BMJ Open cfDNA as a surrogate marker for COVID-19 severity in patients with influenza-like symptoms with and without SARS-CoV-2 infections in general practice: a study protocol for a prospective cohort study

Dorothea Dehnen,1 Elmo Neuberger,2 Jürgen in der Schmitten,1 Ekaterini Giagkou,1 Perikles Simon,2 Suzan Botzenhardt

ABSTRACT

Introduction The clinical course of patients with a SARS-CoV-2 (COVID-19) infection varies widely, from symptom-free to severe courses that can lead to death. Laboratory values of SARS-CoV-2 patients such as lymphocyte counts or C-reactive protein (CRP) do not allow a prediction of the actual course of the disease. To identify a possible predictive marker for the differentiation and prognosis of illness with influenza-like symptoms with and without SARS-CoV-2 infections in general practice, we will analyse the concentrations of cell-free DNA (cfDNA) levels, laboratory and clinical parameters, temperature, oxygen saturation, breathing rate and concomitant symptoms in patients with flu-like symptoms with and without a SARS-CoV-2 infection.

Methods and analysis This is a single-centre, two-arm, parallel longitudinal cohort study with a total of 44 patients. 22 patients with flu-like symptoms without a SARS-CoV-2 infection and 22 patients with flu-like symptoms with a SARS-CoV-2 infection will be recruited. The primary objective is to compare cfDNA levels in ambulatory patients in general practice with flu-like symptoms with SARS-CoV-2 infection with those with influenza like symptoms with and without SARS-CoV-2 infections in general practice, we will analyse the concentrations of cell-free DNA (cfDNA) levels, laboratory and clinical parameters, temperature, oxygen saturation, breathing rate and concomitant symptoms in patients with flu-like symptoms with and without a SARS-CoV-2 infection.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ Measurement of cell-free DNA (cfDNA) is cost-effective and requires only minimal amounts of blood, which in the future can be collected in the primary care physician’s office.
⇒ Clinical and serological parameters (high-sensitive CRP, erythrocyte sedimentation rate and a differential blood count) are collected in a setting (family practice) where patients with mild or moderate symptoms are predominantly treated, but where no suitable prognostic markers are available at this time.
⇒ The success of the study should be ensured by the close follow-up and home visits of the patients by the general practitioner and the short duration of the study.
⇒ cfDNA is already a well-established biomarker that is associated with various diseases and has been used in different research areas, such as oncology, non-invasive prenatal diagnosis, organ transplantation, autoimmune diseases, trauma, coronary heart disease and diabetes.
⇒ The limitation of the study due to missing laboratory values at t₁ and t₄ in the case of a severe disease course requiring hospitalisation is addressed by conducting the analysis according to the intention-to-treat principle and assessing hospitalisation as the main outcome.

INTRODUCTION

The current SARS-CoV-2 (COVID-19) pandemic is confronting humanity with a new dimension of medical, economic and social problems.1 Among humans, the virus spreads rapidly and causes varying degrees of severity of symptoms and illness in patients.2 3 Thus, the clinical course of patients with a SARS-CoV-2 infection can vary widely, from symptom-free courses to severe...
courses that can lead to death. The incubation period of the virus ranges from 1 to 14 days, the duration of viral excretion can last from 8 to 37 days and the time from disease onset to discharge or death ranges from 15 to 25 days. Furthermore, it has been shown that mortality rates correlate with increasing age and pre-existing concomitant diseases, such as cardiovascular disease, diabetes, overweight and hypertension. In the course of analysing laboratory values of SARS-CoV-2 patients, such as measurements of lymphocyte counts, C-reactive protein (CRP), as well as secondary bacterial infections it was found that the analysis and evaluation of these do not allow an assessment of the actual course of the disease. In order to be able to better assess the course of this disease, it would be important to find a predictive marker that could be used to determine the severity of the disease at the earliest possible stage.

One marker that could play an important role in this determination is cell-free DNA (cfDNA), which is usually released from cells by apoptosis, necrosis, NETosis, as well as active secretion. It comprises a high variability of fragmented molecules that contain valuable information about gene expression and the nucleosome pattern in relation to their tissue of origin. Numerous studies have already demonstrated that cfDNA levels are associated with various diseases and have been used in various research areas, such as oncology, non-invasive prenatal diagnosis, organ transplantation, autoimmune diseases, trauma, coronary heart disease and diabetes.

Recent studies have investigated the role of cfDNA as a potential marker for therapeutic targets of SARS-CoV-2 in order to develop new therapeutic strategies for the disease. In their study, Chen et al. first-time plasma cfDNA of mild and severe COVID-19 patients. They found that in comparison between mild and severe COVID-19 patients, Interleukin-37 signalling was one of the most relevant pathways. Their data thus revealed potential tissue involvement, provided insights into mechanism on COVID-19 progression and highlighted utility of cfDNA as a non-invasive biomarker for disease severity inspections. In a further study, Andargie et al. showed that cfDNA levels correlated positively with COVID-19 disease severity, CRP and D-dimer, and that the cfDNA profile at admission identified patients who subsequently required intensive care or died during hospitalisation. They conclude that cfDNA could be used as a potential diagnostic biomarker to map sources of injury and as a prognostic biomarker to predict COVID-19 trajectory and outcome by providing mechanistic information about COVID-19-induced tissue injury. However, in this study, cfDNA levels were not measured in patients before they presented at the hospital. The aim of our study is to determine if cfDNA levels differ in patients with mild or moderate influenza symptoms when either SARS-CoV-2 infection or infection with another common respiratory pathogen is present or if these concentrations are similar to each other.

METHODS
Design
This is a single-centre, two-arm, parallel cohort study with a 1:1 allocation ratio to be conducted between August 2021 and April 2022; 22 patients with flu-like symptoms without a SARS-CoV-2 infection and 22 patients with flu-like symptoms with a SARS-CoV-2 infection will be included (n=44, age over 18 years) (see figure 1). Due to our study design, no randomisation is needed. Thus, no blinding will be performed.

PATIENTS
Setting of the study and characteristics of participants
This single-centre study will be conducted at the University Hospital in Essen. Patients who visit their general practice in Mühlheim an der Ruhr with flu-like symptoms will be asked to complete a screening questionnaire to check if they met the inclusion criteria. If they meet the eligibility criteria, they will be included in the study. Written informed consent will be obtained by the principal investigator from all patients willing to participate in the study.

Inclusion and exclusion criteria
All persons enrolled in the study must provide full written informed consent and are required to complete a baseline screening questionnaire to assess their eligibility. Inclusion criteria are as follows:
1. Age 18–99 years.
2. Consent given by the patient or legal representative for blood draw, oropharyngeal or nasopharyngeal swab for a rapid SARS-CoV-2 antigen test and subsequent qPCR, if applicable.
3. Sufficient knowledge of the German language to understand the study content and instructions.
Exclusion criteria are as follows:
1. Severe acute or chronic illness with known elevated cfDNA levels due to the underlying disease, for example:
   1. Tumour disease.
   2. Severe renal insufficiency
   3. Severe/moderate inflammatory diseases
   4. Autoimmune diseases
   5. Rheumatological diseases

Intervention description
All patients included in this study will receive a point-of-care antigen rapid test for SARS-CoV-2 (Roche SARS-CoV-2 Rapid Antigen Test) and a subsequent RT-qPCR on the day of initial presentation. Afterwards, patients will be assigned either to the group ‘SARS-CoV-2 positive...
Robert Koch Institute indicate the frequency of possible respiratory pathogens. Thus, in the 10th calendar week of 2022, a total of 72 (60%) of the 121 sent in sentinel samples (national reference centre influenza) identified respiratory viruses, including most frequently SARS-CoV-2 (n=32, 26%), followed by rhinoviruses (n=20, 17%), human metapneumoviruses (hMPV) (n=18, 15%), human seasonal coronaviruses (hCoV) (n=6, 5%), influenza A(H3N2) viruses (n=4, 3%), parainfluenza viruses (n=3, 2%) and respiratory syncytial viruses (n=2, 2%).

Taking into account the other weekly reports from 2022, hMPV (7%–16%), rhinoviruses (7%–22%) and hCoV (1%–17%) dominated in addition to SARS-CoV-2. Influenza viruses in particular play a subordinate role as pathogens of respiratory tract infections (<5%).

All patients will have venous blood (15.3 mL) collected on three different visits to determine cfDNA concentrations as well as the inflammation markers high-sensitive CRP (hsCRP), erythrocyte sedimentation rate (ESR) and a differential blood count (t₀=day of recruitment, t₁=after 7 days, t₂=after 14 days). In addition, temperature, oxygen saturation and respiratory rate will be measured at all three visits and patients must complete a questionnaire about current symptoms. The blood samples t₀ were collected at the day when the patients came to the general practice with having symptoms occurring in the last 24–48 hours. Blood samples will then be centrifuged directly at the GP’s office at 1600 rpm to obtain the necessary blood plasma to avoid possible lysis of the cells. The samples will be stored at −20°C until the required case numbers are reached and will thereafter be transported to Mainz. Once there, the samples will be stored in the freezer at −80°C until the further analysis for the same purpose of the study objectives is performed. The cfDNA concentrations, including non-disease-specific qualitative aspects such as integrity of DNA, will be determined at the Department of Sports Medicine, Prevention and Rehabilitation at the Johannes Gutenberg University Mainz using a self-established qPCR. The material will be processed in a S2 laboratory.

After 6 months, patients with a SARS-CoV-2 infection will be contacted again by telephone to inquire about possible long-term symptoms (post-COVID-19-syndrome according to the NICE guidelines) and again to determine cfDNA levels in patients who remain symptomatic. Special note will be made of whether there has been a new SARS-CoV-2 infection or a vaccination against SARS-CoV-2 in the meantime.

Analysis will be done intention to treat, as it may not be possible to retrieve blood samples from the patients at t₁ or t₂ due to hospitalisation or other circumstances. All participants can discontinue their participation in the trial at any time for any reason without prejudice to current or future medical care. The investigator may discontinue patients’ participation in the trial for any reason for their safety or in their best interest. If participants express a desire to withdraw from the study, they will receive instructions to complete an ‘end-of-study’ visit.
which will also be voluntary. All participants can receive any concomitant treatment at any time during the trial. However, participants must indicate at each study visit whether they are receiving any concomitant drug therapy.

In the context of this research project, the participants will be insured for potential damages of the biomaterial collection as well as for commuting accidents in accordance with § 2 Para. 1 No. 13b Social Security Code (SGB) VII. After the trial, the patients will receive their primary care by the GP as usual. There is thus no need to provide any additional post-trial care.

Measurement of cfDNA concentrations

CfDNA concentrations will be determined at the Department of Sports Medicine, Prevention and Rehabilitation at the Johannes Gutenberg University Mainz, in a S2 laboratory. We will use validated qPCR assays to determine the concentrations of a 90 and 222 bp fragment in diluted EDTA plasma samples without prior DNA isolation. As described in Neuberger et al, the assays specifically target repetitive sequences, which facilitates highly sensitive cfDNA detection from small amounts of diluted plasma. No DNA isolation is required which saves time, costs and avoids the loss of DNA due to the isolation procedure. The assays show repeatability ≤11.6% (95% CI 8.1 to 20.3), and intermediate precision ≤12.1% (95% CI 9.2 to 17.7). Moreover, the robustness of the assays was demonstrated by incurred sample reanalysis, indicating sufficient validity and sensitivity to quantify cfDNA in the study samples. The blood samples t0 were collected at the day when the patients came to the general practice with having symptoms occurring in the last 24–48 hours. These samples than were directly centrifuged in the general practice.

Participant timeline

The symptomatic participant timeline is presented in table 1.

| Timepoint | T₀ | T₁ | T₂ | T₃ (when symptomatic 6 months after positive PCR result) |
|-----------|----|----|----|-------------------------------------------------------|
| Consent collection | X  |     |    |                                                       |
| Demographics, medical history, disease characteristics |     | X  |    |                                                       |
| Point-of-care antigen rapid test for SARS-CoV-2 |     |     | X  |                                                       |
| Additional RT-qPCR |     |     |     | X                                                     |
| cfDNA determinations | X  | X  | X  | X                                                     |
| hsCRP, ESR, temperature, oxygen saturation and breathing rate |     |     |     | X                                                     |
| Differential blood count |     | X  | X  | X                                                     |
| Concomitant symptoms | X  | X  | X  | X                                                     |
| Questionnaire | X  | X  | X  | X                                                     |
| cfDNA, cell-free DNA; ESR, erythrocyte sedimentation rate; hsCRP, high-sensitive CRP. |     |     |    |                                                       |

Table 1 Research timeline for each participant

OUTCOMES

Main outcome measures

The primary outcome is the determination of cfDNA concentrations in patients with flu-like symptoms with SARS-CoV-2 infection with those with influenza like symptoms without a SARS-CoV-2 infection in general practice using a self-established qPCR from EDTA plasma. These cfDNA concentrations will be determined at t₀ (day 0), t₁ (day 7) and t₂ (day 14). We investigate the difference in cfDNA at t₀ between the two cohorts as primary outcome.

Secondary outcome measures

The secondary objective is to analyse whether there is a correlation between cfDNA concentrations with symptoms/wellness and the severity of the disease. To this end, we will determine whether there is an association between cfDNA concentrations and the variables mentioned below:

1. High-sensitivity CRP (hsCRP, collected at t₀, t₁ and t₂). hsCRP is a routine inflammatory biomarker and will be measured in the patients’ blood samples.
2. Erythrocyte sedimentation rate (ESR, collected at t₀, t₁ and t₂) at day 0, day 7 and day 14. ESR is a routine inflammatory biomarker and will be measured in the patients’ blood samples.
3. Differential blood count, especially regarding lymphocyte and neutrophil granulocytes. Lymphopenia and neutropenia appear to be associated with a severe COVID-19 course.
4. Temperature taken at day 0, day 7 and day 14. Several studies suggest that high fever increases the risk of acute respiratory distress syndrome and should be controlled accordingly at an early stage.
5. Oxygen saturation measured at day 0, day 7 and day 14, because impairment of oxygen is associated with critical illness. Thus, a target spO₂ of 92%–96% is recommended.
6. Breathing rate measured at day 0, day 7 and day 14. It has long been known that determining the respiratory rate is a simple way to assess the prognosis in pneumonia or other lung diseases.
7. Concomitant symptoms (collected at t₀, t₁ and t₂). Concomitant symptoms will be queried via a self-assessment questionnaire where the patients can report any current concomitant symptoms which then have to be rated on a Likert scale.

To assess the severity of dyspnoea, the American Thoracic Society’s Dyspnoea Scale will be used. The
inflammatory and currently non-acute diseases, such as this can be found, for example, in the group of chronic compared with healthy subjects. A group comparable to variance—a twofold increase in cfDNA concentrations that lupus patients had a mean cfDNA level of 44.7 ng/mL with a SD of 53.5 ng/mL. We hypothesise that a SARS-CoV-2 infection will increase the levels by 100% once more compared with the cohort with influenza-like symptoms without a SARS-CoV-2 infection. Furthermore, we want to be sure enough to obtain group sizes of at least 22 participants in each intervention group (total n=44) and to adjust for any dropouts during the study. Sample size is planned by a two-sample t-test on a two-sided significance level of α=5% to achieve a power of more than 80%. We therefore plan to include a total of around 35 patients of comparable age and sex in both the group with a positive and the group with a negative test.

Recruitment will take place in the general practice in Mülheim an der Ruhr via the respective principal physician. We plan to include a total of 44 patients until April 2022.

SAFETY

Adverse event reporting and harms

The risk of the venous blood sampling required for the cfDNA determination can be considered minimal. As with any other venous blood draw, pain may occur during the blood draw. Bruising may also occur, especially if there is insufficient compression on the puncture site after blood collection. In very rare cases, the blood draw may result in infection of the puncture site (thrombophlebitis) or nerve injury. However, there are no serious complications associated with blood sampling. The swab for the SARS-CoV-2 Ag rapid test and the subsequent RT-qPCR will be performed by a GP who is experienced in this procedure. Rarely, minor bleeding can occur during nasopharyngeal swabbing; serious injuries do not occur if the procedure is performed correctly. The investigator will assess the severity of each adverse event and will report all serious and non-serious adverse events in the electronic case report form. The investigator will also assess the causal relationship of the serious adverse events to the trial intervention. Termination criteria have not been defined, as it does not seem reasonable for the planned study with a short survey period.

Sample size calculation

With respect to the group with flu-like symptoms without a SARS-CoV-2 infection, we expect—with a relatively high variance—a twofold increase in cfDNA concentrations compared with healthy subjects. A group comparable to this can be found, for example, in the group of chronic inflammatory and currently non-acute diseases, such as systemic lupus erythematosus. In one study, we showed that lupus patients had a mean cfDNA level of 44.7 ng/mL with a SD of 53.5 ng/mL. We hypothesise that a SARS-CoV-2 infection will increase the levels by 100% once more compared with the cohort with influenza-like symptoms without a SARS-CoV-2 infection. Furthermore, we want to be sure enough to obtain group sizes of at least 22 participants in each intervention group (total n=44) and to adjust for any dropouts during the study. Sample size is planned by a two-sample t-test on a two-sided significance level of α=5% to achieve a power of more than 80%. We therefore plan to include a total of around 35 patients of comparable age and sex in both the group with a positive and the group with a negative test.

Recruitment will take place in the general practice in Mülheim an der Ruhr via the respective principal physician. We plan to include a total of 44 patients until April 2022.

Plans to promote participant retention and complete follow-up

Participants will benefit from the study as they will know immediately whether a SARS-CoV-2 infection is present via point-of-care diagnostics (Roche SARS-CoV-2 Rapid Antigen Test: sensitivity: 96.52 %, specificity: 99.68 %) which will be subsequently validated by RT-qPCR. During the study, the patients will be cared for in their home environment by their GP. If participants express a desire to withdraw from the study, they will be asked to complete an ‘end-of-study’ visit. Data collected up to the time of withdrawal will remain in the trial database and be included in data analysis, unless otherwise indicated by the participant.

Data management

Trial data will be collected in the electronic case report form by the principal investigator at the Institute of General Practice at the University Hospital in Essen. Source documents, defined as any original document or object making it possible to prove the existence or accuracy of data or facts recorded during the research, will be kept by the principal physician according to the regulations in force. All questionnaire data will be entered twice by two different persons to ensure the dual control principle. Using a software tool, a third person will check the agreement between the two data sets resulting from the double entry. In cases where entries deviate from one another, the third person will determine the correct entry by looking at the questionnaire. In cases where the questionnaire answers are ambiguous, two persons will decide what should be entered by discussion until a consensus is reached. All data concerning participant information will be stored in locked file cabinets accessible only by the principal investigator. All collected data will be pseudonymised and will therefore be traceable only by means of a code. All files containing names or other personal identifiers, such as the informed consent forms, will be stored separately from the data containing this code number.

STATISTICAL METHODS

Statistical methods for primary and secondary outcomes

The statistical analysis of the data will be performed with the statistical programme SPSS, R or SAS using a pseudonymised data set. A correlation of the cfDNA concentration with the presence or absence of a SARS-CoV-2 infection, as well as a correlation of these results with the data obtained from the questionnaires will be determined. Descriptive statistics will summarise all study variables.

All data will be tested for normal distribution before and after log transformation using the Kolmogorov-Smirnov test with the Lilliefors correction. If the assumption of normally distributed data cannot be rejected, arithmetic group means±SD will be calculated. If the normality test fails in at least one of the study groups compared, all data will be expressed as group medians with IQRs given in parentheses. Between-group comparisons of
primary and secondary outcomes will be performed at t₀ and at two consecutive time points independently (t₁, t₂) using unpaired Student’s t-tests (comparing two groups, normally distributed) or the Mann-Whitney U test (comparing two groups, not normally distributed) or analysis of variance (three groups) and χ² tests for numerical and categorical data, respectively. P values will be considered statistically significant if $p \leq 0.05$. We plan to perform a subgroup analysis for the concomitant symptoms and the cfDNA levels at t₀, t₁, and t₂ in the event of differences within each arm.

Statistical analyses will be carried out according to the intention-to-treat approach and therefore will include all participants. The extent of missing data will be analysed. We will explore missing data patterns and determine the type of missing data. We will use multiple imputation to substitute missing values.

**Patient and public involvement**
No patients were involved in the development of the research questions. The results of the temperature measurement, the oxygen saturation and breathing rate serve as the research questions. The results of the temperature measurement, the oxygen saturation and breathing rate will be communicated to the attending physician. The laboratory and clinical parameters and the levels of the cfDNA will be disseminated after they have been measured.

**Acknowledgements**
We acknowledge support by the Open Access Publication Fund of the University of Duisburg-Essen.

**Contributors**
DD is the principal investigator; she conceived the study, obtained funding, led the proposal and protocol development, and assisted in manuscript preparation and revision. EN contributed to the study concept and design and assisted in manuscript preparation and revision. SB contributed to the study concept and design and assisted in manuscript preparation. PS contributed to the acquisition and analysis of quantitative data and the development of the intervention and assisted in manuscript preparation and revision. SJ contributed to the study concept and design and drafted the manuscript. All authors read and approved the final manuscript.

**Funding**
This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. This study is funded entirely by the University of Essen and the University of Mainz.

**Competing interests**
None declared.

**Patient and public involvement**
Patients and/or the public were not involved in the design, or conduct, or reporting or dissemination plans of this research.

**Patient consent for publication**
Consent obtained directly from patient(s).

**Ethics approval**
This study involves human participants and was approved by Ethics Committee Essen under the reference number 21-9916-BO. Participants gave informed consent to participate in the study before taking part.

**Provenance and peer review**
Not commissioned; externally peer reviewed.

**Data availability statement**
The study principal investigator and the coinvestigators will have access to the full study data and materials. The authors will be willing to share the individual-level study data after completion and publication of primary and secondary analyses. No data are available.

**Open access**
This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.
a retrospective study within a multidisciplinary approach. *Int J Infect Dis* 2020;100:258–63.

27 Wu C, Chen X, Cai Y, et al. Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. *JAMA Intern Med* 2020;180:934.

28 Petrilli CM, Jones SA, Yang J, et al. Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study. *BMJ* 2020;369:m1966.

29 Shenoy N, Luchtel R, Gulani P. Considerations for target oxygen saturation in COVID-19 patients: are we under-shooting? *BMC Med* 2020;18:260.

30 Strauß R, Ewig S, Richter K, et al. The prognostic significance of respiratory rate in patients with pneumonia. *Dtsh Aerzteblatt Online* 2014.

31 Marshall JC, Murthy S, Diaz J, et al. A minimal common outcome measure set for COVID-19 clinical research. *Lancet Infect Dis* 2020;20:e192–7.