Aus dem
Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie (IBE)
Institut der Ludwig-Maximilians-Universität München

Patient Stratification and the Short- and Long-Term Health Consequences of SARS-CoV-2 Infection

Dissertation
zum Erwerb des Doctor of Philosophy (Ph.D.)
an der Medizinischen Fakultät
der Ludwig-Maximilians-Universität München

vorgelegt von
Yanyan Shi

aus
Linfen, China

Jahr
2025

Mit Genehmigung der Medizinischen Fakultät der
Ludwig-Maximilians-Universität München

Erstes Gutachten:	Prof. Dr. Eva Grill
Zweites Gutachten:	Prof. Dr. Klaus Jahn
Drittes Gutachten:	Priv. Doz. Dr. Sandra Frank
Viertes Gutachten:	Prof. Dr. Li Deng

Dekan: Prof. Dr. med. Thomas Gudermann

Tag der mündlichen Prüfung: 13.10.2025

Affidavit

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Promotionsbüro
Medizinische Fakultät

MMRS

Affidavit

Shi, Yanyan
Surname, first name

Marchioninistraße 15
Street

81377, Munich, Germany
Zip code, town, country

I hereby declare, that the submitted thesis entitled:

Patient Stratification and the Short- and Long-Term Health Consequences of SARS-CoV-2 Infection

……

is my own work. I have only used the sources indicated and have not made unauthorised use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

I further declare that the dissertation presented here has not been submitted in the same or similar form to any other institution for the purpose of obtaining an academic degree.

Munich, 28.10.2025 Yanyan Shi

place, date Signature doctoral candidate

Confirmation of congruency

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Promotionsbüro
Medizinische Fakultät

MMRS

Confirmation of congruency between printed and electronic version of the doctoral thesis

Shi, Yanyan
_____
Surname, first name

Marchioninistraße 15
_____
Street

81377, Munich, Germany
_____
Zip code, town, country

I hereby declare, that the submitted thesis entitled:

Patient Stratification and the Short- and Long-Term Health Consequences of SARS-CoV-2 Infection

……

is congruent with the printed version both in content and format.

Munich, 28.10.2025 Yanyan Shi

_____ _____

place, date Signature doctoral candidate

Table of content

List of abbreviations

aHR	adjusted Hazard Ratio
BMI	Body Mass Index
BRS	Brief Resilience Scale
CI	Confidence Interval
COVID-19	Coronavirus disease 2019
DGPI	Deutsche Gesellschaft für Pädiatrische Infektiologie
	(German Society of Pediatric Infectious Diseases)
ENT	Ear, Nose, and Throat
ICU	Intensive Care Unit
MICE	Multiple Imputation by Chained Equations
NAPKON	Nationales Pandemie Kohorten Netz
	(The National Pandemic Cohort Study Network)
NICE	UK's National Institute for Health and Care Excellence
OR	Odds Ratio
PCR	Polymerase Chain Reaction
PCS	Post-COVID-19 Syndrome
PH	Proportional hazard
RF	Random Forest
RSV	Respiratory Syncytial Virus
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
WHO	World Health Organization

List of publications

Peer reviewed publications used for this doctoral thesis

- **Shi Y**, Strobl R, Berner R, Armann J, Scheithauer S, Grill E. Six Clinical Phenotypes with Prognostic Implications were identified by Unsupervised Machine Learning in Children and Adolescents with SARS-CoV-2 Infection: Results from a German Nationwide Registry. Respiratory Research. 2024 Oct 30;25(1):392. doi: 10.1186/s12931-024-03018-3.
- **Shi Y**, Strobl R, Apfelbacher C, Bahmer T, Geisler R, Heuschmann P, Horn A, Hoven H, Keil T, Krawczak M, Krist L, Lemhöfer C, Lieb W, Lorenz-Depiereux B, Mikolajczyk R, Montellano FA, Reese JP, Schreiber S, Skoetz N, Störk S, Vehreschild JJ, Witzenrath M, Grill E; NAPKON Study Group. Persistent symptoms and risk factors predicting prolonged time to symptom-free after SARS-CoV-2 infection: an analysis of the baseline examination of the German COVIDOM/NAPKON-POP cohort. Infection. 2023 Dec;51(6):1679-1694. doi: 10.1007/s15010-023-02043-6.

Conference Contributions

- **Shi Y**, Strobl R, Berner R, Armann J, Scheithauer S, Grill E. Uncovering Prognostic Clinical Phenotypes in Children and Adolescents with SARS-CoV-2 Infection. 17th European Public Health Conference 2024, poster. Lisbon, Portugal. doi: 10.1093/eurpub/ckae144.1932.
- **Shi Y**, Strobl R, Apfelbacher C, Bahmer T, Geisler R, Heuschmann P, Horn A, Hoven H, Keil T, Krawczak M, Krist L, Lemhöfer C, Lieb W, Lorenz-Depiereux B, Mikolajczyk R, Montellano FA, Reese JP, Schreiber S, Skoetz N, Störk S, Vehreschild JJ, Witzenrath M, Grill E; NAPKON Study Group. Persistent symptoms and risk factors predicting prolonged time to symptom-free after SARS-CoV-2 infection: an analysis of the baseline examination of the German COVIDOM/NAPKON-POP cohort. 18th Annual Meeting of the German Society for Epidemiology (DGEpi) 2023, oral presentation. Würzburg, Germany

1. My contribution to the publications

1.1 Contribution to publication I

I, Yanyan Shi, was leading the following activities: conceptualization, programming, model development, formal analysis, data visualization, writing the original draft and revising the draft.

I conducted literature review, prepared and adapted the data of the registry from the German Society of Pediatric Infectious Diseases. I conducted data interpretation, developed the methods and statistical analysis strategy, and applied an unsupervised machine learning algorithm – hierarchical clustering analysis. I checked for the number of optimal clusters and finalized the visualization of the data in the form of a profile plot. The statistical methodology was validated and checked by Ralf Strobl (RS). Eva Grill (EG) supervised this work and provided feedback on the design of this study. Variable selection was done with the guidance of Reinhard Berner (RB), Jakob Armann (JA), and Simone Scheithauer (SS). I finished the first draft of the manuscript and incorporated the feedback from coauthors (EG, RS, RB, JA, SS) into the final version. EG approved this manuscript through multiple stages.

1.2 Contribution to publication II

I, Yanyan Shi, was involved in the following activities: conception and design, data interpretation, statistical analysis, drafting and revising the manuscript. Eva Grill (EG) supervised the publication and provided feedback on the conceptualization. I prepared the National Pandemic Cohort Study Network (NAPKON) data, and I was in charge of interpreting the data, developing and conducting the statistical analysis, with collaborative support from Ralf Strobl (RS) and EG. I performed the survival analyses to visualize and test the time to symptom-free between different groups of participants. I checked the proportional hazard (PH) assumption, adapted model by stratification on “symptom burden” and used a stratified Cox proportional hazard regression model to investigate the predictors associated with delayed recovery. During the review phase, I addressed the constructive comments from Stefan Schreiber, Thomas Bahmer, Wolfgang Lieb, Michael Krawczak, Peter Heuschmann, Stefan Störk, Anna Horn, Lilian Krist, Thomas Keil, Martin Witzenrath, Jens Peter Reese, and Felipe A. Montellano. These comments were related to the interpretation of the timeline of the data, potential inclusion of other predictors and exclusion of certain subgroups of the participants from the perspective of the design and data acquisition of the COVIDOM/NAPKON-POP study. Furthermore, I worked on the revision of the manuscript following comments from other coauthors (Jörg Janne Vehreschild, Ramsia Geisler, Nicole Skoetz, Rafael Mikolajczyk, Hanno Hoven, Bettina Lorenz-Depiereux, Christina Lemhöfer, and Christian Apfelbacher). EG made further revisions and approved the manuscript at multiple stages.

2. Introductory summary

2.1 Background and relevance

Since early 2020, the coronavirus disease 2019 (COVID-19) pandemic caused by infection with the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has deeply affected human life on a global scale, imposing unprecedented challenges on individuals, communities, as well as healthcare systems. As of October 2024, over 700 million COVID-19 cases were diagnosed and more than seven million mortality were documented globally [1].

During the acute SARS-CoV-2 infection, individuals present with varying clinical manifestations, ranging from mild symptoms to critical condition including organ failure and death [2]. In children and adolescents, severe cases have been documented [3, 4], despite the majority of this population typically experience milder courses of disease following SARS-CoV-2 infection compared to adults [5, 6]. Many characteristics have been linked to severe disease in children and adolescents, including symptoms during acute infection, preterm birth, pre-existing comorbidities, and co-infection with respiratory syncytial virus (RSV) [7-13]. However, it is still not completely understood how typical characteristics of affected children and adolescents translate into severe disease outcomes at discharge from the hospital. This doctoral thesis uses the outcomes at discharge from the hospital to describe the short-term health consequence of SARS-CoV-2 infection.

Beyond the immediate health concerns and acute cases, COVID-19 has given rise to a troubling set of long-term health issues. Many patients, including those who were initially asymptomatic or experienced only mild-to-moderate symptoms amid the acute infection, developed or continued to suffer from a range of signs and symptoms after the primary infection had resolved [14, 15]. Many terms were developed to describe these signs and symptoms, among which “Post-COVID-19 Syndrome (PCS)”, “Long COVID”, and “Post-COVID condition” are the most widely used [16-18]. “Long COVID” is a term collectively created by patients to describe these signs and symptoms [18]. In the definition of the UK’s National Institute for Health and Care Excellence (NICE), Long COVID is the signs and symptoms lasting over four weeks, while PCS refers to symptoms persisting beyond 12 weeks [17]. This doctoral thesis follows the NICE terminology of PCS to describe the long-term health consequence of SARS-CoV-2 infection.

The world health organization (WHO) estimated that around 10–20% of SARS-CoV-2-infected patients might experience PCS [16]. Many studies revealed that time to recovery exceeded six months [19, 20]. However, the evolving nature of PCS has led to substantial variability in definitions and diagnostic criteria across studies, resulting in limited consensus on the prevalence, the recovery time and preventable risk factors associated with PCS [21, 22].

A deeper understanding of the patient stratification as well as the short- and long-term health consequences of SARS-CoV-2 infection is essential to facilitate optimized care and rehabilitation protocols tailored to the needs of individuals and to guide public health initiatives for SARS-CoV-2-infected patients. This PhD project advances the understanding of patient stratification and the

immediate and long-term effects of SARS-CoV-2 infection on health by examining two different time frames following infection: the acute infection phase and the period from infection to at least six months after infection. We also contribute a broader life course perspective by investigating outcomes of children, adolescents, and adults after the acute infection period. Specifically, we present the clinical phenotypes and the short-term outcomes in SARS-CoV-2-infected children and adolescents, we quantify the prevalence and time to recovery from PCS in adults. Ultimately, our findings can serve as a risk stratification tool for patients with SARS-CoV-2 infection, and as a blueprint for other viral diseases that pose comparable short- and long-term consequences.

2.2 Exploring the short- and long-term health consequences of SARS-CoV-2 infection

2.2.1 Acute SARS-CoV-2 infection and short-term health consequence in children and adolescents

In comparison to adults, the course of SARS-CoV-2 infection is less severe in children and adolescents [5, 6]. However, serious illness necessitating admission to the intensive care unit (ICU) was observed in 0.02% of all infected children and adolescents [4]. Those who underwent severe disease were at an increased risk of unfavorable short-term outcomes, including mortality [23], as well as long-term consequences of PCS [24, 25].

Many factors can affect the severity of acute disease in children and adolescents. Signs and symptoms including respiratory symptoms (dyspnea, abnormal breath sounds, and chest recessions), gastrointestinal symptoms (diarrhea and vomiting), neurological symptoms (seizures), and general symptoms (fever) were shown to be relevant to moderate/severe disease [11, 13]. Apart from signs and symptoms during the acute phase, the most identified risk factors associated with severe disease were preterm birth, comorbidities such as cardiac disease, neurological disease, diabetes, obesity, asthma, and immunocompromised condition [7-11]. One study also showed that RSV coinfection was correlated to severe COVID-19 among those hospitalized patients less than five years old [12]. However, a clear stratification of patients based on those clinical and epidemiological characteristics during acute SARS-CoV-2 infection is still needed.

Clinical phenotype identification is a promising approach to stratify patients. Clinical phenotypes disclose how the population being studied can be divided into clusters with different epidemiological and clinical characteristics [26, 27]. Following the COVID-19 pandemic, some studies sought to identify distinct clinical phenotypes in adults using variables encompassing demographics, symptoms, comorbidities, treatment for comorbidities, radiographic findings, laboratory parameters, and other COVID-19 risk factors like smoking [26-28], but clinical phenotypes focusing on children and adolescents with SARS-CoV-2 infection were not reported.

The short-term outcomes of children and adolescents after acute SARS-CoV-2 infection include full recovery, mortality and complications [29, 30]. How patients' clinical and epidemiological fea-

tures during acute SARS-CoV-2 infection can affect patients' outcomes is still not completely understood. For the clinical phenotype identification approach, although it is different from the traditional method of investigating predictive associations of each risk factor with certain disease outcomes [26], it does play an important role in helping physicians understand disease pathophysiology, predict outcomes and guide personalized case strategies [31]. Some of the identified clinical phenotypes in SARS-CoV-2-infected adults indicated potential prognostic values for short-term outcomes. To give an example, phenotype with older age, high severity of illness, and high frequency of shock were associated with mortality during ICU [28], similarly, phenotype with older age, higher frequency of comorbidities, and laboratory parameters indicating systemic inflammation were associated with mortality after 30 days [26]. However, whether specific clinical phenotypes have similar prognostic short-time validity among SARS-CoV-2-infected children and adolescents remains to be investigated.

2.2.2 Post SARS-CoV-2 infection to at least six months after: prevalence, duration, and risk factors

After acute infection, patients of all ages may experience delayed recovery. PCS can influence patients' health-related quality of life, complicate their recovery trajectory, and affect their ability to return to daily activities or productive work [32-34]. Nevertheless, there is still no consensus on the burden of PCS, including the estimation of the prevalence, the recovery time, and the risk factors prolonging recovery.

The most common symptoms of PCS include fatigue, difficulty breathing/dyspnea, myalgia, difficulty concentrating, post-exertional malaise, sleep disorder, intolerance of effort, cognitive symptoms including memory loss and cognitive impairment [35, 36]. Nevertheless, estimation of PCS prevalence differs across populations. One population-based cohort study in Australia estimated that approximately 5% of patients showed PCS, but most COVID-19 cases experienced fast recovery [21]. Another study from Scotland estimated the prevalence of at least one symptom was 13.8% at six months post-infection [19]. However, prevalence of PCS was found to be 22.5% in the US [22]. A population-based prevalence estimation is still needed.

In terms of time from acute infection to recovery, there are huge discrepancies across different settings. One study reported only 4% of infected individuals not recovered at 120 days post infection [21], another study reported the proportion of not having recovered as 17.2% (14.0% to 20.8%) at 24 months after infection [20]. A study from Ethiopia reported 50% of the participants were recovered within 9 days [37]. Another study reported 50% of individuals with mild acute disease recovered in 63 days and 50% of those with moderate acute disease recovered in 232 days [38]. In another international online cohort, the estimated risk of symptoms persisting longer than 35 weeks was 91.8% [36]. Thus, it is essential to address the between-study heterogeneity and provide a reliable estimation of time to recovery after SARS-CoV-2 infection.

Many studies have explored the risk factors of prolonged recovery. Some studies reported older age, female sex, overweight/obesity, number of pre-existing health conditions, severe COVID-19 disease during the acute phase, and smoking to be risk factors for PCS [21, 38, 39]. However,

whether there are other preventable or modifiable factors that can delay recovery is still awaiting investigation, to better inform public health intervention strategies.

2.2.3 Research questions and objectives

This PhD project was designed to address the following research questions:

How can patients infected with SARS-CoV-2 be stratified? What are the short- and long-term health effects following SARS-CoV-2 infection?

More specifically, we sought to investigate:

A) Do specific clinical phenotypes exist among SARS-CoV-2-infected children and adolescents?

B) Can these clinical phenotypes in children and adolescents inform prognosis?

C) What is the prevalence of COVID-19-related symptoms in adults over six months post SARS-CoV-2 infection?

D) How long does it take for adult patients to recover following infection with SARS-CoV-2?

E) What characteristics lead to delayed recovery in SARS-CoV-2-infected adults?

2.2.4 Overview of the scientific publications encompassed in this cumulative thesis

This cumulative thesis comprises two scientific publications:

- **Shi Y**, Strobl R, Berner R, Armann J, Scheithauer S, Grill E. Six Clinical Phenotypes with Prognostic Implications were identified by Unsupervised Machine Learning in Children and Adolescents with SARS-CoV-2 Infection: Results from a German Nationwide Registry. Respiratory Research. 2024 Oct 30;25(1):392. doi: 10.1186/s12931-024-03018-3. (Hereinafter referred to as “publication 1”)
- **Shi Y**, Strobl R, Apfelbacher C, Bahmer T, Geisler R, Heuschmann P, Horn A, Hoven H, Keil T, Krawczak M, Krist L, Lemhöfer C, Lieb W, Lorenz-Depiereux B, Mikolajczyk R, Montellano FA, Reese JP, Schreiber S, Skoetz N, Störk S, Vehreschild JJ, Witzenrath M, Grill E; NAPKON Study Group. Persistent symptoms and risk factors predicting prolonged time to symptom-free after SARS-CoV-2 infection: an analysis of the baseline examination of the German COVIDOM/NAPKON-POP cohort. Infection. 2023 Dec;51(6):1679-1694. doi: 10.1007/s15010-023-02043-6. (Hereinafter referred to as “publication 2”)

2.3 Methods and results for publication 1

2.3.1 Methods

2.3.1.1 Data source

Data from the DGPI registry (“Deutsche Gesellschaft für Pädiatrische Infektiologie”, German Society of Pediatric Infectious Diseases) were used to answer scientific questions A and B.

The DGPI registry is initiated by DGPI in Germany, which aimed at including children and adolescents who were hospitalized in pediatric hospitals across Germany and had a confirmed SARS-CoV-2 infection [40]. Data on demographics, COVID-19 symptoms, comorbidities, co-infections, and COVID-19 risk factors were collected at admission and during hospitalization, and discharge outcomes were collected upon discharge by pediatricians at the respective hospitals [4]. From March 2020 to November 2022, 6983 participants were included in the registry.

2.3.1.2 Variables for phenotype identification in children and adolescents

Variables including sex, COVID-19 symptoms, comorbidities, coinfection, and COVID-19 risk factors were used to identify phenotypes in children and adolescents. A description of the variables is reported in publication 1 [41].

2.3.1.3 Outcome variables for prognostic evaluation of clinical phenotypes

In publication 1, two main outcomes were examined. The primary outcome was discharge status, which was classified into three categories: unfavorable prognosis (which consists of identified irreversible impairment at discharge from hospital and mortality due to SARS-CoV-2), residual symptoms, and full recovery. The secondary outcome was ICU admission.

2.3.1.4 Statistical methods

In order to determine phenotypes in children and adolescents, hierarchical agglomerative clustering was applied on the above predefined variables in section 2.3.1.2. This method was chosen since it does not predefine the number of phenotypes. This method uses a "bottom-up" approach, starting with each patient in its own cluster, then merging a pair of patients most similar to each as one moves up the hierarchy, and finally stopping until all patients in one cluster [42]. Thirty statistical indices were employed to identify the most appropriate number of clusters, in combination with the evaluation of the clinical explanation by experienced pediatricians. Following the identification of the clinical phenotypes, unadjusted summary statistics including absolute and relative frequencies of the clustering variables were presented stratified by clinical phenotypes. Additionally, for prognosis evaluation, we applied a binary logistic regression to assess their correlation with ICU admission and multinomial logistic regression to evaluate their relationships with discharge status. Odds ratios (ORs) with confidence intervals (CIs) were estimated in these models. Phenotype A was chosen as the reference phenotype, admission to a standard ward as the reference ICU status, and full recovery as the reference discharge status.

2.3.2 Main results

2.3.2.1 Clinical phenotypes during acute SARS-CoV-2 infection among children and adolescents

Six distinct phenotypes were identified among children and adolescents infected with SARS-CoV-2, each differing regarding symptomatology, co-infections, comorbidities, and SARS-CoV-2 risk factors. Each phenotype reveals varying clinical patterns and associated risk factors, emphasizing the heterogeneity of SARS-CoV-2 manifestations in pediatric populations. A detailed description of phenotypes is shown in **Table 1**.

Table 1 Main characteristics of hospitalized children and adolescents with SARS-CoV-2 infection grouped into six distinct clinical phenotypes (hierarchical agglomerative clustering)

Phenotypes	Main characteristics
Phenotype A	Patients presented with diverse symptoms and cannot be characterized by one single typical symptom. This group had higher frequencies of non-pulmonary bacterial infections (9.2%), preterm birth history (9.9%), and smoking exposure (6.2%) compared to other phenotypes.
Phenotype B	Patients were predominantly characterized by gastrointestinal symptoms (95.9%). These patients also had non-pulmonary viral coinfections more frequently (12.5%).
Phenotype C	Patients were mostly asymptomatic (95.9%). This group showed an increased frequency of non-pulmonary bacterial infections (7.1%) and were more likely to have received immunosuppressive treatment prior to the current illness (4.1%).
Phenotype D	The majority of patients had symptoms related to the lower respiratory tract (49.8%) and had a higher prevalence of comorbidities. They were more likely to have pulmonary bacterial infections (3.3%), demonstrate SARS-CoV-2 risk factors (40.6%, including maternal SARS-CoV-2 positivity for newborn patients), and need home oxygen or ventilation therapy before the present hospitalization (8.7%).
Phenotype E	This group showed a combination of lower respiratory tract symptoms (86.2%) and ear, nose, and throat (ENT) symptoms (41.7%). They also had higher rates of pulmonary viral infections (6.8%) and pulmonary bacterial infections (1.9%).
Phenotype F	Patients had predominantly neurological disease, with 99.2% of patients exhibiting neurological symptoms.

2.3.2.2 Prognostic implications of clinical phenotypes

In terms of discharge status, compared to phenotype A, children and adolescents with phenotypes D and E exhibited the highest odds of experiencing residual symptoms upon discharge from hospital, with odds ratios of 1.33 (95% CI: 1.11–1.59) and 1.91 (95% CI: 1.65–2.21), respectively. Additionally, children and adolescents with phenotype D were more likely to have an unfavorable prognosis, with an odds ratio of 4.00 (95% CI: 1.95–8.19). Regarding ICU admission, individuals

with phenotype D demonstrated higher likelihood of ICU care than staying in a standard ward, with an odds ratio of 4.26 (95% CI: 3.06–5.98) compared to those with phenotype A.

2.4 Methods and results for publication 2

2.4.1 Methods

2.4.1.1 Data source

Data from the NAPKON ("Nationales Pandemie Kohorten Netz", the National Pandemic Cohort Study Network) cohort were used to answer scientific questions C, D and E.

NAPKON is a nationwide cohort which was created in 2020, aiming to be the most comprehensive clinical-epidemiological COVID-19 cohort in Germany [43]. COVIDOM/NAPKON-POP is one of the three cohort platforms within NAPKON; it is a population-based platform, focusing on investigating the long-term health effects in SARS-CoV-2-infected adults. It includes adults at least six months after a positive SARS-CoV-2 polymerase chain reaction (PCR) test [44]. Publication 2 project focused on the baseline assessment of the recruited patients, with data collected by pre-onsite and onsite questionnaires, physical examination, and interviews [44]. Between November 2020 and September 2021, 1441 participants were included for the baseline assessment.

2.4.1.2 Variables for time to symptom-free prediction in adults

Participants were asked whether they had the following 22 specific SARS-CoV-2-related symptoms and "other symptoms" during acute infection and during the baseline assessment. 22 symptoms included smell disorders/anosmia, taste disorders/ageusia, abdominal pain, disturbances of consciousness/confusion, hoarseness, dizziness, fever, runny nose, cough, limb pain, shortness of breath/dyspnea, muscle pain, wheezing or wheezing breathing, nausea, skin rash, hair loss, diarrhea, vomiting, headache, chest pain, sore throat/scratching, and chills. If "other symptoms" was selected, the patients were asked to give a specific free-text answer. We consider the fatigue symptom as present if it contained "fatigue" or its synonyms in this free-text answer.

Symptom burden during the acute phase were classified as 1-5 symptoms and ≥ 6 symptoms to predict prolonged time to symptom-free. Other variables used to predict prolonged time to symptom-free included: sex, age, education status, body mass index (BMI), resilience, alcohol drinking, smoking, acute disease treatment, comorbidities, and whether living with a partner. Resilience was assessed by the brief resilience scale (BRS) [45]. BMI was computed using weight and height measurements obtained during on-site baseline assessment, using the equation: BMI = kg/m^2.

2.4.1.3 Outcome variables for prediction of prolonged time to symptom-free

Self-report using the following question was used to measure the time to symptom-free: "How long did it take for you to become symptom-free after your first symptoms?" [46] In this PhD project,

symptom-free time was defined as the number of days from the occurrence of the first symptom to symptom-free.

2.4.1.4 Statistical methods

Unadjusted percentages were used to compare the differences among symptoms during acute infection and symptoms at least six months after infection. To show the difference in time to symptom-free among different subgroups, we used Kaplan-Meier curves to visualize differences between these subgroups, and used log-rank tests to test them. To estimate adjusted hazard ratios (aHRs) and CIs of variables associated with delayed time to recovery, we used a stratified Cox proportional hazard model. Stratified cox regression model was used since variable “symptom burden” did not conform to the PH assumption; no coefficient was estimated for variable “symptom burden” itself in this model, but hazard ratios for other remaining variables were estimated, by means of setting different baseline hazard for each stratum of “symptom burden” [47]. An aHR smaller than one indicated a prolonged time to becoming symptom-free.

2.4.2 Main results

2.4.2.1 Prevalence of symptoms during acute infection and nine months post infection

Although the study protocol aimed to recruit participants over six months post SARS-CoV-2 infection, participants included in this study had a mean observation time of about 9 months (280 days) since the onset of infection. Patient flow in publication 2 is shown in **Figure 1.** After excluding one patient with an implausible PCR test date and 90 patients whose observation time was less than six months, 1350 patients were included for the description of symptom prevalence.

During acute infection, we observed 23 SARS-CoV-2-related symptoms during acute infection; nine months after infection, 22 of them remained persistent at the time of baseline examination, except for vomiting. The most frequently reported symptoms included anosmia/smell disorder (19.3%), dyspnea/shortness of breath (18.9%), fatigue (14.1%), as well as ageusia/taste disorder (13.8%) nine months after infection. Additionally, each of the following symptoms was reported by more than 5% of participants: cough (6.0%), chest pain (6.4%), disturbances of consciousness/confusion (6.6%), dizziness (9.0%), limb pain (9.5%), headache (10.2%), and muscle pain (10.6%). In contrast, fewer than 5% participants continued to report fever, chills, sore throat, and a runny nose at the time of the baseline examination, whereas many had experienced these symptoms during the acute infection phase.

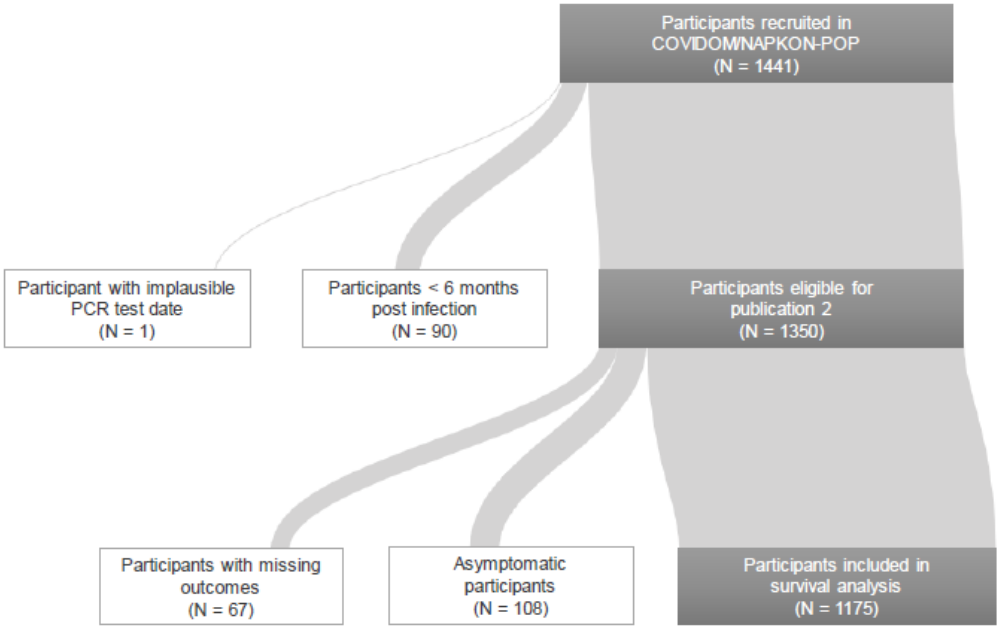


Figure 1 Patient flow in publication 2. This figure is adapted from the study profile in publication 2 [46].

2.4.2.2 Time to full recovery after SARS-CoV-2 Infection

After excluding 108 patients who were asymptomatic and another 67 patients whose outcomes (time to symptom-free or current symptom status) were missing, 1175 patients were included for survival analysis. COVID-19-related symptom resolution occurred in 25% of the study participants in the first 18 days (Interquartile range: 14–21 days). By 28 days post-symptom onset, 34.5% participants had become symptom-free. After this point, symptom resolution slowed significantly. At nine months following the acute infection, 54.1% of participants continued to report COVID-19-related symptoms.

2.4.2.3 Factors correlated with prolonged time to complete recovery following SARS-CoV-2 infection

The adjusted hazard ratios with 95% CI for the variables included in the final stratified Cox proportional hazard model are shown in **Figure 2**. Compared to those under 49 years, patients between 49 and 59 years old were 30% less likely to become symptom-free, while no significant difference was observed for patients who were 60 years or older. Women, individuals with lower educational levels, and those living with a partner also experienced prolonged recovery time. We also found that participants with low resilience had a 35% lower possibility to recover. Additionally, treatment during the acute infection phase also had an impact: steroid treatment and absence of medication were associated with delayed symptom resolution. Pre-existing comorbidities including chronic rheumatologic/immunologic disease, chronic neurological disease, and chronic liver disease were not independent risk factors predicting prolonged recovery.

Characteristics	aHR (95% CI)	P value
Age		
<49	Reference	
49−59	0.70 [0.56; 0.87]	0.0013
≥60	0.92 [0.72; 1.17]	0.4857
Sex		
Male	Reference	
Female	0.78 [0.65; 0.93]	0.0073
Educational level		
University entrance certificate	Reference	
Lower education	0.77 [0.64; 0.93]	0.0062
Living status		
No partner/Not living with a partner	Reference	
Living with a partner	0.81 [0.66; 0.99]	0.0382
Alcohol consumption		
Abstainer	Reference	
Low−risk alcohol consumption	1.31 [0.94; 1.81]	0.1102
Risky alcohol consumption	1.41 [0.98; 2.04]	0.0687
Body Mass Index		
Normal	Reference	
Underweight	1.40 [0.61; 3.17]	0.4259
Pre−obese	1.04 [0.85; 1.27]	0.7237
Obese	0.80 [0.63; 1.03]	0.0826
Resilience		
High resilience	Reference	
Normal resilience	0.83 [0.65; 1.05]	0.1281
Low resilience	0.65 [0.47; 0.90]	0.0090
Treated with medication		
Yes	Reference	
No	0.74 [0.62; 0.89]	0.0013
Steroid treatment		
No	Reference	
Yes	0.22 [0.05; 0.90]	0.0357
Chronic liver disease		
No	Reference	
Yes	0.81 [0.58; 1.15]	0.2385
Chronic rheumatologic/immunologic disease		
No	Reference	
Yes	0.71 [0.50; 1.00]	0.0512
Chronic neurological disease		
No	Reference	
Yes	0.80 [0.64; 1.00]	0.0522

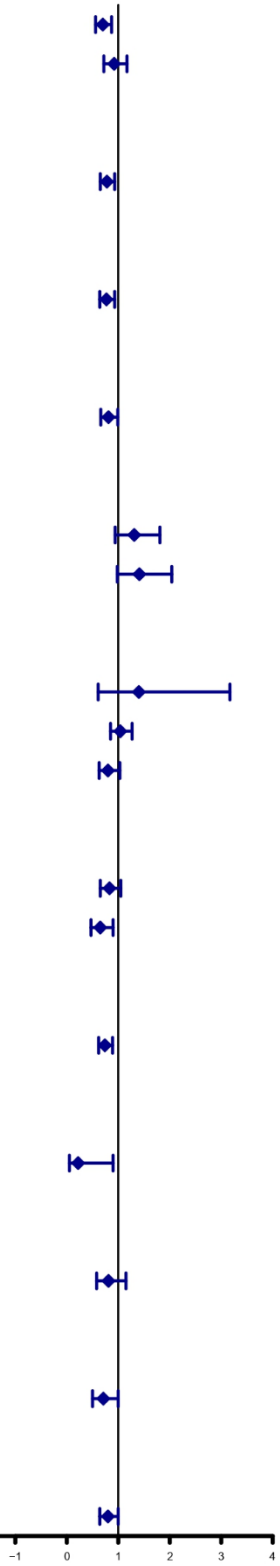


Figure 2 A forest plot of the adjusted hazard ratios with 95% confidence intervals for co-variates in the stratified Cox proportional hazard model.

2.5 Strengths and limitations

This doctoral thesis features several notable strengths: the application of large-scale cohort data, robustness of the study findings, methodological rigor, and a focus on both pediatric and adult populations.

Foremost among these is the large-scale cohort data used in this thesis. We were able to utilize two well-established SARS-CoV-2 cohorts in Germany for this dissertation and therefore could rely on well-curated prospectively collected data for our analysis. The DGPI registry is a unique resource of data from children and adolescents who were confirmed to be infected with SARS-CoV-2. Likewise, the COVIDOM/NAPKON-POP cohort includes data from adults across the whole spectrum of disease. Unlike other COVID-19 cohorts, COVIDOM/NAPKON-POP also included patients identified by local health authorities through positive PCR tests, giving patients the chance to be included even if they did not attend a healthcare provider.

Extensive sensitivity analyses added to the validity of our results. In publication 1, to investigate whether the clinical phenotypes differ in different age groups, we also applied the same set of hierarchical agglomerative clustering and prognosis evaluation to the groups of infants (those under one year old) and to non-infants (those older than or equal to one year old) and showed that clinical phenotypes in infants and non-infants were almost identical except that infant did not exhibit the phenotype characterized by neurological symptoms. In publication 2, since symptom burden and hospitalization are closed related to “disease severity”, we did not include hospitalization in the main model, instead, we explored whether hospitalization had an impact on time to recovery from a SARS-CoV-2 infection in sensitivity analyses and showed that only during the initial four weeks was hospitalization correlated with delayed recovery.

Another major strength is methodological rigor across the two publications. In publication 1 we were able to apply novel methodologies to identify and evaluate the clinical phenotypes among SARS-CoV-2-infected children and adolescents. When determining the most appropriate number of phenotypes, we not only considered the statistically optimal choice but also carefully evaluated their clinical relevance in collaboration with experienced pediatricians. The final phenotypes were selected based on their applicability in clinical practice. In publication 2, we thoroughly reviewed the original NAPKON data, and excluded one participant due to an implausible PCR test date along with 90 cases where the time between the PCR test and the survey was less than six months; and we checked the PH assumption of each predictor and applied stratified Cox proportional hazard model to identify patient characteristics correlated with delayed recovery, by stratifying on “symptom burden” which did not conform to the PH assumption. Furthermore, we handled missing data with Random Forest (RF) in publication 1 and with Multiple Imputation by Chained Equations (MICE) in publication 2, which both provided robust imputations and enhanced the reliability of our study results.

Furthermore, by including both pediatric and adult populations and examining the course from the acute infection phase to nine months post-infection, we provided a comprehensive assessment of disease progression and short- and long-term outcomes. This approach allows for a better

understanding of age-specific differences in recovery patterns and risk factors, ultimately informing more targeted clinical management and public health strategies.

There are also some noteworthy limitations. Firstly, due to the self-reported nature of the study outcomes in publication 2, the results might be subjected to recall bias. However, it is also possible that even after they recover, patients still have a clear memory of the time course. Secondly, extrapolating the study results to the entire infected population should be done with caution. In publication 1, we included only the hospitalized children and adolescents and the sample might not be representative of the whole infected children and adolescent population. In publication 2, we are confident that this sample accurately represents the infected population during the study period in the respective regions. However, the varying responses among infected patients may have introduced bias in estimating the prevalence and persistence of symptoms, so the results may not be transferable to the whole infected adult population. Thirdly, further studies are still needed based on the exploratory research of this doctoral thesis. Although the clinical phenotypes acquired in publication 1 were robust for the current population, they were not externally validated; the estimation for publication 2 mainly included patients with the SARS-CoV-2 wild type and alpha type variant, thus, the estimation of the prevalence of PCS and time to recovery might not be transferable to recent Omicron variant and other variants. To improve generalizability, future studies should validate the clinical phenotypes in independent cohorts and conduct longitudinal studies comparing the prevalence of PCS and the recovery pattern after SARS-CoV-2 infection across different variants.

2.6 Contribution of this doctoral thesis and outlook

This doctoral thesis provides insights into patient stratification and the immediate and over-time health effects of SARS-CoV-2 infection by examining two different time frames: the acute infection phase and the period from infection to nine months post-infection. It also provides a life course perspective by investigating outcomes in children, adolescents, and adults following the initial infection with SARS-CoV-2.

This thesis uncovered six clinical phenotypes in SARS-CoV-2-infected children and adolescents, primarily distinguished by symptomatology: similar symptom presentation as the whole DGPI registration (phenotype A), gastrointestinal symptoms (phenotype B), asymptomatic symptoms (phenotype C), symptoms related to lower respiratory tract (phenotype D), symptoms related to both lower respiratory tract and ENT (phenotype E), as well as neurological symptoms (phenotype F). By identifying these clinical phenotypes, this thesis provides an innovative approach to understand the heterogeneity of the clinical presentations in this population. These phenotypes highlight differences in symptomatology, comorbidities, co-infection, and risk factors, offering clinicians a valuable framework for prognosis and personalized treatment strategies. The thesis addresses a significant gap in knowledge by applying novel methodologies to analyze pediatric data, which had previously been limited compared to adult studies.

The identified clinical phenotypes were evaluated for their associations with ICU admission and discharge outcomes. We found out one phenotype which were typically characterized by lower respiratory tract symptoms, pre-existing comorbidities, and other SARS-CoV-2 risk factors had high possibility of ICU admission, having residual symptoms, and developing unfavorable prognosis. This provides actionable insights for healthcare providers to anticipate the short-term health consequences of pediatric SARS-CoV-2 cases and allocate resources effectively.

Using the COVIDOM/NAPKON-POP cohort, this thesis reveals the prevalence of symptoms was over 50% up to nine months after infection, and also reveals the pattern of symptom resolution in SARS-CoV-2-infected adults. It highlights the PCS burden and identifies key factors associated with prolonged recovery, such as working-age population, female sex, lower educational level, low resilience, as well as steroid treatment or lack of treatment during acute infection. The findings emphasize the complex interplay of clinical and epidemiological factors in PCS recovery and highlight the need for tailored interventions. We found out that although PCS recovery were primarily influenced by factors that are difficult to change, high resilience which is normally considered to be associated with positive mental health [48], it also helps to relieve self-reported symptoms. Future interventional measures for PCS may incorporate resilience development as a component.

In conclusion, this doctoral thesis revealed six clinical phenotypes in pediatric population and identified one phenotype typical of lower respiratory tract symptoms and pre-existing comorbidities to be associated with high risk of severe disease and an unfavorable prognosis. Additionally, this research provided estimates for the time to recovery after SARS-CoV-2 infection among adults and determined that resilience could serve as a potential interventional factor for recovery. These findings contribute to more informed public health strategies and the development of refined care protocols tailored for individuals affected by adverse short- and long-term health consequences.

3. Publication I

SIX CLINICAL PHENOTYPES WITH PROGNOSTIC IMPLICATIONS WERE IDENTIFIED BY UNSUPERVISED MACHINE LEARNING IN CHILDREN AND ADOLESCENTS WITH SARS-COV-2 INFECTION: RESULTS FROM A GERMAN NATIONWIDE REGISTRY

Yanyan Shi, Ralf Strobl, Reinhard Berner, Jakob Armann, Simone Scheithauer, Eva Grill

Respiratory Research

Received 17 July 2024

Accepted 20 October 2024

Published 30 October 2024

DOI: 10.1186/s12931-024-03018-3

Complete Citation

Shi, Y., Strobl, R., Berner, R. *et al.* Six clinical phenotypes with prognostic implications were identified by unsupervised machine learning in children and adolescents with SARS-CoV-2 infection: results from a German nationwide registry. *Respiratory Research* **25**, 392 (2024). https://doi.org/10.1186/s12931-024-03018-3

Shi *et al. Respiratory Research* (2024) 25:392
https://doi.org/10.1186/s12931-024-03018-3

Respiratory Research

RESEARCH Open Access

Six clinical phenotypes with prognostic implications were identified by unsupervised machine learning in children and adolescents with SARS-CoV-2 infection: results from a German nationwide registry

Yanyan Shi[1,2], Ralf Strobl[1,3], Reinhard Berner[4], Jakob Armann[4], Simone Scheithauer[5] and Eva Grill[1,3*]

Abstract

Objective Phenotypes are important for patient classification, disease prognostication, and treatment customization. We aimed to identify distinct clinical phenotypes of children and adolescents hospitalized with SARS-CoV-2 infection, and to evaluate their prognostic differences.

Methods The German Society of Pediatric Infectious Diseases (DGPI) registry is a nationwide, prospective registry for children and adolescents hospitalized with a SARS-CoV-2 infection in Germany. We applied hierarchical clustering for phenotype identification with variables including sex, SARS-CoV-2-related symptoms on admission, pre-existing comorbidities, clinically relevant coinfection, and SARS-CoV-2 risk factors. Outcomes of this study were: discharge status and ICU admission. Discharge status was categorized as: full recovery, residual symptoms, and unfavorable prognosis (including consequential damage that has already been identified as potentially irreversible at the time of discharge and SARS-CoV-2-related death). After acquiring the phenotypes, we evaluated their correlation with discharge status by multinomial logistic regression model, and correlation with ICU admission by binary logistic regression model. We conducted an analogous subgroup analysis for those aged < 1 year (infants) and those aged ⩾ 1 year (non-infants).

Results The DGPI registry enrolled 6983 patients, through which we identified six distinct phenotypes for children and adolescents with SARS-CoV-2 which can be characterized by their symptom pattern: phenotype A had a range of symptoms, while predominant symptoms of patients with other phenotypes were gastrointestinal (95.9%, B), asymptomatic (95.9%, C), lower respiratory tract (49.8%, D), lower respiratory tract and ear, nose and throat (86.2% and 41.7%, E), and neurological (99.2%, F). Regarding discharge status, patients with D and E phenotype had the highest odds of having residual symptoms (OR: 1.33 [1.11, 1.59] and 1.91 [1.65, 2.21], respectively) and patients with phenotype D were significantly more likely (OR: 4.00 [1.95, 8.19]) to have an unfavorable prognosis. Regarding ICU,

*Correspondence:
Eva Grill
Eva.Grill@med.uni-muenchen.de

Full list of author information is available at the end of the article

patients with phenotype D had higher possibility of ICU admission than staying in normal ward (OR: 4.26 [3.06, 5.98]), compared to patients with phenotype A. The outcomes observed in the infants and non-infants closely resembled those of the entire registered population, except infants did not exhibit typical neurological/neuromuscular phenotypes.

Conclusions Phenotypes enable pediatric patient stratification by risk and thus assist in personalized patient care. Our findings in SARS-CoV-2-infected population might also be transferable to other infectious diseases.

Keywords SARS-CoV-2, Clinical phenotype, Clustering, Machine learning, Prognosis

Introduction

Children and adolescents generally experience mild disease and a better prognosis after severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection compared with adults [1, 2]. In some cases, however, severe disease and mortality do occur in the pediatric population as well [3–6]. In Germany, severe disease courses as defined by intensive care unit (ICU) admission occurred in 0.02% of SARS-CoV-2 infections and fatality occurred less than 0.001% during the wild type and the alpha variant [7]. Early stratification of risk groups in order to identify those at highest risk could be beneficial in most appropriate patient care for children and adolescents with SARS-CoV-2 infection.

A promising approach to enhance the patient management of children with SARS-CoV-2 infection involves the identification of distinctive clinical phenotypes, ideally at the time of hospital admission. Phenotypes reveal how the population can be categorized into homogeneous subgroups with distinct clinical features [8]. In addition to description, phenotypes are important for patient classification, disease prognostication, and treatment customization [8, 9]. Methodologically, clustering is a commonly used unsupervised machine learning method, with which hidden objects, patterns, and groupings were found from untagged data [10]. This approach differs from studies focusing on identifying outcome predictors, which assess the independent predictive association of each variable with the outcome [11]. Clustering has previously been employed in the context of disease phenotyping, such as sepsis [12]. Since the appearance of SARS-CoV-2, it was also applied in identifying the clinical phenotypes of COVID-19 [8, 11, 13]. This approach would allow for the tailoring of standard treatment protocols to accommodate the unique requirements associated with each identified phenotype. While this strategy has been proven effective in optimizing treatment for adults with SARS-CoV-2 infection [14], its application in the context of pediatric patients remains to be investigated.

Identification of phenotypes has been utilized in pediatric patients to differentiate severe COVID-19 cases from mild cases and cases with multisystem inflammatory syndrome in Children (MIS-C), also called pediatric multisystem inflammatory syndrome (PIMS), thus enabling more precise treatment according to phenotypes [15]. Our study wanted to adapt this strategy to identify clinical phenotypes with a focus on children and adolescents who tested positive for SARS-CoV-2. Even in the generally low-risk pediatric population, we hypothesized that certain clinical phenotypes representing patient characteristics do exist and that they differ regarding disease severity and an unfavorable prognosis including mortality.

Using data from a German nationwide pediatric registry, we aimed to identify distinct clinical phenotypes of children and adolescents with SARS-CoV-2 infection by clustering, and to assess how the phenotypes differ with regard to disease severity and outcome at discharge.

Methods

Data sources

DGPI registry, initiated by the German Society of Pediatric Infectious Diseases (DGPI), is a nationwide, prospective registry for children and adolescents hospitalized with a SARS-CoV-2 infection in Germany. It included patients with laboratory-confirmed SARS-CoV-2 infections who were admitted to pediatric departments and hospitals. A SARS-CoV-2 infection was confirmed if either a reverse transcriptase polymerase chain reaction (RT-PCR) test or, if a nucleic acid based test was not available, an antigen detection rapid diagnostic test (Ag-RDT) for SARS-CoV-2 was reported positive [16]. Details of the DGPI registry has been published before [7, 16]. This registry was approved by the Ethics Committee of the Technische Universität (TU) Dresden (BO-EK-110032020) [16]. Data of patients reported to the DGPI registry from March 2020 to November 2022 were used in the present analysis. This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guideline [17].

Variables for defining phenotypes

We included variables reported to be associated with severe disease and mortality in children and adolescents with SARS-CoV-2 infection [18–28]. Variables being used for defining phenotypes (Table 1) included sex, SARS-CoV-2-related symptoms at admission, comorbidities at the time of SARS-CoV-2 infection, clinically

Table 1 Description of variables used for defining phenotypes

Domain	Variable	Variable Definition
Demographics	Sex	Sex
Coinfection	Pulmonary viral coinfection	Clinically relevant coinfection with other respiratory viruses
	Pulmonary bacterial coinfection	Clinically relevant pulmonary coinfection with bacterial infectious agents
	Non-pulmonary bacterial coinfection	Clinically relevant non-pulmonary coinfection with bacterial infectious agents
	Non-pulmonary viral coinfection	Clinically relevant non-pulmonary coinfection with viral infectious agents
COVID-19 symptoms on admission	No symptoms (asymptomatic)	No symptoms on admission which were deemed COVID-19 related by corresponding pediatricians of the patients.
	Fever or general symptoms	Fever > 38° Celsius, chills, fatigue, feeling of weakness, weakness to drink / refusal to eat, syncope, dizziness, and other symptoms
	Ear, nose, and throat symptoms	Loss of smell, loss of taste, runny nose, stuffy nose, and other symptoms
	Lower respiratory tract symptoms	Dry cough, productive cough, hemoptysis, retractions of the chest during inhalation, wheezing, tachypnea, shortness of breath, and other symptoms
	Cardiovascular symptoms	Arrhythmias, edema, tachycardia, chest pain, and other symptoms
	Gastrointestinal symptoms	Abdominal pain, diarrhea, nausea, vomiting, loss of appetite, and other symptoms
	Liver symptoms	Jaundice and other symptoms
	Neurological/ neuromuscular symptoms	Disturbance of consciousness or clouding, headache, meningismus, seizure, and other symptoms
	Musculoskeletal Symptoms	Joint pain, muscle pain, inability to walk, and other symptoms
	Other symptoms on admission	Other symptoms on admission which were deemed COVID-19 related by corresponding pediatricians of the patients.
	Unknown admission symptoms	
Comorbidities at the time of SARS-CoV-2 infection	Respiratory disease	Physician diagnosed comorbidities at the time of the current SARS-CoV-2 infection
	Cardiovascular disease	
	Gastrointestinal tract disease	
	Liver disease	
	Kidney disease	
	Neurological/ neuromuscular disease	
	Psychiatric disease	
	Hematologic disease	
	Oncological disease	
	Organ or bone marrow/stem cell transplantation	
	Autoimmunological disease	
	Congenital immunodeficiency	
	Tracheostoma (prior to COVID-19 infection)	
	Other concomitant disease	
COVID-19 risk factors	Home oxygen or ventilation therapy	Oxygen or ventilation therapy before the current SARS-CoV-2 infection
	Preterm birth	The patient was born prematurely
	Exposure to smoking	Both smoking patient and smoking household member were considered to have exposure to smoking
	Immunosuppression	Immunosuppressive medication
	Other COVID-19 risk factors	Other COVID-19 related risk factors (including the newborn patient's mother was SARS-CoV-2 positive, etc.)

Note: Variable type: sex was binary (male/female), and all other variables were binary (yes/no)

relevant bacterial or viral coinfection as decided by the pediatricians (Additional file 1, Table S1) at the time of SARS-CoV-2 infection, home oxygen or ventilation therapy prior to the current illness, preterm birth (regardless of the current age), exposure to smoking, immunosuppression, and other SARS-CoV-2 risk factors (Additional file 1, Table S2).

Outcome variable

Discharge status was the primary outcome of this study. Each patient was assessed at discharge by the pediatricians and was categorized with regard to the admission with a positive SARS-CoV-2 test as: (1) restitutio ad integrum (hereinafter referred to as "full recovery" for easier understanding); (2) residual symptoms that can be considered reversible in the further course of the disease; (3) irreversible consequential damage that has already been identified as potentially irreversible at the time of discharge, such as respiratory failure, heart failure, arrhythmia, renal failure, epilepsy, personality disorder, etc.; (4) transferal to other health facilities; and (5) death, including SARS-CoV-2-related death and non-SARS-CoV-2-related death. In the present analysis, we combined "irreversible consequential damage" and "SARS-CoV-2-related death" as "unfavorable prognosis" due to low case count. Also, patients who were transferred to other health facilities or had non-SARS-CoV-2-related were excluded from further outcome evaluation. Thus, final discharge status was categorized as the three classes: "full recovery", "residual symptoms", and "unfavorable prognosis". ICU stay was the secondary outcome of this study, representing severe disease of COVID-19. It was a binary outcome.

Statistical analysis

We reported median and interquartile range (IQR) for continuous variables and absolute and relative frequencies for categorical variables.

Missing values

We assumed the missing values in our dataset were not missing completely at random (MCAR), and checked this assumption by the Little's MCAR test (R package *naniar*) [29]. The missing information of each variable recorded in the dataset is shown in the footnote of Table 2. Missing values for binary health condition questions (answer: yes/no) were imputed with "no" when physicians skipped the question. The rationale was that non-response indicated a lack of this health condition; this method was also used before [30]. Furthermore, we used random forest (R package *randomForest*) to impute missing values in the variable "sex" as proposed by Breiman [31]. The algorithm starts by imputing missing values with the mode. A Random Forest is fit with this completed data and then used to determine a proximity matrix which is used to update the imputation. The imputed value is the category with the largest average proximity.

Identifying phenotypes

Variables used for defining phenotypes were described in the above section "*Variables for defining phenotypes*". We applied hierarchical agglomerative clustering for phenotype identification in the present study, which does not predefine the number of phenotypes. Hierarchical clustering algorithm initially regards each patient as a single cluster and then gradually merges patients most similar to each to new clusters. This process continues until all patients belong to a single cluster. Similarity was computed using Gower's distance which ranges from 0 to 1, with 0 representing perfect similarity and 1 representing maximum difference [32]. The ongoing merging of clusters was done with respect to minimizing the total within-cluster variance, referred to as Ward´s method [33]. We chose the optimal number of clusters by clinical explanation and the *NbClust* package in R, which offers 30 indices to help decide suitable clustering approach [34]. Hierarchical clustering is usually visualized by dendrogram showing the merging path of each patient (Additional file 1, Figure S1). R (version 4.1.2) and the *cluster* [35] package were used for statistical analysis.

Prognosis of participants with different phenotypes

We included the entire registered population for comprehensive phenotype identification. Subsequently, we conducted prognostic assessments exclusively on patients with relevant outcomes.

We excluded patients who were transferred to other health facilities and those who died from causes unrelated to SARS-CoV-2 infection because these discharge reasons cannot be considered as unfavorable prognosis regarding a SARS-CoV-2 infection. We used a multinomial logistic regression model to evaluate the associations between distinct phenotypes and discharge status. Since no patient with phenotype B had an unfavorable prognosis, we used two methods of handling phenotype B. For the main model, we excluded patients with phenotype B, and evaluated the associations between other phenotypes and discharge status (including full recovery, residual symptoms, and unfavorable prognosis) in the model. Phenotype A was used as the reference phenotype due to large percentage in the total sample and similarity of symptom pattern to the total sample; full recovery was used as the reference discharge status. Age was included in the model as a confounder. Odds ratios greater than one indicate higher possibility of having residual symptoms or having unfavorable prognosis than achieving full recovery, compared with phenotype A. To investigate the effect of phenotype B, we kept patients with phenotype B

Table 2 Characteristics of participants of DGPI registry by discharge status

Variable	All (*n* = 6983)	Discharge Status			
		Full Recovery (*n* = 5352)	Residual Symptoms (*n* = 1526)	Unfavorable Prognosis (*n* = 42)	Transferal/ Non-SARS-CoV-2-related Death (*n* = 63)
Age (years)	1 (0,9)	1 (0,8)	2 (0,11)	7 (3,11.7)	9 (3, 13)
Sex = Female	3236 (46.3)	2465 (46.1)	717 (47.0)	22 (52.4)	32 (50.8)
No symptoms (asymptomatic)	702 (10.1)	677 (12.6)	12 (0.8)	3 (7.1)	10 (15.9)
Fever or general symptoms	4818 (69.0)	3620 (67.6)	1145 (75.0)	24 (57.1)	29 (46.0)
Ear, nose, and throat symptoms	1627 (23.3)	1065 (19.9)	550 (36.0)	7 (16.7)	5 (7.9)
Lower respiratory tract symptoms	2286 (32.7)	1465 (27.4)	770 (50.5)	27 (64.3)	24 (38.1)
Cardiovascular symptoms	226 (3.2)	151 (2.8)	60 (3.9)	9 (21.4)	6 (9.5)
Gastrointestinal symptoms	1884 (27.0)	1440 (26.9)	415 (27.2)	8 (19.0)	21 (33.3)
Liver symptoms	29 (0.4)	20 (0.4)	7 (0.5)	0 (0.0)	2 (3.2)
Neurological/ neuromuscular symptoms	1056 (15.1)	791 (14.8)	240 (15.7)	11 (26.2)	14 (22.2)
Musculoskeletal Symptoms	200 (2.9)	131 (2.4)	69 (4.5)	0 (0.0)	0 (0.0)
Other symptoms on admission	422 (6.0)	292 (5.5)	116 (7.6)	5 (11.9)	9 (14.3)
Unknown admission symptoms	60 (0.9)	43 (0.8)	16 (1.0)	0 (0.0)	1 (1.6)
Respiratory disease	295 (4.2)	191 (3.6)	82 (5.4)	8 (19.0)	14 (22.2)
Cardiovascular disease	261 (3.7)	184 (3.4)	51 (3.3)	11 (26.2)	15 (23.8)
Gastrointestinal tract disease	193 (2.8)	148 (2.8)	34 (2.2)	7 (16.7)	4 (6.3)
Liver disease	65 (0.9)	51 (1.0)	11 (0.7)	1 (2.4)	2 (3.2)
Kidney disease	145 (2.1)	116 (2.2)	21 (1.4)	2 (4.8)	6 (9.5)
Neurological/ neuromuscular disease	445 (6.4)	307 (5.7)	98 (6.4)	20 (47.6)	20 (31.7)
Psychiatric disease	111 (1.6)	87 (1.6)	17 (1.1)	2 (4.8)	5 (7.9)
Hematologic disease	155 (2.2)	118 (2.2)	30 (2.0)	1 (2.4)	6 (9.5)
Oncological disease	106 (1.5)	97 (1.8)	6 (0.4)	0 (0.0)	3 (4.8)
Organ or bone marrow/stem cell transplantation	39 (0.6)	33 (0.6)	5 (0.3)	0 (0.0)	1 (1.6)
Autoimmunological disease	136 (1.9)	106 (2.0)	21 (1.4)	3 (7.1)	6 (9.5)
Congenital immunodeficiency	28 (0.4)	18 (0.3)	8 (0.5)	1 (2.4)	1 (1.6)
Tracheostoma (prior to current infection)	18 (0.3)	13 (0.2)	5 (0.3)	0 (0.0)	0 (0.0)
Other concomitant disease	965 (13.8)	684 (12.8)	243 (15.9)	17 (40.5)	21 (33.3)
Pulmonary viral coinfection	131 (1.9)	72 (1.3)	56 (3.7)	3 (7.1)	0 (0.0)
Pulmonary bacterial coinfection	81 (1.2)	38 (0.7)	27 (1.8)	6 (14.3)	10 (15.9)
Non-pulmonary bacterial coinfection	331 (4.7)	260 (4.9)	54 (3.5)	8 (19.0)	9 (14.3)
Non-pulmonary viral coinfection	136 (1.9)	103 (1.9)	30 (2.0)	1 (2.4)	2 (3.2)
Home oxygen or ventilation therapy	111 (1.6)	58 (1.1)	39 (2.6)	7 (16.7)	7 (11.1)
Preterm birth	357 (5.1)	270 (5.0)	78 (5.1)	6 (14.3)	3 (4.8)
Exposure to smoking	226 (3.2)	149 (2.8)	70 (4.6)	1 (2.4)	6 (9.5)
Immunsuppression	149 (2.1)	124 (2.3)	18 (1.2)	2 (4.8)	5 (7.9)
Other COVID-19 risk factors	465 (6.7)	313 (5.8)	131 (8.6)	9 (21.4)	12 (19.0)
Intensive Care Unit stay	214 (3.1)	107 (2.0)	63 (4.1)	25 (59.5)	19 (30.2)

Note: Number of missing in the above variables: sex (3), respiratory disease (5449), cardiovascular disease (5205), gastrointestinal tract disease (5453), liver disease (5464), kidney disease (5468), neurological/ neuromuscular disease (5443), psychiatric disease (5526), hematologic disease (5462), oncological disease (5466), organ or bone marrow/stem cell transplantation (5472), autoimmunological disease (5472), congenital immunodeficiency (5482), tracheostoma (5476), other concomitant disease (5146), pulmonary viral coinfection (659), pulmonary bacterial coinfection (244), non-pulmonary bacterial coinfection (251), non-pulmonary viral coinfection (1035), home oxygen or ventilation therapy (165), preterm birth (1233), exposure to smoking (1177), immunosuppression (50), other COVID-19 risk factors (1177); other variables did not have missing

but excluded patients with discharge status "unfavorable prognosis", and evaluated the associations between all phenotypes and discharge status (including full recovery and residual symptoms) with a binary logistic regression model.

We used binary logistic regression to evaluate the associations between distinct phenotypes and ICU stay. Odds ratios greater than one indicate higher possibility of ICU admission than staying in a normal ward, compared with phenotype A. Significance level was set to be 0.05.

Subgroup analysis

Based on clinical experience, age significantly influences disease severity and clinical outcome in the study population, thus we divided the DGPI registry into infants (age < 1 year) and non-infants (age $\geqslant$ 1 year). We also

conducted hierarchical agglomerative clustering for phenotype identification, and applied multinomial logistic regression and binary logistic regression as prognostic assessment for discharge status and ICU stay in these two subgroups. Since only five patients had an unfavorable prognosis in infants, we decided to only compare full recovery and residual symptoms with binary logistic regression in this group. See detailed description of data analysis in Additional file 1, Figure S2.

Results

Study participants

The DGPI registry enrolled 6983 patients: 2892 infants and 4091 older children and adolescent. 46.3% of them were female, median age was one year (IQR:0,9). At discharge, 5352 (76.6%) patients were fully recovered, 1526 (21.9%) had residual symptoms, 42 (0.6%) experienced an unfavorable prognosis (including 17 SARS-CoV-2-related deaths), 48 (0.7%) were transferred into another hospital, and 15 (0.2%) had non-SARS-CoV-2-related death. A higher proportion of infants experienced fever or general symptoms, ear, nose and throat (ENT) symptoms, and lower respiratory tract symptoms compared to non-infants, while fewer infants exhibited other symptoms and had comorbidities. See detailed description in Table 2 and Additional file 1, Table S3.

Patient characteristics by phenotypes

Two clusters were proposed as the optimal number of clusters by eight indices in the *NbClust* package, followed by six clusters as the second most frequently proposed optimal number by six indices (Additional file 1, Table S4). To identify the clinically optimal number of phenotypes, we discussed the clinical meaningfulness of the two statistically best solutions, the two phenotypes and the six phenotypes solution, with experienced pediatricians. After this discussion and as a tradeoff between statistical reasoning and better clinical applicability, we decided to report the six phenotype solution as optimal. The six phenotypes varied significantly regarding symptoms on admission, coinfection and SARS-CoV-2 risk factors. Patient characteristics by six phenotypes are shown in Table 3 and Additional file 1, Figure S3.

Difference regarding symptoms at admission

Phenotype A had similar symptom pattern as the total sample. Predominant symptoms of patients with other phenotypes were: gastrointestinal symptoms (95.9% in phenotype B), asymptomatic (95.9% in phenotype C), lower respiratory tract symptoms (49.8% in phenotype D), lower respiratory tract symptoms and ENT symptoms (86.2% and 41.7% in phenotype E), and neurological symptoms (99.2%).

Difference regarding comorbidities

Patients with phenotype D more frequently had comorbidities - respiratory disease (11.3%), cardiovascular disease (11.0%), gastrointestinal disease (5.8%), liver disease (2.0%), neurological disease (34.2%), psychiatric disease (2.4%), hematological disease (3.3%) and other concomitant diseases (74.4%) than phenotype A, phenotype B, phenotype E and phenotype F (see percentages in Table 3). Patients with phenotype C had similar patterns except for less frequently neurological comorbidity (6.8%), more frequently kidney disease (4.2%), psychiatric disease (7.5%) and oncological disease (3.8%).

Difference regarding coinfection

Patients with phenotype A more frequently had non-pulmonary bacterial infection (9.2%, including bloodstream infection, bacterial urinary tract infection/pyelonephritis, and bacterial gastroenteritis); patients with phenotype B more frequently had non-pulmonary viral coinfection (12.5%); patients with phenotype C more frequently had non-pulmonary bacterial infection (7.1%, including bloodstream infection, bacterial arthritis / osteomyelitis, and bacterial urinary tract infection / pyelonephritis); patients with phenotype D more frequently had pulmonary bacterial infection (3.3%, including *Staphylococcus aureus* and *Haemophilus influenzae*); patients with phenotype E more frequently had pulmonary viral infection (6.8%, including respiratory syncytial virus, Influenza virus, human metapneumovirus, human rhinovirus, adenovirus, bocavirus, and enterovirus) and pulmonary bacterial infection (1.9%, including *Streptococcus pneumoniae*, *Haemophilus influenzae*, and Group A *Streptococcus*). See spectrum of coinfection by phenotypes in Additional file 1, Table S1.

Difference regarding home oxygen or ventilation therapy and preterm birth

Overall, compared to patients with other phenotypes, patients with phenotype A more frequently had preterm birth (9.9%) and exposure to smoking (6.2%); patients with phenotype C were more likely to receive immunosuppression before current disease (4.1%); patients with phenotype D were more likely to receive home oxygen or ventilation therapy prior to the current disease (8.7%) and to have other SARS-CoV-2 risk factors (40.6%).

Difference regarding quarter for hospitalization, SARS-CoV-2 variant, SARS-CoV-2 vaccination, and primary reason for hospitalization

After the phenotypes were identified, we presented the distribution of patients with different phenotypes regarding the quarter for hospitalization, SARS-CoV-2 variant, SARS-CoV-2 vaccination, and primary reason for hospitalization. Phenotypes did not differ significantly

Table 3 Characteristics of participants by phenotypes

Characteristics	Median (IQR) / *n* (%)						
	Total sample (*n* = 6983)	Phenotype A (*n* = 2529)	Phenotype B (*n* = 734)	Phenotype C (*n* = 732)	Phenotype D (*n* = 913)	Phenotype E (*n* = 1460)	Phenotype F (*n* = 615)
Sex = Female	3236 (46.3)	1168 (46.2)	378 (51.5)	345 (47.1)	404 (44.2)	665 (45.5)	276 (44.9)
COVID-19 symptoms on admission							
No symptoms (asymptomatic)	702 (10.1)	0 (0.0)	0 (0.0)	702 (95.9)	0 (0.0)	0 (0.0)	0 (0.0)
General symptoms	4818 (69.0)	2182 (86.3)	457 (62.3)	16 (2.2)	655 (71.7)	1002 (68.6)	506 (82.3)
Ear, nose, and throat symptoms	1627 (23.3)	675 (26.7)	18 (2.5)	9 (1.2)	197 (21.6)	609 (41.7)	119 (19.3)
Lower respiratory tract symptoms	2286 (32.7)	414 (16.4)	17 (2.3)	12 (1.6)	455 (49.8)	1259 (86.2)	129 (21.0)
Cardiovascular symptoms	226 (3.2)	183 (7.2)	0 (0.0)	0 (0.0)	30 (3.3)	7 (0.5)	6 (1.0)
Gastrointestinal symptoms	1884 (27.0)	437 (17.3)	704 (95.9)	2 (0.3)	269 (29.5)	308 (21.1)	164 (26.7)
Liver symptoms	29 (0.4)	23 (0.9)	2 (0.3)	1 (0.1)	0 (0.0)	2 (0.1)	1 (0.2)
Neurological / neuromuscular Symptoms	1056 (15.1)	167 (6.6)	13 (1.8)	8 (1.1)	225 (24.6)	33 (2.3)	610 (99.2)
Musculoskeletal Symptoms	200 (2.9)	143 (5.7)	4 (0.5)	1 (0.1)	29 (3.2)	6 (0.4)	17 (2.8)
Other symptoms on admission	422 (6.0)	341 (13.5)	22 (3.0)	1 (0.1)	31 (3.4)	23 (1.6)	4 (0.7)
Unknown symptoms on admission	60 (0.9)	1 (0.0)	0 (0.0)	1 (0.1)	1 (0.1)	57 (3.9)	0 (0.0)
Comorbidities at the time of COVID-19 infection							
Respiratory disease	295 (4.2)	59 (2.3)	8 (1.1)	27 (3.7)	103 (11.3)	96 (6.6)	2 (0.3)
Cardiovascular disease	261 (3.7)	119 (4.7)	6 (0.8)	27 (3.7)	100 (11.0)	9 (0.6)	0 (0.0)
Gastrointestinal tract disease	193 (2.8)	50 (2.0)	34 (4.6)	39 (5.3)	53 (5.8)	16 (1.1)	1 (0.2)
Liver disease	65 (0.9)	16 (0.6)	6 (0.8)	25 (3.4)	18 (2.0)	0 (0.0)	0 (0.0)
Kidney disease	145 (2.1)	86 (3.4)	1 (0.1)	31 (4.2)	23 (2.5)	1 (0.1)	3 (0.5)
Neurological/neuromuscular disease	445 (6.4)	55 (2.2)	1 (0.1)	50 (6.8)	312 (34.2)	23 (1.6)	4 (0.7)
Psychiatric disease	111 (1.6)	11 (0.4)	7 (1.0)	55 (7.5)	22 (2.4)	6 (0.4)	10 (1.6)
Hematologic disease	155 (2.2)	75 (3.0)	2 (0.3)	25 (3.4)	30 (3.3)	19 (1.3)	4 (0.7)
Oncological disease	106 (1.5)	61 (2.4)	3 (0.4)	28 (3.8)	8 (0.9)	6 (0.4)	0 (0.0)
Organ or bone marrow/stem cell transplantation	39 (0.6)	23 (0.9)	0 (0.0)	13 (1.8)	2 (0.2)	0 (0.0)	1 (0.2)
Autoimmunological disease	136 (1.9)	74 (2.9)	0 (0.0)	27 (3.7)	27 (3.0)	7 (0.5)	1 (0.2)
Congenital immunodeficiency	28 (0.4)	11 (0.4)	0 (0.0)	3 (0.4)	7 (0.8)	6 (0.4)	1 (0.2)
Tracheostoma (prior to current infection)	18 (0.3)	1 (0.0)	0 (0.0)	1 (0.1)	14 (1.5)	2 (0.1)	0 (0.0)
Other concomitant disease	965 (13.8)	152 (6.0)	2 (0.3)	96 (13.1)	679 (74.4)	34 (2.3)	2 (0.3)
Coinfection							
Pulmonary viral infection	131 (1.9)	16 (0.6)	3 (0.4)	4 (0.5)	9 (1.0)	99 (6.8)	0 (0.0)
Pulmonary bacterial infection	81 (1.2)	16 (0.6)	1 (0.1)	6 (0.8)	30 (3.3)	28 (1.9)	0 (0.0)
Non-pulmonary bacterial infection	331 (4.7)	232 (9.2)	2 (0.3)	52 (7.1)	31 (3.4)	14 (1.0)	0 (0.0)
Non-pulmonary viral infection	136 (1.9)	12 (0.5)	92 (12.5)	5 (0.7)	16 (1.8)	9 (0.6)	2 (0.3)
COVID-19 risk factors							
Home oxygen or ventilation therapy before the current disease	111 (1.6)	18 (0.7)	1 (0.1)	7 (1.0)	79 (8.7)	6 (0.4)	0 (0.0)
preterm infant	357 (5.1)	250 (9.9)	1 (0.1)	45 (6.1)	46 (5.0)	14 (1.0)	1 (0.2)
Exposure to smoking	226 (3.2)	156 (6.2)	1 (0.1)	31 (4.2)	27 (3.0)	4 (0.3)	7 (1.1)
Immunosuppression	149 (2.1)	89 (3.5)	2 (0.3)	30 (4.1)	21 (2.3)	5 (0.3)	2 (0.3)
Other COVID-19 risk factors	465 (6.7)	46 (1.8)	2 (0.3)	36 (4.9)	371 (40.6)	10 (0.7)	0 (0.0)

in quarter of the year for hospitalization: patients were mostly admitted in the first quarter and least admitted in the third quarter. Additionally, no differences of patients with different phenotypes were observed regarding their infection with different SARS-CoV-2 variants or their vaccination status against SARS-CoV-2. SARS-CoV-2 infection was the primary reason for hospitalization in 3.7% of the patients with phenotype C, 40.5% in

phenotype B, 42.0% in phenotype F, and slightly over 50% in other phenotypes (Additional file 1, Table S5).

Patient characteristics in infants and non-infants

Overall, non-infants and infants exhibited very similar phenotypes to the whole registry. However, phenotype F in infants did not exhibit representative neurological/neuromuscular symptoms at admission as in non-infants (100%) and in the whole registry (99.2%). Instead, phenotype F in infants showed similar attributes as phenotype D whereas with more percentage of patients who had other COVID-19 risk factors (35.6%) (Additional file 1, Table S6 and Table S7).

Association between phenotypes and clinical outcomes

Figure 1 shows the association between phenotypes and clinical outcomes. Compared to full recovery, patients with phenotype C had a lower risk of having residual symptoms (OR: 0.10 [0.06, 0.15]) than those with phenotype A, whereas patients with D and E phenotype had a higher risk of having residual symptoms (OR: 1.33 [1.11, 1.59] and 1.91 [1.65, 2.21], respectively) than those with phenotype A. Additionally, patients with phenotype D were significantly more likely (OR: 4.00 [1.95, 8.19]) to have an unfavorable prognosis and higher possibility of ICU admission than staying in normal ward (OR: 4.26 [3.06, 5.98]), compared to patients with phenotype A. Patients with phenotype B also had lower risk of having residual symptoms (OR: 0.72 [0.58, 0.89]) than those with phenotype A (Additional file 1, Figure S4).

The outcomes observed in the non-infants with phenotype D and phenotype E closely resembled those of the entire registered population, except for less risk of ICU admission of phenotype E (OR: 0.22 [0.08, 0.46]) than phenotype A. In infants, phenotype D, phenotype E, and phenotype F all had higher risk of having residual symptoms than phenotype A (OR: 1.69 [1.06, 2.62], 2.67 [2.15, 3.31], 1.77 [1.25, 2.47], respectively), and phenotype D and E had higher risk of ICU admission (OR: 7.41 [2.65, 19.65], 2.46 [1.11, 5.67], respectively).

Discussion

We identified six distinct phenotypes for children and adolescents with SARS-CoV-2 infection by applying an unsupervised machine learning method in a nationwide registry of Germany. We found that patients with phenotype D and phenotype E had higher risk of having residual symptoms than those with phenotype A, and patients with phenotype D also had 4 times risk of having unfavorable prognosis and 4.26 times risk of ICU admission

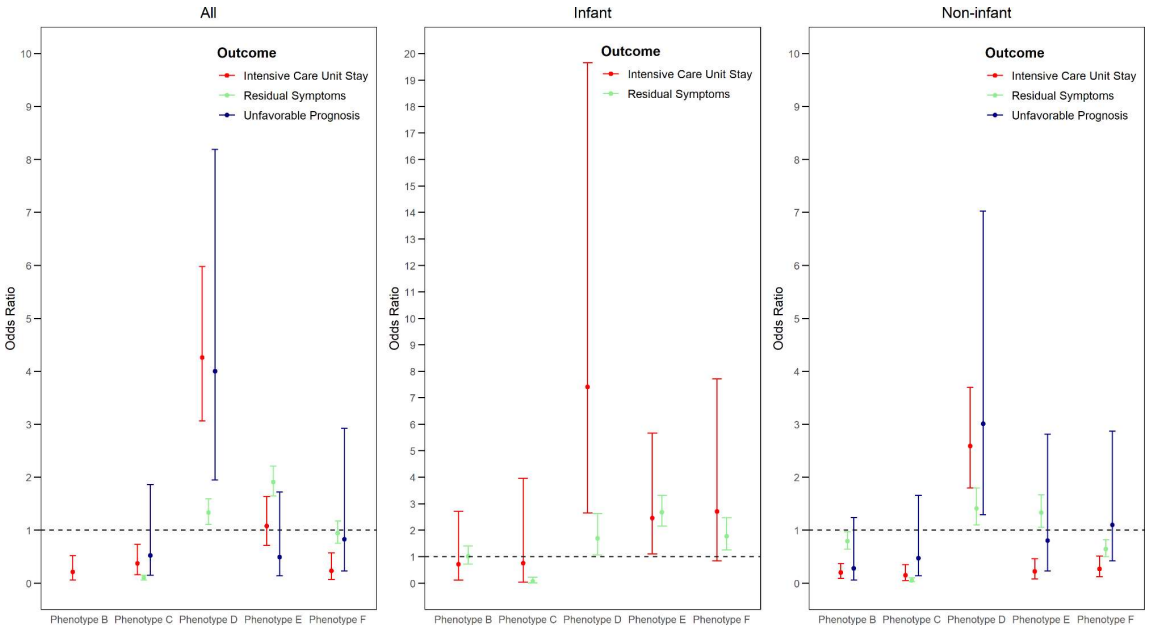


Fig. 1 Risk association between phenotypes and clinical outcomes. (1) Phenotype A was the reference phenotype; full recovery was the reference discharge status for residual symptoms and unfavorable prognosis; staying in normal ward was the reference for intensive care unit stay. (2) Odds ratios for residual symptoms and unfavorable prognosis in all registered population and non-infants were estimated with multinomial logistic regression, odds ratios for intensive care unit stay in all groups and odds ratio for residual symptom in infants were estimated with binary logistic regression. (3) Since no patient with phenotype B had an unfavorable prognosis in all registered population, we excluded phenotype B in multinomial logistic regression and evaluated the associations between other phenotypes and discharge status (including full recovery, residual symptoms, and unfavorable prognosis); age was included in the model as a confounder

than those phenotype A. Compared to the solution with two phenotypes, we were able to offer insights with a finer granularity into the clinical presentation of children and adolescents with SARS-CoV-2 infection. This stratification also found one specific group which had the highest risk of ICU admission and unfavorable prognosis, thus enabling the most appropriate patient care for them.

Patients with phenotype D primarily exhibited lower respiratory tract symptoms, and were at elevated risk of residual symptoms at discharge, developing unfavorable prognosis, and ICU admission than those with phenotype A. Former studies also reported that independent risk factors for moderate/severe disease involves signs and symptoms such as shortness of breath, rash, seizures, temperature on arrival, chest recessions, and abnormal breath sounds [22, 24]. In addition, we found that patients with phenotype D more frequently had pre-existing comorbidities including respiratory disease, cardiovascular disease, gastrointestinal disease, liver disease, neurological disease, psychiatric disease, hematological disease and other concomitant diseases than other phenotypes. This result is in line with former publications. Geva et al. found one phenotype with frequently pre-existing respiratory conditions needed more invasive or non-invasive mechanical ventilation and had more percentage of deaths, compared to other phenotypes [15].

Patients with phenotype C, primarily asymptomatic, had similar comorbidity patterns as phenotype D, except for less frequently neurological comorbidity, more frequently kidney disease, psychiatric disease and oncological disease. This can be explained by the fact that SARS-CoV-2 infection was not the main reason of hospitalization for most of these patients and was found during inpatient stay. Unsurprisingly, patients with this phenotype had lower risk of residual symptoms.

We found that patients with phenotype D more frequently had pulmonary bacterial coinfection and patients with phenotype E more frequently had pulmonary viral coinfection and pulmonary bacterial infection. It has been reported that coinfection with respiratory syncytial virus (RSV) and bacteria was associated with severe illness in infants, and coinfection with RSV was associated with severe illness in COVID-19 patients aged 1 to 4 years [23]. Schober et al. also found that viral coinfection was associated with severe disease of COVID-19 in univariable ordinal logistic regression [19]. Also, patients with phenotype B, characterized mainly by gastrointestinal symptoms, more frequently had non-pulmonary viral infection, and patients with phenotype C, primarily asymptomatic, more frequently had non-pulmonary bacterial infection than patients with other phenotypes. Given that patients with phenotype D had higher risk of both having residual symptoms at discharge and developing unfavorable prognosis than those with phenotype A, we believe pulmonary bacterial coinfection were associated with severe disease of COVID-19 and unfavorable prognosis.

In our study, patients with phenotype D received home oxygen or ventilation therapy before SARS-CoV-2 infection than other phenotypes. Farrar et al. also revealed that pre-existing technology dependence requirements including requirement for home oxygen were associated with severe disease [36].

The vast majority of patients with phenotype B had gastrointestinal symptoms. One systematic review showed that gastrointestinal symptoms have been reported in 17.6% of COVID-19 patients [37], and another review reported these manifestations to be more prevalent in children as compared to adults [38]. These symptoms are generally self-limiting, but supportive treatment is needed [38]. This is in line with our study that patients with this phenotype had lower risk of ICU admission. Also, patients with phenotype B more frequently had non-pulmonary viral coinfection than patients with other phenotypes. This coinfection could possibly be viral gastroenteritis, which also needed supportive treatment other than ICU stay.

Patients with phenotype E showed involvement of both lower respiratory tract and ENT. ENT symptoms including dysosmia, dysgeusia, rhinorrhea have been reported in other studies before [39, 40]. One study from Italy showed that loss of taste/smell existed in 3.3% of the participants from primary care at follow-up of 8 to 36 weeks [41]. Thus, it is self-explanatory that patients with phenotype E had higher risk of residual symptoms in our study. Furthermore, we think that patients with phenotype E generally had fewer pre-existing comorbidities than patients with phenotype D was the reason why patients with phenotype E did not show similar prognosis as patients with phenotype D.

It is understandable that patients with phenotype F exhibited typical neurological symptoms, since neurological complications has been documented before in COVID-19 cases [42, 43]. Possible mechanisms of neurological involvement in SARS-CoV-2 infection included direct viral invasion and immune-mediated damage of nervous system [42]. Although it was shown that most neurological symptoms in children and adolescents with SARS-CoV-2 infection were transient and life-threatening conditions were rare [43], severe neurologic manifestations during hospitalization were shown to be associated with new neurocognitive impairments or functional disabilities at hospital discharge [44]. This might explain why patients with phenotype F showed lower risk of ICU admission, but did not exhibit significant difference to the comparator phenotype regarding unfavorable prognosis.

Age was considered as a mortality risk factor for children and adolescents, with an increased risk of death for those younger than two years and those older than 10 years [25, 27, 28]. Our subgroup analysis revealed that infants and non-infants exhibited nearly identical phenotypic characteristics as observed in all registered population. Nevertheless, within infants, phenotype F did not manifest the typical neurological/neuromuscular symptoms observed in all registered population and in non-infants, but rather similar attributes as phenotype D. Furthermore, infants with phenotype D more frequently had preterm birth history. It has also been reported that prematurity was associated with severe COVID-19 [20, 21, 45].

Our work has limitations. Firstly, epidemiological and clinical parameters were used for identifying phenotypes, but laboratory results were not included. Adding laboratory results might result in more refined phenotypes. Secondly, the inclusion of only hospitalized patients necessitates caution when extrapolating the results to the whole infected population. Thirdly, our phenotypes were not validated with an external cohort. Confirmation is warranted regarding whether individuals from other population demonstrate comparable clustering patterns. Lastly, we did not differentiate patients admitted due to SARS-CoV-2 infection and those with incidental positive SARS-CoV-2 test results [46, 47], but we used symptoms on admission which were highly relevant to whether SARS-CoV-2 infection was the primary reason for hospitalization.

Conclusions

Clustering pediatric patients into phenotypes might help to stratify individuals according to risk and thus assist in tailored patient management. Our findings in SARS-CoV-2-infected population might also be transferable to other infectious diseases.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12931-024-03018-3.

Supplementary Material 1

Author contributions

Y.S.: Conceptualization, Software, Methodology, Formal analysis, Investigation, Visualisation, Writing - Original Draft, Writing - Review & Editing. R.S.: Conceptualization, Software, Data Curation, Methodology, Project administration, Writing - Review & Editing. R.B.: Conceptualization, Writing - Review & Editing. J.A.: Conceptualization, Writing - Review & Editing. S.S.: Conceptualization, Writing - Review & Editing. E.G.: Conceptualization, Supervision, Writing - Review & Editing.

Funding

The study was funded by the National University Medicine Research Network on COVID-19, "NaFoUniMedCovid19" within the framework of the project "CODEX + Monitor" [Grant No. 01KX2121]. The DGPI registry was partially funded by a grant from the Federal State of Saxony, Germany [Grant No. 201.1001]. Yanyan Shi is supported by a scholarship from the China Scholarship Council [202108140048].
Open Access funding enabled and organized by Projekt DEAL.

Data availability

The dataset supporting the conclusions of this article can be shared upon reasonable requests to Dr. Jakob Armann (Jakob.Armann@ukdd.de).

Declarations

Role of the funding source

The funders of this study had no role in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

Ethics approval and consent to participate

The DGPI registry was approved by the Ethics Committee of the Technische Universität (TU) Dresden (BO-EK-110032020).

Consent for publication

Not Applicable.

Competing interest

None.

Author details

[1]Institute for Medical Information Processing, Biometry and Epidemiology (IBE), Faculty of Medicine, Ludwig-Maximilians-Universität München (LMU Munich), Marchioninistr. 15, 81377 Munich, Germany
[2]Pettenkofer School of Public Health, Munich, Germany
[3]German Center for Vertigo and Balance Disorders, University Hospital, Ludwig-Maximilians- Universität München (LMU Munich), Munich, Germany
[4]Department of Pediatrics, University Hospital and Medical Faculty Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany
[5]Department of Infection Control and Infectious Diseases, University Medical Center Göttingen, Göttingen, Germany

Received: 17 July 2024 / Accepted: 20 October 2024
Published online: 30 October 2024

References

1. Ouldali N, Yang DD, Madhi F, et al. Factors Associated with severe SARS-CoV-2 infection. Pediatrics. 2021;147(3):e2020023432. https://doi.org/10.1542/peds.2020-023432.
2. Ludvigsson JF. Systematic review of COVID-19 in children shows milder cases and a better prognosis than adults. Acta Paediatr. 2020;109(6):1088–95. https://doi.org/10.1111/apa.15270.
3. Feldstein LR, Tenforde MW, Friedman KG, et al. Characteristics and outcomes of US children and adolescents with Multisystem Inflammatory Syndrome in Children (MIS-C) compared with severe Acute COVID-19. JAMA. 2021;325(11):1074–87. https://doi.org/10.1001/jama.2021.2091.
4. McCormick DW, Richardson LC, Young PR, et al. Deaths in children and adolescents Associated with COVID-19 and MIS-C in the United States. Pediatrics. 2021;148(5):e2021052273. https://doi.org/10.1542/peds.2021-052273.
5. Bailey LC, Razzaghi H, Burrows EK, et al. Assessment of 135 794 Pediatric patients tested for severe Acute Respiratory Syndrome Coronavirus 2 across the United States. JAMA Pediatr. 2021;175(2):176–84. https://doi.org/10.1001/jamapediatrics.2020.5052.
6. Zhu Y, Almeida FJ, Baillie JK, et al. International Pediatric COVID-19 Severity over the course of the pandemic. JAMA Pediatr. 2023;177(10):1073–84. https://doi.org/10.1001/jamapediatrics.2023.3117.
7. Sorg A-L, Hufnagel M, Doenhardt M, et al. Risk for severe outcomes of COVID-19 and PIMS-TS in children with SARS-CoV-2 infection in Germany. Eur J Pediatrics. 2022;181(10):3635–43. https://doi.org/10.1007/s00431-022-04587-5.
8. Yamga E, Mullie L, Durand M, et al. Interpretable clinical phenotypes among patients hospitalized with COVID-19 using cluster analysis. Front Digit Health. 2023;5:1142822. https://doi.org/10.3389/fdgth.2023.1142822.

9. Loftus TJ, Shickel B, Balch JA, et al. Phenotype clustering in health care: a narrative review for clinicians. Front Artif Intell. 2022;5. https://doi.org/10.3389/frai.2022.842306.
10. Hasan MM, Islam MU, Sadeq MJ, Fung W-K, Uddin J. Review on the Evaluation and Development of Artificial Intelligence for COVID-19 Containment. Sensors. 2023;23(1):527. https://doi.org/10.3390/s23010527.
11. Gutiérrez-Gutiérrez B, Del Toro MD, Borobia AM, et al. Identification and validation of clinical phenotypes with prognostic implications in patients admitted to hospital with COVID-19: a multicentre cohort study. Lancet Infect Dis. 2021;21(6):783–92. https://doi.org/10.1016/S1473-3099(21)00019-0.
12. Seymour CW, Kennedy JN, Wang S, et al. Derivation, validation, and potential treatment implications of Novel Clinical Phenotypes for Sepsis. JAMA. 2019;321(20):2003–17. https://doi.org/10.1001/jama.2019.5791.
13. Ozonoff A, Schaenman J, Jayavelu ND, et al. Phenotypes of disease severity in a cohort of hospitalized COVID-19 patients: results from the IMPACC study. EBioMedicine. 2022;83:104208. https://doi.org/10.1016/j.ebiom.2022.104208.
14. Rodríguez A, Ruiz-Botella M, Martín-Loeches I, et al. Deploying unsupervised clustering analysis to derive clinical phenotypes and risk factors associated with mortality risk in 2022 critically ill patients with COVID-19 in Spain. Crit Care. 2021;25(1):63. https://doi.org/10.1186/s13054-021-03487-8.
15. Geva A, Patel MM, Newhams MM et al. Data-driven clustering identifies features distinguishing multisystem inflammatory syndrome from acute COVID-19 in children and adolescents. eClinicalMedicine. 2021; 40. https://doi.org/10.1016/j.eclinm.2021.101112
16. Armann J, Doenhardt M, Hufnagel M et al. Risk factors for hospitalization, disease severity and mortality in children and adolescents with COVID-19: Results from a nationwide German registry. medRxiv 2021: 2021.06.07.21258488. https://doi.org/10.1101/2021.06.07.21258488
17. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The strengthening the reporting of Observational studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008;61(4):344–9. https://doi.org/10.1016/j.jclinepi.2007.11.008.
18. Williams N, Radia T, Harman K, Agrawal P, Cook J, Gupta A. COVID-19 severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection in children and adolescents: a systematic review of critically unwell children and the association with underlying comorbidities. Eur J Pediatr. 2021;180(3):689–97. https://doi.org/10.1007/s00431-020-03801-6.
19. Schober T, Caya C, Barton M, et al. Risk factors for severe PCR-positive SARS-CoV-2 infection in hospitalised children. BMJ Paediatr Open. 2022;6(1). https://doi.org/10.1136/bmjpo-2022-001440.
20. Prévost B, Retbi A, Binder-Foucard F, et al. Risk factors for admission to the pediatric critical care unit among children hospitalized with COVID-19 in France. Front Pediatr. 2022;10:975826. https://doi.org/10.3389/fped.2022.975826.
21. Graff K, Smith C, Silveira L, et al. Risk factors for severe COVID-19 in children. Pediatr Infect Dis J. 2021;40(4):e137–45. https://doi.org/10.1097/INF.0000000000003043.
22. Nguyen PNT, Thuc TT, Hung NT, et al. Risk factors for disease severity and mortality of children with Covid-19: a study at a Vietnamese children's hospital. J Infect Chemother. 2022;28(10):1380–6. https://doi.org/10.1016/j.jiac.2022.06.010.
23. Choudhary R, Webber BJ, Womack LS, et al. Factors Associated with severe illness in patients aged < 21 years hospitalized for COVID-19. Hosp Pediatr. 2022;12(9):760–83. https://doi.org/10.1542/hpeds.2022-006613.
24. Ng DC-E, Liew C-H, Tan KK, et al. Risk factors for disease severity among children with Covid-19: a clinical prediction model. BMC Infect Dis. 2023;23(1):398. https://doi.org/10.1542/hpeds.2022-006613.
25. Harwood R, Yan H, Talawila Da Camara N, et al. Which children and young people are at higher risk of severe disease and death after hospitalisation with SARS-CoV-2 infection in children and young people: a systematic review and individual patient meta-analysis. eClinicalMedicine. 2022;44. https://doi.org/10.1016/j.eclinm.2022.101287.
26. Bertran M, Amin-Chowdhury Z, Davies HG, et al. COVID-19 deaths in children and young people in England, March 2020 to December 2021: an active prospective national surveillance study. PLoS Med. 2022;19(11):e1004118. https://doi.org/10.1371/journal.pmed.1004118.
27. Oliveira EA, Oliveira MCL, Silva A, et al. Clinical outcomes of Omicron variant (B.1.1.529) infection in children and adolescents hospitalized with COVID-19 in Brazil with Observational Data on the efficacy of the vaccines in adolescents. Pediatr Infect Dis J. 2023;42(3):218–25. https://doi.org/10.1097/INF.0000000000003783.
28. Smith C, Odd D, Harwood R, et al. Deaths in children and young people in England after SARS-CoV-2 infection during the first pandemic year. Nat Med. 2022;28(1):185–92. https://doi.org/10.1038/s41591-021-01578-1.
29. Little RJA. A test of missing completely at Random for Multivariate Data with missing values. J Am Stat Assoc. 1988;83(404):1198–202. https://doi.org/10.2307/2290157.
30. Murray CJ, Richards MA, Newton JN, et al. UK health performance: findings of the global burden of Disease Study 2010. Lancet. 2013;381(9871):997–1020. https://doi.org/10.1016/S0140-6736(13)60355-4.
31. Breiman L. Manual for Setting Up, Using, and Understanding Random Forest V4.0. 2003.
32. Gower JC. A General Coefficient of Similarity and some of its Properties. Biometrics. 1971;27:857–71. https://doi.org/10.2307/2528823.
33. Ward JH. Hierarchical grouping to optimize an objective function. J Am Stat Assoc. 1963;58(301):236–44. https://doi.org/10.1080/01621459.1963.10500845.
34. Charrad M, Ghazzali N, Boiteau V, Niknafs A, NbClust. An R Package for determining the relevant number of clusters in a Data Set. J Stat Softw. 2014;61(6):1–36. https://doi.org/10.18637/jss.v061.i06.
35. Maechler M, Rousseeuw P, Struyf A, Mia. H, Hornik K. cluster: Cluster Analysis Basics and Extensions (R package version 2.1.6). 2023. https://doi.org/10.32614/CRAN.package.cluster
36. Farrar DS, Drouin O, Moore Hepburn C, et al. Risk factors for severe COVID-19 in hospitalized children in Canada: a national prospective study from March 2020-May 2021. Lancet Reg Health Am. 2022;15:100337. https://doi.org/10.1016/j.lana.2022.100337.
37. Cheung KS, Hung IFN, Chan PPY, et al. Gastrointestinal manifestations of SARS-CoV-2 infection and virus load in fecal samples from a Hong Kong Cohort: systematic review and Meta-analysis. Gastroenterology. 2020;159(1):81–95. https://doi.org/10.1053/j.gastro.2020.03.065.
38. Calitri C, Fumi I, Ignaccolo MG, et al. Gastrointestinal involvement in paediatric COVID-19 - from pathogenesis to clinical management: a comprehensive review. World J Gastroenterol. 2021;27(23):3303–16. https://doi.org/10.3748/wjg.v27.i23.3303.
39. Chen Y-c, Wang X, Teng Y-s, Jia D-s, Li L, Pan H-g. Otolaryngology-related symptoms of COVID-19 in children in the post-epidemic era: a cross-sectional web-based survey study. Front Pead. 2023. https://doi.org/10.3389/fped.2023.1190734.
40. Hijazi LO, Alaraifi AK, Alsaab F. Otolaryngology manifestations of COVID-19 in pediatric patients. Int J Pediatr Otorhinolaryngol. 2021;144:110701. https://doi.org/10.1016/j.ijporl.2021.110701.
41. Trapani G, Verlato G, Bertino E, et al. Long COVID-19 in children: an Italian cohort study. Ital J Pediatr. 2022;48(1):83. https://doi.org/10.1186/s13052-022-01282-x.
42. Laçinel Gürlevik S, Günbey C, Ozsurekci Y, et al. Neurologic manifestations in children with COVID-19 from a tertiary center in Turkey and literature review. Eur J Paediatr Neurol. 2022;37:139–54. https://doi.org/10.1016/j.ejpn.2022.02.003.
43. LaRovere KL, Riggs BJ, Poussaint TY, et al. Neurologic involvement in children and adolescents hospitalized in the United States for COVID-19 or Multisystem Inflammatory Syndrome. JAMA Neurol. 2021;78(5):536–47. https://doi.org/10.1001/jamaneurol.2021.0504.
44. Francoeur C, Alcamo AM, Robertson CL, et al. Severe Pediatric neurological manifestations with SARS-CoV-2 or MIS-C hospitalization and New Morbidity. JAMA Netw Open. 2024;7(6):e2414122–e. https://doi.org/10.1001/jamanetworkopen.2024.14122.
45. Woodruff RC, Campbell AP, Taylor CA, et al. Risk factors for severe COVID-19 in children. Pediatrics. 2022;149(1). https://doi.org/10.1542/peds.2021-053418.
46. Misailovski M, Koller D, Blaschke S, et al. Refining the hospitalization rate: a mixed methods approach to differentiate primary COVID-19 from incidental cases. Infect Prev Pract. 2024;6(3):100371. https://doi.org/10.1016/j.infpip.2024.100371.
47. Strobl R, Misailovski M, Blaschke S, et al. Differentiating patients admitted primarily due to coronavirus disease 2019 (COVID-19) from those admitted with incidentally detected severe acute respiratory syndrome corona-virus type 2 (SARS-CoV-2) at hospital admission: a cohort analysis of German hospital records. Infect Control Hosp Epidemiol. 2024;45(6):746–53. https://doi.org/10.1017/ice.2024.3.

Publisher's note

Six Clinical Phenotypes with Prognostic Implications were identified by Unsupervised Machine Learning in Children and Adolescents with SARS-CoV-2 Infection: Results from a German Nationwide Registry

Table of contents

Table S1 Clinically relevant bacterial or viral coinfection by phenotypes

Characteristics	n (%)						
	Total sample (n=6983)	**Phenotype A (n=2529)**	**Phenotype B (n=734)**	**Phenotype C (n=732)**	**Phenotype D (n=913)**	**Phenotype E (n=1460)**	**Phenotype F (n=615)**
Pulmonary viral infection							
Respiratory syncytial virus (RSV)	51 (0.7)	7 (0.3)	1 (0.1)	1 (0.1)	5 (0.5)	37 (2.5)	0 (0.0)
Influenza A or B virus	6 (0.1)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	5 (0.3)	0 (0.0)
Human metapneumovirus (HPMV)	10 (0.1)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	10 (0.7)	0 (0.0)
Human rhinovirus (HRV)	36 (0.5)	4 (0.2)	1 (0.1)	3 (0.4)	1 (0.1)	27 (1.8)	0 (0.0)
Adenovirus (respiratory subtypes)	15 (0.2)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	14 (1.0)	0 (0.0)
Bocavirus	11 (0.2)	1 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)	9 (0.6)	0 (0.0)
Enterovirus (respiratory subtypes)	19 (0.3)	2 (0.1)	2 (0.3)	1 (0.1)	1 (0.1)	13 (0.9)	0 (0.0)
Pulmonary bacterial infection							
Streptococcus pneumoniae	6 (0.1)	1 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)	4 (0.3)	0 (0.0)
Staphylococcus aureus	10 (0.1)	3 (0.1)	0 (0.0)	2 (0.3)	4 (0.4)	1 (0.1)	0 (0.0)
Haemophilus influenzae	11 (0.2)	1 (0.0)	0 (0.0)	0 (0.0)	5 (0.5)	5 (0.3)	0 (0.0)
Group A *Streptococcus*	5 (0.1)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	5 (0.3)	0 (0.0)
Mycoplasma	6 (0.1)	2 (0.1)	1 (0.1)	0 (0.0)	2 (0.2)	1 (0.1)	0 (0.0)
Non-pulmonary bacterial infection							
Bloodstream infection	34 (0.5)	22 (0.9)	0 (0.0)	5 (0.7)	6 (0.7)	1 (0.1)	0 (0.0)
Bacterial meningitis	3 (0.0)	2 (0.1)	0 (0.0)	0 (0.0)	1 (0.1)	0 (0.0)	0 (0.0)
Bacterial arthritis / osteomyelitis	4 (0.1)	0 (0.0)	0 (0.0)	3 (0.4)	1 (0.1)	0 (0.0)	0 (0.0)
Bacterial urinary tract infection / pyelonephritis	132 (1.9)	103 (4.1)	0 (0.0)	13 (1.8)	10 (1.1)	6 (0.4)	0 (0.0)
Bacterial endocarditis	2 (0.0)	2 (0.1)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Bacterial gastroenteritis	32 (0.5)	26 (1.0)	2 (0.3)	0 (0.0)	2 (0.2)	2 (0.1)	0 (0.0)

Table S2 Other SARS-CoV-2 risk factors

Variable	n (%)
Other SARS-CoV-2 risk factors	465 (6.7)
The newborn patient's mother was SARS-CoV-2 positive	9 (0.1)
Not specified	456 (6.6)

Table S3 Characteristics of participants of DGPI registry by age group

Variable	All (n=6983)	Infant (n=2892)	Non-infant (n=4091)
Age (years, median(IQR))	1 (0,9)	0 (0,0)	7 (2,13)
Sex = Female	3236 (46.3)	1274 (44.1)	1962 (48.0)
No symptoms (asymptomatic)	702 (10.1)	165 (5.7)	537 (13.1)
Fever or general symptoms	4818 (69.0)	2401 (83.0)	2417 (59.1)
Ear, nose, and throat symptoms	1627 (23.3)	771 (26.7)	856 (20.9)
Lower respiratory tract symptoms	2286 (32.7)	967 (33.4)	1319 (32.2)
Cardiovascular symptoms	226 (3.2)	54 (1.9)	172 (4.2)
Gastrointestinal symptoms	1884 (27.0)	631 (21.8)	1253 (30.6)
Liver symptoms	29 (0.4)	11 (0.4)	18 (0.4)
Neurological/ neuromuscular symptoms	1056 (15.1)	131 (4.5)	925 (22.6)
Musculoskeletal Symptoms	200 (2.9)	8 (0.3)	192 (4.7)
Other symptoms on admission	422 (6.0)	116 (4.0)	306 (7.5)
Unknown admission symptoms	60 (0.9)	16 (0.6)	44 (1.1)
Respiratory disease	295 (4.2)	53 (1.8)	242 (5.9)
Cardiovascular disease	261 (3.7)	88 (3.0)	173 (4.2)
Gastrointestinal tract disease	193 (2.8)	32 (1.1)	161 (3.9)
Liver disease	65 (0.9)	25 (0.9)	40 (1.0)
Kidney disease	145 (2.1)	37 (1.3)	108 (2.6)
Neurological/ neuromuscular disease	445 (6.4)	38 (1.3)	407 (9.9)
Psychiatric disease	111 (1.6)	1 (0.0)	110 (2.7)
Hematologic disease	155 (2.2)	19 (0.7)	136 (3.3)
Oncological disease	106 (1.5)	4 (0.1)	102 (2.5)
Organ or bone marrow/stem cell transplantation	39 (0.6)	2 (0.1)	37 (0.9)
Autoimmunological disease	136 (1.9)	1 (0.0)	135 (3.3)
Congenital immunodeficiency	28 (0.4)	1 (0.0)	27 (0.7)
Tracheostoma (prior to current infection)	18 (0.3)	0 (0.0)	18 (0.4)
Other concomitant disease	965 (13.8)	349 (12.1)	616 (15.1)
Pulmonary viral coinfection	131 (1.9)	62 (2.1)	69 (1.7)
Pulmonary bacterial coinfection	81 (1.2)	11 (0.4)	70 (1.7)
Non-pulmonary bacterial coinfection	331 (4.7)	121 (4.2)	210 (5.1)
Non-pulmonary viral coinfection	136 (1.9)	38 (1.3)	98 (2.4)

Home oxygen or ventilation therapy	111 (1.6)	37 (1.3)	74 (1.8)
Preterm birth	357 (5.1)	240 (8.3)	117 (2.9)
Exposure to smoking	226 (3.2)	69 (2.4)	157 (3.8)
Immunsuppression	149 (2.1)	4 (0.1)	145 (3.5)
Other COVID-19 risk factors	465 (6.7)	136 (4.7)	329 (8.0)

Table S4 Number of optimal clusters proposed by *NbClust* package

Index	Index Origin	Algorithm for optimal number of clusters	Optimal clusters	Index value
CH	Calinski and Harabasz 1974	Maximum value of the index	2	614.91
DB	Davies and Bouldin 1979	Minimum value of the index	2	1.88
Silhouette	Rousseeuw 1987	Maximum value of the index	2	0.23
Frey	Frey and Van Groenewoud 1972	The cluster level before that index value < 1.00	2	1.28
McClain	McClain and Rao 1975	Minimum value of the index	2	0.18
Dunn	Dunn 1974	Maximum value of the index	2	0.05
SDindex	Halkidi et al. 2000	Minimum value of the index	2	2.87
SDbw	Halkidi and Vazirgiannis 2001	Minimum value of the index	2	0.81
KL	Krzanowski and Lai 1988	Maximum value of the index	6	1.66
Hartigan	Hartigan 1975	Maximum difference between hierarchy levels of the index	6	95.76
TraceW	Milligan and Cooper 1985	Maximum value of absolute second differences between levels of the index	6	159.58
Rubin	Friedman and Rubin 1967	Minimum value of second differences between levels of the index	6	-0.02
Hubert	Hubert and Arabie 1985	Graphical method	6	NA
Dindex	Lebart et al. 2000	Graphical method	6	NA
CCC	Sarle 1983	Maximum value of the index	10	39.65
Scott	Scott and Symons 1971	Maximum difference between hierarchy levels of the index	10	7229.82
Friedman	Friedman and Rubin 1967	Maximum difference between hierarchy levels of the index	10	2.11
Cindex	Hubert and Levin 1976	Minimum value of the index	10	0.16
Ratkowsky	Ratkowsky and Lance 1978	Maximum value of the index	5	0.11
Marriot	Marriot 1971	Max. value of second differences between levels of the index	4	9.43E+78
PtBiserial	Milligan 1980, 1981	Maximum value of the index	4	0.28
TrCovW	Milligan and Cooper 1985	Maximum difference between hierarchy levels of the index	3	50722.28
Ball	Ball and Hall 1965	Maximum difference between hierarchy levels of the index	3	2467.01
Duda	Duda and Hart 1973	Smallest n_c such that index > criticalValue	NA	NA
PseudoT2	Duda and Hart 1973	Smallest n_c such that index < criticalValue	NA	NA
Beale	Beale 1969	n_c such that critical value of the index >= alpha	NA	NA

[REFERENCE]
Malika Charrad, Nadia Ghazzali, Véronique Boiteau, Azam Niknafs. NbClust package for determining the best number of clusters. 2014.

Table S5 Hospitalization, SARS-CoV-2 variant, and SARS-CoV-2 vaccination status by phenotypes

Characteristics	n (%)						
	Total sample (n=6983)	Phenotype A (n=2529)	Phenotype B (n=734)	Phenotype C (n=732)	Phenotype D (n=913)	Phenotype E (n=1460)	Phenotype F (n=615)
Quarter of the year for hospitalization							
Q1	3140 (45.0)	1094 (43.3)	360 (49.0)	347 (47.4)	381 (41.7)	654 (44.8)	304 (49.4)
Q2	1397 (20.0)	537 (21.2)	144 (19.6)	126 (17.2)	177 (19.4)	277 (19.0)	136 (22.1)
Q3	965 (13.8)	374 (14.8)	104 (14.2)	83 (11.3)	120 (13.1)	209 (14.3)	75 (12.2)
Q4	1481 (21.2)	524 (20.7)	126 (17.2)	176 (24.0)	235 (25.7)	320 (21.9)	100 (16.3)
SARS-CoV-2 variant							
Wildtype	10 (0.1)	2 (0.1)	1 (0.1)	3 (0.4)	2 (0.2)	2 (0.1)	0 (0.0)
Alpha	34 (0.5)	10 (0.4)	4 (0.5)	1 (0.1)	7 (0.8)	11 (0.8)	1 (0.2)
Beta	3 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)	1 (0.2)
Delta	128 (1.8)	41 (1.6)	10 (1.4)	25 (3.4)	23 (2.5)	21 (1.4)	8 (1.3)
Omicron	396 (5.7)	143 (5.7)	29 (4.0)	45 (6.1)	53 (5.8)	84 (5.8)	42 (6.8)
Other SARS-CoV-2 variant	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)	0 (0.0)	0 (0.0)
Missing	6411 (91.8)	2332 (92.2)	690 (94.0)	658 (89.9)	827 (90.6)	1341 (91.8)	563 (91.5)
SARS-CoV-2 vaccination = YES	183 (2.6)	65 (2.6)	17 (2.3)	18 (2.5)	33 (3.6)	32 (2.2)	18 (2.9)
SARS-CoV-2 infection as the primary reason for hospitalization = YES	3156 (45.2)	1292 (51.1)	297 (40.5)	27 (3.7)	472 (51.7)	810 (55.5)	258 (42.0)
Missing	297 (4.3)	125 (4.9)	28 (3.8)	18 (2.5)	25 (2.7)	87 (6.0)	14 (2.3)

Table S6 Characteristics of infants by phenotypes

Characteristics	n (%)						
	Total sample (n=2892)	**Phenotype A (n=1248)**	**Phenotype B (n=350)**	**Phenotype C (n=166)**	**Phenotype D (n=124)**	**Phenotype E (n=771)**	**Phenotype F (n=233)**
Sex=Female	1274 (44.1)	569 (45.6)	153 (43.7)	82 (49.4)	47 (37.9)	333 (43.2)	90 (38.6)
COVID-19 symptoms on admission							
No symptoms (asymptomatic)	165 (5.7)	0 (0.0)	0 (0.0)	165 (99.4)	0 (0.0)	0 (0.0)	0 (0.0)
General symptoms	2401 (83.0)	1178 (94.4)	277 (79.1)	0 (0.0)	120 (96.8)	611 (79.2)	215 (92.3)
Ear, nose and throat symptoms	771 (26.7)	298 (23.9)	5 (1.4)	0 (0.0)	45 (36.3)	369 (47.9)	54 (23.2)
Lower respiratory tract symptoms	967 (33.4)	26 (2.1)	53 (15.1)	0 (0.0)	69 (55.6)	717 (93.0)	102 (43.8)
Cardiovascular symptoms	54 (1.9)	20 (1.6)	5 (1.4)	1 (0.6)	7 (5.6)	19 (2.5)	2 (0.9)
Gastrointestinal symptoms	631 (21.8)	95 (7.6)	350 (100.0)	0 (0.0)	14 (11.3)	123 (16.0)	49 (21.0)
Liver symptoms	11 (0.4)	8 (0.6)	2 (0.6)	0 (0.0)	1 (0.8)	0 (0.0)	0 (0.0)
Neurological / neuromuscular Symptoms	131 (4.5)	90 (7.2)	0 (0.0)	0 (0.0)	4 (3.2)	36 (4.7)	1 (0.4)
Musculoskeletal Symptoms	8 (0.3)	4 (0.3)	1 (0.3)	0 (0.0)	1 (0.8)	1 (0.1)	1 (0.4)
Other symptoms on admission	116 (4.0)	86 (6.9)	11 (3.1)	1 (0.6)	2 (1.6)	11 (1.4)	5 (2.1)
Unknown symptoms on admission	16 (0.6)	16 (1.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Comorbidities at the time of COVID-19 infection							
Respiratory disease	53 (1.8)	2 (0.2)	4 (1.1)	9 (5.4)	20 (16.1)	13 (1.7)	5 (2.1)
Cardiovascular disease	88 (3.0)	23 (1.8)	2 (0.6)	9 (5.4)	21 (16.9)	21 (2.7)	12 (5.2)
Gastrointestinal tract disease	32 (1.1)	6 (0.5)	3 (0.9)	5 (3.0)	4 (3.2)	10 (1.3)	4 (1.7)
Liver disease	25 (0.9)	7 (0.6)	0 (0.0)	11 (6.6)	4 (3.2)	3 (0.4)	0 (0.0)
Kidney disease	37 (1.3)	17 (1.4)	1 (0.3)	5 (3.0)	7 (5.6)	3 (0.4)	4 (1.7)
Neurological/neuromuscular disease	38 (1.3)	6 (0.5)	2 (0.6)	8 (4.8)	11 (8.9)	7 (0.9)	4 (1.7)
Psychiatric disease	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)	0 (0.0)
Hematologic disease	19 (0.7)	9 (0.7)	1 (0.3)	2 (1.2)	2 (1.6)	3 (0.4)	2 (0.9)
Oncological disease	4 (0.1)	3 (0.2)	0 (0.0)	1 (0.6)	0 (0.0)	0 (0.0)	0 (0.0)
Organ or bone marrow/stem cell transplantation	2 (0.1)	0 (0.0)	0 (0.0)	2 (1.2)	0 (0.0)	0 (0.0)	0 (0.0)

Autoimmunological disease	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.4)
Congenital immunodeficiency	1 (0.0)	1 (0.1)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Tracheostoma (prior to current infection)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Other concomitant disease	349 (12.1)	35 (2.8)	6 (1.7)	32 (19.3)	18 (14.5)	31 (4.0)	227 (97.4)
Coinfection							
Pulmonary viral infection	62 (2.1)	5 (0.4)	8 (2.3)	0 (0.0)	0 (0.0)	45 (5.8)	4 (1.7)
Pulmonary bacterial infection	11 (0.4)	1 (0.1)	0 (0.0)	1 (0.6)	1 (0.8)	7 (0.9)	1 (0.4)
Non-pulmonary bacterial infection	121 (4.2)	75 (6.0)	3 (0.9)	13 (7.8)	4 (3.2)	17 (2.2)	9 (3.9)
Non-pulmonary viral infection	38 (1.3)	7 (0.6)	21 (6.0)	1 (0.6)	1 (0.8)	4 (0.5)	4 (1.7)
COVID-19 risk factors							
Home oxygen or ventilation therapy before the current disease	37 (1.3)	3 (0.2)	7 (2.0)	4 (2.4)	9 (7.3)	9 (1.2)	5 (2.1)
preterm infant	240 (8.3)	12 (1.0)	30 (8.6)	35 (21.1)	122 (98.4)	24 (3.1)	17 (7.3)
Exposure to smoking	69 (2.4)	39 (3.1)	0 (0.0)	3 (1.8)	1 (0.8)	22 (2.9)	4 (1.7)
Immunosuppression	4 (0.1)	2 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	2 (0.9)
Other COVID-19 risk factors	136 (4.7)	14 (1.1)	2 (0.6)	11 (6.6)	11 (8.9)	15 (1.9)	83 (35.6)

Table S7 Characteristics of non-infants by phenotypes

Characteristics	n (%)						
	Total sample (n=4091)	Phenotype A (n=1522)	Phenotype B (n=720)	Phenotype C (n=507)	Phenotype D (n=371)	Phenotype E (n=439)	Phenotype F (n=532)
Sex=Female	1962 (48.0)	746 (49.0)	405 (56.2)	229 (45.2)	159 (42.9)	185 (42.1)	238 (44.7)
COVID-19 symptoms on admission							
No symptoms (asymptomatic)	537 (13.1)	30 (2.0)	0 (0.0)	507 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)
General symptoms	2417 (59.1)	1076 (70.7)	399 (55.4)	0 (0.0)	242 (65.2)	289 (65.8)	411 (77.3)
Ear, nose, and throat symptoms	856 (20.9)	224 (14.7)	105 (14.6)	0 (0.0)	96 (25.9)	422 (96.1)	9 (1.7)
Lower respiratory tract symptoms	1319 (32.2)	656 (43.1)	147 (20.4)	0 (0.0)	228 (61.5)	223 (50.8)	65 (12.2)
Cardiovascular symptoms	172 (4.2)	129 (8.5)	3 (0.4)	0 (0.0)	20 (5.4)	13 (3.0)	7 (1.3)
Gastrointestinal symptoms	1253 (30.6)	266 (17.5)	682 (94.7)	0 (0.0)	123 (33.2)	47 (10.7)	135 (25.4)
Liver symptoms	18 (0.4)	8 (0.5)	5 (0.7)	0 (0.0)	1 (0.3)	3 (0.7)	1 (0.2)
Neurological / neuromuscular Symptoms	925 (22.6)	145 (9.5)	46 (6.4)	0 (0.0)	83 (22.4)	119 (27.1)	532 (100.0)
Musculoskeletal Symptoms	192 (4.7)	154 (10.1)	4 (0.6)	0 (0.0)	20 (5.4)	5 (1.1)	9 (1.7)
Other symptoms on admission	306 (7.5)	277 (18.2)	9 (1.2)	0 (0.0)	12 (3.2)	4 (0.9)	4 (0.8)
Unknown symptoms on admission	44 (1.1)	44 (2.9)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Comorbidities at the time of COVID-19 infection							
Respiratory disease	242 (5.9)	165 (10.8)	15 (2.1)	14 (2.8)	41 (11.1)	4 (0.9)	3 (0.6)
Cardiovascular disease	173 (4.2)	99 (6.5)	6 (0.8)	14 (2.8)	39 (10.5)	4 (0.9)	11 (2.1)
Gastrointestinal tract disease	161 (3.9)	60 (3.9)	44 (6.1)	28 (5.5)	22 (5.9)	3 (0.7)	4 (0.8)
Liver disease	40 (1.0)	17 (1.1)	7 (1.0)	5 (1.0)	11 (3.0)	0 (0.0)	0 (0.0)
Kidney disease	108 (2.6)	63 (4.1)	13 (1.8)	18 (3.6)	6 (1.6)	3 (0.7)	5 (0.9)
Neurological/neuromuscular disease	407 (9.9)	163 (10.7)	30 (4.2)	37 (7.3)	75 (20.2)	19 (4.3)	83 (15.6)
Psychiatric disease	110 (2.7)	27 (1.8)	19 (2.6)	33 (6.5)	7 (1.9)	15 (3.4)	9 (1.7)
Hematologic disease	136 (3.3)	71 (4.7)	13 (1.8)	17 (3.4)	20 (5.4)	11 (2.5)	4 (0.8)
Oncological disease	102 (2.5)	91 (6.0)	5 (0.7)	3 (0.6)	1 (0.3)	2 (0.5)	0 (0.0)
Organ or bone marrow/stem cell transplantation	37 (0.9)	32 (2.1)	0 (0.0)	3 (0.6)	0 (0.0)	1 (0.2)	1 (0.2)

Autoimmunological disease	135 (3.3)	84 (5.5)	6 (0.8)	23 (4.5)	19 (5.1)	1 (0.2)	2 (0.4)
Congenital immunodeficiency	27 (0.7)	13 (0.9)	5 (0.7)	1 (0.2)	5 (1.3)	2 (0.5)	1 (0.2)
Tracheostoma (prior to current infection)	18 (0.4)	17 (1.1)	0 (0.0)	1 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)
Other concomitant disease	616 (15.1)	140 (9.2)	39 (5.4)	63 (12.4)	360 (97.0)	11 (2.5)	3 (0.6)
Coinfection							
Pulmonary viral infection	69 (1.7)	62 (4.1)	1 (0.1)	3 (0.6)	2 (0.5)	1 (0.2)	0 (0.0)
Pulmonary bacterial infection	70 (1.7)	46 (3.0)	2 (0.3)	5 (1.0)	12 (3.2)	4 (0.9)	1 (0.2)
Non-pulmonary bacterial infection	210 (5.1)	60 (3.9)	81 (11.2)	36 (7.1)	18 (4.9)	15 (3.4)	0 (0.0)
Non-pulmonary viral infection	98 (2.4)	29 (1.9)	52 (7.2)	3 (0.6)	11 (3.0)	2 (0.5)	1 (0.2)
COVID-19 risk factors							
Home oxygen or ventilation therapy before the current disease	74 (1.8)	59 (3.9)	1 (0.1)	1 (0.2)	11 (3.0)	1 (0.2)	1 (0.2)
preterm infant	117 (2.9)	88 (5.8)	11 (1.5)	5 (1.0)	7 (1.9)	4 (0.9)	2 (0.4)
Exposure to smoking	157 (3.8)	87 (5.7)	11 (1.5)	22 (4.3)	14 (3.8)	15 (3.4)	8 (1.5)
Immunosuppression	145 (3.5)	123 (8.1)	5 (0.7)	6 (1.2)	7 (1.9)	0 (0.0)	4 (0.8)
Other COVID-19 risk factors	329 (8.0)	43 (2.8)	4 (0.6)	25 (4.9)	245 (66.0)	7 (1.6)	5 (0.9)

Figure S1 Dendrogram of hierarchical agglomerative clustering

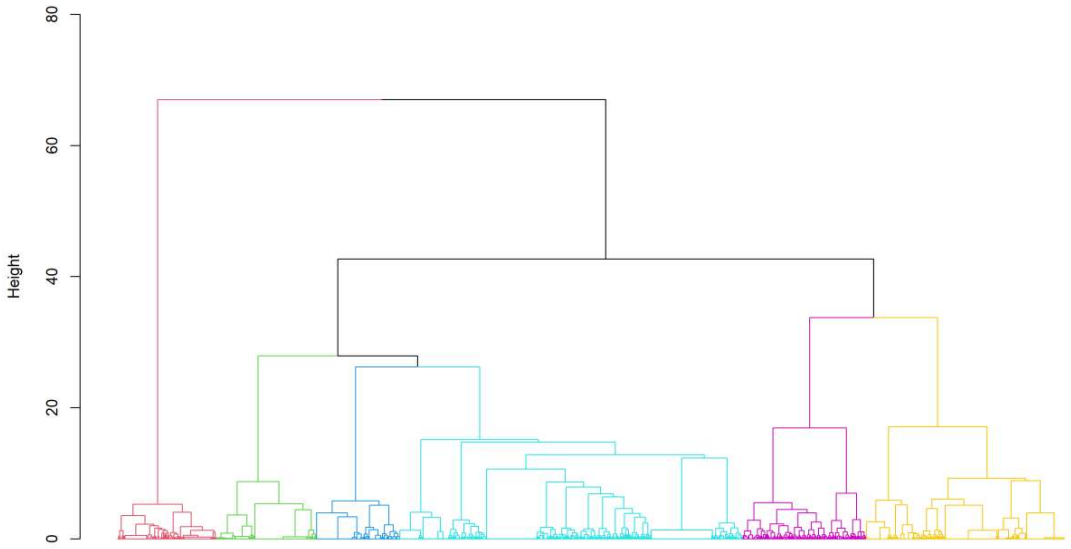
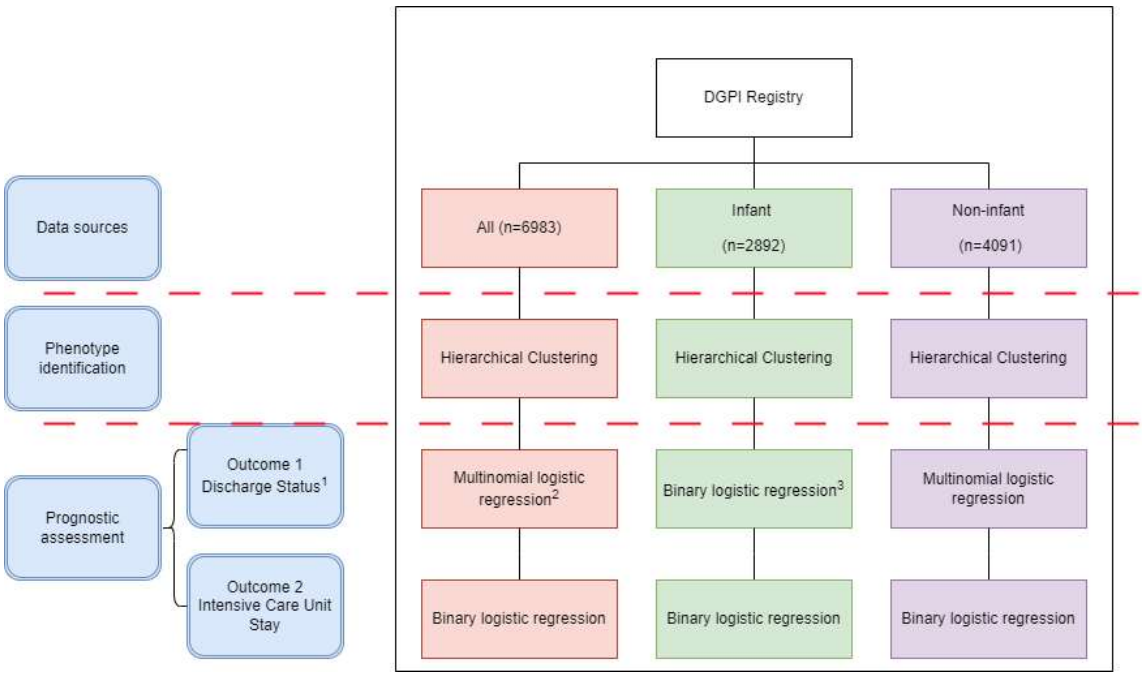


Figure S1 displays the dendrogram resulting from hierarchical clustering, employing Gower' distance and Ward's linkage, utilizing 35 patient characteristics. The y-axis of the dendrogram represents the distance used to cluster the objects ("height"). Initially, each observation is treated as its own cluster, and as the process proceeds in an agglomerative manner, similar objects are grouped together. The height on the y-axis increases as the number of clusters decreases, signifying increased heterogeneity among clusters. Finally, six distinct colors represent the six clusters determined.

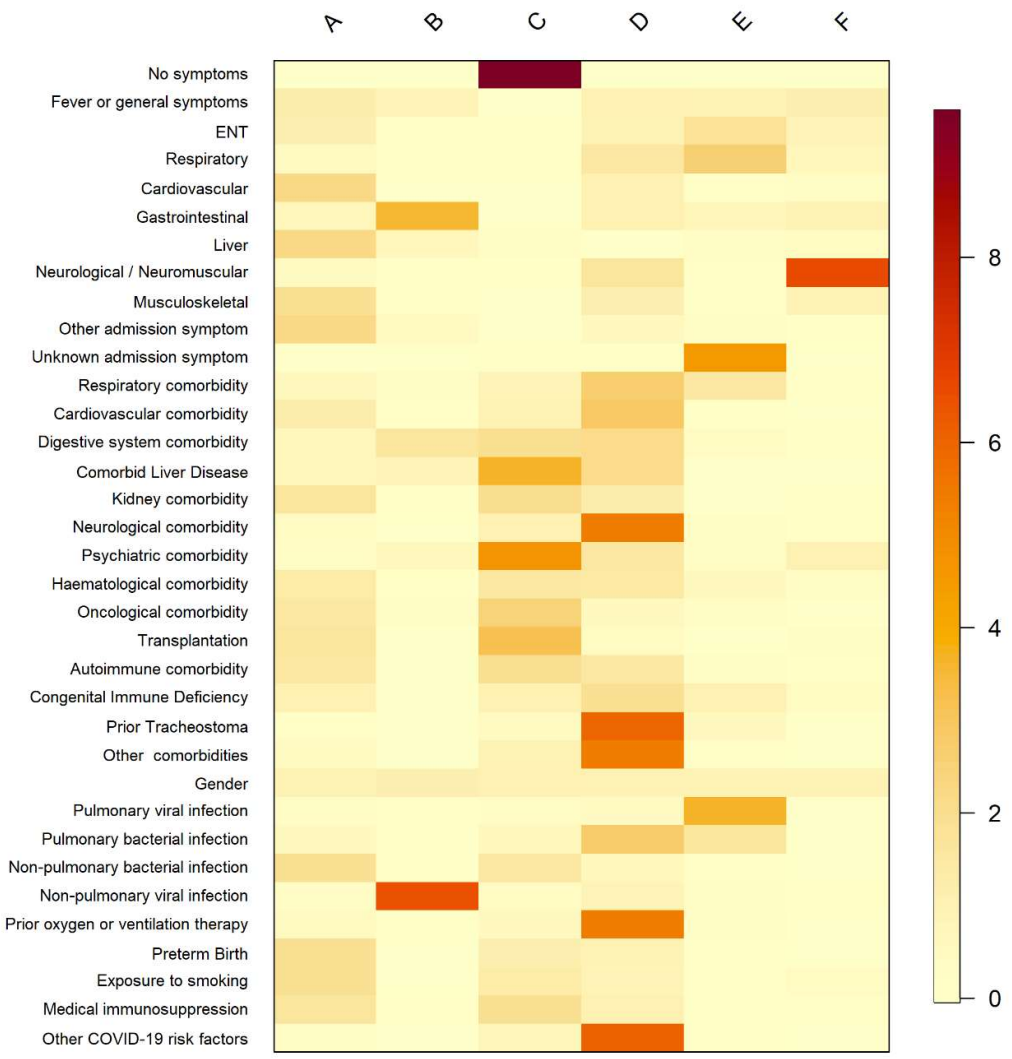
[REFERENCE]
Wang M, Flexeder C, Harris CP, et al. Accelerometry-assessed sleep clusters and cardiometabolic risk factors in adolescents. Obesity (Silver Spring). 2024; 32(1): 200-213.

Figure S2 Data analysis diagram



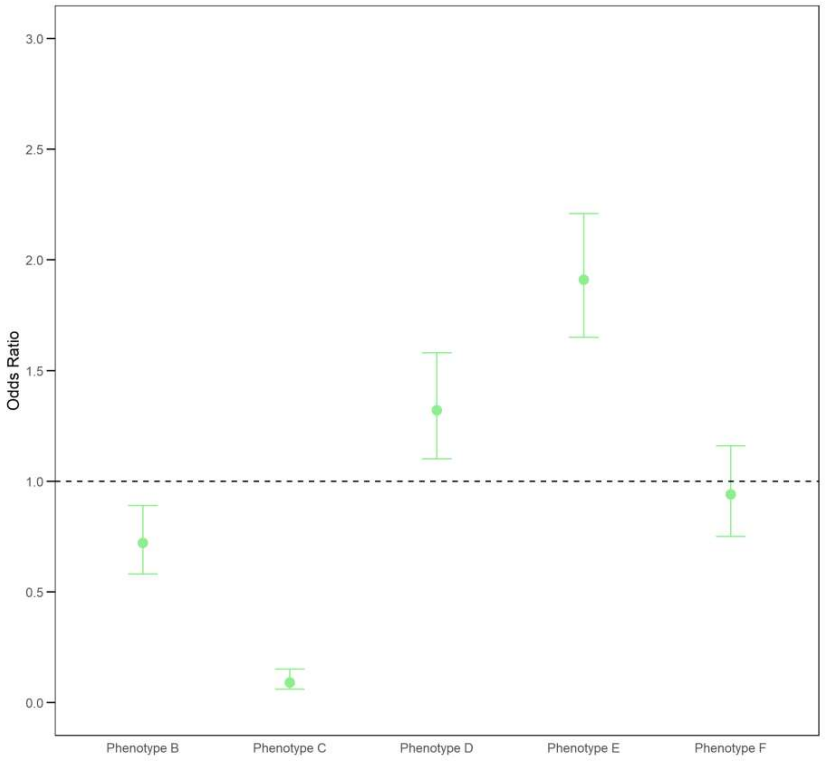
1. Discharge Status included: full recovery, residual symptoms, and unfavorable prognosis; full recovery was used as the reference; patients whose discharge status was transferal or non-SARS-CoV-2-related death were excluded for discharge status assessment. 2. Since no patient with phenotype B had an unfavorable prognosis, we utilized two methods of handling phenotype B. For the main model, we excluded patients with phenotype B, and evaluated the associations between other phenotypes and discharge status (including full recovery, residual symptoms, and unfavorable prognosis) in the model. As a contrast, we excluded patients whose discharge status were "unfavorable prognosis", and evaluated the associations between all phenotypes and discharge status (including full recovery and residual symptoms) with a binary logistic regression model. 3. Since only 5 patients had an unfavorable prognosis in infants, we decided to only compare full recovery and residual symptoms with binary logistic regression in this group.

Figure S3 Heatmap of patient characteristics by clinical phenotypes



This heatmap shows the difference of percentages of variables defining phenotypes in each phenotype in relation to all the registered population.

Figure S4 Risk association between phenotypes and residual symptoms in all registered population



In this model, we excluded patients whose discharge status were “unfavorable prognosis”, and evaluated the associations between all phenotypes and discharge status (including full recovery and residual symptoms) with a binary logistic regression model.

4. Publication II

PERSISTENT SYMPTOMS AND RISK FACTORS PREDICTING PROLONGED TIME TO SYMPTOM-FREE AFTER SARS-COV-2 INFECTION: AN ANALYSIS OF THE BASELINE EXAMINATION OF THE GERMAN COVIDOM/NAPKON-POP COHORT

Yanyan Shi, Ralf Strobl, Christian Apfelbacher, Thomas Bahmer, Ramsia Geisler, Peter Heuschmann, Anna Horn, Hanno Hoven, Thomas Keil, Michael Krawczak, Lilian Krist, Christina Lemhöfer, Wolfgang Lieb, Bettina Lorenz-Depiereux, Rafael Mikolajczyk, Felipe A. Montellano, Jens Peter Reese, Stefan Schreiber, Nicole Skoetz, Stefan Störk, Jörg Janne Vehreschild, Martin Witzenrath, Eva Grill, on behalf of the NAPKON Study Group

Infection

Received 09 March 2023
Accepted 24 April 2023
Published 25 May 2023
DOI: 10.1007/s15010-023-02043-6

Complete Citation

Shi, Y., Strobl, R., Apfelbacher, C. *et al.* Persistent symptoms and risk factors predicting prolonged time to symptom-free after SARS-CoV-2 infection: an analysis of the baseline examination of the German COVIDOM/NAPKON-POP cohort. *Infection* **51**, 1679–1694 (2023). https://doi.org/10.1007/s15010-023-02043-6

Infection (2023) 51:1679–1694
https://doi.org/10.1007/s15010-023-02043-6

RESEARCH

Persistent symptoms and risk factors predicting prolonged time to symptom-free after SARS-CoV-2 infection: an analysis of the baseline examination of the German COVIDOM/NAPKON-POP cohort

Yanyan Shi[1,2] · Ralf Strobl[1,3] · Christian Apfelbacher[4] · Thomas Bahmer[5] · Ramsia Geisler[6] · Peter Heuschmann[7,8] · Anna Horn[7] · Hanno Hoven[9] · Thomas Keil[7,10,11] · Michael Krawczak[12] · Lilian Krist[10] · Christina Lemhöfer[13] · Wolfgang Lieb[14] · Bettina Lorenz-Depiereux[15] · Rafael Mikolajczyk[16,17] · Felipe A. Montellano[7,18] · Jens Peter Reese[7] · Stefan Schreiber[5] · Nicole Skoetz[19] · Stefan Störk[20] · Jörg Janne Vehreschild[6,21,22] · Martin Witzenrath[23,24] · Eva Grill[1,3] on behalf of the NAPKON Study Group

Received: 9 March 2023 / Accepted: 24 April 2023 / Published online: 25 May 2023

Abstract

Purpose We aimed to assess symptoms in patients after SARS-CoV-2 infection and to identify factors predicting prolonged time to symptom-free.

Methods COVIDOM/NAPKON-POP is a population-based prospective cohort of adults whose first on-site visits were scheduled ≥ 6 months after a positive SARS-CoV-2 PCR test. Retrospective data including self-reported symptoms and time to symptom-free were collected during the survey before a site visit. In the survival analyses, being symptom-free served as the event and time to be symptom-free as the time variable. Data were visualized with Kaplan–Meier curves, differences were tested with log-rank tests. A stratified Cox proportional hazard model was used to estimate adjusted hazard ratios (aHRs) of predictors, with aHR < 1 indicating a longer time to symptom-free.

Results Of 1175 symptomatic participants included in the present analysis, 636 (54.1%) reported persistent symptoms after 280 days (SD 68) post infection. 25% of participants were free from symptoms after 18 days [quartiles: 14, 21]. Factors associated with prolonged time to symptom-free were age 49–59 years compared to < 49 years (aHR 0.70, 95% CI 0.56–0.87), female sex (aHR 0.78, 95% CI 0.65–0.93), lower educational level (aHR 0.77, 95% CI 0.64–0.93), living with a partner (aHR 0.81, 95% CI 0.66–0.99), low resilience (aHR 0.65, 95% CI 0.47–0.90), steroid treatment (aHR 0.22, 95% CI 0.05–0.90) and no medication (aHR 0.74, 95% CI 0.62–0.89) during acute infection.

Conclusion In the studied population, COVID-19 symptoms had resolved in one-quarter of participants within 18 days, and in 34.5% within 28 days. Over half of the participants reported COVID-19-related symptoms 9 months after infection. Symptom persistence was predominantly determined by participant's characteristics that are difficult to modify.

Keywords COVID-19 · Long COVID · Post-COVID syndrome · Time to symptom-free · Risk factors

Introduction

As of December 2022, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection has been confirmed in over 600 million people worldwide [1]. Many patients, even those with mild-to-moderate acute symptoms, continue to suffer from symptoms after acute disease [2, 3]. "Long COVID" is increasingly used as an umbrella term for signs and symptoms persisting for 4 weeks or longer after SARS-CoV-2 infection [4].

The most frequently reported persisting symptoms include fatigue, dyspnea, sleep disorders or insomnia, headache, attention disorders, anosmia and ageusia [5–10]. A systematic review of 151 studies revealed that > 50% of COVID-19 patients still had at least one symptom 12 months after a confirmed infection [11]. However, generalizability to the general population is hampered by the fact that many studies

The members of the NAPKON Study Group are listed in Acknowledgements.

Extended author information available on the last page of the article

investigating persisting symptoms after SARS-CoV-2 infection were based on hospitalized patients whilst others drew upon small, selected samples, or lacked a sufficiently long follow-up period [12–16]. The ongoing German COVIDOM/NAPKON-POP population-based study included participants ≥ 6 months after a positive SARS-CoV-2 polymerase chain reaction (PCR) test, regardless of disease severity. Recently, some of us used the first results of this study [9] to develop a severity score to quantify the symptom load associated with post-COVID syndrome (PCS score), which is broadly synonymous with Long COVID. PCS score facilitates an objective assessment of the extent and severity of the condition in the general population. However, detailed information on the health burden of long COVID, specifically on the time to full recovery, remains scarce.

A study from the Netherlands reported a median time to complete recovery of 63 days among individuals with mild, and 232 days among individuals with moderate disease severity [17]. A large international online survey of patients with suspected and confirmed SARS-CoV-2 infection revealed that the probability of time to recovery from symptoms exceeding 35 weeks was 91.8% [18]. Most eminent risk factors for Long COVID were the presence or number of existing comorbidities [2, 17, 19], however, results on risks of individual comorbidities were inconsistent [13, 20–22]. Treatment during acute infection such as steroid or antibiotic medication was not indicative of a complete recovery [23]. Up to date, the time course of COVID-19 symptoms and factors associated with time to recovery are thus still incompletely understood.

Using COVIDOM/NAPKON-POP baseline data, we aimed to retrospectively assess the time course of symptom persistence after SARS-CoV-2 infection. We also investigated factors predicting prolonged time to complete recovery (i.e., to becoming symptom-free) in this multi-center population-based study covering three regions of Germany.

Methods

Study design

The National Pandemic Cohort Study Network ("Nationales Pandemie Kohorten Netz", NAPKON) was established in Germany in 2020 to coordinate and harmonize COVID-19 research at a nation-wide level [24]. NAPKON-POP is the population-based platform that hosts the COVIDOM study aimed at investigating the long-term consequences of COVID-19. Participants in COVIDOM/NAPKON-POP were recruited at three study sites in Germany, namely Kiel, Würzburg, and the Neukölln district of Berlin, covering defined geographical regions in the vicinity.

Participants

All eligible individuals were identified through the mandatory registration of a positive SARS-CoV-2 PCR test by local health authorities. First on-site visits of prospective participants were scheduled ≥ 6 months post PCR test, regardless of their acute disease severity, following procedures detailed elsewhere [25]. Inclusion criteria of participants were: (a) positive PCR for SARS-CoV-2 ≥ 6 months before enrollment, (b) living in one of the three covered regions, (c) ≥ 18 years of age, and (d) written informed consent. Exclusion criterion was an acute SARS-CoV-2 re-infection at the time of the initial questionnaire, or at the scheduled site visit [25]. Recruitment and follow-up of the COVIDOM/NAPKON-POP cohort are still ongoing. For the present analysis, data from participants recruited between November 2020 and September 2021 were used, and only symptomatic participants were included.

Method of data collection

Retrospective data on the acute course of COVID-19, time to symptom-free and current symptoms were collected from self-filled questionnaires before the on-site visit. Later, participants were assessed at the study sites during enrollment into the prospective cohort study, collecting data on body measurement, resilience, COVID-19 treatment, comorbidities, and lifestyles by physical examination, questionnaires, and interviews [25].

Measures

Symptoms

COVID-19-related symptoms were assessed by a self-selection from 22 specific symptoms and "other symptoms" [9]. Participants were asked whether they experienced these symptoms in either the infection/acute period or at the time of the survey ("current symptoms"). Fatigue was considered present when the free-text answer to the prompting question following "other symptoms" contained "fatigue" or its synonyms. A list of all 23 symptoms is provided in Fig. 1. Presence of current symptoms was assessed by the question "Do you still have symptoms currently?".

Time to symptom-free

Time to symptom-free was assessed using the question: "How long did it take you to become symptom-free after the occurrence of first symptoms?" Time to symptom-free was measured as the time from the first appearance

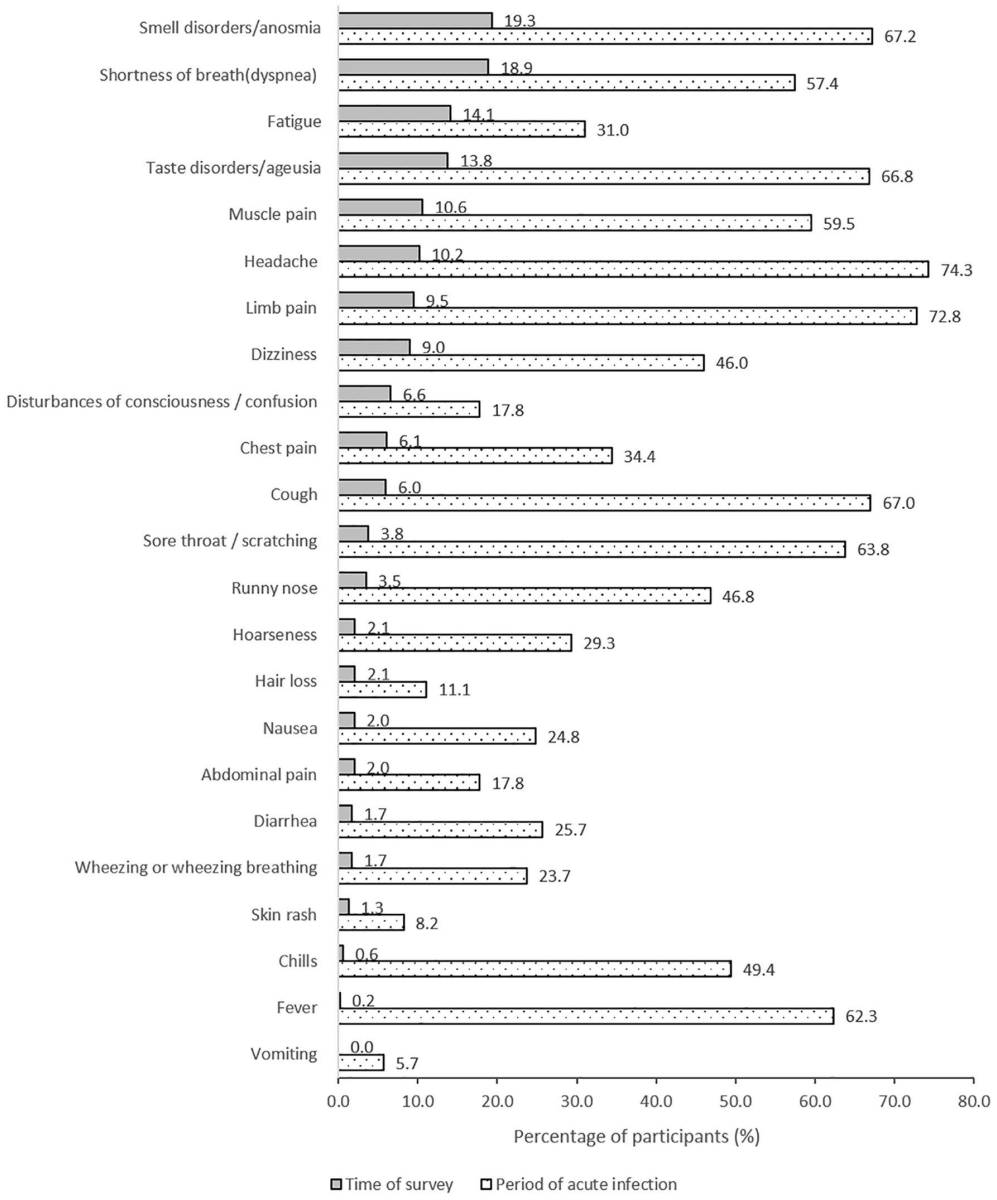


Fig. 1 COVID-19 related symptoms during acute infection and time of survey ($N = 1175$)

of symptoms to symptom-free status in days, weeks or months, re-scaled to days (7 days per week and 30 days per month) for the purpose of the present study.

For those still experiencing symptoms at the time of the survey, time to be symptom-free was considered as censored and was calculated as the time between the appearance of the first symptoms and the survey.

Additionally, we tested for group differences up to 28 days (i.e. before becoming a Long COVID case) by manually censoring data at this time point. In detail, we set the symptom-free time to 28 days and the symptom status to "experiencing symptoms" whenever getting symptom-free took longer than 28 days.

Alcohol consumption

Alcohol consumption was categorized as abstainers, low-risk alcohol consumption, or risky alcohol consumption (i.e. ≥ 5 times per week, or consumption on one occasion ≥ 4 or ≥ 5 glasses for women and men, respectively) [26].

Body Mass Index (BMI)

BMI was calculated from the weight and height measurements taken at the study site with the formula BMI = kg/m^2 and was categorized as: underweight (BMI < 18.5), normal (18.5 ≤ BMI < 25), pre-obese (25 ≤ BMI < 30), or obese (BMI ≥ 30) [27].

Resilience

Resilience was measured by the 6-item Brief Resilience Scale and was categorized as: low (1.00–2.99), normal (3.00–4.30), and high (4.31–5.00). The Brief Resilience Scale can be found in Supplementary Appendix (S Table 1).

COVID-19 treatment

COVID-19 treatment was assessed by the question: "Have you taken any medications for SARS-Cov-2 infection?" together with prompting three treatment categories of steroids, anticoagulation, and anti-infectives. In the present analysis, we merged corticosteroids, steroids (> 0.5 mg/kg prednisone equivalents) and steroids (≤ 0.5 mg/kg prednisone equivalents) into one variable "steroids".

Comorbidities

Comorbidities were self-reported physician-diagnosed diseases. (Detailed in Table 1).

Statistical analysis

Mean, with standard deviation (SD), or median with quartiles were used for the description of continuous variables. Counts and percentages were used for the description of categorical variables.

In the survival analysis, being symptom-free served as the event and time to be symptom-free as the time variable. Since < 50% of symptomatic participants were symptom-free at the time of investigation, we reported the Q1 (25%) time to symptom-free, instead of the median time. Kaplan–Meier estimator served to estimate the survival function and Kaplan–Meier plots served to visualize the survival curves. Log-rank tests were used to test group differences in both overall survival curves and in survival curves up to 28 days.

Missing data were imputed by Multiple Imputation by Chained Equations (MICE) [28], yielding ten imputed datasets. Imputation was based on age, sex, educational level, living status, smoking, alcohol consumption, symptom burden during acute infection, BMI, COVID-19 treatment during acute infection, chronic liver disease, chronic rheumatologic/immunologic disease, tumor/cancer disease, chronic neurological disease, lung disease, ear, nose and throat (ENT) disease, cardiovascular disease, and diabetes. The final model was combined with Rubin's rules, calculating final coefficient as the mean of coefficients estimated from imputed datasets and calculating the variance of estimated coefficients by factoring in the within and between imputation variance [29].

We applied a stratified Cox proportional hazard regression model to explore the factors predicting prolonged time to symptom-free after infection. Proportional hazard (PH) assumption was assessed with the Schoenfeld test [30]. Predictors violating the PH assumption were included as a stratified parameter in the multivariable Cox model [30]. By including a variable as a stratified parameter, the stratified Cox proportional hazard model sets a different baseline hazard corresponding to each stratum as defined by the variable, and then estimates common coefficients for the remaining explanatory variables except for the stratified variable, thus providing hazard ratios controlled for the effect of the stratification variable, but not for the stratification variable itself [30]. Symptom burden and hospitalization both violated the PH assumption and both are closely related to unmeasured disease severity during the acute infection phase. Since only 75 (6.4%) of all patients were hospitalized, we decided to only include symptom burden as a stratification parameter and analyzed the effect of hospitalization in a separate sensitivity analysis (see below). A Generalized Variance Inflation Factor (GVIF) was used to check for multicollinearity among covariates, GVIF$^{1/(2*Df)}$ of ≥ 5 was considered indicative of collinearity [31]. Stepwise variable selection was conducted, selecting the model with the smallest Akaike information criterion. To assess the linearity assumption, we plotted the Martingale residuals against covariates. The adjusted hazard ratios (aHRs) were used to describe the hazard of becoming symptom-free, with aHR < 1 indicating a longer time to symptom free. A multivariate Wald test was used to assess the overall significance of difference for categorical variables with more than three categories. The concordance index (*C*-index) was used to measure the goodness-of-fit of the fitted models with ten imputed datasets; it measures the agreement between observed survival and

Table 1 Characteristics of the final sample and asymptomatic participants

Characteristics	*n* (%)		*P* value
	Symptomatic participants (*n* = 1175)	Asymptomatic participants (*n* = 108)	
Age (years)			< 0.001*
< 49	589 (50.1)	48 (44.4)	
49–59	346 (29.4)	27 (25.0)	
≥ 60	236 (20.1)	26 (24.1)	
Missings	4 (0.3)	7 (6.5)	
Sex			0.0538
Female	659 (56.1)	48 (44.4)	
Male	515 (43.8)	60 (55.6)	
Missings	1 (0.1)	0 (0.0)	
Nationality			< 0.001*
German	1143 (97.3)	63 (58.3)	
Non-German	29 (2.5)	4 (3.7)	
Missings	3 (0.3)	41 (38.0)	
Educational level			< 0.001*
University entrance certificate	665 (56.6)	33 (30.6)	
Lower education	498 (42.4)	32 (29.6)	
Missings	12 (1.0)	43 (39.8)	
Living status			< 0.001*
Living with a partner	820 (69.8)	46 (42.6)	
No partner/not living with a partner	287 (24.4)	19 (17.6)	
Missings	68 (5.8)	43 (39.8)	
Smoking status			< 0.001*
Current-smokers	143 (12.2)	11 (10.2)	
Ex-smokers	436 (37.1)	18 (16.7)	
Non-smokers	587 (50.0)	32 (29.6)	
Missings	9 (0.8)	47 (43.5)	
Alcohol consumption			0.4274
Abstainer	101 (8.6)	12 (11.1)	
Low-risk alcohol consumption	605 (51.5)	49 (45.4)	
Risky alcohol consumption	147 (12.5)	18 (16.7)	
Missings	322 (27.4)	29 (26.9)	
Hospitalization during acute infection			0.8953
Hospitalized	75 (6.4)	6 (5.6)	
Non-hospitalized	1100 (93.6)	102 (94.4)	
Symptom burden during acute infection			< 0.001*
No symptom	0 (0.0)	108 (100.0)	
1–5 symptoms	200 (17.0)	0 (0.0)	
≥ 6 symptoms	975 (83.0)	0 (0.0)	
Body mass index			0.7529
Normal	465 (39.6)	38 (35.2)	
Obese	282 (24.0)	29 (26.9)	
Pre-obese	416 (35.4)	41 (38.0)	
Underweight	10 (0.9)	0 (0.0)	
Missings	2 (0.2)	0 (0.0)	
Resilience			0.0523
Low resilience	212 (18.0)	14 (13.0)	
Normal resilience	690 (58.7)	58 (53.7)	
High resilience	163 (13.9)	18 (16.7)	
Missings	110 (9.4)	18 (16.7)	

Table 1 (continued)

Characteristics	*n* (%)		*P* value
	Symptomatic participants (*n* = 1175)	Asymptomatic participants (*n* = 108)	
COVID-19 treatment			
Treated with medication	641 (54.6)	29 (26.9)	< 0.001*
Antipyretics	540 (46.0)	24 (22.2)	< 0.001*
Missings	17 (1.4)	3 (2.8)	
Steroids	20 (1.7)	0 (0.0)	0.2738
Missings	13 (1.1)	2 (1.9)	
Anticoagulation	64 (5.4)	3 (2.8)	0.3199
Missings	13 (1.1)	2 (1.9)	
Anti-infectives	49 (4.2)	3 (2.8)	0.6167
Comorbidities			
Number of comorbidities			0.5658
0	403 (34.3)	40 (37.0)	
1	364 (31.0)	36 (33.3)	
≥ 2	408 (34.7)	32 (29.6)	
Chronic liver disease	116 (9.9)	11 (10.2)	0.3305
Missings	117 (10.0)	6 (5.6)	
Chronic rheumatologic/immunologic disease	104 (8.9)	7 (6.5)	0.6454
Missings	16 (1.4)	2 (1.9)	
Tumor/cancer disease	21 (1.8)	2 (1.9)	1.0000
Missings	4 (0.3)	0 (0.0)	
Chronic neurological disease	307 (26.1)	23 (21.3)	0.1306
Missings	12 (1.0)	3 (2.8)	
Lung disease	226 (19.2)	16 (14.8)	0.0165*
Missings	13 (1.1)	5 (4.6)	
Ear, nose and throat disease	290 (24.7)	23 (21.3)	0.1650
Missings	24 (2.0)	5 (4.6)	
Cardiovascular disease	346 (29.4)	30 (27.8)	0.0368*
Missings	14 (1.2)	5 (4.6)	
Diabetes	46 (3.9)	5 (4.6)	< 0.001*
Missings	5 (0.4)	47 (43.5)	
Current symptoms			
Symptom-free	539 (45.9)		
Persistent symptoms	636 (54.1)		

P value: Pearson χ^2 test (or Fisher exact test if expected $n < 5$)

*$P < 0.05$

predicted survival, with a value of 0.5 representing a random prediction and a value of 1.0 representing the best possible model prediction [32].

The threshold for statistical significance was set to 0.05. Since this was an exploratory study, no correction for multiple testing was applied. We used R (*version 4.1.1*) with the *dplyr, survival, car, MASS*, and *mice* packages for all statistical analyses. MS Office and R were used to create figures.

Sensitivity analyses

To evaluate the robustness of the final model, we conducted separate Cox proportional hazard models for each potential risk factor adjusted for age and sex. To investigate the effect of hospitalization on time to symptom-free we conducted three separate models: the first model only for patients having been hospitalized during acute infection, the second model for patients not having been hospitalized, and the third model including hospitalization with two different effect estimates, one for the effect in the first four weeks and one afterwards.

Results

Study participants

Data from 1441 COVIDOM/NAPKON-POP participants were available, including 1126 from Kiel, 208 from Würzburg, and 107 from Berlin. After excluding 90 cases with a time between PCR test and survey of < 6 months, and one case with an implausible PCR test date, 1350 participants were eligible for the present analysis. Of these, 108 participants had been asymptomatic during the acute phase, information on the current symptom status or the time to symptom-free of another 67 participants were missing. They were thus excluded from the analyses, resulting in a final sample of 1175 participants (Fig. 2).

Mean time since the onset of infection for 1175 participants was 280 days (SD 68). 54.1% of initially symptomatic participants continued to experience symptoms. Sex, BMI, resilience and most comorbidities of symptomatic participants were comparable to asymptomatic participants, whereas age, nationality, educational level, living status, smoking status, and COVID-19 treatment were not (Table 1).

Persistent COVID-19-related symptoms

At the time of survey, 22 of 23 different symptoms from the acute phase were still persistent: anosmia (19.3%), dyspnea (18.9%), fatigue (14.1%), and ageusia (13.8%) were the most common persisting symptoms. Muscle pain, headache, limb pain, dizziness, disturbances of consciousness/confusion, chest pain, and cough were reported by > 5% of participants each. Over 40% of participants had suffered from sore throat, fever, chills, and a runny nose during acute infection, while only < 5% reported these symptoms at the time of the survey, respectively (Fig. 1).

Time to symptom-free

Figure 3 and Table 2 summarize the observed bivariate differences in symptom persistence. Q1 time to symptom-free was 18 days [quartiles: 14 days, 21 days]. 405 (34.5%) participants had become symptom-free during the first 28 days since symptom onset, and only slow symptom resolution was seen afterwards. Time to symptom-free differed according to age, sex, educational level, living status, alcohol consumption, hospitalization during acute infection, symptom burden during acute infection, BMI, resilience, steroid treatment during acute infection, chronic liver disease, chronic rheumatologic/immunologic disease, chronic neurological

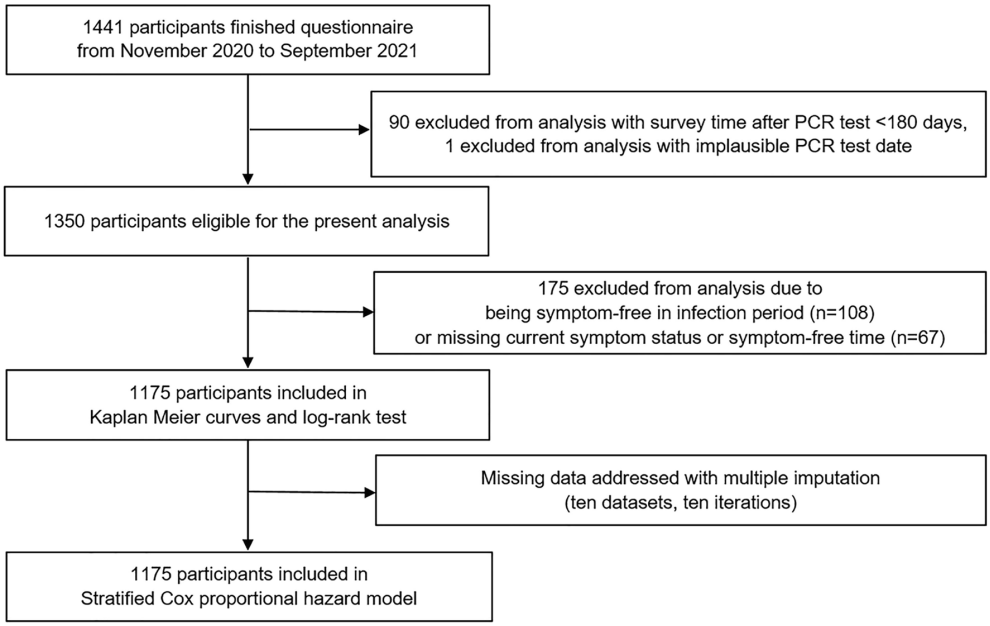


Fig. 2 Study profile

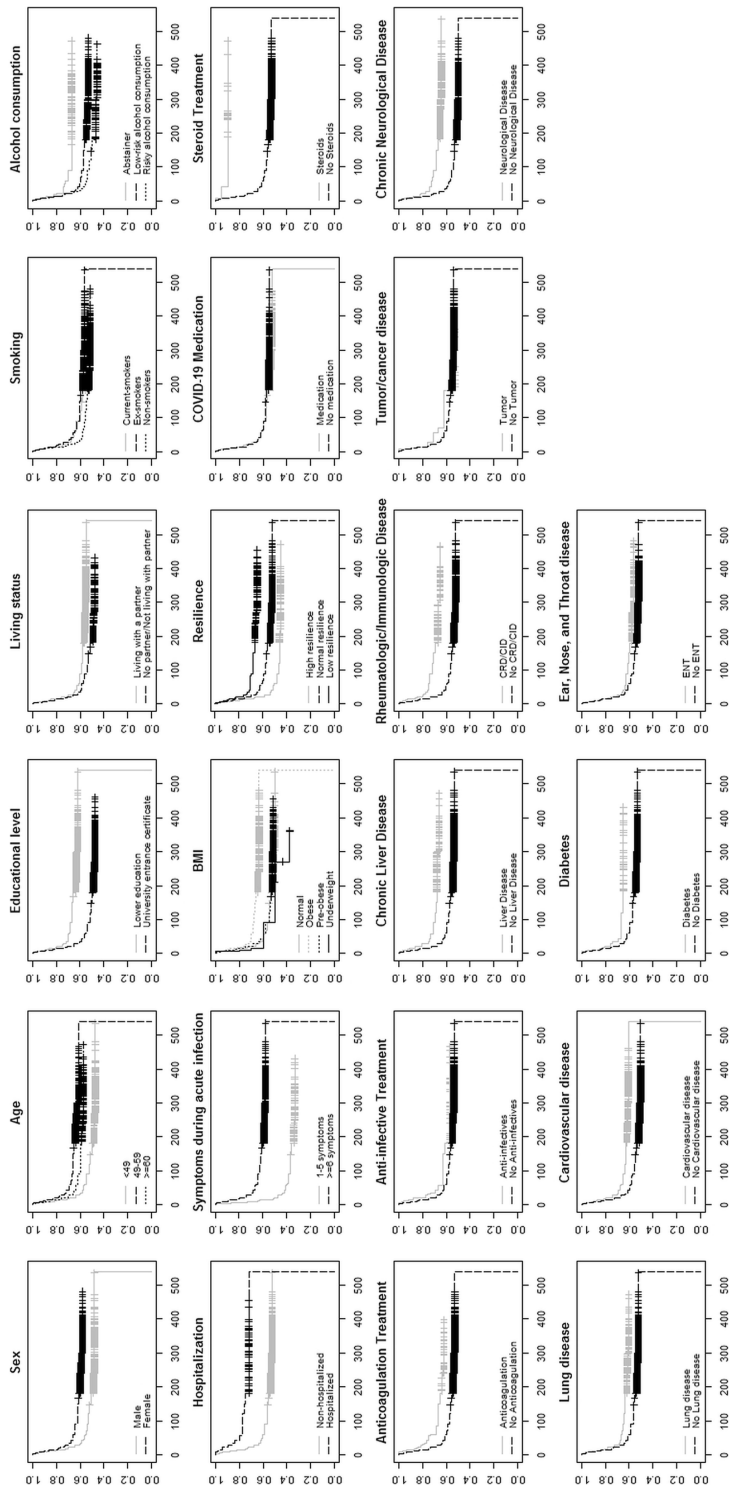


Fig. 3 Survival curves of time to symptom-free status for different patient groups ($N = 1175$). *X*-axis is the time to symptom-free in days, *y*-axis is the percentage of participants not reaching a symptom-free status. *CRD/CID*: chronic rheumatologic/immunologic disease

Table 2 Time to symptom-free status in patients stratified by patient characteristics ($N = 1175$)

Characteristics	Q1 time to symptom-free status	95% confidence interval	% of symptom-free patients 9 months after infection	Difference in survival curves**	
				Whole observation time	First 28 days
Age					
< 49	14	[14; 15]	52.3	< 0.001*	< 0.001*
49–59	28	[21; 42]	37.6		
≥ 60	20	[14; 28]	42.4		
Sex					
Female	21	[18; 28]	41.1	< 0.001*	0.0010*
Male	14	[14; 18]	51.8		
Educational level					
University entrance certificate	14	[14; 28]	52.0	< 0.001*	0.0003*
Lower education	21	[21; 35]	37.3		
Living status					
Living with a partner	21	[14; 21]	44.5	0.0295*	0.0972
No partner/Not living with a partner	14	[14; 20]	51.6		
Smoking status					
Current-smokers	17	[14; 42]	45.5	0.1584	0.0082*
Ex-smokers	21	[20; 28]	43.3		
Non-smokers	14	[14; 18]	48.0		
Alcohol consumption					
Abstainer	21	[18; NA]	32.7	0.0102*	0.0946
Low-risk alcohol consumption	17	[14; 21]	46.3		
Risky alcohol consumption	14	[14; 21]	53.7		
Hospitalization during acute infection					
Hospitalized	150	[42; NA]	29.3	< 0.001*	< 0.001*
Non-hospitalized	14	[14; 21]	47.0		
Symptom burden during acute infection					
1–5 symptoms	7	[6; 10]	66.5	< 0.001*	< 0.001*
≥ 6 symptoms	21	[21; 28]	41.6		
BMI					
Normal	14	[14; 21]	49.0	0.0037*	0.0648
Obese	21	[18; 60]	36.5		
Pre-obese	19	[14; 21]	48.1		
Underweight	10	[7; NA]	60.0		
Resilience					
Low resilience	38	[21; 90]	34.4	< 0.001*	< 0.001*
Normal resilience	17	[14; 21]	47.4		
High resilience	14	[10; 18]	54.6		
Treated with medication					
Yes	20	[14; 21]	46.6	0.8998	0.6708
No	14	[14; 21]	44.9		
Steroids					
Yes	NA	[NA; NA]	10.0	0.0040*	0.0107*
No	17	[14; 21]	46.6		
Anticoagulation					
Yes	49	[21; NA]	37.5	0.1005	0.0145*
No	17	[14; 21]	46.4		
Anti-infectives					

Table 2 (continued)

Characteristics	Q1 time to symptom-free status	95% confidence interval	% of symptom-free patients 9 months after infection	Difference in survival curves**	
				Whole observation time	First 28 days
Yes	30	[21; 180]	42.9	0.4359	0.1079
No	17	[14; 21]	46.0		
Chronic liver disease					
Yes	32.5	[21; NA]	32.8	0.0055*	0.0113*
No	14	[14; 21]	46.0		
Chronic rheumatologic/immunologic disease					
Yes	51	[21; NA]	33.7	0.0051*	0.0026*
No	14	[14; 21]	47.3		
Tumor/cancer diseases					
Yes	28	[10; NA]	47.6	0.9996	0.5793
No	18	[14; 21]	45.8		
Chronic neurological disease					
Yes	28	[21; 90]	35.2	< 0.001*	< 0.001*
No	14	[14; 20]	49.5		
Lung disease					
Yes	21	[18; 28]	38.9	0.0332*	0.2364
No	14	[14; 21]	47.3		
ENT disease					
Yes	21	[14; 28]	43.1	0.2100	0.4942
No	17	[14; 21]	47.4		
Cardiovascular disease					
Yes	21	[21; 28]	39.0	0.0019*	0.0323*
No	14	[14; 20]	48.7		
Diabetes					
Yes	30	[14; NA]	34.8	0.1553	0.1516
No	18	[14; 21]	46.3		

Q1: first quartile; number of days until 25% of participants became symptom-free

*$P < 0.05$

***P*-values were the result of the respective log-rank tests

disease, lung disease, and cardiovascular disease. Similar results were obtained when testing for group differences in survival curves up to 28 days, except for living status, smoking status, alcohol consumption, BMI, anticoagulation treatment and lung disease.

Prognostic analyses

Symptom burden during acute infection was included as a stratification variable in the final model because it violated the PH assumption. All GVIF were smaller than 5. Other variables included in the final model were age, sex, educational level, living status, alcohol consumption, BMI, resilience, COVID-19 medication and steroid treatment during acute infection, chronic liver disease, chronic rheumatologic/immunologic disease, and chronic neurological disease. The concordance indices of the ten fitted models ranged between 0.6305 and 0.6401.

Patients aged 49–59 years had a 30% lower hazard of becoming symptom-free than those aged < 49 years (aHR 0.70, 95% CI 0.56–0.87), while the hazard for patients ≥ 60 years did not differ from that < 49 years. Prolonged time to recovery was also seen in women (aHR 0.78, 95% CI 0.65–0.93), and patients with lower educational level (aHR 0.77, 95% CI 0.64–0.93), or living with a partner (aHR 0.81, 95% CI 0.66–0.99), or with low resilience (aHR 0.65, 95% CI 0.47–0.90). Steroid treatment (aHR 0.22, 95% CI 0.05–0.90) and no medication (aHR 0.74, 95% CI 0.62–0.89) during acute infection also increased time to symptom-free (Table 3).

Age and sex-adjusted coefficients for each potential risk factor can be found in the Supplementary Appendix

Table 3 Risk factors predicting prolonged time to symptom-free status in COVID-19 patients stratified by symptom burden during acute infection ($N = 1175$, stratified Cox proportional hazard model)

Covariates	Adjusted hazard ratio	95% confidence interval	*P* value	Overall *P* value
Age				
< 49	Reference			0.0053*
49–59	0.70	[0.56; 0.87]	0.0013*	
≥ 60	0.92	[0.72; 1.17]	0.4857	
Sex				
Male	Reference			NA
Female	0.78	[0.65; 0.93]	0.0073*	
Educational level				
University entrance certificate	Reference			NA
Lower education	0.77	[0.64; 0.93]	0.0062*	
Living status				
No partner/not living with a partner	Reference			NA
Living with a partner	0.81	[0.66; 0.99]	0.0382*	
Alcohol consumption				
Abstainer	Reference			0.1851
Low-risk alcohol consumption	1.31	[0.94; 1.81]	0.1102	
Risky alcohol consumption	1.41	[0.98; 2.04]	0.0687	
Body Mass Index				
Normal	Reference			0.1596
Underweight	1.40	[0.61; 3.17]	0.4259	
Pre-obese	1.04	[0.85; 1.27]	0.7237	
Obese	0.80	[0.63; 1.03]	0.0826	
Resilience				
High resilience	Reference			0.0327*
Normal resilience	0.83	[0.65; 1.05]	0.1281	
Low resilience	0.65	[0.47; 0.90]	0.0090*	
Treated with medication				
Yes	Reference			NA
No	0.74	[0.62; 0.89]	0.0013*	
Steroid treatment				
No	Reference			NA
Yes	0.22	[0.05; 0.90]	0.0357*	
Chronic liver disease				
No	Reference			NA
Yes	0.81	[0.58; 1.15]	0.2385	
Chronic rheumatologic/immunologic disease				
No	Reference			NA
Yes	0.71	[0.50; 1.00]	0.0512	
Chronic neurological disease				
No	Reference			NA
Yes	0.80	[0.64; 1.00]	0.0522	

Overall *P* value: multivariate Wald test

*$P < 0.05$

(S Table 2). Cox proportional hazard models for hospitalized patients and non-hospitalized patients, together with time-varying effect estimates of hospitalization can be found in the Supplementary Appendix (S Table 3–5). Non-hospitalized patients were more likely to become symptom-free in the first four weeks (aHR 2.42, 95% CI 1.28–4.59). No significant differences were found after this time period.

Discussion

Main findings

We used data from a large population-based multicenter study for the retrospective analysis of the duration of, and risk factors for a prolonged recovery from acute SARS-CoV-2 infection. While 65.5% of included participants reported to still have symptoms 28 days after infection, over half of the symptomatic participants (54.1%) experienced at least one persisting symptom about 9 months post-infection. 22 of 23 different symptoms during the acute phase except for vomiting persisted beyond 9 months, with anosmia, dyspnea, ageusia, and fatigue being the most frequent ones. We found that female sex, age between 49 and 59 years, lower educational level, living with a partner, low resilience, steroid treatment and no medication during acute infection were associated with prolonged time to symptom-free, and being hospitalized was associated with prolonged time only in the first four weeks.

Study findings in context

We found that COVID-19-related symptoms rapidly resolved at the beginning but only incremental improvement was seen beyond 28 days. A former study also demonstrated that symptom load at 1.5 to 6 months was not associated with the length of time since symptom onset, suggesting that improvement in symptoms primarily occurred during the first few weeks after infection [12]. Furthermore, most subgroup differences in time to symptom-free occurred within 28 days after symptom onset in our study.

The most prevalent symptoms including anosmia, dyspnea, ageusia, and fatigue corresponded to those reported in a study of non-hospitalized individuals and another one of patients with mild or moderate symptoms [12, 16]. Long persistence of symptoms is worrying because persisting COVID-19 symptoms are associated with poor health-related quality of life (HRQOL) [9, 33]. Even though the present analysis did not differentiate symptoms according to their severity or their impact on daily life or HRQOL, our previous analysis of COVIDOM/NAPKON-POP data [9] revealed that different symptoms have a different impact on the severity of PCS and, consequently, on HRQOL. Therefore, learning more about symptom persistence and symptom resolution is of utmost clinical relevance.

Our study identified several risk factors for prolonged symptom persistence. An age between 49 and 59 years, being female, lower education, living with a partner, low resilience, steroid treatment, and no medication during acute infection were factors that predicted longer symptom persistence. Some of these factors like age are in line with previous studies [21, 34], although the inverse U-shaped association of age with risk might seem surprising. However, similar results were obtained from 10 longitudinal studies in the UK, with the highest risk noted in the middle age categories, i.e. 45–54 and 55–69 years [20]. Arguably, this might be attributable to competing mortality risks or erroneous attribution of symptoms to other causes in older age [20]. On the other hand, we cannot exclude that participants' differential recall might also have been determined by some of the risk factors in question, especially age, resilience, and education. Hence, the identified predictors still require confirmation by independent longitudinal studies. Consistent with most previous studies [21, 23, 35, 36], we found that female patients were less likely to recover quickly from symptoms than male patients. In contrast to our results, a Swedish study found that the female sex was protective for Long COVID-related sick leave, but only in a subgroup of hospitalized patients [37]. Patients with lower education are more likely to have physically demanding jobs [38], which might have influenced their recovery from symptoms. The effect of living status might be due to recall bias since patients living with a partner might have discussed their symptoms more frequently with their partner, as compared to patients without a partner or not living with a partner. This might result in differential reporting of symptoms in patients without a partner or not living with a partner, thus the observed effect should be interpreted with caution. Moreover, it may be speculated that constant exposure to a partner's infection might have increased virus load. In our previous study [9], we found low resilience and strong acute disease severity to be risk factors for severe PCS. Similarly, patients with more severe acute COVID-19 were also reported to show prolonged symptoms [39]. Likewise, steroid treatment might be an indicator of disease severity that results in prolonged symptoms. Although it has been shown that inhaled corticosteroid treatment improved symptom resolution in COVID-19 patients [40], a meta-analysis demonstrated an association between corticosteroid therapy and increased length of stay, although this finding was only based on subgroup analysis in three randomized controlled trials [41].

Strengths and limitations

A major strength of our study is that we reported a population-based estimate of the status and duration of symptoms drawing upon data from over 1100 COVID-19 patients with an average follow-up of 9 months.

There are some limitations. First and foremost, our use of the COVIDOM/NAPKON-POP time-to-recovery data had to be retrospective in nature because the study did not collect symptoms prospectively starting from infection. Since this might have been subject to recall bias, factors affecting the precision of the derived time-to-recovery data might have confounded some of the relationships between the latter and potential predictors. However, it is also likely that patients remember the time course well even after recovery. Second, as this study is not a representative sample of the total population, selection bias must be taken into account. It has to be mentioned that selection and differential response could have biased the estimates of the prevalence and persistence of symptoms. However, given the nature of the cooperation with the local health authorities, we are confident that the COVIDOM/NAPKON-POP sample is a valid representation of the infected population at the given time in the respective regions. Third, symptom status was collected by self-report, asking participants about COVID-19-related symptoms. However, we cannot rule out the possibility that some symptoms were caused by other respiratory infections. Furthermore, although we assume that most participants would not mention a chronic symptom as it is not noticeably related to the COVID-19 disease, future studies should evaluate the presence of symptoms before COVID-19 and their potential aggravation because of COVID-19. Fourth, long-term symptom status of initially asymptomatic patients was not evaluated. It is still unknown whether this group developed new symptoms after acute infection. Third, patients included in COVIDOM/NAPKON study probably mainly had SARS-CoV-2 wild type or alpha variant infection with a higher burden of symptoms than later variants. Future analyses of the cohort population from 2022 will evaluate how comparable symptom persistence after the omicron variant is to our present findings. Finally, the study does not include a control group, which makes it difficult to know whether the reported symptoms can indeed be attributed to SARS-CoV-2 infection.

Conclusions

Over half of the participants reported COVID-19-related symptoms 9 months after infection. Many patients experienced rapid recovery, but prolonged recovery was also seen particularly among those characterized by middle age, female sex, lower educational level, living with a partner, low resilience, and without medication during acute infection.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s15010-023-02043-6.

Acknowledgements We acknowledge Dr. Nicole S. Erler from Erasmus University Medical Center for assisting with multiple imputations. Data were provided by NAPKON (Nationales Pandemie Kohorten Netz, German National Pandemic Cohort Network) of the Network University Medicine (NUM), funded by the Federal Ministry of Education and Research (BMBF).

We express our gratitude to all collaborators of the NAPKON study group, including study teams, the governance bodies, individuals who made contributions to coordination and infrastructure of NAPKON and the coordination of NUM, and the clinical research support staff. The NAPKON study team members include: Agaplesion Bethesda Krankenhaus Bergedorf gemeinnützige GmbH (Marc Gregor Bota), Elisabeth-Krankenhaus Essen (Ingo Voigt), Frankfurt Universitätsklinikum (Maria J.G.T. Vehreschild, Jörg J. Vehreschild), Hausarztpraxis Dashti (Hiwa Dashti), Hausarztpraxis Egestorf, Bispingen (Barbara Laumerich), Helios Klinik Bad Saarow (Oliver Pociuli), Helios Klinikum Duisburg (Nikolaus Büchner), Helios Klinikum Erfurt (Sabine Adler), Helios Klinikum Krefeld (Mathias Lehmann), Helios Klinikum Siegburg (Selcuk Tasci), Kinder- und Jugendpraxis Jorczyk, Dresden (Maximilian Jorczyk), KJF Klinik Josefinum gGmbH (Thomas Keller), Klinik Hallerwiese-Cnopfsche Kinderklinik (Michael Schroth), Klinikum Dortmund gGmbH, Medizinische Klinik Nord (Martin Hower), Klinikum Leverkusen (Lukas Eberwein), Klinikum Worms (Tim Zimmermann), Krankenhaus Bethanien—Institut für Pneumologie an der Universität zu Köln (Simon-Dominik Herkenrath), Malteser Krankenhaus St. Franziskus-Hospital (SFH) Flensburg (Milena Milovanovic), MVZ am Isartor, München (Ramona Pauli), MVZ im Altstadt-Caree Fulda GmbH Medizinisches Versorgungszentrum (Jörg Simon), OWL: Evangelisches Klinikum Bethel (Eckard Hamelmann), OWL: Klinikum Bielefeld (Christoph Stellbrink), OWL: Klinikum Lippe (Johannes-Josef Tebbe), Petrus Krankenhaus Wuppertal (Sven Stieglitz), Praxis am Ebertplatz (Christoph Wyen), Praxis Bosch und Renner, Köln (Jan Bosch), Praxis Dilltal, Burbach (Mirko Steinmüller), Praxis Dr. Allerlei, Frankfurt am Main (Christoph Allerlei), Praxis Dr. Böbel, Reutlingen (Markus Böbel), Praxis Dr. Elke Heinitz, Grube (Elke Natascha Heinitz), Praxis Dr. med Ariane Roecken (Ariane Roecken), Praxis Dr. med. Andrea Münkle-Krimly (Andrea Münckle-Krimly), Praxis Dr. med. Christiane Guderian (Christiane Guderian), Praxis im Tennental (Ingmar Silberbaur), SHG-Kliniken Völklingen (Harald Schäfer), Tropenklinik Paul-Lechler-Krankenhaus Tübingen (Claudia Raichle), TUM München (Christoph Spinner), UKGM Standort Marburg (Bernd Schmeck), Uniklinik Carl Gustav Carus Dresden (Heidi Altmann, Nicole Toepfner), Uniklinik der Ruhr-Universität Bochum (Wolfgang Schmidt), Uniklinik Düsseldorf (Björn Jensen), Uniklinik Erlangen (Andreas Kremer), Uniklinik Göttingen (Sabine Blaschke), Uniklinik Halle (Jochen Dutzmann), Uniklinik Hamburg-Eppendorf (Marylyn Addo), Uniklinik Homburg (Robert Bals), Uniklinik Leipzig (Sven Bercker), Uniklinik Münster (Phil-Robin Tepasse), Uniklinik Regensburg (Frank Hanses), Uniklinik RWTH Aachen (Dirk Müller-Wieland), Uniklinik Schleswig-Holstein, Kiel (Anette Friedrichs), Uniklinik Schleswig-Holstein, Lübeck (Jan Rupp), Uniklinik Tübingen (Siri Göpel), Uniklinik Würzburg (Jens Maschmann), Universitätsklinikum Augsburg (Christine Dhillon), Universitätsklinikum Bonn (Jacob Nattermann), Universitätsklinikum Essen (Ingo Voigt), Universitätsklinikum Magdeburg (Wilfred Obst), Universitätsmedizin der Johannes Gutenberg-Universität Mainz (Martin Franz Sprinzl), Universitätsmedizin Greifswald (Christian Scheer), Universitätsmedizin Mannheim (Andreas Teufel), Universitätsmedizin Oldenburg (Ulf Günther), Charité—Universitätsmedizin Berlin (Martin Witzenrath, Thomas Keil, Thomas Zoller, Sein Schmidt, Michael Hummel, Lilian Krist, Julia Fricke, Maria Rönnefarth, Denise Treue, Ludie Kretzler, Chantip Dang-Heine, Paul Triller, Andreas Jooß, Jenny Schlesinger, Natalja Liseweski, Christina Pley, Carmen Scheibenbogen), Hannover Medizinische Hochschule (Marius Hoeper), Jena (Philipp A.

Reuken), Klinikum der Universität München (Michael von Bergwelt), UKSH-Kiel (Rainer Noth), UKSH-Lübeck (Daniel Drömann), Universitätsklinikum Frankfurt (Maria J.G.T. Vehreschild), Universitätsklinikum Freiburg (Siegbert Rieg), Universitätsklinikum Giessen / Marburg (Istvan Vadasz), Universitätsklinikum Köln (Philipp A. Koehler), Universtitästklinikum Heidelberg (Uta Merle), Kiel Universitätsklinikum (Stefan Schreiber), Würzburg Universitätsklinikum (Peter Heuschmann, Stefan Störk). The following individuals have supported governance bodies of the NAPKON: Use & Access Committee Anette Friedrichs, Astrid Petersmann, Claudia Ellert, Georg Schmidt, Janne Vehreschild, Katrin Milger, Marie von Lilienfeld, Martin Witzenrath, Oliver Witzke, Patrick Meybohm, Peter Heuschmann, Sabine Blaschke, Sandra Frank, Stefan Schreiber, Thomas Illig. Advisory Board Alexander Hein, Andrea Wittig, Andreas Simm, Anette Friedrichs, Anke Reinacher-Schick, Anna Frey, Antonella Iannaccone, Astrid Petersmann, Benjamin Maasoumy, Benjamin Waschki, Bimba Hoyer, Brigitt van Oorschot, Carolina van Schaik, Christina Lemhöfer, Christina Polidori, Christine Klein, Daniel Medenwald, Eva Christina Schulte, Eva Grill, Felix Meinel, Folke Brinkmann, Ghazal Arabi, Heike Bickeböller, Holger Lindner, Ildiko Gagyor, Jessica Hassel, Jürgen Deckert, Katrin Milger-Kneidinger, Kerstin Ludwig, Marcus Dörr, Marie von Lilienfeld-Toal, Martin Möckel, Martin Weigl, Matthias Nauck, Miriam Banas, Muenevver Demir, Nicole Lindenberg, Nora Hettich, Norma Jung, Oliver Witzke, Orlando Guntinas-Lichius, Patrick Meybohm, Reinhard Berner, Sabine Blaschke, Samuel Knauss, Sandra Frank, Sebastian Baumeister, Sebastian Dolff, Selma Ugurel, Sophia Stöcklein, Stefanie Joos, Winfred Häuser. The coordination and infrastructure of the NAPKON was conducted by: ICU: Jörg Janne Vehreschild, Maximilian Schons, Sina Hopff, Markus Brechtel, Cristina Schmidt-Ibanez, Johannes Schneider, Carolin Jakob, Franziska Voß. BCU: Inga Bernemann, Sonja Kunze, Maike Tauchert, Thomas Illig, Gabriele Anton. ECU: Cornelia Fiessler, Mirjam Kohls, Olga Miljukov, Steffi Jiru-Hillmann, Jens-Peter Reese, Peter Heuschmann. IGCU: Jens-Peter Reese, Peter Heuschmann, Anna-Lena Hofmann, Julia Schmidt, Kathrin Ungethüm, Anna Horn, Michael Krawczak. POP: Thomas Bahmer, Wolfgang Lieb, Daniel Pape, Stefan Schreiber, Anne Hermes, Irene Lehmann, Corina Maetzler, Lukas Tittmann. DZHK: Roberto Lorbeer, Bettina Lorenz-Depiereux, Monika Kraus, Christian Schäfer, Jens Schaller, Mario Schattschneider, Dana Stahl, Heike Valentin, Dagmar Krefting, Matthias Nauck. Pediatric core unit: Nicole Toepfner, Reinhard Berner. GECCO & Interoperability Team: Christof von Kalle, Sylvia Thun, Alexander Bartschke, Liudmila Lysyakova, Stefanie Rudolph, Julian Sass. Kardiology team: Eike Nagel, Valentina Püntmann. We acknowledge the dedicated support of clinical research support staff of the Institute of Experimental and Translational Cardiovascular Imaging, including Tammy Wolf, Thourier Azdad, Franziska Weis, Ira Krückemeier, Simon Bohlender, Deniz Desik, and Layla Laghchioua. Coordination NUM: Ralf Heyder, Silke Wiedmann.

Author contributions YS: statistical analysis; interpretation of results; drafting and revising the manuscript. RS: acquisition and interpretation of data; statistical support; revision of the draft. JPR, FAM: acquisition and interpretation of data; revision of the draft. SSch, TB, WL, MK, PH, SStö, AH, LK, TK, MW: conceptualization of the COVIDOM/NAPKON-POP study; acquisition and interpretation of data; revision of the draft. JJV, RG, NS, RM, HH, BLD, CL, CA: revision of the draft. EG: conception and design of the present study; acquisition and interpretation of data; revision of the draft. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. This work was supported by the German Federal Ministry of Education and Research (BMBF) via the Network University Medicine (FKZ: 01KX2021). Parts of the infrastructure of the Kiel and Würzburg study sites were supported by the federal states of Schleswig–Holstein and Bavaria. Yanyan Shi is supported by a scholarship from the China Scholarship Council (202108140048).

Data availability Data of this study are available upon request to the Use & Access Committee (UAC) of NAPKON (https://proskive.napkon.de).

Declarations

Conflict of interest TB reports personal fees from AstraZeneca, GlaxoSmithKline, Novartis, Roche, Chiesi, Boehringer Ingelheim, MSD and Pfizer outside the submitted work. SStö reports research grants of the Federal Ministry of Education and Research (#01EO1004; #01EO1504), speaker honoraria or advisory board honoraria of AstraZeneca, Boehringer Ingelheim, Novartis, Pfizer, Pharmacosmos, and case payment fees in clinical studies of Alnylam, AstraZeneca, Boehringer Ingelheim, IONIS, MSD, NovoNordisk, SOBI, Servier; all outside the submitted work. All other authors have no competing interests to declare.

Ethical approval COVIDOM/NAPKON-POP is registered at the German Registry for clinical studies (DRKS00023742) and at http://www.clinicaltrials.gov (NCT04679584). NAPKON-POP was approved by the local ethic committees of respective study sites (Kiel, No. D 537/20; Würzburg, No. 236/20_z). According to the professional code of the Berlin Medical Association regarding multi-center studies, approval by the Ethics Committee of the coordinating study center (Kiel) was also valid for the Berlin study site [25]. Written informed consent was obtained prior to all procedures from all participants.

References

1. World Health Organization. WHO coronavirus (COVID-19) dashboard.
2. Michelen M, Manoharan L, Elkheir N, Cheng V, Dagens A, Hastie C, et al. Characterising long COVID: a living systematic review. BMJ Glob Health. 2021;6:e005427. https://doi.org/10.1136/bmjgh-2021-005427.
3. van Kessel SAM, Olde Hartman TC, Lucassen PLBJ, van Jaarsveld CHM. Post-acute and long-COVID-19 symptoms in patients with mild diseases: a systematic review. Fam Pract. 2021;39:159–67. https://doi.org/10.1093/fampra/cmab076.
4. National Institute for Health and Care Excellence: Clinical Guidelines. COVID-19 rapid guideline: managing the long-term effects of COVID-19. London: National Institute for Health and Care Excellence (NICE); 2022.

5. Alkodaymi MS, Omrani OA, Fawzy NA, Shaar BA, Almamlouk R, Riaz M, et al. Prevalence of post-acute COVID-19 syndrome symptoms at different follow-up periods: a systematic review and meta-analysis. Clin Microbiol Infect. 2022;28:657–66. https://doi.org/10.1016/j.cmi.2022.01.014.
6. Nasserie T, Hittle M, Goodman SN. Assessment of the frequency and variety of persistent symptoms among patients with COVID-19: a systematic review. JAMA Netw Open. 2021;4:e2111417. https://doi.org/10.1001/jamanetworkopen.2021.11417.
7. Lopez-Leon S, Wegman-Ostrosky T, Perelman C, Sepulveda R, Rebolledo PA, Cuapio A, et al. More than 50 long-term effects of COVID-19: a systematic review and meta-analysis. Sci Rep. 2021;11:16144. https://doi.org/10.1038/s41598-021-95565-8.
8. Aiyegbusi OL, Hughes SE, Turner G, Rivera SC, McMullan C, Chandan JS, et al. Symptoms, complications and management of long COVID: a review. J R Soc Med. 2021;114:428–42. https://doi.org/10.1177/01410768211032850.
9. Bahmer T, Borzikowsky C, Lieb W, Horn A, Krist L, Fricke J, et al. Severity, predictors and clinical correlates of Post-COVID syndrome (PCS) in Germany: a prospective, multi-centre, population-based cohort study. EClinicalMedicine. 2022;51:101549. https://doi.org/10.1016/j.eclinm.2022.101549.
10. Ballering AV, van Zon SKR, olde Hartman TC, Rosmalen JGM. Persistence of somatic symptoms after COVID-19 in the Netherlands: an observational cohort study. Lancet. 2022;400:452–61. https://doi.org/10.1016/S0140-6736(22)01214-4.
11. Zeng N, Zhao Y-M, Yan W, Li C, Lu Q-D, Liu L, et al. A systematic review and meta-analysis of long term physical and mental sequelae of COVID-19 pandemic: call for research priority and action. Mol Psychiatry. 2023;28:423–33. https://doi.org/10.1038/s41380-022-01614-7.
12. Stavem K, Ghanima W, Olsen MK, Gilboe HM, Einvik G. Persistent symptoms 1.5–6 months after COVID-19 in non-hospitalised subjects: a population-based cohort study. Thorax. 2021;76:405–7. https://doi.org/10.1136/thoraxjnl-2020-216377.
13. Hirschtick JL, Titus AR, Slocum E, Power LE, Hirschtick RE, Elliott MR, et al. Population-based estimates of post-acute sequelae of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection (PASC) prevalence and characteristics. Clin Infect Dis. 2021;73:2055–64. https://doi.org/10.1093/cid/ciab408.
14. Blomberg B, Mohn KG-I, Brokstad KA, Zhou F, Linchausen DW, Hansen B-A, et al. Long COVID in a prospective cohort of home-isolated patients. Nat Med. 2021;27:1607–13. https://doi.org/10.1038/s41591-021-01433-3.
15. Petersen MS, Kristiansen MF, Hanusson KD, Danielsen ME, á Steig B, Gaini S, et al. Long COVID in the Faroe Islands: a longitudinal study among nonhospitalized patients. Clin Infect Dis. 2020;73:e4058–63. https://doi.org/10.1093/cid/ciaa1792.
16. Boscolo-Rizzo P, Guida F, Polesel J, Marcuzzo AV, Capriotti V, D'Alessandro A, et al. Sequelae in adults at 12 months after mild-to-moderate coronavirus disease 2019 (COVID-19). Int Forum Allergy Rhinol. 2021;11:1685–8. https://doi.org/10.1002/alr.22832.
17. Wynberg E, van Willigen HDG, Dijkstra M, Boyd A, Kootstra NA, van den Aardweg JG, et al. Evolution of Coronavirus Disease 2019 (COVID-19) symptoms during the first 12 months after illness onset. Clin Infect Dis. 2022;75:e482–90. https://doi.org/10.1093/cid/ciab759.
18. Davis HE, Assaf GS, McCorkell L, Wei H, Low RJ, Re'em Y, et al. Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. EClinicalMedicine. 2021;38:101019. https://doi.org/10.1016/j.eclinm.2021.101019.
19. The PHOSP-COVID Collaborative Group. Clinical characteristics with inflammation profiling of long COVID and association with 1-year recovery following hospitalisation in the UK: a prospective observational study. Lancet Respir Med. 2022;10:761–75. https://doi.org/10.1016/s2213-2600(22)00127-8.
20. Thompson EJ, Williams DM, Walker AJ, Mitchell RE, Niedzwiedz CL, Yang TC, et al. Long COVID burden and risk factors in 10 UK longitudinal studies and electronic health records. Nat Commun. 2022;13:3528. https://doi.org/10.1038/s41467-022-30836-0.
21. Sudre CH, Murray B, Varsavsky T, Graham MS, Penfold RS, Bowyer RC, et al. Attributes and predictors of long COVID. Nat Med. 2021;27:626–31. https://doi.org/10.1038/s41591-021-01292-y.
22. Wu Q, Ailshire JA, Crimmins EM. Long COVID and symptom trajectory in a representative sample of Americans in the first year of the pandemic. Sci Rep. 2022;12:11647. https://doi.org/10.1038/s41598-022-15727-0.
23. Evans RA, McAuley H, Harrison EM, Shikotra A, Singapuri A, Sereno M, et al. Physical, cognitive, and mental health impacts of COVID-19 after hospitalisation (PHOSP-COVID): a UK multicentre, prospective cohort study. Lancet Respir Med. 2021;9:1275–87. https://doi.org/10.1016/S2213-2600(21)00383-0.
24. Schons M, Pilgram L, Reese J-P, Stecher M, Anton G, Appel KS, et al. The German National Pandemic Cohort Network (NAPKON): rationale, study design and baseline characteristics. Eur J Epidemiol. 2022;37(8):849–70. https://doi.org/10.1007/s10654-022-00896-z.
25. Horn A, Krist L, Lieb W, Montellano FA, Kohls M, Haas K, et al. Long-term health sequelae and quality of life at least 6 months after infection with SARS-CoV-2: design and rationale of the COVIDOM-study as part of the NAPKON population-based cohort platform (POP). Infection. 2021;49:1277–87. https://doi.org/10.1007/s15010-021-01707-5.
26. Federal Center for Health Education, German Medical Association. Addressing alcohol consumption in patients. Medical manual for the prevention and treatment of risky, harmful and dependent consumption. 2021.
27. WHO Consultation on Obesity, World Health Organization. Obesity : preventing and managing the global epidemic: report of a WHO consultation. Geneva: World Health Organization; 2000.
28. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? Int J Methods Psychiatr Res. 2011;20:40–9. https://doi.org/10.1002/mpr.329.
29. Rubin DB. Multiple imputation for nonresponse in surveys. New York: Wiley; 2004.
30. Kleinbaum DG, Klein M. Survival analysis: a self-learning text. Berlin: Springer; 2012.
31. Fox J, Monette G. Generalized collinearity diagnostics. J Am Stat Assoc. 1992;87(417):178–83. https://doi.org/10.1080/01621459.1992.10475190.
32. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. JAMA. 1982;247:2543–6. https://doi.org/10.1001/jama.1982.03320430047030.
33. Malik P, Patel K, Pinto C, Jaiswal R, Tirupathi R, Pillai S, et al. Post-acute COVID-19 syndrome (PCS) and health-related quality of life (HRQoL)—a systematic review and meta-analysis. J Med Virol. 2022;94:253–62. https://doi.org/10.1002/jmv.27309.
34. Kostev K, Smith L, Koyanagi A, Jacob L. Prevalence of and factors associated with post-coronavirus disease 2019 (COVID-19) condition in the 12 months after the diagnosis of COVID-19 in adults followed in general practices in Germany. Open Forum Infect Dis. 2022;9:333. https://doi.org/10.1093/ofid/ofac333.
35. Peghin M, Palese A, Venturini M, De Martino M, Gerussi V, Graziano E, et al. Post-COVID-19 symptoms 6 months after acute infection among hospitalized and non-hospitalized patients. Clin

Microbiol Infect. 2021;27:1507–13. https://doi.org/10.1016/j.cmi.2021.05.033.
36. Subramanian A, Nirantharakumar K, Hughes S, Myles P, Williams T, Gokhale KM, et al. Symptoms and risk factors for long COVID in non-hospitalized adults. Nat Med. 2022;28:1706–14. https://doi.org/10.1038/s41591-022-01909-w.
37. Westerlind E, Palstam A, Sunnerhagen KS, Persson HC. Patterns and predictors of sick leave after Covid-19 and long Covid in a national Swedish cohort. BMC Public Health. 2021;21:1023. https://doi.org/10.1186/s12889-021-11013-2.
38. Warren JR, Hoonakker P, Carayon P, Brand J. Job characteristics as mediators in SES–health relationships. Soc Sci Med. 2004;59:1367–78. https://doi.org/10.1016/j.socscimed.2004.01.035.
39. Kamal M, Abo Omirah M, Hussein A, Saeed H. Assessment and characterisation of post-COVID-19 manifestations. Int J Clin Pract. 2021;75:e13746. https://doi.org/10.1111/ijcp.13746.
40. Chen CH, Wang CY, Wang YH, Chen CY, Chen KH, Lai CC, et al. The effect of inhaled corticosteroids on the outcomes of patients with COVID-19: a systematic review and meta-analysis of randomized controlled trials. Expert Rev Clin Pharmacol. 2022;15:593–600. https://doi.org/10.1080/17512433.2022.2094769.
41. Zhou F, Deng J, Heybati K, Zuo QK, Ali S, Hou W, et al. Efficacy and safety of corticosteroid regimens for the treatment of hospitalized COVID-19 patients: a meta-analysis. Future Virol. 2022;17:463–89. https://doi.org/10.2217/fvl-2021-0244.

Authors and Affiliations

Yanyan Shi[1,2] · Ralf Strobl[1,3] · Christian Apfelbacher[4] · Thomas Bahmer[5] · Ramsia Geisler[6] · Peter Heuschmann[7,8] · Anna Horn[7] · Hanno Hoven[9] · Thomas Keil[7,10,11] · Michael Krawczak[12] · Lilian Krist[10] · Christina Lemhöfer[13] · Wolfgang Lieb[14] · Bettina Lorenz-Depiereux[15] · Rafael Mikolajczyk[16,17] · Felipe A. Montellano[7,18] · Jens Peter Reese[7] · Stefan Schreiber[5] · Nicole Skoetz[19] · Stefan Störk[20] · Jörg Janne Vehreschild[6,21,22] · Martin Witzenrath[23,24] · Eva Grill[1,3] on behalf of the NAPKON Study Group

✉ Eva Grill
Eva.Grill@med.uni-muenchen.de

1 Institute for Medical Information Processing, Biometry and Epidemiology (IBE), Faculty of Medicine, Ludwig-Maximilians-Universität München (LMU Munich), Marchioninistr. 15, 81377 Munich, Germany

2 Pettenkofer School of Public Health, Munich, Germany

3 German Center for Vertigo and Balance Disorders, University Hospital, Ludwig-Maximilians-Universität München (LMU Munich), Munich, Germany

4 Institute of Social Medicine and Health Systems Research, Medical Faculty, Otto von Guericke University Magdeburg, Magdeburg, Germany

5 Internal Medicine Department I, University Hospital Schleswig-Holstein Campus Kiel (UKSH Kiel), Kiel, Germany

6 Department II of Internal Medicine, Hematology/Oncology, Goethe University, Frankfurt, Frankfurt Am Main, Germany

7 Institute for Clinical Epidemiology and Biometry, Julius-Maximilians-University, Würzburg, Würzburg, Germany

8 Clinical Trial Center, University Hospital Würzburg, Würzburg, Germany

9 Institute for Occupational and Maritime Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

10 Institute of Social Medicine, Epidemiology and Health Economics, Charité-Universitätsmedizin Berlin, Berlin, Germany

11 State Institute of Health I, Bavarian Health and Food Safety Authority, Erlangen, Germany

12 Institute of Medical Informatics and Statistics, Kiel University, University Medical Center Schleswig-Holstein, Kiel, Germany

13 Institute of Physical and Rehabilitation Medicine, University Hospital Jena, Jena, Germany

14 Institute of Epidemiology, Kiel University, University Medical Center Schleswig-Holstein, Kiel, Germany

15 Research Unit Molecular Epidemiology, Institute of Epidemiology, Helmholtz Center Munich, Munich, Germany

16 Institute for Medical Epidemiology, Biometrics, and Informatics, Interdisciplinary Center for Health Sciences, Medical Faculty of the Martin Luther University Halle-Wittenberg, Halle (Saale), Germany

17 German Centre for Mental Health, Site Jena-Magdeburg-Halle, Halle, Germany

18 Department of Neurology, University Hospital Würzburg, Würzburg, Germany

19 Evidence-Based Medicine, Department I of Internal Medicine, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany

20 Department of Clinical Research and Epidemiology, Comprehensive Heart Failure Center and Department of Internal Medicine I, University Hospital Würzburg, Würzburg, Germany

21 Department I of Internal Medicine, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany

22 German Centre for Infection Research (DZIF), Partner Site Bonn-Cologne, Cologne, Germany

23 Department of Infectious Diseases, Respiratory Medicine and Critical Care, Charité-Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

24 German Center for Lung Research (DZL), Giessen, Germany

Persistent symptoms and risk factors predicting prolonged time to symptom-free after SARS-CoV-2 infection: An analysis of the baseline examination of the German COVIDOM/NAPKON-POP cohort

Yanyan Shi, Ralf Strobl, Christian Apfelbacher, Thomas Bahmer, Ramsia Geisler, Peter Heuschmann, Anna Horn, Hanno Hoven, Thomas Keil, Michael Krawczak, Lilian Krist, Christina Lemhöfer, Wolfgang Lieb, Bettina Lorenz-Depiereux, Rafael Mikolajczyk, Felipe A. Montellano, Jens Peter Reese, Stefan Schreiber, Nicole Skoetz, Stefan Störk, Jörg Janne Vehreschild, Martin Witzenrath, Eva Grill *on behalf of the NAPKON Study Group*

Corresponding author:

Prof. Dr. Eva Grill, Institute for Medical Information Processing, Biometry and Epidemiology, Ludwig-Maximilians-Universität München, Marchioninistr. 15, 81377 Munich, Germany. **E-Mail:** Eva.Grill@med.uni-muenchen.de

SUPPLEMENTARY DATA

S Table1 Brief Resilience Score

Item
1. I tend to bounce back quickly after hard times.
2. I have a hard time making it through stressful events. (R)
3. It does not take me long to recover from a stressful event.
4. It is hard for me to snap back when something bad happens. (R)
5. I usually come through difficult times with little trouble.
6. I tend to take a long time to get over set-backs in my life. (R)

R **=** reverse coded items
Response options: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree.

S Table 2 Risk factors predicting prolonged time to symptom-free status in COVID-19 patients (N=1175, separate Cox proportional hazard models for each risk factor adjusted for age and sex)

Covariates	Adjusted hazard ratio	95% confidence interval	P value
Educational level			
University entrance certificate	Reference		
Lower education	0.70	[0.58; 0.84]	0.0002*
Living status			
No partner/Not living with a partner	Reference		
Living with a partner	0.85	[0.69; 1.03]	0.1024
Smoking status			
Current-smokers	Reference		
Ex-smokers	1.04	[0.78; 1.39]	0.7860
Non-smokers	1.24	[0.95; 1.63]	0.1205
Alcohol consumption			
Abstainer	Reference		
Low-risk alcohol consumption	1.40	[1.00; 1.96]	0.0493*
Risky alcohol consumption	1.58	[1.08; 2.30]	0.0186*
Body Mass Index			
Normal	Reference		
Underweight	1.22	[0.54; 2.74]	0.6347
Pre-obese	0.97	[0.80; 1.18]	0.7674
Obese	0.69	[0.54; 0.87]	0.0019*
Resilience			
High resilience	Reference		
Normal resilience	0.77	[0.61; 0.98]	0.0342*
Low resilience	0.56	[0.42; 0.76]	0.0003*
Treated with medication			
Yes	Reference		
No	0.92	[0.78; 1.10]	0.3579
Steroid treatment			
No	Reference		
Yes	0.20	[0.05; 0.82]	0.0253*
Anticoagulation			
No	Reference		
Yes	0.74	[0.49; 1.12]	0.1591

Anti-infectives			
No	Reference		
Yes	0.94	[0.60; 1.45]	0.7717
Chronic liver disease			
No	Reference		
Yes	0.67	[0.48; 0.94]	0.0199*
Chronic rheumatologic/immunologic disease			
No	Reference		
Yes	0.68	[0.48; 0.96]	0.0297*
Tumor/cancer diseases			
No	Reference		
Yes	1.16	[0.62; 2.18]	0.6512
Chronic neurological disease			
No	Reference		
Yes	0.69	[0.56; 0.85]	0.0006*
Lung disease			
No	Reference		
Yes	0.80	[0.63; 1.01]	0.0587
ENT disease			
No	Reference		
Yes	0.94	[0.77; 1.15]	0.5403
Cardiovascular disease			
No	Reference		
Yes	0.80	[0.64; 1.00]	0.0468*
Diabetes			
No	Reference		
Yes	0.77	[0.47; 1.27]	0.2998

Note: Age and sex adjusted coefficients for symptom burden during acute infection and hospitalization during acute infection were not presented due to violation of proportional hazard assumption.

* $P < 0.05$.

S Table 3 Risk factors predicting prolonged time to symptom-free status in hospitalized COVID-19 patients (N=75, Cox proportional hazard model)

Covariates	Adjusted hazard ratio	95% confidence interval	P value
Alcohol consumption			
Abstainer	Reference		
Low-risk alcohol consumption	2.90	[0.63; 13.30]	0.1944
Risky alcohol consumption	3.76	[0.71; 19.86]	0.1429
Symptom burden during acute infection			
1-5 symptoms	Reference		
≥6 symptoms	0.09	[0.03; 0.32]	0.0020*
Treated with medication			
Yes	Reference		
No	0.27	[0.08; 0.88]	0.0459*
Chronic rheumatologic/immunologic disease			
No	Reference		
Yes	0.32	[0.07; 1.48]	0.1656

* $P < 0.05$.

S Table 4 Risk factors predicting prolonged time to symptom-free status in non-hospitalized COVID-19 patients stratified by symptom burden during acute infection (N=1100, stratified Cox proportional hazard model)

Covariates	Adjusted hazard ratio	95% confidence interval	P value
Age			
<49	Reference		
49-59	0.71	[0.57; 0.88]	0.0022*
≥60	1.00	[0.77; 1.28]	0.9821
Sex			
Male	Reference		
Female	0.78	[0.65; 0.94]	0.0078*
Educational level			
University entrance certificate	Reference		
Lower education	0.74	[0.61; 0.89]	0.0018*

Living status			
No partner/Not living with a partner	Reference		
Living with a partner	0.82	[0.66; 1.00]	0.0512
Alcohol consumption			
Abstainer	Reference		
Low-risk alcohol consumption	1.26	[0.85; 1.86]	0.2501
Risky alcohol consumption	1.35	[0.89; 2.04]	0.1569
Resilience			
High resilience	Reference		
Normal resilience	0.86	[0.68; 1.10]	0.2229
Low resilience	0.64	[0.46; 0.89]	0.0089*
Treated with medication			
Yes	Reference		
No	0.76	[0.63; 0.92]	0.0039*
Steroid treatment			
No	Reference		
Yes	0.20	[0.03; 1.43]	0.1094
Chronic liver disease			
No	Reference		
Yes	0.79	[0.55; 1.11]	0.1755
Chronic rheumatologic/immunologic disease			
No	Reference		
Yes	0.71	[0.50; 1.02]	0.0661
Chronic neurological disease			
No	Reference		
Yes	0.80	[0.64; 1.01]	0.0594

* $P < 0.05$.

S Table 5 Effect of hospitalization on COVID-19 patients stratified by symptom burden during acute infection (N=1175, stratified Cox proportional hazard model). The model is adjusted for age, sex, educational level, living status, alcohol consumption, BMI, resilience, COVID-19 medication, steroid treatment, chronic liver disease, chronic rheumatologic/immunologic disease, and chronic neurological disease.

Covariates	Adjusted hazard ratio	95% confidence interval	P value
Hospitalization during acute infection			
Hospitalized	Reference		
Non-hospitalized : first four weeks	2.42	[1.28; 4.59]	0.007*
Non-hospitalized : after four weeks	0.79	[0.42; 1.48]	0.457

* $P < 0.05$.

References

[1] World Health Organization. WHO Coronavirus (COVID-19) Dashboard.

[2] Baj J, Karakuła-Juchnowicz H, Teresiński G, Buszewicz G, Ciesielka M, Sitarz R, et al. COVID-19: Specific and Non-Specific Clinical Manifestations and Symptoms: The Current State of Knowledge. J Clin Med. 2020;9.

[3] Feldstein LR, Tenforde MW, Friedman KG, Newhams M, Rose EB, Dapul H, et al. Characteristics and Outcomes of US Children and Adolescents With Multisystem Inflammatory Syndrome in Children (MIS-C) Compared With Severe Acute COVID-19. JAMA. 2021;325:1074-87.

[4] Sorg A-L, Hufnagel M, Doenhardt M, Diffloth N, Schroten H, von Kries R, et al. Risk for severe outcomes of COVID-19 and PIMS-TS in children with SARS-CoV-2 infection in Germany. European Journal of Pediatrics. 2022;181:3635-43.

[5] Zimmermann P, Curtis N. COVID-19 in Children, Pregnancy and Neonates: A Review of Epidemiologic and Clinical Features. Pediatr Infect Dis J. 2020;39:469-77.

[6] Ludvigsson JF. Systematic review of COVID-19 in children shows milder cases and a better prognosis than adults. Acta Paediatr. 2020;109:1088-95.

[7] Williams N, Radia T, Harman K, Agrawal P, Cook J, Gupta A. COVID-19 Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection in children and adolescents: a systematic review of critically unwell children and the association with underlying comorbidities. Eur J Pediatr. 2021;180:689-97.

[8] Schober T, Caya C, Barton M, Bayliss A, Bitnun A, Bowes J, et al. Risk factors for severe PCR-positive SARS-CoV-2 infection in hospitalised children. BMJ Paediatr Open. 2022;6.

[9] Prévost B, Retbi A, Binder-Foucard F, Borde A, Bruandet A, Corvol H, et al. Risk factors for admission to the pediatric critical care unit among children hospitalized with COVID-19 in France. Front Pediatr. 2022;10:975826.

[10] Graff K, Smith C, Silveira L, Jung S, Curran-Hays S, Jarjour J, et al. Risk Factors for Severe COVID-19 in Children. The Pediatric Infectious Disease Journal. 2021;40:e137-e45.

[11] Nguyen PNT, Thuc TT, Hung NT, Thinh LQ, Minh NNQ, Duy DQ, et al. Risk factors for disease severity and mortality of children with Covid-19: A study at a Vietnamese Children's hospital. J Infect Chemother. 2022;28:1380-6.

[12] Choudhary R, Webber BJ, Womack LS, Dupont HK, Chiu SK, Wanga V, et al. Factors Associated With Severe Illness in Patients Aged <21 Years Hospitalized for COVID-19. Hosp Pediatr. 2022;12:760-83.

[13] Ng DC-E, Liew C-H, Tan KK, Chin L, Ting GSS, Fadzilah NF, et al. Risk factors for disease severity among children with Covid-19: a clinical prediction model. BMC Infectious Diseases. 2023;23:398.

[14] Michelen M, Manoharan L, Elkheir N, Cheng V, Dagens A, Hastie C, et al. Characterising long COVID: a living systematic review. BMJ global health. 2021;6:e005427.

[15] van Kessel SAM, Olde Hartman TC, Lucassen PLBJ, van Jaarsveld CHM. Post-acute and long-COVID-19 symptoms in patients with mild diseases: a systematic review. Family Practice. 2021;39:159-67.

[16] World Health Organization. Post COVID-19 condition (Long COVID). 2022.

[17] National Institute for Health and Care Excellence: Clinical Guidelines. COVID-19 rapid guideline: managing the long-term effects of COVID-19. London: National Institute for Health and Care Excellence (NICE); 2022.

[18] Callard F, Perego E. How and why patients made Long Covid. Soc Sci Med. 2021;268:113426.

[19] Hastie CE, Lowe DJ, McAuley A, Mills NL, Winter AJ, Black C, et al. True prevalence of long-COVID in a nationwide, population cohort study. Nature Communications. 2023;14:7892.

[20] Ballouz T, Menges D, Anagnostopoulos A, Domenghino A, Aschmann HE, Frei A, et al. Recovery and symptom trajectories up to two years after SARS-CoV-2 infection: population based, longitudinal cohort study. BMJ. 2023;381:e074425.

[21] Liu B, Jayasundara D, Pye V, Dobbins T, Dore GJ, Matthews G, et al. Whole of population-based cohort study of recovery time from COVID-19 in New South Wales Australia. The Lancet Regional Health - Western Pacific. 2021;12:100193.

[22] Oelsner EC, Sun Y, Balte PP, Allen NB, Andrews H, Carson A, et al. Epidemiologic Features of Recovery From SARS-CoV-2 Infection. JAMA Network Open. 2024;7:e2417440-e.

[23] Shi Q, Wang Z, Liu J, Wang X, Zhou Q, Li Q, et al. Risk factors for poor prognosis in children and adolescents with COVID-19: A systematic review and meta-analysis. EClinicalMedicine. 2021;41:101155.

[24] Harris E. Millions of US Children Experience Range of Long COVID Effects. JAMA. 2024;331:726.

[25] Lopez-Leon S, Wegman-Ostrosky T, Ayuzo del Valle NC, Perelman C, Sepulveda R, Rebolledo PA, et al. Long-COVID in children and adolescents: a systematic review and meta-analyses. Scientific Reports. 2022;12:9950.

[26] Gutiérrez-Gutiérrez B, Del Toro MD, Borobia AM, Carcas A, Jarrín I, Yllescas M, et al. Identification and validation of clinical phenotypes with prognostic implications in patients admitted to hospital with COVID-19: a multicentre cohort study. Lancet Infect Dis. 2021;21:783-92.

[27] Yamga E, Mullie L, Durand M, Cadrin-Chenevert A, Tang A, Montagnon E, et al. Interpretable clinical phenotypes among patients hospitalized with COVID-19 using cluster analysis. Front Digit Health. 2023;5:1142822.

[28] Rodríguez A, Ruiz-Botella M, Martín-Loeches I, Jimenez Herrera M, Solé-Violan J, Gómez J, et al. Deploying unsupervised clustering analysis to derive clinical phenotypes and risk factors associated with mortality risk in 2022 critically ill patients with COVID-19 in Spain. Crit Care. 2021;25:63.

[29] Oliveira EA, Oliveira MCL, Silva A, Colosimo EA, Mak RH, Vasconcelos MA, et al. Clinical Outcomes of Omicron Variant (B.1.1.529) Infection in Children and Adolescents Hospitalized With COVID-19 in Brazil With Observational Data on the Efficacy of the Vaccines in Adolescents. Pediatr Infect Dis J. 2023;42:218-25.

[30] Fink EL, Alcamo AM, Lovett M, Hartman M, Williams C, Garcia A, et al. Post-discharge outcomes of hospitalized children diagnosed with acute SARS-CoV-2 or MIS-C. Frontiers in Pediatrics. 2024;12.

[31] Loftus TJ, Shickel B, Balch JA, Tighe PJ, Abbott KL, Fazzone B, et al. Phenotype clustering in health care: A narrative review for clinicians. Frontiers in Artificial Intelligence. 2022;5.

[32] Carlile O, Briggs A, Henderson AD, Butler-Cole BFC, Tazare J, Tomlinson LA, et al. Impact of long COVID on health-related quality-of-life: an OpenSAFELY population cohort study using patient-reported outcome measures (OpenPROMPT). The Lancet Regional Health – Europe. 2024;40.

[33] Ottiger M, Poppele I, Sperling N, Schlesinger T, Müller K. Work ability and return-to-work of patients with post-COVID-19: a systematic review and meta-analysis. BMC Public Health. 2024;24:1811.

[34] Szewczyk W, Fitzpatrick AL, Fossou H, Gentile NL, Sotoodehnia N, Vora SB, et al. Long COVID and recovery from Long COVID: quality of life impairments and subjective cognitive decline at a median of 2 years after initial infection. BMC Infectious Diseases. 2024;24:1241.

[35] Alkodaymi MS, Omrani OA, Fawzy NA, Shaar BA, Almamlouk R, Riaz M, et al. Prevalence of post-acute COVID-19 syndrome symptoms at different follow-up periods: a systematic review and meta-analysis. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases. 2022;28:657-66.

[36] Davis HE, Assaf GS, McCorkell L, Wei H, Low RJ, Re'em Y, et al. Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. EClinicalMedicine. 2021;38:101019.

[37] Habtamu Tamiru D, Gedef Azene A, Wudie Tsegaye G, Mulatu Mihretie K, Hunegnaw Asmare S, Arega Gete W, et al. Time to Recovery from COVID-19 and Its Predictors in Patients Hospitalized at Tibebe Ghion Specialized Hospital Care and Treatment Center, A Retrospective Follow-Up Study, North West Ethiopia. Glob Health Epidemiol Genom. 2023;2023:5586353.

[38] Wynberg E, van Willigen HDG, Dijkstra M, Boyd A, Kootstra NA, van den Aardweg JG, et al. Evolution of Coronavirus Disease 2019 (COVID-19) symptoms during the first 12 months after illness onset. Clin Infect Dis. 2022;75:e482–e90.

[39] Tsampasian V, Elghazaly H, Chattopadhyay R, Debski M, Naing TKP, Garg P, et al. Risk Factors Associated With Post−COVID-19 Condition: A Systematic Review and Meta-analysis. JAMA Internal Medicine. 2023;183:566-80.

[40] Armann J, Doenhardt M, Hufnagel M, Diffloth N, Reichert F, Haas W, et al. Risk factors for hospitalization, disease severity and mortality in children and adolescents with COVID-19: Results from a nationwide German registry. medRxiv. 2021:2021.06.07.21258488.

[41] Shi Y, Strobl R, Berner R, Armann J, Scheithauer S, Grill E. Six clinical phenotypes with prognostic implications were identified by unsupervised machine learning in children and adolescents with SARS-CoV-2 infection: results from a German nationwide registry. Respir Res. 2024;25:392.

[42] Nielsen F. Hierarchical Clustering. 2016. p. 195-211.

[43] Schons M, Pilgram L, Reese J-P, Stecher M, Anton G, Appel KS, et al. The German National Pandemic Cohort Network (NAPKON): rationale, study design and baseline characteristics. European Journal of Epidemiology. 2022;37:849-70.

[44] Horn A, Krist L, Lieb W, Montellano FA, Kohls M, Haas K, et al. Long-term health sequelae and quality of life at least 6 months after infection with SARS-CoV-2: design and rationale of the COVIDOM-study as part of the NAPKON population-based cohort platform (POP). Infection. 2021;49:1277-87.

[45] Smith BW, Dalen J, Wiggins K, Tooley E, Christopher P, Bernard J. The brief resilience scale: Assessing the ability to bounce back. International Journal of Behavioral Medicine. 2008;15:194-200.

[46] Shi Y, Strobl R, Apfelbacher C, Bahmer T, Geisler R, Heuschmann P, et al. Persistent symptoms and risk factors predicting prolonged time to symptom-free after SARS-CoV-2 infection: an analysis of the baseline examination of the German COVIDOM/NAPKON-POP cohort. Infection. 2023;51:1679-94.

[47] Kleinbaum DG, Klein M. Survival analysis: a self-learning text: Springer; 2012.

[48] Srivastava K. Positive mental health and its relationship with resilience. Ind Psychiatry J. 2011;20:75-6.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to Prof. Dr. Eva Grill, without whom my PhD journey would not have been possible. Thank her for supervising my PhD project, giving constructive and timely feedback, showing me the beauty of rigorous and appealing science, and being supportive in my academic pursuits. Her commitment to academic excellence and encouragement to pursue innovative approaches have been invaluable to my personal and professional growth.

Same gratitude also goes to my other members in my thesis advisory committee: Prof. Dr. Klaus Jahn and PD Dr. Daniela Koller, whose guidance and support have been an indispensable part of this journey. I would like to extend a special thanks to Dr. Ralf Strobl, Dr. Benedict Katzenberger, and all other members in our working group, who generously contributed their time and expertise for my PhD project. I am fortunate to have benefited from such a collaborative environment.

I sincerely thank my sponsor - China Scholarship Council (CSC) for funding my whole PhD studies in Germany. I also thank the European University Alliance for Global Health (EUGLOH) for funding my studies in partner universities across Europe, enabling me to expand my academic perspectives and connections.

My deepest gratitude also goes to my family and friends, who have surrounded me with love, unwavering support, and belief in my capabilities. Thank you, Michael Rückl, for bringing patience, tenderness, and happiness. A heartfelt thank you goes to Faiza Shahid, whose unconditional support and heartfelt prayers have been a source of strength and motivation. I am also grateful to Xianfeng Chen, Miaojv Lai, Junhua Liu, and Chenyv Wang for standing by me during moments of difficulty.

Last but not least, thank myself for diligence, dedication, and resilience. This PhD represents the culmination of years of learning, growth, and determination, and I am proud of the journey that has brought me here.

List of all publications

Yanyan Shi, Ralf Strobl, Reinhard Berner, Jakob Armann, Simone Scheithauer, et al. Six clinical phenotypes with prognostic implications were identified by unsupervised machine learning in children and adolescents with sars-cov-2 infection: Results from a German nationwide registry.Respiratory Research. 2024,25(1):392.

Yanyan Shi, Ralf Strobl, Christian Apfelbacher, Thomas Bahmer, Ramsia Geisler, et al. Persistent symptoms and risk factors predicting prolonged time to symptom-free after SARS-CoV-2 infection: An analysis of the baseline examination of the German COVIDOM/NAPKON-POP cohort. Infection. 2023,51(6):1679-1694.

Yanyan Shi, Yihui Zheng, Yan Liu, Jian Shen, Shengnan Wang, et al. GeneXpert MTB/RIF evaluation in diagnosis of pulmonary tuberculosis. Shanghai Journal of Preventive Medicine. 2020,32(2):151-155. (In Chinese)

Yanyan Shi, Shuhua Ren, Xianfeng Chen, Yi Sun. Effect of Nurses' Risk Perception of Occupational HIV exposure on their Attitude and Behavioral Intention of Providing Nursing Care. Chinese Journal of Social Medicine.2014,31(2):115-117. (In Chinese)

Yan Liu, Shengnan Wang, Yanyan Shi, Lingying Shan, Yihui Zheng. Study on Influencing Factors of the Incidence of Close Contact with Tuberculosis Patients in Putuo District, Shanghai. Health Education and Health Promotion. 2020,15(3):274-276. (In Chinese)

Yan Liu, Yihui Zheng, Jun Chen, Yanyan Shi, Lingying Shan, et al. Tuberculosis-associated mortality and its risk factors in a district of Shanghai, China: a retrospective cohort study. International Journal of Tuberculosis & Lung Disease, 2018,22(6):655-660.

Obiageli Crystal Oluka, Yanyan Shi, Shaofa Nie, Yi Sun. Boosting Cancer Survival in Nigria: Self-management Strategies. Asian Pacific Journal of Cancer Prevention. 2014,15(1):335-341.