Allergic asthma: Environmental factors challenging the immune system

van de Pol, M.A.

Citation for published version (APA):
van de Pol, M. A. (2013). Allergic asthma: Environmental factors challenging the immune system

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Improved validity of cellular and molecular biomarkers for induced sputum in asthma and COPD

Martin Boorsma*, Marianne A. van de Pol*, Julia G. Koopmans, Viviana Turchiarelli, Barbara S. Smids, Tamara Dekker, Henk M. Jansen, Theo A. Out, Peter J. Sterk, Jaring S. van der Zee, René E. Jonkers, René Lutter

* These authors contributed equally to this study.

Submitted to Clinica Chimica Acta
ABSTRACT
Background: Induced sputum biomarkers are important for phenotyping and monitoring of patients with asthma and COPD, but contamination with saliva enhances biomarker variability.
Methods: Squamous cell counts were tested as a measure of dilution by saliva for various biomarkers (total and differential cell count, eosinophil cationic protein, myeloperoxidase, interleukin-8 and alpha2-macroglobulin) in whole sputum samples from 29 COPD patients (247 samples) and 25 asthma patients (235 samples). Further, we determined repeatability of these biomarkers between sputum plugs that were removed from sputum to limit dilution with saliva (8 asthma patients).
Results: Whole sputum samples with ≤90% squamous cells, showed inverse log-linear relationships between absolute cell counts and levels of soluble parameters with percentage squamous cells. These log-linear relationships enabled correction of whole sputum data for saliva contamination, reducing variability and improving repeatability of sputum parameters. Interleukin-8, alpha2-macroglobulin and % neutrophils were heterogeneously distributed between selected plugs. Soluble biomarkers were found to interact dynamically with sputum plug matrix, leading to rapid equilibration with the surrounding fluid phase.
Conclusions: Analyzing multiple plugs and limiting exposure of plugs to fluid reduces biomarker variation in the selected plug method. Analysis of whole sputum with ≤90% squamous cells is improved by correction for the percentage squamous cells. These improvements will further advance the implementation of cellular and molecular monitoring in airway disease.

INTRODUCTION
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 in a range of airway diseases\(^1\)-\(^5\). Standardization of methods for sputum induction, collection and analyses has consolidated and extended its use in experimental studies and to a major extent in clinical trials\(^6\)-\(^10\). The analysis of sputum, however, is frustrated by the contamination with secretions from the upper airways and oral cavity, such as saliva. In studies comparing cell numbers and inflammatory parameters in paired sputum and saliva samples, saliva was found to contain low numbers of inflammatory and immune cells (predominantly squamous epithelial cells), a low protein content and low amounts of soluble inflammatory markers\(^11\),\(^12\). Contamination of sputum with saliva thus primarily leads to dilution, but to which extent is unknown.
Two distinct methods have been advocated to deal with this variable dilution of sputum. In the selected plug method, sputum plugs are selected and physically removed from the sputum and processed. In the whole sputum method, all of the sputum sample is processed in which the relative presence of squamous epithelial cells is considered an indicator of contamination with saliva, labeling the sputum valid or not. Both the selected plugs and whole sputum method are being used, although there appears to be a preference for the selected plug method.

Having performed multiple studies using both methods, we have re-evaluated and performed additional studies aimed at further reducing the variability for sputum analyses. For the whole sputum analysis we assessed whether the presence of squamous cells indeed can serve as a quantitative marker of dilution for soluble and cellular parameters. For the selected plug method we determined the variation of soluble and cellular inflammatory parameters between plugs within a sputum sample and also whether soluble parameters in plugs are affected by surrounding fluid, such as occurs during processing the sample. Implementation of our findings should lead to attenuation of the variation of sputum parameters and thus enhance the discriminative power of sputum parameters in experimental and clinical studies.

METHODS
To determine whether squamous cells can serve as a quantitative marker of dilution for soluble and cellular parameters in whole sputum we analyzed data from 482 sputum samples. To assess the variation of soluble and cellular inflammatory parameters in selected plugs we analyzed plugs from patients with mild asthma.

Whole sputum method: Subjects and data sets
Data sets from two prospective clinical studies, one with asthma and one with chronic obstructive pulmonary disease (COPD) patients, were analyzed. COPD patients participated in a 16-months study and sputum 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 ≥ 15 pack-years, a forced expiratory volume in one second (FEV1)/forced vital capacity (FVC) ratio ≤ 0.70 and FEV1 reversibility ≤ 11% of predicted.

Asthma patients participated in a 6-weeks study and sputum samples (maximally 10) were obtained after treatment with an inhaled corticosteroid and a long-acting beta-agonist or placebo, before and after allergen provocation, also referred to in the text as ‘different study conditions’. Patients were between 18 and 50 years, never- or ex-smokers...
with ≤ 10 pack-years, with a FEV$_1$ ≥ 80% of predicted and a provocative concentration causing FEV$_1$ to fall 20% (PC$_{20}$) histamine ≤ 8 mg/ml. Both studies were approved by the medical ethics committee of the Academic Medical Center; written informed consent was obtained. The induction of sputum was identical to the procedure for the selected plugs described below, and the processing has been described earlier$^{21,22}$.

**Selected plug method: Subjects**

Asthma patients (n=8) were selected among outpatients visiting the Academic Medical Centre, Amsterdam. The diagnosis of asthma was based on the Global Initiative for Asthma (GINA) Guidelines$^{23}$ and asthma severity was defined based on the 2007 consensus$^{24}$. Patients had episodic chest symptoms, pre-bronchodilator FEV$_1$ >75% predicted, documented reversibility in FEV$_1$ >200 ml by 400 µg inhaled salbutamol and/or airway hyperresponsiveness (PC$_{20}$ methacholine <4 mg/mL) during the past 12 months. Patients were non-smokers and were treated with low doses of inhaled corticosteroids (< 1000 µg/day inhaled beclomethasone equivalent) and/or long-acting bronchodilators. The study was approved by the medical ethics committee of the Academic Medical Center and all patients gave their written informed consent.

**Selected plug method: Induction and processing of sputum for selected plugs**

Sputum induction was performed according to ERS recommendations$^{25}$. In short, patients received pre-treatment with 400 µg inhaled salbutamol. FEV$_1$ was measured by standardized spirometry before three episodes of 5 min inhalation of aerosolized 4.5% hypertonic saline solution generated by an ultrasonic nebulizer (KLAVAmed, Bielefeld, Germany). Prior to sputum production, subjects were asked to rinse their mouth with water, swallow, and blow their nose to minimize contamination of sputum with saliva and postnasal drip fluid. Patients were encouraged to cough and expectorate sputum. Sputum was collected in a sterile container and transported on ice to the lab and processed immediately.

Sputum plugs were selected from sputum dispersed in a petri-dish on a black background using a fine forceps, transferred to the lid of the petri-dish where plugs were moved with small circular motions to spread saliva across the lid and keep the plugs in one mass. To determine variability of inflammatory parameters between plugs from one sputum sample multiple (between 3-6) plugs were distributed over 3 to 4 tubes with 100-400 mg of plugs per tube. Per tube, 4 volume-equivalents (with 100 mg of sputum = 100 µl volume) of 0.1% of freshly prepared sputolysin (Calbiochem) were added, vortexed and incubated on a roller bank at 4°C for 15 min. Cells were pelleted (10 minutes, 4°C, 470 x g), and supernatant was stored in aliquots at -80°C till analyses. Cell pellets were re-
suspended in 1 ml in phosphate-buffered saline (PBS). The total cell number was determined by counting manually in a Bürker counting chamber. To determine whether soluble biomarkers associated with sputum plugs were exchanged with a fluid phase, plugs were treated with 8 volume-equivalents of Dulbecco’s PBS (D-PBS) on a roller bank at 4°C for 15 min. after which plugs were collected by centrifugation (10 min, 470 x g, at 4°C) and supernatant was aliquoted for further analyses. The pelleted plugs were treated with sputolysin as indicated above.

Cells were cytocentrifuged for 2 min at 550 rpm in Shandon Cytocentrifuge and stained with Romanovsky (Diff-Quick) and May-Grunwald-Giemsa (MGG). Differential cell counts were based on 500 non-squamous cells. Sputum samples containing > 80% non-squamous cells on differential cell counted were excluded for differential analysis. Squamous and non-squamous epithelial cells, macrophages, lymphocytes, neutrophils and eosinophils were identified. Differential cell counts were expressed as the percentage of non-squamous cells. Absolute cell numbers were calculated as (% cell x total cell count)/sputum weight. Sputum cell counts were performed by one experienced and qualified technician blinded to the details. As an extra control 10% of the samples were analyzed by a second technician.

Levels of eosinophil cationic protein (ECP)²⁶, interleukin (IL)-8²⁷, myeloperoxidase (MPO)²⁷ and alpha2-macroglobulin (A2M)²⁸ were measured in sputum with enzyme-linked immuno-sorbent assays (ELISAs).

Statistical analyses
Whole induced sputa: 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 whole sputum samples with 100% squamous cells were arbitrarily assigned 0.01x10⁶/g. Differences in the presence of squamous cells and albumin content under different study conditions (for explanation see ‘Subjects and data sets: For the whole sputum method’) 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 $10^{b \cdot x}$. 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 dilution on the basis of % squamous cells were tested by t-test.

Selected plugs: To express the variation of soluble and cellular sputum parameters between 2 or 3 tubes with selected plugs for 8 distinct sputa, we determined the repeatability (intraclass correlation coefficient (ICC); two-way mixed, consistency) using SPSS 20. An ICC of ≥ 0.80 was taken to indicate good repeatability.

RESULTS
Whole sputum analyses
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 supplemental 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-discriminable, 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 figs. 1 and 2). 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 figs. 3 and 4).

**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 \( (r = -0.82, p < 0.001; \text{fig. 1A}) \) and asthma samples \( (r = -0.85, p < 0.001, \text{fig. 1B}) \). The regression lines for log-transformed TCC versus % squamous cells showed regression coefficients of \(-0.020 (95\% \text{CI: } -0.022, -0.018) \) for COPD samples and \(-0.017 (95\% \text{CI: } -0.018, -0.016) \) for asthma samples (see Methods). These regression coefficients allowed us to calculate the theoretical TCC at 0% squamous cells (i.e. no contamination with saliva), and the extent of dilution of the samples (see Methods). For TCC we calculated a 100-fold dilution (95% C.I. 63 – 126-fold) in samples with 100% squamous cells relative to samples with 0% squamous cells for COPD samples and a 50-fold dilution (95% C.I. 40 – 63-fold) for asthma samples (supplemental table 2). For an ‘average’ sample, i.e. a sample at the median value of % squamous cells, there was an 8-fold dilution of TCC in asthma samples and a 2.5-fold dilution of TCC in COPD samples compared to samples with 0% squamous cells.

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 \( (r = -0.70 \text{ for COPD and } r = -0.90 \text{ for asthma, both } p < 0.001, \text{see supplemental fig. 5}) \). 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 (fig. 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 \( (r = -0.25 \text{ and } -0.34, \text{for COPD and asthma samples respectively, both } p < 0.001, \text{see supplemental fig. 6}) \). However, this was
Figure 1. Non-squamous Total Cell Count \(\left(10^7/g\right)\) 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 cationic protein (ECP) levels (\(\mu g/g\) 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).
mainly due to samples with >90% squamous cells containing no eosinophils at all. Excluding these samples, % eosinophils marginally, but still significantly 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 myeloperoxidase (MPO), interleukin-8 (IL-8), eosinophil cationic protein (ECP), α-2 macroglobulin (A2M) and albumin (Alb) showed linear decreases with increasing % squamous cells. Data for ECP and MPO are shown in figure 2 and supplemental figure 7 and are representative of other markers. Fold-dilution for soluble parameters at 100% squamous cells relative to 0% squamous cells is shown in supplemental 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.

**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, as was found. (fig. 3, 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).

**Post-hoc analyses**

To determine the effect of correcting sputum data for contamination with saliva we reanalyzed 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
Figure 3. 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, μg/g sputum, log-transformed, middle), and myeloperoxidase (MPO, μg/g sputum, log-transformed, right) with albumin level (μg/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).
improved repeatability (higher intraclass correlation coefficients), and in smaller within-patients and between-patients variability (all \( p < 0.01 \)) (supplemental table 3).

A second post-hoc analysis was done on data from an intervention study on an allergen challenge in allergic asthma patients\(^{21}\). 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\(^{21}\). 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.

**Selected plug analyses**

For each of 8 patients with asthma, multiple sputum plugs were obtained from the sputum sample and distributed over 2 or 3 tubes (100-400 mg of plugs per tube) and subsequently analyzed for soluble and cellular parameters. The repeatability of biomarker data between plugs from the same sample is shown in table 1. In parallel, selected plugs in 1 to 3 tubes were exposed for 15 minutes to isotonic protein-free PBS, after which it was determined whether soluble components from the plugs equilibrated with PBS. Forty to 60\% of the soluble parameters that were originally present in the plugs ended up in PBS (table 2). This indicates that the interaction of soluble parameters with the sputum plug is dynamic.

**DISCUSSION**

Contamination of sputum samples with unknown amounts of saliva increases the variation of sputum biomarkers and thus reduces the discriminative power of these biomarkers in e.g. clinical trials. Sputum processed according to the whole sputum method and contaminated with \( \geq 80\% \) squamous cells are considered to be contaminated markedly with saliva and therefore taken as invalid. We found inverse log-linear relationships for both non-squamous absolute cell numbers (TCC) and soluble inflammatory parameters with \% squamous epithelial cells, indicating that the percentage of squamous cells can serve as a quantitative measure for contamination of sputum samples with saliva. Furthermore, we found that sputum samples with \( \leq 90\% \) squamous epithelial cells are still valid for analyses, increasing the number of valid samples. For the selected sputum plug method we showed that biomarkers are distributed heterogeneously between plugs and thus analyzing few plugs will add to the variability. Further, soluble biomarkers were found to interact dynamically with sputum plug constituents, leading to leakage of biomarkers.
Table 1. Repeatability of soluble and cellular inflammatory parameters between selected sputum plugs from asthmatics.

<table>
<thead>
<tr>
<th>parameter</th>
<th>ICC</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL-8</td>
<td>0.37</td>
<td>8</td>
</tr>
<tr>
<td>MPO</td>
<td>0.78</td>
<td>6</td>
</tr>
<tr>
<td>ECP</td>
<td>0.95</td>
<td>8</td>
</tr>
<tr>
<td>A2M</td>
<td>0.22</td>
<td>8</td>
</tr>
<tr>
<td>% neutrophils</td>
<td>0.31</td>
<td>5</td>
</tr>
</tbody>
</table>

ICC, intraclass correlation coefficient; n, number of patients from whom plugs were analyzed. Sputum plugs from 8 patients were analyzed, but for two patients MPO could not be determined reliably. For three patients % neutrophils were not taken along; for two patients % squamous cells was ≥ 90%, and for one patient the % neutrophils was ≥ 90%.

Table 2. Relative amounts (%) of soluble parameters associated with selected plugs that end up in the surrounding medium.

<table>
<thead>
<tr>
<th>IL-8</th>
<th>MPO</th>
<th>ECP</th>
<th>A2M</th>
<th>mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>50.2</td>
<td>50.2</td>
<td>45.7</td>
<td>42.0</td>
<td>47.7</td>
<td></td>
</tr>
<tr>
<td>8.1</td>
<td>8.1</td>
<td>24.8</td>
<td>8.9</td>
<td>18.8</td>
<td></td>
</tr>
</tbody>
</table>

into the surrounding fluid phase. Thus, plugs should be selected and separated from a fluid phase as soon as possible after sampling.

For the whole sputum method we carefully analyzed data from a large number of sputum samples from COPD and asthma patients. Variation of sputum parameters depends on variable dilution by saliva, on biological variation over time in individual subjects and on differences related to the condition of the patient, such as (corticosteroid) treatment, stable disease and exacerbation. Another important source of variation are inter-individual differences. To limit variation by inter-individual differences 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 of biomarkers with the % squamous epithelial cells are genuine and are not biased by inclusion of multiple samples from each patient.
Apart from dilution by saliva, saline used to induce sputum expectoration may also contribute to dilution. Our data do not allow an estimation of the extent of dilution by saline. Previously, however, it was shown that 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 fluid phase of the sputum sample originated from the nebulized saline. Moreover, dilution with saline would lead to decreased levels of inflammatory parameters with an unaltered % 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.

Previously albumin content has been taken as a measure of dilution of airway secretions. 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 positive relationship of TCC with albumin content was largely reduced (fig. 3), 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 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 variation. Previously a cut-off level of 80% squamous cells was proposed as the presence of large amounts of squamous cells physically obscure other cells and thus reduce the accuracy. 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 (supplementary 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.

The log-linear regression coefficients and thus also the fold-dilution differed between sputum parameters (supplementary 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\textsuperscript{27,28}. Also the charge of the protein may affect its diffusion rate, like positively charged proteins interacting with negative (sulphur groups) charges on mucins. Interestingly, cationic ECP shows a steeper decline with increasing % squamous cells in sputum samples from COPD patients than in that from asthma patients (supplementary table 2). Non-squamous cell counts decreased markedly with increasing % squamous cells, which may indicate that the larger part of squamous cells may be associated with the surface of sputum plugs allowing their easy removal.

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\textsuperscript{29}, correction for dilution increased repeatability and decreased both within-patients variability and between-patients variability (supplementary table 3). In the parent asthma study\textsuperscript{21}, 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.
An important finding in relation to the selected sputum plug method was the rapid exchange of soluble biomarkers in the sputum plugs with the surrounding fluid. In our experimental set up we took several precautions to not promote this exchange. Plugs were incubated with an isotonic buffer to not disturb ionic interactions between soluble biomarkers and the mucin network in the plugs. The incubation was performed at 4°C, which prevents degranulation of associated cells. And finally, the plugs were incubated for 15 min only, which is relatively short compared to the 1 to 2 hours that is usually allowed between sampling and processing of sputum. Despite these conditions, 40 to 60% of the plug-associated soluble biomarkers ended up in the fluid phase. In part this marked exchange may be due to the 8 volume-equivalents used to incubate the plugs with. Nevertheless, this finding reflects that soluble biomarkers interact dynamically with the matrix of the plugs. Taken together these findings imply that selected sputum plugs should be separated as soon as possible after sampling. To limit exchange, additional techniques such as washing of the plugs should be kept to a minimum.

We also found that biomarkers were distributed heterogeneously between the selected plugs. Values for MPO and ECP were relatively consistent between the various plugs, but the variation between the plugs for A2M, IL-8 and % neutrophils was large. This variation is not explained by variation of the employed quantitative assays (ELISAs etc.). The samples were measured in parallel and the intra-assay variation for the assays was below 10%. This heterogeneity underlines the need to analyze multiple if not all plugs for the analyses. We used between 100-400 mg of plugs that revealed this heterogeneity, indicating that more plugs are required to reduce the variation.

In conclusion, we have identified means to reduce variation of sputum biomarkers for both the selected plug and the whole sputum method. For the selected plug method, analyses of a substantial number of plugs and minimizing the interaction with surrounding fluid are crucial to limit variation. For the whole sputum method, the % of squamous cells can be taken as a quantitative measure of contamination with saliva, allowing to correct for dilution. In addition, samples with up to 90% squamous cells can be analyzed, which can provide a significant increment in the number of valid sputa.

Acknowledgements
We gratefully acknowledge Petra Teiwes for lung function tests and collection of sputum. The studies on the selected plug method were facilitated by a grant from ‘Stichting Astmabestrijding’ (2008/003). For the whole sputum method, the parent studies received unrestricted grants from AstraZeneca (the COPD study, BN-00P-0087) and from GlaxoSmithKline (the asthma study, SER98010).
REFERENCES
**SUPPLEMENT**

**Table S1.** 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>MPO (μg/g sputum)</td>
<td>9.4 (0.4 – 188)</td>
<td>0.74 (0.16 – 10.5)</td>
</tr>
<tr>
<td>IL-8 (ng/g sputum)</td>
<td>5.0 (0.4 – 174)</td>
<td>0.25 (0.10 – 1.73)</td>
</tr>
<tr>
<td>ECP (μg/g sputum)</td>
<td>322 (10 – 38260)</td>
<td>57.0 (4.8 – 322)</td>
</tr>
<tr>
<td>Alb (μg/g sputum)</td>
<td>59.2 (1.6 – 2379)</td>
<td>30.6 (0.89 – 97.4)</td>
</tr>
<tr>
<td>A2M (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; MPO, myeloperoxidase; ECP, eosinophil cationic protein; IL-8, interleukin 8; Alb, albumin; A2M, alpha-2-macroglobulin.

**Table S2.** Calculated maximal dilution of inflammatory parameters in sputum samples with 100% squamous cells.

<table>
<thead>
<tr>
<th></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</td>
<td>10 (6.3 – 20)</td>
<td>3.2 (2.0 – 6.3)</td>
</tr>
<tr>
<td>alpha-2-macroglobulin</td>
<td>4.0 (2.5 – 7.9)</td>
<td>2.5 (1.6 – 5.0)</td>
</tr>
<tr>
<td>Myeloperoxidase</td>
<td>20 (13 – 40)</td>
<td>2.0 (1.6 – 3.2)</td>
</tr>
<tr>
<td>Interleukin-8</td>
<td>25 (10 – 50)</td>
<td>N/A</td>
</tr>
<tr>
<td>Eosinophil Cationic Protein</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.
### Table S3. Repeatability of sputum data without and with correction for dilution using % squamous cells.

<table>
<thead>
<tr>
<th></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-sq. TCC (10^6/g &lt;br&gt;¶)</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 (10^6/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>MPO (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>IL-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>ECP (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>Alb (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>A2M (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 of sputum. Non-sq.TCC, non-squamous total cell count; MPO, myeloperoxidase; IL-8, interleukin 8; ECP, eosinophil cationic protein; Alb, albumin; A2M, alpha-2-macroglobulin.
Figure S1. Non-squamous Total Cell Count ($10^6$/g sputum, log-transformed) expressed as a function of % squamous cells in sputum samples in individual patients with COPD who had at least two samples.
Figure S2. Non-squamous Total Cell Count ($10^6/g$ sputum, log-transformed) expressed as a function of % squamous cells in sputum samples in individual patients with asthma who had at least two samples.
Figure S2 (continued). Non-squamous Total Cell Count (10⁶/g sputum, log-transformed) expressed as a function of % squamous cells in sputum samples in individual patients with asthma who had at least two samples.
Figure S3. Squamous Cell count (%, Boxplots) in sputum samples from COPD patients collected under different study conditions. At entry ($n = 31$), after Run-in ($n = 67$), after prednisolone ($n = 42$), after inhaled budesonide ($n = 49$), after inhaled placebo ($n = 43$) or immediate before or after an exacerbation ($n = 15$). Differences between conditions not significant ($p = 0.48$).

Figure S4. Squamous Cell count (%, Boxplots) in sputum samples from asthma patients collected under different study conditions. At entry ($n = 25$), after Run-in ($n = 63$), 6 or 24 hours after allergen challenge ($n = 99$) or 1 week after allergen challenge ($n = 47$). Differences between conditions not significant ($p = 0.38$).
Whole sputum and selected plugs

Figure S5. Non-squamous Total Cell Count ($10^6$/g sputum, log-transformed) expressed as a function of % squamous cells in sputum samples from 29 patients with COPD (triangles) and from 25 patients with asthma (circles), obtained at enrolment in the study. Pearson’s $r = 0.79$, $p < 0.001$. 
Figure S6. 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%.

Figure S7. 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).