Supplementary

Supplementary file (Part 2 paper)

https://dx.doi.org/10.21037/jtd-21-1824

Table S2-1 Adjusted perioperative morbidity and mortality studies

Ordered by confidence that results reflect the effect of the treatment, resection extent

1st Author, year (reference)	Study characteristics						Confid RE Tmt effect	Time period [b]	Perioperative mortality				Gr ≥3 Morbidity			
									Adjusted %		Adjusted HR		Adjusted %		Adjusted HR	
	Source	N	Yrs	Stage [a]	Lobe *vs.*	Comments			Seg/W	Lobe	HR	P	Seg/W	Lobe	HR	P
Linden 2014 (1)	STS	7,466 [c]	09-11	cI-IIIA	W		VH	30	1.2	1.9 [d]	-	-	4.5	9 [d]	-	-
Linden 2014 (1)	STS	5,288 [c]	09-11	cI-IIIA	W	VATS	VH	30	0.8	1.0 [d]	-	-	3.5	6.7 [d]	-	-
Linden 2014 (1)	STS	2,004 [c]	09-11	cI-IIIA	W	Open	VH	30	**1.8**	**2.9** [d]	-	-	**6.8**	**12.4** [d]	-	-
Linden 2014 (1)	STS	1,872 [c]	09-11	cI-IIIA	W	FEV1 <60%	VH	30	**1.4**	**2.9** [d]	-	-	**6.3**	**13.1** [d]	-	-
Linden 2014 (1)	STS	1,068 [c]	09-11	cI-IIIA	W	Age ≥80	VH	30	1.7	2.3 [d]	-	-	5.4	8.4 [d]	-	-
Stokes 2018 (2)	NCDB	75,114	04-13	cI-IIA	SL		H	90	3.3 [e]	3.5 [e]	0.87	-	-	-	-	-
Zhang 2019 (3)	China x10	244 [c]	14-17	cI-IIA	SL	Age ≥65, VATS	H	-	0.8	1.6	<1	NS	-	-	-	-
Bedat 2019 (4)	Swiss x2	690	14-17	Mix	Seg	VATS	M	-	0.8 [e]	0.4 [e]	-	-	9 [e]	8 [e]	0.93 [f]	NS
Khullar 2015 (5)	NCDB	20,944	03-11	cIA1,2	Seg		M	30	1.6 [e]	1.6 [e]	0.87	NS	-	-	-	-
Khullar 2015 (5)	NCDB	27,015	03-11	cIA1,2	W		M	30	1.5 [e]	1.6 [e]	0.72	.005	-	-	-	-
Tsutani 2018 (6)	Japan x1	205	07-15	cI-IIA	SL	Age ≥75	M	-	0 [e]	0.9 [e]	-	-	**5**	**10**	0.29	.048
Husain 2015 (7)	NCDB	71,171	03-11	cI-IIIA	SL		L	30	2 [e,g]	2.2 [e]	0.9	NS	-	-	-	-
Husain 2015 (7)	NCDB	19,083	03-11	cI-IIIA	SL	Age ≥75	L	30	-	-	0.76	.005	-	-	-	-

Inclusion criteria: studies using multivariable or propensity adjustment to compare perioperative morbidity/mortality after segmentectomy or wedge resection vs lobectomy, 2000–21, ≥50 patients per arm. Reference is lobectomy (HR <1 means lower morbidity/mortality for segment/wedge); Light green shading highlights statistically significant difference (lighter shade = univariable; darker = multivariable); Bold highlights differences that are somewhat clinically meaningful (see definition in Part 1 paper).

[a], 8th edition stage classification (reported stage is translated into current 8th edition nomenclature for the sake of uniformity and contemporary application); [b], time period for assessment of Morbidity and Mortality (days); [c], propensity matched pairs (total); [d], includes lobectomy and segmentectomy; [e], Unadjusted data; [f], cardiopulmonary complications (any grade); [g], data for wedge resection.

Confid RE tmt effect, Confidence that results reflect the effect of the treatment (extent of resection) *vs.* confounding factors; FEV1, forced expiratory volume in 1 second; H, high confidence; HR, hazard ratio; L, low confidence; Lobe, lobectomy; M, moderate confidence; Mix, mixture of a variety of diagnoses (NSCLC, metastases, benign); NCDB, US national cancer database; NS, not statistically significant; Seg, segmentectomy; SL, sublobar resection (segmentectomy or wedge); VATS, video-assisted thoracic surgery; VH, very high confidence; W, wedge; Yrs, years (of patient accrual).

https://dx.doi.org/10.21037/jtd-21-1824

Table S2-2 Summary of evidence in generally healthy patients with typical (i.e., solid) tumors

	Segment (*vs.* Lobe)		Wedge (*vs.* Lobe)		Wedge (*vs.* segment)	
	Effect	Conf	Effect	Conf	Effect	Conf
Short-term (90-day) outcomes						
Mortality	=	++++	=	+++	=	+
Morbidity	=	+++	=	+++	=	+
QOL 30-day	= [a]	0	= [a]	0	-	-
QOL 90-day	= [a]	0	= [a]	0	-	-
Pain VATS	= [a]	0	= [a]	0	-	-
Pain open	= [a]	0	= [a]	0	-	-
Intermediate (1-2 year) outcomes						
Δ FEV1	=	++	=/↑	0	-	-
Dyspnea	=/↑ [a]	0	=/↑ [a]	0	-	-
QOL VATS	= [a]	0	= [a]	0	-	-
Pain VATS	= [a]	0	= [a]	0	-	-
QOL open	= [a]	0	= [a]	0	-	-
Pain open	= [a]	0	= [a]	0	-	-
Long-term (5-year) outcomes						
OS	↓	+	↓↓	++	↓	+
LCSS	↓	+	↓	+	↓	+
FFR	=/↓ [a]	0	=/↓ [a]	0	-	-
LR- FFR	=/↓ [a]	0	=/↓ [a]	0	-	-

Qualitative assessment of the impact of treatment approaches on various key outcome measures and the confidence in the evidence. Differences are categorized by degree of clinically meaningful differences as defined in the legend insert. The reference (for improvement or worsening) is the treatment in parentheses.

Effect	
↑↑↑	2x meaningful improvement
↑↑	Meaningful improvement
↑	Somewhat better
=	Similar
↓	Somewhat worse
↓↓	Meaningful worsening
↓↓↓	2x meaningful worsening

Confidence in / consistency of evidence	
++++	Very High
+++	High
++	Moderate
+	Low
0	Very Low
Extpol	Extrapolation

A clinically "meaningful" difference is defined as ≥10-unit difference, with "somewhat" being half of the meaningful difference. The units of measure (for categories in parentheses) are: normalized scale points (QOL); 5-year actuarial rate (OS, LCSS); actuarial rate or simple incidence (recurrence, FFR); incidence of Gr ≥3 treatment related complications (morbidity); absolute change in % FEV1 (PFTs in compromised patients). Different thresholds of "meaningful" are: 90-day mortality (2% difference); PFTs in healthy patients (20% difference in FEV1%).

[a] data for sublobar resection not parsed out to segment or wedge.

Δ FEV1, change in FEV1 ≥6 months; Conf, confidence in the evidence; Extpol, extrapolation (indirect evidence); FFR, freedom from recurrence (only recurrence counts as an event); Gr, grade; HR, hazard ratio; LCSS, lung cancer specific survival (only death due to lung cancer counts as an event); Lobe, lobectomy; LR-FFR, locoregional freedom from recurrence; OS, overall survival; PFT, pulmonary function tests; QOL, quality of life; VATS, video-assisted thoracic surgery.

https://dx.doi.org/10.21037/jtd-21-1824

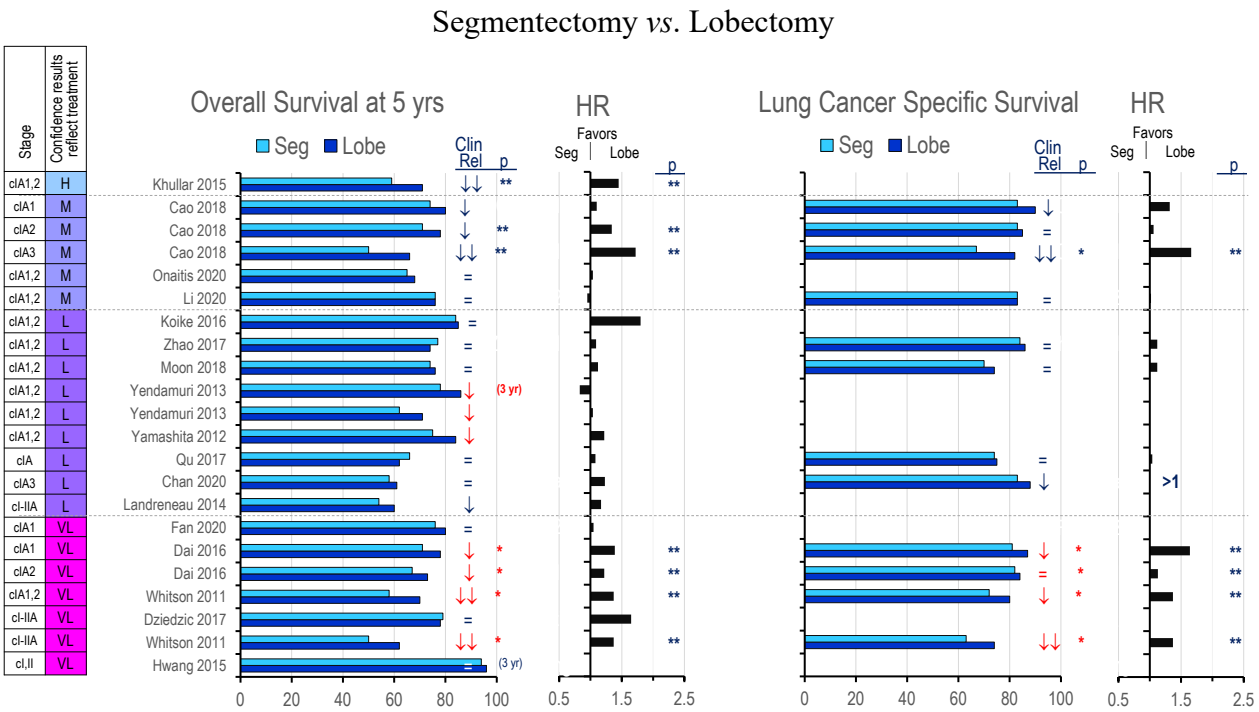

Figure S2-1 Graphic depiction of outcomes in *Table 1*: segmentectomy *vs.* lobectomy.

https://dx.doi.org/10.21037/jtd-21-1824

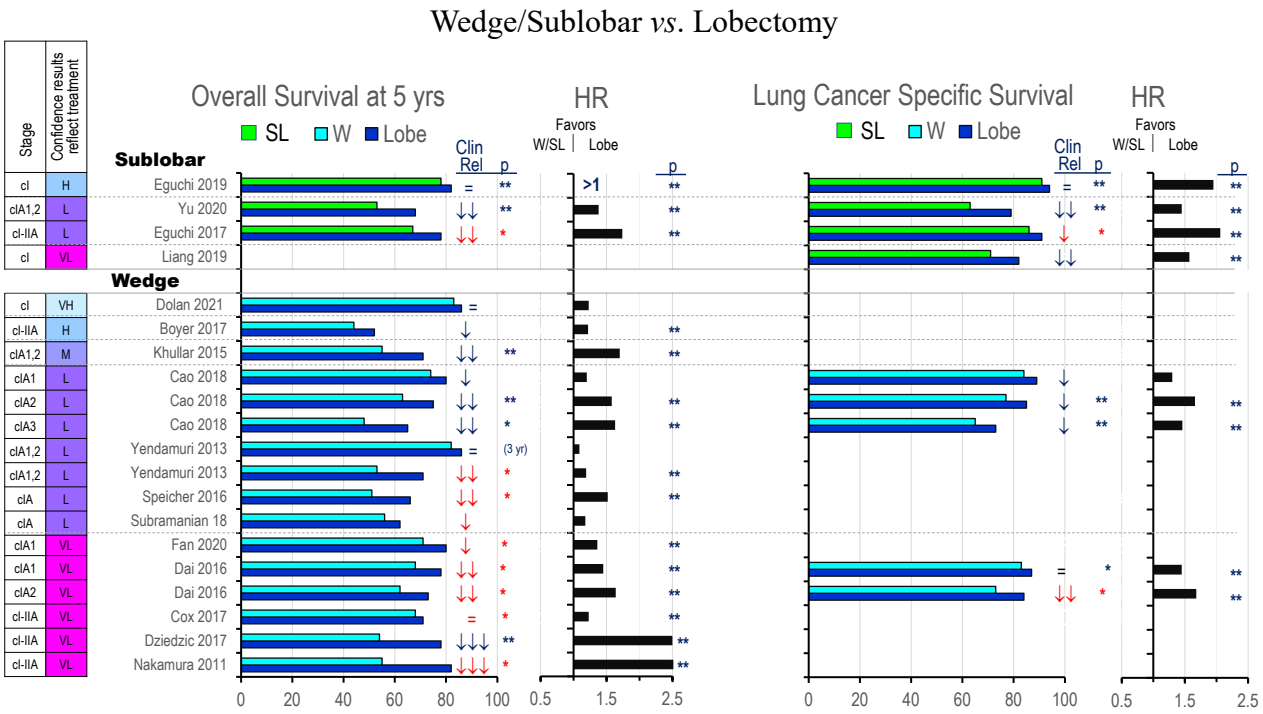

Figure S2-2 Graphic depiction of outcomes in *Table 2*: wedge/sublobar resection *vs.* lobectomy.

https://dx.doi.org/10.21037/jtd-21-1824

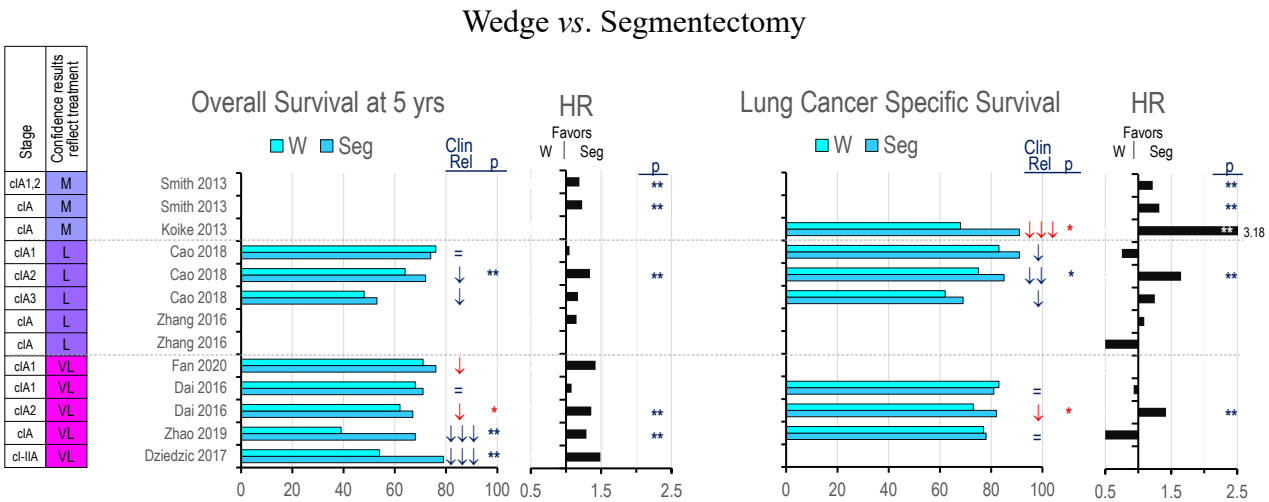

Figure S2-3 Graphic depiction of outcomes in *Table 3*: wedge *vs.* segmentectomy.

Legend (Figures S2-1,S2-2,S2-3): Graphic depiction of outcomes in *Tables 1-3*. Figure rows correspond to the respective table rows. Also depicted is the confidence that the outcomes reflect the treatment (*vs.* confounders), the level of clinical relevance and statistical significance.

Confidence results reflect the treatment	
VH	Very High
H	High
M	Moderate
L	Low
VL	Very Low
See *Table 1* for details	

Relevance of effect	
↑↑↑	2x meaningfully better
↑↑	Meaningfully better
↑	Somewhat better
=	Similar
↓	Somewhat worse
↓↓	Meaningfully worse
↓↓↓	2x meaningfully worse

The HR reference is the larger resection, i.e., HR >1 reflects worse outcome compared with lobectomy (or segmentectomy in Figure S2-3).

Red font indicates unadjusted survival rates.

* reported as statistically significant by univariable analysis; ** reported as statistically significant by multivariable analysis; Clin Rel, clinical relevance of effect. A clinically relevant difference is defined as ≥5-point difference in the 5-year actuarial rate (overall survival, lung cancer specific survival). Details of this categorization is provided in the Part 1 paper (*Tab. S1-1*) (8). HR, hazard ratio; Lobe, lobectomy; Seg, segment; SL, sublobar resection; W, wedge; yrs, years.

https://dx.doi.org/10.21037/jtd-21-1824

Appendix 2-1: Tools to assess confidence in cause and effect attribution to the interventions in question

Assessment for confounding

ROBINS-I assessment

The Risk of Bias in Non-randomized Studies of Interventions (ROBINS-I) tool was used to assess included studies (9). This validated tool has gained acceptance for observational studies. The process involves identification of domains of bias for particular interventions, assessment of each study for potential bias relative to confounders and co-interventions in each domain, and aggregation of individual assessments into an overall risk of bias across studies. Studies are categorized as "low risk" if comparable to a well-done RCT, "moderate" is sound for a NRC but not comparable to a RCT, "serious" if at least one domain is not measured or controlled, and "critical risk" if internal or external data suggests residual confounding. It is suggested that critical studies be excluded from any systematic review (9).

In application of this tool, we found few that were low risk (2%), some that were moderate (18%); most were either serious (34%) or critical risk (45%). This illustrates problematic aspects of the ROBINS-I tool for our purpose. It is a generic tool designed largely to eliminate weak evidence. However, clinical care seeks to glean whatever information can be found; valuation rather than elimination seems more conducive to gaining an understanding of the strengths and pitfalls of the full scope of evidence. Furthermore, assessing the full spectrum of adjusted NRCs promotes uncovering reasons for discrepant results and nuances of which patients, tumors, and settings provide more convincing signs of efficacy.

Adapted assessment tool specific for this project

We adapted the ROBINS-I approach to the specific nature of our project. We identified 7 domains of potential confounding (detailed below) for the major long-term outcomes. We adopted a detailed approach that allows exploration of specific areas of confounding or patient and study characteristics. We adapted the rating of confounding, shifting from eliminating studies with potential confounding to assessing the impact of confounding on attribution of outcomes to the intervention of interest. This recognizes that the impact of unaddressed confounders can sometimes be ameliorated by the setting and study characteristics.

Domains of potential confounding

Non-medical patient-related factors

Non-medical patient factors include age, sex, race, marital status, education level and income level. These factors have all been associated with long term outcomes in lung cancer patients (10,11). They can be thought of as influencing how aggressively patients want to be treated. Examples of factors that can affect the impact of such confounding include age cohorts under consideration, facility location, study region/country (i.e. that might create greater or lesser uniformity of the study cohorts).

Medical patient-related factors

Comorbidities are more common in patients diagnosed with lung cancer than in a general population of similar age (12); these can account for competing causes of death. Most often a general measure of comorbidities such as the Charlson score is available. Such composite measures don't differentiate specific comorbidities or their severity. Ideally, additional information is available (e.g. FEV1, Performance status [PS]). Co-morbidities should not impact LCSS, since only a death due to lung cancer is counted as an event. (Consistent effect for OS and LCSS argues against major comorbidity confounding for OS).

Stage accuracy

The method and thoroughness of stage assessment differs among the interventions in question (e.g., wedge resections are often Nx). Additionally, until recently the SEER database only recorded best stage (clinical for non-surgical interventions, pathologic for surgery). Mitigating factors for discrepancies in stage assessment include use of PET, invasive mediastinal staging, risk of node involvement according to tumor characteristics (size, GG component).

Study time span

Often outcome studies encompass many years. The impact of trends over time is complicated. The proportion of resections involving sublobar resection is increasing as is the use of SBRT and ablation (13-16). The use of VATS is increasing, as is PET (17,18). There is also a trend towards detecting smaller size lung cancers, and an increase in lung cancers with a ground glass component (14,19,20). All of these factors potentially confound interpretation of studies: changing nature of tumors, type of resection, type of surgeon/radiotherpist and facilities at which they are performed—all of which are associated with differences in long-term outcomes.

Examples of factors contributing to the impact include the duration of the time span, whether adjustment is dichotomized or more differentiated, whether PET was

https://dx.doi.org/10.21037/jtd-21-1824

used consistently, interactions with facility characteristics, tumor characteristics (size, GG component) and whether these are accounted for.

Setting characteristics

Facility characteristics are associated with discrepancies in the use of treatment modalities. For example, wedge resection may be associated with both the lowest volume and the highest volume hospitals, non-thoracic surgeons and nonacademic hospitals (13,21), and regional discrepancy in the use of SBRT and ablation is well documented (15). There are likely interactions between the setting and characteristics like details of pre-treatment evaluation, how tumors are detected, timeliness of care. Mitigating factors include the nature of the data source, breadth of facilities in question.

Treatment quality

Different treatment approaches may be associated with differences that affect outcomes, for example margin extent, use of adjuvant therapy, discrepancy in technical treatment factors (e.g. VATS), conversion to lobectomy if margins or nodes are concerning. All of these can produce discrepancies in factors other than the treatment intervention itself that can affect outcomes.

Favorable tumors

It is likely that tumors deemed more favorable are selected for lesser interventions (e.g. mostly GG, low PET activity, slow growth). It is clear that CT screening as well as incidental detection leads to an increased proportion of biologically more indolent tumors (22-24). Tumors with a ground glass (GG) appearance have a better prognosis (25). The presence of even a small GG component is associated with better outcomes (26,27). Prognosis correlates with the size of the solid component, not the GG component (25,28-33).

Methods of multivariable adjustment

Research involving large databases can provide an assessment of effects of a treatment in the "real world." However, ascribing an observed difference in outcomes to an intervention of interest requires assuming that nothing else is different—regarding the patients, the setting, the measurement of the outcomes etc. Since this is almost always not true, adjustment is necessary to mitigate the effect of confounding. It has become common to use propensity score analysis to accomplish this. It is worth explicitly noting several principles of this method. First, it can only adjust for known and observed factors – unmeasured factors remain a problem (e.g. severity of a condition, assessment of frailty). Second, propensity score analysis requires the assumption that any factors not included in the adjustment are "ignorable"—i.e., not associated with who will or will not receive the intervention in question (34). Indeed, derivation of the propensity score should include all factors that may be related to the outcomes and/or the treatment decision (but not those related to outcomes alone) (35). However, most outcomes studies of limited resection or SBRT have omitted adjustment for factors that are clearly related to the choice of treatment (e.g., sicker patients, favorable tumors, type of treatment facility, time period). Third, the ability of propensity scores to mitigate the effect of confounding is variable; it depends on which adjustment method is used, characteristics of the population (e.g., whether treated and control groups are markedly skewed, have a large amount of overlap or one is contained in the other, number of events) (35-37).

There are many ways of using the propensity score to adjust for confounding: the most common are (I) propensity adjustment (PA) that uses the propensity score as an additional variable in a multivariable model, (II) propensity matching (PM), involving creation of 2 subsets (treatment and control) in which each treatment patient is paired with a control patient with an equal (or nearly equal) propensity score, (III) stratification, usually into quintiles, of the entire study population (PQ), with assessment of the treatment effect in each, and 4) inverse propensity weighting (PW) in which treated patients that were less likely to be treated (and vice versa) are weighed more heavily, essentially creating an equalized pseudo-population. Which method is best depends on many factors: e.g., PM is not ideal with small samples, PW does not perform well in skewed populations, and PQ in survival analyses, but this is an oversimplification (35-37).

Because details of the propensity score development and the type of analysis affect how well the process can mitigate confounding effects, it is beneficial to perform additional analyses (different methods of adjustment, age groups, tumor size categories). Such additional analyses do not adjust for unmeasured factors or prove that they are ignorable, but if the observed effect is consistent it provides a degree of increased confidence that it is related to the intervention in question; in contrast if it is inconsistent there should be significant caution in attributing the effect in any one group to the intervention of interest. While

https://dx.doi.org/10.21037/jtd-21-1824

specific techniques can diminish some of the limitations of each method, the complexity underscores that propensity score adjustment does not guarantee that an observed effect is related to the intervention in question.

Finally, it is not clear that propensity analysis adjusts for confounders better than multivariable adjustment models (e.g., cox regression) (35,36). Multivariable regression models the relationship between multiple covariates and outcome. Because simultaneous adjustment for multiple confounders is complex, a substantial sample size is needed—it is generally accepted that about 10 events are required for each included covariate. Propensity scoring models the relationship between confounders and the treatment assignment, thus collapsing all confounders into a single propensity score. In theory, propensity techniques may have an advantage when the number of confounders is large and the number of events is small. However, analyses have not clearly demonstrated that propensity methods provide a more accurate estimate of treatment effect than multivariable methods (35,36).

Assessment process

Two individuals independently assessed each study using the adapted tool; differences were resolved by discussion or a third assessment. There was agreement in most cases or only minor differences regarding adjacent degrees of concern in individual domains. It was rare that resolution of discrepant evaluations changed the overall study rating. Results of the consensus assessment are shown in the relevant tables. Additionally, each study was assessed using the ROBINS-I tool. Our adapted rating was generally consistent with the ROBINS-I rating, although our scale allowed a more differentiated range (we avoided the threshold for a NRC of being comparable to a well-done RCT, and tried to understand critical confounders instead of a threshold of "one and you're out" approach).

Additional information

Further detail (individual rating results, reasons for ratings etc.) available if desired.

https://dx.doi.org/10.21037/jtd-21-1824

References

1. Linden PA, D'Amico TA, Perry Y, Saha-Chaudhuri P, Sheng S, Kim S, et al. Quantifying the safety benefits of wedge resection: a society of thoracic surgery database propensity-matched analysis. Ann Thorac Surg. 2014;98(5):1705-11; discussion 11-2.
2. Stokes WA, Bronsert MR, Meguid RA, Blum MG, Jones BL, Koshy M, et al. Post-Treatment Mortality After Surgery and Stereotactic Body Radiotherapy for Early-Stage Non-Small-Cell Lung Cancer. J Clin Oncol. 2018;36(7):642-51.
3. Zhang Z, Feng H, Zhao H, Hu J, Liu L, Liu Y, et al. Sublobar resection is associated with better perioperative outcomes in elderly patients with clinical stage I non-small cell lung cancer: a multicenter retrospective cohort study. J Thorac Dis. 2019;11(5):1838-48.
4. Bédat B, Abdelnour-Berchtold E, Perneger T, Licker MJ, Stefani A, Krull M, et al. Comparison of postoperative complications between segmentectomy and lobectomy by video-assisted thoracic surgery: a multicenter study. J Cardiothorac Surg. 2019;14(1):189.
5. Khullar OV, Liu Y, Gillespie T, Higgins KA, Ramalingam S, Lipscomb J, et al. Survival after sublobar resection versus lobectomy for clinical stage IA lung cancer: An analysis from the National Cancer Data Base. J Thorac Oncol. 2015;10(11):1625-33.
6. Tsutani Y, Tsubokawa N, Ito M, Misumi K, Hanaki H, Miyata Y, et al. Postoperative complications and prognosis after lobar resection versus sublobar resection in elderly patients with clinical Stage I non-small-cell lung cancer. Eur J Cardiothorac Surg. 2018;53(2):366-71.
7. Husain ZA, Kim AW, Yu JB, Decker RH, Corso CD. Defining the High-Risk Population for Mortality After Resection of Early Stage NSCLC. Clin Lung Cancer. 2015;16(6):e183-7.
8. Detterbeck F, Blasberg J, Woodard G, Decker R, Kumbasar U, Park H, et al. A Guide for Managing Patients with Stage I NSCLC: Deciding between Lobectomy, Segmentectomy, Wedge, SBRT and Ablation. Part 1: A Guide to Decision-Making. J Thor Dis. 2022.
9. Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:i4919.
10. Albano JD, Ward E, Jemal A, Anderson R, Cokkinides VE, Murray T, et al. Cancer mortality in the United States by education level and race. J Natl Cancer Inst. 2007;99(18):1384-94.
11. Wong MCS, Lao XQ, Ho KF, Goggins WB, Tse SLA. Incidence and mortality of lung cancer: global trends and association with socioeconomic status. Sci Rep. 2017;7(1):14300.
12. Cho H, Mariotto AB, Mann BS, Klabunde CN, Feuer EJ. Assessing non-cancer-related health status of US cancer patients: other-cause survival and comorbidity prevalence. American journal of epidemiology. 2013;178(3):339-49.
13. Kim AW, Detterbeck FC, Boffa DJ, Decker RH, Soulos PR, Cramer LD, et al. Characteristics associated with the use of nonanatomic resections among Medicare patients undergoing resections of early-stage lung cancer. Ann Thorac Surg. 2012;94(3):895-901.
14. McMurry TL, Shah PM, Samson P, Robinson CG, Kozower BD. Treatment of stage I non-small cell lung cancer: What's trending? J Thorac Cardiovasc Surg. 2017;154(3):1080-7.
15. Uhlig J, Ludwig JM, Goldberg SB, Chiang A, Blasberg JD, Kim HS. Survival Rates after Thermal Ablation versus Stereotactic Radiation Therapy for Stage 1 Non-Small Cell Lung Cancer: A National Cancer Database Study. Radiology. 2018;289(3):862-70.
16. Vest MT, Herrin J, Soulos PR, Decker RH, Tanoue L, Michaud G, et al. Use of new treatment modalities for non-small cell lung cancer care in the Medicare population. Chest. 2013;143(2):429-35.
17. Detterbeck FC. Maintaining Aim at a Moving Target. J Thor Oncol. 2011;6(3):417-22.
18. Detterbeck F, Molins L. Video-assisted thoracic surgery and open chest surgery in lung cancer treatment: present and future. J Vis Surg. 2016;2:173.
19. Koike T, Yamato Y, Asamura H, Tsuchiya R, Sohara Y, Eguchi K, et al. Improvements in Surgical Results for Lung Cancer from 1989 to 1999 in Japan. J Thorac Oncol. 2009;4(11):1364-9.
20. Okami J, Shintani Y, Okumura M, Ito H, Ohtsuka T, Toyooka S, et al. Demographics, Safety and Quality, and Prognostic Information in Both the Seventh and Eighth Editions of the TNM Classification in 18,973 Surgical Cases of the Japanese Joint Committee of Lung Cancer Registry Database in 2010. J Thorac Oncol. 2019;14(2):212-22.
21. Camposilvan I, Akhtar-Danesh N, Schneider L, Finley CJ. The effect of surgeon volume on procedure selection in non-small cell lung cancer surgeries. J Thorac Cardiovasc Surg. 2015;150(3):507-12.
22. Detterbeck F, Gibson C. Turning Gray: The Natural History of Lung Cancer Over Time. J Thorac Oncol. 2008;3(7):781-92.

https://dx.doi.org/10.21037/jtd-21-1824

23. Schabath MB, Massion PP, Thompson ZJ, Eschrich SA, Balagurunathan Y, Goldof D, et al. Differences in Patient Outcomes of Prevalence, Interval, and Screen-Detected Lung Cancers in the CT Arm of the National Lung Screening Trial. PLoS One. 2016;11(8):e0159880.
24. Quadrelli S, Lyons G, Colt H, Chimondeguy D, Buero A. Clinical characteristics and prognosis of incidentally detected lung cancers. Int J Surg Oncol. 2015;2015:287604.
25. Mase VJ, Jr., Detterbeck FC. Approach to the Subsolid Nodule. Clin Chest Med. 2020;41(1):99-113.
26. Hattori A, Hirayama S, Matsunaga T, Hayashi T, Takamochi K, Oh S, et al. Distinct Clinicopathologic Characteristics and Prognosis Based on the Presence of Ground Glass Opacity Component in Clinical Stage IA Lung Adenocarcinoma. J Thorac Oncol. 2019;14(2):265-75.
27. Miyoshi T, Aokage K, Katsumata S, Tane K, Ishii G, Tsuboi M. Ground-Glass Opacity Is a Strong Prognosticator for Pathologic Stage IA Lung Adenocarcinoma. Ann Thorac Surg. 2019;108(1):249-55.
28. Maeyashiki T, Suzuki K, Hattori A, Matsunaga T, Takamochi K, Oh S. The size of consolidation on thin-section computed tomography is a better predictor of survival than the maximum tumour dimension in resectable lung cancer. Eur J Cardiothorac Surg. 2013;43(5):915-8.
29. Murakawa T, Konoeda C, Ito T, Inoue Y, Sano A, Nagayama K, et al. The ground glass opacity component can be eliminated from the T-factor assessment of lung adenocarcinoma. Eur J Cardiothorac Surg. 2013;43(5):925-32.
30. Sawabata N, Kanzaki R, Sakamoto T, Kusumoto H, Kimura T, Nojiri T, et al. Clinical predictor of pre- or minimally invasive pulmonary adenocarcinoma: possibility of sub-classification of clinical T1a. European Journal of Cardio-Thoracic Surgery. 2014;45(2):256-61.
31. Tsutani Y, Miyata Y, Nakayama H, Okumura S, Adachi S, Yoshimura M, et al. Prognostic significance of using solid versus whole tumor size on high-resolution computed tomography for predicting pathologic malignant grade of tumors in clinical stage IA lung adenocarcinoma: A multicenter study. J Thorac Cardiovasc Surg. 2012;143(3):607-12.
32. Tsutani Y, Miyata Y, Mimae T, Kushitani K, Takeshima Y, Yoshimura M, et al. The prognostic role of pathologic invasive component size, excluding lepidic growth, in stage I lung adenocarcinoma. J Thorac Cardiovasc Surg. 2013;146(3):580-5.
33. Yanagawa N, Shiono S, Abiko M, Ogata S-y, Sato T, Tamura G. New IASLC/ATS/ERS Classification and Invasive Tumor Size are Predictive of Disease Recurrence in Stage I Lung Adenocarcinoma. J Thorac Oncol. 2013;8(5):612-8.
34. Freemantle N, Marston L, Walters K, Wood J, Reynolds MR, Petersen I. Making inferences on treatment effects from real world data: propensity scores, confounding by indication, and other perils for the unwary in observational research. BMJ. 2013;347:f6409.
35. Elze MC, Gregson J, Baber U, Williamson E, Sartori S, Mehran R, et al. Comparison of Propensity Score Methods and Covariate Adjustment: Evaluation in 4 Cardiovascular Studies. J Am Coll Cardiol. 2017;69(3):345-57.
36. Hade EM, Lu B. Bias associated with using the estimated propensity score as a regression covariate. Stat Med. 2014;33(1):74-87.
37. Benedetto U, Head SJ, Angelini GD, Blackstone EH. Statistical primer: propensity score matching and its alternatives. Eur J Cardiothorac Surg. 2018;53(6):1112-7.

https://dx.doi.org/10.21037/jtd-21-1824