



OPEN

Optimising the yield from bronchoalveolar lavage on human participants in infectious disease immunology research

Jane Alexandra Shaw^{1,2,3}✉, Maynard Meiring^{1,2,3,4}, Devon Allies^{1,2,3}, Lauren Cruywagen^{1,2,3}, Tarryn-Lee Fisher^{1,2,3}, Kesheera Kasavan^{1,2,3}, Kelly Roos^{1,2,3}, Stefan Marc Botha^{1,2,3}, Candice MacDonald^{1,2,3}, Andriëtte M. Hiemstra^{1,2,3}, Donald Simon^{1,2,3}, Ilana van Rensburg^{1,2,3}, Marika Flinn^{1,2,3}, Ayanda Shabangu^{1,2,3}, Helena Kuivaniemi^{1,2,3}, Gerard Tromp^{1,2,3,4,5}, Stephanus T. Malherbe^{1,2,3}, Gerhard Walz^{1,2,3}, Nelita du Plessis^{1,2,3} & The Stellenbosch University Immunology Research Group Bronchoalveolar Lavage Study Consortium (SU IRG BAL Study Consortium)*

Bronchoalveolar lavage (BAL) is becoming a common procedure for research into infectious disease immunology. Little is known about the clinical factors which influence the main outcomes of the procedure. In research participants who underwent BAL according to guidelines, the BAL volume yield, and cell yield, concentration, viability, pellet colour and differential count were analysed for association with important participant characteristics such as active tuberculosis (TB) disease, TB exposure, HIV infection and recent SARS-CoV-2 infection. In 337 participants, BAL volume and BAL cell count were correlated in those with active TB disease, and current smokers. The right middle lobe yielded the highest volume. BAL cell and volume yields were lower in older participants, who also had more neutrophils. Current smokers yielded lower volumes and higher numbers of all cell types, and usually had a black pellet. Active TB disease was associated with higher cell yields, but this declined at the end of treatment. HIV infection was associated with more bloody pellets, and recent SARS-CoV-2 infection with a higher proportion of lymphocytes. These results allow researchers to optimise their participant and end assay selection for projects involving lung immune cells.

Bronchoscopy with bronchoalveolar lavage (BAL) is a commonly performed clinical procedure that yields fluid containing cells and other substances that populate the lung mucosa and alveolar surface, assisting in diagnosis and treatment of many lung diseases. A flexible video bronchoscope introduced into a participant's airway, advanced until gently wedged in a target lung lobe or segment, allows warmed sterile saline to be instilled and then aspirated through the working channel of the bronchoscope (Fig. 1). The procedure is not without risk, as often the patient needs intravenous (IV) sedative agents, there may be mechanical trauma from the scope or endobronchial procedures which can cause bleeding or pneumothorax, or residual fluid which cannot be aspirated may, rarely, cause problems such as infection¹. These risks can be mitigated and therefore BAL for research alone, without any envisioned clinical benefit to the participant, is becoming a common procedure. The fluid is used for both qualitative and quantitative measurements to study structural and immune cells, soluble mediators, mucus, and features of tissue remodelling, at the site of disease². Researchers have an ethical duty to ensure that the procedure is as safe as possible for the volunteer, and that the BAL specimen retrieved is of sufficient quantity and quality to be of scientific value. The European Respiratory Society (ERS) in 1999 and the American Thoracic Society (ATS) in 2012 published guidance documents (the latter specifically for patients with suspected interstitial

¹DSI-NRF Centre of Excellence for Biomedical Tuberculosis Research, Stellenbosch University, Cape Town, South Africa. ²South African Medical Research Council Centre for Tuberculosis Research, Stellenbosch University, Cape Town, South Africa. ³Division of Molecular Biology and Human Genetics, Biomedical Research Institute, Faculty of Medicine and Health Sciences, Stellenbosch University, PO Box 241, Cape Town 8000, South Africa. ⁴South African Tuberculosis Bioinformatics Initiative, Stellenbosch University, Cape Town, South Africa. ⁵Centre for Bioinformatics and Computational Biology, Stellenbosch University, Stellenbosch, South Africa. *A list of authors and their affiliations appears at the end of the paper. ✉email: janeshaw@sun.ac.za

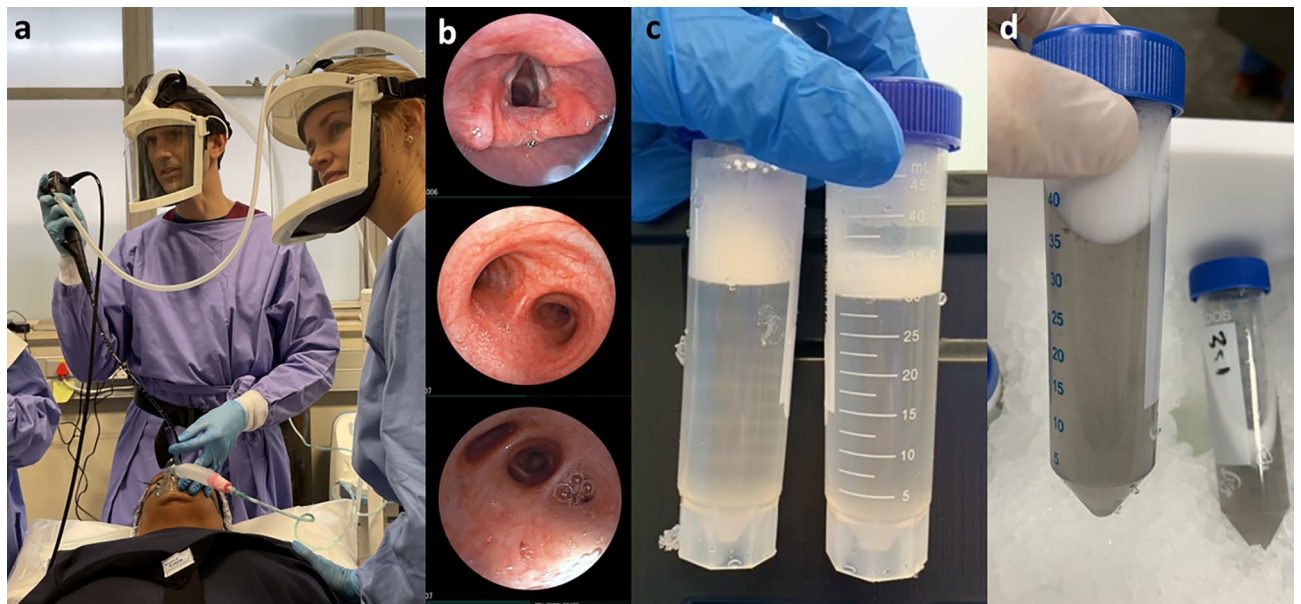


Figure 1. Bronchoscopy and bronchoalveolar lavage. (A) A research participant is about to have a Fujifilm flexible videoscope with 3.2 mm working channel inserted through her nose while under conscious sedation. Clinicians performing the procedure are authors STM and JAS. (B) Sequential photos from within the airways (from top to bottom: vocal cords before intubation, main carina, and right middle lobe medial segment subsegmental bronchi) taken with the bronchoscope. (C) Results of the bronchoalveolar lavage in a non-smoker, aliquots of saline with foamy surfactant on the surface. (D) An example of significant staining of the lavage fluid with black particulate matter from tobacco smoking.

lung disease), in an attempt to standardise BAL procedures, processing and analysis^{3,4}. Research studies may, however, require deviations from these recommendations, and many other factors can influence the outcomes of a BAL procedure beyond these parameters. For example, underlying chronic obstructive pulmonary disease (COPD) is associated with lower BAL volume yield than that from controls, and the yield is worse with declining lung function and disease severity on imaging^{5,6}; tobacco and cannabis smoking both have been shown to increase BAL cell yields^{7,8}; and a manual aspiration technique improves volume and cell yield over wall suction use⁹. Moreover, there are likely to be many factors which influence outcomes that are yet unknown to us. At the time of the bronchoscopy, the only measures of success or failure the physician (operator) has, is the wellbeing of their patient and the BAL volume yield, or percentage of instilled volume returned. It is only after processing of the specimen that the feasibility of using the BAL for specific research assays can be established, as determined by the cell quantity, quality, viability and differential count.

The aim of this study was to optimise bronchoscopy with BAL for immunology research, by identifying factors which affected the main BAL outcomes when procedures are performed according to current recommendations: BAL volume yield, total BAL cell yield and cell concentration, BAL cell viability, pellet colour (to identify blood contamination) and differential cell counts. In addition, procedure-related adverse events during the bronchoscopy were investigated. Special attention was given to investigating the effects (if any) of infections on BAL outcomes, specifically tuberculosis (TB), HIV infection and recent SARS-CoV-2 infection.

Methods

Study design. Participants in this study were recruited from five separate National Institutes of Health (NIH)-funded projects in which BAL was performed between 1 January 2019 and 15 March 2022, for research into host immune responses to *Mycobacterium tuberculosis* (Mtb) by the Biomedical Research Institute (BMRI) Clinical Team and the Immunology Research Group (IRG) laboratory of Stellenbosch University, Cape Town, South Africa. These studies included participants recruited through contact with BMRI community-based research workers from subdistricts in the Cape Town metropolitan region, as well as from the Drakenstein and Saldanha municipal subdistricts of the Western Cape. Data on clinical parameters were collected by research clinicians and nurses as part of the studies in which the participants were enrolled, while the laboratory processing data were collected by the research assistants on the day of processing. All participants provided written informed consent for the parent study into which they were enrolled. Each parent study was approved by the Stellenbosch University Health Research Ethics Committee (reference numbers: N16/03/033, N18/10/118, N19/07/093, N19/10/150 and N20/07/079). Research was performed in accordance with the Declaration of Helsinki. The participant and authors in Fig. 1 have provided informed consent to publish the image in an online open access publication.

In general, participants were included if they met the following criteria: (1) age 18–65 years; (2) weight 35–120 kg; (3) absence of current source of infection apart from those required for specific groups; (4) consented to bronchoscopy with BAL, Positron Emission Tomography-Computed Tomography (PET-CT) scan where

applicable, and nasopharyngeal swab for SARS-CoV-2 reverse transcription quantitative real time-Polymerase Chain Reaction (RT-qPCR); and (5) had a verifiable address or residence location for duration of the study. Exclusion criteria included: (1) pregnancy or breastfeeding; (2) haemoglobin < 9.0 g/dl; (3) any severe systemic condition/co-morbidity that may affect the safety of the participant or the performance of the assays; (4) diabetes (point of care HbA1c \geq 6.5%, random glucose \geq 200 mg/dl (or 11.1 mmol/L), fasting plasma glucose \geq 126 mg/dl (or 7.0 mmol/L), or the concomitant use of any anti-diabetic agent); (5) immunosuppressive medication use, including inhaled corticosteroids, within the past 2 weeks; (6) substance or alcohol abuse that in the opinion of the investigator may interfere with the participant's adherence to study procedures; or any person that the physician/study nurse feels is not appropriate for the study, or that the study is not in the person's best interest; (7) known allergy or intolerance to drugs used in procedural sedation or a known history or family history of malignant hyperthermia; or (8) history of taking antibiotics with anti-TB activity in the past 4 weeks (non-TB groups).

Participants were divided into the following groups according to history, clinical assessment, imaging and laboratory investigations:

1. TB groups: categorised according to treatment duration into 'Pre-Treatment' (treatment not yet initiated), 'Early Treatment' (less than one month) and 'End of Treatment' (completed 6 months of standard treatment). Participants with drug resistant TB or extrapulmonary TB (without predominant intrathoracic TB component) were excluded.
2. Household Contacts: close contact with an individual with active untreated TB within three months; active TB disease ruled out on sputum testing, chest radiograph and symptom screen. Contacts were excluded if they had previous TB disease in the last three years, or current TB symptoms.
3. Community Controls: no active TB disease in previous three years, no recent known exposure to TB, active TB ruled out on sputum testing, chest radiograph and symptom screen.
4. People living with HIV (PLHIV): known or newly diagnosed HIV infected individuals, active TB disease ruled out on sputum testing, chest radiograph and symptom screen.

Research procedures. *Mtb detection.* All participants had sputum testing for Mtb with Auramine smear, liquid culture [Mycobacterial Growth Indicator Tubes (MGIT), BACTEC 960, Becton Dickinson], and GeneXpert MTB/RIF Ultra[®] assay (Cepheid, Sunnyvale, CA, USA), hereafter 'Xpert Ultra'. A diagnosis of TB was assigned if a symptomatic individual had a positive MGIT culture or Xpert Ultra (graded as medium and higher), or positive Xpert Ultra graded as lower than medium but with consistent chest radiograph features.

All participants except those with active TB disease were also tested with an interferon gamma release assay (IGRA), in this case QuantiFERON-TB Gold Plus[®] (QFT) assay (Qiagen, Carnegie, Australia).

SARS-CoV-2 detection. The SARS-CoV-2 positive subgroup was defined as a previous verifiable positive SARS-CoV-2 RT-qPCR in a local reference laboratory, or positive serology on the Abbott SARS-CoV-2 IgG assay (Abbott Laboratories, South Africa, Pty Ltd) or the Abbott SARS-CoV-2 IgG II Quant (Abbott Laboratories, South Africa, Pty Ltd), both done on the Architect i System (Abbott Laboratories, Illinois, USA). All SARS-CoV-2 tests were carried out by the National Health Laboratory Service (NHLS) of South Africa (<https://www.nhls.ac.za/>). The strategy interpreting the serology results is described in Supplementary Table S1 in accordance with study protocol¹⁰. Participants who were previously vaccinated against SARS-CoV-2 who only had a positive anti-S antibody on serologic testing, and no previous positive RT-qPCR (and therefore no evidence of infection-acquired immunity or previous infection), were not included in the SARS-CoV-2 positive subgroup.

Bronchoscopy with bronchoalveolar lavage (BAL). Procedures were performed in accordance with current recommendations^{3,4}. In brief, all participants were pre-screened for fitness for bronchoscopy according to predefined criteria by a study clinician with knowledge of the procedure. A dedicated sedation practitioner achieved conscious sedation using either a propofol/fentanyl or midazolam/fentanyl combination. The participant received pre-oxygenation and continuous supplemental oxygen throughout the procedure, with continuous cardiac, non-invasive blood pressure and peripheral saturation monitoring until full recovery. Topical 1% lignocaine spray was applied to the throat and an anaesthetic gel in the nose. Topical lignocaine was applied to the vocal cords and large airways through the working channel of the flexible video bronchoscope which was then gently wedged in the target segment. Sequential aliquots of 60 ml were individually instilled then manually aspirated with a 20 ml syringe after ten seconds 'dwell time', to a total of 240 ml. The target lobe was chosen by the team prior to the procedure, based on the protocol of the parent study. In general, the right middle lobe (RML) was targeted in Community Controls and Household Contacts with normal chest imaging (with lingula as backup), and the area of most involvement with pulmonary TB or an imaging abnormality in keeping with subclinical TB in a Household Contact was targeted. Two of the studies included bronchial brushings (a single use brush with fine silicone bristles was used to harvest cells and microbiological material by lightly abrading the bronchial mucosa when inserted through the bronchoscope working channel) as part of their procedure, done either before or after the BAL depending on the standard operating procedure at the time. A thoracic ultrasound was performed in the recovery room on participants who underwent brushings, to exclude complications.

Adverse events. Data on adverse events were collected from an additional 20 study participants, resulting in 357 BAL procedures being included in these analyses. Information on adverse events was collected during the procedure and in recovery by clinicians using standardized patient record sheets. Participants were also contacted by study staff at 72 h and 14 days after the procedure to gather information regarding adverse events and persistent

symptoms using a standardised list of questions, captured directly onto the electronic database. If needed a follow up review was arranged.

Processing of BAL fluid. The BAL fluid was placed on ice immediately after aspiration. Processing was initiated within two hours of harvest. Steps differed between studies, but in general the processing steps included: pooling of all aliquots, two sterile filtration steps (Falcon Cell Strainers 70 μm and 40 μm , Corning Inc., NY), then centrifugation for ten minutes at $350 \times g$ at $+8^\circ\text{C}$ to pellet the cells. Two washing steps were performed using supplemented Roswell Park Memorial Institute Medium (RPMI 1640; Biowest, Nuaille, France), that is 46.5 ml RPMI 1640 with 500 μl of 1% L-glutamine (MilliporeSigma, Massachusetts, USA), 2.5 ml of 5% Human AB serum (MilliporeSigma, Massachusetts, USA), and 500 μl of a 1% Penicillin/Streptomycin/Amphotericin B mixture [100 IU/ml Penicillin, 100 $\mu\text{g}/\text{ml}$ Streptomycin and 0.25 $\mu\text{g}/\text{ml}$ Amphotericin B in 0.85% saline (MilliporeSigma, Massachusetts, USA)]. If the pellet was judged contaminated with blood by visual inspection, an additional red cell lysis step was performed in 1 ml of Lonza ACK lysis buffer (1x) [Whitehead Scientific (Pty) Ltd, SA]. The total cell count was done with a haemocytometer and viability check by Trypan Blue exclusion method. A fraction (1% of the total cell count) was placed in the cytocentrifuge for the Cytospin (Simport, Saint-Mathieu-de-Beloil, Canada) and differential count procedure. The Cytospin slide (ELITechGroup, Paris, France) was fixed with a fixative and stained RAPI-DIFF Stain 1 (contains Eosin Y dye in phosphate buffer) and using RAPI-DIFF Stain 2 (Clinical Sciences Diagnostics, Johannesburg, South Africa) containing polychromed methylene blue in phosphate buffer for 30 s each. Cytospin slides were counted manually (to a total of 100 or 200 cells). The BAL was finally fractionated according to the study protocol and cryopreserved in 10% dimethyl sulfoxide (MilliporeSigma, Massachusetts, USA) and 90% fetal bovine serum (Cytiva, Massachusetts, USA) and, by gradual cooling to -80°C in a Nalgene Mr Frosty™ container [Sigma Aldrich (Pty) Ltd, Gauteng, South Africa] with isopropanol for 24 h followed by long term preservation in liquid nitrogen.

PET-CT scans. Household Contacts, PLHIV, and a subset of the Community Controls had PET-CT imaging performed in addition to the routine chest radiographs, according to their study protocol. PET-CT imaging was done at the Nuclear Medicine Research Institute Node for Infection Imaging (NII), Stellenbosch University, South Africa with a Vereos PET-CT scanner (Philips, Amsterdam, Netherlands)¹⁰. 18F-fluorodeoxyglucose (FDG) radiotracer was administered at a dose of 2.8 mBq per kg, 60 min before imaging. Results were reported by consensus between a nuclear medicine physician and a radiologist.

Analyses. Analyses were conducted in R¹¹. Simple descriptive statistics were used to describe the observations, with percentages and proportions for categorical variables, and means with standard deviations or medians with interquartile ranges for continuous data that had a normal and skewed distribution respectively. Robust versions of statistical algorithms as implemented in the R package WRS2¹² were used for analyses since the skewed outcome variables would result in non-normality of residuals.

The predetermined outcomes of interest in the analysis were:

1. BAL volume yield: reported in millilitres of fluid or as the percentage of instilled volume returned with aspiration. An optimal yield from BAL is $>30\%$ of the instilled volume (i.e. >70 ml from a 240 ml BAL)⁴.
2. Total BAL cells (reported as counts) and BAL cell concentration (total cells per millilitre of BAL fluid), and viability (percentage of live cells in total cell count).
3. Pellet colour: categorised as bloody, clear, or black. Of particular importance is the bloody pellet, which implies contamination with blood and therefore trauma to the mucosal surface (a minor adverse event). A bloody pellet requires a cell lysis step during processing to lyse erythrocytes, as the presence of red blood cells may reduce the viability of the immune cells after thawing, and as the cells originate in the peripheral blood rather than the lung, they potentially undermine the results of the immunological end assays done on the BAL fluid.
4. Differential cell count in BAL fluid: reported as the relative proportions of macrophages (normal $>85\%$), lymphocytes (normal 10–15%), neutrophils (normal $\leq 3\%$) and eosinophils (normal $\leq 1\%$). The pattern of abnormality may provide insight into the underlying immune dysfunction⁴.

Data were cleaned and the outcome variables BAL yield and BAL concentration were log transformed and winsorized using a robust median-based adjustment implement in RobustHD¹³ to make linear models tractable. To reduce the number of explanatory variables included in models, the variable selection algorithm Boruta was applied. Variables included in the initial Boruta algorithm were selected based on availability of data, previous evidence of their effect on BAL yields, and the authors' hypothesis of a possible effect. The following variables were subject to selection with the Boruta algorithm for all outcomes: clinical group, smoking status, pellet colour (not included in the analysis for the outcome 'pellet colour'), anatomical lung lobe used for the procedure, age category, gender, recreational substance use, imaging category, Karnofsky score, IGRA status, SARS-CoV-2 status, observed airway abnormality during the procedure, comorbidity, and BAL procedure duration. A stringent cut-off for variable selection ($P < 0.005$), and 3,000 iterations were used to select variables. The final variable list included: anatomical lung lobe, observed airway abnormalities, age, pellet colour, gender, recreational substance use, clinical group and smoking status for BAL volume yield; anatomical lung lobe, pellet colour, age, smoking status, clinical group, imaging category, IGRA status, and gender for BAL total cell yield; smoking status, pellet colour, anatomical lung lobe, clinical group, age, IGRA status, and gender for BAL cell concentration; clinical group, IGRA status, and imaging category for BAL cell viability; and clinical group, Karnofsky score, and

SARS-CoV-2 status for pellet colour. These variables were analysed separately using robust ANOVA methods and post-hoc t-tests if the ANOVA was significant. An ANOVA on the medians was also performed for comparison to the robust ANOVA on trimmed means (5% trim). All P values were corrected for multiple testing using the Benjamini and Hochberg¹⁴ false discovery procedure.

Cell differential counts were analysed differently to the other outcomes. Since the total cell counts varied widely between individuals, cell type counts were converted to percentages of the total count to standardize within subjects and ensure comparability. Robust ANOVA, followed by post-hoc testing and alpha of 0.05, with correction for multiple testing using the Benjamini and Hochberg procedure. Since cell type numbers were correlated, e.g., where there was a higher proportion of lymphocytes, there was a corresponding lower proportion of one or all of the other cell types, a variable that was found to significantly influence the proportion of one cell type, was interpreted as a variable influencing the pattern or overall distribution of the cell differential count.

Adverse events were counted and classified as those occurring during the procedure or in recovery, and those occurring after the procedure, assessed at 72 h and 14 days after the bronchoscopy. Adverse events in different clinical groups were compared using a two-sided Cochran-Armitage test for trend in one-against-the-rest scenarios.

Results

Participant characteristics. A total of 337 bronchoscopy procedures were performed solely for research purposes, on 336 participants, as part of five NIH-funded studies from 1 January 2019 to 15 March 2022 and were included in the analysis. The pertinent characteristics of the study population are described in Table 1.

The most common comorbidity was hypertension (eight in Community Controls, six in Household Contacts, one in the TB groups and four in PLHIV). Four participants had asthma, four had previous chest trauma, and three had allergic rhinitis. The remaining comorbidities were a wide range of minor ailments. Approximately one fifth of the population used recreational substances, of whom 59 (85.5%) smoked cannabis (only five of these did not also smoke tobacco); four (5.8%) used both cannabis and another substance such as methamphetamine or methaqualone (all current tobacco smokers); and six (8.7%) used methamphetamine or methaqualone alone or in combination (all current tobacco smokers).

Of the 78 participants with active TB (the TB Pre-Treatment and TB Early Treatment groups), the mean body mass index (BMI) was 19.5 (14.8–29.0), with 39 (50%) of participants classifying as ‘underweight’ and only four (5.1%) as ‘overweight’. On study entry, 39 (50.0%) participants were sputum smear positive, with scores

	Community controls (n = 150)	Household contacts (n = 57)	TB pre-treatment (n = 24)	TB early treatment (n = 54)	TB end of treatment (n = 22)	PLHIV (n = 29)	Pooled (n = 336)
Age in years	29 (16.8)	34 (17.8)	32 (15.3)	33 (21.5)	31 (17.3)	44 (13.5)*	32 (20)
Female	89 (59.3)	36 (66.7)	7 (35.0)	14 (25.9)	6 (27.3)	25 (86.0)*	178 (54.1)
Current tobacco smoking	101 (67.3)	31 (57.4)	18 (90.0)	40 (74.0)	15 (68.2)	7 (24.1)*	212 (64.4)
Recreational substance use	21 (14.0)	14 (25.9)	6 (30.0)	19 (35.2)	7 (31.8)	2 (6.9)	70 (21.3)
Recent SARS-CoV-2 ⁺	17 (11.3)	14 (24.6)	6 (25.0)	0	0	6 (20.7)	43 (15.9)
Other comorbidity [§]	24 (16)	9 (16.7)	1 (5.0)	8 (14.8)	1 (4.5)	4 (13.8)	47 (13.9)
IGRA status [¶]							
Positive	101 (67.3)	48 (88.9)	ND	ND	ND	21 (72.4)	184 (55.9)
Negative	41 (27.3)	6 (11.1)	ND	ND	ND	8 (27.6)	57 (17.3)
Imaging category							
Normal	141 (94.0)	26 (45.6)	1 (4.2)	0	4 (18.2)	12 (42.9)	184 (54.8)
Abnormal (not TB)	9 (6.0)	20 (35.1)	0	3 (5.6)	14 (63.6)	14 (50.0)	60 (17.9)
Consistent with active TB	0	11 (19.3)	23 (95.8)	51 (94.4)	4 (18.2)	3 (10.7)	92 (27.4)

Table 1. Characteristics of the study population stratified by clinical group. Age is expressed as median (interquartile range); all other variables are expressed as number followed by the percentage (in parentheses) that this number represents of valid samples in the specific study group for the specific variable. Imaging category consists of chest radiograph only in the TB Pre-Treatment, TB Early Treatment and TB End of Treatment groups. In the other groups this is a composite report from both chest radiograph and Positron Emission Tomography-Computed Tomography (PET-CT). IGRA interferon gamma release assay, ND not determined, PLHIV people living with HIV, SARS-CoV-2 severe acute respiratory syndrome coronavirus 2, TB tuberculosis. *Two-sided Cochran-Armitage tests found no statistically significant differences between the first five clinical groups. On ANOVA, age was significantly different between PLHIV and the other clinical groups ($P < 0.001$, F-statistic = 7.842). Fisher's exact test found that the PLHIV group had significantly more females ($P < 0.001$, 99.5% CI for odds ratio 1.77–107.9) and fewer current smokers ($P < 0.001$, 99.5% CI for odds ratio 0.03–0.5) than the other clinical groups. ⁺One study only performed SARS-CoV-2 RT-PCR, not serology, and thus participants were unable to be classified. The unclassified participants were not included in the analysis for this group. [§]Excluding TB, HIV, and previous SARS-CoV-2 infection. [¶]Several participants' IGRA results were ‘indeterminate’ or unknown; IGRA was not performed for participants with active TB.

as follows: five (6.4%) ‘scanty’ (less than 10 AFB in all fields); 17 (21.8%) ‘1+’ (less than 1 AFB per field); 13 (16.7%) ‘2+’ (1–10 AFB per field); and four (5.1%) ‘3+’ (more than 10 AFB per field). Of the 76 who had sputum MGIT culture, 71 (93.4%) were culture-positive, with a mean time to positivity of 10.6 days. Of the 70 that had sputum Xpert Ultra results, 66 (84.6%) were positive, with the following ranking: 17 (24.3%) ‘high’; nine (12.9%) ‘medium’; 13 (18.6%) ‘low’; six (8.6%) ‘very low’; three (4.3%) ‘trace’; 18 unknown. All participants were symptomatic of TB, but well enough to undergo bronchoscopy. Chest imaging findings are presented in Table 1. Of note is that 11 (19.3%) of the Household Contacts group had chest imaging (which included both PET-CT and chest radiographs) which was suggestive of TB. This may be either related to previous TB episodes, or implying possible subclinical TB, as participants were asymptomatic and tested negative for TB on sputum. There were 28 PLHIV included, one of whom had two procedures. Three (10.7%) of these also had chest imaging results consistent with TB despite their sputum negative and asymptomatic status. All PLHIV were on antiretroviral therapy (ART) at the time of bronchoscopy, for a median of 10 years duration (range 0.5–17 years). Of the 18 participants who had CD4 counts within six months before the bronchoscopy, the median was 602 cells/ μL (range 149–842; normal range approximately 500–2000 cells/ μL). Of the 16 participants with HIV viral loads within 6 months, the median was < 20 copies per ml (range 0–242 copies per ml).

SARS-CoV-2 subgroup. The distribution of SARS-CoV-2 participants across clinical groups is shown in Table 1. Of the 42 participants (43 procedures), 31 (72.1%) were females, and the mean age was 36 years. There were similar proportions of current (20; 46.5%) and never (18; 41.9%) smokers. The majority (37; 86.0%) did not use recreational substances, four (9.3%) used cannabis, and two (4.7%) used methamphetamines. Fourteen (32.6%) had a comorbidity, of which the most common was hypertension. Using a composite of both chest imaging modalities, 15 (34.9%) were normal, six (13.9%) were classified as in keeping with active TB, six (13.9%) had subclinical TB, and 16 (37.2%) were abnormal but not in keeping with active TB. Forty participants (95.2%) tested either anti-N antibody positive, anti-S antibody positive, or both. Four (9.5%) tested RT-qPCR positive in the 3 weeks prior to the bronchoscopy, however none of them had symptoms of coronavirus disease-2019 (Covid19) at the time of testing. Only 16 participants (38.1%) reported symptoms of Covid19 at any time (mean time between symptoms and bronchoscopy was 6.4 months; in 18 symptom status was unknown); two of these (4.8% of the group) required hospital admission; none of them had active TB.

BAL yields. Overall, the median BAL volume collected from the procedures was 130 ml (IQR 49 ml), which represents 54.2% of the instilled volume. The median number of BAL cells collected from the procedures was 30.2×10^6 (IQR 38.7×10^6). The median BAL cell concentration (cells/ml) was 0.24×10^6 (IQR 0.34×10^6). The median cell viability (proportion of live cells in the total cell yield) was high, at 96.8% (IQR 2.8%). The analysis showed no significant variability in BAL cell yield over time overall, or when stratified by parent study. There was no significant difference in BAL cell yields between operators performing the bronchoscopy, and no significant variability in their performance over time (Supplementary Fig. S1). We investigated the effect of the BAL volume on the total BAL cell yield and found that when the whole sample was considered there was no significant relationship between the two outcomes. However, when we performed group-specific testing, we found a log-linear relationship in current smokers from the Household Contacts and Community Controls ($P = 0.04$) and in the TB treatment groups ($P < 0.0001$ for TB Early Treatment and $P = 0.044$ for TB Pre-Treatment. This suggests that the true relationship was confounded in sample population analysis, as shown in Fig. 2. There were 333 data points

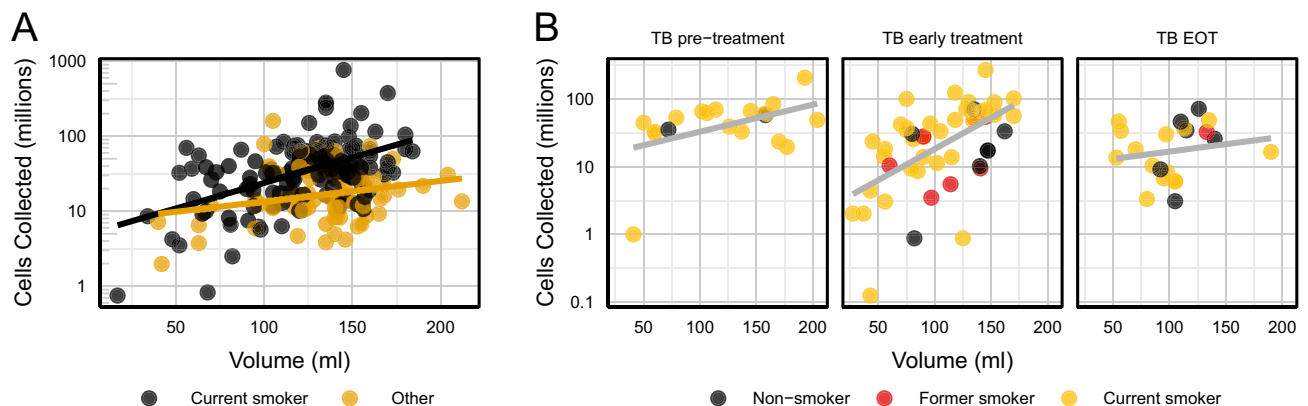


Figure 2. Scatter plot with regression trends showing the relationship between BAL total cell count and BAL volume yield. (A) The correlation between the log cell count and BAL volume in the Household Contacts and Community Control groups (the two groups without any current active infection), stratified by smoking status, either ‘current’ or ‘other’ (prior smokers and never smokers). The trend line for current smokers indicates an increase of 0.61×10^6 cells collected per ml increase in volume (SE 0.21×10^6 cells, $P = 0.004$) (B) Correlation between the log cell count and BAL volume among TB groups. The cell yield increases by 0.32×10^6 cells (SE 0.15×10^6 , $P = 0.044$) per ml BAL volume among participants before TB Pre-Treatment (TB pre-treatment), by 0.54×10^6 cells (SE 0.14×10^6 , $P < 0.001$) per ml among TB Early Treatment participants, and only 0.10×10^6 cells (SE 0.13×10^6 , $P = 0.451$) per ml BAL volume among participants at End of Treatment (TB EOT).

available for the pellet colour analysis: 215 (64.6%) pellets were black, 54 (16.2%) were clear, and 65 (19.5%) were bloody.

Factors affecting BAL yields. Both the Boruta and Random Forest classification identified anatomical lung lobe as having the highest impact on BAL volume yield. In a simple linear model the RML yielded significantly higher volumes compared to all other lobes except the lingula (with RML as reference, RUL -35.4 ml difference in mean, SE 6.4, $P < 0.0001$; LUL -22.5 ml difference in mean, SE 6.8, $P = 0.001$; LLL -30.8 ml difference in mean, SE 13.4, $P = 0.022$; RLL -25.6 ml difference in mean, SE 12.5, $P = 0.042$; lingula -12.0 ml difference in mean, SE 9.3, $P = 0.198$). Increasing age was associated with a lower volume yield when considered as a continuous variable in a linear model with BAL volume as the outcome (-0.7 ml/year, SE 0.2, $P < 0.001$), as was male gender (-8.2 ml, SE 4.1, $P = 0.0447$) though the latter became non-significant after correction for multiple testing. When participant age was divided into three categories, pairwise comparisons (robust median-based ANOVA) showed a relationship between these categories and BAL volume, cell yield and concentration, which remained significant after correction for multiple testing (Fig. 3). The Boruta algorithm identified three variables that were most highly associated with total BAL cell yields: smoking status, gender, and pellet colour. On a simple linear model, current smoking caused a 26×10^6 difference in mean cells from never smokers (SE 7.5×10^6 , $P < 0.0001$); male gender a 16×10^6 cells difference in mean cells from female gender (SE 5.9×10^6 , $P = 0.007$); bloody pellets had a -19.6×10^6 cells difference in mean cells from black pellets (SE 7.3×10^6 , $P = 0.0119$); and clear pellets had -20.1×10^6 cells difference in mean cells from black pellets (SE 9.0×10^6 , $P = 0.027$). The findings for BAL cell concentration were similar. The effect of gender did not stand up to correction for multiple testing, and further interrogation found that the three variables smoking, gender and pellet colour, were highly correlated, explained by the fact that significantly more males were smokers, and smoking is the commonest cause of a black BAL pellet (Fig. 4). To further investigate the effects of smoking on BAL characteristics, pairwise comparisons (robust median-based ANOVA with robust post-hoc t-tests) were performed, showing clearly that although current smokers give the lowest BAL volume yields, they have the highest overall cell yields (Fig. 5). These effects were seen for the black BAL pellet category as well (Supplementary Fig. S2).

After feature selection identified clinical group as important in the BAL yields, pairwise comparisons (one-way ANOVA on the medians with robust post-hoc t-test) found that the PLHIV group was very different from the other clinical groups (Supplementary Fig. S3). To further interrogate differences between other clinical groups which might have been obscured by the analysis with the PLHIV group, we performed pairwise comparisons (robust 2-way ANOVA on the medians) on the other groups. This suggested some interesting trends in BAL volume and cell yields, from participants' TB treatment status (Fig. 6). The TB End of Treatment group yielded lower BAL volumes than the Community Controls group and the Household Contacts group: -60.5 ml difference in modified M-estimator (MOM) ($P < 0.001$) and -71.3 ml difference in MOM ($P = 0.010$) respectively. The TB Pre-Treatment group had higher cell yields than the Community Control group (a difference of 41.9×10^6 in MOM of cell counts, $P < 0.001$), the Household Contacts group (a difference of 29.8×10^6 in MOM of cell counts, $P = 0.038$), the TB Early Treatment group (a difference of 30.4×10^6 in MOM of cell counts, $P = 0.038$) and the TB End of Treatment group (a difference of 45.6×10^6 in MOM of cell counts, $P = 0.030$). The results for comparisons of concentrations were similar. These findings suggest the TB treatment duration affected the cellular content of the BAL. The trend was confirmed on direct comparison of paired Early Treatment and End of Treatment samples (Supplementary Table S2). We performed a two-way ANOVA on MOM between groups with smoking status taken into account, and found that smoking did not affect the interactions reported for BAL volume yield, BAL cell yield or concentrations (Supplementary Table S3). This additional analysis also suggested that differences in smoking status might have confounded a real interaction in the BAL cell yields between the TB End of Treatment group and the Community Controls and Household Contacts (Supplementary Table S3).

Variables which had no effect on the outcomes in the linear models performed after the feature selection suggested they were important, included SARS-CoV-2 status, imaging findings (using a composite of chest radiograph and PET-CT when performed, categorised as 'normal', 'abnormal, in keeping with active TB', or 'abnormal but not active TB'), any airway abnormalities observed during the procedure, IGRA result, recreational substance use, and the presence of other comorbidities. Interestingly, one-way ANOVA on the medians (with post-hoc robust t-test) revealed that participants who had airways which collapsed on minimal suction during the procedure (suggesting underlying airways disease) or abnormal secretions (suggesting chronic lung disease or current infection) had a lower volume yield than those with no observable airways abnormalities (-10.83 ml, SE 4.14, $P = 0.037$). Both abnormal imaging categories ('abnormal but not active TB' and 'abnormal, in keeping with active TB') were associated with a lower BAL volume yield than those with normal imaging results ($P = 0.004$) before correction for multiple testing, but this became non-significant after correction.

Factors affecting BAL pellet colour. ANOVA testing showed that the highest cell counts are found in black pellets, followed by bloody pellets, then clear pellets (Supplementary Fig. S2). If the bloody pellet's higher cell count represents contamination with blood immune cells rather than a higher number of lung immune cells, this has potentially negative implications for the post-thaw BAL cell viability and the interpretation of any BAL immunoassays. Logistic regression analysis found that the PLHIV group had a 70% greater chance of producing a bloody pellet compared to the Community Controls (Odds ratio 0.3 for bloody pellet in Community Controls compared to PLHIV, $P = 0.013$). As the only two parent studies to include PLHIV also had brushings during their procedures, we investigated whether the order in which the procedures were performed had any effect on the pellet colour and found that performing the brushing after the BAL (in the same lung lobe) was associated with a significantly lower rate of bloody BAL pellets compared to performing the brushing before the BAL (26.9% bloody pellets and 69.4% bloody pellets respectively, $P = 0.001$). As the reduced visibility and enhanced cough

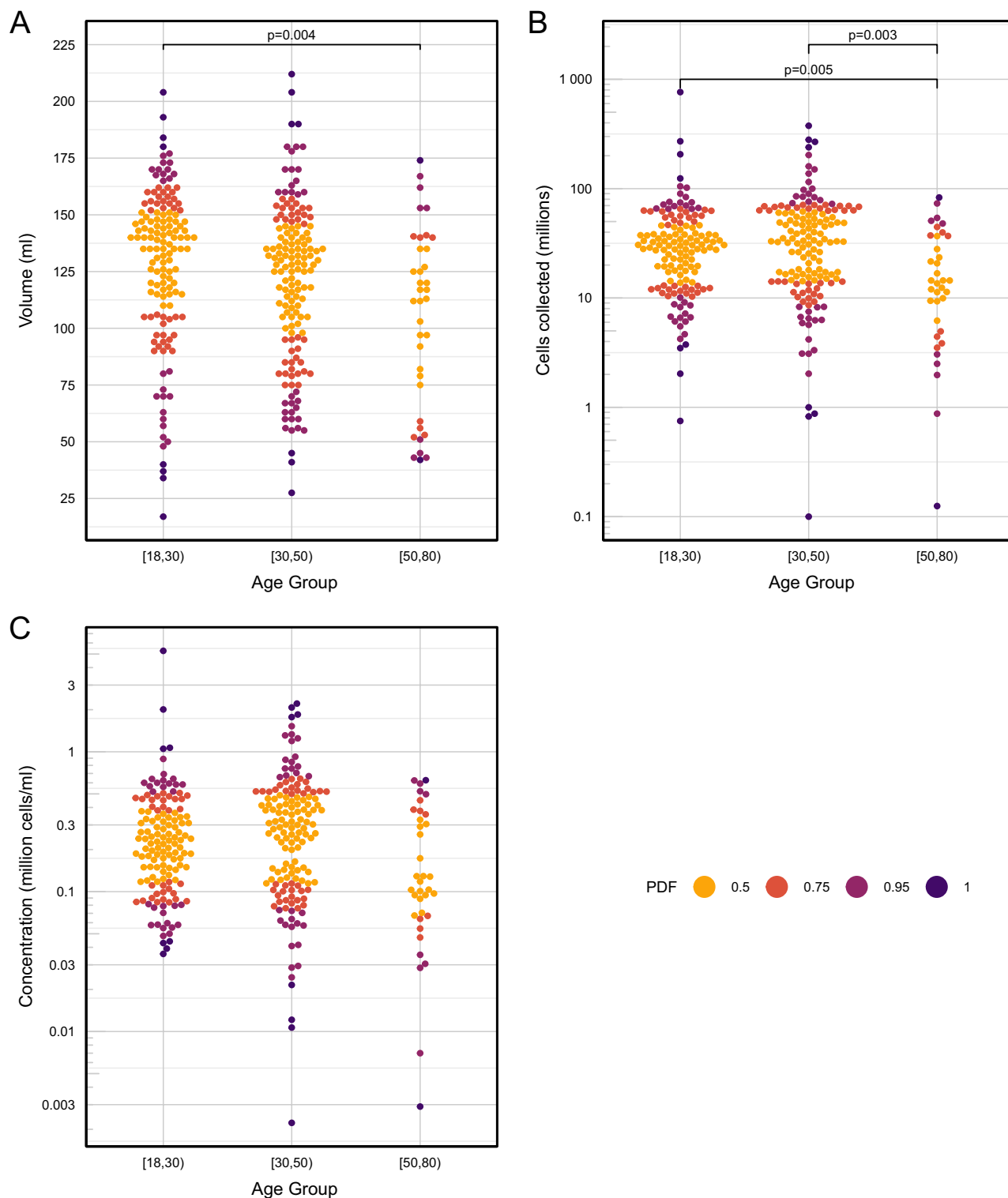


Figure 3. Bee swarm plots showing the effect of participant age on BAL cell yield, volume, and cell concentration. Age was binned into three categories (18–30, 30–50 and 50–80 y). **(A)** BAL volume yield; **(B)** BAL cell yield; **(C)** BAL cell concentration. PDF, a distance metric based on the probability density function showing distance from the mean. ANOVA on medians: **(A)** $F=4.2957$, $P=0.0195$; **(B)** $F=8.202$, $P=0.001$; **(C)** $F=5.2743$, $P=0.004$.

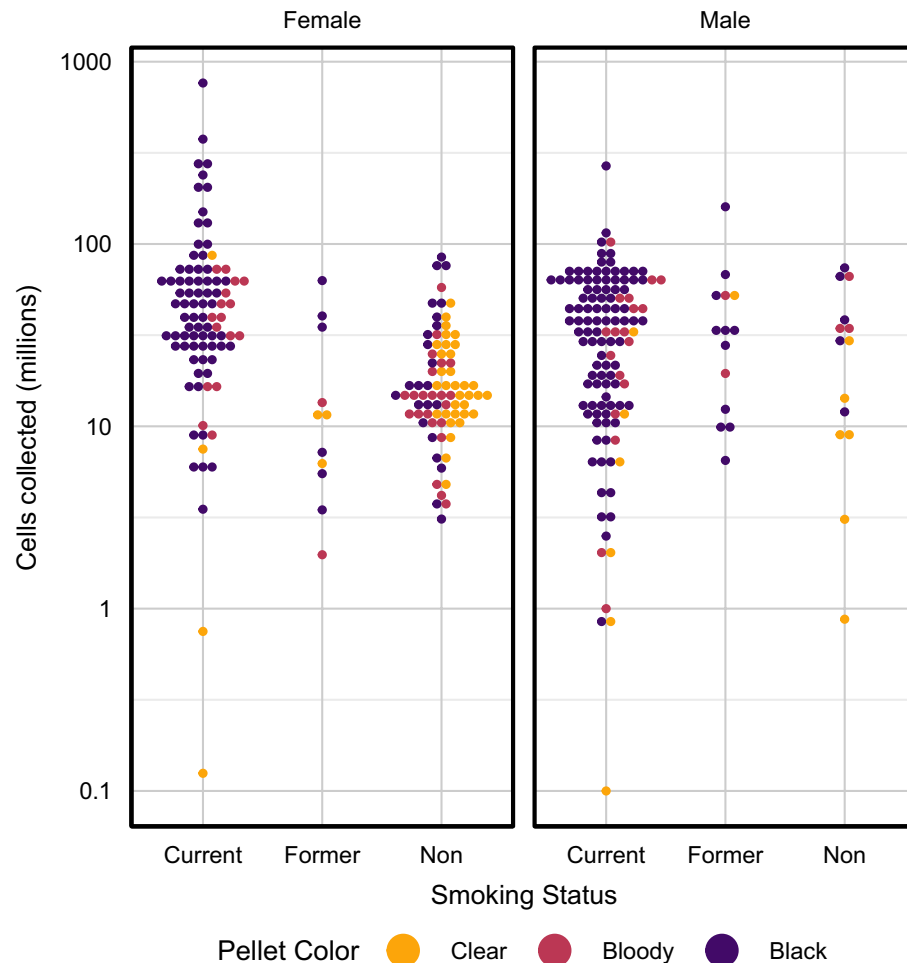


Figure 4. Bee swarm plots showing the relationship between BAL pellet colour, smoking status and gender.

stimulus after the BAL may compromise participant safety when performing a brushing, our group opted to try a third technique in sampling the ipsilateral lower lobe before the BAL, but this was not successful in reducing the proportion of bloody pellets (Supplementary Fig. S4).

BAL cell differential count. It has been shown that smaller instilled BAL volumes sample only the distal airways whereas larger volumes sample the alveolar compartment, resulting in differences in measured protein and cellular contents of BAL fluid³. We therefore investigated whether the BAL volume yield affected cell differential counts, but found that there was no relationship between the two. On pairwise ANOVA of the means, several significant differences in the differential cell counts between groups were found: (1) older age categories were associated with higher neutrophil proportions ($P=0.003$) though the effect size was small; (2) the TB Early Treatment group had the highest proportions of neutrophils of the clinical groups ($P=0.03$); (3) clear BAL pellets had the highest proportions of lymphocytes ($P=0.002$), black pellets had the highest proportions of neutrophils ($P=0.01$), and bloody pellets did not have any significantly enriched cell populations; (4) current smokers had significantly higher proportions of neutrophils and eosinophils; and (5) SARS-CoV-2 positive participants had higher proportions of BAL lymphocytes than SARS-CoV-2 negative participants ($P=0.023$). These results are depicted in Fig. 7, and Supplementary Fig. S5. These interactions became non-significant after correction for multiple testing. The BAL differential cell counts in each clinical group are presented in Table 2.

Adverse events. The mean procedure duration (counted from intubation with the bronchoscope to extubation) was 11.5 min (range 4–22), with a mode of 10 min. Procedures in which a brushing was done with the BAL were significantly longer than those without (1.23 min; 99.5% CI 0.40–2.04; $p<0.001$); the two studies including brushings had more procedures lasting >12 min ($P<0.05$), >14 min ($P<0.05$) and >15 min (Supplementary Fig. S6).

There were no severe adverse events in this study. Minor adverse events were recorded for 118 (33.1%) participants. Supplementary Fig. S7 shows the minor adverse events, stratified by clinical group. Minor bronchial mucosal bleeding was significantly more common in the PLHIV group ($P<0.001$), all of whom also had brushings in the same procedure. Overall, 69 (23.5%) of 293 participants with available data reported still

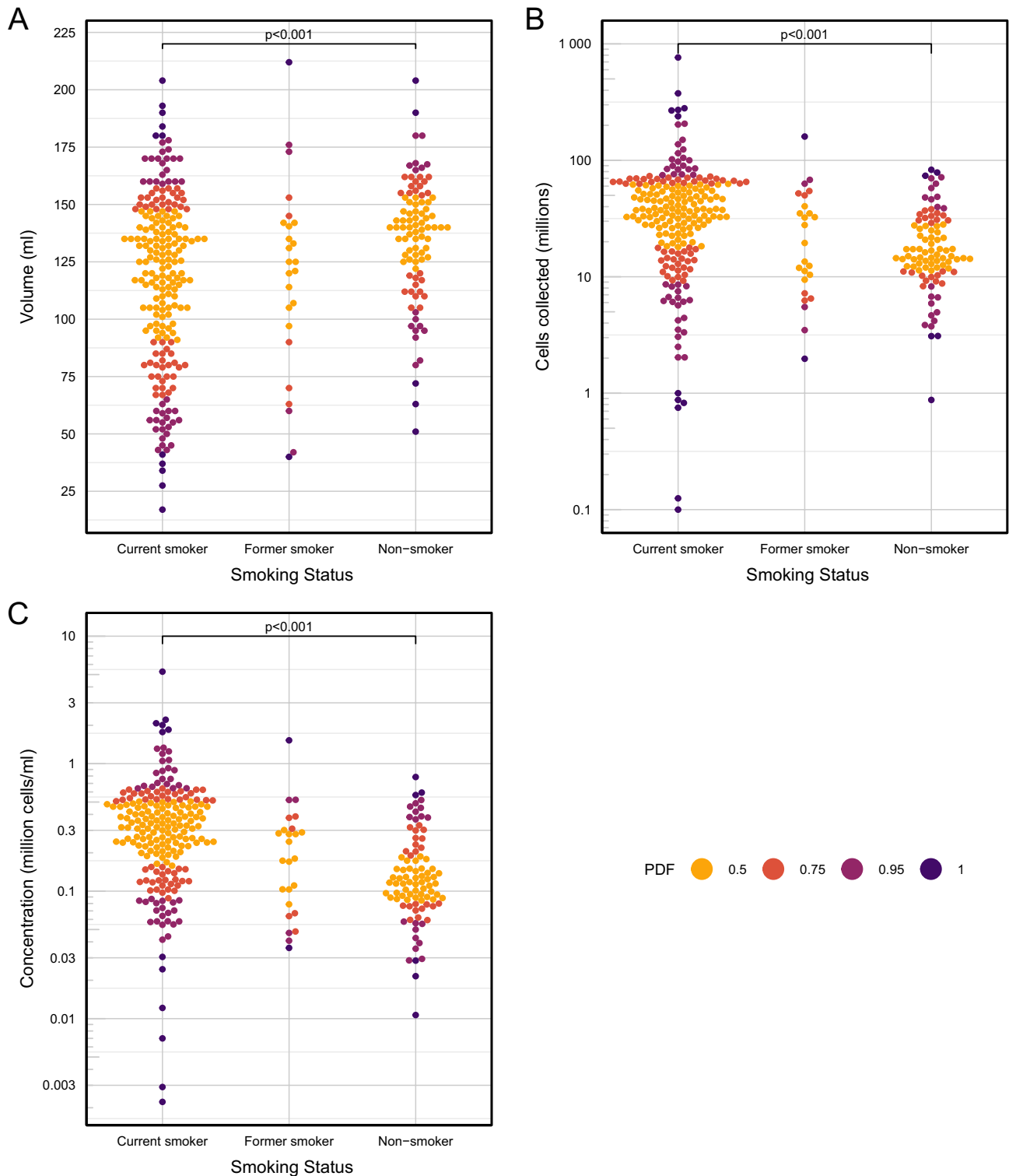


Figure 5. Bee swarm plots showing the effect of smoking on BAL volume, cell yield and cell concentration. (A) BAL volume yield; (B) BAL cell yield; (C) BAL cell concentration. PDF, see Fig. 3. ANOVA on medians: (A) $F = 6.6511$, $P = 0.0075$; (B) $F = 25.754$, $P < 0.001$; (C) $F = 31.0721$, $P < 0.001$.

experiencing symptoms at 72 h after bronchoscopy (Supplementary Fig. S8). ‘Throat discomfort’ and post procedure cough were significantly different between groups, being less frequently reported in the TB Pre-Treatment group ($P = 0.0264$), TB Early- and End of Treatment groups ($P = 0.044$) respectively. PLHIV were more likely to experience dizziness than other groups ($P = 0.010$).

Of those who experienced any symptom, eight (2.7%) sought medical attention for the problem. One (0.3%) study participant was admitted overnight to hospital, three (1.0%) received antibiotics, and six (2.0%) received

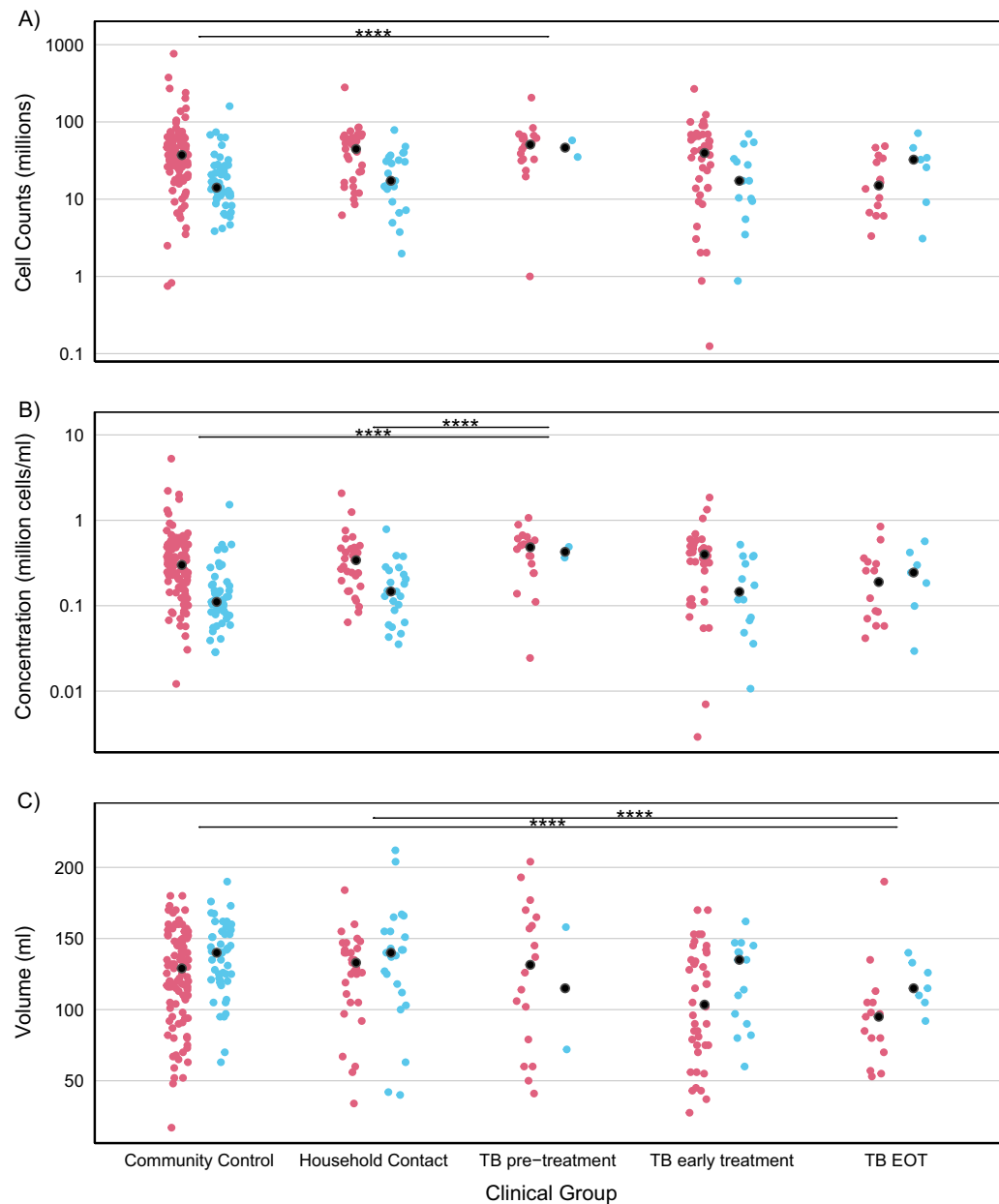


Figure 6. The effect of participant TB clinical group on BAL yields with smoking status taken into account. (A) BAL cell yield; (B) BAL cell concentration; (C) BAL volume yield. Red dots are current smokers; blue dots are prior smokers and never smokers. P values < 0.005 for interactions between groups (including both smoking categories) are shown as ****. The PLHIV group is not included in this analysis.

other medications (including anti-inflammatories, antihistamines, steroids or analgesics). At the day 14 assessment, no participants reported any additional problems. Importantly, at the 72 h assessment, 234 (96.7%) of the 242 participants surveyed stated they would be willing to have a repeat procedure; 236 (97.5%) reported that their overall experience of the bronchoscopy was tolerable; and 239 (98.8%) participants reported that the bronchoscopy procedure was better than they expected. Results were similar at the day 14 assessment.

Discussion

Our findings show that in a sample of over 330 research BAL procedures across diverse clinical groups made up of predominantly current tobacco smokers, the research team may expect a median yield of approximately 130 ml BAL fluid, or 54% of instilled volume, containing 30 million immune cells, with some exceptions. Using our methods, the pre-cryopreservation cell viability was consistently high between procedures, laboratory personnel and bronchoscopy operators. The procedures conducted in a fully equipped, dedicated bronchoscopy suite were safe, with no severe adverse events, with only 2.7% of participants experiencing symptoms which required

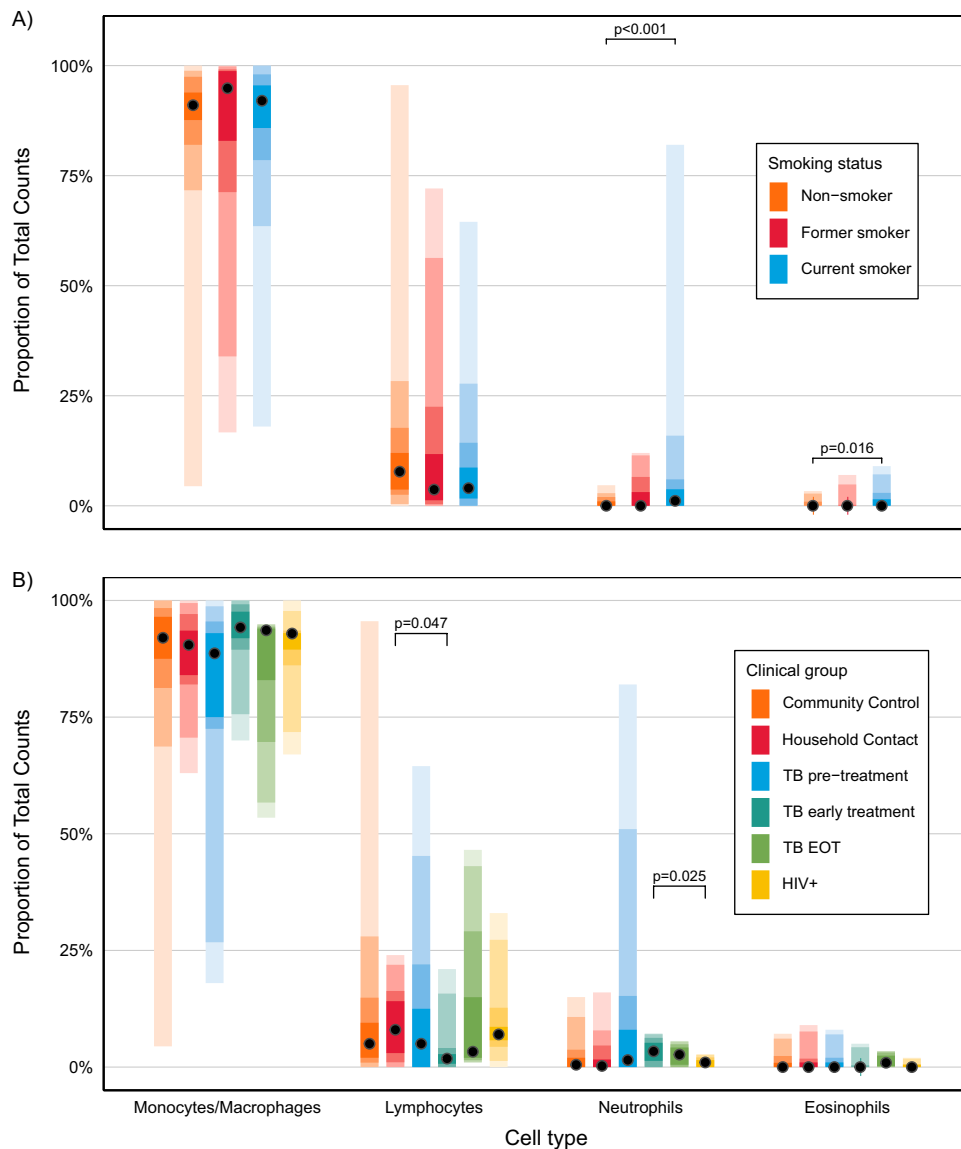


Figure 7. The effect of participant clinical group and smoking status on BAL cell differential counts. **(A)** The effect of participant smoking status on BAL cell differential counts. Orange is never-smokers; red is prior smokers; blue is current smokers. **(B)** The effect of participant clinical group on BAL cell differential count. Clinical groups are represented by colours, in order: orange is Community Controls; red is Household Contacts; blue is TB Pre-Treatment; forest green is TB Early treatment; olive green is TB End of Treatment; yellow is PLHIV. Figures show the relative proportions of each cell type within the total number of cells counted (which was either 100 or 200 for each sample).

Clinical Group	Monocytes/macrophages		Lymphocytes		Neutrophils		Eosinophils	
	Mean ± SD	% (IQR)	Mean ± SD	% (IQR)	Mean ± SD	% (IQR)	Mean ± SD	% (IQR)
Community controls	150 ± 48.6	92.0 (9.0)	13.5 ± 20.6	5.0 (7.5)	2.2 ± 3.7	0.5 (2.0)	1.2 ± 2.1	0 (0.9)
Household contacts	142 ± 44.6	90.5 (9.5)	14.1 ± 11.4	8.0 (11.1)	1.9 ± 3.1	0.3 (1.6)	1.3 ± 2.3	0 (1.0)
TB pre-treatment	134 ± 49.5	88.7 (18.0)	19.9 ± 29.9	5.0 (12.5)	11.9 ± 26.8	1.5 (8.0)	1.3 ± 2.3	0 (1.0)
TB Early treatment	103 ± 15.1	94.2 (5.7)	3.3 ± 5.1	1.8 (2.1)	3.7 ± 2.8	3.4 (3.9)	2.7 ± 2.5	0 (0.0)
TB end of treatment	92.2 ± 9.0	93.6 (11.1)	13.8 ± 20.8	3.3 (13.0)	2.8 ± 2.6	2.7 (3.7)	1.8 ± 1.6	0.9 (2.3)
PLHIV	147 ± 42.3	92.9 (3.5)	15.4 ± 15.6	7.0 (2.9)	1.7 ± 1.5	1.0 (1.5)	0.6 ± 0.9	0 (0.5)

Table 2. Cell differential counts in clinical groups. All values are given as mean absolute cell numbers with standard deviation (SD), and median percentages (%) with interquartile range (IQR) that these numbers represented of the total cells counted. For each sample, either 100 or 200 cells in total were counted. *TB* tuberculosis, *PLHIV* people living with HIV.

medical attention after the bronchoscopy. Performing brushings before the BAL, in the same segment or in the ipsilateral lower lobe, is not advised because of the high rate of blood contamination compared to performing it after the BAL, when the risk of serious adverse events is theoretically higher. Our current practice—informed by these results—is to do the brushing after the BAL, with predetermined safety checks both during the procedure and afterwards, including both a thoracic ultrasound and prolonged observation period.

Consistent with the existing guidelines, the RML yielded the highest BAL volume. However, the factors determining how many cells were present in that volume of fluid were more complex. Smoking appeared to be the single most important factor affecting yields from these procedures. Although current smokers gave consistently lower volume yields than never smokers, their immune cell yields were significantly higher—presumably across all cell types except lymphocytes (which were highest in never smokers)—and increased proportionately with volume yield. Participant age appears to be another important factor. Increasing age was associated with not only lower volume yield but also lower cell yield (as well as lower cell concentration, indicating that the drop in cells is independent of the drop in volume), and a higher proportion of neutrophils within their cell population. Our findings confirm those of Koda and colleagues¹⁵ who identified an anatomical lung lobe target other than the RML or lingula, having a high pack year smoking history and older age as being negatively correlated with BAL volume yield, which they defined as a less than 30% return of instilled BAL volume in their group of 401 patients with diffuse parenchymal lung disease. Olsen et al.¹⁶ found that there was a similar age-dependent decrease in the BAL volume yield, but, in their group of 295 never-smoking healthy volunteers, the cell yield and differential counts were independent of age, gender, season and anatomical lung lobe. Karimi et al.⁷ compared the cell yields in 132 current smokers, 44 prior-smokers and 295 never smokers and concluded that BAL cell recovery was dependent on cumulative smoking history, which they also quantified in pack years. Whilst the pack-years information was available to us in this analysis, in our experience this metric is highly unreliable as an objective measure of smoking intensity, and we therefore chose to exclude it. Karimi et al.⁷ also found that BAL volume recovery was negatively correlated with age, and interestingly that smoking men had a lower BAL volume recovery than smoking women⁷. Men also yielded lower volumes and cells in our results, though neither of these stood up to correction for multiple testing and the high correlation between gender and smoking status suggested this finding was more likely attributable to the difference in smoking status between genders.

The effects of TB on these outcomes were interesting: participants with TB who were untreated or early in treatment yielded the highest cell counts, and there was a steep proportional increase in the number of cells with increasing volume. The practical outcome is that a participant with TB may disappoint the bronchoscopist with a mere 80 ml volume yield, but within that volume are more immune cells than a 'decent' 130 ml yield from a healthy Community Control participant, particularly if that Community Control is a non-smoker. This was not the case for participants who had reached the end of their TB treatment. In general, those at the end of treatment had both lower volume and lower cell yields, possibly attributable to chronic TB-related airway inflammation and damage reducing the volume return, but with little residual active inflammation at the time of lavage, and so fewer immune cells. A previous publication from our group which examined the performance and immune characteristics of BAL in 58 participants with active TB, other lung diseases, household contacts and healthy controls, found a similar significantly lower BAL volume and higher BAL cell yield in the TB group compared to the others, with higher proportion of neutrophils¹⁷, but the current study is the first to provide information on the effects of treatment on these outcomes. An important caveat is that the observed differences in cell differential counts may not translate into differences in immune cell phenotypes and function, and that smoking is still the factor with the highest impact on the cell yields.

From the data one might infer that HIV infection results in a high volume yield with low immune cell content, which has a high proportion of lymphocytes, but there may be a selection bias to consider, as the group of PLHIV that was included were mainly non-smoking middle-aged women who were well-controlled on ART. The lung is a known reservoir for HIV, resulting in pulmonary immune dysregulation, chronic inflammation, and a higher rate of infectious and non-infectious lung diseases¹⁸. Since the dysregulation is only partially restored by ART, our findings would need to be confirmed by a comparison to participants with untreated HIV, smokers living with HIV, and never-smoking community controls. Performing BAL on participants with untreated HIV may, however, result in a higher rate of post-procedural infections, representing a real safety risk for these people, and it should therefore probably be avoided outside of a clinical context¹⁹.

Reassuringly for research projects in the post-Covid-19 pandemic era, there were no significant observable effects of recent mild SARS-CoV-2 infection on our BAL yield outcomes, other than a slightly higher proportion of lymphocytes in the BAL fluid. Importantly though, we cannot make the same conclusion across all severities of SARS-CoV-2 infection, and since no functional assays were included in this analysis it remains unclear how the host lung immune profile is affected by recent SARS-CoV-2.

Overall, our study was limited by the need to select generally healthy candidates for bronchoscopy, potentially excluding important groups such as those with advanced HIV infection and the elderly. As demonstrated above, our results are confounded by numerous important and interrelated variables including the participant's smoking status, recreational drug use, whether a brushing was done in the same procedure, and even age and gender in some groups. Teasing apart the individual effects of these variables on the BAL outcomes is complex, and there may be additional confounders which we have not accounted for. Regarding our assessment of factors affecting the BAL total cell yield and viability, the use of the red cell lysis buffer ACK may adversely affect both these outcomes. As red cell lysis was only included as an optional step in our BAL processing, this may have affected the results in some samples. There was a significant reduction in the cell yield from red pellets compared to black pellets, but not compared to clear pellets. This is reassuring, but there is a possibility that some pellets which were both black and bloody were labelled as bloody only, meaning that the effect size of the red cell lysis step could be minimized by the fact that black pellets start out with more cells. Lastly, our BAL cell viability was only assessed pre-cryopreservation, and post-thaw viability is yet to be established for these samples.

Conclusions

Bronchoscopy with BAL is an invaluable research tool in the study of host immune responses to respiratory infectious disease. Our results show that BAL for research purposes alone, without anticipated clinical benefit, is safe, and even in the presence of current or recent infection with TB, HIV and SARS-CoV-2, the BAL volume yield, cell yield, concentration and viability are consistent and reasonably predictable. Researchers specifically need to consider their target population's smoking status (and possibly gender), age range, HIV status, and duration on TB treatment when selecting end assays which require a large number of immune cells. Clinicians performing the bronchoscopy should wherever possible aim for the RML (with lingula as an acceptable alternative), anticipate a lower volume yield and consider a higher instilled BAL volume wherever a different lobe is targeted. Additionally, if a brushing is planned during the same procedure, then a strategy to avoid contaminating the BAL with blood whilst optimising participant safety, should be identified beforehand. Whether the clinical variables which this study has identified as important in determining BAL volume yield, BAL cell yield, and pre-cryopreservation viability also impact the immune cell phenotype, function, and post-thaw viability, is yet to be established.

Data availability

The datasets generated and analysed during the current study are available from the corresponding author on reasonable request.

Received: 23 January 2023; Accepted: 23 May 2023

Published online: 31 May 2023

References

- Du Rand, I. A. *et al.* British Thoracic Society guideline for diagnostic flexible bronchoscopy in adults: accredited by NICE. *Thorax* **68**, i1 LP-i44 (2013).
- Rahimi, R. A. *et al.* Advancing lung immunology research: An Official American thoracic society workshop report. *Am. J. Respir. Cell Mol. Biol.* **67**, e1-18 (2022).
- Haslam, P. L. & Baughman, R. P. Report of ERS Task Force: Guidelines for measurement of acellular components and standardization of BAL. *Eur Respir J.* **14**, 245–248 (1999).
- Meyer, K. C. *et al.* An official American thoracic society clinical practice guideline: the clinical utility of bronchoalveolar lavage cellular analysis in interstitial lung disease. *Am. J. Respir. Crit. Care Med.* **185**, 1004–1014 (2012).
- Löfdahl, J. M., Cederlund, K., Nathell, L., Eklund, A. & Sköld, C. M. Bronchoalveolar lavage in COPD: fluid recovery correlates with the degree of emphysema. *Eur. Respir. J.* **25**, 275 LP–281 (2005).
- Martin, T. R., Raghu, G., Maunder, R. J. & Springmeyer, S. C. The effects of chronic bronchitis and chronic air-flow obstruction on lung cell populations recovered by bronchoalveolar lavage. *Am. Rev. Respir. Dis.* **132**, 254–260 (1985).
- Karimi, R., Tornling, G., Grunewald, J., Eklund, A. & Sköld, C. M. Cell recovery in bronchoalveolar lavage fluid in smokers is dependent on cumulative smoking history. *PLoS ONE* **7**, e34232 (2012).
- Barbers, R. G., Gong, H. J., Tashkin, D. P., Oishi, J. & Wallace, J. M. Differential examination of bronchoalveolar lavage cells in tobacco cigarette and marijuana smokers. *Am. Rev. Respir. Dis.* **135**, 1271–1275 (1987).
- Seijo, L. M. *et al.* A prospective randomized study comparing manual and wall suction in the performance of bronchoalveolar lavage. *Respiration* **91**, 480–485 (2016).
- Hiemstra, A. M. *et al.* Cascade Immune Mechanisms of Protection against Mycobacterium tuberculosis (IMPAC-TB): Study protocol for the Household Contact Study in the Western Cape South Africa. *BMC Infect. Dis.* **22**, 381 (2022).
- R core team. A language and environment for statistical computing. *R Found. Stat. Comput. Vienna, Austria* (2016).
- Mair, P. & Wilcox, R. Robust statistical methods in R using the WRS2 package. *Behav. Res. Methods* **52**, 464–488 (2020).
- Alfons, A. robustHD: An R package for robust regression with high-dimensional data. *J. Open Source Softw.* **6**, 3786 (2021).
- Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).
- Koda, K. *et al.* Predictors for bronchoalveolar lavage recovery failure in diffuse parenchymal lung disease. *Sci. Rep.* **11**, 1682 (2021).
- Olsen, H. H., Grunewald, J., Tornling, G., Sköld, C. M. & Eklund, A. Bronchoalveolar lavage results are independent of season, age, gender and collection site. *PLoS ONE* **7**, e43644 (2012).
- Young, C. *et al.* Performance and immune characteristics of bronchoalveolar lavage by research bronchoscopy in pulmonary tuberculosis and other lung diseases in the Western Cape, South Africa. *Transl. Med. Commun.* **4**, 7 (2019).
- Alexandrova, Y., Costiniuk, C. T. & Jenabian, M.-A. pulmonary immune dysregulation and viral persistence during HIV infection. *Front. Immunol.* **12**, e808722 (2022).
- Collins, A. M. *et al.* Bronchoalveolar lavage (BAL) for research; obtaining adequate sample yield. *J. Vis. Exp.* <https://doi.org/10.3791/4345> (2014).

Author contributions

J.A.S., N.dP., S.T.M., G.W. conceived and designed the study; A.S., M.F., D.A., L.C., T.F., K.K., K.R., S.B., C.M., A.H., D.S., H.K., I.vR., J.A.S. collected and contributed to the data; M.M. and G.T. performed the statistical analyses; the SU, IRG BAL Study Consortium members were instrumental in study execution at various levels; J.A.S. wrote the paper, which was critically reviewed, edited and approved in the final draft by all authors.

Funding

The bronchoscopy studies reported in this publication were supported by the National Institute of Allergy and Infectious Diseases (NIAID), NIH, through contracts #75N93019C0070 and #75N93019C00071, as well as grants R01 AI134790; R01 AI129980; R01 AI150850 and R01 AI124349. STM received funding from the EDCTP2 program supported by the European Union (grant number CDF1576). GW received funding from the South African National Research Foundation (SANRF; SARChI TB Biomarkers #86535) and the South African Medical Research Council (SAMRC). The Research Bronchoscopy Suite was built and set up by funds from the Faculty of Medicine and Health Sciences (FMHS), Stellenbosch University (SU), South Africa. JAS is funded by the SAMRC through its Division of Research Capacity Development under the Bongani Mayosi National Health Scholars Programme from funding received from the Public Health Enhancement Fund/South African

National Department of Health. The content hereof is the sole responsibility of the authors and does not necessarily represent the official views of any of the funders.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-35723-2>.

Correspondence and requests for materials should be addressed to J.A.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

The Stellenbosch University Immunology Research Group Bronchoalveolar Lavage Study Consortium (SU IRG BAL Study Consortium)

Jane Alexandra Shaw^{1,2,3}✉, Maynard Meiring^{1,2,3,4}, Devon Allies^{1,2,3}, Lauren Cruywagen^{1,2,3}, Tarryn-Lee Fisher^{1,2,3}, Kesheera Kasavan^{1,2,3}, Kelly Roos^{1,2,3}, Stefan Marc Botha^{1,2,3}, Candice MacDonald^{1,2,3}, Andriëtte M. Hiemstra^{1,2,3}, Donald Simon^{1,2,3}, Ilana van Rensburg^{1,2,3}, Marika Flinn^{1,2,3}, Ayanda Shabangu^{1,2,3}, Helena Kuivaniemi^{1,2,3}, Gerard Tromp^{1,2,3,4,5}, Stephanus T. Malherbe^{1,2,3}, Gerhard Walz^{1,2,3}, Nelita du Plessis^{1,2,3}, Elisa Nemes⁶, Léanie Kleynhans^{1,2,3}, Shirley McAnda^{1,2,3}, Charlene Kruger^{1,2,3}, Tracey Richardson^{1,2,3}, Firdows Noor^{1,2,3}, Lauren Benting^{1,2,3}, Nosipho Mtala^{1,2,3}, Robin-Lee Storm^{1,2,3}, Gertrude Persence^{1,2,3}, Elsabe Molima^{1,2,3}, Alicia Chetram^{1,2,3}, Kim Stanley^{1,2,3}, David M. Lewinsohn⁷, Deborah A. Lewinsohn⁸, Kevin B. Urdahl⁹, Erwin Schurr^{10,11}, Marianna Orlova^{10,11}, W. Henry Boom¹² & Sarah Fortune¹³

⁶South African Tuberculosis Vaccine Initiative (SATVI), Institute of Infectious Disease and Molecular Medicine and Division of Immunology, Department of Pathology, University of Cape Town, Cape Town, South Africa. ⁷Division of Pulmonary and Critical Care Medicine, Department of Medicine, Oregon Health and Science University, and Portland Veterans Administration Medical Center, Portland, USA. ⁸Division of Infectious Disease, Department of Paediatrics, Oregon Health and Science University, Portland, USA. ⁹Seattle Children's Research Institute and Departments of Paediatrics and Immunology, University of Washington School of Medicine, Washington, Seattle, USA. ¹⁰Program in Infectious Diseases and Immunity in Global Health, The Research Institute of the McGill University Health Centre, Montreal, QC, Canada. ¹¹McGill International TB Centre and Departments of Medicine, Biochemistry and Human Genetics, Faculty of Medicine and Health Sciences, McGill University, Montreal, QC, Canada. ¹²Department of Medicine, Case Western Reserve University, Cleveland, Ohio, USA. ¹³Department of Immunology and Infectious Disease, Harvard T. H. Chan School of Public Health, Boston, MA, USA.