Online Supplement

A Design-Based Stereologic Method To Quantify the Tissue Changes Associated with a Novel Drug-Eluting Tracheobronchial Stent

Labib Debiane, MD[1]; Ruth Reitzel, PhD[2]; Joel Rosenblatt, PhD[2]; Mihai Gagea, DVSc[3]; Miguel A. Chavez, MD[1, 4]; Roberto Adachi, MD[1]; Horiana Grosu, MD[1]; Ajay Sheshadri, MD[1]; Lori R. Hill DVM[3]; Issam Raad, MD[2]; and David E. Ost, MD[1]

Departments of [1]Pulmonary Medicine, [2]Infectious Diseases, Infection Control and Employee Health, and [3]Veterinary Medicine and Surgery, The University of Texas MD Anderson Cancer Center. [4] Tecnologico de Monterrey, Escuela de Medicina y Ciencias de la Salud

Corresponding Author: David Ost, MD, MPH

Email: dost@mdanderson.org

Address: Department of Pulmonary Medicine, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd, Unit 1462, Houston TX 77030

Conflict of Interest: The authors have no conflicts of interest to declare.

Funding Information: This study was funded in part by an Institutional Research Grant from The University of Texas MD Anderson Cancer Center.

Methods

Rigid bronchoscopy and anesthesia

Swine were placed under general anesthesia, intubated, and mechanically ventilated. Isoflurane 2% was used until it was time to insert the rigid bronchoscope, at which point a 3-mL propofol intravenous bolus was given, followed by additional 2-mL boluses as needed up to a maximum of 6 mg/kg during the entire procedure. Using a Dumon rigid bronchoscope (Bryan Corporation, Woburn, MA), a stent was placed in the trachea 1 to 2 cm proximal to the cranial lobe (tracheal) bronchus. Postoperatively, the animals were monitored, and respiratory infection was considered if they demonstrated 2 of the following 3 criteria: 1) fever, 2) decreased activity/feeding, 3) cough or labored breathing. A computed tomography scan of the chest was ordered in animals meeting 2 criteria. If an infiltrate was present, pneumonia was considered, and the animal was euthanized. If the images showed no infiltrate and the swine were comfortable, they were followed clinically without treatment and only euthanized if they developed clear evidence of respiratory distress, pneumonia, or sepsis (e.g., hypotension, respiratory distress). Surviving swine were euthanized 1 month after stent placement. All euthanized animals underwent necropsy, at which time a definitive diagnosis of pneumonia was made.

Microbiology

The antimicrobial activity of the drug-eluting stent (DES) on nearby tissues was assessed by culturing 1-cm portions from the proximal, middle, and distal portions of the

trachea covered by the stent. Two control samples were obtained from each pig: one from the region cephalad to the proximal end of the stent and another from the regional caudal to the distal end of the stent. The weight of each sample was measured, and each sample was then homogenized in 5 mL of 0.9% sterile saline and cultured for bacteria on trypticase soy agar + 5% sheep blood (nonselective), MacConkey agar (selective for Gram-negative bacteria), and Columbia CNA agar (selective for Gram-positive bacteria). Bacterial growth was expressed as colony-forming units per gram of tissue.

Residual antimicrobial activity of the gendine-filled stents was assessed by culturing swabs of the areas around the eluting holes. Cultures of swabs from areas located away from the elution holes both proximally and distally were used as controls.

Additionally, the inner lining of the stent was removed and any residual gel was assessed for activity via zone-of-inhibition testing.[1] Briefly, a fresh culture of methicillin-resistant *Staphylococcus aureus* (MDA 120) was diluted to 0.5 McFarland and streaked 3× on a Muller-Hinton agar plate. Residual gel was placed in the center of the inoculated plate. After incubation at 37°C for 24 hours, the diameters of the zones-of-inhibition were measured.

Histology

Tracheal sections were fixed in 10% neutral buffered formalin for a minimum of 72 hours at room temperature for the fixation of tissues. Tissue sections arising from the same region were mounted on the same glass slide. The five regions were cephalad to the proximal end of the stent, the proximal stented region, the middle stented region, the

distal stented region, and caudal to the distal end of the stent. The assessor of histology (MG) was blinded to the type of stent used. Each of the 5 regions of trachea was sectioned longitudinally into 2- to 3-mm-thick slices, yielding 4 to 6 slices from each region. These tissue sections were processed and embedded into paraffin blocks, from which 4-µm-thick sections were cut and mounted on glass slides. The sections were stained with hematoxylin and eosin and scanned with an Aperio AT2 digital pathology slide scanner (Leica Biosystems, Buffalo Grove, IL) into digital images. The images were uploaded into the Aperio eSlide Manager database and assessed microscopically with Aperio ImageScope digital image analysis software.

Qualitative Histology

Previous studies of the effects of airway stenting have mainly used qualitative descriptions of histologic findings.[2-7] Only 1 prior study of airway stents used a systematic scoring system, shown in e-Table 3.[2] Our modified scoring system is shown in Table 1 of the main text. Note that while the "score" is indicated using numbers, it is in fact a categorical variable; the numbers do not indicate a scale or ratio and cannot be combined into a meaningful aggregate score. Within a given category, for instance epithelial erosions, different stents can be compared. However, this system does not provide a meaningful overall single global score of the severity of histopathologic changes.

Scoring systems still have merit in that they provide a framework for describing changes. However, they are not suitable for many types of comparisons and are unlikely to detect meaningful differences. Moreover, these systems do not include a

measurement of tissue volume or the relative volumes of different tissue types and are therefore insufficient for comparative clinical studies. Finally, these types of scoring systems cannot be used to weight changes in terms of their relative importance.

We modified the way the original histologic scoring system accounted for cellular infiltrates. The original histology scoring system[2] (e-Table 3) graded cellular infiltrates in an ordinal manner, and within that category it gave precedence to polymorphonuclear (PMN) infiltrates over lymphocytic infiltrates, such that a finding of 50 or fewer PMN infiltrates per high-power field (HPF) would be considered to indicate more inflammation than a finding of more than 50 lymphocytes per HPF. However, no prior studies of tracheal stenting have compared the role and significance of lymphocytic inflammation with that of PMN inflammation, so there is no *a priori* reason to consider lymphocytic infiltrates as less significant than PMN infiltrates. We therefore used a modified scoring system (Table 1 in the main text) that evaluated lymphocytes and PMN infiltrates separately.

Orthogonal Intercepts Method to Determine Tissue Thickness

Tracheal soft tissue thickness was measured from the inner luminal surface of the epithelium to the inner aspect of the airway cartilage using the method of orthogonal intercepts.[8] In Aperio ImageScope (v12.3.0.5056), 4 to 5 random equidistant horizontal guide lines were drawn over each tissue section (e-Figure 1). Next, a line was drawn from the point of intersection between the guide lines and the innermost surface of the tracheal cartilage towards and perpendicular to the basement membrane, and its length was recorded. When more than 1 line perpendicular to the basement membrane could

be drawn, the shortest line was selected. The number of actual measurements for each region was determined by the length of the tissue and the random placement of the guide lines. The estimated thickness was calculated as: (mean length of perpendicular lines) × π/4; this factor was introduced to correct for the overestimation secondary to the imperfect orthogonality of the tissue sectioning to the basement membrane (http://www.stereology.info/orthogonal-intercepts-thickness/).[8]

Stereology for Determination of Volume Density and Relative Volumes

Stereology depends on random unbiased sampling. Each slide contained all the sections from a given region of a single pig's trachea. To facilitate random sampling, the sections on a given slide were systematically labeled following a topographic rule, going from bottom to top and left to right. Digital images of each sample were obtained using ImageScope software. Before image capture, ImageScope software's tools were used to rotate the slides to orient them so that the longitudinal axis of each section was at 20° to 45° from the vertical, with the epithelium facing the bottom-left corner. If this was not possible using ImageScope, images were captured and manually rotated using Photoshop software (Adobe Systems, San Jose, CA), abiding by the same orientation rules.

Each image was labeled with a PigID<region number>_<image number>. Images were imported into the Image Path program. We used a line-pairs grid. Each image had 2 counting areas, each counting area had 100 tiles, and each tile had 2 probes, for a total of 400 probes per image. The tile number represented the number of tiles that appeared on the screen during point counting. The tile number setting had to

be large enough to ensure that the least common structure of interest (i.e., the structure with the smallest volume) received enough counts. In this study, epithelium was the smallest-volume tissue component. Therefore, a count of 400 probes per image ensured that the number of points falling on epithelium would provide sufficient precision (e-Figures 2 and 3).

The stereology assessors (RA, MC, LD) were blinded to the stent type. Each point was classified as surface epithelium, submucosal connective tissue, submucosal glands, smooth muscle, or cartilage. The point-counting rules were as follows: the grid consisted of lines and circles (e-Figure 3)**,** and any tile that fell at the inner-corner intersection of the horizontal and vertical red lines was given a value according to the type of tissue it fell upon. Only tiles that fell over the image were counted. Tiles that fell on blank space (i.e., no tissue) were not counted.

The volume density (Vv) of a given tissue type was determined by taking the percentage of all counted points that were determined to fall on that tissue type. It is important to note that stereological methods cannot estimate volumes, but rather volume densities. Stenting does not affect cartilage volume,[2] so we expressed the relative volume of each tissue type with reference to cartilage (Vv tissue / Vv cartilage). This allowed us to compare stented and nonstented areas within the same animal and between different animals.

Statistical Analysis

To analyze the data from the histologic scoring system, we used ordered logistic regression to compare the effect of DES and control stents for each type of histologic

change. For comparisons between nonstented and stented tracheal regions in the same pig using histologic scoring, we used asymptotic symmetry and marginal homogeneity tests because this was paired data. For comparisons of tracheal thickness between control stents and DES, we used the Mann-Whitney rank sum U test. For comparisons of tracheal thickness between nonstented and stented regions within the same pig, we used the signed-rank test for paired data. We used a correlation matrix to evaluate the relationship between the relative volumes of different types of tissue. We compared the variances between stented and nonstented airways using the *F* test. All reported *p* values are unadjusted. We chose an α of 0.05 to determine statistical significance, and all tests were 2-sided. We used a Bonferroni correction for multiple comparisons and report the original *p* values and the relevant threshold *p* values (α/number of comparisons), which defined what constituted a statistically significant result. When Bonferroni adjustments were used, we report confidence intervals accordingly (e.g., if the threshold *p* value is 0.001, we reported the 99.9% confidence interval). All analyses were conducted using STATA software (StataCorp LLC, College Station, TX).

Sample Size Calculation for Animal Study

The study objective was to use stereology for assessment of granulation tissue in order to obtain initial point estimates of the volume fraction of each type of soft tissue that develops in response to stenting. Our secondary goal was to use our DES to make sure it did not disintegrate or malfunction during or after placement. For pilot studies traditional sample size calculations are not always relevant but careful consideration of sample size is still warranted in order to ensure that the data generated is sufficient to

guide future studies.[11] We therefore did not do a traditional power calculation for the comparison of DES to normal. Instead we considered how precisely we needed our estimate of the volume densities to be in order for the study to be useful for future research. We did not actually know what the relative volume densities of each tissue (submucosal tissue, submucosal glands, epithelium) would be relative to cartilage and there is no prior published data in this area. Clinically bronchoscopists see the aggregate volume of tissue, so we based our calculations on the estimate of the total soft-tissue volume density relative to cartilage (VvTotal/Vvc). We estimate that in the stented areas the VvTotal/VvC=2. We wanted to be able to estimate the mean with a 95% CI margin of ±0.2. So the total width of the 95% CI would be 0.4. We estimated based on clinical experience that there is a large amount of variability between patients in terms of the volume of granulation tissue, so we estimated the standard deviation at 0.6. For a 95% CI width of 0.4 and a standard deviation of 0.6, N=35. Based on published studies of stent migration rates, we estimated that 10% of all stents would migrate and not be evaluable at study completion. Therefore the total sample size of the DES was 40. We chose a 4:1 ratio of DES to normal controls because there is abundant experience with normal controls and we were therefore interested in more precise estimates of the DES than the control stents.

Because the ultimate goal is to make better stents, we expect to go through many revised versions of the DES in the future. Each version of the DES will require quantification of its outcomes and a control arm. However, if the sample size of the control arm is equal to the DES arm in each individual experiment, then the total number of controls will be many times more than the number of observations on the

final version of the DES. This will result in more precise estimates of the control arm, but less precise knowledge of the DES arm for the *series* of all experiments planned. The return on investment in terms of increasing precision of the measurement for the normal controls is not good enough to warrant a 1:1 ratio in each individual experiment. By decreasing the control arm sample size in each individual experiment, it allows resources to be diverted to more precisely measure the intervention arm, and over the series of planned experiments the precision of the estimates of outcomes for the control arm will still gradually improve (albeit we will need to add a context variable to adjust). So for example if we conduct the same experiment after revising the DES five different times, then by version 5 of the DES there will be as many controls as DES. If each individual experiment is randomized 1:1 then costs will go out of control. Since there is no reasonable expectation that version 1 will be definitive, it is not cost-effective to do an equal number of controls for the very first experiment.

Clinical Trial Sample Size Power Calculations

Previous histologic scoring systems, as outlined above, are unsuitable for determining sample sizes for clinical trials because they are not quantitative, lack precision, and cannot combine different domains (e.g., cilia vs. epithelial erosions) into a single score. We examined our stereology data to determine whether it is usable for calculating the sample size needed to provide a defined level of statistical power in a clinical trial.

The sample size chosen for a given study depends on the statistical power required, the α level, the effect size, the underlying distribution of the data, and the statistical tests used. For our calculations of the sample size needed for a hypothetical clinical trial, we made the following choices and assumptions:

- The data are not normally distributed.
- Two-sided α = 0.05.
- Effect size was expressed using Cohen's *d* (the difference between the means divided by the standard deviation). The interpretation of Cohen's *d* has the following conventions: *d* = 0.2 indicates a small effect size, *d* = 0.5 indicates a medium effect size, and *d* = 0.8 indicates a large effect size.[9]
- Because the data were nonparametric, we used the Mann-Whitney rank sum test. To compute sample size, we used the Pitman asymptotic relative efficiency (PARE) approximation. With this method, we computed the sample size needed for a 2-sample *t* test and divided by a coefficient determined by the PARE from the underlying distribution of the data. For example, if the data were normally distributed, the coefficient would be 0.955, indicating that the Mann-Whitney rank sum test is less efficient for a normal distribution than a *t* test. Conversely, if the distribution were logistic, the coefficient would be 1.097; that is, the Mann-Whitney test would be more efficient for a distribution more likely to produce outliers. Fortunately, the PARE coefficient is never less than 0.864, so we assumed a worst-case scenario and divided by 0.864.

Results

Animals

A total of 50 swine were processed. Two were excluded during the intake period (one for seizures and the other for a respiratory infection). A total of 48 swine were randomized. There were no intraoperative complications. Deployment of the DES was easy and essentially identical to that of the control stents. One DES malfunctioned on day 2 —the inner lining separated from the stent due to failure of the adhesive and dislodged, leading to respiratory distress and death of the pig. We reinforced the adhesive between the inner lining and the outer wall in subsequent versions of the DES, which was sufficient to prevent delamination in all animals afterwards. Among animals receiving a control stent, one was found at necropsy to have no stent—presumably, this pig had coughed it out—and another had a stent that had migrated up to the vocal cords. These 5 swine were excluded from the analysis. The final cohort consisted of 36 pigs with a DES and 9 with control stents.

Microbiology

We found no difference in the number or type of bacteria cultured from the proximal normal control (nonstented) tracheal tissue and the stented tracheal samples. There was also no difference in tracheal tissue cultures obtained from animals treated with control stents or DES. All these tissues were colonized by normal gram-positive and gram-negative bacterial flora.

The DES removed at necropsy demonstrated loss of color (Figure 1 in the main text) and had a variable amount of visible residual gel, with some having no gel. This

was consistent with successful release of drug from the reservoir. When residual gel was present, we sampled it and performed zone-of-inhibition testing. The residual gel showed antimicrobial activity, with 8- to 9-mm zones of inhibition when tested against methicillin-resistant *S. aureus*. In one of the stents that contained visible residual gel, cultures from areas at the elution holes demonstrated a mean of 14 CFU comprising 2 gram-negative bacterial species. In comparison, areas of the same stent that were remote from the elution holes (either distally or proximally) demonstrated more than 1000 CFU of the same 2 gram-negative organisms.

Histology

Conventional histologic analysis found no significant differences between control stents and DES (e-Table 4). However, the importance of measuring each relevant domain and recording it separately rather than combining domains is highlighted when comparing cellular infiltrates in the proximal and distal stented areas. Comparison of the proximal and distal regions of the stented airways showed that the distal region had significantly less cilia loss ($p = 0.003$) and fewer lymphocytic infiltrates ($p = 0.0003$). If we had used the original scoring system (e-Table 3) that combined lymphocytic and PMN infiltrates into a single measure, this difference would not have been discernible. Even sparse PMN infiltrates would lead to higher scores, and the difference in lymphocytic infiltrates would go unrecognized.[2] These results underline the importance of taking primary measurements rather than aggregating categories.

Power Calculations for Clinical Trials

In this study, the mean soft-tissue relative volume per unit of cartilage volume in stented airways was 1.36 ± 0.53. The normal nonstented airway soft-tissue relative volume per unit of cartilage volume was 0.78 ± 0.56. The distributions had right-sided tails and were not normal, so we reported medians and interquartile ranges in the primary analysis, but use the mean and standard deviation for sample size calculations along with the PARE.

For our calculations of the sample size needed for a hypothetical clinical trial, we made the following assumptions based on our data:

- Mean soft-tissue volume per unit volume of cartilage with the control stent = 1.36.
- Standard deviation of the mean soft-tissue volume per unit of cartilage volume = 0.55 for both stented and control populations.
- Two-sided α = 0.05.
- Effect size was expressed using Cohen's *d* (the difference between the means divided by the standard deviation). With a standard deviation of 0.55, an effect size of 0.5 using Cohen's *d* is equivalent to the difference between the 2 means: 0.55 × 0.5 = 0.275. If the mean total soft-tissue volume per unit volume of cartilage of the proximal stented airway is 1.36, the intervention arm would have a mean of 1.36 − 0.275 = 1.085. Applied to our data, a Cohen's *d* of 0.5 represents roughly a 48% reduction of excess granulation tissue volume, with excess tissue defined as the soft-tissue volume exceeding that of an equivalent nonstented airway.
- PARE coefficient 0.864.

The sample sizes required for different effect sizes and levels of power are shown in e-Figure 6. For example, achieving 80% power to detect a moderate effect size (Cohen's *d* = 0.5) with a mean soft-tissue relative volume of 1.36, a standard deviation of 0.55, and a 2-sided α of 0.05 would require 148 cases (74 per arm), assuming a nonparametric worst-case scenario (e-Figure 6).

Design-Based vs. Model-Based Stereology

Design-based stereology refers to newer methods in stereology in which the probes and the sampling scheme are designed a priori to the investigation in a manner that the methods are independent of the size, shape, spatial orientation, and spatial distribution of the geometrical features being studied. This eliminates bias, provided the rules are followed to achieve independence.

The alternative to design-based would be model-based stereology. This means that the method uses models based on the geometric properties of the objects being studied. It requires information about the geometry of the objects being studied, and sometimes this information takes the form of a simplifying assumption. The methods can work only as well as the models truly represent the actual objects. Model-based stereology is not set up to be assumption free. Instead there is an attempt to change the data after-the-fact to make up for the biased nature of the counting/experiment.

An example of a model-based Stereologic system is the Abercrombie cell count correction. In this system, instead of developing a rule based sampling system ensuring that only one cell is counted at a time, the system plows ahead and counts cell pieces. After the data is collected a formula is used to correct for the potential bias introduced.

But for you to correct accurately, you need certain information. In this particular case, the information required is the size of the particle in the z-dimension. If your model/assumption is off, you can introduce systematic bias. An example of a design-based approach to this same problem of cell counting in a given region that is model free is the optical fractionator probe. You make no assumption about the size of the particle in the z-dimension.

There are four basic assumptions that design-based approaches *avoid:*

I. No assumption about shape.

II. No assumption about size.

III. No assumption about orientation.

IV. No assumption about distribution.

References

1 Isenberg HD. Antimicrobial Susceptibility Testing, 5.1 - Disk Diffusion Susceptibilty Testing. *Clinical Microbiolog Procedures Handbook*. ASM Press, 1995.

2 Ruegemer JL, Perkins JA, Azarow KS, O'Bryant LK, Nielsen RE, Thomas RW. Effect of the Palmaz balloon-expandable metallic stent in the trachea of pigs. Otolaryngol Head Neck Surg. 1999; **121**: 92-7.

3 Mair EA, Parsons DS, Lally KP, Van Dellen AF. Comparison of expandable endotracheal stents in the treatment of surgically induced piglet tracheomalacia. Laryngoscope. 1991; **101**: 1002-8.

4 Marquette CH, Mensier E, Copin MC, Desmidt A, Freitag L, Witt C, Petyt L, Ramon P. Experimental models of tracheobronchial stenoses: a useful tool for evaluating airway stents. Ann Thorac Surg. 1995; **60**: 651-6.

5 Grewe PH, Muller KM, Lindstaedt M, Germing A, Muller A, Mugge A, Deneke T. Reaction patterns of the tracheobronchial wall to implanted noncovered metal stents. Chest. 2005; **128**: 986-90.

6 Wiswell TE, Wiswell SH. The effect of 100% oxygen on the propagation of tracheobronchial injury during high-frequency and conventional mechanical ventilation. Am J Dis Child. 1990; **144**: 560-4.

7 Weinberg M, Sandbank J, Flumenblit Y, Klin B, Vinograd I. Tracheal reaction to three different intraluminal stents in an animal model of tracheomalacia. Journal of laparoendoscopic & advanced surgical techniques Part A. 2005; **15**: 333-7.

8 Jensen EB, Gundersen HJ, Osterby R. Determination of membrane thickness distribution from orthogonal intercepts. J Microsc. 1979; **115**: 19-33.

9 Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Lawrence Erlbaum Asociates, United States, 1988.

e-Table 1. Comparison of tracheal thickness in swine with control stents and those with drug-eluting stents.

Tracheal region	Median tracheal thickness (μm)* (IQR 25%-75%)		*P* value†
	Control stents	Drug-eluting stents	
Proximal nonstented control	501 (357-696)	606 (463-750)	0.32
Proximal stent	818 (759-952)	813 (650-941)	0.69
Mid-stent	580 (511-723)	649 (574-743)	0.18
Distal stent	722 (526-774)	654 (560-747)	0.59
Distal nonstented control	384 (315-527)	430 (356-606)	0.23

IQR, interquartile range

* Tracheal thickness was measured from the inner luminal surface of the airway epithelium to the inner aspect of the tracheal cartilage.

†*P* values determined by Mann-Whitney rank sum test; Bonferroni correction threshold *p* value is 0.05/5 = 0.01

e-Table 2. Tissue volume per unit volume of cartilage with control stents and drug-eluting stents

Tracheal region	Vv of tissue / Vv of cartilage Median (IQR 25%-75%)		*P* value*
	Control stents	Drug-eluting stents	
	Epithelium		
Proximal nonstented control	0.053 (0.041-0.076)	0.059 (0.047-0.073)	0.90
Proximal stent	0.094 (0.083-0.116)	0.106 (0.077-0.144)	0.68
Mid-stent	0.098 (0.076-0.123)	0.082 (0.066-0.099)	0.44
Distal stent	0.097 (0.650-0.140)	0.090 (0.058-0.109)	0.41
Distal nonstented control	0.056 (0.043-0.080)	0.059 (0.045-0.072)	0.95
	Submucosal glands		
Proximal nonstented control	0.141 (0.027-0.235)	0.164 (0.129-0.232)	0.56
Proximal stent	0.204 (0.173-0.393)	0.235 (0.161-0.380)	0.95
Mid-stent	0.200 (0.112-0.298)	0.258 (0.169-0.356)	0.35
Distal stent	0.218 (0.164-0.282)	0.215 (0.173-0.323)	0.82
Distal nonstented control	0.157 (0.107-0.216)	0.186 (0.137-0.251)	0.33
	Submucosal tissue		
Proximal nonstented control	0.264 (0.206-0.580)	0.340 (0.251-0.529)	0.96
Proximal stent	0.645 (0.340-0.906)	0.719 (0.585-0.947)	0.28
Mid-stent	0.386 (0.298-0.529)	0.560 (0.328-0.690)	0.08
Distal stent	0.494 (0.383-0.693)	0.528 (0.386-0.725)	0.65

Distal nonstented control	0.307 (0.203-0.595)	0.317 (0.240-0.388)	0.93
	Smooth muscle		
Proximal nonstented control	0.126 (0.051-0.139)	0.095 (0.073-0.133)	0.99
Proximal stent	0.113 (0.086-0.158)	0.147 (0.124-0.166)	0.15
Mid-stent	0.136 (0.055-0.165)	0.121 (0.104-0.160)	0.85
Distal stent	0.142 (0.104-0.163)	0.135 (0.113-0.144)	0.72
Distal nonstented control	0.098 (0.090-0.114)	0.105 (0.085-0.116)	0.62

Vv, volume density; IQR, interquartile range

**P* values determined using Mann-Whitney rank sum test; Bonferroni correction threshold *p* value is 0.05/5 = 0.01

e-Table 3. Original* histopathologic scoring system from Ruegemer et al.[2]

Parameter	Score				
	0	1	2	3	4
Intraepithelial mucus loss	Absent	Focal	Diffuse		
Cilial loss	Absent	Focal	Diffuse		
Submucosal hemorrhage	Absent	Minimal	Severe		
Surface epithelial changes	Absent	Hyperplasia	Squamous Metaplasia	Epithelial loss	
Infiltrates*	Absent	< 50 lymphocytes/HPF	> 50 lymphocytes/HPF	< 50 PMN/HPF	> 50 PMN/HPF
Epithelial erosions	Absent	1 area	2-3 areas	Extensive	Total
Necrosis	Absent	1 area	2-3 areas	Extensive	Total
Granulation tissue	Absent	Focal	Moderate	Small polyps	Large polyps
Cartilage changes	Absent	Focal	Moderate	Focal necrosis	Extensive necrosis

HPF, high-power field; PMN, polymorphonuclear cell

* Note that in the original scoring system, lymphocytes and PMN infiltrates are reported as a single variable as cells per HPF.

e-Table 4. Histopathologic scores for stented and nonstented control sections of the trachea

	OR for drug-eluting stent having more significant changes* (99.9% CI) ***P* value**				
Histologic characteristic	**Proximal nonstented trachea**	**Proximal stent**	**Mid-stent**	**Distal stent**	**Distal nonstented trachea**
Intraepithelial mucus loss	3.93 (0.29-52.3) *P* = 0.08	0.48 (0.03-7.85) *P* = 0.39	1.64 (0.14-19.0) *P* = 0.51	1.5 (0.12-19.5) *P* = 0.60	1.83×10^7 (0-∞) † *P* = 0.99
Cilial loss	12.83 (0.69-239.5) *P* = 0.004	1.55 (0.03-86.8) *P* = 0.72	2.38 (0.16-36.4) *P* = 0.30	2.16 (0.17-27.4) *P* = 0.32	3.41 (0.20-59.5) *P* = 0.16
Submucosal hemorrhage	2.07×10^7 (0-∞)ǂ *P* = 0.998	0.97 (0.07-13.8) *P* = 0.97	1.91 (0.14-25.2) *P* = 0.41	0.79 (0.07-8.3) *P* = 0.74	1.87×10^7 (0-∞) † *P* = 0.99
Surface epithelial changes	4.54 (0.40-51.2) *P* = 0.04	2.03 (0.10-39.8) *P* = 0.43	2.36 (0.14-39.8) *P* = 0.32	3.29 (0.24-45.1) *P* = 0.13	2.74×10^7 (0-∞) † *P* = 1.00
Infiltrates, PMN	9.78 (0.45-212.4) *P* = 0.015	2.81 (0.23-34.9) *P* = 0.18	3.23 (0.32-32.7) *P* = 0.10	2.19 (0.22-21.4) *P* = 0.26	16.2 (0.78-338) *P* = 0.003
Infiltrates, lymphocytes	10.27 (0.73-145.4) *P* = 0.004	1.48 (0.11-20.1) *P* = 0.62	3.67 (0.40-33.4) *P* = 0.05	1.05 (0.11-10.0) *P* = 0.94	9.91 (0.25-397) *P* = 0.04
Epithelial erosions	4.94 (0.13-194.8)	1.82 (0.18-18.4)	5.19 (0.47-57.4)	2.38 (0.23-24.6)	2.67 (0.07-109)

	P = 0.15	*P* = 0.39	*P* = 0.02	*P* = 0.22	*P* = 0.39
Necrosis	2.07×10^7 (0-∞)† *P* = 0.98	2.40 (0.24-23.7) *P* = 0.21	6.80 (0.57-80.7) *P* = 0.01	1.89 (0.18-20.0) *P* = 0.38	1.87×10^7 (0-∞)† *P* = 1.00
Granulation tissue	1.16 (0.02-53.9) *P* = 0.90	1.76 (0.15-21.2) *P* = 0.46	4.77 (0.46-49.3) *P* = 0.03	1.86 (0.19-17.7) *P* = 0.37	1.96×10^7 (0-∞)† *P* = 1.00
Cartilage changes	1‡ *P* = 1.00	1.40 (0.16-11.9) *P* = 0.60	1.04 (0.12-9.2) *P* = 0.95	2.05 (0.20-21.2) *P* = 0.31	1‡ *P* = 1.00

OR, odds ratio; CI, confidence interval; PMN, polymorphonuclear cell

* Ordered logit regression odds ratios.

† Multiple categories were possible, but all observations in the control arm were the same, so the data were completely determined and confidence intervals under ordered logit assumptions are questionable. However, the *P* value using ordered logit and Fisher's exact test were both 1.00, so no significant difference between the groups was demonstrated.

‡ All values were the same ("absent" or "no changes"), so there were too few categories for ordered logit regression. *P* value is from Fisher's exact test. With Bonferroni correction, the threshold *P* value with $\alpha=0.05/50 = 0.001$.

e-Figure 1. Tracheal thickness measurement using orthogonal lines. Note that each region of the trachea (e.g., proximal nonstented control) is divided into multiple longitudinal sections (5 shown). These were all placed on 1 slide. The individual sections are labelled topographically from bottom to top and left to right. In this case, there is only one row, the sections are labeled from left to right.

e-Figure 2. Representative image from a section of trachea showing gridlines for point counting. Each point was assigned to one of the 5 categories of tissue types. This was done using the keypad, and the software kept track of the total number of points for each tissue type (upper left). The arrows indicate the 5 different tissue types used in our model: epithelium (red), submucosal glands (blue), submucosal tissue (green), smooth muscle (black), and cartilage (orange).

e-Figure 3. Point-counting rule. The value for each point was determined at the inner-corner intersection of the horizontal and vertical red lines (black arrows). Note that the green circles did not serve a particular function in this study.

e-Figure 4. Box-and-whisker plot showing volume of tracheal soft tissue per unit volume of cartilage in nonstented control and stented airways. Soft tissue includes epithelium, submucosal glands, submucosal tissue, and smooth muscle. Vv, volume density. Signed-rank test $p < 0.0001$ (proximal nonstented region vs. proximal, mid, and distal stented regions).

Boxes represent interquartile range, horizontal line in the box is the median, and the whiskers represent the minimum and maximum values.

e-Figure 5. Correlation matrix of relative tissue volume density for (A) proximal nonstented control tracheal region and (B) proximal stented tracheal region. Numbers represent relative volumes of the indicated type of tissue per unit volume of cartilage. Bonferroni correction threshold *p* value is 0.05/10 = 0.001.

e-Figure 6. Sample sizes required for different effect sizes and levels of power. Sample size calculations assume a 2-sided α of 0.05, nonparametric Mann-Whitney rank sum test, and 1:1 randomization. The sample size shown is the total sample size, so each arm of the study would have half of the number shown.