MJA
The Medical Journal of Australia

Appendix

This appendix was part of the submitted manuscript and has been peer reviewed.
It is posted as supplied by the authors.

Appendix to: Currow DC, You H, Aranda S, et al. What factors are predictive of surgical resection and survival from localised non-small cell lung cancer? *Med J Aust* 2014; 201: 475-480.
doi: 10.5694/mja14.00365.

Appendix - Methodological details

Data sources and items

The NSWCCR operates under authority of the NSW Public Health Act (1991), which mandates cancer notification by hospitals, pathology laboratories, nursing homes and ancillary sources for invasive cancers (apart from non-melanoma skin cancers) diagnosed in the NSW residential population [14]. Operational details of the Registry are described elsewhere [14].

Registry data were extracted for this study for lung cancers coded to ICD-O-3 C34.0 – main bronchus; C34.1 – upper lobe; C34.2 – middle lobe; C34.3 – lower lobe; C34.8-overlapping lesions; and C34.9 – lung not otherwise specified [15, 19]. Extracted data items included:

- Demographic and patient characteristics - sex; date of birth; census collection district of residence at diagnosis; country of birth (classified as Australia, other English speaking, or other non-English speaking country); and ICD-10 coded cause of death [15, 16, 19]. Residential census collection district was used to classify cases by remoteness according to the Australian Standard Geographical Classification (major city, inner regional, outer regional, remote and very remote) [16, 17]; and to derive the SEIFA Index of Relative Socio-economic Disadvantage by equal-population quintile [18]. Place of residence was also classified by Local Health District (LHD) to examine differences by LHD.
- Cancer characteristics - date of diagnosis; histology type expressed as adenocarcinoma (ICD-O-3: 8140, 8211, 8230-8231, 8250-8260, 8310, 8323, 8333, 8480-8490, 8550-8551, 8560. 8570-8574 and 8576); squamous cell carcinoma (8050-8078 and 8083-8084); large cell carcinoma (8010-8021, 8123, 8046, 8082, 8022-8033, 8035, 8980); and other NSCLC (remaining codes in the 8000-8941 range, excluding 8040-8045, 8246, 8800, 8801, 8802, 8811, 8920) [15, 19]; degree of spread, using the largest degree of spread recorded within four months of diagnosis, classified as localized (confined to the primary organ site), regional (spread to regional nodes), and distant (spread to distant organ sites), using Surveillance, Epidemiology and End Results (SEER) summary stage criteria [19]; and lung location classified as main bronchus; upper, middle or lower lobe; overlapping location or not specified [15, 19].

Hospital inpatient data

Treatment data were extracted from all NSW hospitals that the cancer patients were admitted to and included:

- Co-morbidity status – inferred from hospital separation diagnoses for periods of up to five years before cancer diagnosis, by combining Charlson scores for individual diagnoses to indicate the presence (combined score≥1) or absence of co-morbidity [20].
- Resection codes – extracted for periods of up to six months after diagnosis using the Australian Classification of Health Intervention codes for wedge resection, segmental resection, lobectomy or pneumonectomy [16, 19]. Dates of resections were extracted, using the first for each case in the event of multiple resections.
- Public/Private (or Veterans) payment status - cases were classified by payment status at first hospital payment for their lung cancer management, as recorded in the NSW Admitted Patient Data Collection.

The data used in this study applied to all localized NSCLC cases diagnosed in NSW in 2003-2007. Their descriptive characteristics are shown in Tables 1 and 2.

Data linkage

Cancer Registry and hospital admitted patient data were linked by the Centre for Health Record Linkage [21, 22]. ChoiceMaker probabilistic linkage software was used and names and other identifiers were replaced with Unique Project Person Numbers (PPNs) for privacy protection in the de-identified linked data files forwarded to the data analyst. Further details of the linking protocols are described elsewhere [21, 22].

Research ethics approval for the study and for the data linkage was obtained from the NSW Population and Health Services Research Ethics Committee.

Statistical analyses

Cross-tabulations were used initially to explore associations of resection and resection type with demographic, patient, and cancer characteristics, and co-morbidity status, employing Pearson chi-square tests for variables with nominal or binary distributions (substituting the likelihood ratio chi-square for small cell sizes]), and Mann-Whitney U tests or Kruskal-Wallis ANOVA for ordinal and non-normally distributed continuous variables [23].

Multivariate logistic regression analysis was used to determine the relative odds (i.e., odds ratios) with 95% confidence limits of "resection" compared with "no resection" [23].Multi-variate competing risk regression modelling, as described by Fine and Gray [24], was used to calculate sub-hazards ratios with 95% confidence limits) for lung cancer death from time of diagnosis to death (in months) or to December 31st, 2008, whichever occurred first. Sub-hazards ratios were determined for resection types and cases not having a resection, adjusting for age at diagnosis, sex, payment status, residential remoteness, area-based socio-economic disadvantage, LHD, country of birth, histology type, and co-morbidity status. For resection cases, duration (months) from diagnosis to resection and annual resection volume of hospital were included as additional model co-variates. Co-linearity assumptions were tested and log-normal plots examined to test proportionality assumptions [23, 24]. Missing values of predictor variables were coded as dummy variables in the modelling. The stability of sub-hazards ratio estimates was tested in repeated modelling by restricting the data to cases without missing values (i.e., using "complete case" analysis). Sensitivity analysis was also undertaken using all-cause mortality rather than lung cancer death as the outcome variable.

Adjusted cumulative probabilities of survival from lung cancer were derived from the competing risk regression base model for periods of up to five years from diagnosis [24].