Sputum Cytologic Atypia Predicts Incident Lung Cancer: Defining Latency and Histologic Specificity

Tim Byers,[1] Holly J. Wolf,[1] Wilbur A. Franklin,[1] Sarah Braudrick,[1] Daniel T. Merrick,[1,2] Kenneth R. Shroyer,[1] Fred R. Hirsch,[1] Chan Zeng,[1] Anna E. Barón,[1] Paul A. Bunn,[1] York E. Miller,[1,2] and Timothy C. Kennedy[3]

[1]School of Medicine, University of Colorado Health Sciences Center, Aurora, Colorado; [2]Denver Veterans Affairs Medical Center; and [3]Lung Cancer Institute of Colorado, Denver, Colorado

Abstract

Background: **There is a need for early detection methods for lung cancer. Radiologic imaging may be more sensitive for peripheral cancers than for cancers arising in the central airways, from which bronchial epithelial cells are exfoliated into the sputum.**
Methods: **Sputum samples were collected at baseline and periodically thereafter in a cohort of smokers and former smokers with chronic obstructive lung disease. The association between cytologic atypia and incident lung cancer was assessed by hazard ratios (HR; 95% confidence intervals) using Cox regression and by odds ratios (95% confidence intervals) using logistic regression, adjusting for potential confounding factors.**
Results: **We observed 174 incident lung cancers in a cohort of 2,521 people over 9,869 person-years of observation. Risk for incident lung cancer was increased among those with cytologic atypia graded as moderate or worse (adjusted HR, 2.37; 1.68-3.34). The association between sputum atypia and lung cancer incidence was greatest for those sputum samples collected 5 months or less before the diagnosis of lung cancer (odds ratio, 10.32; 5.34-19.97). The association was substantially stronger for squamous cell lung cancers (HR, 5.13; 2.89-9.10) than for adenocarcinomas (HR, 1.85; 0.94-3.65).**
Conclusion: **Cytologic atypia is a marker for increased lung cancer risk. These cytologic changes seem to arise from late events that are most apparent for cancers arising in the central respiratory airways. Whether cytologic atypia might complement radiologic imaging in a combined approach to lung cancer, early detection requires additional evaluation of those two methods used together.** (Cancer Epidemiol Biomarkers Prev 2008;17(1):158–62)

Introduction

The potential value of earlier diagnosis of lung cancer using computerized tomographic (CT) radiologic imaging is now under intensive investigation (1-3). If CT imaging is shown to reduce lung cancer mortality with an acceptable balance of benefit and risk, it may well be found to be more sensitive for cancers arising in the peripheral lung fields than for those arising in the more central respiratory airways, where interpretation of CT imaging is less certain. Therefore, any method that might identify lesions in the central airways could be clinically useful to either complement radiologic imaging for case finding, and/or to aid in clinical decisions about the management of CT-detected abnormalities of undetermined importance.

Cytologic atypia of exfoliated cells has been shown to be associated with both prevalent and incident lung cancer (4-10). In the Johns Hopkins Lung Project, 14% of the participants with sputum cytologic atypia graded as moderate or worse later progressed to lung cancer as compared with only 3% of those without aytpia (5). In earlier studies in Colorado, we have also found cytologic atypia to predict lung cancer incidence prospectively in a cohort at high risk for lung cancer (11). At this point, however, the clinical utility of sputum cytologic atypia is uncertain, as the screening parameters of positive and negative predictive values have not been good enough for clinical utility, and there is not yet any evidence that earlier lung cancer detection by sputum cytology reduces lung cancer mortality risk (12).

As new radiologic imaging methods are being developed for lung cancer detection, it seems timely to consider whether biomarkers in the sputum might complement imaging as a modality to assist in early lung cancer diagnosis. As a study within the University of Colorado Lung Cancer Specialized Program of Research Excellence, we are following a high-risk cohort to assess whether either sputum cytology or other molecular changes in sputum might be useful early biomarkers of lung cancer risk. This is a report of our analyses of the latency between sputum cytologic changes and lung cancer incidence, and our assessment of whether the association between cytologic changes in sputum might vary according to the histologic type of lung cancer.

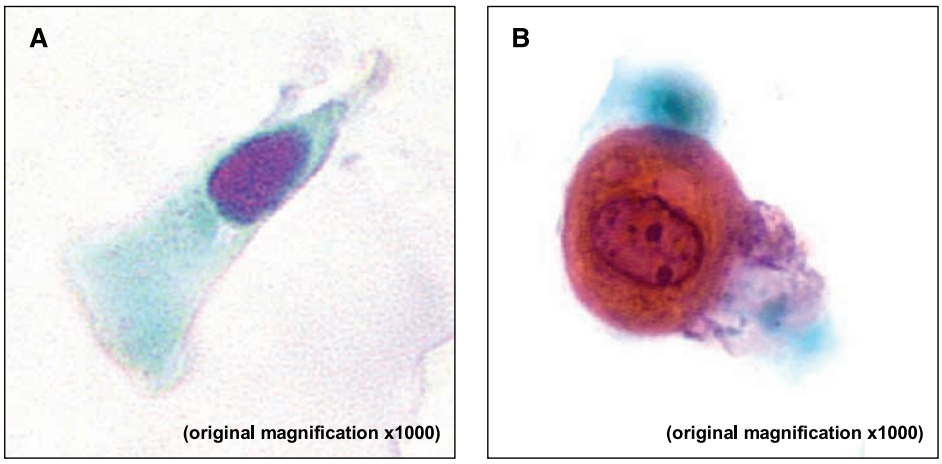
Received 5/14/07; revised 10/3/07; accepted 10/25/07.

Grant support: National Cancer Institute Specialized Programs of Research Excellence grant P50CA58187, Prevention Award from the International Association for the Study of Lung Cancer/Cancer Research Foundation of America (F.R. Hirsch), and by a research grant from the Tibotec-Virco Corporation.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: Tim Byers, School of Medicine, University of Colorado Comprehensive Cancer Center, Box F-519, 13001 East 17th Place, Aurora, CO 80045. Phone: 303-315-5169. E-mail: tim.byers@uchsc.edu

doi:10.1158/1055-9965.EPI-07-0436



Panel A: Pap stain of a normal bronchial cell in the sputum, with an eccentric nucleus and abundant apical cytoplasm.

Panel B: Pap stain of a bronchial cell in the sputum with moderate atypia, a brightly eosinophilic cell with high nuclear/cytoplasmic ratio, irregular nuclear membrane, and distinct nucleolus.

Figure 1. Examples of normal and moderate cytologic atypia in bronchial cells in the sputum.

Materials and Methods

The University of Colorado Lung Specialized Programs of Research Excellence Cohort Study is an ongoing prospective study initiated in 1993 (11). Briefly, subjects have been recruited from community and academic pulmonary clinics in the greater Denver, Colorado metropolitan area. The cohort includes adults at high lung cancer risk due to their tobacco smoking history and chronic obstructive pulmonary disease. At the time of enrollment, all subjects were ages 25 years or older, with a cigarette smoking history of 30 or more pack-years, pulmonary airflow obstruction documented by forced expiratory volume in 1 s of 75% or lower than predicted for age, and a ratio of forced expiratory volume to forced vital capacity of 0.75 or lower. Excluded were those who had a diagnosis of cancer within 5 years prior to the time of recruitment (except for non–melanoma skin cancer), a current acute respiratory infection, or who were judged by their physician to have a life expectancy of <5 years.

Patient information, tobacco history, and pulmonary function test results were obtained by interview and examination at the time of enrollment. There were 3,269 patients enrolled into the cohort as of December 31, 2006, but 729 (22.3%) of those enrollees did not submit at least one readable sputum specimen for analysis, 7 (0.2%) were missing a lung cancer diagnosis date, and 12 (0.4%) had lung cancer diagnosed at the time of the baseline examination. This analysis is therefore limited to those 2,521 cohort members who were enrolled in the cohort between 1993 and 2006, who submitted at least one sputum sample for cytologic analysis, and who were not known to have lung cancer at the time of enrollment.

Participants were provided with two containers filled with a fixative solution of 2% carbowax and 50% alcohol and instructed to collect an early morning, spontaneous cough sputum specimen for six consecutive days (3 days in the first container and 3 days in the second container). The second 3-day pooled sputum samples were those examined in this study (13). Specimens were homogenized, then centrifuged at 1,500 rpm for 15 min, and four slide smears were prepared from the resuspended cell pellets. Slides were then air-dried, fixed with 95% alcohol, and stained using the Papanicolaou technique. Slides were initially screened and categorized by trained cytotechnologists. Final diagnosis was determined in all cases by review of the screened slides by a cytopathologist, and cases were classified as not adequate for diagnosis, normal, squamous metaplasia, mild atypia, moderate atypia, severe atypia, or carcinoma (14). Interreader reliability of cytology diagnosis among cytopathologists was 85% (15). Sputum cytology diagnoses were first categorized into five groups (unreadable, normal or squamous metaplasia, mild atypia, moderate atypia, or worse than moderate atypia). Examples of grades of atypia are shown in Fig. 1.

Subsequent to their enrollment and baseline assessments, each year, all cohort members were mailed a postcard and telephoned to request their continued participation. All cohort members, even those declining to submit serial sputum samples, were also followed annually by matching to the Colorado Central Cancer Registry, to the Colorado Department of Public Health and Environment Vital Statistics records system, and to the National Death Index.

Both in previous studies by others and in our previous analyses, we used moderate or greater cytologic atypia to define the principal cytologic exposure; therefore, that is also the dichotomized cytologic classification in this analysis. Categorical variables were first examined, and differences between lung cancer cases and non-cases were assessed statistically by χ^2 testing. With age as the time scale, cohort members were included in a Cox proportional hazards regression model from age at first sputum sample, then censored at the age of death or lung cancer incidence, or at the age at last date of contact if lost to follow-up (16). Models were developed in two stages for individual risk factors and for sputum cytology. First, crude models for the association between individual risk factors and lung cancer risk were fit, then multivariate models were fit, adjusting for the potential confounding factors of gender, race, smoking status, pack-years of smoking, and year of enrollment into the cohort. In order to use information from all the sputum samples for the

Table 1. Lung cancer incidence by selected characteristics of the University of Colorado Lung Cancer Specialized Programs of Research Excellence cohort (1993-2007)

	Incident lung cancer cases	Non-cases	Person-years	Lung cancer incidence* (95% CI)	HRs for incident lung cancer	
					Crude HR (95% CI)	Adjusted HR † (95% CI)
Total cohort	174	2,347	9,869	1.76 (1.50-2.03)		
Gender						
Female	45	749	3,068	1.47 (1.04-1.90)	1.00 (Reference)	1.00 (Reference)
Male	129	1,598	6,801	1.90 (1.57-2.22)	1.21 (0.86-1.70)	1.15 (0.82-1.63)
Age at baseline (y)						
30-59	34	648	2,593	1.31 (0.87-1.75)	1.00 (Reference)	1.00 (Reference)
60-69	64	915	4,177	1.53 (1.16-1.91)	1.16 (0.76-1.75)	1.11 (0.72-1.71)
70+	76	784	3,099	2.45 (1.90-3.00)	1.95 (1.30-2.92)	2.03 (1.31-3.14)
Race/ethnicity						
Caucasian	161	2,100	9,091	1.77 (1.50-2.04)	1.00 (Reference)	1.00 (Reference)
Hispanic	6	119	360	1.67 (0.33-3.00)	1.06 (0.47-2.39)	1.19 (0.52-2.70)
African American	6	87	299	2.01 (0.40-3.61)	1.26 (0.56-2.85)	1.36 (0.60-3.08)
Other	1	41	119	0.84 (0-2.48)	0.45 (0.06-3.22)	0.51 (0.07-3.63)
Baseline smoking status						
Former smoker	86	1,293	5,457	1.58 (1.24-1.91)	1.00 (Reference)	1.00 (Reference)
Current smoker	88	1,054	4,412	1.99 (1.58-2.41)	1.63 (1.19-2.24)	1.55 (1.13-2.13)
Baseline sputum cytology						
All grades < moderate atypia	124	1,947	8,064	1.54 (1.27-1.81)	1.00 (Reference)	1.00 (Reference)
Moderate atypia or worse	50	400	1,805	2.77 (2.00-3.54)	1.86 (1.34-2.59)	1.84 (1.31-2.57)

*Rate per 100 person-years.
† Cox regression model adjusted for gender, race/ethnicity, pack-years of smoking, smoking status at baseline, and enrollment year.

analyses of sputum cytology, a time-varying covariate analysis of the repeated measures of sputum cytology was applied. To examine latency between the time of the most recent sputum collection preceding lung cancer incidence, the distribution of latency times for the samples collected closest to the date of lung cancer diagnosis was partitioned into quartiles for the cases. The first available sputum was used for controls (sputum cytologic atypia did not vary across time for controls; data not shown). For each quartile of latency time, a logistic regression model was fit using the cases in that quartile and all controls to examine the crude and adjusted associations between sputum cytology and lung cancer incidence. Covariate adjustment was made for age at enrollment, gender, smoking status, pack-years of smoking, and enrollment year. Those associations were expressed as odds ratios and their corresponding 95% confidence interval (CI). All analyses were carried out using Statistical Analysis Software (SAS, version 9.1, SAS Institute, Inc.). The study protocol was approved by the Colorado Multi-Institutional Review Board.

Results

The cohort was predominantly older (73% were age 60 or older at baseline), Caucasians (90%), males (69%), and

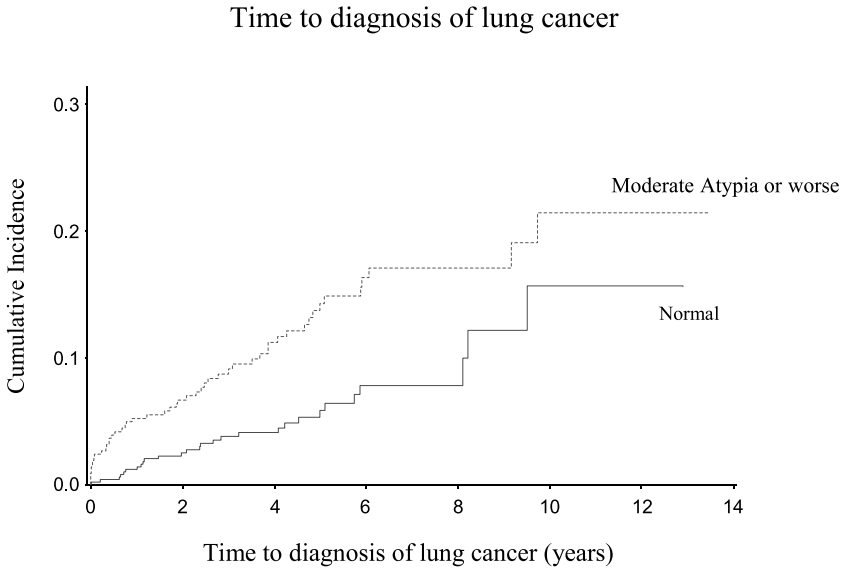


Figure 2. Cumulative distribution of time to diagnosis of lung cancer by baseline sputum cytology findings among 2,521 members of the University of Colorado Lung Cancer Specialized Programs of Research Excellence Cohort. The time scale for this variable was observation time from date of enrollment to date of lung cancer incidence, death, or last follow-up.

Table 2. The association between sputum cytologic atypia (moderate or worse vs. less than moderate) and lung cancer incidence, by the latency time between sputum collection and lung cancer diagnosis, the University of Colorado Lung Cancer Specialized Programs of Research Excellence cohort (1993-2007)

	Cases (*n*)	Non-cases (*n*)	Moderate atypia or worse (% cases/% non−cases)	Risk ratios for incident lung cancer	
				Crude HR (95% CI)	Adjusted HR* (95% CI)
All of the baseline samples (Table 1)	174	2,347	28.7:17.0	1.86 (1.34-2.59)	1.84 (1.31-2.57)
All of the available samples†	174	2,347	33.3:17.0	2.46 (1.77-3.44)	2.37 (1.68-3.34)
Time of sample collection relative to lung cancer diagnosis				Crude odds ratio (95% CI)	Adjusted odds ratio‡ (95% CI)
Quartile 1 (<5 mo)	43	2,347	67.4:17.0	10.08 (5.28-19.25)	10.32 (5.34-19.97)
Quartile 2 (5-14 mo)	43	2,347	27.9:17.0	1.88 (0.96-3.70)	1.95 (0.99-3.87)
Quartile 3 (14-38 mo)	44	2,347	27.2:17.0	1.83 (0.93. 3.57)	1.73 (0.87-3.43)
Quartile 4 (38 mo or longer)	44	2,347	11.4:17.0	0.62 (0.24-1.59)	0.57 (0.22-1.46)

*Cox regression model adjusted for gender, race/ethnicity, pack-years smoking, smoking status at baseline, and enrollment year.
† Modeled as a time-varying covariate in a Cox regression model.
‡ Logistic regression model adjusted for age at enrollment, gender, race/ethnicity, pack-years smoking, smoking status at baseline, and enrollment year.

former smokers (55%). A finding of moderate atypia or worse in the sputum was apparent among 17.8% of the cohort at the time of the baseline sample collection (Table 1). Moderate atypia or worse was more common among continuing smokers than among former smokers, but there was no association between sputum atypia and either pack-years of smoking or airflow obstruction in the baseline multivariate models (data not shown).

Among the 2,521 cohort members in this analysis, there were 174 documented incident lung cancers, for an incidence of lung cancer of 1.76 per 100 person-years in the entire analytic cohort (95%CI, 1.50-2.03; Table 1). Lung cancer risk was higher among males than among females (1.90 versus 1.47), and also higher among continuing smokers than former smokers (1.99 versus 1.58). Lung cancer incidence was also associated with cytologic atypia in the sputum sample collected at baseline [adjusted hazard ratios (HR), 1.84; 1.31-2.57]. The association between sputum cytologic atypia and lung cancer risk was largely independent of gender, time of enrollment, and tobacco use (crude HR, 1.86; adjusted HR, 1.84). The cumulative distribution of time to diagnosis of lung cancer for those with moderate or greater sputum atypia at the time of the baseline sample collection is shown in Fig. 2, which displays 1 minus the Kaplan-Meier probability of being lung cancer−free. This figure shows that the cumulative lung cancer incidence among those with moderate atypia or worse reaches 10% at ~3 years and 17% at ~6 years.

The association between sputum atypia and incident lung cancer is stronger when the repeated samples for an individual are modeled as a time-varying covariate, as compared with a single sputum sample taken at the baseline enrollment (HR, 2.37 versus 1.84; Table 2).

Table 3. The association between sputum cytologic atypia (moderate or worse vs. less than moderate) and lung cancer incidence by selected risk factors and by histologic type of lung cancer

	Controls, *n* (%)	Cases, *n* (%)	HRs for incident lung cancer	
			Crude HR (95% CI)	Adjusted HR* (95% CI)
Total cohort	2,347 (100)	174 (100)		
Gender				
Female	749 (31.9)	45 (25.9)	1.69 (0.83-3.43)	1.61 (0.78-3.33)
Male	1,598 (68.)	129 (74.1)	2.69 (1.88-3.85)	2.70 (1.89-3.88)
Age at baseline (y)†				
30-59	648 (27.6)	34 (19.5)	2.46 (1.23-4.96)	2.51 (1.20-5.24)
60-69	915 (39.0)	64 (36.8)	2.34 (1.41-3.88)	2.38 (1.42-3.98)
70+	784 (33.4)	76 (43.7)	1.87 (1.14-3.08)	1.95 (1.18-3.23)
Baseline smoking status				
Former smoker	1,293 (55.1)	98 (56.3)	1.73 (1.10-2.72)	1.77 (1.12-2.79)
Current smoker	1,054 (44.9)	76 (43.7)	3.51 (2.22-5.54)	3.42 (2.15-5.43)
Histologic type of lung cancer				
Squamous cell	2,347 (100)	48 (27.6)	5.26 (2.98-9.30)	5.13 (2.89-9.10)
Adenocarcinoma	2,347 (100)	43 (24.7)	1.91 (0.98-3.73)	1.85 (0.94-3.665)
Small cell	2,347 (100)	16 (9.2)	0.32 (0.05-2.58)	0.33 (0.04-2.55)
All other types	2,347 (100)	42 (24.1)	1.62 (0.81-3.26)	1.73 (0.86-3.47)
Unknown	2,347 (100)	25 (14.4)	3.24 (1.45-7.23)	3.14 (1.40-7.07)

*Adjusted for gender, race/ethnicity, pack-years of smoking, smoking status at baseline, and enrollment year.
† The time scale for this variable was observation time from date of enrollment to date of lung cancer incidence, death, or last follow-up.

Furthermore, when the most proximal samples are categorized by quartiles of the time between collection and diagnosis, the association is strongest for the cases whose samples were collected within 5 months of diagnosis (adjusted odds ratio, 10.32; 5.34-19.97). The association between atypia and lung cancer is somewhat stronger for men and for current smokers, but the association is considerably stronger for squamous cell cancers (HR, 5.13; 2.89-9.10; Table 3).

Discussion

This ongoing cohort study continues to confirm our earlier observation that cytologic atypia predicts incident lung cancer (11). This analysis strongly suggests that atypia is a late event, however, and that it is most strongly associated with squamous cell cancers. Cytologic atypia could be associated with increased lung cancer risk due to morphologic changes in bronchial epithelial cells that correlate with broad changes in the bronchial epithelium rather than that they arise from cancers per se. Indeed, the observation that atypia is modestly associated with incident lung cancer, even several years before diagnosis, suggests that cytologic morphology could result from field effects in the bronchial epithelium that are linked to increased lung cancer risk. However, our observation that this association is considerably stronger for samples collected within a few months of the diagnosis of lung cancer suggests that late events, perhaps lung cancer itself, also exfoliate abnormal cells into the sputum.

Our observation of a much stronger association between sputum atypia and squamous cell lung cancer is consistent with the notion that either cancers per se or fields of abnormal cells that are found with cancer are more likely to exfoliate abnormal cells into the sputum if they are located in the more central airways than in the periphery of the lung. Because CT imaging may well be found to perform better for detecting peripheral cancers than those in the larger, central airways, biomarkers of lung cancer risk in the sputum may be found to be complementary to CT imaging (1, 2).

An important weakness of this study is also a weakness of the entire field of cytology—the uncertainties of pattern recognition and classification by cytopathologists (17). Importantly, however, as these readings were all made prior to the diagnosis of lung cancer, so the cytologic readings were not biased relative to the later diagnosis and any misclassification in the readings would have biased our findings toward the null.

In conclusion, this prospective study of individuals at high risk for lung cancer shows that sputum cytologic atypia predicts increased lung cancer risk. These findings may be due both to the exfoliation of cytologically abnormal cells coming from fields of abnormal respiratory epithelium that give rise to lung cancer, and/or from early lung cancers themselves. Because cytologic morphology might eventually be found to have insufficient sensitivity and specificity for early changes leading to lung cancer, we are examining other biomarkers in the sputum in this same cohort, including fluorescence *in situ* hybridization assays of chromosomal changes and methylation of various cancer-related genes (18, 19). Further studies are needed to assess the clinical utility of multiple biomarkers in sputum, including changes in cytology as well as changes in chromosomal and epigenetic biomarkers in the sputum. We suspect that these biomarkers may together constitute a useful complement to CT radiologic imaging for early lung cancer diagnosis.

References

1. Henschke CI, Yankelevitz DF, Libby DM, et al.; International Early Lung Cancer Action Program Investigators. Survival of patients with stage I lung cancer detected on CT screening. N Engl J Med 2006; 355:1763–71.
2. Mulshine JL, Sullivan DC. Clinical practice. Lung cancer screening. N Engl J Med 2005;352:2714–20.
3. Bach P, Jett J, Tockman M, Swensen S, Begg C. Computed tomography screening and lung cancer outcomes. JAMA 2007;297: 995–7.
4. Flehinger BJ, Melamed MR, Zaman MB, Heelan RT, Perchick WB, Martini N. Early lung cancer detection: Results of the initial (prevalence) radiologic and cytologic screening in the Memorial Sloan-Kettering Study. Am Rev Respir Dis 1984;130:555–60.
5. Frost JK, Ball WC, Levin ML, et al. Early lung cancer detection: results of the initial (prevalence) radiologic and cytologic screening in the Johns Hopkins Study. Am Rev Respir Dis 1984;130:549–54.
6. Fontana RS, Sanderson DR., Taylor WF, et al. Early lung cancer detection: results of the initial (prevalence) radiologic and cytologic screening in the Mayo Clinic Study. Am Rev Respir Dis 1984;130: 561–5.
7. Tockman M. Advances in sputum analysis for screening and early detection of lung cancer. Cancer Control 2000;7:19–24.
8. Kubik A, Parkin DM, Khlat M, Erban J, Polak J, Adamec M. Lack of benefit from semi-annual screening for cancer of the lung: follow-up report of a randomized controlled trial on a population of high-risk males in Czechoslovakia. Int J Cancer 1990;45:26–33.
9. Hirsch FR, Franklin WA, Gadzar AF, Bunn PA, Jr., Early detection of lung cancer: clinical perspectives of recent advances in biology and radiology. Clin Cancer Res 2001;7:5–22.
10. Petty TL, Tockman MS, Palcic B. Diagnosis of roentgenographically occult lung cancer by sputum cytology. Clin Chest Med 2002;23:59–64.
11. Prindiville S, Byers T, Hirsch F, et al. Sputum cytological atypia as a predictor of incident lung cancer in a cohort of heavy smokers with airflow obstruction. Cancer Epidemiol Biomarkers Prev 2003;12:987–93.
12. Marcus P, Bergstralh E, Fagerstrom R, et al. Lung cancer mortality in the Mayo Clinic Lung Project: impact of extended follow-up. J Natl Cancer Inst 2000;92:1308–15.
13. Kennedy TC, Proudfoot SP, Piantadosi S, et al. Efficacy of two sputum collection techniques in patients with air flow obstruction. Acta Cytol 1999;43:630–6.
14. Kennedy TC, Proudfoot SP, Franklin WA, et al. Cytopathological analysis of sputum in patients with airflow obstruction and significant smoking histories. Cancer Res 1996;56:4673–8.
15. Archer PG, Koprowska I, McDonald JR, Naylor B, Papanicolaou GN, Umiker WO. A study of variability in the interpretation of sputum cytology slides. Cancer Res 1966;26:2122–44.
16. Cox DR. Regression models and life table. J R Stat Soc 1972;B34: 187–220.
17. Holiday DB, McLarty JW, Farley ML, et al. Sputum cytology within and across laboratories. A reliability study. Acta Cytol 1995;39:195–206.
18. Varella-Garcia M, Kittelson J, Schulte A, et al. Multi-target interphase fluorescence *in situ* hybridization assay increases sensitivity of sputum cytology as a predictor of lung cancer. Cancer Detect Prev 2004;28:244–51.
19. Belinsky S, Liechty K, Gentry F, et al. Promoter hypermethylation of multiple genes in sputum precedes lung cancer incidence in a high risk cohort. Cancer Res 2006;66:3338–44.