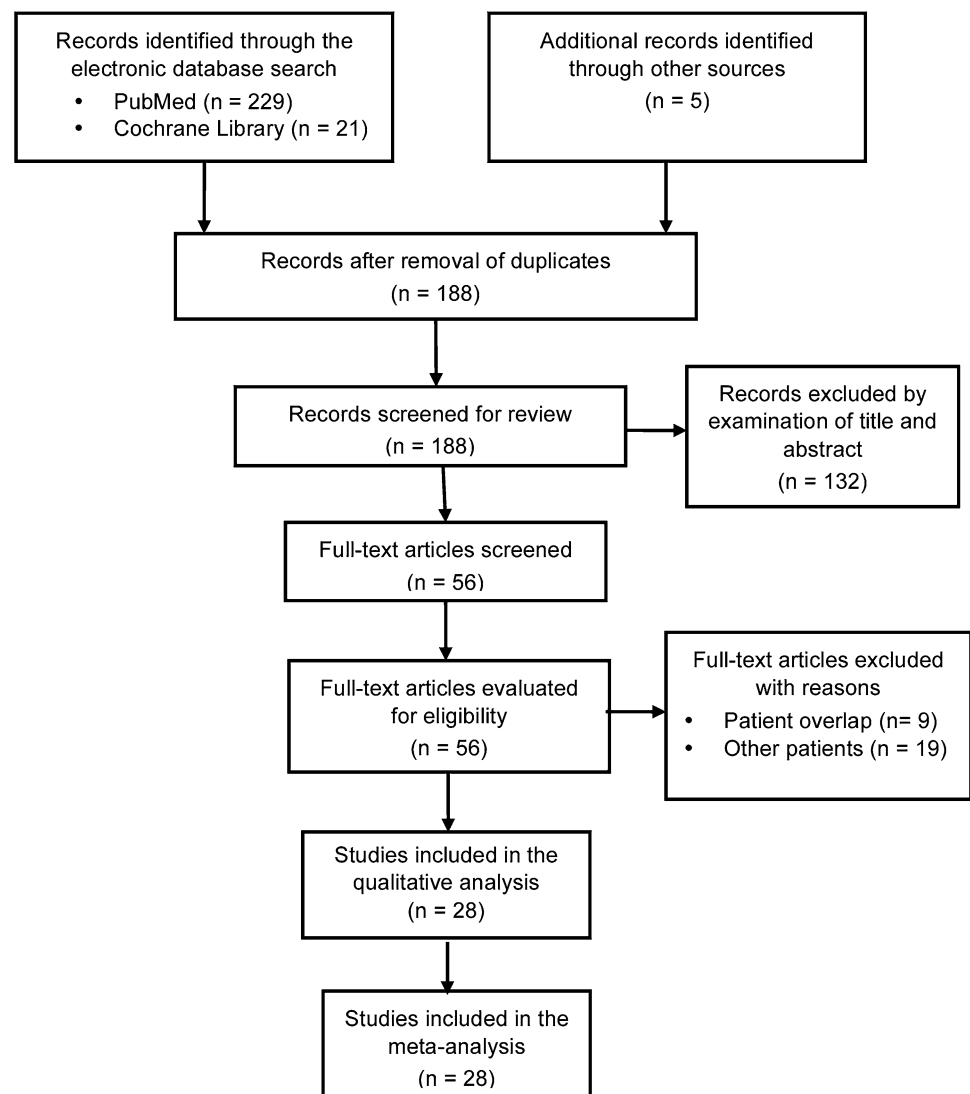
The studies are heterogeneous with respect to inclusion criteria, previous surgeries, nationalities of patients, protocols for the use of antacids, monitoring of PPI use and follow-up time, which ranged from 3 to 120 months. Thus, these data represent a wide range of clinical situations over a clinically significant time. To test the stability of outcomes for follow-up time, we performed weighted linear regression (weights from the random effects model) of the treatment effects versus time for each outcome.

### Meta-analyses of outcomes

#### *Use of proton pump inhibitors*

PPI use is an important outcome. Studies most commonly reported the number of patients who were no longer using PPI, so we derived the number relying on PPI at any frequency of use by subtraction of those patients from the number of patients enrolled (intent-to-treat analysis). We then stratified the data according to the type of study (RCTs, cohort studies, sham or PPI and LF controlled). The 23 studies comprised 1795 Stretta patients. The treatment effect for PPI use was calculated as the risk ratio between baseline and follow-up fractions of patients reliant on PPIs; the risk ratio can be interpreted as the change in the fraction of patients using PPI. These risk ratios are presented as forest plots in Fig. 3.

**Fig. 1** Flow diagram of study search and selection process. This flow diagram of study search and selection process for the Stretta RCT and cohort studies details the progression from the original PubMed, Central and other source research ( $n=255$ ) through duplication removal ( $n=188$ ), eligibility screening ( $n=56$ ) and finally to study inclusion ( $n=28$ ). ( $n$  = number of studies)



Collectively at baseline, 97.1% (1743) of patients in the Stretta group were using PPI. After Stretta treatment, less than half (850) of these patients were using PPI, producing a pooled estimate of the risk ratio of 0.49 [0.40, 0.60] ( $P < 0.001$ ) by the random effects model. LF treatment yielded a pooled estimate of the risk ratio of 0.10 ([0.01, 1.35],  $P = 0.08$ ), and all patients in sham or PPI control groups remained on PPIs (i.e. no treatment effect). The pooled estimate of Stretta treatment effect (0.49 [0.40, 0.60]) was significantly greater than the pooled sham and PPI controls (1.00 [0.92, 1.08]), and when considered alone, the two RCT studies showed a smaller but significant treatment effect for Stretta therapy, with a risk ratio of 0.86 ([0.74, 1.00],  $P = 0.05$ ).

Heterogeneity was not observed within either Stretta or control group in the RCT subgroups or within the Stretta subgroup of the LF comparative studies. Considerable

heterogeneity was observed within the cohort subgroup ( $I^2 = 95\%$ ,  $P < 0.001$ ), and there was a significant difference among the pooled estimated risk ratios among the three Stretta subgroups ( $P < 0.001$ ). A meta-regression of change in risk ratio with follow-up time was not statistically significant ( $P = 0.65$ ).

#### Health-related quality of life

Two RCTs and 9 cohort studies, comprising a total of 507 patients, reported HRQL at baseline and after Stretta treatment. The Stretta and sham treatment effects are the changes in HRQL score from baseline to follow-up. The 95% confidence intervals show that the 11 studies had individually very significant treatment effects for Stretta therapy as reported in Fig. 4. In the 11 studies, Stretta reduced (thus improved) the pooled estimate of the change

**Table 1** Characteristics of the included studies

Studies	Countries	Number of sites	Study design	Sample size	Follow-up (months)
Arts et al. [26]	Belgium	1	Cohort	13	6
Arts et al. [27]	Belgium	1	RCT	22	3
Aziz et al. [28]	USA	1	RCT	12	12
Cipolletta et al. [29]	Italy	1	Cohort	32	12
Corley et al. [30]	USA	8	RCT	35	6
Coron et al. [19]	France	1	RCT	23	12
DiBaise et al. [31]	USA	1	Cohort	18	6
Dughera et al. <sup>a</sup> [32]	Italy	1	Cohort	69	48
Dughera et al. <sup>a</sup> [33]	Italy	1	Cohort	86	96
Dundon et al. [34]	USA	1	Cohort	37	53
Gao et al. <sup>b</sup> [35]	China	1	Cohort	505	12
Go et al. [36]	USA	1	Cohort	50	> 3
Higuchi et al. [23]	Japan	1	Cohort	9	6
Houston et al. [37]	USA	1	Cohort	41	6
Liang et al. <sup>b</sup> [38]	China	1	Cohort	132	60
Liang et al. [21]	China	1	Cohort with LF comparison	85	36
Liu et al. [39]	China	1	Cohort	90	12
Lufti et al. [40]	USA	1	Cohort	86	26
Mansell [41]	USA	1	Cohort	29	4
Meier et al. [42]	Germany	1	Cohort	60	12
Noar and Lofti-Emran [43]	USA	1	Cohort	109	48
Noar et al. [44]	USA	1	Cohort	217	120
Reymunde and Santiago [45]	Puerto Rico	1	Cohort	83	48
Richards et al. [46]	USA	1	Cohort with LF comparison	65	6
Tam et al. [47]	Australia	1	Cohort	20	12
Torquati et al. [48]	USA	1	Cohort	82	27
Triadafilopoulos et al. [22]	USA	13	Cohort	118	12
Wolfsen and Richards [49]	International	33	Registry	558	8

<sup>a</sup>Studies with overlapping patient populations. Esophagitis and adverse events data are from Dughera et al. [32], and HRQL and heartburn data are from Dughera et al. [33]

<sup>b</sup>Studies with overlapping patient populations. Adverse events data are from Gao et al. [35], and PPI and heartburn data are from Liang et al. [38]

**Table 2** Quality assessment of RCT studies using Cochrane Collaboration criteria

	Coron et al. (2008)	Corley et al. (2003)	Arts et al. (2012)	Aziz et al. (2010)
Randomization risk	Low	Low	Low	Low
Allocation concealment risk	Low	Low	Low	Low
Performance (blinding) risk	Unclear	Low	Low	Low
Detection bias risk	Unclear	Low	Low	Low
Attrition bias risk	Low	Low	Low	Low
Reporting bias risk	Low	Unclear	Unclear	Low

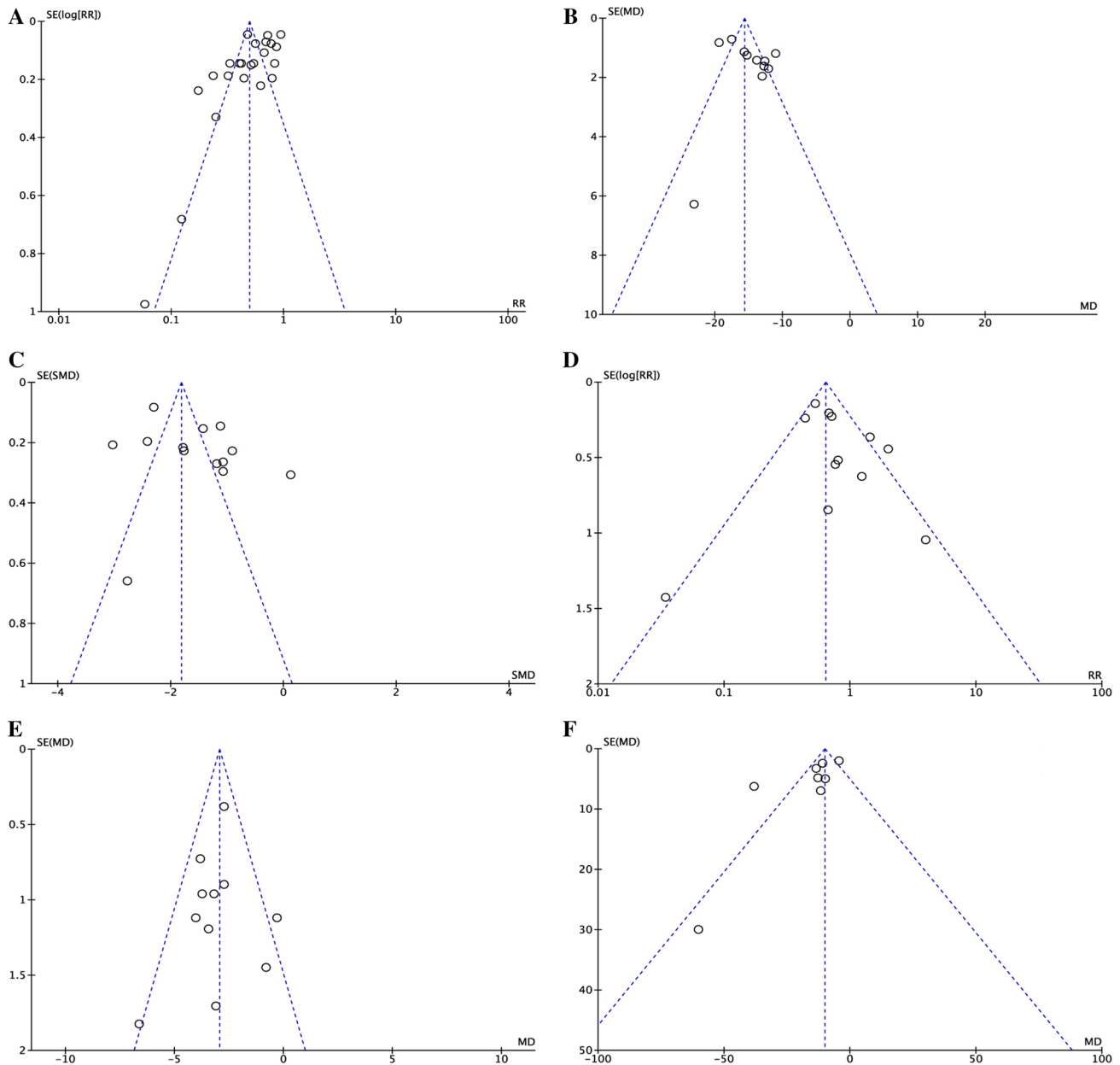
in HRQL score by a mean of  $-14.60$  [ $-16.48, -12.73$ ] (random effects model,  $P < 0.001$ ). When the HRQL treatment effects were analysed based on study design (RCTs and cohort studies), the pooled estimate of Stretta treatment

effects were similar between the RCT ( $-14.56$  [ $-16.63, -12.48$ ]) and cohort studies ( $-14.69$  [ $-16.90, -12.47$ ]), and the difference between these subgroups was not significant ( $P = 0.93$ ). The sham treatment also had a significant

**Table 3** Assessment of cohort study quality using modified Newcastle–Ottawa scale (Appendix 1)

Sources	Representativeness of exposed cohort	Selection of non-exposed cohort	Ascertainment of exposure	Incident disease	Comparability	Assessment of outcome	Length of follow-up	Adequacy of follow-up
Arts et al. [26]	A	A	A	A	A	C	A	A
Cipolletta et al. [29]	A	A	A	A	A	C	A	A
DiBaise et al. [31]	A	A	A	B	A	C	A	A
Dughera et al. <sup>a</sup> [32]	A	A	A	A	A	C	A	B
Dughera et al. <sup>a</sup> [33]	A	A	A	A	A	C	A	B
Dundon et al. [34]	A	A	A	A	A	C	A	B
Gao et al. [35]	A	A	A	C	A	C	A	A
Go et al. [36]	B	A	A	A	A	C	A	C
Higuchi et al. [23]	A	A	A	A	A	C	A	A
Houston et al. [37]	A	A	A	A	A	C	A	C
Liang et al. [38]	A	A	A	A	A	C	A	B
Liang et al. [21]	C	C	B	B	C	C	A	D
Liu et al. [39]	A	A	A	A	A	C	A	D
Lufi et al. [40]	B	A	A	A	A	C	A	C
Mansell [41]	A	A	A	B	A	C	A	D
Meier et al. [42]	A	A	A	B	A	C	A	C
Noar and Lofiti-Emran [43]	A	A	A	C	A	C	A	A
Noar et al. [44]	A	A	A	A	A	C	A	C
Reymunde and Santiago [45]	A	A	A	A	A	C	A	B
Richards et al. [46]	A	A	A	A	A	C	A	A
Tam et al. [47]	A	A	A	B	A	C	A	B
Torquati et al. [48]	A	A	A	A	A	C	A	C
Triadafilopoulos et al. [22]	A	A	A	B	A	C	A	C
Wolfsen and Richards [49]	B	B	B	B	B	C	A	A

<sup>a</sup>Studies with overlapping patient populations. Esophagitis and adverse events data are from Dughera et al. [32], and HRQL and heartburn data are from Dughera et al. [33]



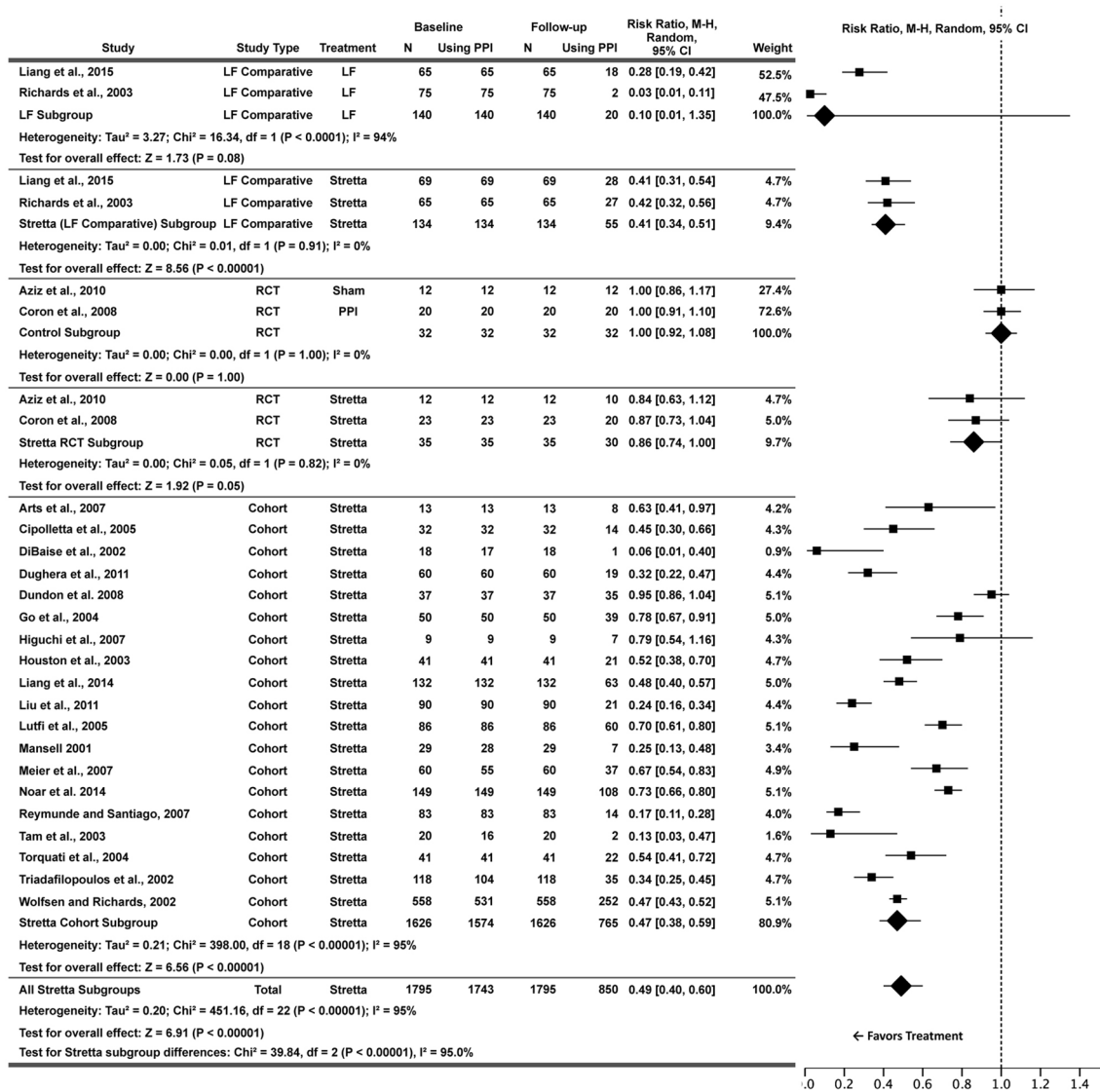
**Fig. 2** Funnel plots of the standard error of the treatment effect for the study outcomes. This series of six funnel plots was developed to assess publication bias for each of the six study outcomes within the meta-analyses. Each funnel plot was labelled **A–F** as follows: diagram **A** plots the SE of the log risk ratio versus the risk ratio (log scale) for PPI use, diagram **B** plots the SE of the mean difference versus the mean difference for HRQL score, diagram **C** plots the SE of the standardized mean difference versus the standardized mean differ-

ence for heartburn score, diagram **D** plots the SE of the log risk ratio versus the risk ratio (log scale) for erosive esophagitis frequency, diagram **E** plots the SE of the mean difference versus the mean difference for acid exposure time, and diagram **F** plots the SE of the mean difference versus the mean difference for LES pressure. *PPI* proton pump inhibitor, *HRQL* health-related quality of life, *LES* lower esophageal basal pressure, *SE* standard error, *RR* risk ratio

treatment effect ( $-4.95$  [ $-7.15$ ,  $-2.75$ ],  $P < 0.001$ ). However, the sham treatment effect was only one-third as strong as the Stretta treatment effect, and the Stretta treatment effect for the RCT subgroup was significantly larger than the sham effect ( $P < 0.001$  by  $\chi^2$  test, not shown in the forest plot).

Heterogeneity was not significant in the RCT subgroups, but was high in the Stretta cohort subgroup ( $I^2 = 85\%$ ,  $P < 0.001$ ). The pooled estimate of treatment effect for Stretta was not significantly different between the RCT and cohort subgroups ( $P = 0.93$ ). A meta-regression of treatment effect versus time was not significant ( $P = 0.51$ ).





**Fig. 3** Forest plots for change in reliance on PPI use by patients following Stretta or sham. Chart plots treatment effect in 20 pre/post-treatment cohort trials, 2 pre/post-treatment RCTs and 2 pre/post-treatment trials with LF comparison. The treatment effect is the risk ratio of reliance on PPI use (at any frequency) at baseline and reliance on PPI use at the longest follow-up period of each study (3–120 months). Summary statistics using the random effects model for

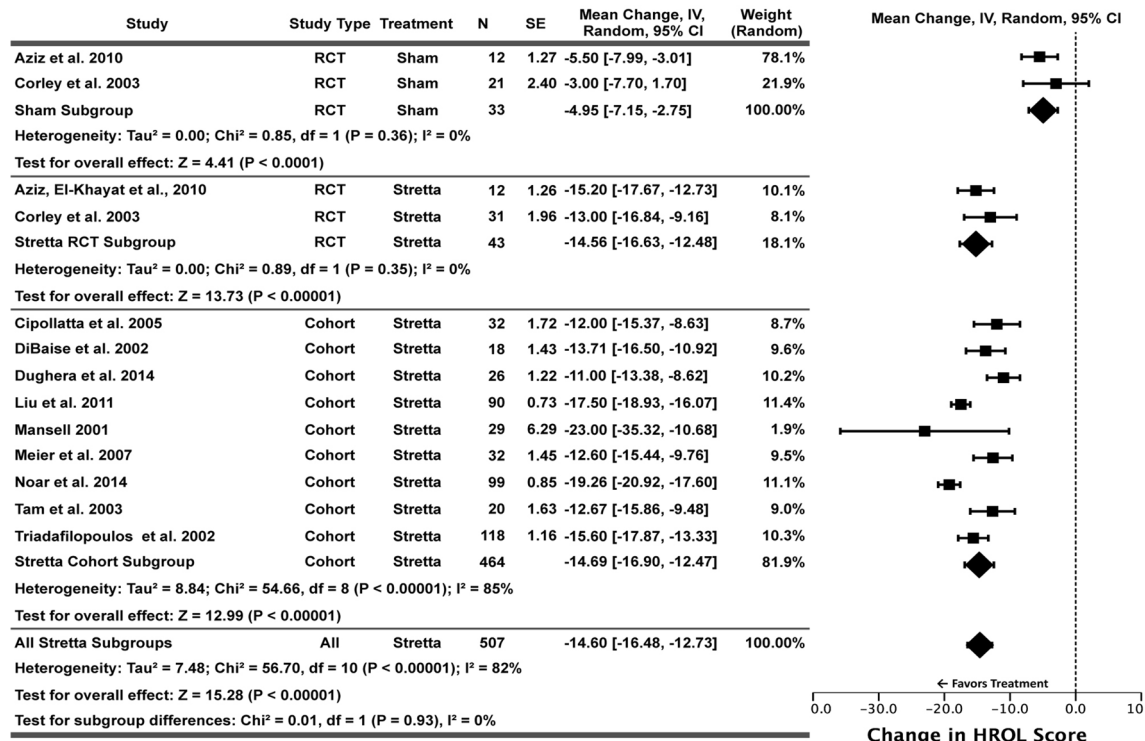
assigning weights to studies are presented for subgroups comprising LF treatment, sham treatment, Stretta treatment in LF comparative trials, Stretta treatment in RCTs and Stretta treatment in cohort trials. Weights are determined by Mantel–Haenszel method. Lower risk ratio represents fewer patients using PPI at follow-up and favors treatment. *N* number of patients, *SE* standard error, *CI* confidence interval, *LF* laparoscopic fundoplication

*Heartburn score*

Heartburn was reported in 12 studies with a total of 637 patients and, because different measuring scales were used (six point Likert, five point Likert or a product of severity and frequency), we used standardized variables to compare scores between studies as described in “Methods” section. We calculated and analysed Stretta and sham treatment effects, which are the standardized changes in heartburn score from baseline to follow-up. As depicted in Fig. 5, the 95% confidence intervals show

that only 1 of the 12 studies did not have an individually significant treatment effect for Stretta treatment. One study [21] included a LF comparison, which also showed a significant treatment effect.

In the RCTs, a statistically significant pooled estimate of treatment effect was not found for either the Stretta subgroup (−0.53 [−1.58, 0.52], *P* = 0.32) or the control subgroup (+0.17 [−1.27, 1.61], *P* = 0.82). However, when we pooled the Stretta arm of the RCT studies with the cohort subgroup, Stretta treatment reduced (thus improved) the heartburn standardized score significantly (*P* < 0.001,



**Fig. 4** Forest plots for change in self-reported health-related quality of life (HRQL) following Stretta or sham. Chart plots treatment effect in nine pre/post-treatment cohort trials and two pre/post-treatment RCTs. The treatment effect is the mean change from baseline at the longest follow-up period of each study (4–120 months). Summary statistics using the random effects model for assigning weights

$N=12$  studies) by  $-1.53$  [ $-1.97, -1.09$ ], which is statistically better than the control subgroup ( $P=0.01$ ).

Heterogeneity is highly significant ( $P<0.001$ ) in all Stretta subgroups, and there is also a significant difference in mean Stretta treatment effect between the RCT and cohort subgroups ( $P=0.04$ ). A meta-regression of standardized score versus time was not significant ( $P=0.90$ ).

#### Erosive esophagitis incidence

The frequency of erosive esophagitis was reported in 12 studies that performed upper endoscopy at baseline ( $N=500$ ) and follow-up ( $N=486$ ), and we stratified studies according to study design (RCTs and cohort studies) and treatment (Stretta and sham) on a per-protocol basis. Because studies either used the Los Angeles or the Savary–Miller classification, we pooled erosive esophagitis of any severity and calculated the treatment effect for erosive esophagitis as the risk ratio between baseline and follow-up frequencies. The risk ratio can be interpreted as the change in the fraction of patients with esophagitis (of any severity) between baseline and follow-up.

to studies are presented for subgroups comprising sham treatment, Stretta treatment in RCTs and Stretta treatment in cohort trials. Weights are determined by inverse variance. Negative change from baseline favors treatment.  $N$  number of patients,  $SE$  standard error,  $CI$  confidence interval

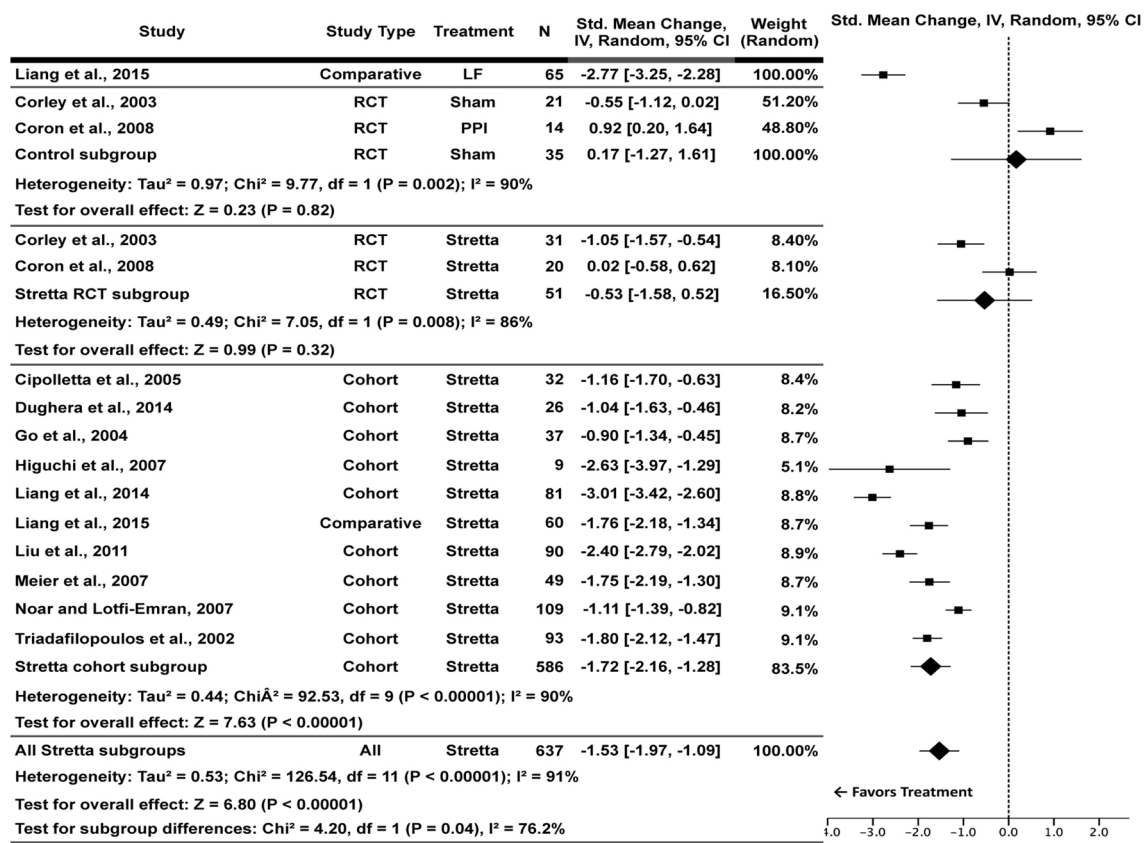
We found a substantial difference between the fixed effects and random effects models for erosive esophagitis. For the random effects model, Fig. 6A indicates that Stretta treatment marginally reduced the pooled estimate of frequency of erosive esophagitis at follow-up in all Stretta subgroups by a risk ratio of 0.76 [0.56, 1.04] ( $P=0.08$ ,  $N=12$ ). The pooled estimate of treatment effect was similar between Stretta RCT and cohort subgroups ( $P=0.37$ ,  $N=2$  subgroups).

However, when these data are analysed by the fixed effects model (Fig. 6B), the treatment effect for Stretta on erosive esophagitis is statistically significant ( $P<0.00001$ ).

Heterogeneity was not significant in the RCT subgroups, but significant heterogeneity was seen in the Stretta cohort subgroups ( $I^2=55%$ ,  $P=0.01$ ). A meta-regression of treatment effect versus time was not significant ( $P=0.31$ ).

#### Esophageal acid exposure

Eleven studies with 364 Stretta patients reported the percentage total time pH less than 4.0 (over 24-h period). For esophageal acid exposure, the treatment effect was the change in percentage of time of acid exposure



**Fig. 5** Forest plots for change in self-reported heartburn symptom score following Stretta, sham or LF treatment. Chart plots treatment effect in 11 pre/post-treatment cohort trials, 2 pre/post-treatment RCTs and 1 pre/post-treatment trial with LF comparison. The treatment effect is the standardized mean change from baseline at the longest follow-up period of each study (3–96 months). Summary statistics using the random effects model for assigning weights to stud-

ies are presented for subgroups comprising sham treatment, Stretta treatment in RCTs and Stretta treatment in cohort trials (pooled with the LF comparative trial). Weights are determined by inverse variance. Negative change from baseline favors treatment. *N* number of patients, *SE* standard error, *CI* confidence interval, *LF* laparoscopic fundoplication

between baseline and follow-up. As presented in Fig. 7, Stretta treatment reduced (thus improved) the pooled estimate of esophageal acid exposure by  $-3.01$  [ $-3.72$ ,  $-2.30$ ], which is highly significant ( $P < 0.001$ ,  $N = 11$ ).

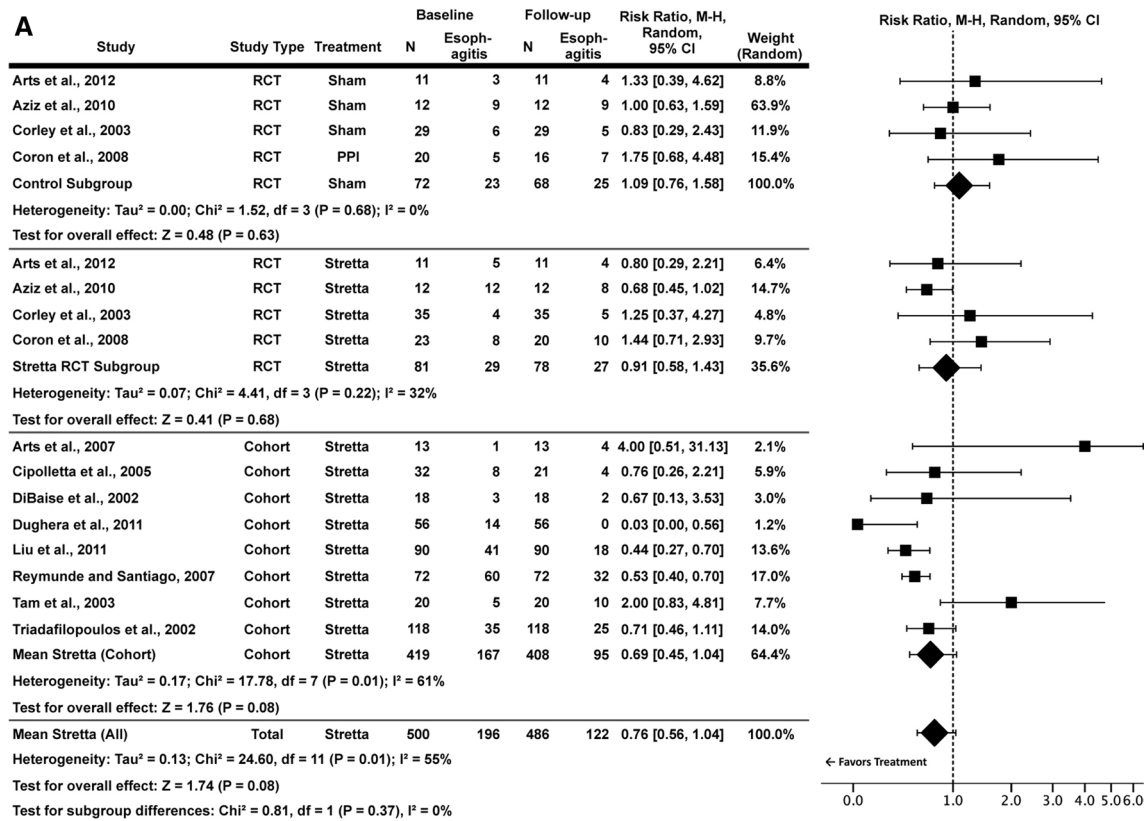
In the RCT subgroups, the pooled estimate of Stretta treatment effect was  $-1.45$  [ $-3.05$ ,  $0.15$ ] ( $P = 0.08$ ,  $N = 3$ ), while the sham treatment effect was  $-1.63$  [ $-2.88$ ,  $-0.37$ ] ( $P = 0.01$ ). These estimates are not significantly different ( $P = 0.86$ ). However, in the cohort subgroup, the treatment effect ( $-3.20$  [ $-3.74$ ,  $-2.66$ ]) was highly significant ( $P < 0.001$ ,  $N = 8$ ). Heterogeneity was not significant in any subgroup. However, there was a significant difference in Stretta treatment effect between RCT and cohort subgroups ( $P = 0.04$ ). A meta-regression of Stretta treatment effect versus time was not significant ( $P = 0.80$ ). In addition, the number of subjects with normalization of 24-h acid exposure at follow-up following Stretta treatment was reported in two RCTs and six cohort studies. These studies reported a total of 43

subjects with normalized acid exposure out of 144 subjects tested (30%).

Although not reported in a separate forest plot, the DeMeester score is an alternative way of expressing esophageal acid exposure that was reported in 8 cohort studies with 407 patients. As with acid exposure time, the pooled estimate of Stretta treatment effect on DeMeester score ( $-13.79$  [ $-20.01$ ,  $-7.58$ ], random effects model) was highly significant ( $P < 0.001$ ); however, considerable heterogeneity ( $I^2 = 77\%$ ) was also present.

#### Lower esophageal sphincter basal pressure

Six cohort and 3 RCT studies with a total of 269 patients reported LES basal pressure at baseline and follow-up. The analyses and forest plots for these studies appear in Fig. 8. Stretta treatment increased (thus improved) the pooled estimate of LES basal pressure (in mmHg) by  $+1.73$  [ $-0.29$ ,  $3.74$ ] ( $P = 0.09$ ,  $N = 9$  studies). In the RCT subgroups, the



**Fig. 6 A** Forest plots for change in frequency of erosive esophagitis following Stretta or sham treatment using random effects model. Chart plots treatment effect in eight pre/post-treatment cohort trials and four pre/post-treatment RCTs. The treatment effect is the risk ratio of frequency at baseline and at the longest follow-up period of each study (3–48 months). Summary statistics using the random effects model for assigning weights to studies are presented for subgroups comprising sham treatment, Stretta treatment in RCTs and Stretta treatment in cohort trials. Weights are determined by Mantel–Haenszel method. Lower risk ratio favors treatment. *N* number of patients, *SE* standard error, *CI* confidence interval. **B** Forest plots

for change in frequency of erosive esophagitis following Stretta or sham treatment using fixed effects model. Chart plots effect in eight pre/post-treatment cohort trials and four pre/post-treatment RCTs. The treatment effect is the risk ratio of frequency at baseline and at the longest follow-up period of each study (3–48 months). Summary statistics using the fixed effects model for assigning weights to studies are presented for subgroups comprising sham treatment, Stretta treatment in RCTs and Stretta treatment in cohort trials. Weights are determined by Mantel–Haenszel method. Lower risk ratio favors treatment. *N* number of patients, *SE* standard error, *CI* confidence interval

pooled estimate of treatment effect for Stretta was +3.00 [1.02, 4.98] ( $P=0.003$ ,  $N=3$  studies) while it was +2.80 [0.13, 5.47] ( $P=0.04$ ,  $N=3$  studies) for sham. Stretta is not significantly different than sham ( $P=0.09$ ).

There is significant heterogeneity in the sham and Stretta cohort studies ( $P=0.03$ ,  $N=3$  and  $P<0.001$ ,  $N=6$ , respectively). The pooled estimate of Stretta treatment effect is not different between RCT and cohort subgroups ( $P=0.24$ ).

The LES basal pressure change reported by Triadafilopoulos et al. [22] may be considered an outlier. If this study is removed from the analysis, the cohort treatment effect becomes +2.00 [0.21, 3.79], which is significant ( $P=0.03$ ). However, it is still not significantly different than sham ( $P=0.63$ ,  $N=5$ ).

A meta-regression of treatment effect versus time in months shows an increase in pressure of 0.35 mmHg/month ( $P=0.05$ ).

#### Adverse events

As presented in Table 4, 26 studies comprising 2468 Stretta procedures, 52 sham procedures and 195 LF procedures reported on AEs. The AE rate for the Stretta procedure was 0.93%, whereas it was 7.18% for the LF procedure. For Stretta, small erosions and mucosal lacerations was the most frequent AE at less than 1%, while for LF procedures, subcutaneous emphysema was the most frequent AE at approximately 3%.

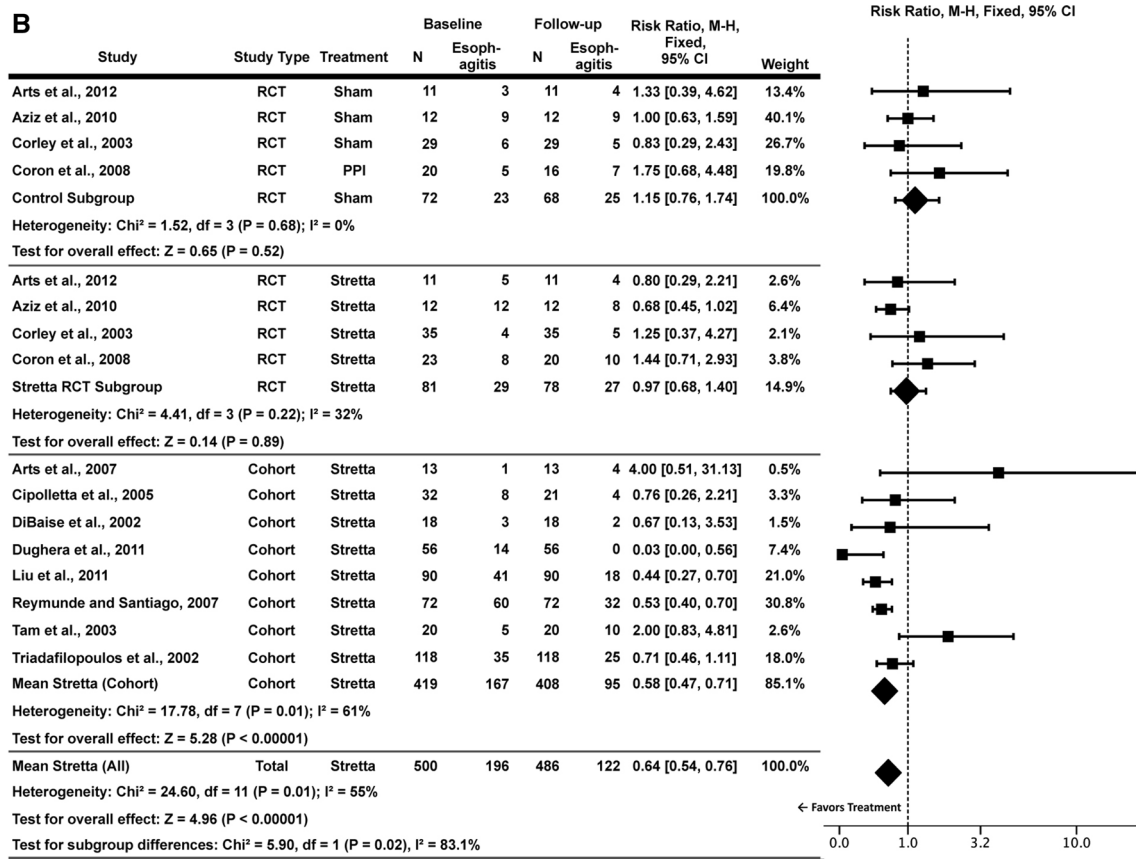


Fig. 6 (continued)

**Discussion**

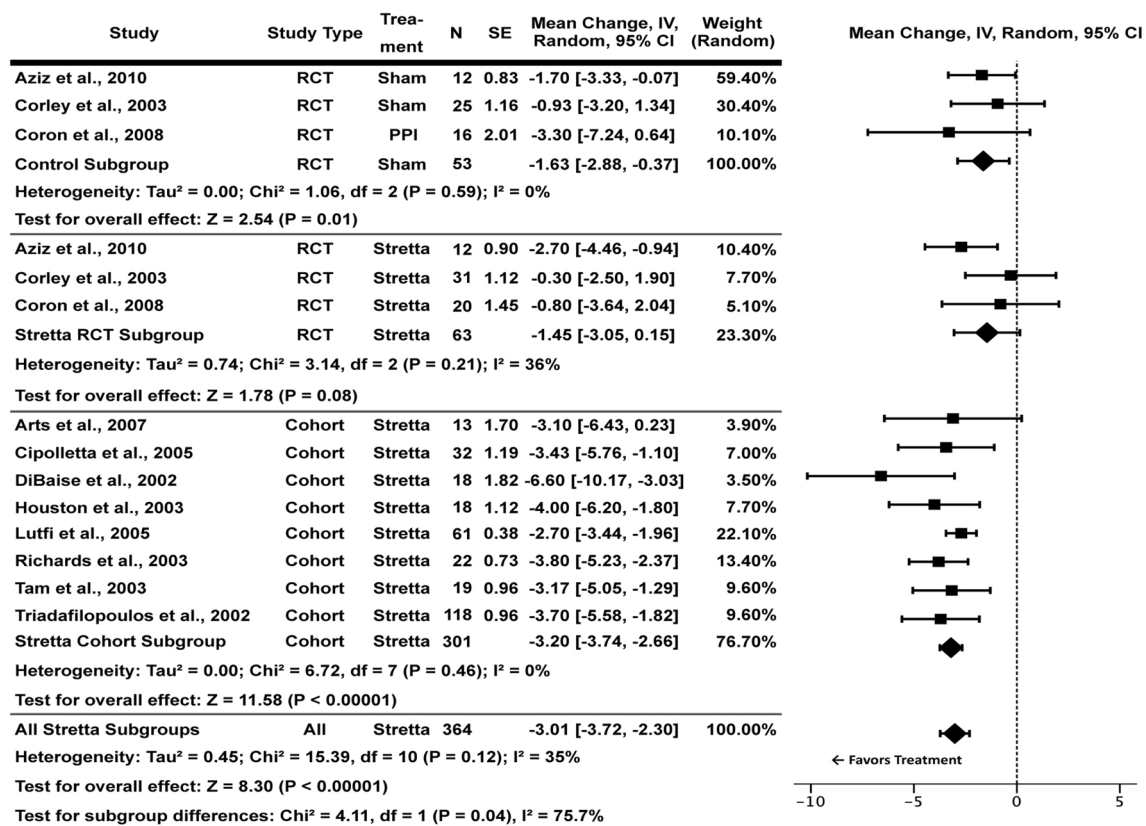
Our meta-analysis demonstrated that the Stretta procedure significantly improved HRQL, heartburn score and erosive esophagitis incidence. In addition, the technique significantly reduced the use of PPI’s and esophageal acid exposure but appears to have no significant effect on LES basal pressure. Overall, the safety profile of the Stretta technique was excellent with only approximately 1% AE rate.

The strength of our meta-analysis includes greater number of studies with a wide range of patient populations, health care systems, investigators, and a much higher total number of subjects. Therefore, the results are likely more representative of real-world effectiveness of Stretta than an analysis of only highly controlled RCTs. Since PPI use, HRQL and heartburn are self-reported, there is little risk of investigator bias due to lack of blinding.

The Stretta procedure is the endoscopic technique with the longest duration of clinical experience [23]. However, most of the studies available thus far are cohort trials with only four RCTs available. Whilst RCTs typically provide the highest level of evidence; one cannot ignore

the many prospective cohort studies that enrolled large number of patients, who underwent the Stretta procedure.

Lipka et al. has recently performed a systematic review and meta-analysis of the RF ablation endoscopic technique for the treatment of GERD [7]. The authors included only the 4 RCTs of the Stretta procedure with a total of 153 patients available for analysis. The study concluded that the overall quality of the evidence was very low and that the pooled results showed no difference between Stretta and sham or management with a PPI in patients with GERD for the outcomes of mean total time pH less than 4, LES basal pressure, ability to discontinue PPI and HRQL. However, for HRQL our study pooled two RCTs and nine cohort studies, revealing a statistical significant improvement post Stretta procedure. Similar results were achieved by Perry et al. ( $P=0.001$ ) [6]. Even when only the RCTs were pooled (limited to two studies which reported HRQL both at baseline and after the Stretta procedure), there was a statistically significant improvement in HRQL post Stretta procedure ( $P=0.002$ ). In many endoscopic trials of GERD patients, HRQL is a primary endpoint, assessing the effect of treatment on patients’ many aspects of daily function.



**Fig. 7** Forest plots for change in acid exposure time following Stretta or sham. Chart plots treatment effect in eight pre/post-treatment cohort trials and three pre/post-treatment RCTs. Acid exposure time is the percent of time over a 24-h period that the esophagitis is exposed to pH levels  $<4.0$ . The treatment effect is the mean change from baseline at the longest follow-up period of each study (6–12

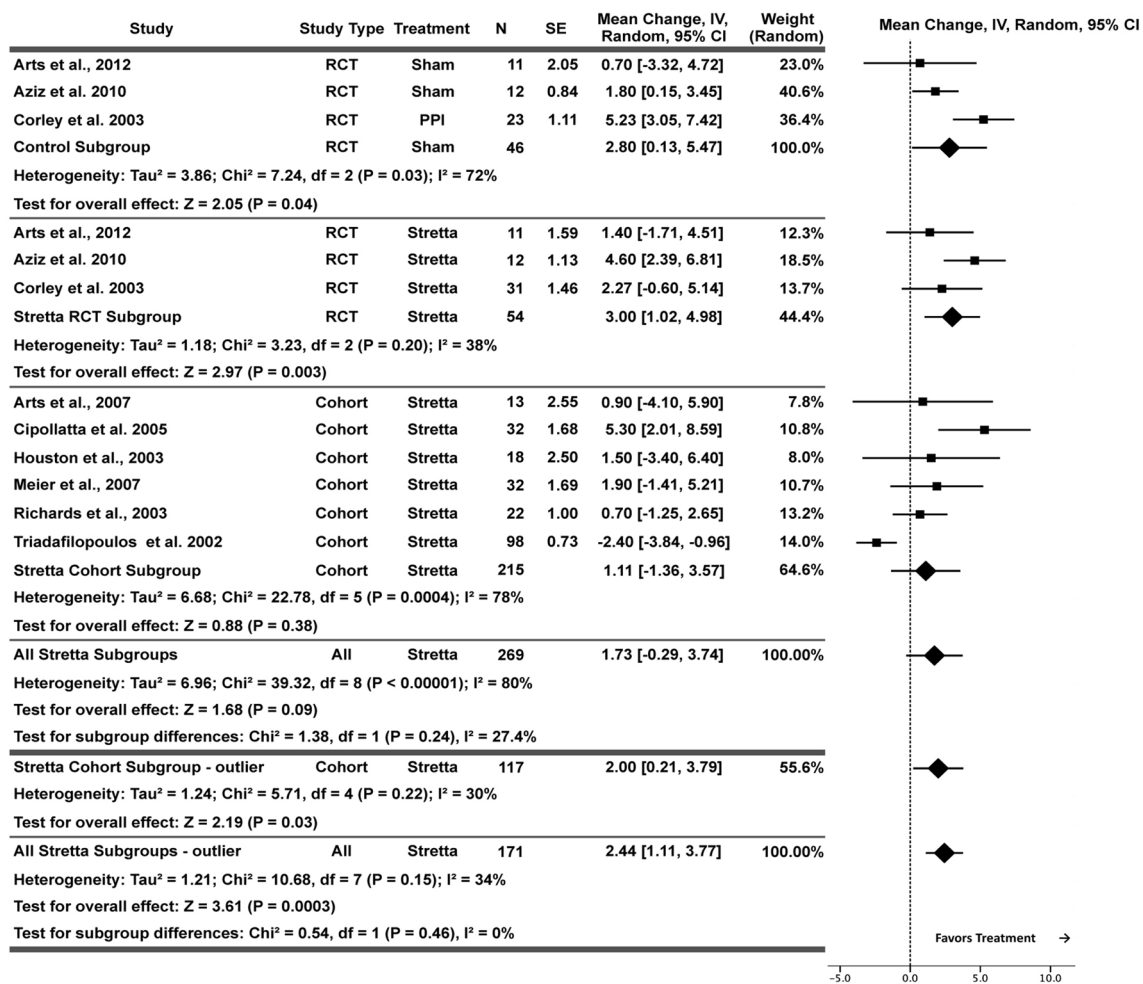
months). Summary statistics using the random effects model for assigning weights to studies are presented for subgroups comprising sham treatment, Stretta treatment in RCTs and Stretta treatment in cohort trials. Weights are determined by inverse variance. Negative change in acid exposure time favors treatment. *N* number of patients, *SE* standard error, *CI* confidence interval

Our study also assessed the use of PPIs, pooling only RCTs and revealing a small treatment effect that did not reach statistical significance. The other clinical endpoints, (except LES basal pressure) revealed clinical significance only when the RCTs were pooled together with the cohort studies.

The study by Lipka et al. argues for only the use of RCTs when performing a meta-analysis of an intervention. The main argument behind this approach is the assumption that RCTs have a valid study design for causal inference as compared with observational study design. However, a recent study critically examined the principal elements underlying this claim which include that randomization removes the chance of confounding and the double-blind process minimizes biases caused by the placebo effect [24]. The authors concluded that both RCTs and observational studies have strengths and weaknesses and including information from observational studies may improve the inference based on only RCTs. The authors also found that review of empirical studies suggests that meta-analysis based on observational trials

generally produces estimates of effect similar to those from meta-analyses based on RCTs. Importantly, the authors determined that the advantages of including both observational studies and randomized studies in a meta-analysis are likely to outweigh the disadvantage in many situations and that observational studies should not be excluded a priori [24]. Thus, we argue, similar to Perry et al., that in the meta-analysis of the Stretta procedure both data from RCTs and cohort studies should be pooled together. If this is done, as has been shown in our study, all clinical endpoints, except LES baseline resting pressure reached statistical significance.

Limitations of our study include the lack of contemporaneous control groups in most of the studies. Researchers might expect more heterogeneity of treatment effect in this combination of RCTs and cohort studies compared to a meta-analysis solely of RCTs studies because of broadening of eligibility criteria and inclusion of more practitioners who have differing levels of expertise; however, several of the measures in our study exhibited less heterogeneity than that of the RCTs.



**Fig. 8** Forest plots for change in LES pressure following Stretta or sham. Chart plots treatment effect in six pre/post-treatment cohort trials and three pre/post-treatment RCTs. The treatment effect is the mean change from baseline at the longest follow-up period of each study (6–12 months). Summary statistics using the random effects

The four RCTs considered alone have limitations: they enrolled a total of 92 Stretta-treated patients, whereas the cohort trials and registry enrolled 2376 Stretta-treated patients. Only one of the outcome measures (erosive esophagitis) was measured in all four RCTs. Three of the outcome measures (HRQL, heartburn, PPI use) were measured in only two RCTs. Furthermore, the longest follow-up time in the RCTs was 12 months, whereas cohort studies included data up to 120 months (average 23 months). Thus, the larger sample sizes and longer follow-up may balance the theoretical advantages of the RCT design. Further, while placebo effect does not bias magnitude of the treatment effect expected in clinical practice, a significant placebo effect could demonstrate that the actual mechanism of action of a treatment is not understood, and/or that there could be less complex and invasive, or less costly alternatives that achieve similar effectiveness. In this

model for assigning weights to studies are presented for subgroups comprising sham treatment, Stretta treatment in RCTs and Stretta treatment in cohort trials. Weights are determined by inverse variance. Positive change in LES pressure favors treatment.  $N$  number of patients,  $SE$  standard error,  $CI$  confidence interval

meta-analysis, sham treatment data from the four RCTs were available to estimate placebo effects. For example, we observed no treatment effect for the sham procedure relative to PPI use. Furthermore, the treatment effects we observed are very large (e.g.  $P < 0.001$ ), making it unlikely that they were due to statistical biases. Therefore, while there is less historical precedent than for meta-analyses solely of RCTs, we believe our methodology is well justified.

In conclusion, our meta-analysis demonstrated that the Stretta procedure significantly reduced the use of PPIs while improving esophageal acid exposure time, heartburn symptoms, and HRQL. The observed 24% reduction in erosive esophagitis incidence approached, but did not reach, statistical significance under the random effects ( $P = 0.08$ ), but did reach statistical significance under fixed effects ( $P < 0.001$ ). There was no significant effect on LES basal pressure. Overall, it appears that the Stretta procedure is

**Table 4** Comparison of adverse events among patients who underwent the Stretta technique, sham procedure or laparoscopic fundoplication

	Stretta N=2468	Sham N=52	LF N=195
Small erosions	9 (0.36)	0 (0.00)	0 (0.00)
Mucosal lacerations	7 (0.28)	0 (0.00)	0 (0.00)
Gastroparesis (prolonged)	3 (0.12)	0 (0.00)	0 (0.00)
Bleeding esophageal ulcer	1 (0.04)	0 (0.00)	0 (0.00)
Mediastinal inflammation	1 (0.04)	0 (0.00)	0 (0.00)
Pleural effusion	1 (0.04)	0 (0.00)	0 (0.00)
Pneumonia	1 (0.04)	1 (1.45)	0 (0.00)
Subcutaneous emphysema	0 (0.00)	0 (0.00)	6 (3.08)
Enterotomy	0 (0.00)	0 (0.00)	2 (1.03)
Incisional hernia	0 (0.00)	0 (0.00)	2 (1.03)
Dysphagia (prolonged)	0 (0.00)	0 (0.00)	1 (0.51)
Paraesophageal hernia	0 (0.00)	0 (0.00)	1 (0.51)
Pneumothorax	0 (0.00)	0 (0.00)	1 (0.51)
Slipped-Nissen	0 (0.00)	0 (0.00)	1 (0.51)
Death	0 (0.00)	0 (0.00)	0 (0.00)
Perforation	0 (0.00)	0 (0.00)	0 (0.00)
Totals	23 (0.93)	1 (1.92)	14 (7.18)

Results are expressed as number of patients (percentage of events)

LF laparoscopic fundoplication

efficacious in improving both objective and subjective clinical endpoints.

Our current meta-analysis, combined with recently published data [25] demonstrating that the Stretta procedure can result in cost savings, ranging from 7.3 to 50.5% in the 12-month time period following the index procedure, provides important evidence to support the utilization of the Stretta procedure in clinical practice as an alternative therapeutic modality for GERD patients seeking non-surgical options.

#### Compliance with ethical standards

**Disclosures** Dennis J. Scotti is a Part-time Consultant with Baker Tilly Virchow Krause, LLC. Baker Tilly Virchow Krause, LLC is a Business Advisor to Mederi Therapeutics, Inc. David A. Gregory is a principal with Baker Tilly, a Business Advisor to Mederi Therapeutics. Frederick Cahn is a principal with BioMedical Strategies, a Business Advisor to Baker Tilly. Ronnie Fass is an Advisor to Ironwood and Mederi Therapeutics, Speaker for AstraZeneca, Dr. Reddy, Mederi Therapeutics and Takeda and receives Research Grant from Ironwood.

#### Appendix 1: Modified Newcastle–Ottawa scoring system

We used the modified Newcastle–Ottawa scale to evaluate the following study characteristics, with the definition of

exposure being Stretta treatment, and the measures of interest are the self-reported outcomes:

- *Representativeness of exposed cohort* A=truly representative of the average patient with GERD, B=some-what representative of the average patient with GERD, C=selected group and D=no description of the derivation of the cohort.
- *Selection of non-exposed cohort* A=drawn from the same community as the exposed cohort, B=drawn from a different source and C=no description of the derivation of the non-exposed cohort.
- *Ascertainment of exposure* A=secure record (e.g. surgical records), B=structured interview, C=written self-report and D=no description.
- *Demonstration that outcome of interest was not present at start of study (with respect to PPI use)* A=yes, B=no, C=not applicable because outcome was not assessed at follow-up.
- *Comparability of cohorts on the basis of the design or analysis* A=study controls for comorbidities, B=study controls for additional risk factors (such as age and severity of illness) and C=not done.
- *Assessment of outcome* A=independent blind assessment, B=record linkage, C=self-report and D=no description.
- Was follow-up long enough for outcomes to occur A=yes, and B=no.
- *Adequacy of follow-up of cohorts* A=complete follow-up—all subjects accounted for, B=subjects lost to follow-up unlikely to introduce bias (small number lost), follow-up rate higher than 90% or description provided of those lost, C=follow-up rate 90% or lower and no description of those lost and D=no statement.

#### Appendix 2: Assessment of studies using the PICOTS framework

1. The patient population (“P”) in all trials were GERD patients. GERD is a common chronic and stable condition with a clinical context and natural history that is consistent with the criteria of Glasziou et al. [50] for the validity of non-randomized studies.
2. The intervention (“I”) is in all included studies a single physician-administered outpatient procedure using one specific device. Patient compliance was not a factor affecting treatment effectiveness.
3. In both RCT and cohort studies, the comparison (“C”) is a pre–post-treatment difference that directly measures a treatment effect. This design also reduces the risk of bias from confounding variables. Also, because of the stable natural history of GERD there is a low



risk of bias due to regression-to-the-mean or spontaneous improvement. None of the studies use historical controls, which is known to have substantial risk of bias [24].

4. Both RCT and cohort studies report the same patient-reported effectiveness outcomes (“O”) (HRQL, heartburn score, PPI use) and in the case of erosive esophagitis an objective physician-measured effectiveness outcome.
5. The timeframe (“T”) of the treatment is short and the timeframe of follow-up is significant in terms of natural history of GERD. The follow-up times for the cohort studies overlap those of the RCT studies.
6. The setting (“S”) for the treatment in all studies is the outpatient clinic and the evaluation of HRQL, heartburn and PPI use is by the patient in the course of their normal activities. This setting matches that of normal clinical practice.

## References

1. Maradey-Romero C, Fass R (2014) New and future drug development for gastroesophageal reflux disease. *J Neurogastroenterol Motil* 20(1):6–16
2. Locke GR III, Talley NJ, Fett SL, Zinsmeister AR, Melton LJ III (1997) Prevalence and clinical spectrum of gastroesophageal reflux: a population-based study in Olmsted County, Minnesota. *Gastroenterology* 112(5):1448–1456
3. El-Serag HB, Sweet S, Winchester CC, Dent J (2014) Update on the epidemiology of gastro-oesophageal reflux disease: a systematic review. *Gut* 63(6):871–880
4. Franciosa M, Triadafilopoulos G, Mashimo H (2013) Stretta radiofrequency treatment for GERD: a safe and effective modality. *Gastroenterol Res Pract* 2013:783–815
5. Triadafilopoulos G (2014) Stretta: a valuable endoscopic treatment modality for gastroesophageal reflux disease. *World J Gastroenterol* 20(24):7730–7738
6. Perry KA, Banerjee A, Melvin WS (2012) Radiofrequency energy delivery to the lower esophageal sphincter reduces esophageal acid exposure and improves GERD symptoms: a systematic review and meta-analysis. *Surg Laparosc Endosc Percutan Tech* 22(4):283–288
7. Lipka S, Kumar A, Richter JE (2015) No evidence for efficacy of radiofrequency ablation for treatment of gastroesophageal reflux disease: a systematic review and meta-analysis. *Clin Gastroenterol Hepatol* 13(6):1058.e1–1067.e1
8. Higgins JP, Green S (2014) Cochrane handbook for systematic reviews of interventions. *Online Kensington* 35(3):154–155
9. Moher D, Liberati A, Tetzlaff J, Altman DG (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 151(4):264–269
10. Velanovich V (2007) The development of the GERD-HRQL symptom severity instrument. *Dis Esophagus* 20(2):130–134
11. Velanovich V (1998) Comparison of generic (SF-36) vs. disease-specific (GERD-HRQL) quality-of-life scales for gastroesophageal reflux disease. *J Gastrointest Surg* 2(2):141–145
12. Hedges LV (1982) Statistical methodology in meta-analysis: ERIC Clearinghouse on Tests, Measurement and Evaluation. Report No.: ERIC/TM Report 83, December 1982. Princeton
13. Higgins JP, Green S (2008) Cochrane handbook for systematic reviews of interventions. Wiley Online Library
14. Wells G, Shea B, O’Connell D et al (2011) The Newcastle–Ottawa scale (NOS) for assessing the quality of nonrandomized studies in meta-analysis. The Ottawa Health Research Institute, Ottawa
15. Abdel-Latif A, Bolli R, Tleyjeh IM et al (2007) Adult bone marrow-derived cells for cardiac repair: a systematic review and meta-analysis. *Arch Intern Med* 167(10):989–997
16. Fu R, Vandermeer BW, Shamliyan TA et al (2014) Chapter 14: handling continuous outcomes in quantitative synthesis. In: *Methods guide for effectiveness and comparative effectiveness reviews*. Agency for Healthcare Research and Quality, Rockville, p 285–311
17. Hozo SP, Djulbegovic B, Hozo I (2005) Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol* 5(1):13
18. Balk EM, Earley A, Patel K, Trikalinos TA, Dahabreh IJ (2012) In: Agency for Healthcare Research and Quality (ed) *Empirical assessment of within-arm correlation imputation in trials of continuous outcomes*. Agency for Healthcare Research and Quality, Rockville
19. Coron E, Sebillé V, Cadiot G et al (2008) Clinical trial: radiofrequency energy delivery in proton pump inhibitor-dependent gastro-oesophageal reflux disease patients. *Aliment Pharmacol Ther* 28(9):1147–1158
20. The Agency for Healthcare Research and Quality (2014) In: The Agency for Healthcare Research and Quality (ed) *Methods guide for effectiveness and comparative effectiveness reviews*. The Agency for Healthcare Research and Quality, Rockville, p 384
21. Liang WT, Yan C, Wang ZG et al (2015) Early and midterm outcome after laparoscopic fundoplication and a minimally invasive endoscopic procedure in patients with gastroesophageal reflux disease: a prospective observational study. *J Laparoendosc Adv Surg Tech A* 25(8):657–661
22. Triadafilopoulos G, DiBaise JK, Nostrant TT et al (2002) The Stretta procedure for the treatment of GERD: 6 and 12 month follow-up of the U.S. open label trial. *Gastrointest Endosc* 55(2):149–156
23. Higuchi K, Fujiwara Y, Okazaki H et al (2007) Feasibility, safety, and efficacy of the Stretta procedure in Japanese patients with gastroesophageal reflux disease: first report from Asia. *J Gastroenterol* 42(3):205–210
24. Shrier I, Boivin J-F, Steele RJ et al (2007) Should meta-analyses of interventions include observational studies in addition to randomized controlled trials? A critical examination of underlying principles. *Am J Epidemiol* 166(10):1203–1209
25. Gregory D, Scotti D, Buck D, Triadafilopoulos G (2016) Budget impact analysis to estimate the cost dynamics of treating refractory gastroesophageal reflux disease with radiofrequency energy: a payer perspective. *Manag Care (Langhorne Pa)* 25(5):42
26. Arts J, Sifrim D, Rutgeerts P, Lerut A, Janssens J, Tack J (2007) Influence of radiofrequency energy delivery at the gastroesophageal junction (the Stretta procedure) on symptoms, acid exposure, and esophageal sensitivity to acid perfusion in gastroesophageal reflux disease. *Dig Dis Sci* 52(9):2170–2177
27. Arts J, Bisschops R, Blondeau K et al (2012) A double-blind sham-controlled study of the effect of radiofrequency energy on symptoms and distensibility of the gastro-esophageal junction in GERD. *Am J Gastroenterol* 107(2):222–230
28. Aziz AM, El-Khayat HR, Sadek A et al (2010) A prospective randomized trial of sham, single-dose Stretta, and double-dose Stretta for the treatment of gastroesophageal reflux disease. *Surg Endosc* 24(4):818–825
29. Cipolletta L, Rotondano G, Dughera L et al (2005) Delivery of radiofrequency energy to the gastroesophageal junction (Stretta

- procedure) for the treatment of gastroesophageal reflux disease. *Surg Endosc* 19(6):849–853
30. Corley DA, Katz P, Wo JM et al (2003) Improvement of gastroesophageal reflux symptoms after radiofrequency energy: a randomized, sham-controlled trial. *Gastroenterology* 125(3):668–676
  31. DiBaise JK, Brand RE, Quigley EM (2002) Endoluminal delivery of radiofrequency energy to the gastroesophageal junction in uncomplicated GERD: efficacy and potential mechanism of action. *Am J Gastroenterol* 97(4):833–842
  32. Dughera L, Navino M, Cassolino P et al (2011) Long-term results of radiofrequency energy delivery for the treatment of GERD: results of a prospective 48-month study. *Diagn Ther Endosc* 2011:507157
  33. Dughera L, Rotondano G, De Cento M, Cassolino P, Cisarò F (2014) Durability of Stretta radiofrequency treatment for GERD: results of an 8-year follow-up. *Gastroenterol Res Pract* 2014:531907
  34. Dundon JM, Davis SS, Hazey JW, Narula V, Muscarella P, Melvin WS (2008) Radiofrequency energy delivery to the lower esophageal sphincter (Stretta procedure) does not provide long-term symptom control. *Surg Innov* 15(4):297–301
  35. Gao X, Wang Z-g, Wu J-m et al (2011) Radiofrequency treatment on respiratory symptoms due to gastroesophageal reflux disease. *Chin Med J (Engl)* 124(7):1006–1009
  36. Go MR, Dundon JM, Karłowicz DJ, Domingo CB, Muscarella P, Melvin WS (2004) Delivery of radiofrequency energy to the lower esophageal sphincter improves symptoms of gastroesophageal reflux. *Surgery* 136(4):786–794
  37. Houston H, Khaitan L, Holzman M, Richards WO (2003) First year experience of patients undergoing the Stretta procedure. *Surg Endosc* 17(3):401–404
  38. Liang WT, Wu JM, Wang F, Hu ZW, Wang ZG (2014) Stretta radiofrequency for gastroesophageal reflux disease-related respiratory symptoms: a prospective 5-year study. *Minerva Chir* 69(5):293–299
  39. Liu HF, Zhang JG, Li J, Chen XG, Wang WA (2011) Improvement of clinical parameters in patients with gastroesophageal reflux disease after radiofrequency energy delivery. *World J Gastroenterol* 17(39):4429–4433
  40. Lutfi R, Torquati A, Kaiser J, Holzman M, Richards W (2005) Three year's experience with the Stretta procedure: Did it really make a difference? *Surg Endosc Other Interv Tech* 19(2):289–295
  41. Mansell DE (2001) Community practice evaluation of the effectiveness on the Stretta procedure for the treatment of GERD. *Am J Gastroenterol* 96(9):S21
  42. Meier PN, Nietzschmann T, Akin I, Klose S, Manns MP (2007) Improvement of objective GERD parameters after radiofrequency energy delivery: a European study. *Scand J Gastroenterol* 42(8):911–916
  43. Noar MD, Lotfi-Emran S (2007) Sustained improvement in symptoms of GERD and antisecretory drug use: 4-year follow-up of the Stretta procedure. *Gastrointest Endosc* 65(3):367–372
  44. Noar M, Squires P, Noar E, Lee M (2014) Long-term maintenance effect of radiofrequency energy delivery for refractory GERD: a decade later. *Surg Endosc* 28(8):2323–2333
  45. Reymunde A, Santiago N (2007) Long-term results of radiofrequency energy delivery for the treatment of GERD: sustained improvements in symptoms, quality of life, and drug use at 4-year follow-up. *Gastrointest Endosc* 65(3):361–366
  46. Richards WO, Houston HL, Torquati A, Khaitan L, Holzman MD, Sharp KW (2003) Paradigm shift in the management of gastroesophageal reflux disease. *Ann Surg* 237(5):638–647 (discussion 648–649)
  47. Tam WC, Schoeman MN, Zhang Q et al (2003) Delivery of radiofrequency energy to the lower oesophageal sphincter and gastric cardia inhibits transient lower oesophageal sphincter relaxations and gastro-oesophageal reflux in patients with reflux disease. *Gut* 52(4):479–485
  48. Torquati A, Houston HL, Kaiser J, Holzman MD, Richards WO (2004) Long-term follow-up study of the Stretta procedure for the treatment of gastroesophageal reflux disease. *Surg Endosc* 18(10):1475–1479
  49. Wolfsen HC, Richards WO (2002) The Stretta procedure for the treatment of GERD: a registry of 558 patients. *J Laparoendosc Adv Surg Tech A* 12(6):395–402
  50. Glasziou P, Chalmers I, Rawlins M, McCulloch P (2007) When are randomised trials unnecessary? Picking signal from noise. *BMJ* 334(7589):349–351