Inflammation in chronic obstructive pulmonary disease: its assessment and the effects of corticosteroids
Boorsma, M.

Citation for published version (APA):
Chapter 4

Improved validity of cellular and molecular biomarkers from whole induced sputum in asthma and COPD

Martin Boorsma¹, Marianne A. van de Pol¹,², Julia G. Koopmans¹, Barbara Smids¹,², Tamara Dekker¹,², Henk M. Jansen¹, Theo A. Out,² Peter Sterk¹, Jaring S. van der Zee¹, René E. Jonkers¹, René Lutter¹,²

¹Department of Pulmonology and ²Department of Experimental Immunology, Academic Medical Center, Amsterdam

Parts of the data were presented at the 2005 congress of the European Respiratory Society (Boorsma M, Koopmans J, van de Pol MA, Jansen HM, van der Zee JS, Jonkers RE, Lutter R. Are invalid induced sputum samples characterized by 80% or more of squamous cells? [Abstract]. Eur Respir J. 2005;26 (Suppl 49):373s ).
Abstract

Background: Biomarkers in induced sputum samples are increasingly important in phenotyping and monitoring of patients with asthma and COPD. Whole induced sputum samples, however, are contaminated with saliva as reflected by the presence of squamous epithelial cells. We aimed to improve the validity of sputum biomarkers in asthma and COPD by taking contamination into account.

Methods: Total and differential cell counts, levels of soluble markers of inflammation (myeloperoxidase, interleukin-8 and eosinophil cationic protein) and airway permeability (alpha-2-macroglobulin and albumin) in 482 sputum samples, i.e. 247 samples of 29 COPD patients (mean age: 64 yr; FEV1 predicted: 61%) and 235 samples of 25 asthma patients (25 yr; FEV1:102%) were related to squamous cells.

Results: For sputum samples with ≤90% squamous cells, absolute cell numbers (TCC) and levels of soluble parameters showed inverse log-linear relationships with percentage squamous cells, indicative of dilution, in COPD and asthma. TCC and soluble parameters were calculated to be reduced 8- and 1.4 to 3.2-, and 2.5- and 1.3 to 2.1-fold in samples with 48% and 20% (i.e. median %) squamous cells in asthma and COPD, respectively. These log-linear relationships enabled correction of sputum data for saliva contamination, reducing variability of sputum parameters, and improving repeatability of sputum parameters (improved intraclass correlation coefficients) and the discriminative power in two former patient studies.

Conclusions: These findings show that adjustment for percentage squamous cells significantly improved validity of biomarkers in whole induced sputum in asthma and COPD, advancing the implementation of cellular and molecular monitoring in airway disease.
Background

Analysis of induced sputum, which is produced by subjects after being exposed to nebulized saline, is a safe and relatively non-invasive method to assess airway inflammation and effects of interventions\(^1\)\(^-\)\(^5\). Standardization of methods for sputum induction, collection and analyses has consolidated and extended its use\(^6\)\(^-\)\(^10\). The analysis of sputum, however, is complicated by the potential contamination with secretions from the upper airways and oral cavity, such as saliva. The presence of squamous epithelial cells in whole sputum samples is considered an indicator of contamination with saliva and various percentages of squamous cells (15%, 20%, 50% and most often 80% of total cell count) have been taken to designate sputa valid or invalid for further use\(^6\)\(^;\)\(^11\)\(^-\)\(^17\). In studies assessing cell numbers and inflammatory parameters in paired sputum and saliva samples, saliva was found to contain low cell numbers (predominantly squamous epithelial cells), a low protein content and very low amounts of soluble inflammatory markers, compared to sputum\(^18\)\(^;\)\(^19\). Thus contamination of sputum with saliva primarily leads to dilution reducing the amounts of soluble parameters and the number of inflammatory cells per g sputum, but to which extent is unknown. Provided that saliva contains predominantly squamous cells, contamination with saliva will not affect cell differentials for non-squamous inflammatory cells. Contamination with large quantities of squamous cells, however, can physically obscure the detection of inflammatory cells on cytospin slides and thus prohibit accurate assessment of cell differentials.

It is not clear whether any parameter and, if so, which parameter in sputum samples can serve as a measure of dilution by saliva. Previously, albumin content in airway secretions has been taken as a measure of dilution, based on relative stable albumin levels in serum and limited variation in permeation across the epithelial barrier at stable disease\(^20\)\(^;\)\(^21\). However, correction for albumin in sputum is controversial since albumin permeation may fluctuate with disease severity and treatment. We reasoned (i) that dilution of sputum samples by saliva proportionally reduces values of sputum parameters and hypothesized (ii) that squamous epithelial cells in sputum samples can serve as a measure of dilution. Therefore, levels of cellular and soluble parameters were expressed relative to squamous cells for a large set of whole sputum samples from asthma and COPD patients obtained in two clinical studies\(^22\)\(^;\)\(^23\). Percentage (\(\leq 90\%\)) of squamous cells and inflammatory parameters were correlated in a log-linear fashion, which facilitated correction of sputum samples for dilution by saliva. This correction improved the repeatability of sputum parameters and the discriminative power in two former studies.
Methods

Subjects and data sets

Data sets from two prospective clinical studies\textsuperscript{22,23}, one with asthma and one with COPD patients, were analyzed. COPD patients participated in a 16-months study and samples (maximally 14) were obtained after placebo or systemic and inhaled corticosteroid treatment and some during exacerbations, referred to as different study conditions in the text. Patients were between 40 and 75 years, with middle-age onset of symptoms, a cigarette consumption of $\geq$ 15 Pack-Years, a FEV$_1$/VC ratio $\leq$ 0.70 and FEV$_1$ reversibility $\leq$ 11% of predicted. Asthma patients participated in a 6-weeks study and samples (maximally 10) were obtained after treatment with an inhaled corticosteroid and a long-acting beta-agonist or placebo, before and after allergen provocation\textsuperscript{22}, also referred to in the text as different study conditions. Patients were between 18 and 50 years, never- or ex-smokers with $\leq$ 10 Pack-years, with a FEV$_1$ $\geq$ 80% of predicted and PC$_{20}$ histamine $\leq$ 8 mg/ml. Both studies were approved by the medical ethics committee of the Academic Medical Center; written informed consent was obtained.

Statistical analyses

Sputum data (except % neutrophils and % squamous cells) were base 10 log-transformed prior to all analyses in order to obtain normal distributions. Values of 0% eosinophils were arbitrarily assigned 0.05% before log-transformation. Non-squamous Total Cell Count in 9 samples with 100% squamous cells were arbitrarily assigned 0.01x10$^6$/g. Differences in the presence of squamous cells and albumin content under the different study conditions (for explanation see Subjects and data sets) were checked by one-way analysis of variance of % squamous cells and of albumin levels over the study conditions within each patient group. Pearson’s test was used to correlate inflammatory parameters, linear regression was performed on the data (log-transformed when appropriate) versus percentage squamous cells and versus log-transformed albumin levels. Bonferroni’s correction was applied to compensate for multiplicity when analyzing 8 parameters from the same sputum sample. P-values below 0.00625 were considered statistically significant. An explorative Mixed Models analysis was performed post hoc to investigate whether incorporating multiple samples from the same patient (despite sampling under different study conditions) had a significant impact on the correlation between non-squamous total cell count and % squamous cells. Similarly, an explorative analysis was performed with only one sample, i.e. the first sample, per patient.

On basis of the observed correlations we were able to correct sputum parameters for dilution by saliva. A regression coefficient was calculated for each parameter. We used the formula $y = a + b \cdot x$ to calculate the theoretical (log-transformed) value of the parameter in non-diluted sputum at 0% squamous cells ($y$), from the (log-transformed) measured value of the parameter ($a$), the (positive) regression coefficient for that parameter ($b$) and
the % of squamous cells (x). The extent of dilution was calculated with 10^{b.x}. Regression coefficients were also calculated for the relation between log-transformed inflammatory parameters and log-transformed albumin levels and the magnitude of dilution was determined relative to log-transformed albumin levels over the range from the 5th to the 95th percentile (2.5 to 628 μg/g sputum). For graphic display, data is also shown within ten subsets of data points of equal size for increasing % squamous cells (cut-off values 4.0, 8.7, 14.4, 23.0, 30.0, 45.0, 63.0, 78.4 and 90.0%) and for increasing albumin content (cut-off values 3.3, 7.1, 14.9, 26.3, 41.8, 61.1, 88.3, 142.0 and 281.7 μg/g).

In the post-hoc analyses, repeatability (intraclass correlation coefficient), within-patients variability (standard deviation of the absolute difference between the two samples of the log-transformed data) and the between-patients variability (standard deviation of log-transformed data in the first sample) before and after correction for the dilution on the basis of % squamous cells were tested by t-test.

Results

Patients

Data of 247 and 235 induced sputum samples, from 29 patients with COPD and 25 with asthma respectively, were analyzed. Patient characteristics and baseline sputum data are given in Table 1.

Percentage squamous epithelial cells.

The median % squamous epithelial cells was higher in samples from asthma patients than from COPD patients (48% versus 20%, p<0.001): 22% of asthma samples and 15% of COPD samples contained ≥80% squamous cells. In 9 samples (all from 4 patients with COPD) there was a surplus of non-discernible, mainly squamous cells, for which % squamous cells was arbitrarily set at 100%. The % squamous cells in subsequent samples from most patients differed widely, but for some patients the % squamous cells were similar in all samples (see supplemental figures S1 and S2). There were no significant differences in % squamous cells in samples obtained under different study conditions (see ‘Subjects and data sets’ in Methods; p=0.38 and p=0.48 for asthma and COPD, respectively; see supplemental figures S3 and S4).

Absolute and relative cell counts

Non-squamous Total Cell Count (TCC) decreased significantly in a log-linear mode with increasing % squamous cells, both for COPD samples (r = -0.82, p<0.001; Figure 1A) and asthma samples (r = -0.85, p<0.001, Figure 1B). The regression lines for log-transformed TCC versus % squamous cells showed regression coefficients of -0.020 (95% Confidence Interval (CI): -0.022, -0.018) for COPD samples and -0.017 (95% CI: -0.018, -0.016) for
A post hoc Mixed Model analysis showed that correction for taking multiple samples from the same patient yielded similar log-linear relationships as shown in Figures 1A and B, with a regression coefficient of \(-0.0189\) (\(p < 0.001\)), corresponding with a 78-fold dilution at 100% squamous cells and non-significant differences in regression coefficients between patients with COPD or asthma (\(p=0.21\)). Similar significant relationships as shown in Figures 1A and B were found when the analysis was restricted to the parameter values for the first sputum sample of each patient (\(r= -0.70\) for COPD and \(r= -0.90\) for asthma, both \(p<0.001\), see supplemental figure S5).

Table 1. Demographic and baseline sputum data of patients at enrolment into the study.

<table>
<thead>
<tr>
<th></th>
<th>COPD</th>
<th>Asthma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male / female</td>
<td>21/8</td>
<td>8/17</td>
</tr>
<tr>
<td>Age (yrs)</td>
<td>64.7 (51 - 76)</td>
<td>25.2 (19 - 35)</td>
</tr>
<tr>
<td>FEV(_1) (% predicted)</td>
<td>61.1 (29 - 97)</td>
<td>102.0 (79 - 120)</td>
</tr>
<tr>
<td>Total Cell Count (10(^6) / g sputum)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- including squamous cells</td>
<td>1.90 (0.30 – 21.8)</td>
<td>1.04 (0.25 – 3.64)</td>
</tr>
<tr>
<td>- excluding squamous cells</td>
<td>0.98 (0.11 – 21.5)</td>
<td>0.58 (0.02 – 3.48)</td>
</tr>
<tr>
<td>Squamous cells</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- %</td>
<td>19.6 (1.2 – 91.3)</td>
<td>48.0 (3.6 – 95.1)</td>
</tr>
<tr>
<td>- 10(^6) / g sputum</td>
<td>0.35 (0.02 – 4.15)</td>
<td>0.33 (0.10 – 1.12)</td>
</tr>
<tr>
<td>Neutrophils (% of non-squamous cells)</td>
<td>70.6 (31.1 – 96.5)</td>
<td>30.0 (4.1 – 83.5)</td>
</tr>
<tr>
<td>Eosinophils (% of non-squamous cells)</td>
<td>1.2 (0.1 – 21.6)</td>
<td>5.2 (0 – 72.2)</td>
</tr>
<tr>
<td>Myeloperoxidase (µg/g sputum)</td>
<td>9.4 (0.4 – 188)</td>
<td>0.74 (0.16 – 10.5)</td>
</tr>
<tr>
<td>Interleukin-8 (ng/g sputum)</td>
<td>5.0 (0.4 – 174)</td>
<td>0.25 (0.10 – 1.73)</td>
</tr>
<tr>
<td>Eosinophil Cationic Protein (µg/g sputum)</td>
<td>322 (10 – 38260)</td>
<td>57.0 (4.8 – 322)</td>
</tr>
<tr>
<td>Albumin (µg/g sputum)</td>
<td>59.2 (1.6 – 2379)</td>
<td>30.6 (0.89 – 97.4)</td>
</tr>
<tr>
<td>alpha-2-Macroglobulin (µg/g sputum)</td>
<td>1.10 (0.34 – 50.4)</td>
<td>1.35 (0.21 – 8.92)</td>
</tr>
</tbody>
</table>

Data expressed as absolute numbers or mean (range) for age and FEV\(_1\) and for all sputum data as median (95% Confidence Interval), obtained in a stable state without corticosteroid treatment; FEV\(_1\): post-bronchodilator forced expiratory volume in the first second.
Figure 1. Non-squamous Total Cell Count ($10^6$/g sputum, log-transformed) expressed as a function of % squamous cells in sputum samples as individual data from patients with COPD (A, top) and from patients with asthma (B, middle), and grouped in ten equal sized subsets with increasing % squamous cells (C, bottom).

Figure 2. Eosinophil count (% log-transformed) expressed as a function of % squamous cells in sputum samples as individual data from patients with COPD (A, top) and from patients with asthma (B, middle), and grouped in ten equal sized subsets with increasing % squamous cells (C, bottom). Values of 0% eosinophils have been assigned arbitrarily the value of 0.05%.
The box-plot figure indicates that variation of TCC data per subset is similar for all decades with increasing % squamous cells (see Methods) except for the subset with % squamous cells above 90%, both in COPD and asthma (Figure 1C).

The % neutrophils did not significantly change with increasing % squamous cells, both for COPD ($r = 0.002$, $p = 0.97$) and asthma samples ($r = 0.05$, $p = 0.48$). The % eosinophils slightly decreased with increasing % squamous cells ($r = -0.25$ and $-0.34$, for COPD and asthma samples respectively, both $p<0.001$, Figure 2). However, this was mainly due to samples with >90% squamous cells containing no eosinophils at all. Excluding these samples, % eosinophils marginally, decreased with increasing % squamous cells ($r = -0.16$, $p=0.017$ and $r = -0.14$, $p=0.039$, respectively for COPD and asthma).

### Soluble inflammatory markers

Log-transformed data for MPO, IL-8, ECP, A2M and albumin showed linear decreases with increasing % squamous cells. Data for ECP and MPO are shown in Figure 3 and 4 and are representative of other markers. Fold-dilution for soluble parameters at 100% squamous cells relative to 0% squamous cells are shown in Table 2. We could not calculate fold-dilution for IL-8 for asthma as in half of these samples the IL-8 level was below the detection limit. Samples with undetectable IL-8, however, had significantly higher % squamous cell counts than samples with detectable IL-8 (mean 55% versus 45%, t-test $p<0.01$). At median squamous cell counts the fold-dilution for soluble parameters ranged between 1.4 and 3.2-fold in asthma, in COPD between 1.3 and 2.1-fold.

### Table 2. Calculated maximal dilution of inflammatory parameters in sputum samples with 100% squamous cells.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>COPD</th>
<th>Asthma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-squamous Total Cell Count</td>
<td>100 (63 – 158)</td>
<td>50 (40 – 63)</td>
</tr>
<tr>
<td>% Neutrophils</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>% Eosinophils</td>
<td>4.0 (2.0 – 10)</td>
<td>7.9 (4.0 – 16)</td>
</tr>
<tr>
<td>Albumin (Alb)</td>
<td>10 (6.3 – 20)</td>
<td>3.2 (2.0 – 6.3)</td>
</tr>
<tr>
<td>alpha-2-Macroglobulin (A2M)</td>
<td>4.0 (2.5 – 7.9)</td>
<td>2.5 (1.6 – 5.0)</td>
</tr>
<tr>
<td>Myeloperoxidase (MPO)</td>
<td>20 (13 – 40)</td>
<td>2.0 (1.6 – 3.2)</td>
</tr>
<tr>
<td>Interleukin-8 (IL-8)</td>
<td>25 (10 – 50)</td>
<td>N/A</td>
</tr>
<tr>
<td>Eosinophil Cationic Protein (ECP)</td>
<td>40 (20 – 79)</td>
<td>10 (6.3 – 16)</td>
</tr>
</tbody>
</table>

Data from 247 COPD samples and 235 asthma samples, dilution expressed as n-fold decrease (95% Confidence Interval) at 100% squamous cells relative to 0% squamous cells; % neutrophils and % eosinophils as % of non-squamous cells; N/A: not applicable, as there is no significant correlation with % squamous cells; #: data expressed as absolute decrease and 95% Confidence Interval; *: relation became non-significant (N/A) when restricted to samples with <90% squamous cells.
Figure 3. Eosinophil cationic protein (ECP) levels (µg/g sputum, log-transformed) expressed as a function of % squamous cells in sputum samples as individual data from patients with COPD (A, top) and from patients with asthma (B, middle), and grouped in ten equal sized subsets with increasing % squamous cells (C, bottom).

Figure 4. Myeloperoxidase (MPO) levels (µg/g sputum, log-transformed) expressed as a function of % squamous cells in sputum samples as individual data from patients with COPD (A, top) and from patients with asthma (B, middle), and grouped in ten equal sized subsets with increasing % squamous cells (C, bottom).
Correction using % squamous cells

In line with the negative correlation between inflammatory markers and % squamous cells for dilution by saliva we found a positive correlation between inflammatory parameters and albumin levels. This indicates that albumin may too be taken as a surrogate marker of dilution with saliva. We argued that correction for dilution on basis of % squamous cells (see Methods) would make parameters independent of albumin levels, which was found. (Figure 5, top versus bottom). In contrast to our findings for % squamous cells, albumin levels tended to differ between study conditions (p=0.066 and p=0.095 for COPD and asthma samples, respectively).

Figure 5. Effect of correction for dilution on the relationship between non-squamous total cell count (TCC, $10^6$/g sputum, log-transformed, left), eosinophil cationic protein (ECP, $\mu$g/g sputum, log-transformed, middle), and myeloperoxidase (MPO, $\mu$g/g sputum, log-transformed, right) with albumin level ($\mu$g/g, log-transformed, split in ten equal subsets of equal size with increasing albumin level), in sputum samples from patients with COPD (open bars) and asthma (filled bars). Top: without correction, bottom: with correction using % squamous cells (see Methods).
Post-hoc analyses

To determine the effect of correcting sputum data for contamination with saliva we re-analyzed data from a previously published study on repeatability of sputum data. In that study, sputum was obtained twice within one week in 21 clinically stable COPD patients. The total number of sputa in this sub-analysis comprised 17% of the COPD sputa that were used for the correlation studies. Nine inflammatory parameters were studied: non-squamous TCC, numbers of neutrophils, eosinophils and macrophages per g sputum, levels of MPO, IL-8, ECP, A2M and Alb. Data were corrected for % squamous cells with a mean regression coefficient of 0.0155 (see Methods). Correction resulted in a markedly improved repeatability (higher intraclass correlation coefficients), and in smaller within-patients and between-patients variability (all p<0.01) (Table 3). A second post-hoc analysis was done on data from an intervention study on an allergen challenge in allergic asthma patients. Data were obtained from 43% of the asthma sputa of the correlation study. The effect of intervention on the number of eosinophils and ECP levels per g sputum were analyzed before and after correction. Previously we observed no difference (p=0.3) after single-dose pretreatment with salmeterol/fluticasone propionate compared to fluticasone alone. After correcting the data for dilution we observed a tendency (p=0.06) towards a reduced increase of sputum eosinophil counts at 24 hours following the allergen provocation in patients treated with salmeterol/fluticasone as compared to with fluticasone alone.

Table 3. Repeatability of sputum data without and with correction for dilution using % squamous cells

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Ri Within-patient variability*</th>
<th>Between-patients variability#</th>
<th>Ri after correction$</th>
<th>Within-patient variability after correction$</th>
<th>Between-patients variability after correction$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-squamous TCC (10^6 /g)</td>
<td>0.52</td>
<td>0.287</td>
<td>0.556</td>
<td>0.70</td>
<td>0.156</td>
</tr>
<tr>
<td>Neutrophil count (10^6 /g)</td>
<td>0.52</td>
<td>0.306</td>
<td>0.580</td>
<td>0.70</td>
<td>0.193</td>
</tr>
<tr>
<td>Eosinophil count (10^3 /g)</td>
<td>0.35</td>
<td>0.549</td>
<td>0.773</td>
<td>0.62</td>
<td>0.390</td>
</tr>
<tr>
<td>Macrophage count (10^6 /g)</td>
<td>0.51</td>
<td>0.369</td>
<td>0.630</td>
<td>0.57</td>
<td>0.303</td>
</tr>
<tr>
<td>Myeloperoxidase (μg/g)</td>
<td>0.47</td>
<td>0.544</td>
<td>0.754</td>
<td>0.63</td>
<td>0.354</td>
</tr>
<tr>
<td>Interleukin-8 (ng/g)</td>
<td>0.34</td>
<td>0.676</td>
<td>0.781</td>
<td>0.46</td>
<td>0.384</td>
</tr>
<tr>
<td>Eosinophil Cationic Protein (ng/g)</td>
<td>0.52</td>
<td>0.552</td>
<td>0.804</td>
<td>0.74</td>
<td>0.323</td>
</tr>
<tr>
<td>Albumin (μg/g)</td>
<td>0.70</td>
<td>0.367</td>
<td>0.737</td>
<td>0.86</td>
<td>0.211</td>
</tr>
<tr>
<td>alpha-2-Macroglobulin (μg/g)</td>
<td>0.60</td>
<td>0.315</td>
<td>0.533</td>
<td>0.81</td>
<td>0.211</td>
</tr>
</tbody>
</table>

Data from sets with paired samples from 21 COPD patients obtained within one week; Ri: intraclass correlation coefficient; *:within-patients variability is the standard deviation of the absolute difference of the base-10 logarithmic transformed data; #:between-patients variability is the standard deviation of the log-transformed data from the first sample; $:correction for dilution by adding to the log-transformed data the value of “ % squamous cells x 0.0155 ”; ¶:gram sputum.
Discussion

Contamination of whole induced sputum samples with saliva, as reflected by the presence of squamous epithelial cells, leads to exclusion of sputum samples and biases the analyses of sputum biomarkers. To assess whether squamous cells can serve as a quantitative measure of contamination of sputum samples with saliva, we retrospectively analyzed data from 247 and 235 sputum samples from COPD and asthma patients respectively. We found inverse log-linear relationships for both non-squamous absolute cell numbers (TCC) and soluble inflammatory parameters with % squamous epithelial cells. These linear relationships facilitated correction of sputum samples for contamination which resulted in an enhanced repeatability and reduced variability of biomarkers, and enhanced the discriminative power of the sputum parameters. Our analyses further provide a rationale for designating sputum samples with ≤90% squamous epithelial cells as valid.

Saliva contains squamous epithelial cells and only low amounts of proteins and virtually no inflammatory cells\(^{18,19}\). In fact, the number of inflammatory cells was previously estimated to be 40-fold lower in saliva than in sputum, while ECP levels were 9-fold lower in saliva\(^{19}\). Thus, contamination of sputum with saliva predominantly dilutes sputum. Previously, albumin content has been taken as a measure of dilution of airway secretions\(^{20,21,25}\). Indeed, we found that lower values of sputum albumin were correlated with lower values of other inflammatory parameters in sputum. And further, when TCC was corrected for dilution using % squamous cells, the strength of the relationship of TCC with albumin content was largely reduced (Figure 5), indicating that increasing % squamous cells and decreasing albumin content can serve both as measures of dilution. Correction for albumin in sputum, however, is controversial since albumin permeation may differ with disease severity and treatment. Indeed, we found a tendency for albumin content to vary with study conditions (see ‘Subjects and data sets’ in Methods). Therefore, we propose that % squamous cells provides a better measure of dilution of whole induced sputum samples by saliva than albumin content.

In sputum samples with >90% squamous cells the variation of sputum parameters, in particular that of cellular parameters, was larger than in samples with ≤90% squamous cells. This indicates that, for samples with >90% squamous cells, parameters are not only affected by dilution but also by other factors, which enhance the variation. Previously a cut-off level of 80% squamous cells was proposed as the presence of large amounts of squamous cells physically obscure the counting of other cells and thus reduce the accuracy\(^{17}\). We report a higher, 90% cut-off level which may relate to our extended procedure for counting cells (see Methods). The validity of a cut-off value of 90% is reinforced by cell differentials that remain virtually unaffected in sputum samples with up to 90% squamous cells. Therefore we propose that, following our cell count protocol, samples with ≤90% squamous cells are valid for statistical analyses. By adopting 90% as cut-off level, the number of samples in our study deemed invalid was reduced markedly.
in comparison to the widely used cut-off level of 80% squamous cells: from 14.6 % to 8.9 % excluded samples for COPD and from 21.7 % to 11.5 % respectively for asthma. The dilution of an ‘average’ sputum sample by saliva is far smaller than that calculated for a sputum sample with 100% squamous cells (Table 2), but still is considerable: for a sample with median % squamous cells a 8-fold and 2.5-fold dilution in total cell counts was calculated for asthma and COPD, respectively, and a maximal 3-fold dilution for soluble markers both for asthma and COPD. Overall, the extent of dilution of sputum in an “average” sample at median % squamous cells was larger in asthma than in COPD samples even though maximal dilution, calculated to occur at 100% squamous cells, was larger within COPD than in asthma samples.

We applied this novel approach to re-evaluate two clinical studies from our institution. Sputum data from samples with ≤90% squamous cells were included and data were corrected for dilution. In the parent COPD study assessing repeatability of sputum parameters in two subsequent sputum samples from clinically stable patients 24, correction for dilution increased repeatability and decreased both within-patients variability and between-patients variability (Table 3). In the parent asthma study22, investigating the effects of two therapeutic interventions on allergen-induced inflammation, correction led to a reduced variability of parameters and improved the power of the study, revealing a tendency (p=0.06) for a difference between treatments, which was not found before (p=0.3). Taken together, we have shown that when the % squamous cells is taken as a measure of dilution by saliva such a correction enhances the discriminative power of sputum parameters.

Variation of sputum parameters depends on variable dilution, on biological variation over time in individual subjects and on differences due to the condition of the patient, such as (corticosteroid) treatment, stable disease, exacerbation or allergen provocation. To limit biological variation we analyzed data from multiple sputum samples from a limited number of patients rather than one sputum sample from many patients. With respect to the variation due to differences in the condition of the patient, retrospective analysis indicated that these different conditions did not underlie the observed relationship between sputum parameters and % squamous epithelial cells. Furthermore, we also showed that a similar correlation was found when only single samples for each patient were analyzed. Therefore, we propose that the observed relationships are reflecting dilution of sputum and are not biased by the inclusion of multiple samples from each patient.

Apart from dilution by saliva, saline that is being used to induce sputum expectoration may also contribute to dilution. Our data do not allow us to estimate the extent of dilution by saline. A previous study, however, showed that the chloride content of sputum samples obtained after induction by hypertonic saline was only slightly higher than in samples
obtained after induction by isotonic saline, indicating that only a small proportion of the sputum sample originated from the nebulized saline\(^1\). Moreover, dilution with saline would lead to decreased levels of inflammatory parameters with an unaltered percentage of squamous cells. Thus we propose that the inverse relationship between % squamous cells and inflammatory markers reflects dilution of induced whole sputum samples predominantly by saliva.

The log-linear regression coefficients and thus also the fold-dilution differed between sputum parameters (Table 2). These differences between sputum parameters may be explained by assuming that saliva contains different levels of sputum parameters. In addition, but not excluding the previous explanation, sputum parameters may display differential interactions with mucin structures, influencing the diffusion rate of components during transport of sputum in the airways. Diffusion of A2M, a large protein with a molecular mass of 725 kDa, is restricted by the tight mucin network, as a consequence of which the fold-decrease is relatively low. In fact, A2M levels were markedly higher in the gel phase as opposed to the soluble phase of spontaneous sputum in COPD\(^2\). Also the charge of the protein may affect its diffusion rate, like positively charged proteins interacting with negative (sulphur groups) charges on mucins. Interestingly, ECP, a cationic protein, shows a steeper decline with increasing % squamous cells in sputum samples from COPD patients than in that from asthma patients (Table 2). Non-squamous cell counts decreased markedly with increasing % squamous cells, which may indicate that although some cells may be entangled by sputum, the larger part may be associated with the surface of sputum plugs allowing their easy removal. As yet we can merely speculate about the exact process of dilution that underlies the log-linear relationships between sputum parameters and % squamous cells.

In conclusion, our data are compatible with % squamous cells being a marker of dilution of sputum for whole induced sputum samples with ≤ 90% squamous cells. Correction for dilution increases the repeatability and decreases the variation of sputum biomarkers, thereby enhancing the discriminative power of sputum parameters. It is important to realize that this correction has a profound effect also for sputum biomarkers of samples with a low to median % of squamous cells and not only for those samples with high percentages of squamous cells. By providing a rationale for determining sputum samples valid or invalid, the number of invalid samples will be reduced and because of reduced variability the effects of interventions may be more easily distinguished as was exemplified by our post-hoc analyses. In all adjustment for percentage squamous cells significantly improved validity of biomarkers in whole induced sputum in asthma and COPD, and may improve usage of biomarkers in phenotyping and monitoring patients with asthma and COPD.
Acknowledgements
We gratefully acknowledge Petra Teiwes for lung function tests and collection of sputum. Prof. Dirkje S. Postma is acknowledged for critically reading the manuscript. Part of the present data was presented at the European Respiratory Society Congress, September 2005\textsuperscript{28}.

Supplement
In the supplement are presented: Figures S1 to S5, showing data per patient (S1 and S2) as well as analyses per study condition (S3 and S4) and correlation between % squamous cells versus log of non-squamous cell count for first sputum sample of each patient (S5).
References


