Ecole Doctorale BioSE (Biologie-Santé-Environnement)

Thèse

Présentée et soutenue publiquement pour l'obtention du titre de

DOCTEUR DE l'UNIVERSITE DE LORRAINE

Mention : « Sciences de la Vie et de la Santé »

Par **Aurélie FISCHER**

Titre :

Développement d'une solution digitale de santé basée sur la voix pour le suivi en vie réelle de symptômes liés au COVID Long

Présentée et soutenue publiquement le 11 décembre 2024

Membres du jury :

Présidente du jury :	
Pr Nelly AGRINIER	Professeur, HDR, APEMAC, Université de Lorraine
Rapporteurs :	
Pr Sylvie GROSJEAN	Professeur, Université d'Ottawa, Ottawa
Pr Jade GHOSN	PU-PH, HDR, Université Paris Cité, Paris
Examinateurs :	
Dr Pauline OUSTRIC	PhD, Centre Léon Bérard, Lyon
Dr Charles BENOY	PhD, Centre hospitalier neuro-psychiatrique, Ettelbruck
Directeur de thèse :	
Dr Guy FAGHERAZZI	PhD, HDR, Luxembourg Institute of Health

Membre invité :

Dr Manon GANTENBEIN — PhD, Luxembourg Institute of Health

Deep Digital Phenotyping Research Unit - Luxembourg Institute of Health
1AB, rue Thomas Edison
L-1445 Luxembourg

Table des matières

Remerciements

Cette thèse est le fruit d'un long cheminement personnel. Se lancer dans un projet de thèse sur le tard, avec une grande famille à charge et en étant déjà en poste, retrouver un statut d'étudiant, n'était pas tous les jours de tout repos, de nombreux doutes ont traversé ma route et je suis heureuse d'avoir réussi à mener ce projet à terme. La thèse est bien plus que l'obtention d'un titre, c'est aussi apprendre, dans sa discipline de recherche mais aussi sur soi-même et cela a contribué à mon développement personnel au-delà de mes attentes.

Je tiens en premier lieu à remercier ***Dr Guy Fagherazzi****, directeur du département de santé de précision et de l'équipe de recherche Deep Digital Phenotyping du Luxembourg Institute of Health (LIH), mon superviseur. Sans toi et ta constante confiance en mes capacités, ce projet de thèse n'aurait jamais existé. Tu as su me guider durant les trois dernières années, chasser les doutes qui m'empêchaient parfois d'avancer et me pousser à me dépasser et à prendre confiance en moi. Je t'en serai toujours reconnaissante et j'espère que notre collaboration durera encore longtemps.*

Je remercie également les membres de mon jury de thèse. Je remercie tout particulièrement ***Dr Sylvie Grosjean*** *et* ***Dr Jade Ghosn*** *pour avoir accepté la charge de rapporteurs de cette thèse. Un grand merci également à* ***Dr Pauline Oustric*** *et* ***Dr Charles Benoy*** *d'avoir accepté le rôle d'examinateurs. Je suis également très reconnaissante envers* ***Dr Nelly Agrinier*** *d'avoir accepté d'évaluer ma thèse comme examinateur et de présider ce jury.*

Je tiens également à remercier l'école doctorale BIOSE de l'université de Lorraine et en particulier sa directrice, ***Dr Sophie Rahuel-Clermont****,* ***Manon Sage*** *et* ***Mélanie Neveu****, pour avoir patiemment répondu à toutes mes questions. Je remercie le LIH de m'avoir permis de mener ce projet et pour avoir financé cette thèse.*

Merci aux membres de mon comité de suivi de thèse, ***Dr Nelly Agrinier, Dr Manon Gantenbein et Dr Charles Benoy,*** *pour vos précieux conseils et encouragements qui m'ont maintenue motivée pendant tout ce parcours.*

Je remercie aussi toute l'équipe Deep Digital Phenotyping. Travailler avec vous tous est un plaisir, nous avons partagé bien plus que du professionnel ensemble et c'est grâce à l'ambiance détendue, consciencieuse et chaleureuse que j'y sens si bien et que j'ai été portée pendant ces dernières années. Tout d'abord, merci ***Gloria*** *de partager avec moi notre bureau, merci pour ta bonne humeur constante, tes conseils avisés et nos échanges plus personnels.* ***Abir****,* ***Mégane****,* ***Maurane*** *et* ***Charline****, les amoureuses des chats, merci aussi pour votre bonne humeur, nos échanges de photos de nos animaux préférés, mais aussi pour vos conseils et toutes nos discussions professionnelles et extra-professionnelles. Merci enfin à* ***Hanin****,* ***Dulce, Noémie*** *et* ***Kevser****, toujours souriantes et à l'écoute.*

Merci ***Vincent****, toi qui partage ma vie depuis près de 15 ans. Merci pour ton amour et ton soutien sans faille qui m'ont porté pendant ce cheminement. Et merci pour nos discussions animées qui ont contribué à développer mon esprit critique !*

***Anaïs**, **Solène**, **Corentin** et **Jules**, mes enfants chéris, je sais que cela n'a pas toujours été facile de supporter une maman étudiante, mais c'est aussi grâce à vous, à la famille imparfaite que nous formons que j'ai pu aller jusqu'au bout de cette thèse et j'espère que vous serez fiers de moi.*

Merci à tous les amis qui m'entourent, je ne les citerai pas tous mais vous m'avez aidée à décompresser quand j'en avais besoin et votre présence m'est précieuse.

*Enfin, je voudrais dédier cette thèse à **mes parents** qui sont partis bien trop tôt. Je vous remercie pour l'éducation que vous m'avez donnée, les valeurs essentielles dans la vie, le respect du travail et de la famille. Vous vous êtes mis en quatre pour que je puisse faire tout ce dont j'avais envie, vous avez toujours été plus que présents pour moi et les enfants. **Papa**, tu as été le pilier sur lequel je pouvais me reposer, la force tranquille. **Maman**, lors de notre dernière discussion, pendant laquelle je t'ai annoncé que j'allais me lancer dans une thèse, tu m'as demandé en quoi cela consistait, je t'ai simplement dit que celà me permettrait de me faire appeler Docteur. Bien sûr c'est beaucoup plus que celà et j'espère que de là-haut vous êtes fiers de moi.*

Abréviations

ACE2	Enzyme de Conversion de l'Angiotensine 2
ARNm	*Acide RiboNucléique messager*
BV	*Biomarqueur Vocal*
CAH	Classification Ascendante Hiérarchique
CE	Conformité Européenne
CL	COVID Long
CNER	*Comité National d'Ethique de la Recherche*
CoVaLux	*COVID-19, Vaccination & Long-term health consequences of COVID-19 in Luxembourg*
COVID-19	*Maladie à Coronavirus 2019*
CRF	Case Report Form
DDP	Deep Digital Phenotyping
DPIA	Analyse d'Impact sur la Protection des Données
EM/SFC	Encéphalomyélite Myalgique/Syndrome de Fatigue Chronique
EMA	European Medicines Agency
EQ5D-5L	EuroQoL 5-Dimension 5-Level
ESCMID	European Society of Clinical Microbiology and Infectious Diseases
FDA	Food and Drug Administration
FSS 9	Fatigue Severity Scale 9
GAD 7	Generalized Anxiety Disorder 7
GISAID	*Global Initiative on Sharing All Influenza Data*
HCP	*Healthcare Professional*
HHV-6	*Herpès Virus humain 6*
IACC	*Infection-Associated Chronic Condition*
IRM	Imagerie à Résonance Magnétique
LCMM	Modèles Mixtes à Classes Latentes
LIH	*Luxembourg Institute of Health*

NASEM	National Academies of Sciences, Engineering, and Medicine
NICE	*National Institute for Health Excellence*
NIHR	National Institute for Health and Care Research
OECD	*Organisation for Economic Co-operation and Development*
OMS	Organisation Mondiale de la santé
PCR	Réaction de Polymérisation en chaîne
PET scan	Tomographie par Émission de Positrons
PIPEDA	*Personal Information Protection and Electronic Documents Act*
POTS	Syndrome de Tachycardie Orthostatique Posturale
PPI	*Patient and Public Involvement*
PSQI	*Pittsburgh Sleep Quality Index*
PSS 4	*Perceived Stress Scale 4*
PWLC	*People living With Long COVID*
RGPD	Règlement Général sur la Protection des Données
SARS-Cov 2	Coronavirus du Syndrome Respiratoire Aigu Sévère
SF12	*Short Form 12-items*
UUID	IDentifiant Utilisateur Unique
VOC	Variant of Concern
VOI	Variant of Interest

Production scientifique

Articles faisant partie de cette thèse

Aurélie Fischer, Abir Elbéji, Gloria Aguayo, Guy Fagherazzi. Recommendations for Successful Implementation of the Use of Vocal Biomarkers for Remote Monitoring of COVID-19 and Long COVID in Clinical Practice and Research. Interact J Med Res. 2022 Nov 15;11(2):e40655. DOI : 10.2196/40655 **(Publié)**

Aurélie Fischer, Lu Zhang, Abir Elbéji, Paul Wilmes, Pauline Oustric, Therese Staub, Petr V. Nazarov, Markus Ollert, Guy Fagherazzi. Long COVID symptomatology after twelve months and its impact on quality of life according to initial COVID-19 disease severity. Open Forum Infect Dis. 2022 Aug 5;9(8):ofac397. DOI : 10.1093/ofid/ofac397 **(Publié)**

Aurélie Fischer, Nolwenn Badier, Lu Zhang, Abir Elbéji, Paul Wilmes, Pauline Oustric, Charles Benoy, Markus Ollert, and Guy Fagherazzi. Long COVID classification: findings from a clustering analysis in the predi-COVID cohort study. Int. J. Environ. Res. Public Health 2022, 19(23), 16018; DOI : 10.3390/ijerph192316018 **(Publié)**

Aurélie Fischer, Lu Zhang, Abir Elbéji, Paul Wilmes, Chantal Snoeck, Jérôme Larché, Pauline Oustric, Markus Ollert, Guy Fagherazzi. Trajectories of persisting COVID-19 symptoms up to 24 months after acute infection: Findings From The Predi-COVID Cohort Study. BMC Infectious Diseases. 2024 **(Soumis)**

Aurélie Fischer, Gloria Aguayo, Pauline Oustric, Laurent Morin, Charles Benoy, and Guy Fagherazzi. Co-design of a voice-based digital health solution to monitor persisting symptoms related to COVID-19: protocol for the UpcomingVoice mixed-methods study. JMIR Res Prot. 2023. DOI : 10.2196/46103 **(Publié)**

Aurélie Fischer, Gloria Aguayo, India Pinker, Pauline Oustric, Jérôme Larché, Charles Benoy, and Guy Fagherazzi. Co-design of the Long COVID Companion voice-based app to monitor Long COVID symptoms with its end-users: A mixed-method study. Digital Health. 2024. DOI : 10.1177/20552076241272671 **(Publié)**

Articles associés aux travaux de cette thèse

Pour les articles suivants j'ai été impliquée dans la conception de l'étude, la préparation des données, l'interprétation des résultats, la relecture, et/ou la coordination des travaux.

Guy Fagherazzi, **Aurélie Fischer**, Fay Betsou, et al. Protocol for a prospective, longitudinal cohort of people with COVID-19 and their household members to study factors associated with disease severity: the Predi-COVID study. BMJ Open. 2020;10(11):e041834. Published 2020 Nov 23.

Guy Fagherazzi and Lu Zhang and Abir Elbéji and Eduardo Higa and Vladimir Despotovic and Markus Ollert and Gloria Aguayo and Petr Nazarov and **Aurélie Fischer**. A Voice-Based Biomarker for Monitoring Symptom Resolution in Adults with COVID-19: Findings from the Prospective Predi-COVID Cohort Study. PLOS Digital Health 2022 Oct; 1(10): e0000112.

Abir Elbéji, Lu Zhang, Eduardo Higa, **Aurélie Fischer**, Vladimir Despotovic, Petr V Nazarov, Gloria Aguayo, Guy Fagherazzi . Vocal biomarker predicts fatigue in people with COVID-19: results from the prospective Predi-COVID cohort study. BMJ Open. 2022;12(11):e062463. Published 2022 Nov 22.

Eduardo Higa, Abir Elbéji, Lu Zhang, **Aurélie Fischer**, Gloria A Aguayo, Petr V Nazarov , Guy Fagherazzi. Discovery and Analytical Validation of a Vocal Biomarker to Monitor Anosmia and Ageusia in Patients With COVID-19: Cross-sectional Study. JMIR Med Inform. 2022;10(11):e35622. Published 2022 Nov 8.

Gloria A. Aguayo, **Aurélie Fischer,** Abir Elbéji, Nyan Linn, Markus Ollert and Guy Fagherazzi. Association between use of psychotropic medications prior to SARS-COV-2 infection and trajectories of COVID-19 recovery: Findings from the prospective Predi-COVID cohort study. Front. Public Health, 16 March 2023 Sec. Public Mental Health Volume 11 - 2023

Autres publications

Abir Elbéji, Mégane Pizzimenti, Gloria Aguayo, **Aurélie Fischer**, Hanin Ayadi, Franck Mauvais-Jarvis, Jean-Pierre Riveline, Vladimir Despotovic, Guy Fagherazzi. A voice-based algorithm can predict type 2 diabetes status in USA adults: Findings from the Colive Voice study. PLOS Digital Health (Submitted)

Laurent Malisoux, Anne Backes, **Aurélie Fischer**, Gloria Aguayo, Markus Ollert, Guy Fagherazzi. Associations between physical activity prior to infection and COVID-19 disease severity and symptoms: results from the prospective Predi-COVID cohort study. BMJ Open. 2022;12(4):e057863. Published 2022 Apr 29.

Hanin Ayadi, Charline Bour, **Aurélie Fischer**, Myriam Ghoniem, Guy Fagherazzi. The Long COVID experience from a patient's perspective: a clustering analysis of 27,216 Reddit posts. Frontiers in Public Health, 11, 1227807.

Charline Bour, Adrian Ahne, Gloria Aguayo, **Aurélie Fischer**, David Marcic, Philippe Kayser, Guy Fagherazzi. Global diabetes burden: analysis of regional differences to improve diabetes care. BMJ Open Diabetes Res Care. 2022; 10(5): e003040.

Communications scientifiques

Présentations orales conférences

Long COVID in Luxembourg: a translational project to develop a digital health app to monitor Long COVID symptoms. (Conférence du consortium CoVaLux. 15 Novembre 2024)

Biomarqueurs vocaux et COVID Long. (Séminaire Parcours de soins territorial et pluridisciplinaire COVID long en Occitanie. Bilan à deux ans et perspectives. CHU Montpellier, Mai 2024)

Co-design of a digital app to monitor Long COVID symptoms and their impact on quality of life. (2nd Congress of the Medical Association Long COVID, Germany, Novembre 2023, Invited speaker)

Frequencies, co-occurrence and risk factors of Long COVID symptoms after one year according to the initial COVID-19 disease severity: results from the prospective Predi-COVID cohort study (Physiological society, Long COVID: Mechanisms, Risk Factors, and Recovery, 22 février 2022)

Development of a digital health solution based on vocal biomarkers to remotely monitor frequent symptoms in real-life in people with Long COVID (Journée de rentrée de l'école doctorale BIOSE, Nancy, Novembre 2022)

Posters

Co-design of a digital application to monitor Long COVID symptoms: results from the UpcomingVoice study (ESCMID. Barcelone, 27-30 avril 2024)

Long-term trajectories of COVID-19 symptoms and their associations with quality of life : results from the 2-year-follow-up of the prospective Predi-COVID study (ESCMID. Barcelone, 27-30 avril 2024)

Long-term trajectories of COVID-19 symptoms and their associations with quality of life : results from the 2-year-follow-up of the prospective Predi-COVID study (Journée scientifique de l'école doctorale BIOSE, Nancy, 23 avril 2024)

UpcomingVoice : attentes des utilisateurs et co-construction d'une application de santé digitale utilisant la voix pour la détection et le suivi de symptômes liés au COVID long. (Journée Nationale COVID Long. Nancy, 8 Décembre 2022)

Frequencies and impact of Long COVID symptoms one year after the acute phase: results from the prospective Predi-COVID cohort study (ESCMID conference, online, 25 Avril 2022)

Autres communications

Participation à une table ronde intitulée **"Navigating the clinical landscape, research gaps and priorities: an international perspective on long COVID"** lors de la conférence sur le COVID Long organisée par l'OECD (Organisation for Economic Co-operation and Development). (10 Septembre 2024).

Présentation des travaux de recherche sur les biomarqueurs vocaux lors de la réunion trimestrielle du département Precision Health (Septembre 2022)

Présentation des travaux sur le COVID long lors d'une réunion du consortium CoVaLux (COVID-19, Vaccination & Long-term health consequences of COVID-19 in Luxembourg. Novembre 2022)
Chapitre d'ouvrage: Your health goes through your voice: an introduction to vocal biomarkers. (1er auteur). LIH Precision Health Book (2022). https://precisionhealth.lu/en

Présentation de l'étude Predi-COVID lors de la Journée de la Recherche Translationnelle à destination d'un public de lycéens (Décembre 2021)

Activité d’enseignement et d’encadrement stagiaires

“Patient Reported Outcome MeasureS, Patient Reported Experience MeasureS and digital biomarkers” - Cours dans le cadre du bachelor de Médecine - Université de Luxembourg - Mars 2024

“Digital epidemiology” - Master 2 Digital Public Health Bordeaux - 15 Octobre 2024

Encadrement stage Master 1 Mlle Nolwenn Badier (Mai-Août 2023) : Long COVID clusters and determinants

Encadrement stage Master 2 Mlle Zoi Mavroedi (Avril-Octobre 2023) : Improvement of explainability of Artificial Intelligence algorithms

Activité de révision d’articles

BMC Public Health (2024) : "Follow-up of long COVID based on the definition of WHO: a multi-centre cross-sectional questionnaire study"

BMC Medicine (2024) : “The dream is that there's one place you go”: A Qualitative Study of Women’s Experiences Seeking Care from Long COVID Clinics in the U.S.

J. Clin. Epidemiology (2024) : “Clusters of post-acute COVID-19 symptoms: a latent class analysis across 9 databases and 7 countries”

1. Introduction générale

1.1 Le COVID Long

Origine

Pandémie de COVID-19 - Depuis Février 2020, la pandémie de COVID-19 a mobilisé les acteurs politiques et de recherche dans le monde entier. En effet, la maladie de COVID-19 causée par le coronavirus du syndrome respiratoire aigu sévère (SARS-Cov 2) a touché plus de 775 millions de personnes et causé plus de 7 millions de morts dans le monde[1]. En décembre 2022, la surmortalité due à la pandémie a été estimée à 14,83 millions par l'Organisation Mondiale de la Santé (OMS), soit plus de 2 fois l'estimation des décès directement imputables au SARS-Cov 2[2].

Variants - Comme pour la majorité des virus, différents variants du SARS-Cov 2 se sont succédé depuis le début de la pandémie. Le premier variant apparu fin 2020 était le variant Alpha, puis d'autres variants nommés selon l'alphabet grec ont suivi jusqu'à ce que le variant Omicron apparaisse en novembre 2021 et devienne rapidement majoritaire[3]. Ces variants dits préoccupants (*Variant of Concern*, VOC) sont aujourd'hui considérés comme non circulants. Cependant de nombreux autres variants dits d'intérêt (*Variants of Interest*, VOI) sont aujourd'hui circulants mais aucun n'est considéré actuellement comme préoccupant. La dénomination des VOI est basée sur 3 différentes nomenclatures scientifiques des virus (Global Initiative on Sharing All Influenza Data (GISAID), Nextstrain et Pango) ce qui rend leur suivi difficile[4].

Impact de la pandémie - Pour tenter d'enrayer la pandémie, différentes mesures ont été mises en place par les pays, comme des mesures de confinement, de distanciation sociale ou de traçage des contacts[5]. Le monde scientifique s'est en parallèle mobilisé sur la recherche de traitements et un vaccin a été développé comme mesure additionnelle de prévention.

Manifestations cliniques de l'infection initiale à SARS-Cov 2 - L'infection initiale à SARS-Cov 2 peut se présenter sous des formes cliniques de sévérité variable, allant de la forme asymptomatique jusqu'à des formes sévères en passant par des formes plus modérées. Les symptômes les plus courants de la forme aigüe de la maladie sont la fièvre, la toux, la dyspnée, les douleurs musculaires et la fatigue[6].

Définitions du COVID Long - Dans la majorité des cas, les symptômes se résolvent dans les 3-4 semaines suivant le début de la maladie. Cependant, une fraction des personnes touchées par une infection à SARS-Cov 2 voient leurs symptômes persister ou de nouveaux symptômes apparaître plusieurs semaines ou mois après.

Face à la difficulté de faire reconnaître la réalité de leur condition par le système de santé pendant les premiers mois de la pandémie, ce sont les patients eux-mêmes qui sont à l'origine du terme COVID Long (CL), qui englobe les conséquences à long terme d'une infection COVID-19[7].

Une définition du CL a ensuite été proposée par l'OMS en Novembre 2021 et indique que "L'affection post COVID-19 survient chez des personnes présentant des antécédents d'infection probable ou confirmée par le SARS-Cov 2, généralement 3 mois après l'apparition de COVID-19 avec des symptômes qui persistent au moins 2 mois et qui ne peuvent être expliqués par un autre diagnostic."[8]

La définition du CL a été sujette à discussions multiples et en juin 2024 une nouvelle définition a été proposée par la NASEM (*National Academies of Sciences, Engineering, and Medicine*) et semble faire consensus : "Le COVID long est une maladie chronique associée à une infection (*Infection-Associated Chronic Condition, IACC*) qui survient après une infection par le SARS-Cov 2 et qui est présente pendant au moins trois mois sous la forme d'une maladie continue, récurrente et rémittente, ou progressive affectant un ou plusieurs systèmes d'organes."[9]

Epidémiologie

Le CL peut toucher des personnes de tout âge, adultes et enfants, quel que soit le genre, l'état de santé initial ou le pays d'origine[10]. Les estimations de sa prévalence sont très variables car soumises à l'hétérogénéité des données disponibles. Une revue systématique sur 120 études a montré que la prévalence des symptômes persistants après 12 semaines était comprise entre 0 et 93% avec une prévalence groupée de 42.1%[11]. Une autre méta-analyse sur 194 études a montré un résultat similaire avec 45% des survivants de COVID-19 qui présentaient des symptômes non résolus à 4 mois, et ce indépendamment du statut d'hospitalisation initial[12].

Il est aujourd'hui généralement accepté que 10 à 20% des personnes touchées par le COVID-19 développent un CL[13], ces chiffres variant de 10 à 30% pour les personnes non hospitalisées, à 50 - 70% pour les personnes hospitalisées, et à 10 - 12% chez les personnes vaccinées[14]. Une récente étude estime que 7% de l'ensemble des personnes adultes aux Etats-Unis a ou a été touchée par le CL[15]. Une étude en France a évalué la prévalence du CL dans la population générale adulte à 4%[16].

Finalement, la prévalence du CL semble converger aux environs de 6 - 7% chez les adultes et 1% chez les enfants, ce qui représente plus de 400 millions de cas cumulés en 2024, ainsi qu'un impact économique annuel d'environ 1% de l'économie globale, soit environ un milliard de milliard de dollars américains[17]. Le CL représente ainsi un enjeu majeur de santé publique, avec de nombreuses problématiques liées à la prise en charge médicale, à la reprise du travail et aux arrêts maladies, ou à la reconnaissance du CL comme affection de longue durée.

Une étude prospective en France a mis en évidence que 62.2% des personnes touchées par le CL ont dû arrêter leur activité professionnelle et seules 32.5% ont pu reprendre une activité professionnelle à plein temps[18]. Aux Etats-Unis, il a été chiffré qu'à tout moment plus d'un million de personnes ont arrêté de travailler à cause du CL, ce qui conduit à une perte de revenus d'environ 50 000 milliards de dollars par an[19].

Facteurs de risques

Les principaux facteurs de risque de développer un CL semblent être la présence de **comorbidités** comme l'asthme, le diabète de type 2, l'obésité, l'hypertension, l'hypothyroïdie, **l'âge**, le fait d'**être une femme**, et une **infection initiale sévère**[20–22]. Sur ce dernier point, bien que le risque relatif soit plus élevé après une forme initiale sévère, il a été estimé que plus de 90% des cas de CL concernent des personnes ayant eu une forme initiale modérée, cette forme étant la plus répandue[10]. Enfin, le risque d'hospitalisation et de développer au moins une séquelle à long terme augmente à chaque nouvelle infection par le SARS-Cov 2[23,24] et la **réinfection** est associée à un plus grand nombre de symptômes du CL[25].

Impact de la vaccination - Plusieurs études se sont penchées sur l'impact de la vaccination contre le SARS-Cov 2 sur le risque de développer un CL. La majorité d'entre elles s'accordent sur le fait qu'une vaccination complète, soit 2 doses, pré-infection, permet de limiter le risque de développer un CL et les séquelles à long terme de la majorité des catégories de symptômes, avec cependant des Odds-Ratios variables, dépendant du design de l'étude, et de la définition du CL utilisée [26,27]. Une récente étude aux Etats-Unis a estimé qu'une vaccination complète avant infection diminuait la prévalence de CL de 40 à 60%[28]. De même, deux études de cohortes en Corée du Sud et Japon ont plus récemment montré un effet protecteur de la vaccination pré-infection à SARS-Cov 2 sur les conséquences à long terme respiratoires[29] et neuropsychiatriques[30].

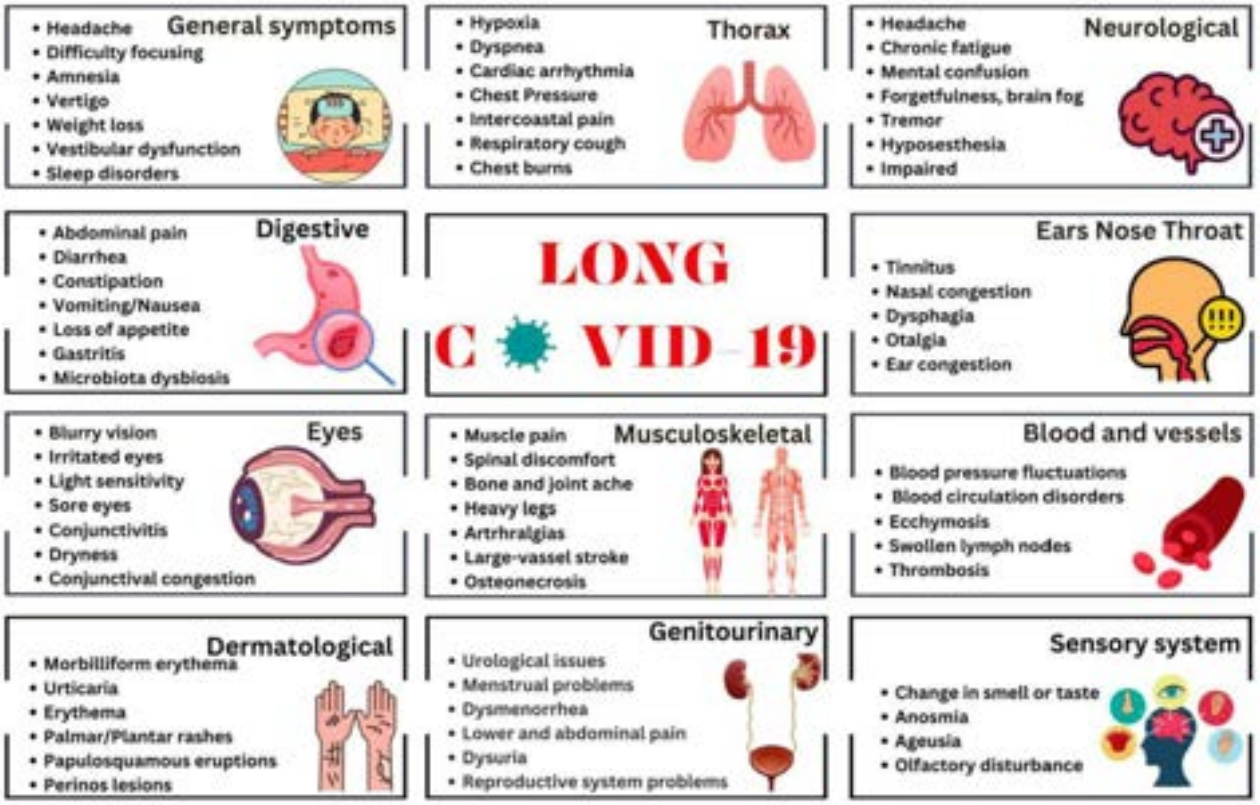
Les effets de la vaccination post-infection à SARS-Cov 2 sont eux moins clairs[31]. De même, les effets de la vaccination après le développement d'un CL n'ont été clairement établis[32]. Une étude sur un petit nombre de patients a montré un effet bénéfique avec une diminution du nombre de symptômes et une amélioration du bien-être[33], alors qu'une autre concluait que les effets de la vaccination différaient selon les personnes, avec une amélioration des symptômes pour 16.7% des personnes, une dégradation pour 21.4% et une absence de changement pour les autres personnes[34].

Impact des variants - Les données sont limitées concernant l'impact du variant sur le risque de développer un CL. Quelques études suggèrent que les infections avec les variants précédant Omicron, et en particulier le variant Delta, étaient des facteurs de risque de CL[32,35–37]. Cependant une étude récente a montré que cet effet protecteur du variant Omicron sur le risque de CL diminuait avec le temps et que finalement le risque d'évolution vers un CL était similaire entre les différents variants[38].

Manifestations cliniques

D'un point de vue clinique, le CL est multi-systémique, et plus de 200 symptômes touchant différents organes ont été décrits[14]. Les symptômes typiques du CL sont des symptômes généraux comme la fatigue, la perte de goût et d'odorat, les problèmes de sommeil, les atteintes cardio-respiratoires, neurocognitives, musculo-squelettiques, psychologiques, circulatoires, dermatologiques et gastro-intestinales. La **Figure 1** synthétise les symptômes groupés par catégories.

Figure 1 : Synthèse des principaux symptômes



Gheorgita, R. et al. The knowns and unknowns of long COVID-19: from mechanisms to therapeutical approaches. Front. Immunol., 15:1344086 (2024). https://doi.org/10.3389/fimmu.2024.1344086

Fatigue - La fatigue est le symptôme le plus fréquemment rapporté, avec des fréquences allant de 11 à 87% en fonction des études, dépendant de la sévérité de l'infection initiale et du temps de suivi[20,39,40]. On peut distinguer la fatigue chronique qui se caractérise par un manque d'énergie constant du phénomène de malaise post-effort qui décrit une intolérance aux efforts physiques et intellectuels et aux événements émotionnels et environnementaux stressants et qui peut être déclenchée par des activités de la vie de tous les jours. La fatigue chronique et les malaises post-effort ressentis par les personnes souffrant de CL peuvent évoquer ceux ressentis par les personnes souffrant d'encéphalomyélite myalgique/syndrome de fatigue chronique (EM/SFC) mais malgré des similarités, des différences ont été rapportées au niveau de la manifestation clinique du malaise post-effort entre les deux maladies[41]. Plusieurs études décrivant l'évolution des symptômes du CL ont montré que la fatigue reste prédominante 1 an voire 2 ans après l'infection initiale[42–44].

Symptômes neurocognitifs - Les symptômes neurocognitifs typiques sont les problèmes de concentration, une confusion mentale souvent décrite comme un brouillard mental, les troubles de la mémoire, les maux de tête ou migraine, les problèmes d'équilibre, la paresthésie (sensations de picotements ou d'engourdissements dans les extrémités) ou une hypersensibilité au bruit et à la lumière[45]. Une étude publiée en Juillet 2024 a également mis en évidence une perte mesurable du déficit cognitif, avec une diminution équivalente à une perte de 6 points de QI chez les personnes ayant des symptômes persistants non résolus[46]. On retrouve également la perte ou l'altération du goût et de l'odorat qui touche entre 10 et 20% des personnes avec un CL[20], mais le symptôme semble s'améliorer avec le temps avec une prévalence d'environ 5% à 3 ans[47]. Les problèmes de sommeil comme les insomnies ou une somnolence diurne accrue sont fréquemment rapportés et impactent fortement la qualité de vie des personnes concernées. Enfin, des acouphènes, des vertiges ou des pertes d'audition ont également été rapportés[20].

Symptômes cardio-respiratoires - Parmi les atteintes cardio-respiratoires du CL on retrouve la dyspnée, les douleurs thoraciques, la toux, les expectorations, mais aussi les palpitations et les troubles du rythme cardiaque, en particulier la tachycardie. Cette dernière peut être au repos, une réponse exagérée à un exercice physique, ou apparaître de manière inappropriée lors de la station debout, alors appelée syndrome de tachycardie orthostatique posturale (POTS)[20]. La fréquence globale des symptômes cardiorespiratoires est de 15%[48].

Symptômes musculo-squelettiques - Les troubles musculo-squelettiques comme les douleurs musculaires ou articulaires sont également fréquemment rapportés par les personnes touchées par le CL, une méta-analyse de 36 études a estimé leurs fréquences à 13.3% et 28.8%, respectivement[48].

Symptômes vasculaires - Au niveau vasculaire, des troubles de la circulation, des fluctuations de la pression sanguine, mais aussi des thromboses et des gonflements des ganglions lymphatiques ont été rapportés.

Symptômes gastro-intestinaux - Des symptômes gastro-intestinaux peuvent également persister à long terme, en particulier la constipation, la diarrhée, les douleurs abdominales, les nausées et vomissements et les brûlures d'estomac[49].

Santé mentale - Plusieurs symptômes liés à la santé mentale sont décrits dans le cadre du CL, comme l'anxiété et la dépression, chacune rapportées par 15.7% des personnes avec un CL 24 mois après l'infection initiale[50]. Les problèmes de sommeil sont également fréquents, avec une prévalence estimée de 28.98%[51], incluant l'insomnie, les troubles respiratoires du sommeil, l'hypersomnolence diurne ou les troubles du rythme circadien veille-sommeil.

Symptômes dermatologiques - Des manifestations dermatologiques comme des éruptions cutanées (morbilliformes, papulosquameuses, urticaire) ou des engelures ont également été rapportées[52].

Symptômes urinaires et du système reproductif - Enfin, des problèmes urinaires et des troubles des systèmes reproductifs féminins (troubles du cycle menstruel, douleurs pré-menstruelles) et masculins (dysfonction érectile, réduction de la libido, entre autres) ont été décrits[53,54].

Durée des symptômes - La durée des symptômes du CL est variable, pouvant aller de quelques mois à plusieurs années. Les atteintes comme la dysautonomie et l'EM/SFC peuvent quant à elles être permanentes[55]. Des périodes d'amélioration et de rechutes sont fréquentes et la présentation des symptômes peut varier au cours du temps. Les symptômes neurocognitifs par exemple peuvent apparaître avec un délai de plusieurs semaines après l'infection initiale, et peuvent empirer avec le temps[56]. D'autres symptômes peuvent rester stables dans le temps ou s'améliorer[57].

De nombreux parallèles peuvent être faits avec d'autres syndromes post-infectieux survenant après des infections par d'autres coronavirus, les virus du Nil occidental ou Epstein-Barr, ce qui laisse présager une morbidité et une mortalité accrue plusieurs années après l'infection[58].

Finalement, bien que les connaissances sur la présentation clinique du CL augmentent, il reste encore beaucoup d'incertitudes et en particulier sur l'évolution et la probabilité de résolution complète.

Un ou plusieurs COVID Long?

La définition du CL adoptée aujourd'hui est par nature très générique et large, il est probable qu'elle regroupe en fait plusieurs réalités physiopathologiques sous-jacentes. Différentes méthodes de classification ont été utilisées dans plusieurs études.

Classification des personnes affectées par le CL - Certaines études ont catégorisé les personnes présentant des symptômes persistants selon des caractéristiques et des présentations cliniques communes et ont mis en évidence des sous-groupes de personnes avec un CL de sévérité croissante, mesurée par un nombre de symptômes croissant[59–62].

Une autre étude a effectué une classification des différents profils de personnes avec un CL sur base de données collectées par le réseau social Reddit et a classifié les personnes en fonction de catégories de symptômes et de leurs préoccupations, montrant l'hétérogénéité des profils ainsi que la prédominance de la fatigue qui était commune aux différents profils[63].

Trajectoires de l'évolution du CL dans le temps - D'autres études se sont intéressées à l'évolution de la maladie entre l'infection initiale et au maximum 24 mois après, avec des designs très variés, certaines s'intéressant à l'évolution de la fréquence des symptômes au cours du temps[44,64], alors que d'autres ont cherché à modéliser des trajectoires de symptômes, individuellement ou groupés en un score de symptômes.

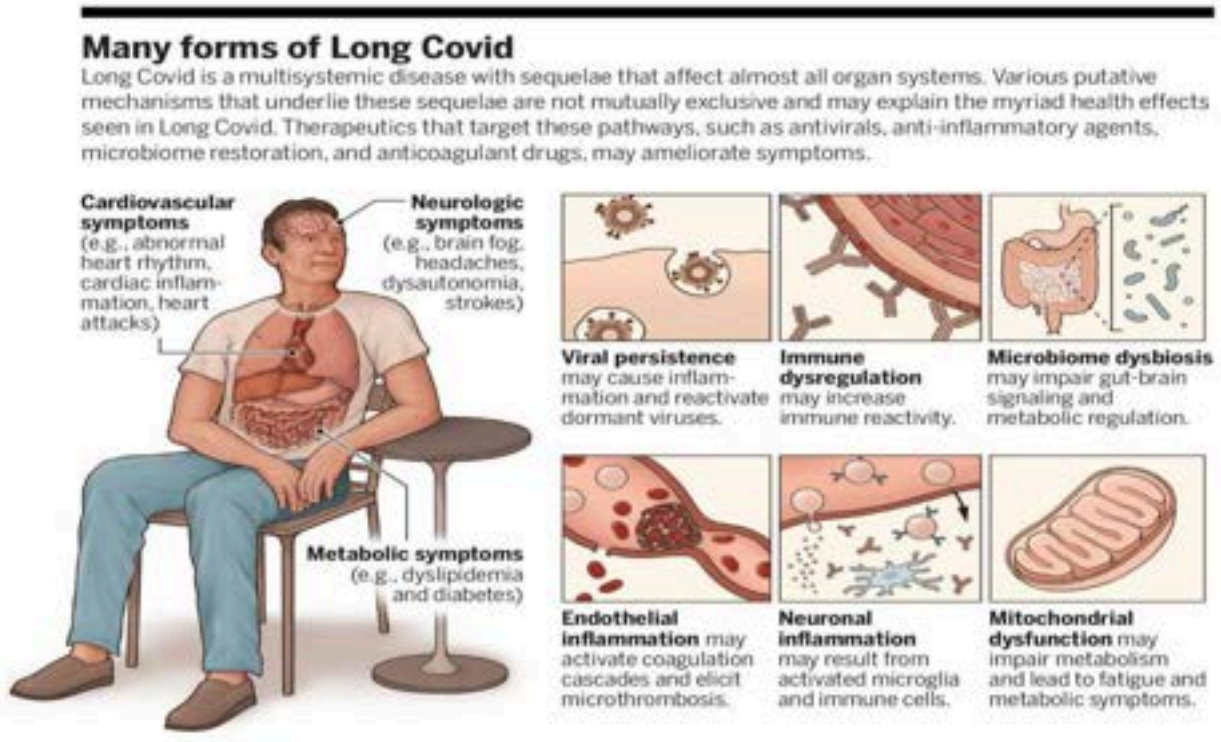
Parmi ces dernières, deux études de cohorte ont modélisé l'évolution du nombre total de symptômes, une en France qui a identifié 3 trajectoires : (1) "symptômes hautement persistants", (2) "symptômes diminuant rapidement" et (3) "symptômes diminuant lentement" comprenant 91% des participants[42]. La deuxième étude, aux Pays-Bas, a identifié 4 trajectoires différentes, avec un nombre croissant de symptômes selon la trajectoire. Dans chaque trajectoire l'évolution restait stable au cours du temps. La trajectoire la plus haute englobait 8.9% des participants et ils rapportaient plus de 6 symptômes[65].

Finalement, tous ces résultats sous-tendent qu'il ne faut probablement pas parler d'un CL mais de plusieurs CLs, avec des présentations cliniques, des sévérités et des évolutions variables. Cependant, au vu de la diversité des méthodologies d'étude, des populations d'études (hospitalisées ou non), des temps choisis pour l'évaluation des symptômes, et des symptômes choisis pour effectuer les classifications, il s'avère nécessaire de réaliser des études supplémentaires pour caractériser plus précisément les différents CLs.

Physiopathologie

Plusieurs hypothèses sur les mécanismes biologiques responsables du CL ont été avancées et sont étudiées actuellement (voir **Figure 2**).

Figure 2 : Synthèse des différents mécanismes pathophysiologiques impliqués dans le CL



Ziyad Al-Aly, Eric Topol, Solving the puzzle of Long Covid. Science 383,830-832(2024). DOI:10.1126/science.adl0867

Les 3 mécanismes principaux sont les suivants :

Persistance virale - Tout d'abord, un phénomène de persistance virale a été décrit par plusieurs études montrant que des protéines virales ou de l'ARNm du SARS-Cov 2 ont été retrouvés plusieurs mois après l'infection initiale dans différents tissus et organes, suggérant l'existence d'un réservoir viral chez certaines personnes et pouvant expliquer la persistance des symptômes[66–68]. La réactivation de virus latents de la famille des Herpès virus humains comme le virus d'Epstein-Barr ou le Herpès Virus humain 6 (HHV-6) a également été décrite comme mécanisme potentiel d'explication de l'apparition du CL en provoquant un dysfonctionnement mitochondrial et métabolique[69,70].

Dérégulation immunitaire - Ensuite, une dérégulation immunitaire pourrait aussi être impliquée dans l'apparition du CL. Une activation immunitaire permanente pourrait être induite par l'infection par SARS-Cov 2 et amener au phénomène d'épuisement immunitaire, ce qui pourrait être lié à la persistance de SARS-Cov 2 dans l'organisme. D'autre part, de nombreuses études ont montré la présence à des niveaux élevés d'auto-anticorps, notamment dirigés contre l'enzyme de conversion de l'angiotensine 2 (ACE-2, le récepteur cellulaire du SARS-Cov 2), chez des personnes présentant un CL ou une infection COVID-19 de manière plus générale[71,72]. Cependant l'impact de cette auto-immunité dans le développement du CL n'est pas encore clair[73]. Un déséquilibre dans le système Rénine-Angiotensine pourrait provoquer une tempête cytokinique et un état inflammatoire persistant et expliquer les dommages à long termes d'une infection COVID-19[66].

Inflammation endothéliale - Lors d'une infection COVID-19, la coagulation sanguine microvasculaire et la formation de micro-caillots peuvent entraîner une hypoxie des différents organes, participer indirectement au maintien de l'état inflammatoire chronique et ainsi expliquer certains symptômes persistants comme le syndrome de tachycardie orthostatique posturale[74,75].

D'autres mécanismes ont été avancés, comme la dysbiose du microbiote intestinal[76], une signalisation neurologique altérée au niveau du nerf vague et du tronc cérébral, ou une altération du métabolisme mitochondrial[58].

Finalement, tous ces mécanismes potentiels ne sont pas indépendants et peuvent interagir. Certains mécanismes sont partagés avec ceux intervenant dans le développement de l'EM-SFC, ce qui peut expliquer les similarités cliniques entre les deux pathologies[77].

Diagnostic et prise en charge du COVID Long

Malgré l'augmentation des connaissances sur le CL, il reste de nombreux défis liés au diagnostic et à la prise en charge des personnes touchées. De plus, le caractère très hétérogène du CL rend complexe son diagnostic et également l'identification de cibles thérapeutiques et de traitements.

Diagnostic - Tout d'abord, il n'existe à l'heure actuelle pas de test de diagnostic du CL. Le diagnostic repose donc aujourd'hui principalement sur la symptomatologie, complétée par l'utilisation de différents outils diagnostiques comme des techniques d'imageries (radiographies, imagerie à résonance magnétique (IRM), PET scan..)[78,79], des tests de tolérance à l'effort si appropriés, ou le test d'inclinaison (Tilt test) qui permet d'identifier les POTS. Il existe quelques recommandations pour le diagnostic et la prise en charge du CL[80,81].

Des recherches sur de potentiels biomarqueurs sanguins sont en cours. Une méta-analyse de 28 études a mis en évidence 113 biomarqueurs significativement associés au CL, incluant des cytokines (33.6%), des biomarqueurs biochimiques (21.2%), des biomarqueurs vasculaires (17.7%), mais aussi des biomarqueurs neurologiques et des protéines de l'inflammation. L'utilisation d'un panel de ces biomarqueurs pourrait aider au diagnostic du CL ou de ses différents sous phénotypes[82].

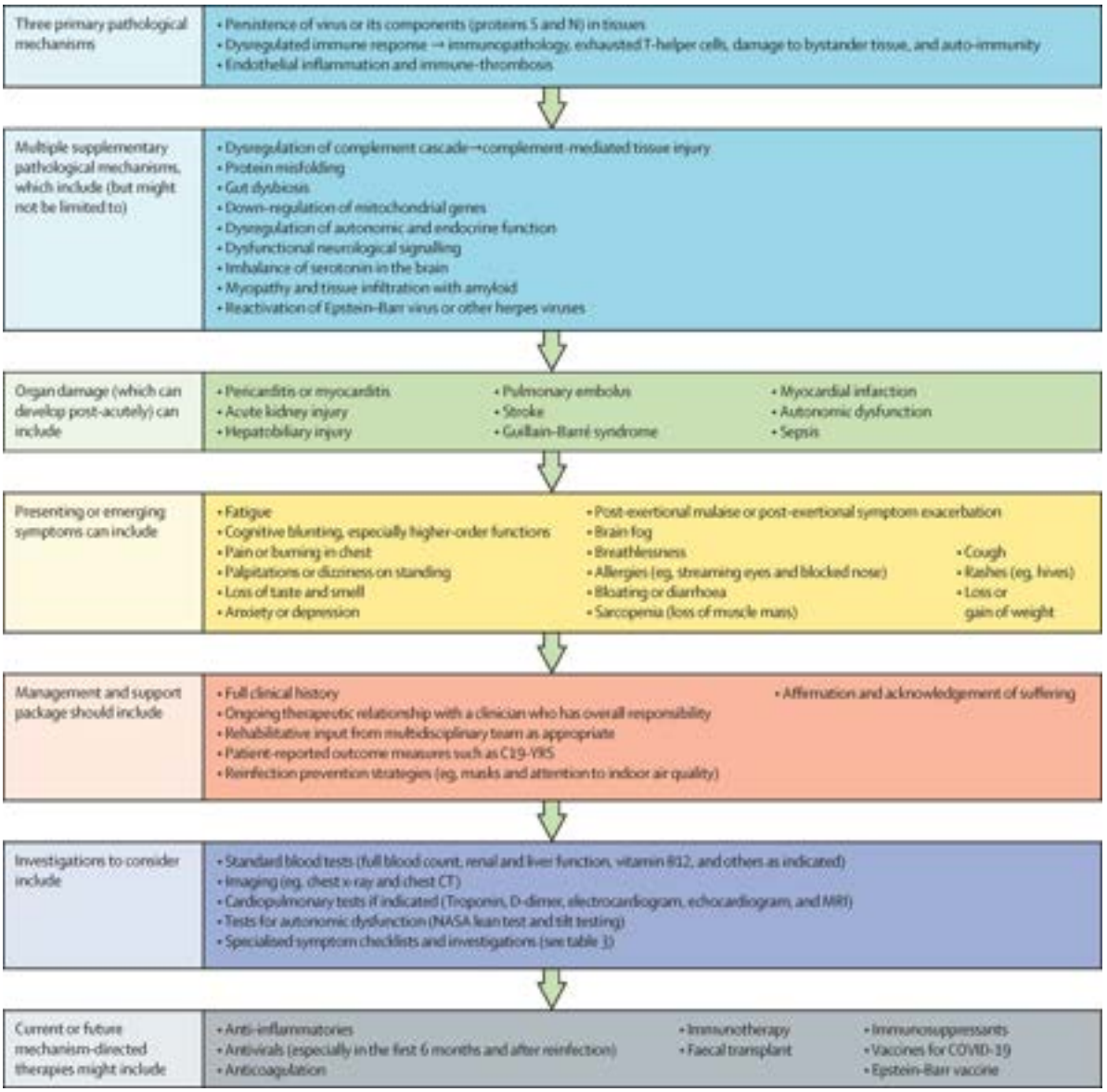
Prise en charge - De nombreuses personnes touchées par le CL sont confrontées à des difficultés pour faire reconnaître leur maladie par les professionnels de santé qui les prennent en charge, et sont parfois victimes de scepticisme sur la réalité de leurs symptômes qui sont souvent attribués à tort à des problèmes psychosomatiques préexistants. Ceci induit des difficultés à trouver une prise en charge adaptée, une errance médicale et une certaine inquiétude quant aux options thérapeutiques possibles et à l'avenir[83].

Différents modèles de prise en charge ont été développés en fonction des pays et impliquent généralement en première intention les médecins traitants, puis dans un second temps des médecins spécialistes et parfois des centres spécialisés dans la prise en charge de patients avec un CL[84]. La prise en charge actuelle du CL a pour but de soulager les différents symptômes et intègre ainsi de multiples spécialités médicales (médecins généralistes, infectiologues, neurologues, psychologues, orthophonistes etc) et différents types de rééducations (cardiorespiratoire, olfactive et gustative, à l'effort etc.).

Pacing - Le pacing est une stratégie de gestion des efforts fréquemment utilisée pour limiter les risques de malaises post-efforts. Cette approche est parfois compliquée à mettre en place mais elle a montré un bénéfice avec une réduction du nombre d'exacerbations de symptômes post-effort[85]. Finalement, ces prises en charge multiples induisent une charge importante en termes de temps et d'énergie pour les patients.

La **Figure 3** résume les mécanismes physiopathologiques, présentations cliniques, diagnostics et méthodes de prise en charge.

Figure 3 : Résumé des différents mécanismes physiopathologiques, présentations cliniques, diagnostics et méthodes de prise en charge



T. Greenhalgh, M. Sivan, A. Perlowski, J. Nikolich. Long COVID: A clinical update. The Lancet. Online (31 Juillet 2024). https://doi.org/10.1016/S0140-6736(24)01136-X

Traitements - Il n'existe aujourd'hui pas de traitement curatif du CL, la stratégie de traitement se base sur le traitement des différents symptômes, par des stratégies médicamenteuses ou non.

Au 7 Août 2024, la recherche du terme "Long COVID" sur le site Clinicaltrials.gov liste 318 essais cliniques interventionnels, dont 80 pour des traitements pharmacologiques et les autres regroupant diverses interventions non médicamenteuses comme des programmes d'entraînement, de rééducation, ou des dispositifs médicaux.

Le **Tableau 1** résume les principaux traitements médicamenteux en cours d'étude.

Tableau 1 : Principaux traitements médicamenteux en cours d'étude (en date du 7 Août 2024)

Catégorie médicamenteuse	Traitements	Mécanisme visé	Symptômes visés
Antiviraux	Nirmatrelvir/ritonavir, Remdesivir, Amantadine, Larazotide, Ensitrelvir, Tenofovir disoproxil/Emtricitabine, Maraviroc, Valaciclovir, Rintatolimod	Persistance virale	Tous
Antiinflammatoires	Ibudilast, Tonabersat, Naltrexone, Methylprednisolone, Celecoxib, Baricitinib, Montelukast, Deupirfenidone	Inflammation, persistance virale	Tous
Antihistaminiques	Cétirizine, lévocétirizine, famotidine	Inflammation	Tous, respiratoires, fatigue, gastrointestinaux
Anticancéreux	Imatinib	Dérégulation immunitaire, persistance virale	Pulmonaires
Anticorps monoclonaux	Infliximab , Indevimab, Casirivimab	Inflammation, persistance virale	Tous
	Ipilimumab, Nivolumab	Dérégulation immunitaire, inflammation	
Antidépresseurs	Vortioxetine, Fluvoxamine, Cyclobenzaprine, prégabaline	Inflammation, dérégulation immunitaire, stimulation cellules souches, potentielle sur persistance virale	Troubles anxieux et dépressifs, fatigue, neurocognitifs
Traitements cardiovasculaires	Pentoxifylline, Ivabradine, Vericiguat, Métoprolol, Pravastatine, S-1226	Inflammation endothéliale	Cardiovasculaires, respiratoires, hypotension orthostatique
Antidiabétiques	Metformine, liraglutide	Dysfonction métabolique	Fatigue, neurocognitifs
Modulateurs immunitaires	Immunoglobulines intraveineuses, Anakinra, Efgartigimod	Dérégulation immunitaire, stimulation cellules souches	Fatigue, cardiovasculaires, Hypotension orthostatique, tous
Traitement pour le sommeil	Modafinil, Solriamfetol, Mélatonine		Sommeil
Antipsychotique	Pimozide		Acouphènes
Hormones, cellules souches	Liquide amniotique humain purifié, Cellules souches mésenchymateuses du cordon ombilical, Somatropine	Stimulation cellules souches, dysfonction métabolique	Tous
Oxygénothérapie hyperbare	N/A	Inflammation, dérégulation immunitaire, stimulation cellules souches	Neurocognitifs et cardiovasculaires, tous

Résultats des essais - Seuls quelques essais enregistrés ont publié leurs résultats. Tout d'abord, un essai randomisé de phase III a montré que la metformine, un antidiabétique, permettrait une diminution de 41% du risque de développer un CL si administré précocément après l'infection[86]. Dans la même étude, l'ivermectine et la fluvoxamine n'ont montré aucun bénéfice par rapport au placébo. L'administration d'antiviraux comme le nirmatrelvir boosté au ritonavir précocement après une infection SARS-Cov 2 semble diminuer le risque de développer un CL ainsi que sa sévérité[87,88].

Ensuite, une étude sur la famotidine, un antihistaminique qui pourrait réduire l'inflammation et l'activation mastocytaire, a montré un bénéfice pour le traitement de l'anxiété et des troubles cognitifs[89]. Un essai pilote d'une combinaison maraviroc (antiviral) et pravastatine (anticholestérolémiant) sur 18 personnes touchées par le CL a montré des résultats prometteurs avec une amélioration clinique perçue par les personnes traitées[90].

Les anticorps monoclonaux dirigés contre la protéine spike sont également une piste de recherche, avec la publication de 3 cas de rémissions de CL sévère après administration de casirivimab et d'imdevimab au moment d'une réinfection[91]. L'oxygénothérapie hyperbare a montré une amélioration de la qualité de vie, de sommeil et de symptômes cognitifs, avec un effet persistant 1 an après[92].

Enfin, les anticoagulants sont également une piste de traitement étudiée, avec notamment une étude sur un traitement combinant 3 anticoagulants et qui a montré une amélioration des symptômes et une amélioration des paramètres sanguins chez 91 patients en Afrique du Sud[93].

Impact sur la qualité de vie

Plusieurs études ont montré un impact important du CL sur la qualité de vie globale des personnes touchées, en utilisant des questionnaires d'évaluation de la qualité de vie globale comme les questionnaires SF12 ou EQ5D-5L[94,95]. D'autres études se sont intéressées à différents volets de la qualité de vie comme le sommeil, le niveau d'anxiété ou le niveau de fatigue et ont montré une altération de ces différents paramètres en raison de symptômes du CL[96,97]. Les femmes rapportent plus fréquemment des problèmes d'anxiété et de dépression, une qualité de vie liée à la santé physique altérée et sont moins nombreuses que les hommes à avoir repris le travail[22].

Enfin, du point de vue des personnes touchées par le CL, le fardeau de la maladie sur leur vie quotidienne est induit, outre l'intensité et la variabilité des symptômes, par la difficulté d'accéder à une prise en charge adaptée, la multiplicité des rendez-vous médicaux et pour la rééducation, voire la difficulté à faire reconnaître leurs symptômes par certains professionnels de santé qui méconnaissent la maladie. Toutes ces difficultés entraînent un sentiment d'abandon et impactent négativement la santé mentale, avec d'importants niveaux de stress, d'anxiété et de peur de l'avenir[98].

1.2 Les solutions digitales de santé

Définition des solutions digitales de santé

Selon l'OMS, les solutions digitales de santé réfèrent à l'utilisation des nouvelles technologies de l'information et de la communication pour des applications dans le domaine de la santé, facilitent l'échange et le stockage de données, et peuvent contribuer à l'amélioration des diagnostics médicaux, des décisions de traitement fondées sur des données, des thérapies digitales, des essais cliniques ou de l'autogestion des soins[99]. Les solutions digitales en santé peuvent donc être utilisées tout au long du cycle de la médecine, de la recherche de nouveaux traitements jusqu'à leur utilisation en vie réelle, et permet d'augmenter l'implication des patients dans leur traitement ou parcours de soins (cf **Figure 4**).

Figure 4 : Utilisation des solutions digitales en santé dans le cycle de vie de la médecine (https://www.efpia.eu/about-medicines/development-of-medicines/digital-health/)

Elles peuvent se présenter sous forme d'applications mobiles de suivi et de gestion des symptômes, de dispositifs portables (*wearables*) pour surveiller les signes vitaux et l'activité physique ou de plateformes en ligne pour l'éducation des patients, la télésurveillance et la communication avec les professionnels de santé par exemple. La **Figure 5** résume les différentes catégories de solutions digitales de santé.

Figure 5 : Différentes catégories de solutions digitales de santé

Les icônes de cette figure ont été conçues avec Freepik (www.freepik.com)

Dans le contexte particulier des maladies chroniques, les solutions digitales de santé participent au développement d'une médecine centrée sur le patient qui devient acteur à part entière dans la gestion de sa maladie, le but étant d'améliorer le suivi et de réduire le fardeau de la maladie.

Voici quelques exemples de solutions digitales pour des maladies chroniques. Ces maladies impliquent, tout comme le CL, un suivi régulier entre les visites médicales et ont un impact important sur la vie quotidienne des personnes concernées.

Dans le cadre du diabète, et en particulier le **diabète de type 1**, une maladie où les personnes touchées sont très investies dans leur suivi, il existe de nombreuses applications, comme "Diabetes Tracker" ou "mySugr", avec des fonctionnalités multiples, comme le suivi de la glycémie, des conseils nutritionnels, ou de suivi de l'activité physique, ainsi que des systèmes de surveillance continue de la glycémie comme FreeStyle Libre.

Pour les **maladies cardio-vasculaires et respiratoires**, il existe de nombreuses applications permettant par exemple un suivi de la pression artérielle, comme l'application "suivi HTA", des moniteurs de fréquence cardiaque comme "Heart Rate Monitor", ainsi que des dispositifs de surveillance à domicile pouvant faire suite à une hospitalisation. Pour les **maladies respiratoires chronique**s, il existe des systèmes de capteurs connectés pour surveiller la fonction pulmonaire, comme "Respiratory Rate Monitor", ainsi que des applications d'auto-gestion de l'**asthme,** comme "FindAir".

Intérêt des solutions digitales par rapport aux méthodes traditionnelles

Le développement et l'utilisation des solutions digitales de santé contribuent à la numérisation du système de santé et de la recherche clinique[100,101]. Cette numérisation présente plusieurs avantages par rapport aux approches traditionnelles basées sur des visites médicales en personne, sur site, avec une utilisation majoritaire de documents "papier" et des évaluations cliniques effectuées par des professionnels de santé. Parmi les avantages, on peut citer une diminution des temps de visites et des évaluations entre les visites sur site, une facilitation du recrutement aux études et un accès facilité aux consultations pour des régions géographiques isolées.

De plus, en unifiant et en normalisant les solutions de télémédecine avec des téléconsultations augmentées comprenant une évaluation en temps réel des principaux paramètres physiologiques, le temps de consultation pourrait être optimisé et un suivi personnalisé pourrait être proposé grâce à une communication accrue entre les professionnels de la santé et les patients.

Solutions digitales et COVID-19 et COVID Long

La pandémie de COVID-19 a accéléré le développement de solutions digitales de santé. En effet, les consultations par système de visioconférence ou par téléphone ont parfois remplacé, pendant les périodes de confinement, les visites traditionnelles chez le médecin et ont montré l'utilité de ces formats pour suivre les patients qui ne peuvent pas se déplacer ou qui se trouvent dans des régions géographiques dépourvues de médecins[102].

De plus, des centaines d'applications mobiles de recherche de contacts et des technologies radiologiques basées sur l'intelligence artificielle pour faciliter la détection précoce du COVID-19 sont apparues dès le début de la pandémie[103]. Parallèlement, plusieurs technologies digitales ont été mises au point pour répondre aux différents besoins des patients, qu'il s'agisse du diagnostic, de la prévention, du traitement, de l'observance, du mode de vie ou de l'engagement des patients[104].

Par exemple, une application reliée à une montre connectée a été développée en Allemagne pour faciliter le diagnostic du COVID-19 à partir de quelques signes vitaux[105].

Dans le contexte du CL, étant donné l'absence de traitement spécifique et les difficultés relatives à la prise en charge des personnes touchées, un groupe d'expert du National Institute for Health Excellence (NICE) a émis des recommandations en faveur du développement de systèmes de télésurveillance et d'autogestion des symptômes de l'infection aigüe et du CL. Ces systèmes devant être accessibles au plus grand nombre et leur utilisation à domicile est encouragée[106,107].

Deux systèmes de télésurveillance, Telecare-COVID et CareSimple-COVID ont été développés au Canada et étaient bien acceptés par les utilisateurs[105,108].

Il existe aussi quelques applications dédiées au CL mais elles ne sont disponibles que dans certaines régions ou certaines langues (voir **Tableau 2**). Par exemple, l'application "Visible"[109], qui permet un suivi des symptômes et de l'activité, n'est disponible qu'aux Etats-Unis et en anglais. "Living with COVID Recovery"[110] est une application très complète qui propose le suivi de symptômes par des questionnaires standardisés, des programmes de rééducation, des ressources informatives et un lien avec des professionnels de santé mais n'est disponible qu'au Royaume-Uni, en anglais et sur invitation.

Il existe donc un manque pour une application accessible à tous, en différentes langues pour offrir un support au quotidien à un maximum de personnes concernées.

Tableau 2 : Applications dédiées au CL existantes

Nom	Fonctionnalités principales	Pays création	Disponibilité	Langues
Visible	● Suivi de symptômes ● Suivi de l'activité (avec bracelet connecté)	Etats-Unis	Etats-Unis	Anglais
Living with COVID Recovery	● Suivi de symptômes ● Questionnaires standardisés ● Exercices de rééducation ● Messagerie avec professionnels de santé	Royaume-Uni	Royaume-Uni et sur invitation uniquement	Anglais
Responsum for Long COVID	● Informations sur le CL ● Chat entre personnes avec CL ● Informations médicales ● Messagerie avec professionnels de santé	Etats-Unis	Monde	Anglais
Long COVID Tagebuch	● Journal médical ● Journal vie quotidienne ● Informations sur le CL	Suisse	Monde	Allemand, Anglais, Français

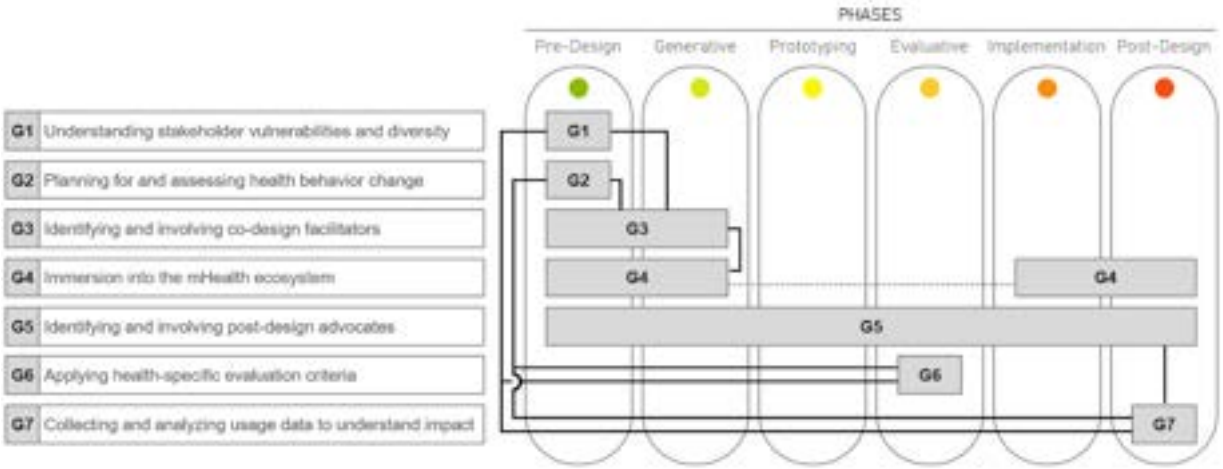
Importance du co-design

Il est crucial d'impliquer les futurs utilisateurs dans le développement de la solution digitale qui doit idéalement être réalisé selon un processus de co-construction afin de s'assurer de répondre aux besoins et aux attentes, de garantir sa pertinence clinique mais aussi son adoption future par les utilisateurs[111–113]. En effet, l'utilisation des solutions digitales de santé est soumise au phénomène d'attrition, tout comme les essais cliniques[114], et impliquer les futurs utilisateurs dans le développement contribue à diminuer ce phénomène et à garantir une utilisation sur le long terme.

Il existe de nombreuses interprétations et mises en œuvre pratiques de recherches participatives pour le développement de solutions digitales de santé, avec des degrés variables d'engagement des patients, allant du "simple" testeur, à un rôle de co-chercheur ou co-concepteur[115].

La **Figure 6** montre un exemple de co-design pour le développement d'une solution digitale de santé impliquant les patients ou utilisateurs finaux tout au long du processus, depuis la phase de pré-conception jusqu'à la phase de post-conception qui consiste à utiliser les données des utilisateurs pour évaluer l'utilisabilité et l'impact de la solution ainsi développée.

Figure 6 : Exemple de processus de co-design pour le développement d'une solution de santé digitale



Noorbergen TJ, Adam MTP, Teubner T, Collins CE. Using Co-design in Mobile Health System Development: A Qualitative Study With Experts in Co-design and Mobile Health System Development. JMIR Mhealth Uhealth 2021;9(11):e27896[115]

1.3 Les biomarqueurs digitaux

Définition - La collecte de données digitales grâce aux solutions digitales de santé permet de dériver des critères d'évaluation (*end-point*) digitaux quantifiables grâce à des biomarqueurs digitaux. Ces derniers sont définis par l'*European Medicines Agency* (EMA) comme des "*mesures objectives et quantifiables d'un paramètre physiologique et/ou du comportement, utilisées comme indicateurs d'un processus biologique ou pathologique ou d'une réponse à une exposition ou à une intervention, et dérivées d'une mesure digitale. Sa signification clinique est établie par une relation fiable avec un critère d'évaluation existant et validé"*[116]. Ils permettent donc une mesure indirecte de la santé grâce aux données digitales collectées par les dispositifs portables, les applications de santé, voire même les réseaux sociaux.

Tout comme pour les solutions digitales de santé, il est crucial de développer des mesures digitales qui soient pertinentes pour les patients. Pour cela il est important de les impliquer pour comprendre quelles sont les problématiques liées à leur pathologie qu'ils souhaiteraient améliorer, quels sont les symptômes les plus impactants ou de combien devrait être amélioré un paramètre pour que cela ait un réel intérêt pour eux[117].

L'équipe de recherche Deep Digital Phenotyping du Luxembourg Institute of Health s'intéresse à un biomarqueur digital particulier qui s'appelle biomarqueur vocal (BV).

1.4 Les biomarqueurs vocaux

Intérêt de la voix pour le suivi de la santé - La voix est un moyen de communication entre individus qui permet d'interagir avec efficacité et impact. Elle joue un rôle majeur dans nos interactions sociales en transmettant nos émotions et nos sentiments, en modulant différents paramètres que sont la hauteur, la tonalité, ou l'intensité. La voix est en elle-même un moyen riche en informations sur l'état de santé et les émotions, permettant une caractérisation plus riche des patients grâce à l'utilisation de ce que l'on appelle les biomarqueurs vocaux.

De plus, la voix est peu contraignante à collecter pour les patients ou les participants à une étude, nécessite moins de temps que le remplissage d'un questionnaire, n'est pas invasive.

Ceci ouvre de multiples perspectives basées sur l'utilisation de la voix en plus de son utilisation comme simple outil de collecte d'informations de santé, et en particulier pour le suivi à distance de patients ou de participants à un essai clinique.

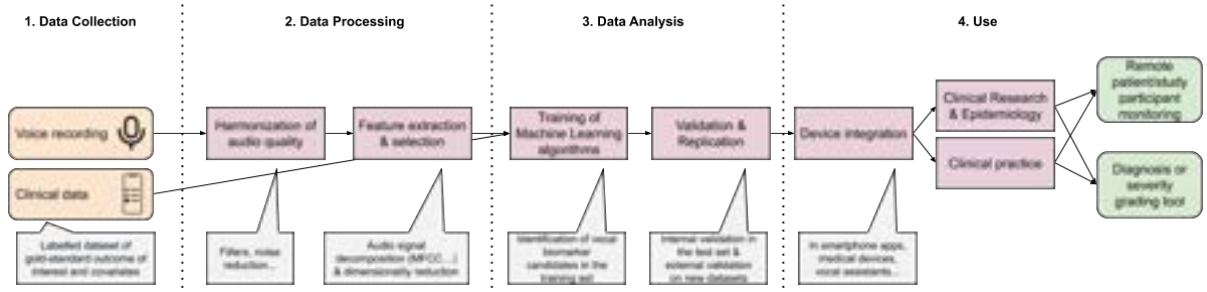
Définition d'un biomarqueur vocal - La voix dérive du signal vocal émis par les cordes vocales. Ce signal vocal peut être décomposé en des milliers de caractéristiques pouvant être affectées par une modification de notre état de santé. Ces caractéristiques peuvent être perçues par l'oreille humaine (comme l'intensité et l'amplitude par exemple), mais la plupart ne le sont pas. Un BV peut être défini comme une signature, une caractéristique ou une combinaison de caractéristiques du signal audio de la voix, associées à une maladie ou un symptôme. Ainsi, un BV est une extension d'un biomarqueur classique, un facteur objectivement mesuré et évalué représentant un processus biologique ou pathogène ou une réponse pharmacologique à une intervention thérapeutique. Il doit donc posséder toutes les propriétés d'un biomarqueur traditionnel et doit être validé à l'aide d'évaluations cliniques rigoureuses. Un BV peut être utilisé pour suivre les patients, diagnostiquer une maladie ou évaluer la gravité d'une maladie[118].

Domaines d'application - Les premiers travaux sur l'identification de biomarqueurs vocaux datent de 2010 et se sont intéressés aux maladies dégénératives et en particulier à la maladie de Parkinson[119], dans laquelle la prévalence de problèmes de parole a été estimée à 89%[120,121]. Depuis quelques années, la recherche sur les biomarqueurs vocaux s'est étendue à d'autres maladies comme l'hypertension, les maladies mentales, le diabète ou l'insuffisance cardiaque[118].

Seuls quelques-uns d'entre eux ont à l'heure actuelle été intégrés dans des applications pour smartphones, actuellement disponibles sur les magasins d'applications, comme l'application Real Time Voice Analyser pour surveiller le bien-être respiratoire[122], ou l'application Sonde Mental Health[123], basée sur un BV de la condition mentale. Cependant très peu d'informations sont disponibles sur la méthodologie de développement et de validation de ces applications.

Développement d'un BV - Dans l'équipe de recherche nous avons décrit le processus de développement d'un BV, constitué de 4 phases successives décrites dans la **Figure 7**[118].

Figure 7 : Processus de développement d'un biomarqueur vocal



Fagherazzi et al. Voice For Health: The Use Of Vocal Biomarkers From Research To Clinical Practice. Digital Biomarkers 2021

La première étape est une étape de **collecte de données**, pendant laquelle des enregistrements de voix sont collectés au même moment que des données cliniques afin de les labelliser. Typiquement cela permet d'obtenir des jeux de données de personnes présentant un symptôme d'intérêt et de personnes ne présentant pas ce même symptôme. Différents types d'enregistrements peuvent être utilisés : soit des enregistrements dits standardisés pour lesquels les personnes effectuent des tâches comme lire un texte prédéfini, dire une voyelle le plus longtemps possible, compter de 1 à 20, ou tousser par exemple, soit des enregistrements semi-guidés comme la description d'une image, soit des enregistrements libres pour lesquels les personnes sont invitées à parler du sujet de leur choix pendant un certain temps.

Enfin, les enregistrements peuvent être réalisés dans différentes conditions, dans un studio d'enregistrement avec un micro et des conditions environnantes très contrôlées et identiques pour tous les participants, ou dans des conditions de vie réelle en utilisant des applications web, le téléphone ou un smartphone, avec pour conséquence des différences importantes en terme de qualité des enregistrements.

Des précautions particulières doivent être prises lors de cette phase de collecte de voix car la voix est considérée comme une donnée identifiante et sensible et sa collecte est soumise à des réglementations spécifiques en fonction des régions géographiques, comme le Règlement Général sur la Protection des Données (RGPD)[124] en Europe ou la loi sur la protection des renseignements personnels et les documents électroniques (Personal Information Protection and Electronic Documents Act, PIPEDA)[125] au Canada. Aux États-Unis, il n'existe pas de loi unique sur la protection des données, mais de multiples lois promulguées au niveau fédéral ou au niveau de l'État. Ces différentes lois et réglementations ne confèrent pas le même niveau de protection aux individus et il est donc fortement recommandé d'appliquer le plus haut niveau de protection et de demander un consentement explicite avant la collecte d'enregistrements vocaux.

La deuxième étape est une étape de **pré-traitement des enregistrements audios**, de manière à harmoniser la qualité entre les enregistrements, en éliminant le bruit de fond ou en les calibrant à la même durée. Puis des méthodes de traitement du signal sont appliquées pour extraire des caractéristiques du signal audio. Ainsi plusieurs milliers de caractéristiques acoustiques peuvent être extraites. Ensuite, un ensemble de caractéristiques d'intérêt sera sélectionné en fonction du symptôme ou de la maladie d'intérêt.

Pendant l'étape d'**analyse des données**, les caractéristiques sélectionnées sont ensuite utilisées pour entraîner des algorithmes d'apprentissage automatique ou profond à prédire ou classifier un enregistrement en fonction du paramètre clinique étudié. Les candidats biomarqueurs vocaux ainsi identifiés doivent ensuite être validés sur des jeux de données externes avant de pouvoir être intégrés dans un dispositif comme une application smartphone, un dispositif médical, ou une montre connectée par exemple.

La dernière étape est l'**utilisation** concrète du dispositif embarquant le BV dans la pratique clinique comme outil de dépistage ou d'évaluation de la sévérité, ou en recherche clinique ou en épidémiologie pour effectuer du suivi à distance des participants à une étude.

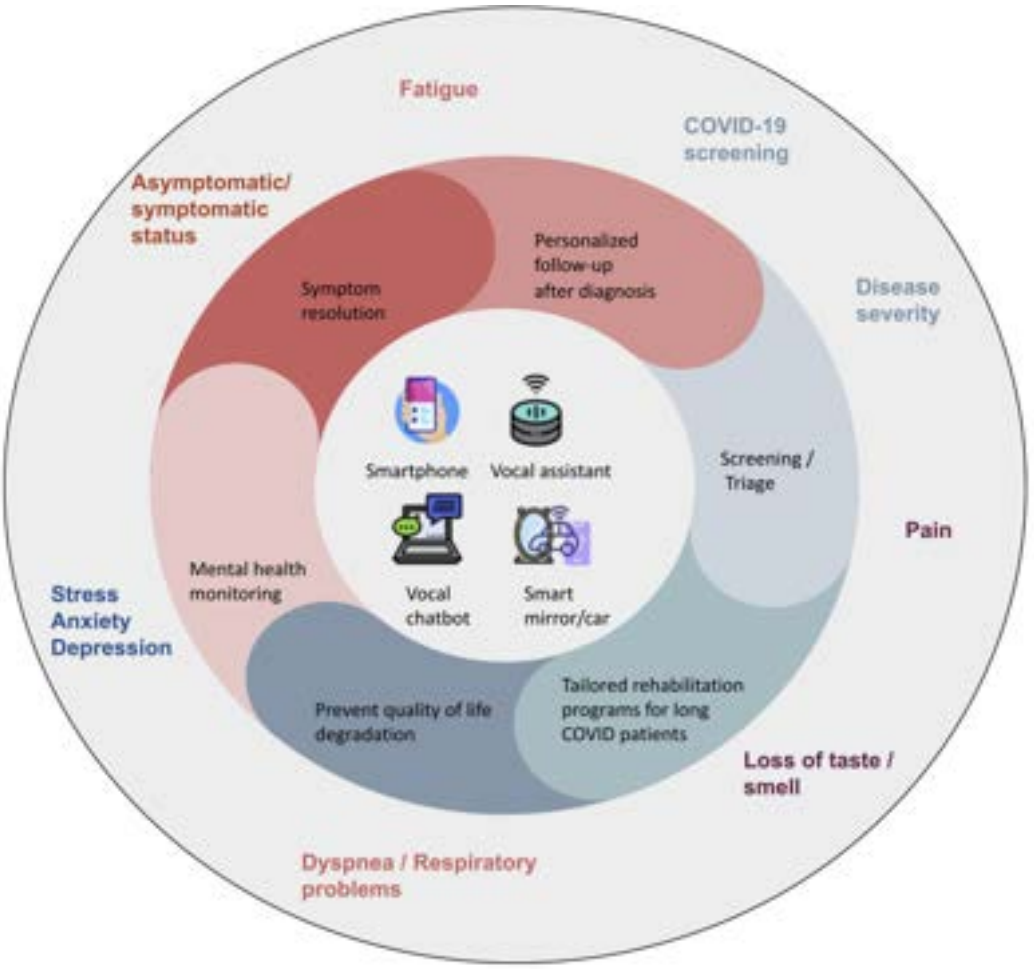
1.5 Utilisation des biomarqueurs vocaux dans le contexte du COVID Long

En parallèle des travaux principaux de cette thèse, j'ai contribué à d'autres travaux dans l'équipe de recherche visant à identifier des biomarqueurs vocaux pour différents symptômes liés à COVID-19 comme la fatigue[126], la perte de goût et odorat[126,127] ou pour différencier le statut symptomatique ou asymptomatique d'une personne[128]. Nous avons ainsi identifié des candidats biomarqueurs vocaux prometteurs, mais il subsistait une problématique majeure qui était de définir comment les amener à une utilisation concrète en clinique ou dans la vie de tous les jours.

J'ai tenté de répondre à cette question dans un article de type "perspective" dans lequel j'ai analysé et présenté les différents cas d'usage des biomarqueurs vocaux et proposé des recommandations pour une implémentation pratique réussie des biomarqueurs vocaux dans le contexte de COVID-19 et du CL[129] (article complet en Annexe 1).

Cas d'usage - Tout d'abord, j'ai proposé différents cas d'usage des biomarqueurs vocaux dans le contexte de COVID-19 et du CL, présentés dans la **Figure 8**. Ils pourraient être implémentés dans des dispositifs tels que des applications smartphone, des chatbots, ou des assistants vocaux, afin de surveiller la résolution des symptômes (avec un BV de l'état symptomatique ou asymptomatique) ou la dégradation de la santé mentale et de la qualité de vie (avec un BV du stress, de l'anxiété, de la dépression, de la fatigue chronique ou de la dyspnée). Ils pourraient aussi être utilisés pour proposer un suivi personnalisé après le diagnostic (avec un BV de gravité de la maladie), pour effectuer un dépistage et un triage des patients à l'hôpital (avec un BV de sévérité de la maladie, de douleur ou de perte de goût et d'odorat), et pour proposer des programmes de réadaptation sur mesure pour les patients atteints de CL (en utilisant des biomarqueurs vocaux de perte de goût et d'odorat, de douleur, de dyspnée ou de fatigue).

Figure 8 : Différents cas d'usage des biomarqueurs vocaux pour le suivi de symptômes du COVID et du CL



Fischer A, Elbeji A, Aguayo G, Fagherazzi G. Recommendations for Successful Implementation of the Use of Vocal Biomarkers for Remote Monitoring of COVID-19 and Long COVID in Clinical Practice and Research. Interact J Med Res 2022;11(2):e40655

Développement d'une solution digitale - J'ai ensuite proposé un processus pour le développement d'une solution digitale de santé basée sur les biomarqueurs vocaux dans le cadre de COVID-19 et du CL (voir **Figure 9**).

Figure 9 : Processus de développement d'une solution digitale de santé basée sur les biomarqueurs vocaux (BV).

La première étape (**phase d'identification**) est de bien identifier les symptômes et/ou problèmes de santé à suivre et de collecter les enregistrements vocaux. Cette étape importante comporte des enjeux spécifiques dans le cas de COVID-19 et du CL, d'abord au niveau de l'identification des symptômes au vu du caractère multiple du CL, et ensuite au niveau de la collecte des enregistrements de voix car le port du masque ou les symptômes eux-mêmes peuvent rendre difficile la réalisation des enregistrements. Par contre, le port du masque n'a semble-t-il pas d'incidence sur les paramètres vocaux. Ainsi, les enregistrements peuvent être réalisés sans avoir à diminuer les protections individuelles nécessaires[130].

Une fois la collecte de données et l'identification des biomarqueurs vocaux d'intérêt réalisés, ils doivent être soumis à une **phase de validation**. Cette phase de validation externe doit idéalement se faire en comparaison avec le Gold Standard d'évaluation du symptôme en question qui peut être une échelle validée pour le cas de la fatigue ou du stress, ou des paramètres biologiques comme la détection de la présence de virus dans le sang ou d'autres cellules du corps.

Cette validation des biomarqueurs vocaux doit être similaire à tout autre biomarqueur biologique afin d'évaluer sa sensibilité, spécificité et reproductibilité[131].

L'étape suivante est la **phase d'intégration** dans une solution digitale de santé. Cette phase d'intégration englobe le développement de la solution digitale, avec ses futurs utilisateurs dans une démarche de co-construction, l'élaboration d'un prototype et des tests fonctionnels du prototype. Cela est particulièrement important pour les technologies intégrant la voix, en raison de leur caractère innovant, afin de garantir leur acceptabilité future. Enfin, des boucles de rétroaction devraient être mises en œuvre pour améliorer à la fois la solution digitale et l'algorithme grâce aux enseignements tirés des études auprès des populations cibles.

Enfin, la dernière étape (**phase d'évaluation clinique de la solution digitale de santé**) consiste à évaluer l'intérêt clinique de la solution digitale par des essais cliniques (allant d'études cliniques contrôlées en double-aveugle jusqu'à des études en vie réelle), son utilisabilité et son acceptabilité grâce à des études qualitatives ou mixtes. Puis, en fonction du type de solution développée, sa qualification en tant que dispositif médical pourra être demandée. La définition de l'utilisation future de la solution de santé doit être faite le plus tôt possible à ce stade afin de définir les évaluations cliniques nécessaires pour obtenir le marquage CE ou la certification par la Food and Drug Administration (FDA). Une fois la certification obtenue, il est alors possible de demander le remboursement de la solution digitale de santé par les organismes nationaux d'assurance santé après avoir démontré son intérêt clinique et économique. Enfin, dépendant du type de solution digitale de santé, il peut être nécessaire de l'intégrer dans les systèmes informatiques des hôpitaux ou des professionnels de santé, ce qui présente des défis spécifiques à prendre en compte également le plus tôt possible dans le processus de développement.

2. Objectifs du projet de thèse

Ce travail de thèse a pour but d'apporter une solution concrète aux personnes touchées par le CL, sous la forme d'une solution digitale de santé intégrant la voix. Il se base sur une **recherche translationnelle**, démarrant d'une problématique touchant la population cible qui est celle touchée par le CL, et qui est impliquée dans la totalité du projet à la fois comme participants à la recherche mais aussi comme partenaires à la recherche.

Le projet de thèse s'articule en 2 axes principaux, divisés en 5 chapitres. Le schéma général du déroulement de ce projet de thèse est représenté par la **Figure 10**.

NB: Cette thèse est basée sur mes publications. Selon les recommandations de l'Université de Lorraine, la langue principale du manuscrit est le Français, avec les articles insérés dans leur version finale, publiée ou soumise, en Anglais.

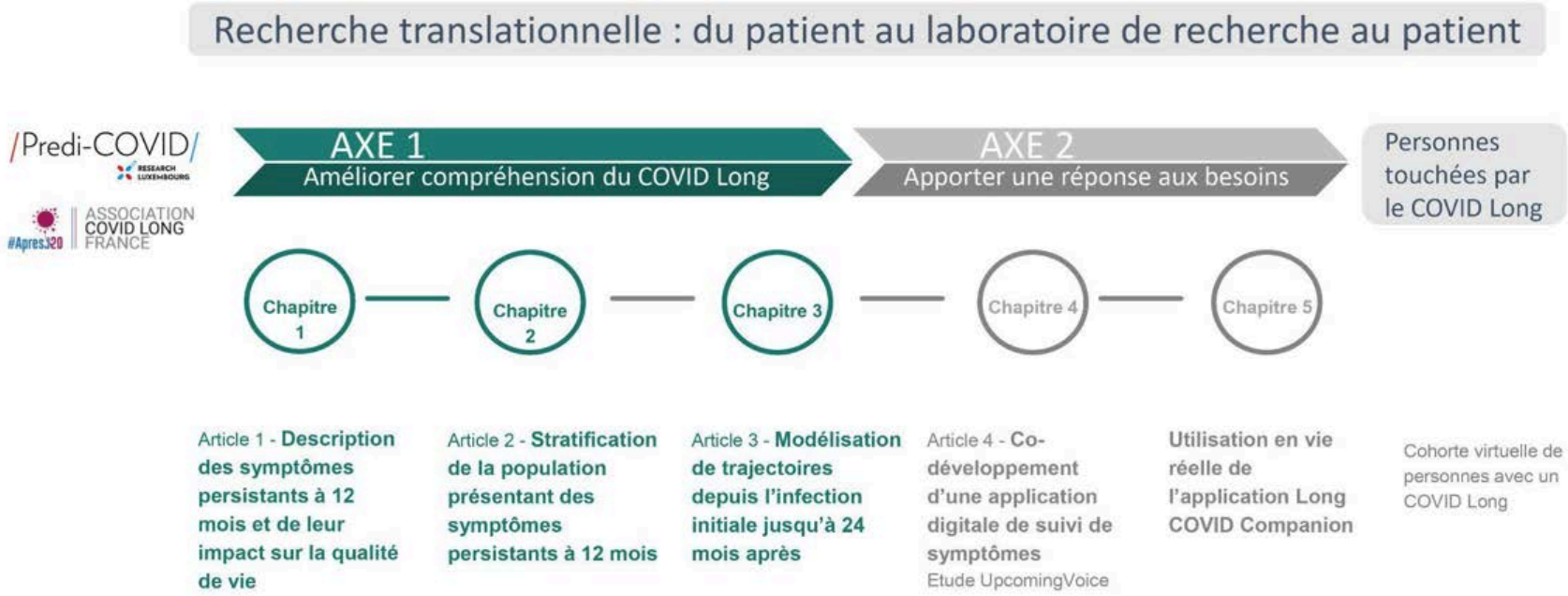


Figure 10 : Schéma général de la thèse

Axe 1 - Améliorer la compréhension du COVID Long - Chapitres 1 à 3

L'Axe 1 correspond à la phase d'**identification** du besoin (cf. processus présenté plus haut, **Figure 9**) et a pour but d'améliorer la compréhension de la problématique du CL sur base des données de la cohorte Predi-COVID au Luxembourg. L'identification de candidats BV a été faite antérieurement et leur validation se poursuit en parallèle à ces travaux de thèse.

Le **chapitre 1** se focalise sur la description des symptômes persistants et leur impact sur la qualité de vie des personnes touchées 12 mois après l'infection initiale. Le **chapitre 2** a pour but de stratifier la fraction de la population présentant encore des symptômes à 12 mois, et le **chapitre 3** vise à modéliser des trajectoires d'évolution de symptômes depuis l'infection initiale jusqu'à 24 mois après.

Axe 2 - Apporter une réponse aux besoins - Chapitres 4 et 5

L'Axe 2 correspond à la phase d'**implémentation** de la solution digitale et a pour but de proposer une solution concrète aux besoins identifiés dans l'Axe 1.

Le **chapitre 4** vise à co-construire une application de suivi de symptômes liés au CL en vie réelle, basée sur la voix, avec ses futurs utilisateurs. Nous avons tout d'abord défini les besoins et les attentes des personnes touchées par le CL en termes d'application de suivi de symptômes et d'utilisation de la voix. Puis, nous avons défini les spécifications d'une telle application. Le **chapitre 5** est axé sur le développement technique et la mise à disposition de l'application Long COVID Companion en vie réelle.

Enfin, la **validation clinique** de la solution digitale développée dans l'Axe 2 fera l'objet des perspectives de ces travaux de thèse.

3. Matériels et Méthodes

3.1 Matériels - La cohorte Predi-COVID

Description de la cohorte

L'étude Predi-COVID[132] est une cohorte prospective, longitudinale de personnes testées positives au SARS-COV 2 au Luxembourg. La cohorte a été lancée dès le début de la pandémie avec un premier participant inclus le 5 mai 2020. Les objectifs de Predi-COVID étaient d'identifier les caractéristiques épidémiologiques, cliniques, digitales et sociodémographiques, les biomarqueurs prédictifs de la sévérité ainsi que les conséquences à long terme d'une infection COVID-19. En Février 2022, l'étude Predi-COVID a été intégrée à l'initiative nationale luxembourgeoise CoVaLux (COVID-19, Vaccination & conséquences sanitaires à long terme du COVID-19), visant à évaluer en particulier l'impact de la vaccination et le CL[133].

Le protocole de l'étude a été publié le 23 novembre 2020 dans la revue BMJ Open[134]. L'étude a obtenu l'avis favorable du Comité National d'Ethique de la Recherche (CNER) du Luxembourg (ref 202003/07) et l'autorisation du Ministère de la Santé du Luxembourg en tant qu'autorité compétente en Avril 2020. L'étude est également enregistrée dans ClinicalTrials.gov (NCT04380987).

En bref, toute personne testée positive au SARS-COV 2 par Réaction de Polymérisation en Chaîne (PCR) au Luxembourg, hospitalisée ou non, était appelée par téléphone par les services de l'Inspection Sanitaire qui lui demandait son accord pour transmettre ses coordonnées au Luxembourg Institute of Health (LIH) à des fins de recherche. Les personnes ayant accepté étaient alors contactées par des infirmiers de recherche spécialisés du Centre d'Investigation et d'Epidémiologie Clinique du LIH pour leur expliquer l'étude. Toute personne acceptant de participer à l'étude signait un consentement éclairé électronique.

Collecte de données

La collecte de données était effectuée de 3 manières différentes :

- Des données étaient collectées à l'inclusion par les infirmiers de recherche, par téléphone pour les personnes à domicile et lors d'une visite d'inclusion pour les personnes hospitalisées.
- Des questionnaires électroniques sur les symptômes et l'état de santé général, complétés en ligne par les participants, tous les jours pendant les 14 jours après l'inclusion, puis à la semaine 3 et 4, et enfin à 12, 15 et 24 mois après l'inclusion.

- Des enregistrements de voix standardisés (dire la voyelle “a” le plus longtemps possible et lire un article de la déclaration des droits de l’Homme) collectés en même temps que les questionnaires électroniques, via l’application smartphone CoLive LIH (développée en interne par le LIH.

Les participants pouvaient choisir de participer à tout ou partie de la collecte de données.

Données collectées à l’inclusion dans la cohorte

Les données collectées à l’inclusion étaient adaptées du cahier d’observation patient (Case Report Form, CRF) ISARIC[135] et étaient groupées selon les catégories suivantes :

- Caractéristiques individuelles,
- Comorbidités
- Symptômes présents à l’inclusion,
- Traitements réguliers avant l’infection COVID-19
- Paramètres cliniques pendant l’hospitalisation, le cas échéant

Données relatives aux conséquences à long terme

Un questionnaire détaillé inspiré de celui développé par Tran et co [136] était proposé aux participants à 12, 15 et 24 mois après l’inclusion dans l’étude. Il consistait en une liste de 64 symptômes les plus courants liés au CL, des questionnaires standardisés et des questions ad-hoc.

Symptômes - La liste des 64 symptômes étaient classés en 8 catégories principales : symptômes généraux, oto-rhino-laryngologiques (ORL), cardiovasculaires, neurologiques, gastro-intestinaux, vasculaires, urinaires et cutanés. Pour chacun des symptômes, les participants étaient invités à répondre à la question suivante : "Avez-vous remarqué les symptômes ou maladies suivants depuis votre diagnostic COVID-19 ? Les modalités de réponse étaient les suivantes 1/ Oui et je le ressens encore aujourd'hui, 2/ Oui, je l'ai eu mais je ne l'ai plus et 3/ Non, je n'ai jamais eu ce symptôme. Une variable "Nombre total de symptômes" qui correspond à la somme de tous les symptômes encore présents à 12 mois a été créée.

Questionnaires standardisés - La qualité du sommeil a été évaluée à l'aide de l’index de qualité du sommeil de Pittsburgh (Pittsburgh Sleep Quality Index, PSQI). Une variable catégorielle a été générée à partir du score PSQI : une mauvaise qualité de sommeil était définie par un score PSQI total > 5[20].

La qualité de vie respiratoire a été évaluée à l'aide du questionnaire de qualité de vie VQ11. Un score global et trois sous-scores (fonctionnel, psychologique et relationnel) ont été calculés comme décrit précédemment[21][22] et des variables catégorielles ont été générées. Une qualité de vie respiratoire altérée a été définie par un score global ≥ 22, une altération de l'autonomie physique par une composante fonctionnelle ≥ 8, une altération de la qualité de vie psychologique par une composante psychologique ≥ 10 et une altération de la qualité de vie relationnelle par une composante relationnelle ≥ 10.

Le niveau de stress a été évalué à l'aide de l'échelle de stress perçu (Perceived Stress Scale, PSS 4). Le score final allait de 0 à 16, le score le plus élevé correspondant à un niveau de stress plus élevé. Un score PSS 4 supérieur ou égal à 6 a été utilisé pour identifier les participants présentant un niveau de stress élevé[137].

L'échelle de gravité de la fatigue (Fatigue Severity Scale, FSS 9), récemment validée dans la population COVID-19, a été utilisée pour mesurer le niveau de fatigue[138]. Le score total FSS 9 correspondait à la moyenne des scores des 9 questions. Un niveau élevé de fatigue était défini par un score total supérieur ou égal à 36[139].

L'échelle du trouble anxieux généralisé à 7 questions (Generalized Anxiety Disorder, GAD 7) a été utilisée pour évaluer le niveau d'anxiété. Un score supérieur ou égal à un seuil de 10 a été considéré comme identifiant un trouble anxieux généralisé[140].

Question ad-hoc - Il a aussi été demandé aux participants s'ils pouvaient gérer leur état de santé actuel à long terme, en tenant compte de tous les symptômes qu'ils ont ressentis au cours des 30 derniers jours en termes de fréquence, d'intensité et d'impact sur leur vie et qui peuvent être attribués à COVID-19 (Oui/Non).

Le questionnaire complet est fourni en annexe 2.

Collecte d'échantillons biologiques

En parallèle de la collecte de données, les participants pouvaient également participer à une collecte d'échantillons biologiques en donnant un consentement éclairé spécifique pour ce volet de l'étude.
Les visites pour la collecte des échantillons étaient organisées à l'inclusion, à 3 semaines puis à 12, 15 et 24 mois.

Les échantillons collectés étaient des frottis nasopharyngés, du sang, des selles, de la salive et des cheveux. Les participants pouvaient choisir de donner tous les types d'échantillons ou seulement une partie.
Les échantillons étaient traités puis stockés à l'Integrated BioBank de Luxembourg (IBBL) puis distribués aux chercheurs en fonction de leurs besoins.

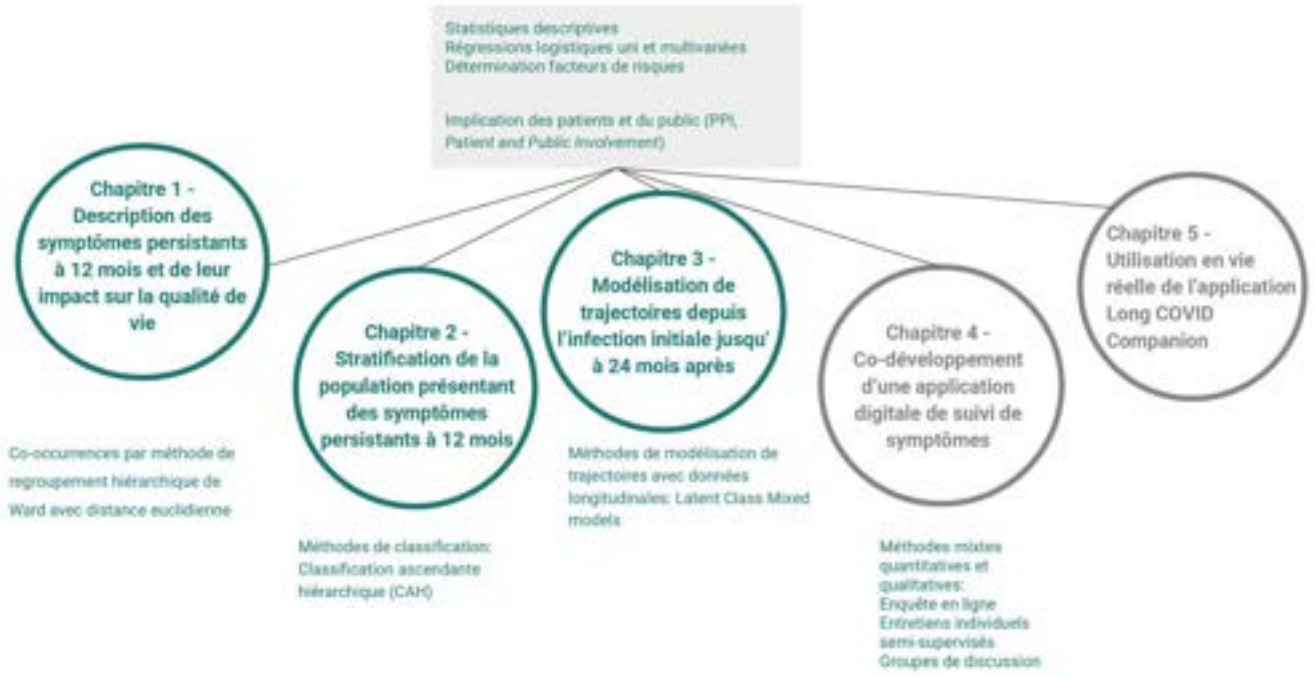
Mon implication

J'ai été impliquée dès mars 2020 dans la conception et la mise en place de la cohorte, en participant à la rédaction du protocole et des formulaires de consentements éclairés destinés aux participants ainsi qu'au choix et au développement des questionnaires. J'ai également coordonné la mise en place des bases de données de l'étude et à la formation des data managers et des infirmiers de recherche à l'étude. J'ai ensuite coordonné le bon déroulement de l'étude entre les différentes équipes impliquées, été point de contact pour toutes les questions relatives aux données collectées et j'ai participé aux réunions du consortium CoVaLux pour assurer le suivi et présenter les résultats de nos recherches.

3.2 Méthodes

Les méthodes utilisées pour les analyses des résultats des 2 axes sont schématisées par la **Figure 11**. Les différentes méthodes sont décrites ici de manière générale puis sont détaillées dans les chapitres respectifs.

Figure 11 : Méthodes utilisées pour chaque chapitre



Implication des patients et du public (*Patient and Public Involvement*, PPI)

L'implication des patients et du public peut se définir par une recherche réalisée "avec" ou "par" des membres du public par opposition à une recherche "sur" ou "pour" eux. Il s'agit d'un partenariat actif dans le cadre duquel les personnes peuvent partager leurs opinions et leurs expériences afin d'influencer et de façonner la recherche. La définition du public du National Institute for Health and Care Research (NIHR) inclut les patients, les patients potentiels, les soignants et les personnes qui utilisent les services de santé et de soins sociaux, ainsi que les personnes issues de communautés spécifiques et d'organisations qui représentent les personnes qui utilisent les services[141]. Le public n'est alors plus uniquement sujet de recherche mais peut être considéré comme acteur à part entière de la recherche.

Le but de l'implication des patients et du public dans la recherche est de réduire le décalage entre ce qui est important pour les patients et ce qui est réellement fait dans la recherche et finalement d'en augmenter la valeur et l'impact[142].

Ainsi l'implication peut être mise en place à tous les stades d'un projet de recherche :

- Avant le début du projet pour définir les objectifs de la recherche, identifier les besoins et problématiques spécifiques à une pathologie, obtenir un retour critique sur le protocole de recherche et en particulier les procédures que les participants devront suivre; pour s'assurer de la faisabilité, faire lire et obtenir l'avis de patients sur les formulaires d'informations et de consentement ou encore les questionnaires, pour s'assurer qu'ils sont clairs et compréhensibles.
- Pendant la réalisation du projet pour aider au recrutement des participants par exemple
- À la fin du projet pour discuter des résultats, ou participer à la dissémination des résultats par des publications et communications scientifiques et par des communications à destination du grand public).

Nous avons appliqué au maximum ces principes tout au long de ce projet de thèse. En particulier, un partenariat de collaboration a été mis en place avec l'association #ApresJ20 COVID Long France, ce qui a permis d'impliquer des représentants de l'association à divers degrés dans les différents chapitres.

Pour les chapitres 1 à 3 la collaboration consistait principalement à une relecture critique des articles et à une participation active en tant que co-auteur des publications.

La collaboration s'est ensuite intensifiée pour les chapitres 4 et 5. Des membres de l'association, mais aussi des professionnels de santé en charge de patients avec un CL ont été impliqués comme co-chercheurs dès le début du projet de développement de l'application digitale, dans la définition des objectifs de l'étude de co-construction, la revue du protocole, des questionnaires et des documents d'information et de consentement destinés aux participants. Des membres de l'association ont également participé à la validation des contenus définitifs de l'application et effectué les premiers tests de l'application avant sa mise en ligne finale. Ils ont aussi contribué largement à la communication autour de ce projet, des résultats et de l'application.

Des professionnels de santé, en France et au Luxembourg, ont également été impliqués dans tous les chapitres, comme contributeurs aux articles, ou co-chercheurs pour les chapitres 4 et 5.

Méthodes statistiques descriptives

Des méthodes statistiques descriptives ont été utilisées pour tous les chapitres 1 à 5.
Classiquement, les variables quantitatives continues normalement distribuées ont été présentées par la moyenne et l'écart-type comme indicateur de dispersion. Les variables quantitatives continues non normalement distribuées ont été présentées par la médiane et les valeurs minimales et maximales comme indicateur de dispersion. Les variables qualitatives ont été décrites par les effectifs et les fréquences des différentes modalités.

Méthodes de classification

Classification Ascendante Hiérarchique (CAH) - Afin de répondre aux objectifs du chapitre 2 de stratifier la population des personnes présentant au moins un symptôme 12 mois après l'infection initiale, nous avons utilisé la méthode de Classification Ascendante Hiérarchique (CAH). La CAH est un algorithme non supervisé faisant partie des méthodes permettant de regrouper des données en sous-groupes présentant des caractéristiques similaires. Cette méthode est adaptée pour des jeux de données de taille limitée, ce qui a guidé notre choix, et permet un classement des objets ce qui est utile pour la représentation graphique[143,144].
La classification est dite ascendante car elle part des individus séparément et cherche à les regrouper en classes homogènes. Pour cela, chaque point du jeu de données est comparé et fusionné à son plus proche voisin. Après chaque regroupement, les distances entre les points et groupes sont recalculées et de nouvelles fusions sont opérées de manière itérative. Le nombre idéal de classes (k) est déterminé à postériori en utilisant la méthode du coude : la classification est effectuée pour différentes valeurs de k, puis pour chaque valeur de k la variabilité intra-classe est calculée puis représentée graphiquement. Le nombre idéal de classe est identifié sur le graphique par le point de cassure (ou coude) pour lequel la variabilité intra-classe est la plus faible et pour lequel la taille des classes contient un nombre suffisant d'individus pour maintenir la relevance clinique .

Modèles mixtes à classes latentes (LCMM) - Pour la modélisation de trajectoires de symptômes du chapitre 3, nous avons utilisé la méthode des modèles mixtes à classes latentes (LCMM) qui permettait d'intégrer toutes les données, y compris celles de participants n'ayant pas complété les données à tous les points[145]. Les modèles mixtes à classes latentes visent à découvrir les profils cachés des trajectoires possibles au sein d'une population diversifiée. Ces modèles fusionnent la théorie des modèles mixtes pour considérer les liens individuels entre les mesures répétées d'une caractéristique et les modèles à classes latentes pour identifier des groupes similaires de participants.

La fonction LCMM disponible dans R[146] est utilisée pour estimer des modèles mixtes lorsque la variable n'a pas une distribution gaussienne. Elle traite les marqueurs continus curvilignes ou non gaussiens en utilisant des fonctions de lien continues, et les marqueurs binaires et ordinaux en utilisant le cadre probit avec des fonctions de lien constantes par morceaux. Cette méthode permet donc de traiter des profils hétérogènes de trajectoire.
Dans le cas des données analysées dans cette thèse, nous avions d'une part une variable continue qui était le score total de symptômes, et des variables binaires qui étaient la présence ou l'absence des symptômes pris individuellement.

Méthodes mixtes quantitatives et qualitatives

Afin de répondre aux objectifs du chapitre 4 qui a pour objectif le co-développement de la solution digitale de santé, nous avons combiné des méthodes quantitatives et qualitatives.

La **recherche quantitative** donne habituellement accès à des mesures, des descriptions généralisables, et des tendances générales (p. ex., analyses statistiques sur un grand groupe de personnes). Elle a pour objectif la mesure et l'**explication** d'un phénomène et utilise des méthodes d'analyses statistiques des données. Elle repose principalement sur la collecte de données digitales observées ou mesurées qui sont analysées pour proposer une description d'un phénomène, effectuer des prédictions, trouver des relations de cause à effet, ou tester une hypothèse[147]. La collecte de données peut être expérimentale, basée sur des enquêtes ou questionnaires, ou au travers d'études de cohorte longitudinales.

La **recherche qualitative** vise à explorer des questions pour lesquelles les méthodes quantitatives sont moins adaptées. Elle repose sur la collecte de données à travers des entretiens, des observations et une documentation systématique (notes, photos, vidéos, etc.). La recherche qualitative n'a pas pour objectif d'être représentative d'une population mais cherche une **compréhension approfondie** et une contextualisation des phénomènes étudiés. Il n'y a donc pas de standard pour calculer la taille d'échantillon idéale et la collecte de données est généralement stoppée quand la saturation des données est atteinte, c'est-à-dire quand il n'y a plus de nouveaux thèmes qui émergent[148–150].

L’analyse des données peut se baser sur différentes méthodes. Parmi les plus utilisées on retrouve la théorie ancrée (ou Grounded Theory) qui permet de faire émerger une théorie nouvelle à partir des données collectées grâce à une méthodologie inductive et dans laquelle l’analyse des données doit se faire avec le moins d’idées préconçues sur le potentiel résultat de la recherche. Cette méthode a été développée dans les années 1960 par Glaser et Strauss et peut être complexe à mettre en œuvre[151]. L’analyse thématique est également une méthode très utilisée en recherche qualitative pour identifier les thèmes les plus pertinents pour l’objectif de recherche au travers des données, et qui semble plus accessible et flexible que la théorie ancrée[152].

En combinant les deux approches, la recherche utilisant des **méthodes mixtes** intègre les atouts et peut potentiellement compenser les limites de chaque méthode pour comprendre un phénomène de manière plus complète et nuancée. Les méthodes mixtes peuvent être mises en place selon les modèles de base décrits par Creswell[153] :

- les modèles séquentiels qui impliquent l'utilisation d'une méthode pour élaborer ou développer les résultats d'une autre. Cette approche permet une intégration des méthodes qualitatives et quantitatives de manière progressive et logique, avec chaque phase de la recherche informant et guidant la suivante. La recherche peut commencer par une phase qualitative pour explorer un phénomène en profondeur. Par exemple, des entretiens ou des groupes de discussion peuvent être utilisés pour identifier des thèmes, des concepts ou des hypothèses qui seront ensuite testés quantitativement. La recherche peut également commencer par une enquête quantitative pour tester une théorie ou examiner des relations statistiques puis être suivie d'une phase qualitative pour approfondir la compréhension des résultats quantitatifs.
- les méthodes mixtes convergentes, dans lesquelles les phases qualitatives et quantitatives sont réalisées de manière indépendante, c’est-à-dire que la collecte et l’analyse des données quantitatives et qualitatives ne dépendent pas l’une de l’autre, contrairement aux modèles séquentiels. Les deux phases ont théoriquement le même poids dans la recherche.

Dans tous les cas, les résultats des phases quantitatives et qualitatives seront combinés et intégrés selon un processus appelé triangulation pour maximiser les connaissances issues des méthodes mixtes.

Plusieurs techniques ont été décrites pour effectuer cette triangulation. De manière générale, les résultats de chaque composante d'une étude sont listés puis examinés de manière à déterminer si les résultats de chaque méthode concordent (convergence), offrent des informations complémentaires sur la même question (complémentarité) ou semblent se contredire (divergence ou dissonance)[154].

Enfin, la dernière étape de ce processus d'intégration est la restitution des résultats. Elle peut se faire de manière narrative mais également grâce à une présentation visuelle conjointe ("Joint display"), sous forme de tableau, de figure ou de représentation graphique résumant les résultats importants des 2 parties quantitatives et qualitatives et de leur intégration[155,156].

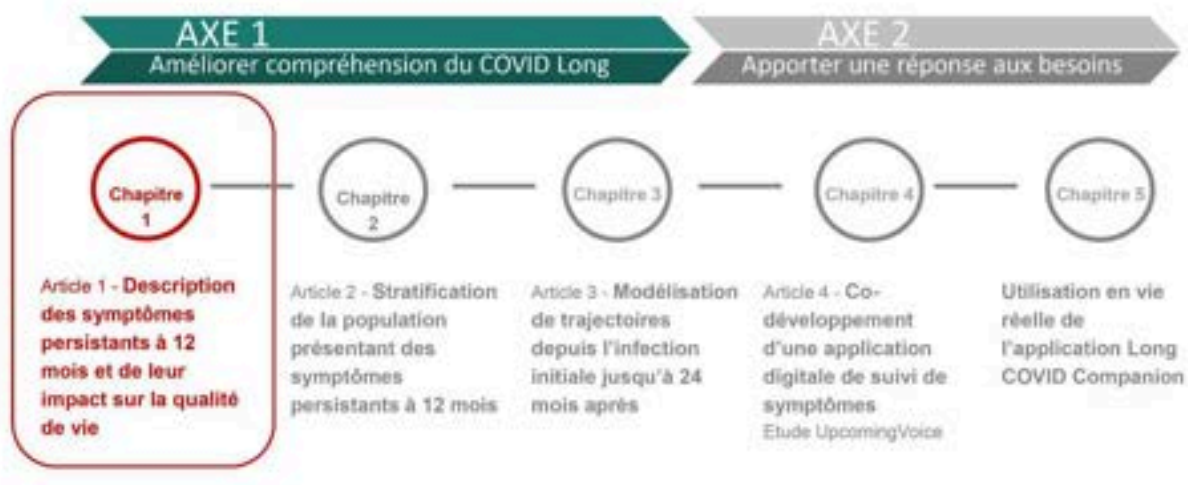
Etant donné la nouveauté de la problématique du CL et de la technologie des biomarqueurs vocaux, il nous a semblé pertinent d'utiliser des méthodes mixtes pour s'assurer d'obtenir la meilleure compréhension globale possible et de ne pas passer à côté d'éléments déterminants.

4. **Résultats de l'Axe 1 :** Description de la problématique du COVID Long dans le contexte du Luxembourg

4.1 Chapitre 1 : Description des symptômes persistants à 12 mois et de leur impact sur la qualité de vie

Article 1 - Publié

Aurélie Fischer, Lu Zhang, Abir Elbéji, Paul Wilmes, Pauline Oustric, Therese Staub, Petr V. Nazarov, Markus Ollert, Guy Fagherazzi. **Long COVID symptomatology after twelve months and its impact on quality of life according to initial COVID-19 disease severity.** Open Forum Infect Dis. 2022 Aug 5;9(8):ofac397. doi: 10.1093/ofid/ofac397. eCollection 2022 Aug. DOI : 10.1093/ofid/ofac397



4.1.1 Résumé en français

Introduction

Le CL se caractérise par une variété de symptômes et représente un fardeau important pour les personnes touchées. Les connaissances sur cette pathologie évoluent rapidement et l'identification de marqueurs prédictifs et de facteurs de risque des séquelles à long terme de COVID-19 a été définie comme une priorité de recherche. Afin de répondre à cette demande nous avons utilisé les données de la cohorte prospective Predi-COVID pour cette première étude.

Notre hypothèse était que la symptomatologie du CL pouvait différer en fonction de la sévérité initiale de la maladie COVID-19 et que les symptômes pouvaient se regrouper et qu'il était dès lors possible de définir des sous-types de CL. Pour vérifier cette hypothèse, nous avons voulu 1) décrire les symptômes rapportés 12 mois après l'infection aiguë par les participants adultes de la cohorte Predi-COVID et présentant différentes formes de gravité de la maladie pendant la phase aiguë, 2) effectuer une analyse de co-occurrence des symptômes et 3) évaluer l'impact du CL sur la qualité de vie.

Méthodes

Les participants à l'étude de cohorte Predi-COVID, recrutés au moment de leur infection aiguë COVID-19, ont été invités à remplir un questionnaire détaillé sur leurs symptômes et leur qualité de vie 12 mois après l'infection initiale (questionnaire M12). La fréquence et la co-occurrence des symptômes ont été étudiées. La qualité de vie a été évaluée grâce aux questionnaires PSQI et VQ11 ainsi qu'en analysant la réponse à la question "En tenant compte de tous les symptômes pouvant être attribués au COVID-19 que vous avez ressentis au cours des 30 derniers jours (fréquence, intensité, impact sur votre vie), diriez-vous que vous pourriez vivre au long court dans votre état de santé actuel ?".

Principaux résultats et conclusions

289 participants ont rempli le questionnaire M12, et 172 d'entre eux (59.5 %) ont rapporté ressentir au moins un symptôme, avec une médiane de 6 symptômes. Par rapport aux participants initialement asymptomatiques qui ont développé des symptômes après l'infection aigüe, ceux dont l'infection initiale était modérée ou sévère ont déclaré plus fréquemment un ou plusieurs symptômes (82.6% contre 38.6 %, $p<0.001$) et présentaient en moyenne 6,8 symptômes de plus (IC 95% [4.18 ; 9.38]). Les fréquences de la plupart des symptômes augmentaient avec la sévérité de l'infection initiale, les plus fréquents étant la fatigue, la dyspnée et l'anxiété. La qualité de vie des participants était fortement impactée avec 12.5% déclarant qu'ils ne pouvaient pas envisager de vivre à long terme dans leur état de santé actuel, une mauvaise qualité de sommeil rapportée par 155 (54.2%) participants et 37 (12.9%) participants rapportant une qualité de vie respiratoire altérée. Les symptômes avaient tendance à se regrouper en clusters. Un premier groupe était principalement constitué de symptômes neurologiques comme la confusion mentale, des pertes de mémoire, des maux de tête, des troubles de l'équilibre ou des tremblements, associés à des symptômes plus généraux comme l'irritabilité, l'anxiété ou la fatigue. Un deuxième groupe est constitué par les symptômes liés à la douleur et les symptômes cardiovasculaires : les douleurs musculaires ou articulaires des membres supérieurs étaient fréquemment rapportées avec les douleurs musculaires ou articulaires des membres inférieurs, les douleurs dorsales, la sensation d'oppression thoracique ou l'arythmie (respectivement 68.9%, 60.5%, 41% et 47.4%). Certains symptômes étaient moins fréquemment rapportés dans l'ensemble, mais avaient tendance à se regrouper les uns avec les autres lorsqu'ils étaient présents, en particulier les symptômes gastro-intestinaux d'une part, et la perte de goût et d'odorat d'autre part.

Finalement nous avons montré que la fréquence et l'impact des symptômes à 12 mois étaient dépendants de la sévérité de la maladie initiale, que les symptômes avaient tendance à se regrouper et que le CL semble donc consister en plusieurs sous-entités et non en un syndrome unique.

Mon implication

J'ai participé à la conceptualisation de cette étude, identifié les données de l'étude Predi-COVID à utiliser, effectué le nettoyage des données, et défini la stratégie d'analyse en collaboration avec la plate-forme de bio-informatique du LIH qui a réalisé les analyses. J'ai ensuite mené les discussions sur l'interprétation des résultats, rédigé l'intégralité du manuscrit, et effectué les corrections suite à la revue des pairs.

4.1.2 Article

ABSTRACT

Introduction: Long COVID is characterized by a variety of symptoms and an important burden for affected people. Our objective was to describe Long COVID symptomatology according to initial COVID-19 severity.

Methods: Predi-COVID cohort study participants, recruited at the time of acute COVID-19 infection, completed a detailed 12-month symptoms and quality of life questionnaire. Frequencies and co-occurrences of symptoms were assessed.

Results: Among the 289 participants who fully completed the 12-month questionnaire, 59.5% reported at least one symptom with a median of 6 symptoms. Participants with an initial moderate or severe acute illness declared more frequently one or more symptoms (82.6% vs 38.6%, $p<0.001$) and had on average 6.8 more symptoms (CI 95% [4.18;9.38]) than initially asymptomatic participants who developed symptoms after the acute infection. Overall, 12.5% of the participants could not envisage coping with their symptoms in the long term. Frequently reported symptoms, like neurological and cardiovascular symptoms, but also less frequent such as gastrointestinal symptoms, tended to cluster.

Conclusion: Frequencies and burden of symptoms present 12 months after acute COVID-19 infection increased with the severity of the acute illness. Long COVID likely consists of multiple sub-categories rather than one single entity. This work will contribute to the better understanding of Long COVID and to the definition of precision health strategies.

Keywords: COVID-19; SARS-CoV-2; Long COVID ; symptoms; severity; cluster

BACKGROUND

Since March 2020, the Covid-19 pandemic has disrupted the entire world population. The novel SARS-CoV-2 virus infected more than 315 million people and caused more than 5.5 million deaths worldwide as of January 2022[157] and, given the current body of knowledge, there are major unsolved questions around the long-term health consequences of Covid-19, which has been termed as “Long COVID” by patients themselves to characterise the multisystemic, fluctuant and debilitating symptoms[7]. Long COVID is defined by the WHO as “Post COVID-19 condition occurs in individuals with a history of probable or confirmed SARS-CoV-2 infection, usually 3 months from the onset of COVID-19 with symptoms that last for at least 2 months and cannot be explained by an alternative diagnosis. Common symptoms include fatigue, shortness of breath, cognitive dysfunction but also others which generally have an impact on everyday functioning. Symptoms may be new onset, following initial recovery from an acute COVID-19 episode, or persist from the initial illness. Symptoms may also fluctuate or relapse over time.”[158]

Indeed, a fraction of the people with COVID-19 experience continuing effects of the disease, with complaints such as tachycardia, extreme fatigue and inability to perform daily physical tasks[159]. The signs and symptoms are diverse and related to multiple organs [160]. Between 53[136] and 203[161] symptoms have been described in 10 organ systems such as systemic, thorax, neurological, digestive but also ears/nose/throat, eyes, vascular, hair and skin or genitourinary. A recent meta-analysis showed that the most prevalent symptoms vary in prevalence over time but the most prevalent have been shown to be fatigue, memory loss, and dyspnea and that the prevalence of Long COVID was higher in people hospitalized during the acute phase[162].

Long COVID can affect adults and children and many of them did not return to the same level of work and quality of life as before COVID-19 infection[161].

Both hospitalized and non-hospitalized persons during acute infection are likely to develop Long COVID. However, it remains unclear to which extent the persistence of symptoms or the appearance of new symptoms is related or not to the severity of acute illness[163,164].

An increasing number of studies report long-term health consequences in asymptomatic or mildly symptomatic patients or hospitalized patients or in mixed cohorts with COVID-19[165][166][167][168][169][170].

For all these reasons, the identification of predictive markers and risk factors of the long-term sequelae of COVID-19 has been defined as a research priority[171].

Our hypothesis is that Long COVID symptomatology may differ according to the initial COVID-19 disease severity and that symptoms can cluster and define sub-types of Long COVID. To test it, we aimed to 1) provide an overview of the symptoms reported 12 months after the acute infection among a cohort of COVID-19 positive adults in Luxembourg, composed of study participants with various forms of disease severity during the acute illness, 2) perform a co-occurrence analysis of the symptoms, and 3) evaluate the impact of Long COVID on quality of life.

METHODS

Population and study design

Data were obtained from participants from the Predi-COVID study, a prospective hybrid cohort study of persons with a PCR-confirmed diagnosis of COVID-19 in Luxembourg. The Predi-COVID study design and analysis plan has been published previously[134]. In brief, all people with a positive COVID-19 PCR test performed by one of the certified laboratories in Luxembourg were contacted by the Health Inspection to explain the measures to be respected and ask them if they agree to be contacted for research purposes. If yes, an experienced research nurse from the LIH contacted the person to explain the study objective and procedures and to collect the participants' consent.
Baseline characteristics were collected at study inclusion, which was performed in the 5 days after the PCR test result. Participants were then followed digitally 12 months after inclusion with self-reported questionnaires on symptoms and quality of life.

Patient consent statement

The written consent of all participants was obtained before inclusion in the study. Data collection in Predi-COVID follows the best practices guidelines from the German Society of Epidemiology[172]. Predi-COVID is registered in ClinicalTrials.gov (NCT04380987) and was approved by the National Research Ethics Committee of Luxembourg (study number 202003/07) and by the Luxembourg Ministry of Health as the authorizing body in April 2020.

Study Design

This study is an analysis of participants' symptoms and health status 12 months after the acute infection. All participants included between May, 1st and November 8th, 2020 were eligible for the present study (N = 539) and were invited to complete a detailed 12-month questionnaire. Among

them, 330 completed the questionnaire (Response rate = 61.2%). We further excluded 41 participants with incomplete data and the final study population was composed of 289 participants. Characteristics of the respondents were compared to those of the non-respondents (Supplementary Table 1).

Symptoms and quality of life

The detailed 12-month questionnaire was inspired from the one co-developed by Tran et co[136] and consisted in a list of the 64 most common symptoms related to Long COVID (full list available in Figure 1), classified in 8 main categories: general, ear nose and throat (ENT), cardiovascular, neurological, gastrointestinal, vascular, urinary and skin symptoms.

Participants were also requested to answer the question "Have you noticed the following symptoms or illnesses since your COVID-19 diagnosis?". Response modalities were: 1/ Yes and I still feel it today, 2/ Yes, I had it but I no longer have it and 3/ No, I have never had this symptom. A variable "Total number of symptoms" has been created that corresponds to the sum of all symptoms still present at 12 months.

Sleep quality was assessed using the PSQI questionnaire. A categorical variable was generated using the PSQI score: a poor sleep was defined as PSQI total score > 5[173].

The respiratory quality of life was assessed using the VQ11 questionnaire. One global score and 3 sub-scores (functional, psychological and relational) were calculated as described elsewhere[174][175] and categorical variables were generated (see Table 1).

Finally participants were asked if they could manage their current state of health in the long run, taking into account all the symptoms that they have experienced in the last 30 days in terms of frequency, intensity, and impact on their life and that can be attributed to COVID-19 (Yes/No).

Covariates

Our analysis took into account the following set of covariates: age, gender, body mass index (BMI), smoking status (never, former and current smoker) and comorbidities (diabetes, asthma, cardiovascular diseases, and hypertension). The disease severity at inclusion was also used as a potential determinant of developing a more severe form of Long COVID.

Disease severity during the acute phase of the disease was defined in 3 categories according to an adapted version of the NIH severity classification[176]: asymptomatic, mild illness and

moderate/severe illness, as previously described here[177]. Hospitalized patients were included in the moderate/severe category.

Descriptive statistics

We described the continuous variables, which were normally distributed, as mean±SD, while the categorical variables as numbers (percentage). We used the student t test and the one-way analysis of variance (ANOVA) to determine the differences of distribution for continuous variables and Fisher's exact tests for categorical variables. Kruskal-Wallis test was used to determine the differences for the variable total number of symptoms.

We studied the co-occurrences of Long COVID symptoms at 12 months and clustered the symptoms using Ward's hierarchical clustering method with Euclidean distance[178].

We performed all the analysis using the R language[179] and generated the figures using the ggplot2 R package[180].

RESULTS

Characteristics of study participants

Individual characteristics of respondents and non respondents to the 12-months questionnaire were similar, except for gender with women being slightly over-represented in the study population (50.2 vs 40.4% p=0.029, supplementary Table 1). Demographic and clinical characteristics of included participants according to the disease severity during the acute phase are provided in **Table 1**.

Table 1: Study population characteristics (Predi-COVID cohort study, N = 289)

	Overall population (N = 289)	Disease severity at inclusion[α]			P-value
		Asymptomatic (N = 44)	Mild (N = 174)	Moderate/severe (N = 58)	
Age (Years)	40.2±12.5	45.4±14.7	39.7±12.1	37.7±11	0,006
Female N(%)	144(50.2)	13(29.6)	91(52.3)	36(62.1)	0,004
Body Mass Index (kg/m2)	25.6±4.8	25.4±3.5	25.5±4.7	26.1±6.1	0,667
Former smoker N(%)	55(19.9)	9(20.5)	35(20.1)	11(19.0)	0,925
Current smoker N(%)	48(17.4)	9(20.5)	30(17.2)	9(15.5)	0,773
Diabetes N(%)	8(2.9)	1(2.3)	5(2.9)	2(3.5)	1,000
Asthma N(%)	8(2.9)	0(0.0)	5(2.9)	3(5.2)	0,382
Cardiovascular disease N(%)	9(3.3)	4(9.1)	4(2.3)	1(1.7)	0,065
Hypertension N(%)	26(9.4)	8(18.2)	13(7.5)	5(8.6)	0,098
Poor sleep[#] N(%)	155(54.2)	17(38.6)	93(54.1)	37(63.8)	0,040
Altered respiratory quality of life[&] at 1 year N(%)	37(12.9)	0(0.0)	16(9.3)	18(31.0)	< 0.001
Altered physical autonomy* at 1 year N(%)	21(7.3)	0(0.0)	7(4.0)	12(20.7)	< 0.001
Altered psychological quality of life** at 1 year N(%)	37(12.9)	0(0.00)	16(9.3)	18(31.0)	< 0.001
Altered relational quality of life*** at 1 year N(%)	11(3.9)	0(0.00)	2(1.2)	9(15.5)	< 0.001
Could not live in their current health status in the long run N(%)	36(12.5)	4(9.1)	22(12.6)	9(15.5)	0,655

[#]Sleep quality was assessed using the PSQI questionnaire. A categorical variable was generated using the PSQI score: # poor sleep was defined as PSQI total score > 5.

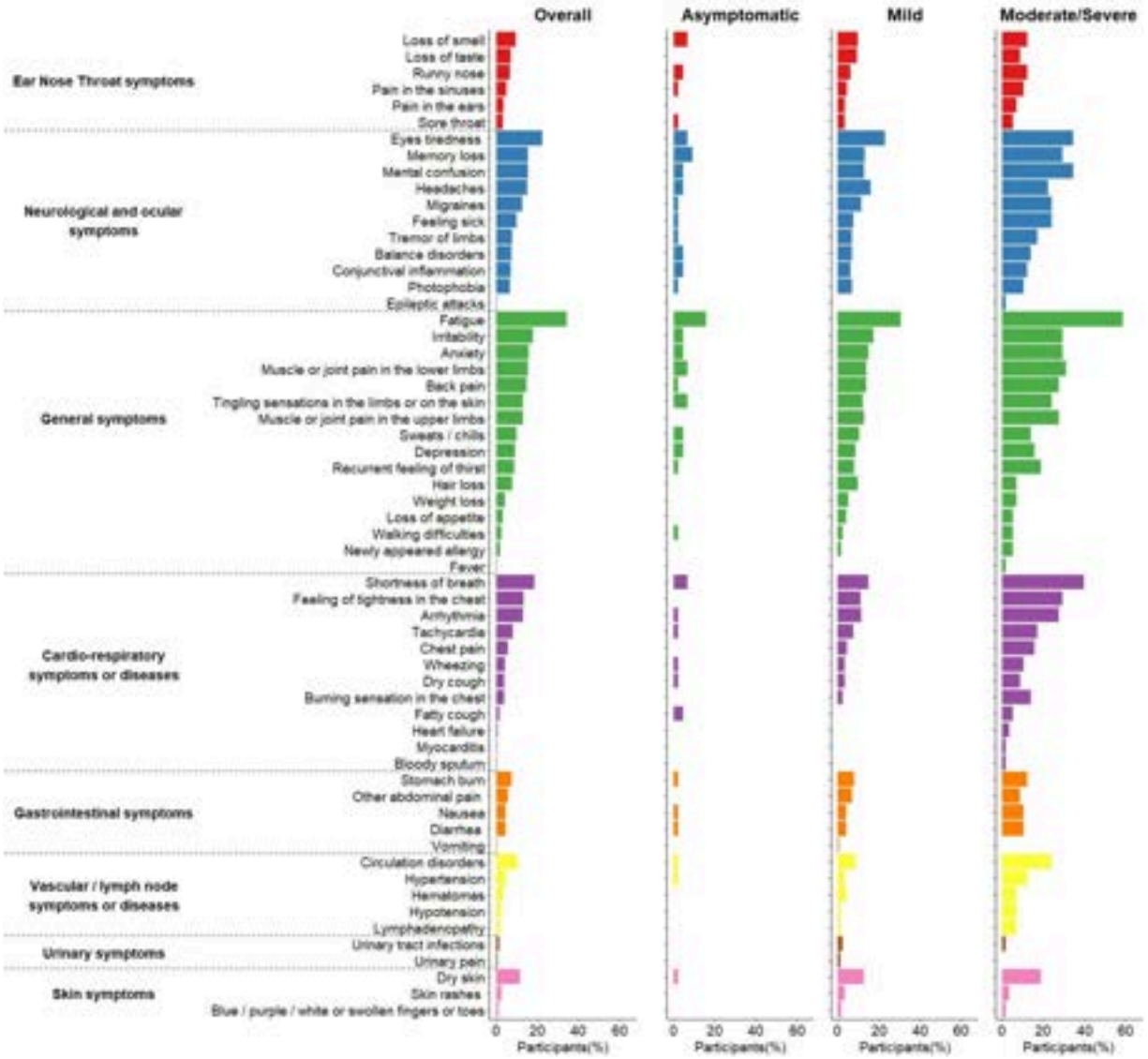
[&]The respiratory quality of life was assessed using the VQ11 questionnaire, initially developed for COPD patients. One global score and 3 sub-scores (functional, psychological and relational) were calculated as described elsewhere and categorical variables were generated. An altered respiratory quality of life& was defined as VQ11 global score > 22, an altered physical autonomy* as functional component > 8, an altered psychological quality of life** as psychological component > 10 and an altered relational quality of life*** as relational component > 10. P-values are determined using the ANOVA Significant Difference test for continuous variables (age and BMI) and the Fisher's exact test for categorical variables. [α]Information on disease severity at inclusion was missing for 13 participants.

Description of persisting symptoms 12 months after acute infection

Among study respondents with complete data, 172 (59.5%) participants presented at least one persisting symptom 12 months after acute infection. Among them, the median total number of persisting symptoms was 6 [2,11] and one third experienced more than 10 symptoms (Supplementary Table 2).

Four out of the 64 symptoms were not reported by any participant (infarction, stroke, hallucinations and necessity of dialysis). The most frequently reported symptoms were fatigue (34.3%), eye tiredness (22.5%), shortness of breath (18.7%) and irritability (18.0%) (see **Figure 1**).

Figure 1: Description of 60 persisting symptoms, 12 months after the acute infection



Symptoms are grouped by symptom category according to disease severity during acute infection. Four symptoms are not presented as no participants reported them: infarction, stroke, hallucinations and necessity of dialysis.

Participants affected by a moderate or severe acute infection reported more frequently at least one symptom 12 months after acute infection in comparison with initially asymptomatic participants that developed symptoms during the follow-up time (82.8% vs 38.6%, p<0.001) and many symptom frequencies were increased (See supplementary Table 2).

Co-occurrences

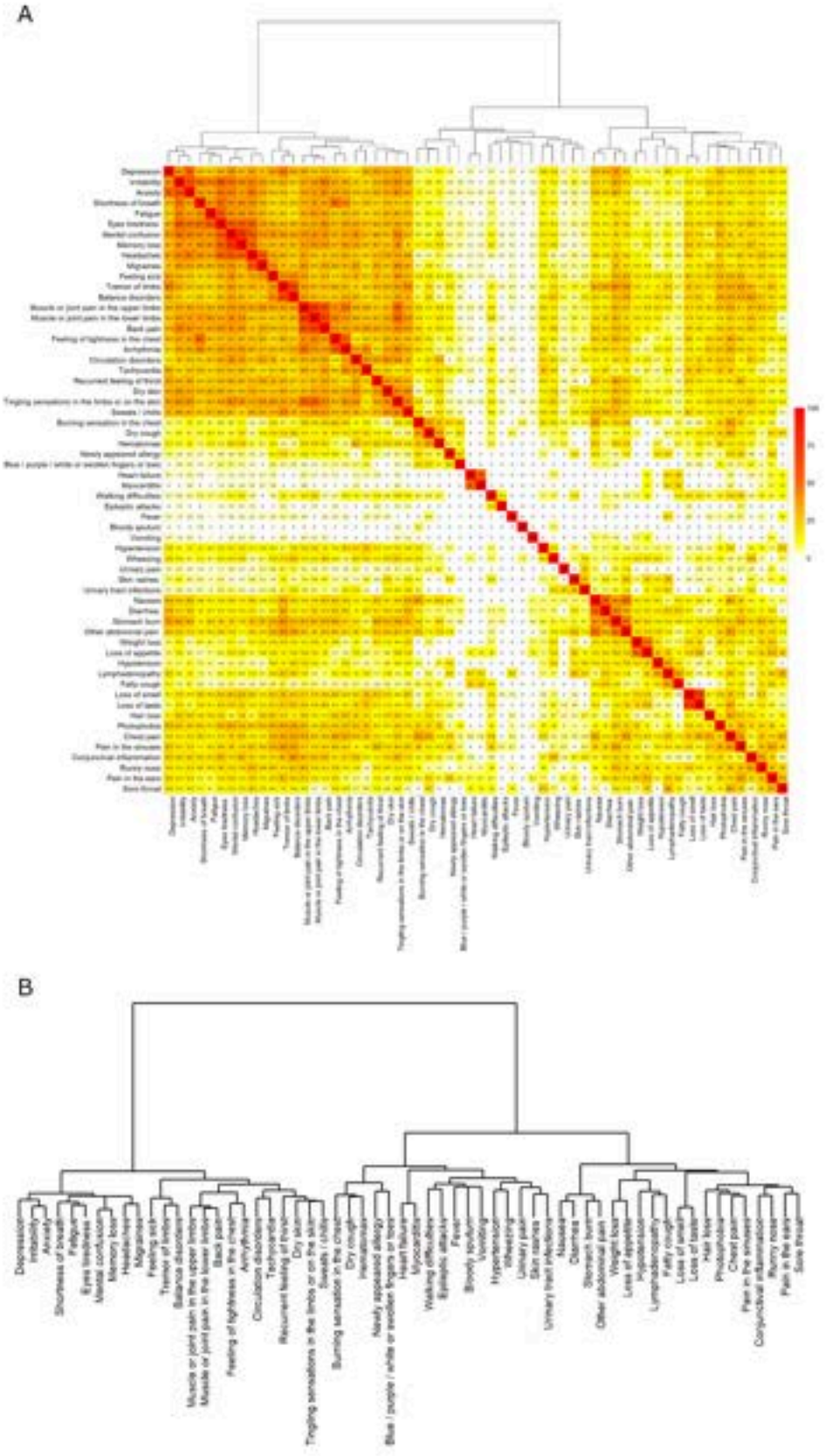
Figure 2 shows the co-occurence rates in the 60 persisting symptoms reported by at least one participant. The 24 most frequent symptoms, including mainly neurological and cardiovascular symptoms, tended to cluster.

A first cluster was mostly constituted by neurological symptoms: when mental confusion was present, memory loss was also reported in 73.3%, headaches in 53.3%, eyes tiredness in 50.8%, irritability in 46.2%, anxiety in 43.5%, shortness of breath in 40.7% and fatigue in 39.4% of the cases. When balance disorders were present, tremor of the limbs was also reported in 60.9% of the cases.

A second cluster was constituted by pain-related and cardiovascular symptoms: muscle or joint pain in the upper limbs was frequently reported together with muscle or joint pain in the lower limbs, back pain, feeling of tightness in the chest or arrhythmia (68.9%, 60.5%, 41% and 47.4%, respectively).

Some symptoms were less frequently reported overall, but tended to cluster with each other when present. In particular, gastrointestinal symptoms constituted one cluster as 50% of the participants with nausea reported also diarrhea, 36.4% of them had also stomach burns and 47.1% of them declared other abdominal pain. Finally when loss of taste was present, loss of smell was also reported in 67.9% of the cases.

Figure 2: Co-occurrence heatmap of the 60 most frequent symptoms present 12 months after acute infection.
(A: Co-occurrences are presented in % per line (for example line “Depression” : we show the % of participants reporting depression and having the second symptom). B: The dendrogram indicates the order in which the symptoms were clustered. The smaller height reflects the earlier joining and the higher similarity between the clusters/symptoms.)



Impact on quality of life

Poor sleep was reported by 155 (54.2%) participants and 37 (12.9%) participants had an altered respiratory quality of life 12 months after the acute infection. The physical autonomy, the psychological and the relational components of the respiratory quality of life score were affected in 7.3%, 12.9% and 3.9% of the participants, respectively. Participants with a moderate/severe form of the disease during the acute phase reported more frequently a poor sleep (63.8% vs 38.6%, $p = 0.04$) and had a more altered respiratory quality of life (31% vs 0%, $p<0.001$) when compared to asymptomatic participants.

Moreover, 36 participants (12.5% of the respondents) answered “No” to the question on whether they could manage their current health status in the long run (Table 1). This represents 20.9% of the 172 symptomatic participants at 12 months. These participants had similar individual characteristics than participants who answered yes to this question (see Supplementary table 3). All symptom frequencies were increased in participants who declared not being able to cope with their symptoms in the long run compared to the overall study population (Supplementary table 4).

DISCUSSION

Our study provides a comprehensive overview of the symptoms reported by people with Long COVID 12 months after an acute SARS-CoV-2 infection. We found that almost 60% of the participants experienced at least one symptom with a predominance of fatigue, dyspnea, and anxiety. The presence of symptoms at 12 months was associated with a moderate or severe form of the acute infection. We also observed clusters of symptoms frequently co-occurring and showed that long term symptoms had a high impact on quality of life.

Comparison with literature

In our cohort 59.5% of the participants who fully completed the 12-month questionnaire declared at least 1 symptom one year after the acute infection. This proportion was higher for participants with moderate/severe form of acute illness (82.8%) than for participants with mild or asymptomatic initial form of the illness (56.3% and 38.6%, respectively). Other prospective cohort studies report similar frequencies of participants experiencing symptoms 12 months after acute infection[181,182]

We obtained similar patterns as reported in other studies, with a variety of symptoms and a predominance of fatigue, dyspnea and anxiety[40,166,183][71]. The individual frequencies of these frequent symptoms were similar to their pooled prevalence described in a recent meta-analysis of 12-months follow-up cohort studies, however those of less frequent symptoms like loss of taste, loss of smell or gastro-intestinal symptoms were higher in our study[184].

Several mechanisms can explain this wide panel of symptoms: SARS-CoV-2 virus can enter cells of many organs via the ACE2 receptors and thus provoke multi-organ damage. Chronic fatigue can be explained by a dysfunction in inflammatory response pathways but other factors may be involved too. Neurological and cognitive impairments may be explained by chronic neuronal inflammation and damage since SARS-CoV-2 is able to pass the blood-brain barrier[160]. An acute infection can also induce chronic disturbance in immune subsets or an activation of an autoimmune response, which is in line with the relapses observed in Long Covid[185]. In particular autoimmunity can be induced by SARS-CoV-2 virus and is dependent on the initial viral load[71] and the severity of the disease[177].
The total number of symptoms present 12 months after acute infection is difficult to compare with existing studies as the cohort types, the timepoints and the way to collect symptoms are different[164][39][136]. However, it was striking that one third of the participants of the Predi-COVID cohort who were symptomatic 12 months after the acute infection were experiencing more than 10 symptoms.

Participants with a moderate or severe form of acute illness had a significantly higher propensity to declare symptoms 12 months after and had a more altered quality of life than those with asymptomatic or mild form of acute illness. Previous studies reported contradictory findings as some studies did not find an association between Long COVID and initial disease severity during acute COVID-19[163][186] but other studies showed that the severity of the acute COVID-19 illness was associated with an increase in Long COVID features[187][188]. In contrast, we also showed that an initial mild or asymptomatic form of acute disease did not prevent from experiencing Long COVID-associated symptoms at 12 months, in particular fatigue, and limited studies are proposing potential mechanisms to explain it[189][190]. This finding is consistent with a review stating that 30 to 60% of patients with an initial asymptomatic or mild form of COVID-19 developed a Long COVID[191].

We observed several clusters of symptoms that often co-occurred. Among frequently reported symptoms, we observed a distinct cluster mainly constituted by neurological symptoms and another by pain-related symptoms. Cardiovascular symptoms were present in both clusters but did not constitute a separate one. In less frequently reported symptoms, we have seen that gastrointestinal symptoms (nausea, diarrhea, stomach burn and other abdominal pain) on one hand and loss of taste and loss of smell on the other hand usually co-occurred. This is consistent with findings from other studies confirming that different clusters of Long COVID can be described and could benefit from adapted treatments and diagnosis [187][192].

Symptoms present at 12 months had an important impact on participants' quality of life, with sleep disturbances, altered respiratory quality of life and discouragement concerning their state of health. People who experienced a moderate or severe form of initial COVID-19 were the most impacted in terms of sleep and respiratory quality of life. Participants who declared not being able to cope with their symptoms in the long run had similar individual characteristics than those who declared the opposite, and had reported higher symptom frequencies. This further underlines the impact of long term symptoms on quality of life.

Strengths and limitations

Our study has several strengths. Firstly, it is a prospective cohort study including COVID-19 positive persons combining various acute illness stages and followed 12 months after acute infection compared to other studies[193][166]. All study participants had a PCR-based COVID-19 diagnosis, which avoided the risk of false negatives in the asymptomatic group of participants. Women were slightly overrepresented in the respondent group, which was expected since it has already been shown that women are more frequently affected by Long COVID compared with men[194]. We focused our analysis on symptoms present at the time of questionnaire completion to limit the risk of recall bias and as such, it was not possible to clearly describe the dynamics of the various symptoms during the year of follow-up.

This study also has several limitations. The response rate to the Long COVID questionnaire was 61.2%, which could lead to an overestimation of symptoms since people without any symptoms might have been less inclined to complete the questionnaire. Participants who responded to the questionnaire had the same individual characteristics as participants who did not respond, which limits but does not completely rule out the risk of selection bias, which could then lead to an overestimation of some symptom frequencies.

We can not fully state that all symptoms reported in the 12-month questionnaire are associated with COVID-19, in particular for symptoms newly appeared in participants asymptomatic during the acute phase of the infection. However the wording of the question ("Have you noticed the following symptoms or illnesses since your COVID-19 diagnosis?") limits the risk of reporting symptoms not related to COVID-19. Moreover, the appearance of symptoms in asymptomatic patients or after recovery of the acute illness has been reported by other studies[195][191]. It can be questionable whether gastrointestinal symptoms could be associated with COVID-19 one year after infection. A "gastrointestinal post-acute COVID-19 syndrome" has been recently characterized with putative physiological mechanisms including persistent viral antigen in the gut, persistent abnormalities in the mucosa and blood, or increase in organ-specific autoantibodies[196].

Symptom intensity was not assessed, except for sleep quality and respiratory quality of life, for which we used validated scales. We did not have the information on potential medication taken by the participants to reduce their symptoms and this could lead to an underestimation of some symptoms.
Study participants were included before the COVID-19 vaccination campaign started in Luxembourg and thus were not vaccinated at inclusion, with unknown vaccination status after 12 months of follow-up. Therefore, we could not assess the impact of COVID-19 vaccination on the development of Long COVID. This would be of high interest to replicate our study as some studies show an association between vaccination and a lower Long COVID risk whereas others showed no association[197][198].

CONCLUSIONS

Our study provides an extensive description of symptoms present 12 months after COVID-19 infection and their impact on the quality of life of COVID-19 patients in a well-characterized prospective cohort. We highlighted a significant burden for people living with Long COVID 12 months after the initial infection. An initial moderate or severe illness was associated with an increase in long term consequences of COVID-19. Finally, our study helps to define Long COVID and confirms that it is multisystemic and presents different clusters of symptoms. These results will ultimately help to better identify long COVID in clinical settings and contribute to the definition of precision health strategies.

Notes

Acknowledgments: We thank the Predi-COVID participants for their involvement in the study, the members of the Predi-COVID external scientific committee for their expertise, as well as the project team, and the nurses in charge of recruitment, data and sample collection, and management on the field.

Author contributions: Fischer and Fagherazzi had full access to all of the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis.

Fischer, Fagherazzi and Zhang conceived the idea and designed the study. Fischer, Fagherazzi, Zhang drafted the manuscript. Zhang and Fischer realized the statistical analysis. Fagherazzi and Ollert obtained funding. Fischer ensured administrative, technical and material support. All authors contributed to the acquisition, analysis, or interpretation of data. All authors contributed to writing and revised the manuscript for important intellectual content, and approved the final version of the manuscript.

Potential Conflicts of Interest: Paul Wilmes is co-speaker of the Research Luxembourg COVID-19 Task Force and a member of the Luxembourg Government-appointed expert group on vaccination mandates. Paul Wilmes is a member of the scientific steering committee for a clinical trial by 4D Pharma plc, unrelated to the present research. Other authors have no conflict of interest.

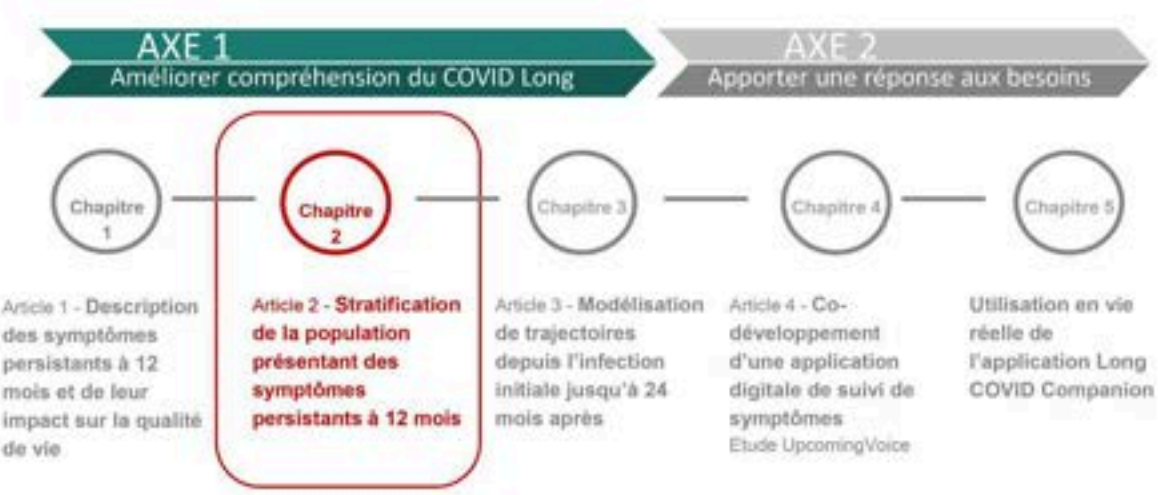
Funding: The Predi-COVID study is supported by the Luxembourg National Research Fund (FNR) (Predi-COVID, grant number 14716273), the André Losch Foundation, and the Luxembourg Institute of Health. The related work of Paul Wilmes is supported by the Luxembourg National Research Fund (FNR), and the European Union's Horizon 2020 research and innovation program including the European Research Council and the Luxembourg Government.

Role of the Funder/Sponsor: The study funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

4.2 Chapitre 2 - Stratification de la population présentant des symptômes persistants à 12 mois

Article 2 - Publié

Aurélie Fischer, Nolwenn Badier, Lu Zhang, Abir Elbéji, Paul Wilmes, Pauline Oustric, Charles Benoy, Markus Ollert, and Guy Fagherazzi. **Long COVID classification: findings from a clustering analysis in the Predi-COVID cohort study.** Int. J. Environ. Res. Public Health 2022, 19(23), 16018; DOI : 10.3390/ijerph192316018



4.2.1 Résumé en français

Introduction-contexte

Le chapitre 1 a fourni une description détaillée des symptômes présents 12 mois après l'infection COVID-19 initiale dans une cohorte prospective bien caractérisée. Les résultats étaient similaires à ceux d'autres études de cohorte avec la fatigue, la dyspnée et l'anxiété étant les symptômes les plus fréquemment rapportés.

Nous avons mis en évidence un impact important des symptômes 12 mois après l'infection initiale sur la vie des personnes concernées. Une infection initiale modérée ou sévère était associée à une augmentation des conséquences à long terme du COVID-19. Cette première étude a ainsi participé à la définition du CL et a confirmé qu'il était multisystémique et constitué de symptômes fréquemment présents simultanément.

Au vu de ces résultats, il était important d'approfondir cette caractérisation du CL en investiguant la possibilité de stratifier la population des personnes touchées par le COVID en sous-groupes présentant des manifestations cliniques différentes.
L'objectif de cette étude était de concevoir une classification du CL, facile à utiliser, pour aider à stratifier les personnes touchées et à personnaliser leur prise en charge.

Méthodes

Nous avons collecté les caractéristiques individuelles et une liste détaillée de 62 symptômes persistants autodéclarés ainsi que des indices de qualité de vie 12 mois après l'infection initiale par COVID-19 chez les participants de la cohorte Predi-COVID. Une classification ascendante hiérarchique (CAH) a été utilisée pour identifier les sous-groupes de personnes.

Principaux résultats et conclusions

Nous avons identifié 3 profils de personnes touchées par le CL avec un gradient de sévérité de la maladie. Le groupe "léger" englobait près de 50 % de la population étudiée et était composé de participants présentant une infection initiale moins grave, moins de comorbidités et moins de symptômes persistants (moyenne = 2.9). Le groupe "modéré" se caractérisait par une moyenne de 11 symptômes persistants et une mauvaise qualité de vie sur le plan du sommeil et de la respiration. Le groupe "sévère" se caractérisait, lui, par une plus grande proportion de femmes et de fumeurs que dans les autres groupes, avec un plus grand nombre de symptômes liés au CL, en particulier des symptômes vasculaires, urinaires et cutanés. Notre étude a montré que le CL peut être stratifié en trois sous-catégories de sévérité croissante.
Après reproduction et validation dans d'autres populations, cette classification simple aidera les cliniciens à améliorer les soins prodigués aux personnes souffrant de CL.

Mon implication

J'ai participé à la conceptualisation de cette étude, identifié les données l'étude Predi-COVID à utiliser, effectué le nettoyage des données, et défini la stratégie d'analyse. J'ai encadré pour cette étude une étudiante en Master 1 qui a réalisé les analyses. J'ai ensuite mené les discussions sur l'interprétation des résultats, rédigé le manuscrit, et effectué les corrections suite à la revue des pairs.

4.2.2 Article

ABSTRACT

The increasing number of people living with Long COVID requires the development of more personalized care, as for now limited treatment options and rehabilitation programs adapted to the variety of Long COVID presentations are available. Our objective was to design an easy-to-use Long COVID classification to help stratifying people with Long COVID. Individual characteristics and a detailed set of 62 self-reported persisting symptoms together with quality of life indexes 12 months after initial COVID-19 infection were collected in a cohort of SARS-CoV-2 infected people in Luxembourg. A hierarchical ascendant classification (HAC) was used to identify clusters of people. We identified 3 patterns of Long COVID symptoms with a gradient in disease severity. Cluster-Mild encompassed almost 50% of the study population and was composed of participants with less severe initial infection, fewer comorbidities, and fewer persisting symptoms (mean=2.9). Cluster-Moderate was characterized by a mean of 11 persisting symptoms and poor sleep and respiratory quality of life. Cluster-Severe was characterized by a higher proportion of women and smokers as in the other clusters, with a higher number of Long COVID symptoms, in particular vascular, urinary, and skin symptoms. Our study evidenced that Long COVID can be stratified into 3 sub-categories in terms of severity. If replicated in other populations, this simple classification will help clinicians to improve the care of people with Long COVID.

KEYWORDS: Clustering; COVID-19; Long COVID; disease severity

1. Introduction

It is now estimated that a mean of 10 to 20% of the people infected by the SARS-CoV-2 experience persisting and fluctuating symptoms more than 12 weeks after the acute infection[199][200]. This syndrome has been called “Long COVID” by patients themselves and has a high impact on the quality of life of the affected people and as a consequence on the whole healthcare system.

Long COVID has been defined by WHO as a condition that occurs 3 months after infection with SARS-CoV-2, with symptoms that last at least 2 months and cannot be explained by any other diagnosis[201], but this definition does not account for the substantial intra-group variability in the different presentations of Long COVID.

Many studies described Long COVID in post-hospitalization cohorts[39][202][203] and in population-based studies of less severe forms of COVID-19 with similar results[204][205], the most common reported symptoms being fatigue, shortness of breath, cognitive dysfunction, with usually a major impact on daily life [206][204][205]. Long COVID affects many organs with pulmonary, cardiac, thromboembolic, neurologic, and renal sequelae. However, their distribution and intensity in the general population are largely heterogeneous [207].

A one-size-fits-all care strategy for people with Long COVID is therefore not possible and a better understanding of sub-forms of Long COVID would allow to develop personalized care for people with Long COVID or could be integrated as a screening tool for future clinical trials[208]. To date, few studies used clustering analysis to identify and characterize different Long COVID phenotypes[61,62,205].

In this study, we hypothesized that Long COVID can be stratified into different clinically relevant sub-groups. We applied hierarchical clustering to study participants with Long COVID from the Predi-COVID cohort study to test this hypothesis.

2. Materials and Methods

STUDY POPULATION

We used data from the Predi-COVID study, a prospective cohort study of persons with a PCR-confirmed diagnosis of COVID-19 in Luxembourg. The study design and objectives have been published previously[134]. Participants were followed-up at 12 months with a self-reported questionnaire to update their general health status, persisting symptoms, and quality of life. The Predi-COVID study was approved by the National Research Ethics Committee of Luxembourg (study number 202003/07) and by the Luxembourg Ministry of Health as the authorizing body in April 2020.

Individual characteristics, comorbidities, and initial symptoms were collected at inclusion in the Predi-COVID study. Initial COVID-19 disease severity ("Asymptomatic," "Mild illness," and "Moderate/severe illness") has been previously assessed as described elsewhere[209][177].

Persisting symptoms were collected using a list of 62 symptoms[208], further divided into 8 categories: ear/nose/throat symptoms, neurological and ocular symptoms, general symptoms, cardio-respiratory symptoms or diseases, gastrointestinal symptoms, vascular and ganglionic symptoms or diseases, urinary symptoms, and skin symptoms (see Online Supplementary Table 1 for the full list).

Sleep quality was assessed using the Pittsburgh Sleep Quality Index [173]. The respiratory quality of life was assessed with the VQ11 questionnaire (global score and 3 sub-scores) [174]. Finally, participants were asked whether they could envisage coping with their current health status in the long term (yes/no).

Inclusion criteria for our analysis were: adult participants with a complete 12-month questionnaire and baseline data available and who declared at least one persisting symptom.

CLUSTERING AND STATISTICAL ANALYSIS

The clustering was based on the following features: sociodemographic characteristics, initial classification of COVID-19 disease severity, comorbidities, symptoms at inclusion, and quality of life (See Online Supplementary Table 2 for the full list).

A Hierarchical Ascendant Classification (HAC) was used to construct clusters[205]. The optimal number of clusters has been determined using the "elbow" method calculating the distortion depending on the number of clusters with the objective to maintain clinical interpretability and sufficient cluster size. The cluster stability was assessed with the Jaccard Similarity Index. A simple imputation was done for variables if they had less than 5% of missing data (using median for quantitative variables and main modality for categorical variables) and multiple imputations using the mice package from R otherwise. Data were described with numbers and percentages for categorical variables and with mean and standard deviation for numerical variables. We performed all the analysis using the R software[210] and generated the figures using the ggplot2 R package[180].

3. Results

3.1. POPULATION STUDY CHARACTERISTICS

We initially included 545 participants between May 2020 and May 2021 with an available follow-up questionnaire 12 months after their primary infection. Participants with incomplete questionnaires were excluded (n=54) as well as participants with an age below 18 (n=1), and participants without any information about their study inclusion (n=19) or about their initial COVID-19 severity classification (n=3). Participants who did not experience any symptoms at 12 months were removed (n=180). Finally, 288 participants were considered in the analysis (see Online Supplementary Figure 1).

The majority of the overall study participants were females (59%) and were not hospitalized at the time of COVID-19 (97%). The average age was 43 years (sd=12) and 16% of the participants were smokers. One-third (33 %) of the participants had a moderate/severe form of the initial COVID-19. Sixty percent of the participants experienced poor sleep quality (PSQI total score>5) and 28% had a poor respiratory quality of life (VQ11 global score>22). Few participants had comorbidities prior to COVID-19 diagnosis (14%) and they had an average of 2.38 (sd=0.33) comorbidities. Hypertension was the most frequent one (13%). At the time of inclusion, the most frequent symptoms were fatigue/malaise (47%), fever (34%), cough (33%), cephalea (27%), and rhinorrhea (26%).

On average, participants declared 8 symptoms (sd=8) after 12 months. Most participants had general symptoms (80%), neurological and ocular symptoms (65%), and cardio-respiratory symptoms (55%).

3.2. CLUSTERS

Based on the elbow curve (see **Figure 1**), we determined the optimal cluster number as 3, which simultaneously allows good cluster stability (Cluster-Mild, Jaccard=0.5707; Cluster-Moderate, Jaccard=0.7556; and Cluster-Severe, Jaccard=0.8297), clinical interpretability, and sufficient cluster size for each cluster.

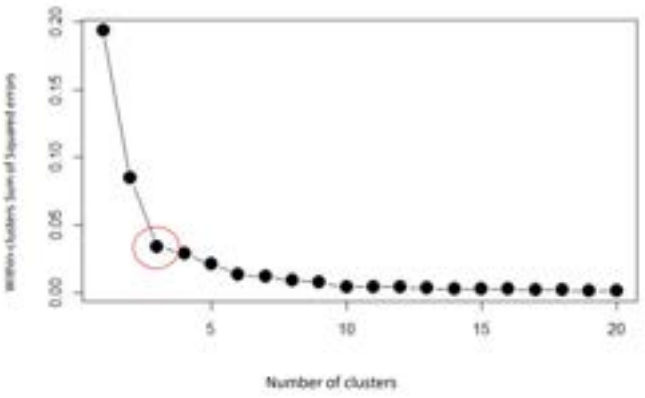


Figure 1. Determination of optimal cluster number. The optimal number is visualized by the inflection point that corresponds to 3.

We labeled them according to their distinguishing characteristics. The characteristics of the overall study population and of the 3 clusters are shown in **Table 1**.

Table 1. Participants' characteristics in the overall study population and by cluster.

		Overall Population N = 288	Cluster – Mild N = 139 (48.26%)	Cluster – Moderate N = 106 (36.81%)	Cluster – Severe N = 43 (14.93%)	P value*
Sociodemographic Characteristics and Initial Severity Classification	Female N(%)	170 (59%)	73 (53%)	66 (62%)	31 (72%)	0.053
	Age (Years)	43 ±12	42 ±12	43 ±12	45 ±14	0.360
	Body Mass Index (kg/m2)	26.4 ±5.5	25.8 ±5.1	27.0 ±5.8	26.7 ±5.7	0.224
	Smoker N(%)	45 (16%)	16 (12%)	15 (14%)	14 (33%)	0.027
	Moderate/severe illness N(%)	95 (33%)	34 (24%)	41 (39%)	20 (47%)	0.015
Comorbidities	At least one comorbidity N(%)	40 (14%)	12 (8.6%)	16 (15%)	12 (28%)	0.007
	Number of comorbidities Mean(SD)	2.38 ±0.33	2.37 ±0.25	2.34 ±0.16	2.48 ±0.68	0.001
	Hypertension N(%)	38 (13%)	14 (10%)	12 (11%)	12 (28%)	0.015
	Cardiac diseases N(%)	11 (3.8%)	3 (2.2%)	6 (5.7%)	2 (4.7%)	0.311
	Asthma N(%)	14 (4.9%)	4 (2.9%)	8 (7.5%)	2 (4.7%)	0.200
	Diabetes N(%)	13 (4.5%)	3 (2.2%)	4 (3.8%)	6 (14%)	0.009
Symptoms at inclusion N(%)	Fever	98 (34%)	45 (32%)	36 (34%)	17 (40%)	0.688
	Cough	96 (33%)	41 (29%)	38 (36%)	17 (40%)	0.362
	Cough Sputum	27 (9.4%)	11 (7.9%)	9 (8.5%)	7 (16%)	0.279
	Sore throat	50 (17%)	17 (12%)	24 (23%)	9 (21%)	0.076
	Rhinorrhea	76 (26%)	35 (25%)	31 (29%)	10 (23%)	0.708
	Earache	22 (7.6%)	8 (5.8%)	10 (9.4%)	4 (9.3%)	0.490
	Chest Pain	19 (6.6%)	4 (2.9%)	11 (10%)	4 (9.3%)	0.036
	Myalgia	51 (18%)	11 (7.9%)	28 (26%)	12 (28%)	<0.001
	Arthralgia	25 (8.7%)	4 (2.9%)	14 (13%)	7 (16%)	0.001
	Fatigue	136 (47%)	47 (34%)	60 (57%)	29 (67%)	<0.001
	Dyspnea	33 (11%)	10 (7.2%)	16 (15%)	7 (16%)	0.067
	Cephalea	77 (27%)	27 (19%)	36 (34%)	14 (33%)	0.022
	Abdominal pain	14 (4.9%)	4 (2.9%)	3 (2.8%)	7 (16%)	0.004
	Nausea	13 (4.5%)	5 (3.6%)	4 (3.8%)	4 (9.3%)	0.289
	Diarrhea	20 (6.9%)	5 (3.6%)	8 (7.5%)	7 (16%)	0.019
Persisting symptoms at 12 months N(%)	Ear Nose Throat (ENT) symptoms	110 (38%)	24 (17%)	65 (61%)	21 (49%)	<0.001
	Neurological symptoms	188 (65%)	51 (37%)	101 (95%)	36 (84%)	<0.001
	General symptoms	229 (80%)	80 (58%)	106 (100%)	43 (100%)	<0.001
	Cardio-respiratory symptoms	159 (55%)	33 (24%)	87 (82%)	39 (91%)	<0.001
	Gastrointestinal symptoms	63 (22%)	7 (5.0%)	32 (30%)	24 (56%)	<0.001
	Vascular symptoms	76 (26%)	10 (7.2%)	29 (27%)	37 (86%)	<0.001
	Urinary symptoms	16 (5.6%)	2 (1.4%)	0 (0%)	14 (33%)	<0.001
	Skin symptoms	66 (23%)	17 (12%)	12 (11%)	37 (86%)	<0.001
Number of persisting symptoms at 12 months Mean(SD)	Total number of symptoms	8 ±8	2.89 ±2.15	11.5 ±5.7	18 ±9	<0.001
	Number ENT symptoms	0.70 ±1.11	0.25 ±0.63	1.12 ±1.24	1.09 ±1.44	0.079
	Number neurological symptoms	2.12 ±2.28	0.72 ±1.27	3.27 ±2.07	3.79 ±2.63	<0.001
	Number general symptoms	3.02 ±2.86	1.19 ±1.48	4.04 ±2.30	6.44 ±3.13	<0.001
	Number cardio-respiratory symptoms	1.36 ±1.72	0.42 ±0.92	2.02 ±1.65	2.81 ±2.11	0.002
	Number gastrointestinal symptoms	0.39 ±0.87	0.079 ±0.382	0.48 ±0.86	1.19 ±1.35	0.010
	Number vascular symptoms	0.39 ±0.75	0.09 ±0.33	0.41 ±0.73	1.35 ±0.95	0.356
	Number urinary symptoms	0.07 ±0.32	0.01 ±0.11	0.00 ±0.00	0.44 ±0.70	0.610
	Number skin symptoms	0.27 ±0.54	0.14 ±0.38	0.13 ±0.39	1.05 ±0.62	0.570
Quality of life N(%)	Could not envisage coping with symptoms long term	45 (16%)	11 (7.9%)	24 (23%)	10 (23%)	0.002
	Poor sleep[#]	239 (83%)	102 (73%)	99 (93%)	38 (88%)	<0.001
	Altered respiratory quality of life[&] at 1 year	81 (28%)	8 (5.8%)	51 (48%)	22 (51%)	<0.001

Sleep quality was assessed using the PSQI questionnaire. A categorical variable was generated using the PSQI score: [#] poor sleep was defined as PSQI total score > 5.

The respiratory quality of life was assessed using the VQ11 questionnaire, initially developed for COPD patients. One global score and 3 sub-scores (functional, psychological, and relational) were calculated as described elsewhere and categorical variables were generated. Altered respiratory quality of life[&] was defined as VQ11 global score > 22. *P-values are determined using the ANOVA Significant Difference test for continuous variables (age and BMI) and Fisher's exact test for categorical variables.

Cluster-Mild contains 139 participants (48.26%). Compared with the overall study population, the initial disease severity was classified as moderate/severe for only 24% of the members of Cluster-Mild. Individuals in this cluster had a less impacted quality of life than the overall study population: only 7.9% declared that they could not envisage coping with their symptoms in the long term, 40% of them had poor sleep quality and 5.8% had poor respiratory quality of life. Overall, participants in Cluster-Mild had fewer comorbidities (8.6%). At 12 months, participants declared fewer symptoms overall (mean number=2.89, sd=2.15). The symptoms were mostly grouped in the following categories: general symptoms (58%), neurological and ocular symptoms (37%), and cardio-respiratory symptoms or diseases (24%).

Cluster-Moderate contains 106 participants (36.81%). Compared with the overall study population, members were slightly more frequently female (62%) and presented more frequently a moderate/severe form of the initial illness (39%). Quality of life was more impacted with 23% of Cluster-Moderate declaring that they could not envisage coping with their symptoms in the long term, 78% of them having a poor sleep quality, and 48% having a poor respiratory quality of life. Comorbidities were similar in Cluster-Moderate and in the overall study population but participants declared a higher number of symptoms at 12 months (mean=11.5, sd=5.7). All participants had general symptoms (100%), and a large majority also had neurological and ocular symptoms (95%) and cardio-respiratory symptoms or diseases (82%). Most participants also had ENT symptoms (61%).

Cluster-Severe contains 43 participants (14.93%). Compared with the overall study population, members were a majority of females (72%). Participants were more frequently smokers (33%) and 47% had an initial moderate/severe acute illness. As for Cluster-Moderate, the quality of life in Cluster-Severe was highly impacted with 84% of them having poor sleep quality and 51% having a poor respiratory quality of life. Overall, participants in Cluster-Severe presented more comorbidities at inclusion (28%), hypertension being the most frequent one (28%). At 12 months, participants had a high number of symptoms (mean=18, sd=9). The presentation of symptoms was similar to Cluster-Moderate for general, neurological, and cardio-respiratory symptoms: all participants had general symptoms (100%), 84% had neurological and ocular symptoms or diseases, and 91% had cardio-respiratory symptoms or diseases. What characterizes Cluster-Severe is the high frequencies of vascular, skin, and urinary symptoms (86%, 86%, and 33%, respectively).

The symptom distribution by symptom categories in the 3 clusters is represented in **Figure 2** which shows the differences between the clusters.

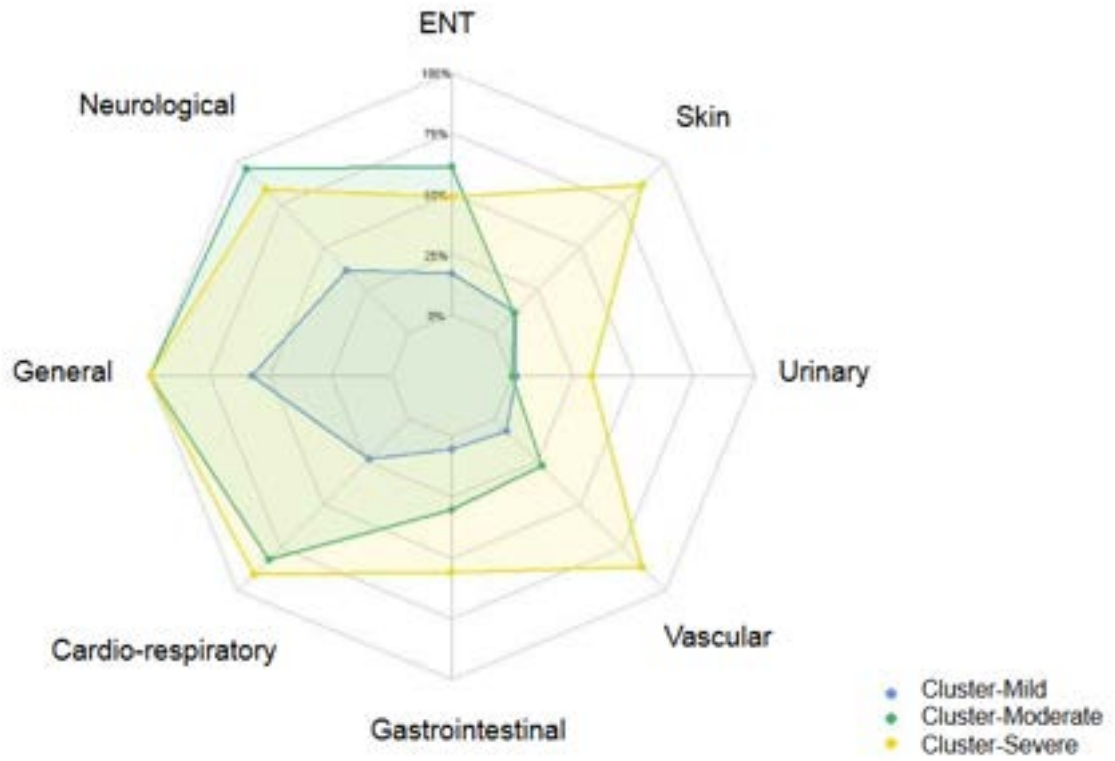


Figure 2: Distribution of Long COVID symptoms (in %) by symptom categories in the 3 clusters.

4. Discussion

In this study, we identified 3 clusters of Long COVID in people with persisting symptoms 12 months after acute infection with a clear gradient in Long COVID severity. Cluster-Mild represented almost half of the study population and was composed of participants with less severe initial infection, fewer comorbidities, and with few persisting symptoms (mean=2.9), mainly in the general, neurological, or cardiorespiratory categories. Individuals in Cluster-Moderate declared a mean of 11.5 persisting symptoms and had a poor quality of sleep and of respiratory quality of life. Cluster-Severe was characterized by a higher proportion of women, of smokers, with a higher number of preexisting comorbidities than in Clusters-Mild and Moderate. Strikingly, participants from Cluster - Severe declared more persisting symptoms in total than those from Cluster-Moderate (mean=18), with a similar pattern of general, neurological, and cardio-respiratory symptoms, but is distinct by higher occurrences of vascular, urinary, and skin symptoms.

General symptoms were predominant in all 3 clusters. This is in line with previous findings showing that general symptoms were the most frequently reported symptoms in people with persisting symptoms at 12 months, with a predominance of fatigue (34.3%), irritability (18%), anxiety (15.9%), muscle or joint pain in the lower limb (15.6%), back pain (14.9%) [61].

Few studies investigated clustering analysis of Long COVID patients. Kenny et al. applied similar clustering methods to a prospective cohort of 233 COVID-19-infected patients with ongoing symptoms at least 4 weeks after acute infection and described also 3 clusters: the larger one constituted by participants with a lower number of persisting symptoms (mean=2) and 2 characterized by a higher number of persisting symptoms (mean=4 and 6) and more functional impairments. As in our study, the distribution of persisting symptoms was different between the 2 most severe clusters, with one cluster grouping cardio-respiratory and general symptoms, and the other one with a predominance of pain-related symptoms.
The time and method of symptom evaluation were different as it was done in person during a visit to a clinic and the median time of symptom duration was 18 weeks[61]. Another study identified 3 different clusters among a cohort of 1969 post-hospitalized COVID-19 patients in Spain[62]: one cluster grouped patients with fewer comorbidities and symptoms at the hospital inclusion, less persisting symptoms, and had a preserved quality of life, and the 2 other clusters were constituted of patients with more pre-existing comorbidities, a higher number of symptoms during the acute phase, a higher number of persisting symptoms and higher impact on quality of life (higher level of anxiety and altered sleep quality). One cluster was also characterized by respiratory symptoms (dyspnea at rest 73.4%) and particularly high limitations in daily activities (92.1% for social activities and 93.3% for instrumental daily activities). The overall number of symptoms in each cluster was lower than in our clusters because their clustering included also participants without persisting symptoms.

Another study conducted in the United Kingdom in 2022 also described groups of people with Long COVID. More participants (N=2550) were recruited, via an online survey, with a mean duration of illness of 7.2 months (sd=1.8). The mean age was similar to our participants, as was the greater presence of women and comorbidities. The most common first symptoms (fatigue, headache, chest pain, shortness of breath, and cough), persistent symptoms (fatigue, cognitive dysfunction, chest pain, shortness of breath, headache, and muscle pain), number of symptoms experienced, and organ systems affected were also similar. Participants were asked to report the presence or absence of 35 symptoms, and two groups were identified.

The first group (88.8%) had mainly cardiopulmonary, cognitive, and fatigue symptoms and the second group had more multisystem symptoms [205], which is relatively well aligned with our findings.

Reese et al applied an adapted Phenomizer algorithm to classify patients with Long COVID, based on the ICD-10 diagnosis code U09.9 for Post COVID-19 condition, and identified 6 clusters[60]. Although the clustering method was different and based on medical records data, this study also identified 2 "severe" clusters with more pre-existing comorbidities, an increased initial illness, and a wide range of Long COVID symptoms.

The overrepresentation of women in the most severe cluster is consistent with findings from other studies[61,205].

Finally, despite different analysis time points, similar results were found in these different studies which confirm that our findings are relevant despite the fluctuating character of Long COVID.

STRENGTHS AND LIMITATIONS

This study has several strengths. First, a large list of 62 symptoms was considered, distributed in 8 categories that cover the complex symptomatology of Long Covid. Participants with different forms of initial illness severity were represented. All participants had a documented initial COVID-19 infection, confirmed by a PCR test and their symptoms were assessed 12 months after acute infection.

This study also has some limitations. The analyses were done on a moderate sample size and, as in any selected study population, results may not be directly extrapolated to all people with Long COVID. External validation in a larger population would be of the highest interest to confirm these results. Information on pre-existing symptoms before COVID-19 infection was missing and symptoms were self-reported which could lead to biases in the estimation of the number of persisting symptoms attributable to COVID-19. However, this may not affect the main message of our findings.

The participants in the present study were included before the Omicron wave, thus we can not ensure that our results can be extended to Long COVID following infection by the Omicron variant. Recent studies demonstrated that infection by Omicron variants leads to a 24 to 50% risk reduction of developing Long COVID, however, there were no differences in the distribution of Long COVID symptoms and the risk of neurological and psychiatric sequelae remains the same after infection by Omicron[36][211][56].

5. Conclusions

Our study highlighted 3 clinically relevant subgroups of people with Long COVID of increasing severity, but also with different patterns of symptoms. Such stratification of Long Covid will help healthcare professionals to improve the triage and care of people with Long COVID.

Supplementary Materials: Not applicable.

Author Contributions: N.B., A.F., and G.F. wrote the manuscript. All the authors interpreted the data, critically revised the manuscript, and approved the final version. All authors have read and agreed to the published version of the manuscript.

Funding: The Predi-COVID study is supported by the Luxembourg National Research Fund (FNR) (Predi-COVID, grant number 14716273), the André Losch Foundation, and the Luxembourg Institute of Health.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the National Research Ethics Committee of Luxembourg (study number 202003/07).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are available from the corresponding author upon reasonable request.

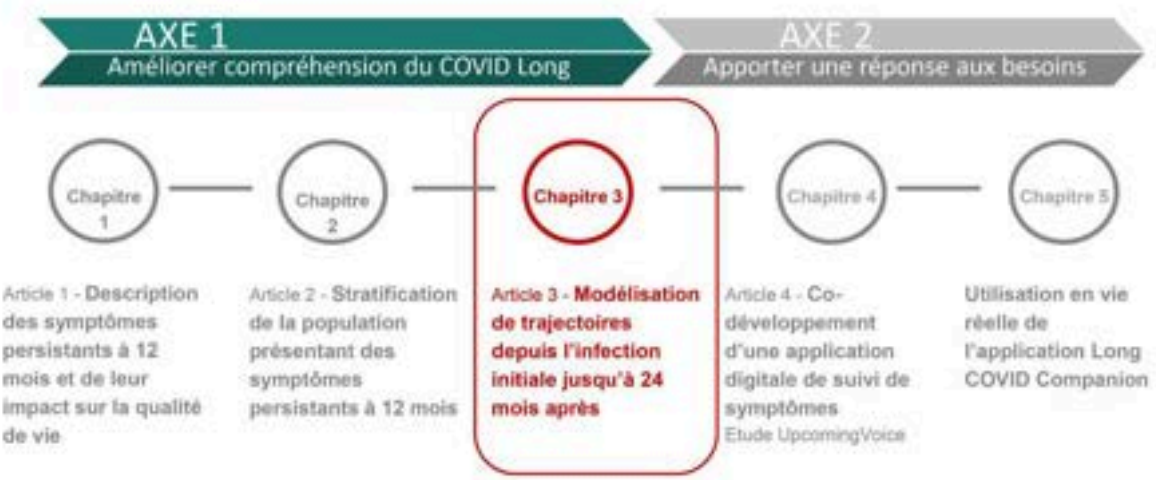
Acknowledgments: In this section, you can acknowledge any support given which is not covered by the author contribution or funding sections. This may include administrative and technical support, or donations in kind (e.g., materials used for experiments).

Conflicts of Interest: The authors declare no conflict of interest.

4.3 Chapitre 3 - Modélisation de trajectoires depuis l'infection initiale jusqu'à 24 mois après

Article 3 - Soumis à BMC Infectious Diseases

Aurélie Fischer, Lu Zhang, Abir Elbéji, Paul Wilmes, Chantal Snoeck, Jérôme Larché, Pauline Oustric, Markus Ollert, Guy Fagherazzi. **Trajectories Of Persisting COVID-19 Symptoms Up To 24 Months After Acute Infection: Findings From The Predi-COVID Cohort Study.** BMC Infectious Diseases



4.3.1 Résumé en français

Introduction - contexte

Le chapitre 2 a montré que les personnes touchées par le CL pouvaient être regroupées en sous-groupes de sévérité croissante et avec des présentations différentes de certains symptômes. Cette étude était une image de la situation des personnes avec un CL 12 mois après l'infection COVID-19 initiale, or il était également important d'obtenir une vue longitudinale de l'évolution des symptômes sur une durée plus longue.

Le chapitre 3 avait donc pour objectif de décrire si les symptômes persistants suite à une infection à SARS-CoV-2 évoluaient selon une trajectoire unique chez toutes les personnes touchées, ou si différentes trajectoires pouvaient être identifiées parmi les participants de la cohorte Predi-COVID entre l'infection initiale et 24 mois plus tard.

Méthodes

Les données des participants inclus dans la cohorte Predi-COVID entre mai 2020 et septembre 2021 et ayant complété le questionnaire à l'inclusion et au moins un questionnaire parmi les questionnaires à 12, 15 ou 24 mois ont été utilisées pour cette analyse.

Sur base d'une liste de 10 symptômes collectés à tous les temps de suivi (fatigue, toux, maux de gorge, diarrhée, douleurs thoraciques, myalgie, essoufflement, conjonctivite, éruption cutanée et fièvre) nous avons créé un score total de symptôme correspondant à la somme des symptômes déclarés par les participants à un temps donné.

Les trajectoires individuelles des 10 symptômes entre l'infection initiale et 24 mois plus tard ont également été décrites.

La méthode de modélisation de trajectoire LCMM a été utilisée pour ces analyses. Les facteurs de risque de faire partie de l'une ou l'autre des trajectoires ont été identifiés par des régressions linéaires univariées. Les caractéristiques individuelles des participants dans les différentes trajectoires ont été décrites.

La qualité de vie des participants 2 ans après l'infection initiale a été décrite en utilisant 5 échelles standardisées (fatigue, sommeil, anxiété, stress et qualité de vie respiratoire) ainsi que des questions ad-hoc.

Principaux résultats et conclusions

Les données de 555 participants ont été utilisées pour modéliser les trajectoires de la variable "score total de symptômes" ainsi que les trajectoires individuelles des 10 symptômes.

Nous avons identifié 2 trajectoires d'évolution de la variable "score total de symptômes". Une première trajectoire T1 "Symptômes modérés, résolution rapide" rassemblant deux tiers des participants et une autre T2 "Symptômes élevés, et hautement persistants" concernant un tiers des participants. Les facteurs de risque d'être dans la trajectoire T2 par rapport à la trajectoire T1 étaient le fait d'être une femme, l'âge, la multi-comorbidité, les comorbidités pré-existantes (hypertension, diabète, obésité), ainsi que les traitements chroniques comme les anti-hypertenseurs, les anti-douleurs, les anti-dépresseurs, les anxiolytiques, les anti-diabétiques et les anti-cholestérolémiants.

Nous avons aussi mis en évidence que les symptômes prédominants dans la trajectoire T2 étaient la fatigue, les douleurs thoraciques et musculaires, ainsi que la dyspnée. Les fréquences de ces symptômes avaient tendance à augmenter entre l'infection initiale et un an après, puis restaient élevées jusqu'à 24 mois. La fatigue était également le symptôme principal dans la trajectoire T1 mais sa fréquence avait tendance à diminuer entre le mois 15 et le mois 24.

En s'intéressant aux trajectoires de symptômes individuels nous avons mis en évidence que la fatigue évoluait selon 4 trajectoires distinctes entre l'infection initiale et 2 ans plus tard.
La qualité de vie de 142 participants ayant complété le questionnaire 24 mois a été étudiée et nous avons mis en évidence que les participants dans la trajectoire T2 "Symptômes élevés et hautement persistants" avaient une qualité de vie fortement impactée, en s'intéressant à 5 indicateurs qui étaient le niveau de fatigue, de stress, d'anxiété, la qualité du sommeil et la qualité de vie respiratoire.
Ces résultats viennent renforcer l'hypothèse que le CL n'est pas unique, que c'est une entité qui affecte les personnes avec des intensités différentes et que son évolution diffère selon les personnes. Ceci apporte un outil supplémentaire aux professionnels de santé pour identifier les personnes les plus à risque de développer une forme sévère de CL pour leur proposer une prise en charge personnalisée.

Mon implication

J'ai participé à la conceptualisation de cette étude, identifié les données de l'étude Predi-COVID à utiliser, effectué le nettoyage des données, et défini la stratégie d'analyse en collaboration avec la plate-forme de bio-informatique du LIH qui a réalisé les analyses. J'ai ensuite mené les discussions sur l'interprétation des résultats, rédigé le manuscrit, et effectué les corrections suite à la revue des pairs.

4.3.2 Article

Abstract

Introduction: Long COVID is a multisystemic, fluctuating condition inducing a high burden on affected people. Despite the existence of some guidelines, its management remains complicated. We aimed to demonstrate that Long COVID evolution follows different trajectories from the initial infection until 24 months after and to identify the determinants of these trajectories.

Methods: Study participants from the Predi-COVID cohort included between May 2020 and September 2021 were digitally followed from their acute COVID-19 infection until a maximum of 24 months. Data from 10 common symptoms were collected at study inclusion, and months 12, 15, and 24 and used to create a total symptom score. Impact of symptoms on quality of life (Sleep, respiratory quality of life, anxiety, stress, and fatigue) was assessed at month 24 using standardized questionnaires and ad-hoc questions. Latent classes mixed models were used to identify total score symptom trajectories and individual symptoms trajectories.

Results: We included 555 participants with at least 2 different time points available during follow-up. We identified 2 trajectories: T1 “Mild symptoms, fast resolution” (N=376; 67.7%), and T2 “Elevated and persisting symptoms” (N=179; 32.3%). Symptom severity was worse in T2 than in T1 at 24 months (high fatigue level: 64.8% vs 19.5%, altered respiratory quality of life: 42.6% vs 4.6%, anxiety: 24.1% vs 4.6%, stress: 57.4% vs 35.6%, and bad sleep: 75.9% vs 51.1%). Fatigue and pain-related symptom frequencies in T2 increased between acute infection and month 12, and remained elevated until 24 months. Women, elevated body mass index, diabetes, and chronic medications were associated with T2.

Conclusion: A third of our study population was in the T2 “Elevated and persisting symptoms” trajectory, presenting high symptom frequencies up to 24 months after initial infection, with a significant impact on quality of life. This work underlined the urgent need to better identify individuals most vulnerable to long-term complications to develop tailored interventions for them.

BACKGROUND

Four years since the pandemic started, it has been estimated that more than 65 millions of people are still suffering from long-term sequelae grouped under the term Long COVID or Post COVID which became a major public health issue worldwide[14,212]. It has been estimated that 10-20% of people infected by SARS-CoV-2 develop Long COVID. All age categories are concerned and people with mild acute illness represent a majority of them. The economic impact of Long COVID is also important with a varying number of people with Long COVID that had to stop working, reduce working time, or retire earlier than foreseen, depending on the countries[18,213]. In the US, the annual total cost of Long COVID taking into account the cost of reduced earnings, of medical spendings and of reduced quality of life, has been estimated around $3.7 trillion, representing 17% of the GDP[214].

In the absence of medical treatment, the management of Long COVID primarily involves the incorporation of various strategies that encompass symptom-specific care such as neurocognitive issues, physical rehabilitation for senses like taste and smell, along with dietary and activity adjustments. Pacing stands out as the primary recommendation for managing activities, emphasizing the importance of balancing exertion with rest to prevent worsening of symptoms[215]. Vaccination has been consistently shown by studies to be an effective prevention measure with a decrease of 15 to 75% of Long COVID risk, with an average risk reduction of around 40%[10,28,216].
Despite progressing knowledge about biological mechanisms, epidemiology, clinical manifestation, and risk factors, Long COVID care still faces many challenges and unmet needs[10].

Long COVID has also been shown to be heterogeneous[217], with a wide variety of symptoms[14], and affected people could be classified into different sub-groups of various Long COVID severity[96,218]. Only a few studies reported long-term evolution (up to 24 months or more)[42,44,65] and it is crucial to better understand how and why some people with Long COVID evolve differently over time and could inform physicians to personalize the care of people with Long COVID.

In this study, we hypothesized that COVID-19 persisting symptoms evolved following different trajectories with a differential impact on the quality of life of affected people.
We thus aimed at 1) identifying symptom trajectories from the acute infection until 24 months after, among a cohort of initially COVID-19 positive adults, 2) describing individual characteristics and identifying the main determinants of the trajectories, and 3) assessing multi-dimensions of the quality of life of people in the different trajectories.

METHODS

Population and study design

In this study we analyzed the data from participants in the Predi-COVID study, a prospective cohort study of persons with a PCR-confirmed SARS-CoV-2 infection in Luxembourg. The Predi-COVID study design and analysis plan has been published previously[134]. The study is registered in ClinicalTrials.gov (NCT04380987) and was approved by the National Research Ethics Committee of Luxembourg (study number 202003/07) in April 2020. All participants signed an informed consent before inclusion in the study.

Data were collected longitudinally, from baseline to a maximum of 24 months. Baseline data were collected by phone by an experienced clinical research nurse at study inclusion, which was done in the 5 days after the PCR test result and consisted of individual characteristics and symptoms. Participants were then invited to complete detailed self-reported questionnaires on symptoms and quality of life at months 12, 15 and 24 after inclusion in the study.

In addition to data collection, participants were invited to participate in an optional biological sample collection. For those who agreed, an inclusion visit was organized in the 5 days after the positive PCR test. Nasopharyngeal swabs were collected for viral load measurement.

Study Design

This study is a longitudinal analysis of participants' symptoms and health status from acute infection to a maximum of 24 months after. Participants included between May 1st, 2020 and September 30th, 2021, who provided the baseline data and completed at least one of the M12, M15 or M24 questionnaires were eligible for the present study (N = 555).

Symptoms and quality of life

We used a list of 10 symptoms (fatigue, cough, sore throat, diarrhea, chest pain, myalgia, short breath, conjunctivitis, rash, and fever) at baseline, M12, M15, and M24. This list was elaborated based on the symptoms available at baseline, with the limited level of knowledge available at the pandemic’s start. The addition of symptoms reported by the participants at each time point corresponds to the “total symptom score” variable.

Trajectories modeling

We used latent class mixed modeling (LCMM)[146] to identify and describe distinct trajectories in the evolution of the total symptom score and of individual symptoms from baseline to M24. This method characterizes trajectories in repeated measurements, with the assumption that several underlying subpopulations or latent classes exist. The LCMM does not require the same number of measurements per participant or measurement time points. The time metric used was the time in days from baseline. We first tested different link functions, including linear and splines with different number of nodes and nodes location, to identify the best-fitting model with one class, which had the lowest Bayesian information criterion (BIC). We then estimate the model with selected link function with two to four classes to determine the optimal number of latent trajectories, appraising the entropy of the model. We applied a gridsearch to ensure the convergence of the model. We did not include covariates to predict latent class membership.

Covariates

The following covariates were used as potential determinants of belonging to a given trajectory: age, gender, body mass index (BMI), smoking status (never, former and current smoker), comorbidities (diabetes, asthma, cardiovascular diseases, and hypertension), regular treatments at time of study inclusion, and disease severity at inclusion proxied by the total number of symptoms.

We fit a generalized linear model on each imputed dataset and pooled the models for a single set of estimates following the Rubin's rules to explore the association of a characteristic and the different trajectories. Each characteristic was explored with the adjustment of the other characteristics in the model. Regression coefficients (Beta) with 95% Confidence intervals were estimated.

Descriptive statistics

We described the continuous variables, when the skewness was between -1 and 1, as mean±SD, otherwise, as median[min,max], while the categorical variables as numbers (percentage). To determine the differences of distribution we used the student t-test for normally distributed continuous variables, the Wilcoxon test for non normally distributed continuous variables and the Fisher's exact test for categorical variables.

Missing values

We did not need to impute missing values for the trajectories modeling as we only included participants who responded to the entire dataset of 10 symptoms. However, participants were included in this study if they completed at least 2 out of the 4 timepoints.

We imputed the missing values in the covariates and generated 45 imputed datasets.

We performed all the analysis with the R version 4.3.0[219]. We used lcmm R package for trajectory analysis, the mice R package for missing covariate values imputation, and the ggplot2 R package.

Sensitivity analysis

Impact of missing timepoints on total score trajectories

To assess the impact of missing timepoints on the total score trajectories, we compared the trajectories obtained on data from the 555 participants who completed at least baseline data and one monthly questionnaire with trajectories obtained on 84 participants who completed the 4 timepoints.

Quality of life evaluation

We described the impact of symptoms on quality of life in a subpopulation of 141 participants who completed the M24 questionnaire.

Sleep quality was assessed using the PSQI questionnaire. A categorical variable was generated using the PSQI score. Poor sleep was defined as PSQI total score ≥ 5[173].

The respiratory quality of life was assessed using the VQ11 questionnaire, initially developed for COPD patients. One global score and 3 sub-scores (functional, psychological and relational) were calculated as described elsewhere and categorical variables were generated[174][220]. An altered respiratory quality of life was defined as VQ11 global score ≥ 22, an altered physical autonomy as functional component ≥ 8, an altered psychological quality of life as psychological component ≥ 10 and an altered relational quality of life as relational component ≥ 10.

The stress level was assessed using the Perceived Stress Scale 4 (PSS 4) questionnaire. The final score ranged from 0 to 16, the highest score corresponding to a higher stress level. A PSS4 score of 6 and above was used to identify participants with high levels of stress[137].

The Fatigue Severity Scale (FSS9) which has recently been validated in COVID-19 population was used to measure the fatigue level[138]. The FSS9 score corresponded to the mean of the scores from the 9 items. A high level of fatigue was defined as a total score ≥36.

The Generalized Anxiety Disorder 7-item (GAD7) has been used to grade the level of anxiety. A score above or equal to a cut-off of 10 was considered to identify generalized anxiety disorder[140].

SARS-CoV-2 viral load

We described the viral load at inclusion in a subsample of participants who provided nasopharyngeal swabs at inclusion. Briefly, SARS-CoV-2 viral RNA was extracted from 140 µL of respiratory swab supernatant and quantified using EDX SARS-CoV-2 Standard (BioRad) synthetic RNA transcripts containing five gene targets (E, N, ORF1ab, RdRP and S Genes of SARS-CoV-2) quantified by ddPCR by the manufacturer at $2x10^5$ copies/mL. The limit of detection (LOD) and limit of quantification (LOQ) of the SARS-CoV-2 N gene RT-qPCR were determined using 3-fold dilution series of the standard with 48 to 60 replicates of each dilution. The LOD was the lowest concentration with at least 95% detection rate and the LOQ as the lowest concentration quantifiable with a coefficient of variation below 35%. Both parameters were calculated using curve-fitting methods implemented in the R script developed by Merkes et al. 2019[221]. The LOD was 3.6 viral RNA copies/reaction and LOQ was 16.0 copies/reaction. Three replicates of 6 points 3-fold dilution series of the standard were included in each experiment to quantify SARS-CoV-2 viral RNA in clinical samples. Standard curves were analyzed and outliers were excluded when necessary to reach PCR efficiency ranging from 90-110% and R^2 above 0.98 in agreement with MIQE guidelines[222]. RNA extracts were tested in duplicates and average values were used for downstream analyses.

Samples with Cq values below 40 were considered positive for SARS-CoV-2. When viral RNA concentration exceeded the upper range of the standard curves, RNA extracts were diluted in 10-fold series and retested in duplicates.

When viral RNA concentrations were lower than the LOQ, samples were considered positive but no viral load was calculated. Viral loads were expressed in viral RNA copies/mL of swab supernatant.

RESULTS

Study population characteristics

The study population was composed of 51.5% of women, mean age was 41.6 years (±12.6), and mean BMI was 25.1kg/m^2 [16.7,55.1]. Thirty-two percent of the participants took at least one regular treatment and 6.3% had at least 2 comorbidities prior COVID-19 infection.

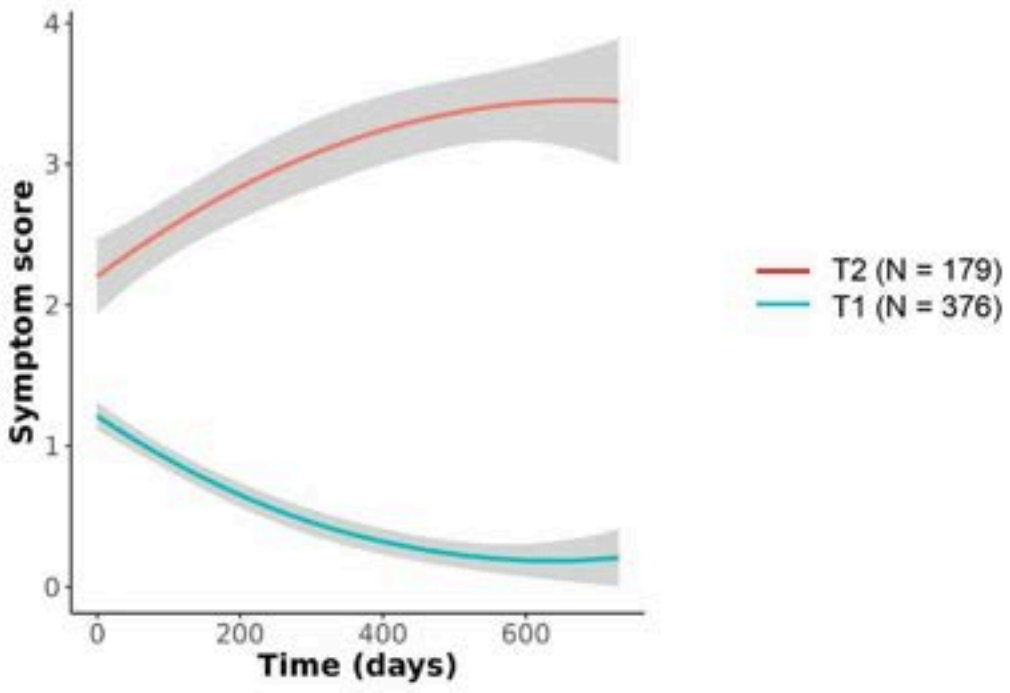
The most frequent treatments were anti-hypertensive (10.4%), antibiotics (10.4%), and anti-cholesterol (7.4%).

Total symptom score trajectories

Based on the lowest BIC and the highest entropy, the optimal number of total score trajectories was identified as 2 (see Supplementary table 1, additional file 1).

The total score trajectories were named according to their characteristics: T1, mild symptoms, fast resolution, and T2, elevated and persisting symptoms. The trajectories are presented in **Figure 1**.

Figure 1: Total symptom score trajectories



Total symptom score evolution in T1 “Mild symptoms, fast resolution”, and T2 “Elevated and persisting symptoms”, from baseline up to 24 months after (in days).

The gray areas show the 95% confidence intervals.

The number of participants in each trajectory was 376/555 (67.7%) in T1 and 179/555 (32.3%) in T2. Participants in the T2 “Elevated and persisting symptom” trajectory were more frequently female (61.5%), had a higher BMI, were older, had more comorbidities, and took more medications than participants in the T1 “Mild symptoms, fast resolution” trajectory.

Participants characteristics in total study population and in each trajectory are summarized in **Table 1**.

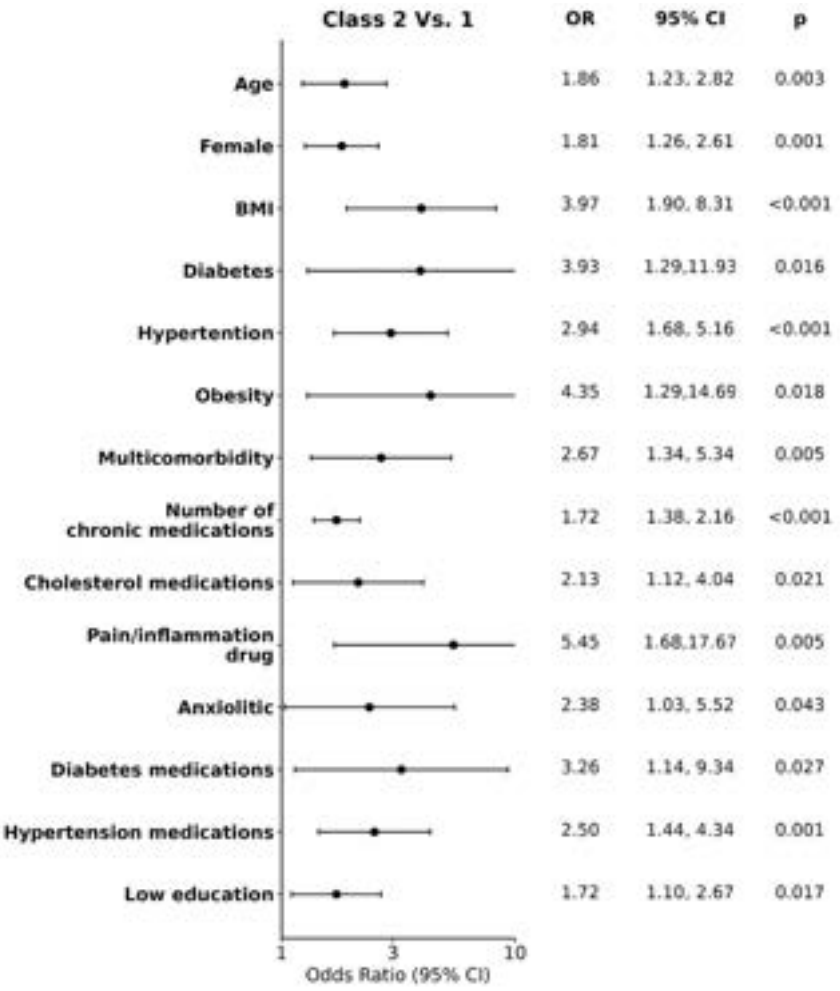
Table 1: Participants individual characteristics at baseline

Variable Mean (±SD) for normally distributed continuous variables ; Median [min;max] for not ormally distributed continuous variables ; N (%) for categorical variables	**Total study population N=555**	**T1: Mild symptoms. rapid resolution N = 376 (67.7%)**	**T2: Elevated and persisting symptoms N=179 (32.3%)**	**P-Value**
BMI (kg/m^2)	25.1[16.7;55.1]	24.7[16.7;55.1]	26.3[18.4;48.9]	<0.001
Female (yes)	286(51.53)	176(46.81)	110(61.45)	0.001
Age (years)	41.6±12.6	40.5±12.4	44±12.7	0.002
At least 2 comorbidities (yes)	35(6.31)	16(4.26)	19(10.61)	0.008
Current smoker (yes)	90(16.25)	56(14.89)	34(19.10)	0.275
Total symptom score initial infection	1[1;8]	1[0;6]	2[0;8]	<0.001
Total symptom score M12	0[0;9]	0[0;4]	3[0;9]	<0.001
Total symptom score M15	1[0;10]	0[0;4]	3[1;10]	<0.001
Total symptom score M24	1[0;9]	0[0;2]	3[1;9]	<0.001
Participants with at least 1 medication before COVID-19 infection	179 (32.3)	99 (26.3)	80 (44.7)	<0.001
Sleep aids	6(1.08)	0(0.00)	6(3.35)	0.001
Anti hypertensive	58(10.45)	28(7.45)	30(16.76)	0.002
Anti pain/inflammation	14(2.53)	4(1.07)	10(5.59)	0.003
Anti cholesterol	41(7.40)	21(5.60)	20(11.17)	0.024
Diabetes treatment	15(2.70)	6(1.60)	9(5.03)	0.026
Treatment for anxiety	23(4.14)	11(2.93)	12(6.70)	0.042
Antibiotic in the 2 months before COVID-19 initial infection	57(10.36)	33(8.85)	24(13.56)	0.100
Anti coagulant	18(3.25)	9(2.40)	9(5.03)	0.125
Anti depressant	22(3.96)	12(3.19)	10(5.59)	0.243
Anti convulsivant	6(1.08)	3(0.80)	3(1.68)	0.393

P-values between T2 and T1 were calculated using the student t-test for normally distributed continuous variables. the Wilcoxon test for non normally distributed continuous variables and the Fisher's exact test for categorical variables

The main determinants of experiencing a T2 "Elevated and persisting symptoms" trajectory were older age, being a female, higher BMI, multi comorbidities, diabetes, hypertension, the number and type of chronic medications (for pain, diabetes in particular) (see **Figure 2**).

Figure 2: Determinants of being in T2 (Elevated and persisting symptoms) versus T1 (Mild symptoms, fast resolution)



When exploring symptom frequencies at each time point in the 2 trajectories we observed that fatigue, cough and fever were the most frequent symptoms at baseline in the “Mild symptoms, fast resolution” trajectory, and in the “Elevated and persisting symptoms” trajectory. Frequencies decreased in T1 from baseline until M24 whereas in T2 fatigue frequency increased between baseline and M12 and remained elevated until M24. Pain-related symptoms (chest pain, myalgia, and short breath) frequencies also increased between baseline and M12 and remained elevated until M24. Symptom frequencies in both trajectories are shown in **Figure 3**.

Figure 3: Symptom frequencies in T1 and T2 trajectories

T1: Mild symptoms, fast resolution

Symptom	Baseline	M12	M15	M24
Fatigue	34.6	14.7	20.5	7.5
Cough	26.1	0.5	2.6	0.0
Sore Throat	9.8	0.8	2.6	0.0
Diarrhea	3.7	0.5	1.3	1.1
Chest Pain	1.9	1.3	3.8	1.1
Myalgia	9.8	4.0	5.1	2.2
Short Breath	5.1	6.2	7.7	3.2
Conjunctivitis	2.9	1.9	2.6	2.2
Rash	1.3	2.1	2.6	1.1
Fever	26.3	0.0	0.0	0.0

T2: Elevated and persisting symptoms

Symptom	Baseline	M12	M15	M24
Fatigue	53.6	82.8	76.2	84.4
Cough	38.0	17.8	19.0	29.7
Sore Throat	22.3	12.1	11.9	18.8
Diarrhea	10.6	15.5	19.0	12.5
Chest Pain	11.2	43.1	38.1	34.4
Myalgia	25.1	52.3	50.0	59.4
Short Breath	12.3	61.5	61.9	54.7
Conjunctivitis	5.6	21.8	26.2	32.8
Rash	1.7	11.5	16.7	15.6
Fever	39.7	2.3	2.4	3.1

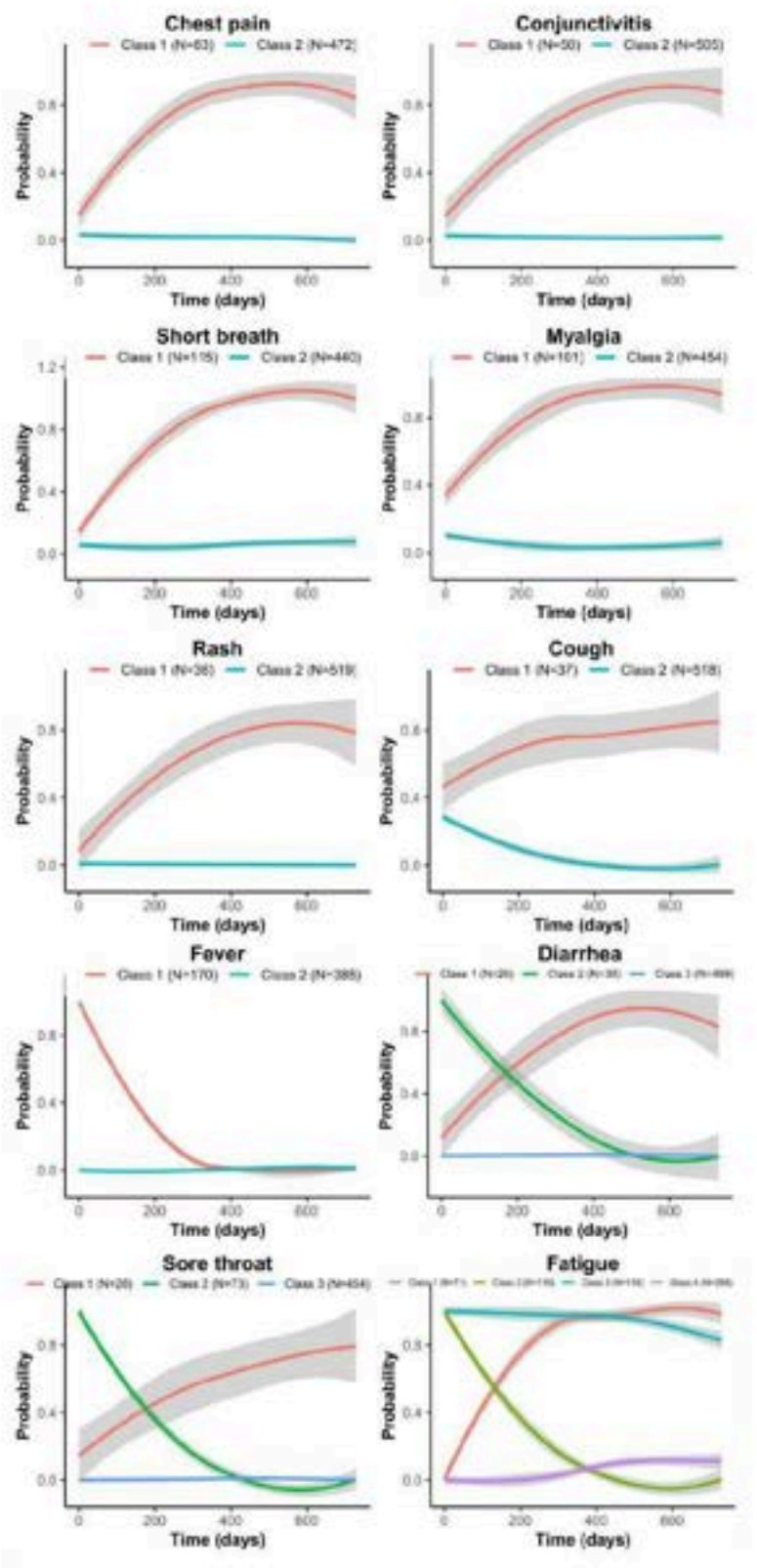
Symptom frequencies are provided for each trajectory at baseline, M12, M15, and M24 (%).

Individual symptom trajectories

Individual symptom trajectories from baseline up to M24 were also identified and are summarized in **Figure 4**. Briefly, some symptoms evolved following 2 trajectories, one trajectory remaining at a low level and the other one increasing over time (chest pain, conjunctivitis, short breath, myalgia, rash and cough). Diarrhea and sore throat evolved following 3 trajectories, one being low, one increasing and one decreasing. Fever and fatigue had particular patterns of evolution. Fever followed 2 trajectories, one including participants with low level and the other one with fever decreasing in a fast way after baseline.

Fatigue was the most complex symptom in terms of individual trajectories as we identified 4 different trajectories: one with half of the participants experiencing low level of fatigue, but with a slight increase over time, the second trajectory with initial low level of fatigue but increasing and remaining at a high level until M24, the third one with initial high level of fatigue but decreasing rapidly over time, and the last one with fatigue being highly present from baseline until M24. Individual characteristics of participants in the 4 fatigue trajectories are provided in supplementary table 2 (see additional file 2).

Figure 4: Individual symptom trajectories from baseline until month 24



Individual symptom trajectories were modeled for the 555 participants from baseline until month 24 (in days).

Sensitivity analysis

The trajectories obtained on 84 participants with complete data at each timepoint were similar to those obtained on the population of 555 participants described above (See supplementary figure 1, additional file 3).

We also described the quality of life of 138 participants who completed the month 24 questionnaire, in the total population and in the 2 trajectories. In brief, participants in the T2 “Elevated and persisting symptoms” trajectory had higher stress, fatigue and anxiety levels, and were more likely to experience poor sleep quality and poor respiratory quality of life than participants in the T1 “Mild symptoms, fast resolution” trajectory. They also less frequently recovered a similar life rhythm and professional activity as before COVID-19 infection, and they were more likely to experience a worsening of their relationships with their family or friends (see **Table 2**).

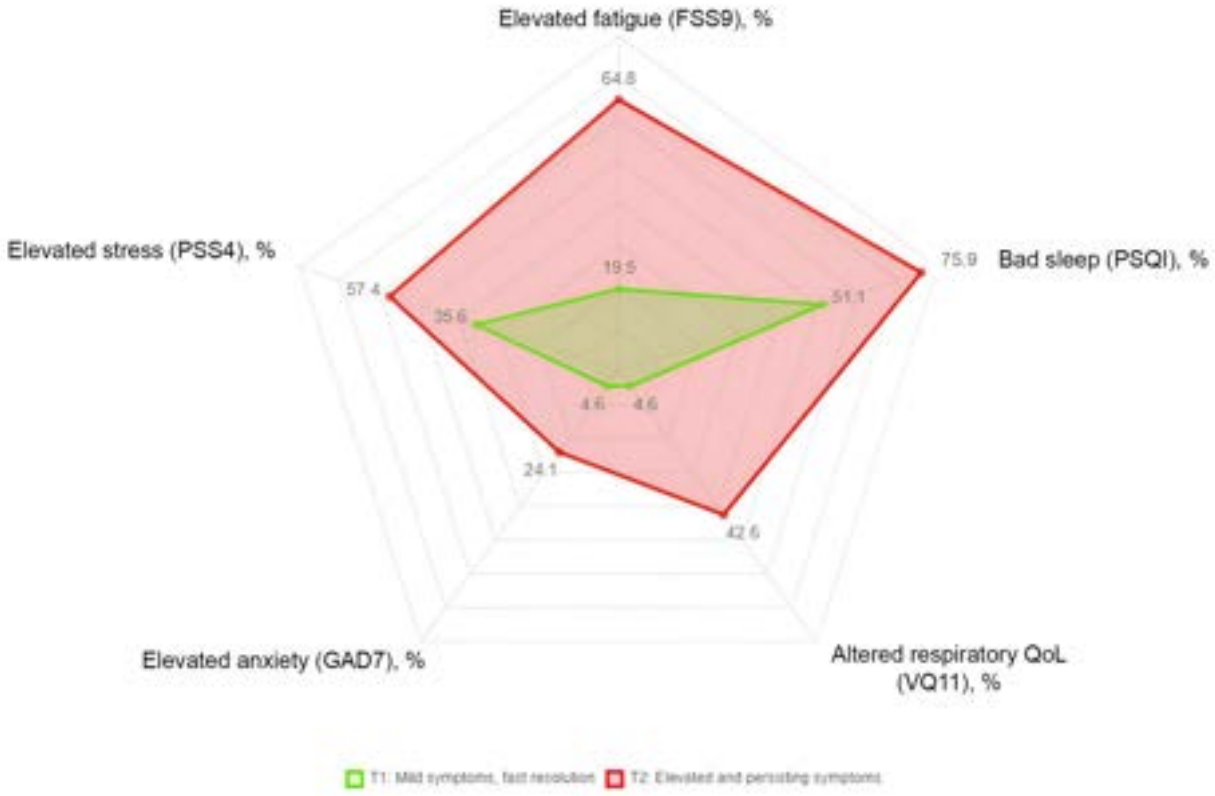
Table 2: Quality of life 24 months after initial COVID-19 infection, in a subpopulation of 141 participants who completed the M24 questionnaire

Variable	Total study population (N=141)	T1: Mild symptoms, fast resolution (N = 87)	T2: Elevated and persisting symptoms (N=54)	p value (T1 vs T2)
Stress level (PSS4 score)	5.3±3.4	4.5±3.2	6.6±3.4	0.001
Participants with high stress level N(%)	62(43.97)	31(35.63)	31(57.41)	0.015
Fatigue level (FSS9 score)	29.1±16.4	22.1±13.3	40.3±14.8	<0.001
Participants with high fatigue level, N(%)	52(36.88)	17(19.54)	35(64.81)	<0.001
Anxiety level (GAD7 score)	4[0,21]	1[0,21]	6.5[0,21]	<0.001
Participants with high anxiety level, N(%)	17(12.06)	4(4.60)	13(24.07)	0.001
Sleep (PSQI)	5[0,20]	5[0,14]	8[2,20]	<0.001
Poor sleep, N(%)	86(60.56)	45(51.14)	41(75.93)	0.004
Respiratory quality of life (VQ11)	14[11,47]	12[11,34]	19.5[11,47]	<0.001
Altered respiratory quality of life, N(%)	27(19.15)	4(4.60)	23(42.59)	<0.001
Life rhythm recovered as before COVID-19 (Yes)	132(88.00)	88(97.78)	44(73.33)	<0.001
Professional activity unrecovered, N(%)	6(4.00)	1(1.11)	5(8.33)	0.038
Relationship with family or friends worsened, N(%)	10(6.67)	1(1.11)	9(15.00)	0.001
No symptoms anymore, N(%)	75(50.00)	69(76.67)	6(10.00)	<0.001
Symptoms occuring under the form of crisis, N(%)	40(26.67)	10(11.11)	30(50.00)	<0.001
Symptoms being constants, N(%)	35(23.33)	11(12.22)	24(40.00)	<0.001

P-values are determined using the student t-test or Wilcoxon test for continuous variables and the Fisher's exact test for categorical variables.

The percentage of participants above the cut-off in each of the PSS4, FSS9, GAD7, PSQI and VQ11 scales is summarized in **Figure 5** and shows a degradation of these 5 indicators in participants from the “Elevated and persisting symptoms” trajectory.

Figure 5: Participants with high fatigue, anxiety, and stress, and poor sleep and respiratory quality of life (%)



Radar diagram showing the percentage of participants with high levels of fatigue, stress, anxiety and with poor sleep and respiratory quality of life in each trajectory using the specific cut-off score of each scale.

The viral load was measured in nasopharyngeal swabs from 172 participants, collected during study inclusion visit. Among them, 145 (84.3%) had a detectable viral load, and 129 (75%) had a measurable viral load. Viral RNA levels were below LoQ cut-off for 16 participants (9.3%) and no viral load could be calculated.

The median viral load at baseline was 1.2E6 [1.4E3,1.8E9] RNA copies/ml in the total participants, and was higher in T2 than in T1 (2.6E6 [1.5E3,1.8E9] and 9.3E5[1.4E3,1.3E9] RNA copies/ml respectively ; p=0.139).

DISCUSSION

In this study we described the evolution of a score based on 10 COVID-19-related symptoms, from the initial infection up to 24 months after. We have observed two trajectories, with one third of our study participants experiencing an “Elevated and persisting symptoms” trajectory, with some symptoms having increasing frequencies until month 24, and having their quality of life heavily impacted. Fatigue was the most frequent symptom in both total score trajectories and we identified 4 trajectories of fatigue taken individually.

Comparison with literature

Although an increasing number of studies describe Long Covid prevalence, subphenotypes and related symptoms at 12 or 24 months[22,96,223,224], few of them aimed at modeling the long-term trajectories of Long COVID evolution[42,65]. Our results are in coherence with these studies which showed also that a subpopulation of people with Long COVID experienced very long lasting symptoms with few recovery over time. Other studies focused on trajectories from specific symptoms like neurological or respiratory symptoms[43,56,64].

We found that fatigue was predominant in both trajectories. Its frequency increased over time in the T2 “Elevated and persisting symptoms”, whereas in the T1 “Mild symptoms, fast resolution” trajectory it remained on a higher level than other symptoms until M15 and decreased at M24. Looking at fatigue independently from other symptoms we identified 4 different trajectories, with 34% of our participants with either a high and persisting level of fatigue from the acute infection until 24 months after, or an initial low level of fatigue importantly increasing until month 12 and reaching a maximum between month12 and month 24. This tendency of fatigue persistence has been recently described in a recent meta-analysis on the neurological symptoms of Long COVID at 12 months[43] and another study also described a worsening of fatigue over time[225].

Being a woman and of higher age were risk factors to experience the T2 persisting Long COVID trajectory. We also showed that preexisting comorbidities like diabetes, obesity and hypertension, and associated treatments, but also treatments for pain, inflammation and anxiolytics, were associated with a higher risk of developing a severe form of Long COVID. These findings are in line with results previously described[42,226].

There are few studies describing the quality of life of people with Long COVID, and they generally focus on overall quality of life using questionnaires like SF12, EQ-5D-3L, or EQ-5D-5L[94,227] or on only one specific aspect like fatigue[97]. A recent study described the quality of life of people with Long COVID at a median time of 197.5 days after initial infection using various scales (including GAD7, PHQ9, MOCA) and showed subpopulations with a higher impact on quality of life[96]. Our study is providing additional information on the multiple aspects of quality of life that are impacted by Long COVID 24 months after acute infection. We showed that being in the T2 "Elevated and persisting symptoms" was associated with a multidimensional alteration of quality of life (altered sleep and respiratory quality of life, increase of fatigue, stress and anxiety).
The impaired respiratory quality of life observed at month 24 in people belonging to the T2 highly persisting trajectory could be explained by a limited recovery in lung function 2 years after initial infection[64].

Participants in the T2 persisting trajectory had a higher SARS-CoV-2 viral load during acute infection, even though this result was not statistically significant due to the low number of data available. Previously, some studies found no relation between viral load and early COVID-19 clinical outcomes [228,229], however another study suggested a correlation between higher viral load during acute infection and Long COVID[230]. It would be of interest to deeper investigate this finding as it may provide new insight on Long COVID determinants and biological mechanisms.

Strengths and limitations

Our study has several strengths. First, all study participants had an initial PCR-confirmed COVID-19 infection and were prospectively followed up to 24 months after acute SARS-Cov 2 infection. Trajectories have been modeled based on 10 symptoms collected systematically at each timepoint from day 0 to month 24. Finally, study participants were in majority non hospitalized individuals, enhancing the result's generalizability since the majority of people with Long COVID undergo mild infections.

This study also has some limitations. The high number of participants who did not complete the questionnaire at months 15 and 24 might have led to an overestimation of Long COVID symptoms at 24 months, as people who completed the questionnaire were experiencing more symptoms than participants who completed only the questionnaire at 12 months. However, our sensitivity analysis on participants who completed the full set of questionnaires showed similar trajectories, confirming the reliability of our results.

In addition, symptoms were self-reported, and we could not fully assert that reported symptoms were linked to Long COVID and we could not exclude that other concomitant health issues could have interfered.

CONCLUSIONS

Our findings demonstrated a high diversity in the long-term evolution of Long COVID. One-third of study participants are still suffering from symptoms 24 months after the acute illness with a significant impact on various dimensions of their quality of life. This work underlined the need to identify the individuals most vulnerable to long-term sequelae to develop tailored care interventions.

Ethics approval and consent to participate: The study is registered in ClinicalTrials.gov (NCT04380987) and was approved by the National Research Ethics Committee of Luxembourg (study number 202003/07) in April 2020. All participants signed an informed consent before inclusion in the study.

Consent for publication: Not applicable

Availability of data and materials: The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interest: The authors declare that they have no competing interests.

Funding: This work was supported by the Luxembourg Government through the CoVaLux Programme and the Luxembourg Institute of Health. The Predi-COVID study was supported by the Luxembourg National Research Fund (FNR) (Predi-COVID, grant number 14716273), the André Losch Foundation and by European Regional Development Fund (FEDER, convention 2018-04-026-21).

Author Contributions: Fischer and Fagherazzi had full access to study data and took responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Fischer, Fagherazzi, Zhang.

Acquisition, analysis, or interpretation of data: All authors.

Drafting of the manuscript: Fischer, Fagherazzi, Zhang.

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: Zhang, Fischer.

Obtained funding: Fagherazzi, Ollert, Wilmes.

Administrative, technical, or material support: Fischer.

Acknowledgments: We are thankful to all the participants of the Predi-COVID study. We also acknowledge the involvement of the interdisciplinary and inter-institutional study team that contributed to Predi-COVID. The full list of the Predi-COVID team can be found here: https://www.lih.lu/en/predi-covid-project-team/.

5. Résultats de l'Axe 2 : Développement d'une application de suivi de symptômes persistants liés à COVID-19 basée sur la voix

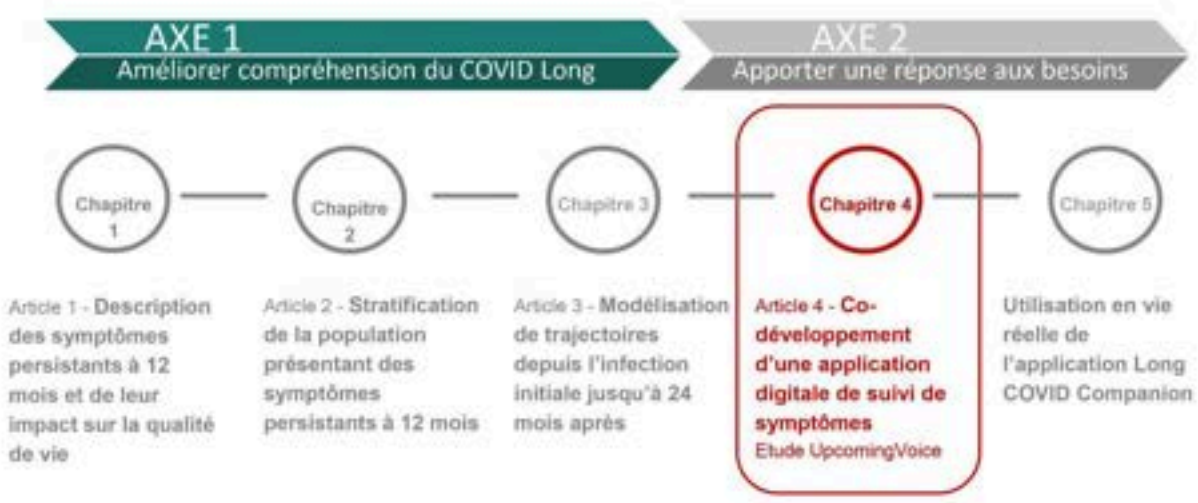
5.1 Chapitre 4 - Co-développement d'une application digitale de suivi de symptômes. Etude UpcomingVoice

Article 4 - **Publié**

Aurélie Fischer, Gloria Aguayo, India Pinker, Pauline Oustric, Jérôme Larché, Charles Benoy, and Guy Fagherazzi. **Co-design of the Long COVID Companion voice-based app to monitor Long COVID symptoms with its end-users: A mixed-method study.** Digital Health. 2024 DOI : 10.1177/20552076241272671

Le protocole de l'étude UpcomingVoice a été publié dans JMIR Research protocols (article complet en Annexe 6).

Aurélie Fischer, Gloria Aguayo, Pauline Oustric, Laurent Morin, Charles Benoy, and Guy Fagherazzi. Co-design of a voice-based digital health solution to monitor persisting symptoms related to COVID-19: protocol for the UpcomingVoice mixed-methods study. JMIR Res Prot. 2023. DOI : 10.2196/46103 (**Publié**)



5.1.1 Résumé en français

Introduction - contexte

L'ensemble des résultats de l'axe 1 souligne le fait que le CL n'est pas une entité unique mais qu'il peut se subdiviser en différents phénotypes de sévérité variable et que les caractéristiques individuelles des personnes influençaient l'appartenance à l'un ou l'autre de ses sous-phénotypes. Nous avons également identifié les symptômes les plus fréquents, et décrit qu'ils se présentaient souvent par groupe de symptômes. Enfin, nous avons montré que le CL impactait négativement différents volets de la qualité de vie 12 mois et même 24 mois après l'infection initiale.

Bien que les connaissances sur cette maladie aient significativement augmenté dans les 4 années suivant le début de la pandémie, il reste de nombreuses zones mal connues, les possibilités de traitements sont limitées et la prise en charge est souvent multidisciplinaire et difficile à mettre en place. Ceci induit une charge importante pour les patients et souligne le besoin de méthodes innovantes pouvant aider au suivi de la santé des personnes avec un CL et limiter leurs déplacements.

Dans ce cadre, et avec les données préliminaires dont nous disposions sur l'intérêt de la voix et des biomarqueurs vocaux, il était justifié de se demander dans quelle mesure nous pourrions apporter une solution concrète pour aider les personnes touchées par le CL à gérer leur maladie au quotidien et comment la voix pourrait être utile dans ce contexte. Nous avons souligné également la nécessité d'impliquer les futurs utilisateurs potentiels dans un processus de co-construction pour s'assurer de la pertinence clinique et répondre à leurs besoins.

L'étude UpcomingVoice a donc pour objectif de co-développer une solution digitale de santé pour le suivi de symptômes du CL avec ses futurs utilisateurs potentiels que sont les personnes touchées par le CL et avec des professionnels de santé en charge de patients CL, et d'évaluer dans quelle mesure la voix et les biomarqueurs vocaux pourraient être une valeur ajoutée à ce type de solution digitale.

Méthodes

Cette étude est basée sur des méthodes mixtes quantitatives et qualitatives, selon un design séquentiel exploratoire. Le protocole de l'étude a été publié dans JMIR Research Protocols[231]. Des personnes touchées par le CL et des professionnels de santé en charge de patients CL ont été impliqués à chaque étape de cette étude d'une part comme co-chercheurs et d'autre part comme participants à la recherche.

La partie quantitative de l'étude a consisté en une enquête anonyme en ligne et visait à définir les principaux aspects de la vie quotidienne les plus impactés par le Long COVID, les besoins des personnes touchées par le CL, à évaluer l'acceptabilité et les attentes vis-à-vis de l'utilisation de la voix dans une application digitale de santé, ainsi qu'à définir les grandes lignes de l'application digitale.

La phase qualitative de l'étude comprenait des entretiens individuels semi-structurés et un groupe de discussion. Les entretiens individuels semi-structurés visaient à approfondir la définition des problèmes, des besoins, des craintes et des attentes des personnes vivant avec le CL, ainsi que les caractéristiques de l'application digitale.

Une première version de l'application, développée sur base des résultats de l'enquête et des entretiens individuels a été présentée à un groupe de patients et de professionnels de santé volontaires, pour recueillir leurs retours et suggestions d'amélioration.

Principaux résultats et conclusions

Cette étude nous a permis, tout d'abord, d'identifier les besoins et les attentes des personnes avec un CL. Celles-ci étaient principalement d'améliorer la vie quotidienne et la prise en charge médicale des personnes avec un CL.

Ensuite nous avons décrit les préoccupations et les facilitateurs par rapport à l'utilisation d'une telle application en les regroupant sous quatre thèmes principaux : engagement de l'utilisateur, contenu de l'application, accessibilité et confiance de l'utilisateur.

Enfin, nous avons défini les principales fonctionnalités attendues, qui peuvent être résumées en 1/ un module de suivi de symptômes basé sur une auto-évaluation, des questionnaires standardisés ciblant certains symptômes et des candidats BV, 2/ un journal permettant de consigner les événements médicaux et de la vie courante importants à corréler avec les symptômes, 3/ un module proposant des informations sur le CL, des liens vers des associations de patients et des réseaux de prise en charge spécialisée, et 4/ la possibilité de générer un rapport PDF reprenant toutes les informations afin de servir de support aux discussions entre patients et professionnels de santé.

Nous avons ainsi co-développé, avec des patients et des professionnels de santé, une application intégrant la voix pour suivre à distance des symptômes du Long COVID, en nous basant sur les attentes et les besoins des personnes touchées par le CL. Cette solution de santé digitale comble une lacune majeure dans la prise en charge et apporte un soutien quotidien aux personnes touchées par le CL. Nous avons également démontré que ces personnes avaient beaucoup d'attentes par rapport aux BV et espéraient une intégration future de BV dans l'application. De futures études seront nécessaires pour évaluer l'acceptabilité, l'utilisabilité et l'efficacité de l'application Long COVID Companion pour les soins et la gestion de la maladie.

Mon implication

Pour cette étude j'ai réalisé la conception de l'étude, les démarches réglementaires nécessaires, et élaboré différentes stratégies pour atteindre les objectifs de recrutement : mise en place de collaborations avec l'association #ApresJ20 COVID Long France et avec le réseau de consultations spécialisées CL au Luxembourg, réalisation de matériel de communication (vidéo et affiches de présentation de l'étude), de campagnes de communication et présentations de l'étude et de mes travaux à plusieurs médecins spécialistes du CL.
J'ai également réalisé toute la collecte de données avec le développement de l'enquête dans l'outil Redcap et la réalisation des entretiens individuels et du focus groupe. L'analyse et l'interprétation de ces données a été faite en binôme avec ma collègue le Dr Gloria Aguayo. J'ai coordonné le développement de la première version de l'application en partenariat avec des patients partenaires de l'association #ApresJ20 COVID Long France et le prestataire externe développeur de l'application. Enfin j'ai rédigé l'article et effectué les révisions demandées par les pairs.

5.1.2 Article

Abstract

Background: People living with Long COVID (PWLC), which is still a poorly understood disease, often face major issues accessing proper care and frequently feel abandoned by the healthcare system. PWLC frequently report impaired quality of life because of the medical burden, the variability and intensity of symptoms, and an insecurity towards the future. These particular needs justify the development of innovative, minimally disruptive solutions to facilitate the monitoring of this complex and fluctuating disease. Voice-based interactions and vocal biomarkers are promising digital approaches for such health monitoring.

Methods: Based on a mixed-method approach, this study describes the entire co-design process of Long COVID Companion, a voice-based digital health app to monitor Long COVID symptoms. Potential end-users of the app, both PWLC and healthcare professionals (HCP) were involved to 1) understand the unmet needs and expectations related to Long COVID care and management, 2) to assess the barriers and facilitators regarding a health monitoring app, 3) to define the app characteristics, including future potential use of vocal biomarkers and 4) to develop a first version of the app.

Results: This study revealed high needs and expectations regarding a digital health app to monitor Long COVID symptoms and the readiness to use vocal biomarkers from end-users. The main expectations included improved care and daily life, and major concerns were linked to accessibility and data privacy. Long COVID Companion was developed as a web application and is composed of a health monitoring component that allows auto-evaluation of symptoms, global health, and scoring relevant symptoms and quality of life using standardized questionnaires.

Conclusions: The Long COVID Companion app will address a major gap and provide day-to-day support for PWLC. However, further studies will be needed following its release, to evaluate its acceptability, usability and effectiveness.

Keywords: Long COVID, Digital Health app, mixed methods, vocal biomarkers, remote symptom monitoring

INTRODUCTION

Three years after the start of the COVID-19 pandemic, it has been estimated that a mean of 10-20% of COVID-19 patients will develop Long COVID, which represents at least 65 million people worldwide[232–234]. The WHO definition of Long COVID states: “Post-COVID-19 condition occurs in individuals with a history of probable or confirmed SARS-CoV-2 infection, usually 3 months from the onset of COVID-19 with symptoms that last for at least 2 months and cannot be explained by an alternative diagnosis”[235]. PWLC present complaints such as tachycardia, extreme fatigue, dyspnea, and inability to perform daily physical tasks[159]. More than 200 symptoms have been associated with Long COVID, and multiple organs are affected[14]. Our previous work showed that 59% of COVID-19-infected people from the Predi-COVID cohort study reported 1 or more persisting symptoms after a year. The number of persisting symptoms increased with the initial disease severity, and the quality of life of those participants was notably impacted by sleep disorders (54%) and compromised respiratory function (12.9%). Nonetheless, individuals with an initially asymptomatic or mild form of COVID-19 infection could also be affected[208].

At an individual level, the impact on the daily life of people affected by Long COVID is high, with many people who do not return to the same physical or professional activity level. Impact on society, in general, is also considerable with increased costs related to sick leaves, handicapped workers and long-term condition recognition[236]. In particular, 62.2% of people with Long COVID stopped working, and only 32.5% resumed full-time professional activities[18]. People experiencing long-term consequences of COVID-19 frequently encounter problems in obtaining a Long COVID diagnosis and accessing specialized health services[237–239]. Despite the increase in knowledge on Long COVID and an increasing number of medications under clinical investigation, there is still no evidence-based treatment[49].

Long COVID disease management consists mainly of the integration of different strategies combining treatments for some specific symptoms, such as neurocognitive, physical, taste and smell rehabilitation, and dietary or activity recommendations. Pacing is the main activity management recommendation and is an approach to balancing activities with rest to avoid exacerbation of symptoms[240]. In the absence of validated treatment, people experiencing persisting COVID-19-related symptoms are, therefore, in need of a tool to monitor the progression of their symptoms and to help them with general disease management.

In addition, a panel of experts from the National Institute for Health and Care Excellence (NICE) recommended the development of telemonitoring and encouraged self-management of acute and Long COVID symptoms in a tailored and accessible way for each patient[106]. Upon the availability of new treatment, this tool could also be used as a proxy to measure the improvement of global health and specific symptoms.

Long COVID care, as for other chronic diseases, should align with the concept of minimally disruptive medicine, aiming for a reduced burden on patients' lives while maximizing health outcomes. PWLC frequently have several healthcare professionals in charge of the different aspects of their care, with many appointments and travels to manage. They are also regularly asked to complete long standardized questionnaires or scales to evaluate their symptoms, which generates an avoidable burden on their lives if care is not coordinated. The development of innovative methods to integrate multiple Patient Report Outcomes in a portable, versatile way, to reduce travels for medical care, and overall for remote health's monitoring is therefore of the highest importance.

Voice is a promising candidate increasingly used in mobile health (mHealth) interactions in chronic diseases[241–243]. Even if there is no clear evidence so far, we believe that it may be easier and more inclusive for patients with chronic diseases than completing questionnaires for example. Voice is an easy and cheap medium to collect and can be easily integrated into a device like a smartphone, now widely used by people of different ages and education levels. Voice dictation has also entered into the habits with the development of voice messages and home assistants like Google Home or Alexa[118]. It can also be used for vocal biomarker assessment as new clinical endpoints for relevant symptoms. However, no validated vocal biomarkers are currently available as a standard of care and there is a need to clinically validate vocal biomarker candidates to bring them into clinical and real-life practice. Vocal biomarkers have already been described in different therapeutic areas like Parkinson's disease[244], mental health, cardiovascular diseases or diabetes[118]. Some of them have been successfully integrated in smartphone apps, currently available on the app stores, like the Real Time Voice Analyser app to monitor respiratory wellness[245], or the Sonde Mental Health app, based on a vocal biomarker of mental fitness[123,246].

In Luxembourg, participants from the Predi-COVID hybrid prospective cohort study were invited to perform voice recordings simultaneously as the filled-in online questionnaires regarding their symptoms and health status. To date, almost 6000 voice recordings from more than 500 COVID-19 patients have already been collected in the Predi-COVID study[134]. These voice recordings have been analyzed, and vocal biomarker candidates have already been identified with performances above 80% to detect fatigue, loss of taste and smell, and symptomatic status in COVID-19-infected people[126],[127,247]. However, further progress is required before these vocal biomarkers can be used in clinical practice. They must be validated with new investigations and there is still a need to develop vocal biomarkers for other symptoms such as respiratory problems.

Many symptom monitoring apps exist for different use-cases including those tailored for patients undergoing chemotherapy treatments, people with mental health conditions or those experiencing chronic pain. They represent an additional tool for managing symptoms that could be integrated into the care of patients, as a regular use may alleviate the symptom burden and improve quality of life by limiting travels for medical appointments. In the case of Long COVID, people with persisting symptoms often experience difficulties communicating with healthcare professionals and sometimes do not feel believed. Therefore, a monitoring app could also facilitate the communication between them.

Regarding the existing apps that could meet the needs of people with Long COVID, some already exist, such as “Visible[248]” or “Living With”[249]. However, these apps are only available in the US and in the UK, respectively. Furthermore, the “Living with” app is available only by invitation. This limits their availability for PWLC in Europe. Some other apps designed for other chronic conditions, like myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) Pacing app, are also used by some people with Long COVID to manage their daily energy levels. However, these apps do not fully meet the specific needs of people with Long COVID, and are only available in specific countries, in English, and some of them only on the iOS operating system, which restricts their potential users. Moreover, to our knowledge none of the above-mentioned apps were the subject of a scientific peer-reviewed publication describing their development and none of them are embarking vocal biomarkers or voice in general.

Co-designing mHealth solutions with end-users and healthcare professionals is crucial to ensure adherence and clinical relevance. Various study designs, based on qualitative methods alone or on mixed-methods, blending qualitative and quantitative methods, have already been used to assess the user's needs and develop digital health solutions in different therapeutic areas.[250–253]

The overall objective of our study was to co-design a voiced-based digital health app available to the greatest number, with people affected by Long COVID and healthcare professionals in charge of Long COVID patients.

To achieve this objective the study consisted in 4 sub-objectives: 1) to understand the unmet needs and expectations related to Long COVID care and management, 2) to assess the barriers and facilitators regarding a health monitoring app, 3) to define the app characteristics, including future potential use of vocal biomarkers and 4) to develop a first version of the app.

METHODS

PATIENT AND PUBLIC INVOLVEMENT

Integrating patient and public involvement approaches in research and e-health co-design processes is key to ensure maximal adoption and to make sure the solution responds to patients' needs and priorities[254,255].

The collaboration with patient partners of the association #ApresJ20 Covid Long France[256] was established in previous research[208,218], facilitating their involvement from the beginning of the study. By raising their need for a monitoring tool, they were at the origin of the app development. This app is therefore made by and for people with Long Covid.

More specifically 3 Long COVID patients and 2 HCPs were involved as co-researchers in the study design and at all research stages by reviewing the protocol, the interview guide, the survey questions, and the informed consent forms, to ensure that the objectives were pertinent and that the study participation induced no exceeding burden. They were also involved in the recruitment by disseminating the information on the study via social media and internal communication within the #ApresJ20 association members. Finally they will be involved in the dissemination of study results through the #ApresJ20 association communication channels and by directly informing their patients (for HCPs).

STUDY PARTICIPANTS

Study participants were adult people (women and men) experiencing Long COVID (diagnosed or not, to ensure a good coverage of all PWLC) and healthcare professionals (HCPs) in charge of Long COVID patients. All participants were native French speakers.

To recruit participants, electronic and paper flyers presenting the study were disseminated via social media, Long COVID dedicated consultations, and Long COVID patient associations in Luxembourg and France. Participants from the Predi-COVID cohort study[134] who declared persisting symptoms one year after the acute infection were also invited to participate. The detailed recruitment and enrolment process in the study is detailed in the published study protocol[231].

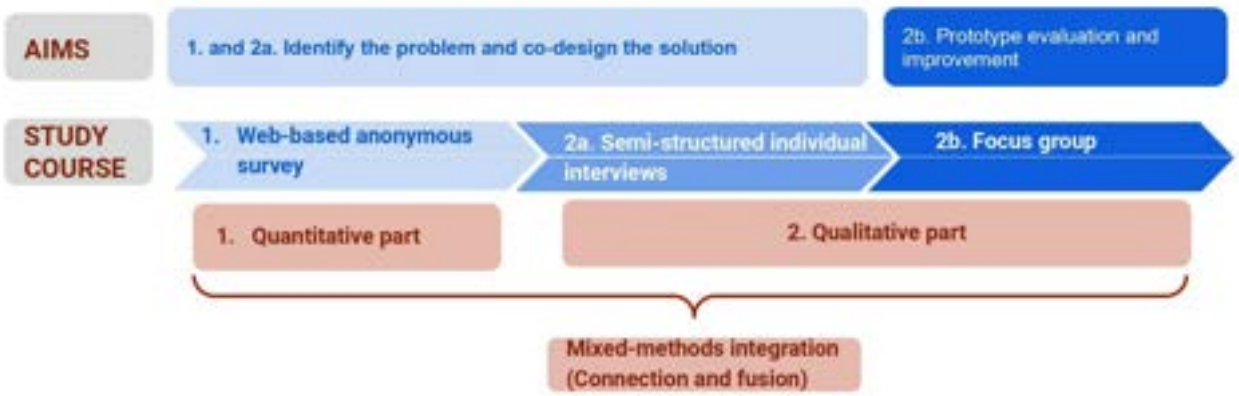
STUDY DESIGN

The UpcomingVoice study is a cross-sectional study based on a mixed-method sequential exploratory design. Quantitative and qualitative data were collected concurrently and analyzed separately. It consisted of 2 successive phases:

1. a quantitative phase based on a web-based anonymous survey (descriptive approach).
2. a qualitative phase based on semi-structured individual interviews (inductive pragmatic approach) and a focus group.

The study design and aims are summarized in **Figure 1**. The study was conducted in French, as participants were coming from French-speaking countries.

Figure 1: Study design



1. Quantitative phase

The quantitative part of the study consisted of an anonymous web-based survey addressed to both PWLC and HCPs. The survey was conducted between October 2022 and April 2023.

This part aimed to define the main aspects of daily life most impacted by Long COVID, the needs of PWLC, to assess acceptability, and expectancies towards the use of voice in a digital health app, and to define the general outlines of the smartphone app.

We used ad-hoc questions covering topics like Long COVID symptoms and impacts, fears and expectancies towards a smartphone app to monitor symptoms alongside the use of voice in such an app, and the 5 main aspects of the User Version of the Mobile Application Rating Scale (uMARS)[257]: Engagement, Functionality, Aesthetics, Information, and Subjective items ("Would you recommend", "Would you be interested in", etc..). We chose to base our questionnaire on these 5 uMARS themes because uMARS is the most used questionnaire for evaluating the quality of mHealth apps, so it was important to take it into account as soon as possible in the app development[258].

The survey questions were adapted depending on the participants' type (PWLC or HCP). Some questions or answer modalities were common, but others were specific to one participant's type. The detailed PWLC and HCP questionnaires were provided in the study protocol[231]. Survey completion took a maximum of 20 minutes to ensure a high acceptability rate, particularly in people with Long COVID who frequently experience fatigue and trouble concentrating. Survey data were collected and managed using REDCap electronic data capture tools hosted at the Luxembourg Institute of Health[259].

2. Qualitative phase

The qualitative phase of the study included semi-structured individual interviews (2a) and a focus group (2b).

2a. The semi-structured individual interviews aimed to go deeper into the definition of PWLCs' problems, needs, fears and expectancies, and the characteristics of the smartphone app. To ensure good interview quality, the following measures were taken. First, the interviewers introduced themselves and reminded the overall study aims and what were the specific aims of individual interviews. Second, participants were reminded that there was no good or false answer and that they could refuse to answer a question they did not want to answer. Finally, PWLC participants were informed to ask at any time to take a break or to stop the interview in case they felt too tired, to limit

the risk of post-exertional malaise. Survey results on related questions were not communicated to participants to not influence their answers and opinions.

The interview guide was developed to assess the study objectives in more depth than the survey. Moreover, an intermediary analysis of survey results was used to identify the most important concerns and to discuss them at the beginning of the interviews to ensure sufficient time to address them correctly. Finally, interviews were organized in the following sections: Understanding Long COVID impact on life, identifying expectations, fears, barriers and facilitators regarding a smartphone app and the use of voice, and defining the app characteristics. Individual interviews were conducted between March and April 2023. Individual interviews had a mean duration of 44 minutes (range: 24-68 min).

2b. Focus group: A first version of the app has been developed based on specifications defined with survey and individual interview results. It was presented to a panel of PWLC and HCPs during a focus group of 90 minutes organized in January 2024. The focus group was structured as follows: After a short reminder of study and focus aims, the moderator made a demo of the app, then each main functionality was individually discussed. Finally, feedback and improvement suggestions were collected during a semi-guided discussion.

The individual interview and the focus group guides are provided in Supplementary file 1. Each main topic was introduced by a leading question. Ad-hoc, open-ended questions were used to encourage the discussion and the emergence of potential new topics. Stimuli questions were prepared and used only in case participants were silent or did not get the topic clearly.

Both parts were conducted by an experienced researcher in epidemiology and infectious diseases with additional training in mixed methods, who had no prior relationship with the participants. Individual interviews and the focus group were done using a web-based teleconferencing system, and where content was audio recorded, transcribed into verbatims and anonymized. Pilot sessions of both the interviews and the focus group were organized to limit the risk of technical issues and to ensure respect for the duration of time. The participants could participate in the survey, individual interview and focus group, but their data were not linked.

3. Population size

Quantitative part: Due to the exploratory nature of the study, a formal size calculation for the survey was not possible. The quantitative data collected through the survey will be solely utilized for descriptive analysis and to assess the significance of various topics for the participants.

Qualitative part: Qualitative research does not explicitly aim to secure a representative sample of individuals, but aims to capture diversity. No gold standard to estimate the correct sample size presently exists[260]. Individual interviews were continued until data saturation was achieved, which can be defined by the point at which no new themes emerge[149,150]. A recent systematic review showed that data saturation was achieved with a mean of 9 to 17 interviews, but with high disparities between studies[150]. We estimated that including at least 15 PWLC and 5 HCP for the individual interviews would allow us to achieve data saturation. We also actively monitored the inclusions for the individual interviews to ensure a balanced representation across various age groups and gender categories, thereby enhancing the diversity within our study population. For the focus group, it was estimated that the ideal sample size was between 6 and 10 participants, particularly when involving persons highly informed on the topic[261]. Participants of the individual interviews who expressed their interest in participating in the focus group were invited to this.

DATA ANALYSIS

Survey

Quantitative data resulting from the web-based survey were analyzed using descriptive statistics. We described the normally distributed continuous variables as mean (min-max), while the categorical variables as numbers (percentage). We used the Student t-test, the one-way analysis of variance (ANOVA) to determine the differences of distribution for continuous variables and Fisher's exact tests for categorical variables. We performed all the analysis using the R software version 4.3.1[262].

Semi-structured individual interviews

The Maxqda software was used for transcription of interviews recordings and for the coding and analysis of transcriptions. The quotes presented in the present study were translated from French to English and are cited using an arbitrary participant number, for example PWLC1 or HCP1.

We applied inductive reflexive thematic analysis to find patterns in the data and define the main topics of interest among the 3 aspects discussed (understanding Long COVID impact, identifying the needs, and defining the app characteristics). Two experienced researchers were involved in the coding process (AF and GA). The first step was the familiarization with the whole data set by the two researchers. The second step was the generation of initial codes separately by AF and GA. Codes were then compared and refined to group them by themes. An iterative process of discussions between the 2 researchers allowed coding structure consolidation.

After 16 interviews with PWLC and 5 with HCPs being realized, transcribed and analyzed, the research team (AF and GA) determined that data saturation was achieved as no new theme or subtheme reached out anymore[149].

Focus group

The Maxqda software was used for the transcription of the focus group audio recording and for analysis of the transcript. The feedback obtained was summarized and grouped by the 5 themes from the uMARS: engagement, functionality, information, aesthetic, and subjective items.

Mixed method integration

Data integration was done through 2 mechanisms[263]:

Connection: survey results were used to elaborate the individual interview guide so that the most important aspects could be discussed in priority during the interviews.

Fusion: data from survey and individual interviews were grouped by themes and compared to determine the areas of convergence, divergence and expansion[155]. 'Convergence' describes a positive alignment between the survey respondents and the interview participants. 'Divergence' describes a disagreement between qualitative and quantitative findings. Expansion indicates that qualitative and quantitative data addressed the same concept but in different and complementary ways.

Ethics

The study was approved by the National Research Ethics Committee of Luxembourg (study number 202208/04) in August 2022. No explicit informed consent was required for the participation in the quantitative part, based on an anonymous survey. An explicit informed consent was electronically signed by all participants of the qualitative phase before participating in the individual interview.

RESULTS

Study population

The survey was completed by 201 PWLC and 15 HCP. PWLC were mostly women (83.1%), with a mean age of 46.9 years (min 18 - max 92), and were in majority of higher education (undergraduate or more). HCP were also mostly women (60%), mean age was 40.5 years (min 27 - max 77) and 93.3% of them were of higher education. 15 PWLC and 5 HCP participated in the individual interviews and among them 6 PWLC and 3 HCP also participated in the focus group.

Survey, interviews and focus group participants characteristics are summarized in **Table 1**.

		Survey		Interviews		Focus group	
Variable		HCP (n=15)	PWLC (n=201)	HCP (n=5)	PWLC (n=16)	HCP (n=3)	PWLC (n=6)
Females, n (%)		9 (60.0%)	167 (83.1%)	2 (40%)	10 (62.5%)	1 (33.3%)	2 (33.3%)
Age, mean (min-max)		40.5 (27-77)	46.9 (18-92)	42.0 (35-54)	51.4 (28-68)	41.3 (35-54)	49 (28-60)
Level of education, n(%)	Lower secondary education	-	27 (13.4%)	-	3 (18.7%)	-	-
	Upper secondary education	-	32 (15.9%)	-	2 (12.5%)	-	1 (16.7%)
	Bachelor's degree	6 (40%)	61 (30%)	1 (20%)	4 (25%)	1 (33.3%)	2 (33.3%)
	Master's degree or	9 (60%)	80(40%)	4 (80%)	7 (43.8%)	2 (66.7%)	3 (50%)

Table 1: Participants' characteristics

PWLC survey participants were mostly not hospitalized during their initial COVID-19 infection (80%). Initial symptoms were mild for 55% and severe for 37% of them. 79% experienced severe persisting symptoms and 91% already obtained a Long COVID diagnosis.

For all survey answers provided below there was no influence of age, gender or level of education for PWLC nor for HCP (data not shown, available upon request).

Diagnose - identify the problem and need

The results of the survey and individual interviews were first grouped in categories: LC impact on life, interest and acceptability regarding a voice-based app, and considerations for a digital health app. For each category we grouped results in theme and sub-themes.

Long COVID impact on life

For this category we identified 2 main themes: daily life impact and disease management.

In relation to the theme "daily life impact", 88% of the survey participants declared that they did not recover the same level of professional and leisure activity as before COVID-19 infection.

Five sub-themes from the "daily life impact theme" were identified from the individual interview transcriptions : feelings, financial issues, work issues, social life, and daily functioning. In relation to the psychological impact, both HCP and PWLC reported that Long COVID was a source of anxiety and stress. As mentioned by HCP 12, "It's a source of psychological difficulties for them, as they find themselves in a completely unexpected state".

Moreover, many PWLC underlined that they often feel alone or misunderstood, either by their close-relatives or by HCPs. “Honestly, I'm lost. I feel completely alone and powerless” (PWLC 57). Financial issues that were mentioned were financial precarity, due to an income reduction, or the cost of medical care. This was closely related to work issues, with half of the PWLC explaining that they had to stop, reduce or adapt their professional lives. Some of them also felt they were at risk of losing their position because of committing many errors due to cognitive problems. The high impact on social life was also often reported by HCP and PWLC, with fewer leisure activities, fewer activities with family and friends, less pleasure in eating or going out to restaurants due to loss of taste or smell or to cognitive problems. Finally, simple tasks from daily functioning were impaired, as PWLC 62 explained: "On a day-to-day basis, I have to choose between taking a shower and taking out the rubbish." Extreme fatigue was the center problem that induced difficulties in daily tasks.

Disease management was an additional important theme related to the impact of Long COVID on lives. 64% of survey participants with Long COVID already benefited from a specialized Long COVID consultation. However, 60% of the survey participants declared that access to care was difficult or very difficult. Three sub-themes could be identified from the individual interviews: Access to care, disease recognition, and rehabilitation. Some inequities in the access to care were reported by PWLC, with people having problems obtaining an appointment with the right specialist, with specialized Long COVID centers being overwhelmed and with people not even knowing that specialized Long COVID centers exist. Disease recognition was also reported as a problem, by PWLCs and HCPs, who mentioned that some HCP don’t know much about Long COVID and some of them even don’t believe their patients and the existence of Long COVID. “At least half the doctors I see don't believe in my symptoms and tell me that it's all in my head" (PWLC 23). The recognition of Long COVID as a long-term disease was also a problem with administrative issues coming in addition to the lack of knowledge on Long COVID from some HCPs. The last theme of concern was rehabilitation. While some rehabilitation exercises and pacing seem to be helpful, some PWLC and some HCP underlined that rehabilitation may not be adapted "You can tell people to do sudoku or memory, but it won't necessarily help them with their shopping list or their job." (HCP 7) or even harmful with a risk of post-exertional malaise "I believed in exercise rehabilitation, and then I had a relapse." (PWLC 43). Rehabilitation also induces many appointments which can be tiring and burdensome.

Interview results were in convergence and expanded upon the survey results, and provided more in-depth information on areas impacted by Long COVID in their lives. Survey and interview themes and verbatims are summarized in Supplementary table 1.

Based on these results, a conceptual framework of key aspects of Long COVID impact on affected people has been elaborated based on these results **(Figure 2**).

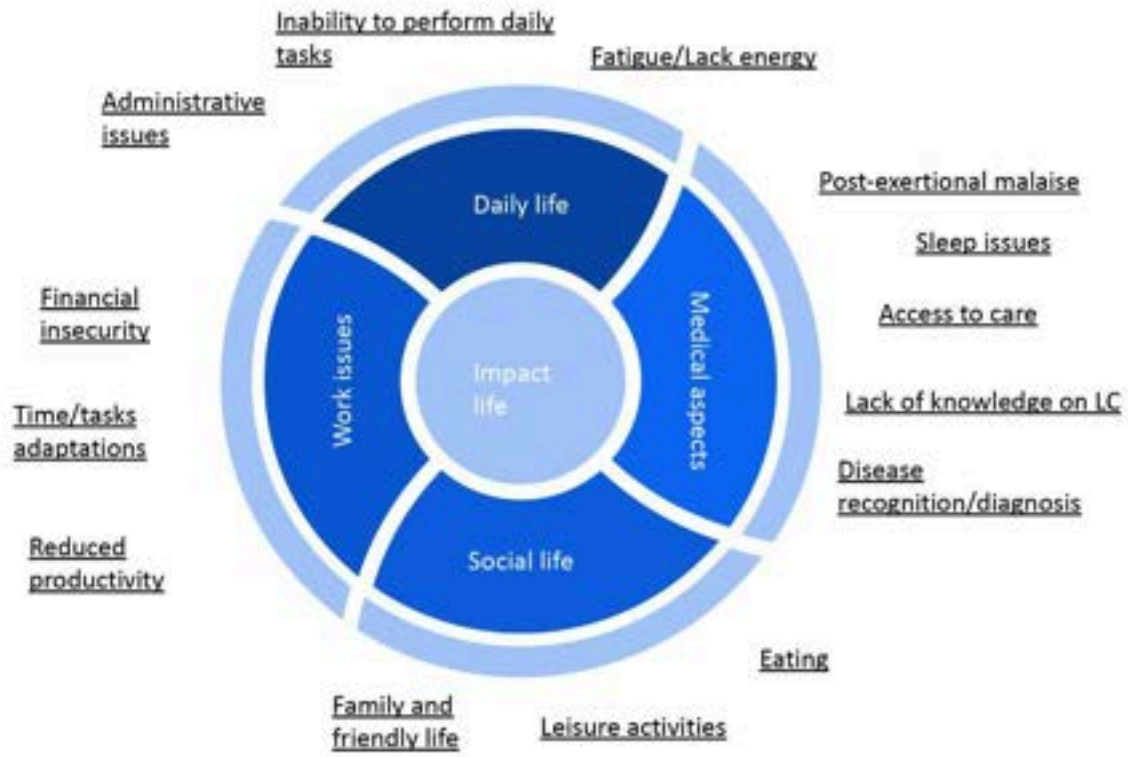


Figure 2: Conceptual framework of key aspects of Long COVID impact on daily life

Interest in and acceptability of a voice-based digital health app

People with Long COVID were 51% to think that a digital health app could be useful to manage their health in the long term, and 43% answered that they did not know.

However, 82% of PWLC declared that they would be interested or very interested in a digital health app based on vocal biomarkers to monitor symptoms.

HCPs were 82% to think that a voice-based monitoring app would be useful for people with Long COVID, regardless if they are already in dedicated Long COVID care or not, and 73% of them would recommend the use of this kind of app.

Finally, 70% of PWLC and 60% HCP agreed that the app should include both voice recordings and questionnaires.

Complete survey answers regarding interest and acceptability of an app based on vocal biomarkers (VB) are provided in Supplementary table 2.

During the interviews, it appeared that all HCPs and PWLC were favorable to the use of voice and vocal biomarkers. Among the advantages that were raised during interviews was the reduction of travels for medical visits: “Using voice is intuitively logical for patients, to limit travels” (HCP 66). The ease-of-use of voice compared to writing was also raised by several participants: “I like the idea of voice, because it could be a way of reducing screen fatigue, rather than writing.” (PWLC 16). “There's an aspect of simplicity and reproducibility that's quite easy for patients, who can do it whenever they want. They can do it at home and not just when they see the doctor.” (HCP 9) “I'd like to have voice monitoring, because I find it less tiring than having to write down all the symptoms.” (PWLC 14)

HCP 37 also positively regarded vocal biomarkers: “It's something that can be more powerful than standardized questionnaires.” However, it was highlighted that it is important to demonstrate that vocal biomarkers are accurate to gain trust from the users, as mentioned by PWLC 19: “What's important here is to gain trust and demonstrate that it can be useful and how it can be useful.”

Another advantage of vocal biomarkers would be the early detection of symptom worsening or improvement. For example, HCP 7 said: “If it can detect maybe a little bit in advance, maybe even a little bit before people realize and that would be really very useful for our patients”, meaning that patients could slow down their activities before the worsening becomes too much disabling. On the other hand, HCP 37 saw it differently and said: “If things get worse, the patient realizes it anyway. A biomarker can be useful precisely to show that things are getting better. That could be interesting and encouraging.”

Finally, an expectation raised from PWLC is that vocal biomarkers could be more objective than an auto-evaluation. “It's something measurable, something more objective, in quotation marks, than the personal assessment we can make of our symptoms.” (PWLC 19) “I just think it's more objective than giving a score to everything over time.” (PWLC 21)

Considerations regarding the digital health app

Expectations

Main expectations identified in the survey were: 1/ a better symptom monitoring (PWLC 71.1%, and HCP 80%) and to obtain an objective measure of symptoms with vocal biomarkers (PWLC 56%, and HCP 53%, 2/ the evaluation of a rehabilitation program (PWLC 48%, and HCP 73%), 3/ to limit the medical visits (PWLC 31%, and HCP 47%, and 4/ to obtain a Long COVID diagnosis (PWLC 43%, and HCP 33%).

During the interviews, two main themes related to expectations were identified: care and daily life. The “Care” theme can be divided in 4 sub-themes: symptom monitoring and identification, improvement of communication between PWLCs and HCPs, improvement of care and access to care, and the limitation of medical visits. This confirmed survey results except for the Long COVID diagnosis, which did not emerge during the interviews as an expectation. Symptom monitoring was the central expectation. It also came out that identification of symptoms was an important aspect for some PWLC who find it difficult to describe their symptoms with the right words. Some PWLC mentioned that they already tried to monitor their symptoms using a paper diary but find it not practical. (“I have a diary to write down my symptoms. So, I write them down sometimes, but it's true that it's not super practical." PWLC 44). This confirmed that a digital health app could be an added value for them. A HCP suggested that symptom monitoring and its visualization may even have a therapeutic interest ("Having a calendar of symptoms like this can be very therapeutic (...) to show the fluctuation of symptoms, to show and visualize the course of certain symptoms" HCP 66). Symptom monitoring over time was also expected to be more objective than the subjective perception that a person can have about his/her health status.

The “Daily life” theme comprised 2 sub-themes: psychological support and administrative support. PWLC participants were a majority to report high levels of stress and anxiety. Some of them said that they expect the digital health app to provide them psychological support in one way or another: "I don't know where or how it can be set up, but here it is, psychological help." (PWLC 55) "Measuring and sending back information in relation to what we have can be very interesting and reassuring too." (PWLC 44). HCPs had a similar opinion, as they were 53% to think that this kind of app could be a companion tool for PWLCs and HCP 7 said during the interview that it could help them "To feel less alone, to feel understood, to not feel abandoned". The “administrative support” theme included administrative information linked to the medical aspect, return to work and patient’s association contacts.

Concerns

Survey results showed that the most frequently reported barriers by PWLCs were: a wrong result interpretation (67.7%), symptom intensity (37.3%), and data protection (36.3%). Among HCPs, the 3 most important barriers were: a wrong result interpretation (66.7%), data protection (53.3%), and age (26.7%). Age was reported as a barrier by only 24% of PWLC and symptom intensity by only 6.7% HCP.

The interviews confirmed the survey findings, with 2 main themes. The first theme was “accessibility” encompassing illiteracy, vision or language problems, and an excessive volume of content that could complexify too much the app use. The second theme was “user confidence” and covered confidentiality issues and the trust in a new technology like vocal biomarkers. The confidentiality issues were either general to the digital health app with participants expressing some concerns with sharing personal and health data ("I'm always a bit suspicious, I prefer to keep the collection of personal and health data to the strict minimum. PWLC62 or "We're perhaps a little afraid, a little suspicious, so I think it's really a question of developing trust, of saying that this is really for the user, … and also to show that RGPD is respected."PWLC 19") or specific to the use of voice that could be a barrier in the app use settings ("Everyone around you hears what you're saying, and I think that can be a bit of a barrier for some people, who will wait until they're alone, until they're well isolated, to use the app, because they have to use their voice."HCP 7). Finally, as vocal biomarkers are a recent technology it was raised as a barrier by some participants who need to be convinced that they can trust it, as mentioned by PWLC 16 "Since I don't know anything about it, I'm wondering about its reliability." and PWLC51 "It remains to be seen whether it really works".

Facilitators

The main facilitators identified in the survey were the app ease-of-use (PWLC 72% ; HCP 73%), the vocal biomarkers results visualization (PWLC 61% ; HCP 47%) and reliability (PWLC 54% ; HCP 80%). During the interviews, 2 themes emerged (user engagement and app content). Facilitators linked to user engagement were adaptability/personalization ("It has to be general, but I'd like to adapt it to my own needs."PWLC 57 ; "The more customized, the better"PWLC 19 ; "Adaptability to each patient is the key word in care. For an application, I think it could be good"HCP 7), an easy-to-use conception ("I'd like something very simple."PWLC 16 ; "In terms of access and interactivity, it has to be simple."PWLC 44), integration of voice dictation ("I do almost all my stuff orally because I'm tired of writing,"PWLC 14 ; "it might help some people who are less at ease with the written word."PWLC 4), and result visualization from vocal biomarkers but also generally speaking about symptom evolution("It could be something really appreciated by patients to have regular feedback with nice graphics."HCP 66 ; "when you see a curve that goes up and down, it's worth a thousand words."HCP 9).

Expectations, facilitators and barriers, with verbatim quotes are summarized in Supplementary table 3.

To summarize these results, concerns and facilitators were grouped in 4 main themes (User engagement, app content, accessibility, and user confidence) and presented in **Figure 3**.

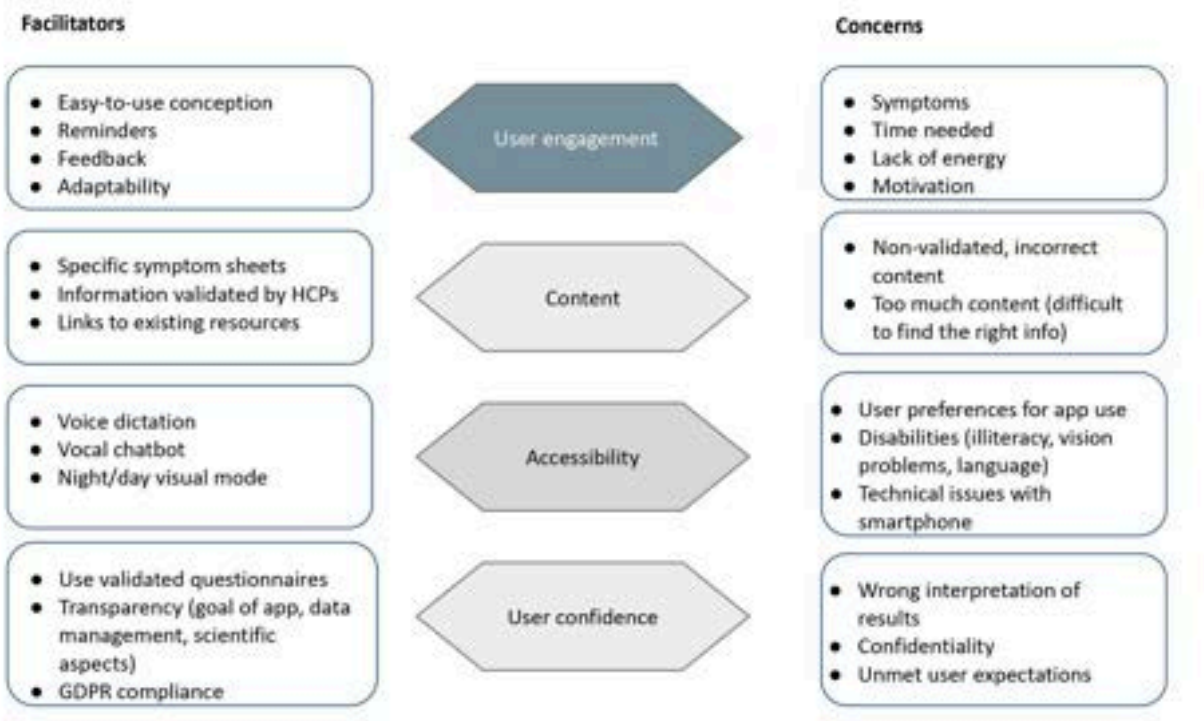


Figure 3: Concerns and facilitators regarding a digital hBlth app

Define solution - App characteristics

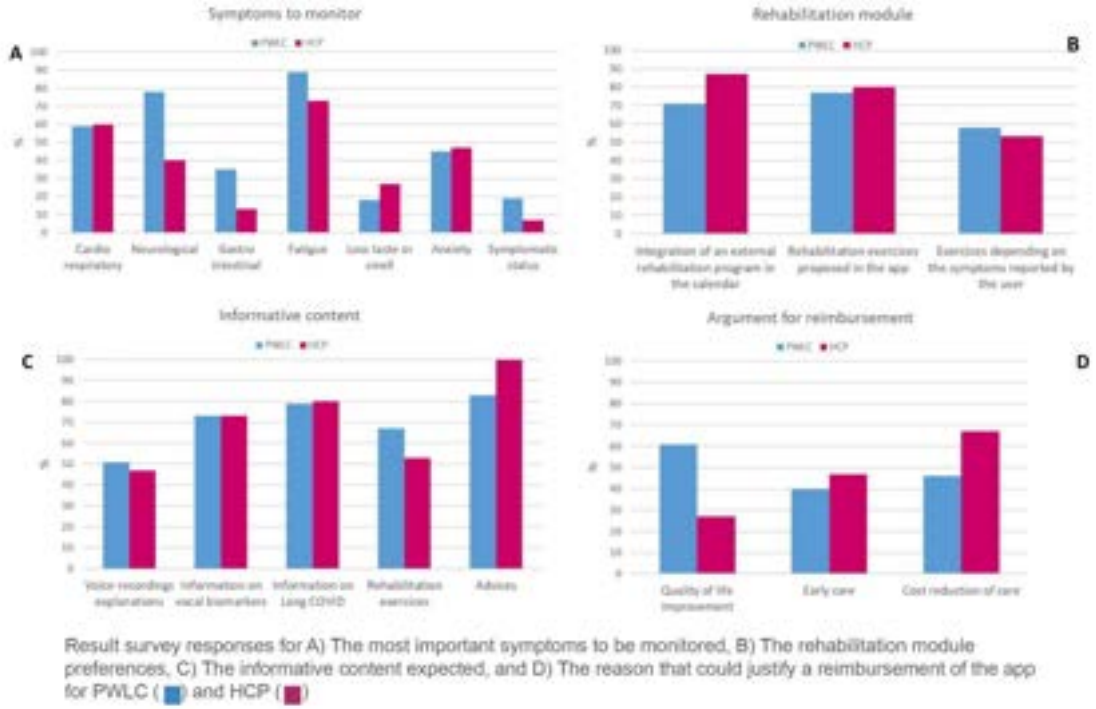
App features

Survey results and interview verbatim quotes were grouped according to the 5 themes of the uMARS: engagement, functionality, aesthetics, information, and generalities.

Regarding the engagement theme, interviews revealed that personalisation options were not essential and that interaction with other app users, namely other PWLC, was rather seen negatively, with the need of a moderator, the presence of potentially toxic people and did not appear to be a real need for PWLC.

App functionalities were extensively discussed with all interview participants. Survey results and individual interview verbatims related to app characteristics are provided in Supplementary table 2 and some detailed survey answers are summarized in **Figure 4**.

Figure 4: Survey findings.



Result survey responses for A) The most important symptoms to be monitored, B) The rehabilitation module preferences, C) The informative content expected, and D) The reason that could justify a reimbursement of the app for PWLC ( ) and HCP ( )

The symptom monitoring was essential and the main symptoms to monitor were cardiorespiratory and neurological symptoms, fatigue and anxiety. It came out that the app should allow the monitoring of several symptoms and should be based on vocal biomarkers and validated questionnaires. Participants also reported that it would be interesting to have a list of symptoms and to be able to indicate the perceived symptom intensity.

The alert system component which seemed to be important for 63% of PWLC and 73% of HCP survey respondents was not consistent among the interview participants. Several participants mentioned that this could be interesting if interactions with HCP were implemented but otherwise, if the user received the alert it could be a source of additional anxiety. Participants suggested an alternative option would be to simply give advice to rest or to consult a doctor.

The integration of rehabilitation exercises in the app was expected by 77% of PWLC and 80% HCP survey respondents. However, when discussing this topic during interviews, it was less clear. Indeed, some participants thought that it could be good to propose breathing, cognitive, or even physical reinforcement exercises ("Suggest, for example, different types of activities, relaxation strategies." PWLC 19 ; “perhaps neurocognitive re-education exercises could be added”HCP9), but several others mentioned that there could be a risk of post-exertional malaise, and that in cases where exercises were proposed they should be accompanied through an extensive amount of detailed explanations and pacing should be explained (“Not pushing yourself too hard, or restricting yourself too much. And that's not easy. So, with the right explanations about post-exertion discomfort and all that, yes”PWLC72 ; "(NB: when discussing rehabilitation exercises): It contributed to me having another Post-Effort Malaise crash, you see? And a big crash."PWLC44).

Among PWLC survey respondents, 76% answered that they would have the option to share their data and results with their family or with HCP. 34% expected to be able to send emails to HCP through the app. Interview participants, either PWLC or HCP, were not that much convinced about the necessity and feasibility of implementing the possibility of a direct interaction between PWLC and HCP through the app. The main need seemed to have a synthetic and graphical view of app data and results to serve as a discussion basis during consultations. Therefore, the option to generate a downloadable PDF report emerged as an efficient and easy alternative.

During the interviews, an additional need emerged, namely the integration of a medical and daily life diary, so that PWLC could record important information like treatments, medical assessments, medical events, physical and cognitive activity or level of stress.

Regarding the informative app content, survey results showed that informative content was expected by 87% of PWLC and 87% of HCP. Advice was expected to be present in the app by 83% of PWLC and 100% HCP. During interviews, these results were confirmed. However, it also came out that the app should not contain too much information and that information and advice should be validated and up to date. Some participants mentioned that internet sites with a lot of information already exist; "You could list a number of sites where people can connect, I don't think the application should become a Long Covid Encyclopedia. " (HCP 9). Personalized advice would be interesting as mentioned by PWLC 4 “Advice on matters that concern me" but many participants agreed that advice on pacing, physical activity, sleep and diet would be interesting for the majority of users.

Finally, 83% of PWLC and 73% of HCP survey respondents thought that the app should be reimbursed, but this point has not been discussed during interviews and did not emerge spontaneously. The main reasons in favor of the reimbursement were quality of life improvement for 61% of PWLC and the reduction of care costs for HCP.

Survey results and individual interview verbatims related to app characteristics are provided in Supplementary table 4.

The main functionalities expected to be present in the app were summarized in **Figure 5**.

Figure 5: Main app functionalities

App development

Based on the app features defined above, a first version of a web-application has been developed with the help of patients at each stage to ensure that patients' priorities, needs and the specificity of the disease were met. The web app format was chosen so that the app will be accessible using a smartphone or a computer.

The app was designed to be user-friendly, with a clear and guided manner to fill in the daily symptoms and health status. Briefly, after account creation, the users arrive at a home page with a message inviting them to complete their daily data.

Daily data include 3 steps: auto-evaluation of global health on a 0 to 10 likert scale, auto-evaluation of symptoms chosen in an extensive list with their perceived intensity on a 0 to 5 scale, and finally a vocal diary allowing users to record the main information and feelings they experience.

Validated questionnaires for some specific symptoms or quality of life are accessible on the home page, as well as resources and the possibility to fill in a medical and daily life diary.

As vocal biomarkers are too early in their development and not validated yet, we decided to integrate them in a module with the option to do four voice recordings for research purposes, available on the home page.

A calendar view allows users to have an overview of their global health status using smileys, and to easily retrieve the days with medical or daily life events recorded.

Users also have the possibility to visualize the evolution of each symptom they recorded and of the questionnaire score in a simple and graphical manner, with the option to see the medical and daily life events so that a correlation could be done.

PDF reports summarizing data from a chosen period of time can be generated and downloaded with all these results to serve as a basis of discussion between PWLC and HCP.

The overall app user journey is presented in **Figure 6**.

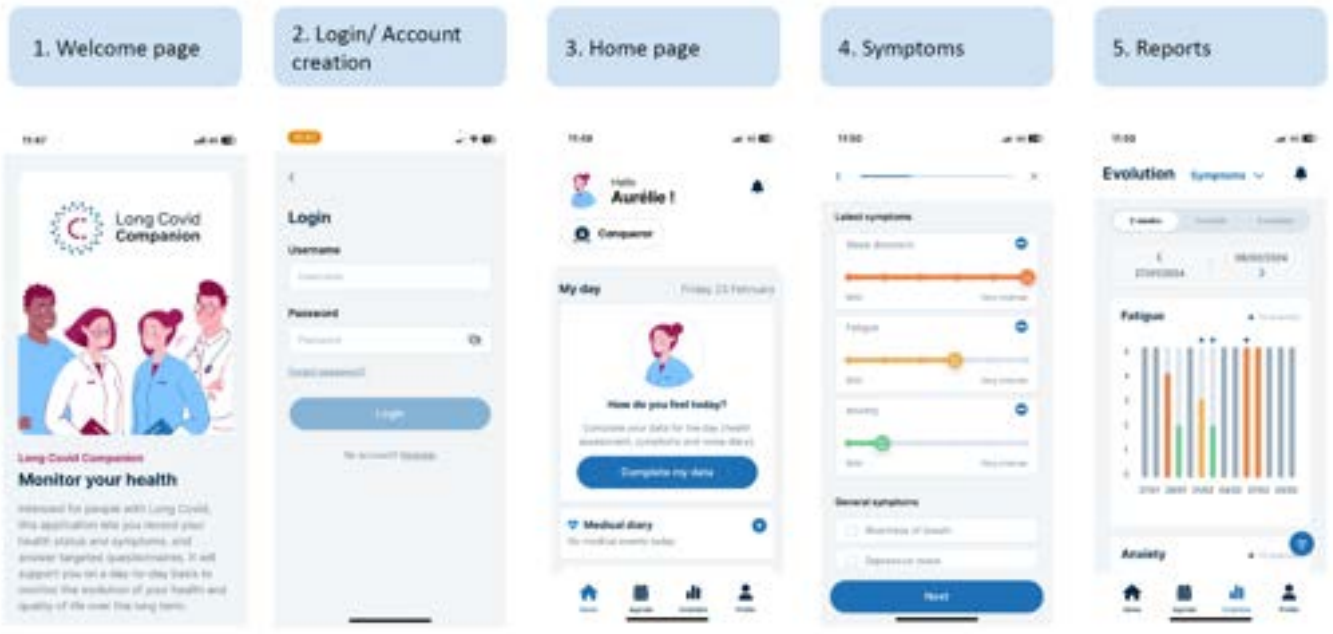


Figure 6: App user journey

App contents, i.e. the list of symptoms, the validated questionnaires and the resources to integrate in the app, have been validated with our patient partners from the #ApresJ20 Covid Long France Association who also tested a beta version of the app to identify and correct potential technical issues before organizing the focus group.

The app was presented during a focus group to 6 PWLC and 3 HCP. After a demonstration of the app, each main functionalities were discussed one by one before opening the discussion. Feedback and improvement suggestions are summarized in **Table 2**.

Table 2: Focus group feedbacks

Themes		**Feedback (problems/suggestions)**
Engagement	Reminders	"It would be nice, if possible, for us as patients to be able to choose whether or not to receive notifications." "Studies that have been published on apps for therapeutic compliance, in which there are regular reminders, have shown that they can improve compliance." "I completely agree that the intrusive aspect should be avoided" "I believe that a minimum recall frequency will make it useful." "The goal would be to make it really customizable." "Once a week is more reasonable (than every day)."
	Gamification	"Useful for some people, but not necessarily indispensable for everyone. It really depends on the person."
Functionality	Completion daily information	"The fact that you can actually go into one place and have successive screens, that's very good. I really like it. Yes, I like it a lot. It's very clear. We go straight to the point." "Likewise, I think it's clear. What's more, you move from the global to the more specific."
	Self-evaluation global health perception	"It's good to have a tool that lets you quickly assess your health and how you feel." "Being able to evaluate yourself like this, you can see that over the last month, there were times when things weren't going so well, but overall, it wasn't too bad."
	Self-evaluation symptom intensity	"You can assess a lot of symptoms. It's pretty cool. " "Just be careful between the terminology of symptoms and diagnosis. Depression is a diagnosis, I don't know what word could be used instead, but depressed mood, for example." "It allows us to see the whole

		picture of our symptoms." "It would be great to differenciate physical and mental health symptoms" "For the symptom list, did you used an official list from WHO or something else?"
	Standard questionnaires with scoring	"Indicate the scientific name of questionnaires on PDF reports, so that physicians can more likely know it" Only questionnaire scores will be available on reports, would it be interesting for HCP to have the details of responses to all questions? "The aim of the score is to have a sufficient value at the end." "The doctor doesn't necessarily have to look at the details of all the questions"
	Vocal diary	"I think that for the patients, it's something that's both very interesting and perhaps essential for them. Creating a narrative of their evolution in their own words, a kind of diary that also has a personal value." "I find it interesting, but I don't necessarily see what it can bring me" "It's a cathartic function for the patient, and in a concrete way, it helps reduce the deleterious effects of these complicated pathologies." "It's like a personal diary, but vocal."
	Calendar	"That's good. It gives a really global view over a whole month." "it's very concise, very visual. And with the color codes, you can actually see the days when you're better, when you're worse, and that really stands out." "it also lets you know whether I've been more or less involved in using the application, and whether I've invested more in it at one time or another." It's good to have the option to add medical appointments in the future.
	Graphical report for symptoms and questionnaires	"It's very visual too, the colors, the evolution, you can follow the evolution over time very easily." "It's a great benefit, it allows us to follow our progress over the long term."
	Research module with voice recordings	"That's the heart of your project, I think, voice biomarkers." "I'm rather curious to see what it will be like" "I'm all in favor" "Will we get feedback on the results of the recordings?" All participants would be willing to use this function.
	Daily life and medical diary	"It's quite important for pacing to be able to note on a daily basis the duration of our various activities and not only to mention "Intense" "Moderate" or "Mild"." "it's always possible to use the

		"other" fields or the voice diary" "there are days when I'm a little more willing. This can be recorded in the life diary too, in the other field, or in the voice diary."
Information	Ressources	"The way resources are displayed as vignettes and links, it's very clear and visual." What's interesting is that you use the word encouragement and motivation, not reward." We shouldn't feel bad if we don't use the application every day and fill in our diary every day."
	Chatbot integration	"Very good too, it's coherent in fact, it's a follow-up tool for the Long Covid, at the same time we also get instant answers to any questions we might have." "You just have to make sure that the answers remain up to date with the latest data."
Aesthetic	App colors	"Dark and clear mode is ergonomic"
Subjective items		"Well done, very well done. Honestly, there are a lot of points, the essential points of what we experience." "Will the data be use for research purposes? "It is important to be informed on how and where our data will be stored" Does the app meet the needs? "Yes, a big yes" "There might also have been one point I'd have liked to see a little more, and that's the link with healthcare professionals." "If we are too focused on our monitoring, it can also have, I think, a negative impact." "Thank you very much for giving us a voice, for listening to us" "We're trying to provide solutions to a real pathology. It's not the app that's going to get you hooked on your condition."

The app received very good overall feedback from both PWLC and HCP: The app seemed to meet the needs of PWLC in many aspects and some of them expressed notable gratitude for the provision of this new tool. Participants appreciated the daily data completion journey, in particular to go from the more general (global health status), to the more specific (symptoms and vocal diary).

The symptom list was validated by the participants. One participant noticed that one item in the list was more a diagnosis than a symptom and recommended replacing “Depression” by “Depressed mood”.

The possibility to complete specific symptoms and quality of life questionnaires were mandatory for all participants. Some participants wondered if the questionnaires’ scientific name and the detailed responses to each item would be present on the PDF report. Whilst the scientific name was present (e.g. fatigue evaluation was labeled FSS9), only the overall score was integrated in the report because inclusion of detailed responses would have overloaded the reports. 2 HCP present in the focus group ensured that the total score would be sufficiently informative to have a good overview.

One participant wondered whether there was a risk that the app would prompt users to stay focused on their disease and symptoms. A HCP present in the meeting answered that "We're trying to provide a solution to a real pathology. It's not the app that's going to get you hooked on your condition".

Some improvements suggested during the focus group were to provide additional ways to monitor mental health, to add or rename some symptoms, and to be able to record physical, intellectual or meal activity in greater detail.

When presenting the research module with standardized voice recording for vocal biomarkers, participants confirmed that they were in favor of it and willing to have and use it. However, some participants were a bit disappointed that vocal biomarkers to measure some symptoms were not integrated in the app.

DISCUSSION

This study showed the entire co-design process of the Long COVID Companion app, designed with PWLC and HCP. The use of mixed methods allowed us 1) to obtain an in-depth comprehension of the daily problems of PWLC, the needs and expectations related to a voice-based monitoring tool, 2) to define the main characteristics the app, and 3) to validate a first version of the Long COVID Companion app.

The main needs and expectations identified in our study were to provide daily support on both daily life (psychological and administrative support) and medical aspects (help symptom identification and monitoring, improve communication between PWLC and HCP, limit medical visits, etc.).

The use of voice was generally well perceived and the expected added-value of using vocal biomarkers was to limit the fatigue of completing questionnaires or typing symptom description and to objectivize symptoms in comparison to auto-evaluation.

Barriers and fears for the use of a voice-based symptom monitoring app were principally linked to data privacy and the reliability of this new technology. This finding could be explained by the high level of education of participants and is in line with a previous study in the context of diabetes distress [264]. This is an important aspect to be taken into account when developing voice-based technologies as voice is considered as sensitive and identifying data. Voice collection and analysis falls therefore under different regulations or laws, in particular GDPR in Europe or PIPEDA in Canada. In the United States, several laws exist at federal or state level. Finally, these different regulations do not protect individuals at the same level so it is highly recommended to obtain explicit informed consent for the users before voice collection[129].

Based on our results, an ideal voice-based symptom monitoring app for PWLC should include the following items: 1) self-monitoring of symptoms, 2) standardized questionnaires for frequent symptoms or for quality of life assessment, 3) visualization and individual reports, 4) journaling, and 5) provide informative resources.

Regarding the self-monitoring module, it was critical to PWLC and HCP to provide a list of the main Long COVID related symptoms to facilitate easier selection of those relevant to future users' experience. In addition, we added the option to auto-evaluate perceived global health on a 0 to 10 likert scale, which was meaningful for PWLC as it allowed them to evaluate their health at a global level and not focus only on individual symptoms. The use of vocal biomarkers for symptom monitoring was perceived as a promising approach, the advantages being the ease of use and the possibility to have a more objective evaluation than the auto-evaluation alone. There is, however, still a need to validate vocal biomarkers and to provide proof of their reliability.

The use of standardized questionnaires was an important feature to add as it was seen as a complement tool to assess symptoms and quality of life. The scientifically validated nature of the questionnaires was reassuring for PWLC. Moreover, obtaining scores on these questionnaires was perceived as tangible support facilitating discussion between PWLC and HCP. Despite standardized questionnaires being subject to bias (recall bias, respondent conscious or subconscious reaction, etc.)[265] they remain perceived by PWLC as more objective than self-evaluation.

The graphical representation of symptom intensity and questionnaire score evolution as shown in Figure 6 was essential for both PWLC and HCP. This should help PWLC to interpret their data generated in the app and obtain a more objective overview of their health status in the past days or weeks. This function has already been shown to have an impact on engagement and motivation to use such a technology in the long-term[113]. HCP also found that the option to generate a PDF report with this graphical representation could be of interest during consultations, by summarizing their patient's symptom trajectories at a glance.

The journaling was very important for PWLC because of the fluctuant character of their disease and because many of them have cognitive impairments. The diary would recall what happened in the past and have reminders for upcoming medical appointments. The possibility to indicate the important medical and life events in parallel to the symptom monitoring could allow them to identify potential correlations between an event and symptom relapse or improvement. We also choose to integrate a vocal diary which will allow the users to easily record all their concerns and feelings. Previous studies suggested that written journaling could have a positive impact on mental health[266], however there is a lack of consolidated results and long term assessment of the benefit of using a diary. In other settings like postoperative period or patients with neuroendocrine tumors it has been described that the use of a diary may reduce the recall biases and provide a broader picture of patients' quality of life than traditional questionnaires[267,268].

Voice journaling may have an additional value compared to written journaling by facilitating the diary completion and ultimately increasing the adherence to journaling.

Finally, different types of resources should be provided, in particular scientifically validated information on Long COVID. The aim of this module, however, would not only be to inform but also to help the app users to manage their disease. For example, tailored advice or rehabilitation exercises according to recorded symptoms could allow the app users to manage these symptoms. Information on specialized Long COVID consultation and administrative content on return to work and long-term disease recognition procedures could also be valuable.

The transferability of our results was also considered, defined as the extent to which it can be applied in other contexts or settings[269]. The essential app functionalities identified in this study correspond to the 3 steps of the self-care process described by Grosjean et al in the development of the e-CARE-PD study: monitoring, interpretation, and action[113]. Some of the identified needs are specific to PWLC, such as obtaining help in symptom identification, but other concerns are common to people affected by other chronic diseases. In particular, among the most impacting symptoms, fatigue and anxiety were common to the vast majority of our PWLC study participants. These are common problems encountered in other chronic diseases like diabetes, migraine or cancer[270,271]. The use of voice and of vocal biomarkers was well accepted, the main identified advantages being to reduce fatigue due to the completion of questionnaires, to limit travel for medical appointments and to give an objective assessment of a symptom. As fatigue and mental health issues are central problems in many other chronic conditions these findings could be transferable to other diseases.[271–273].

Finally, we co-developed the first version of Long COVID Companion with the objective to release it as soon as possible to offer a supportive tool for PWLC who are in need of helpful, concrete solutions. The app was released as a companion tool and not yet as a medical device. Further functionalities could be added later, such as a module of rehabilitation exercises. Further developments will enable the certification of the app as a medical device and to negotiate the app reimbursement. Although the use of voice and vocal biomarkers was expected by many participants, we could not integrate vocal biomarkers in this first version as our vocal biomarker candidates require more validation before implementation. A direct interaction between PWLC and HCP was expected by some of the PWLC participants but could not be developed easily and rapidly as there were considerable feasibility issues to tackle, for example the willingness of HCP to be involved in such a tool and the medical device regulations that would apply [274]. This will be explored in future development of the Long COVID Companion app.

The study had several **strengths**. First, it was based on mixed methods, combining quantitative and qualitative data, providing a more in-depth comprehension of the unmet needs and the potential solutions to them. Secondly, we involved PWLC and HCP in charge of Long COVID patients as study participants, which is highly recommended when co-designing digital health solutions[112]. In addition, we also involved PWLC and HCP as co-researchers throughout the entire study course, from the study conception to the study realization and valorization. They contributed to a better and complementary understanding of patient's needs and expectations that emerged from study results.

The involvement of patients as co-researcher underlined the importance of checking for post-exertional malaise and raising the awareness of pacing to learn how to manage day to day activities. They also proposed to integrate directly in the app their chatbot on Long COVID previously developed with PWLC from ApresJ20 Covid Long France and HCP in order to facilitate tailored answers to patients' needs. Our study participants, in both PWLC and HCP categories, had a wide age and education level range, showing that older people or those with lower levels of education, usually under-diagnosed, were interested in and willing to use such a voice-based monitoring app. Long Covid Companion app is, to our knowledge, the first app intended for PWLC entirely co-designed with them that will be available in 3 languages and free of charge in Europe. Similar co-design of digital health intervention studies exist in other therapeutic areas like Parkinson's disease, type 2 diabetes, or endometriosis[113,250,275], however our study is to our knowledge the only one intended for PWLC.

The study also had some **limitations**. There was an overrepresentation of women and of highly educated people in the survey and in the interviews. This is not surprising, as women are more frequently affected by Long COVID and more willing to participate in studies[276]. A majority of PWLC participating in the survey were already diagnosed, had severe Long COVID symptoms and were in a specialized Long COVID care. Therefore the interest in a symptom monitoring app might be different, and the needs and expectations might not be generalizable to the entire population of Long COVID patients, particularly those who are not in specialized care and lack an official Long COVID diagnosis whose needs may differ to our study population. The study was conducted only in French, so the app evaluation in the context of other languages and target populations will be necessary. We used ad-hoc questionnaires for the survey and to guide the interviews and no validated ones, moreover self-reported questionnaires are prone to desirability, recall, and response bias. However, as questionnaires were reviewed by PWLC and HCP before the study started, this mitigated the risk and allowed to obtain a comprehensive understanding of the problems and needs of PWLC and the risk is limited to impact app use and benefit. Finally there are also some limitations regarding vocal biomarkers that need clinical validation before their use in daily practice and with most of them still at a research stage.

CONCLUSION

We co-developed, with patients and HCPs, a voice-based app to monitor Long COVID symptoms based on the expectations and needs of PWLC. This digital health solution addresses a major gap and provides day-to-day support for people affected by Long Covid. We have also demonstrated that they have high expectations for the use of vocal biomarkers. Future longitudinal studies will be necessary to evaluate acceptability, usability and effectiveness of the Long COVID Companion app for Long COVID care and disease management, but also to validate the vocal biomarkers candidate of Long COVID symptoms. Future app certification as a medical device will also allow to integrate it in the healthcare pathway of PWLC.

Conflicting interests

There are no conflicts of interest to declare.

Funding

This work was supported by the Luxembourg Government through the CoVaLux programme and the Luxembourg Institute of Health.

Ethical approval

The study was approved by the National Research Ethics Committee of Luxembourg (study number 202208/04) in August 2022. An explicit informed consent was signed by participants of the qualitative phase.

Guarantor

AF

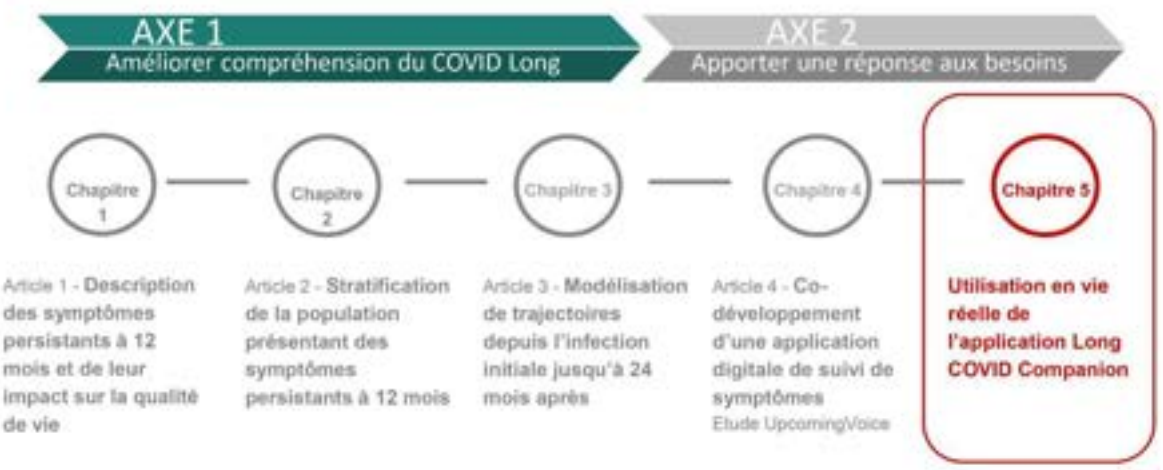
Contributorship

AF and GF researched literature and conceived the study. AF, GA, PO, TL, CB and JL were involved in protocol and study documents development and patient recruitment. AF and GA performed the data analysis. AF wrote the first draft of the manuscript. GF and PW contributed to funding acquisition. All authors reviewed and edited the manuscript and approved the final version of the manuscript.

Acknowledgements

We thank all the study participants and the patient partners who were involved in the study.

5.2 Chapitre 5 - Utilisation de l'application Long COVID Companion en vraie vie



5.2.1 Développement technique

Si le chapitre 4 a permis de définir les grandes lignes des spécifications attendues dans l'application, le développement technique a représenté un travail spécifique conséquent.

Généralités

Tout d'abord nous avons décidé qu'au vu du besoin, nous devions mettre l'application à disposition gratuitement, pour qu'elle soit accessible au plus grand nombre. Ensuite nous avons choisi de développer une application web et non une application classique mise à disposition via les stores. Ce choix a été fait pour 2 raisons principales. Tout d'abord le format application web permet l'utilisation de l'application sur smartphone et sur ordinateur ce qui était une demande des participants à l'étude UpcomingVoice, pour permettre aux personnes moins à l'aise avec les smartphones d'avoir malgré tout accès à l'application. De plus, ce format a permis une mise à disposition plus rapide de l'application en évitant les démarches de mise à disposition dans les stores.

Ensuite il a fallu s'intéresser à la problématique du logiciel comme dispositif médical. En effet, en fonction de la finalité choisie par le fabricant, un logiciel peut être considéré comme un dispositif médical et en cela, peut tomber sous le règlement (UE) 2017/745 relatif aux dispositifs médicaux en application depuis le 26 mai 2021[277]. Afin de pouvoir proposer une mise à disposition rapide nous avons fait le choix de ne pas demander la qualification de l'application comme dispositif médical, dans un premier temps. De ce fait, certaines fonctionnalités ne pouvaient pas être intégrées et en particulier les biomarqueurs vocaux car ils représentent une mesure physiologique. Il a également fallu être attentif à la formulation des messages dans l'application par souci de transparence vis-à -vis des futurs utilisateurs.

Il a également fallu décider de l'hébergement de l'application. Le Luxembourg Institute of Health, en tant qu'institut de recherche, n'a pas pour vocation à mettre à disposition et à gérer une application de ce type. Une possibilité envisagée alors était de proposer une application associative, gérée par l'association #ApresJ20 en France. Cependant cette solution impliquait pour l'association des coûts liés à l'hébergement et à la maintenance de l'application et de fait une recherche de financements qui ralentirait la mise à disposition. Finalement nous avons décidé que le LIH assurerait l'hébergement et la maintenance de l'application dans un premier temps, et que les données collectées pourraient être utilisées secondairement pour continuer les recherches sur le CL et les biomarqueurs vocaux, et pour étudier l'intérêt clinique potentiel de l'application.

Protection des données

L'hébergement de l'application impliquant la gestion d'une quantité conséquente de données sur le long terme, une analyse d'impact sur la protection des données (DPIA) a été réalisée en interne au LIH. Il en est ressorti la nécessité de concevoir l'application dans une perspective de protection de la vie privée prise en compte **dès la conception ("by design") et par défaut ("by default")** pour assurer aux futurs utilisateurs le plus haut niveau de protection de la vie privée et de sécurité des données personnelles.

De manière générale, les données sont traitées et stockées dans le respect du Règlement Général sur la Protection des Données n°2016/679 entré en vigueur en mai 2018 (RGPD).

Les mesures mises en œuvre dans le développement de l'application et de l'infrastructure informatique **dès la conception et par défaut** sont les suivantes.

1/ Les utilisateurs sont informés de leurs droits conformément au RGPD au moyen d'un document d'information et de consentement intégré à l'application.

2/ Une option permettant le téléchargement des données de son compte utilisateur ou la suppression des données locales et à distance a été intégrée.

3/ Une double protection de l'identité personnelle des données a été mise en œuvre dans le cadre d'une procédure en deux étapes :

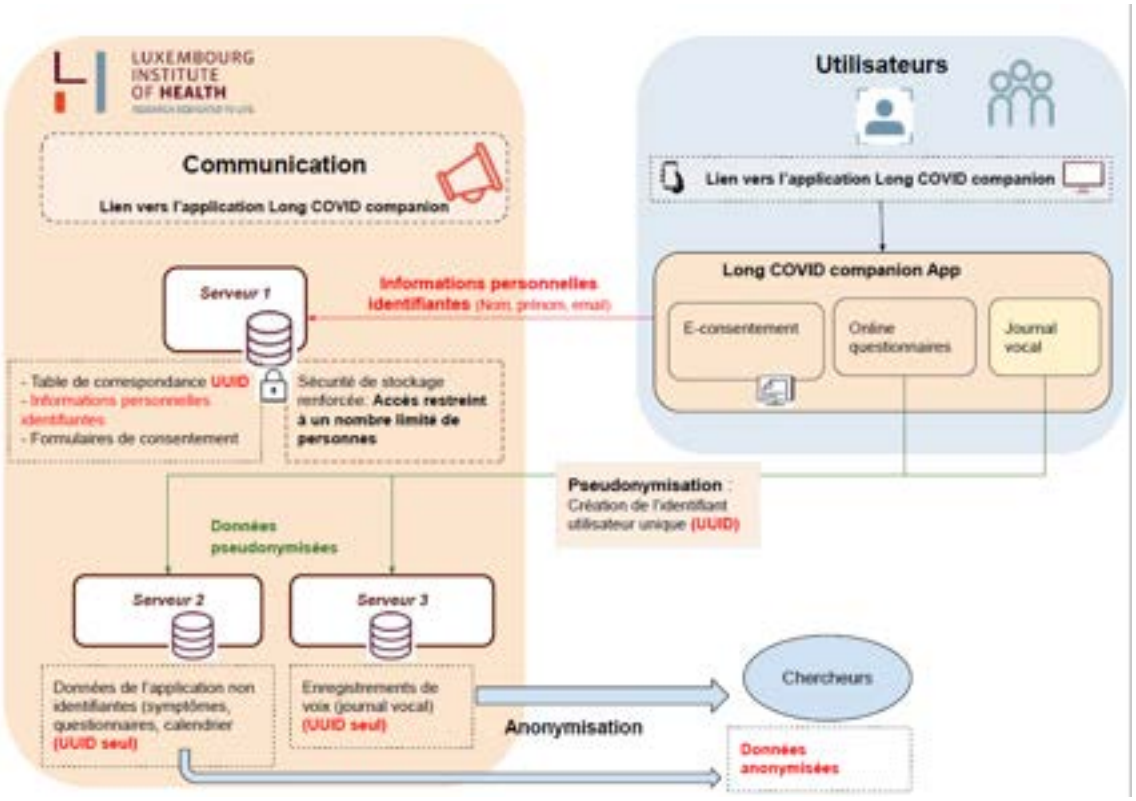
- Une pseudonymisation est effectuée en interne au LIH en attribuant à chaque utilisateur un identifiant utilisateur unique (UUID) qui permet de faire le lien entre les données dites identifiantes (nom, prénom, adresse email) et les autres données entrées dans l'application (symptômes, questionnaires, données du calendrier etc.). La table de correspondance entre les données identifiantes et l'UUID est stockée sur un serveur interne au LIH à accès restreint (= serveur 1).

- Le principe de séparation des données a été appliqué : toutes les données collectées sont stockées sur des serveurs séparés, avec un haut niveau de confidentialité et de sécurité. Le lien entre les données stockées sur les différents serveurs est l'UUID. La confidentialité est assurée par un accès restreint aux seuls professionnels autorisés dans le respect de leur obligation de secret.

Ainsi les données identifiantes du serveur 1 sont physiquement stockées dans un lieu différent des données cliniques (serveur 2) et des enregistrements de voix du journal vocal (serveur 3), ce qui rend le risque de ré-identification faible. Par défaut, l'accès aux données des différents serveurs est restreint à un nombre limité de personnes formées au traitement de données sensibles.

L'architecture mise en place pour le traitement des données collectées via l'application est schématisée par la **Figure 9**.

Figure 9 : Architecture du traitement et du stockage des données collectées via Long COVID Companion



Suivi du développement par le prestataire externe

Le développement technique de l'application a été sous-traité à un prestataire externe, le LIH n'ayant pas les ressources humaines nécessaires pour répondre à ce besoin. J'ai assuré le suivi de ce travail, depuis l'élaboration du cahier des charges, les demandes de devis, le choix du prestataire et les réunions régulières de suivi de l'avancement avec le prestataire choisi.

De nombreux choix devaient être effectués à chaque étape d'avancement, notamment le choix des visuels, la représentation graphique des différents éléments (échelles d'auto-évaluation de la santé de 0 à 10, de l'intensité des symptômes de 0 à 5, de l'évolution des symptômes au cours du temps etc.) et la présentation des informations du modules ressources.

Il a également fallu préparer tout le contenu textuel de l'application, du message de bienvenue aux textes explicatifs, jusqu'aux questionnaires standardisés, dans les 3 langues (français, anglais et allemand).

Une fois une version de test élaborée, des tests ont été organisés en interne dans l'équipe Deep Digital Phenotyping (DDP) du LIH pour identifier d'éventuels problèmes ou fautes de frappe. Plusieurs cycles de correctifs ont été nécessaires avant d'obtenir la version finale.

Test de sécurité (pen-test) avant la mise en production

Une fois la version finale prête, un test visant à détecter d'éventuelles failles de sécurité a été effectué avant sa mise en production par un prestataire externe indépendant du LIH et du prestataire ayant développé l'application. Ce test, appelé test de pénétration ou d'intrusion (pen-test) s'est déroulé pendant 3 semaines au mois de mars 2024. Pendant toute la durée du test des tentatives d'intrusion de la plus simple à la plus élaborée ont été simulées par le prestataire.

Le rapport de pen-test n'a soulevé aucune faille de sécurité majeure, mais a relevé deux points d'importance moyenne. Le premier était qu'il n'y avait pas de limite de tentatives de connexion lors de la saisie d'un mauvais mot de passe, avec pour risque le fait qu'un robot tente de se connecter en générant automatiquement un nombre important de mots de passe. Le second était qu'il n'y avait pas de contrôle du type de fichier des enregistrements vocaux, ce qui induisait un risque d'envoi de fichier malveillant par ce canal. Ces deux points ont été corrigés avant la mise en production (mise en ligne) de l'application.

5.2.2 Mise à disposition et données préliminaires d'utilisation de l'application

La mise en ligne de l'application a été faite le 5 avril 2024, avec une communication limitée dans un premier temps afin de collecter les premiers retours et d'avoir la possibilité de faire des modifications si nécessaire.

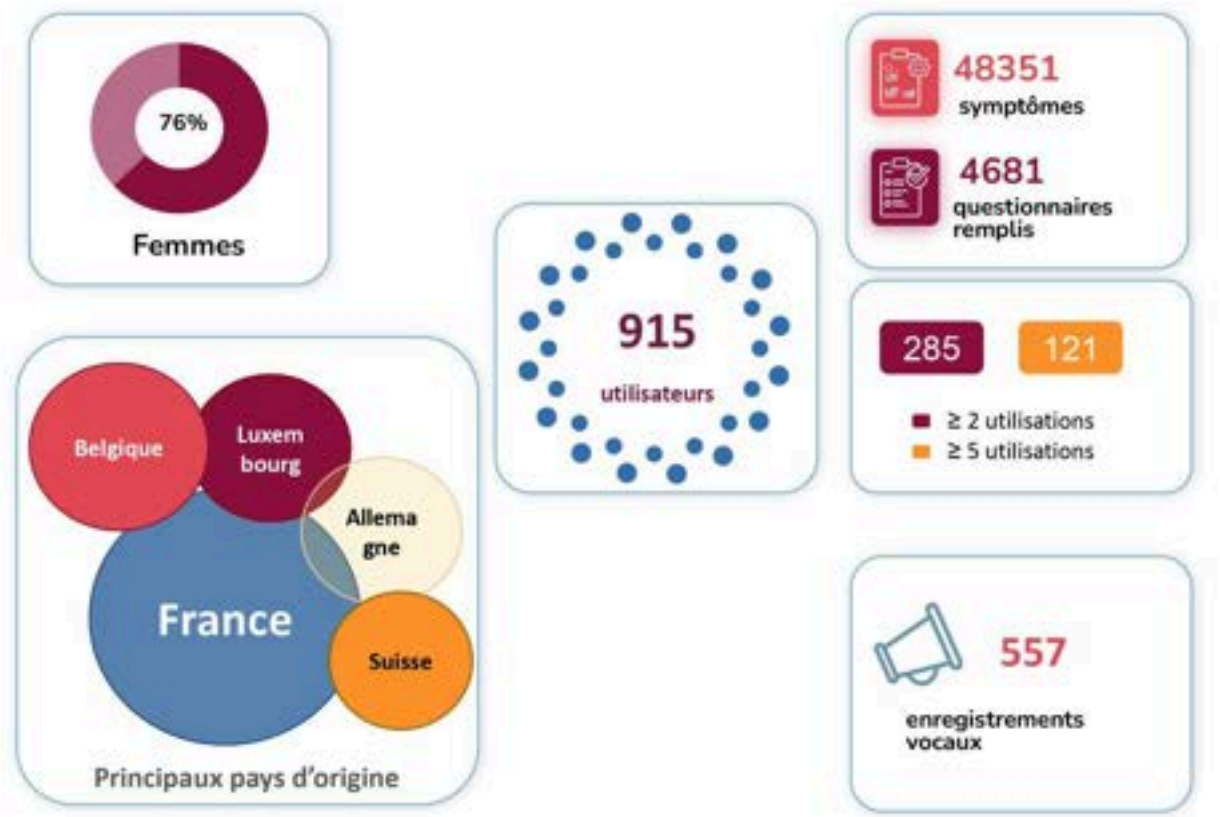
Ceci nous a ainsi permis de corriger quelques problèmes résiduels comme la durée de connexion à l'application sans avoir à se reconnecter, des problèmes de couleur d'affichage des résultats de certains questionnaires et quelques fautes de frappe dans certains messages.

Certains choix initiaux sur le format de question sur la prise en charge médicale en cours ont été corrigés, par exemple il était initialement possible de ne choisir qu'une seule réponse aux questions "Bénéficiez-vous d'une prise en charge médicale pour votre CL?" et "Bénéficiez-vous d'un programme de rééducation en lien avec votre CL?". Les premiers retours ont fait remonter le besoin de pouvoir choisir plusieurs réponses étant donné la pluralité de leurs prises en charge.

Début mai la communication autour de l'application s'est intensifiée, avec des posts sur LinkedIn, Facebook, des communiqués de presse au Luxembourg, et des distributions de dépliants de présentation de l'application dans les différents centres de prise en charge de personnes touchées par le CL.

Finalement, au 31 Août 2024, c'est-à-dire environ 5 mois après la mise en ligne de Long COVID Companion, 915 personnes avaient créé un compte. Parmi elles, 285 ont utilisé l'application au moins 2 fois et 121 au moins 5 fois. Plus de 48000 symptômes et 4600 questionnaires standardisés étaient renseignés. La fonction journal vocal comptait 557 enregistrements vocaux. Les utilisateurs sont majoritairement des femmes (N=698), résidant en France (N=751), Luxembourg (N=52) et Belgique (N=23). Les données préliminaires d'utilisation de Long COVID Companion au 31 Août 2024 sont présentées dans l'infographie de la **Figure 10**.

Figure 10 : Situation d'utilisation au 31 Août 2024



Ces résultats sont encourageants et sont une première preuve de l'acceptabilité de l'application et de ses différentes fonctionnalités.

6. Discussion générale et perspectives

Principaux résultats

L'**Axe 1 - Améliorer la compréhension du COVID Long -** de cette thèse a mis en évidence le caractère multiple du CL ainsi que son impact important sur la vie quotidienne des personnes touchées. Nous avons d'abord montré qu'un an après l'infection initiale, les symptômes persistants étaient présents chez 6 personnes sur 10 de la cohorte Predi-COVID, que les fréquences de symptômes augmentaient avec la sévérité initiale de la maladie et que les symptômes avaient tendance à se manifester de façon concomitante. Nous avons également proposé une classification des personnes touchées par des symptômes persistants en 3 sous-groupes de sévérité croissante de CL. Enfin nous avons modélisé des trajectoires de score total de symptômes et de symptômes individuels depuis l'infection initiale jusqu'à 2 ans après. Nous avons ainsi montré que le score total de symptômes évoluait suivant 2 trajectoires distinctes, l'une regroupant deux tiers des participants avec une évolution favorable et une régression rapide des symptômes et l'autre regroupant un tiers des participants et dans laquelle l'évolution du score de symptômes était défavorable et restait à un niveau élevé 2 ans après l'infection. L'ensemble de ce travail souligne la nécessité d'une prise en charge personnalisée en fonction des différentes formes de CL.

L'**Axe 2 - Apporter une réponse aux besoins -** a abouti à la co-construction, au développement technique et à la mise à disposition de l'application Long COVID Companion comme outil de soutien au quotidien pour les personnes touchées par le CL. Parmi ses fonctionnalités on retrouve une auto-évaluation de l'état de santé général et des symptômes ressentis, le remplissage de questionnaires validés pour certains symptômes, un calendrier d'événements médicaux et de la vie quotidienne, un journal vocal et la possibilité de générer des graphiques visuels pouvant servir de support lors des consultations avec les professionnels de santé. Enfin, l'application propose un volet "Ressources" avec des liens utiles vers des sites d'informations officiels sur le CL, des associations de patients et des réseaux de prise en charge spécialisés, en France, au Luxembourg et pays limitrophes. En date du 31 Août 2024, l'application a déjà été utilisée par plus de 900 personnes avec un CL.

Discussion générale

Points forts

Les travaux de cette thèse ont mis en lumière la complexité et l'hétérogénéité du CL, ses présentations cliniques multiples et son impact important sur la qualité de vie des personnes touchées, jusqu'à 2 ans après l'infection initiale. Nos résultats proposent un outil permettant d'identifier les personnes les plus à risque de forme sévère de CL et de leur proposer une meilleure prise en charge.

L'application Long COVID Companion est à ce jour, à notre connaissance, la première application permettant un suivi des symptômes, spécifique au CL, disponible en 3 langues et gratuitement. Elle a été développée suivant une démarche de recherche participative rigoureuse impliquant des personnes touchées par le CL à tous les stades du projet pour s'assurer de répondre aux attentes et donc de son acceptabilité future.

Nous avons également étudié et montré l'acceptabilité de l'utilisation de la voix et des biomarqueurs vocaux comme outil de suivi de la santé, à la fois par les personnes touchées par le CL et par les professionnels de santé qui se sont dit convaincus de l'utilité de cette technologie et prêts à l'utiliser. Il sera nécessaire dans les développements et utilisations futurs de cette technologie de prendre en compte les inquiétudes concernant les aspects de confidentialité et de protection des données, en proposant des informations extensives et des explications sur la démarche scientifique sous-jacente aux biomarqueurs vocaux.

Transférabilité

Les résultats de ce projet de thèse sont potentiellement transférables à d'autres cas d'usage. D'abord il était important pour l'équipe de recherche d'obtenir des informations sur l'acceptabilité des biomarqueurs vocaux pour le suivi de l'état de santé. Bien que la bonne acceptabilité des personnes touchées par le COVID pour cette nouvelle technologie puisse être liée à leur grande implication dans la recherche, il n'en reste pas moins intéressant de disposer de ce premier retour, et les informations recueillies seront utiles pour d'autres projets de recherche futurs. D'autre part, l'outil développé dans le cadre de cette thèse, l'application Long COVID Companion, pourra être adapté et transféré à d'autres thématiques, comme outil de collecte de données en vie réelle, ou de développement et de validation des biomarqueurs vocaux. De même, les outils et méthodes développés dans le cadre du co-design et de l'implication des patients dans la recherche pourront être utiles à d'autres projets de recherche du laboratoire.

Limites et défis

Tout d'abord, les études réalisées sur les données de la cohorte Predi-COVID ont été effectuées sur des jeux de données limités, en raison de la taille du pays. Bien que nos résultats soient en accord avec d'autres travaux sur le CL dans le monde, il serait utile de les valider sur des jeux de données plus larges et multinationaux. De plus, les données collectées grâce à l'application Long COVID Companion concernent actuellement une majorité de personnes en France, parlant français, ce qui est une limite pour le développement de nouveaux biomarqueurs vocaux, qui devront ensuite être validés dans d'autres langues.

L'application Long COVID Companion a été développée de manière à être accessible au plus grand nombre, en faisant le choix d'une application web qui permet une utilisation sur différents supports comme le smartphone, la tablette mais également l'ordinateur et en la proposant en 3 langues. Cependant, en tant que nouvelle technologie de santé il y a un risque d'accès limité à certaines populations comme les personnes âgées ou défavorisées et il sera important de décrire les caractéristiques des utilisateurs pour mettre en place des actions de diffusion ciblées.

La voix est une donnée considérée comme potentiellement identifiante au regard du RGPD, ce qui nécessite des mesures de protection des données particulières, en particulier le principe d'information et de consentement des personnes et celui de séparation des données lors du stockage. Le module de l'application contenant 4 enregistrements de voix standardisés pour des fins de recherche a volontairement été inactivé temporairement, car cette collecte sort du cadre de l'utilisation normale de l'application et nécessitera une évaluation par le comité d'éthique de la recherche et du ministère de la santé au Luxembourg.

Certaines autres fonctionnalités attendues par les personnes ayant un CL n'ont pas été intégrées à l'application. Par exemple, la possibilité de communiquer avec des professionnels de santé via l'application est un développement qui présente de nombreux défis comme celui de la disponibilité de professionnels de santé, l'intégration à des outils informatiques déjà existants pour le télésuivi des personnes, et les question réglementaires sous-jacentes à la télémédecine. L'intégration d'exercices de rééducation est aussi une problématique qui pourra être évaluée par la suite afin de proposer une solution sans risque de malaise post-effort pour les utilisateurs. Enfin, certains participants de l'étude UpcomingVoice étaient demandeurs de l'intégration d'un système de communication sous forme de chat ou de forum de discussion entre utilisateurs. Cette option pose également des défis propres, avec la nécessité de modération à tout instant et nous avons choisi de proposer à la place des liens vers des sites d'associations de patients et de groupes déjà existants sur les réseaux sociaux.

Du fait du caractère innovant de cette solution de santé et d'une réglementation mal définie autour du logiciel comme dispositif médical, il reste à clarifier si l'application Long COVID Companion pourrait être qualifiée de dispositif médical. Ceci nécessitera des investigations supplémentaires pour évaluer son efficacité en regard du bénéfice proposé et de sa sécurité. De probables ajustements des fonctionnalités de l'application seront nécessaires, mais pourraient permettre d'envisager son intégration officielle dans le parcours de soin et augmenter ainsi son impact pour les personnes avec un CL.

Perspectives

Cohorte virtuelle de personnes touchées par le COVID Long

Tout d'abord, les utilisateurs de l'application Long COVID Companion constituent désormais une cohorte virtuelle de personnes touchées par le CL, permettant de collecter de nombreuses données longitudinales sur la perception de leur état de santé, les symptômes rapportés ainsi que leur évolution, mais aussi l'impact du CL sur leur vie de tous les jours grâce aux données du journal de vie et médical. Ces informations seront très utiles pour générer de nouvelles données sur le CL et améliorer les connaissances sur cette maladie. Une extension de 18 mois du programme CoVaLux permettra de réaliser ces études. Une première étude basée sur les résultats de l'article sur les trajectoires de symptômes est en cours afin d'évaluer si la réponse immunitaire au moment de l'infection initiale est associée à l'appartenance à l'une ou l'autre des trajectoires de score total de symptômes ou de symptômes individuels.

D'autre part, les enregistrements vocaux du journal vocal seront utilisés dans le cadre d'études de validation des candidats biomarqueurs vocaux existants, en particulier celui de la fatigue qui est le symptôme le plus impactant dans le cadre du CL. Ces enregistrements constituent également une ressource cruciale pour le développement de nouveaux BV car ils représentent la première collecte longitudinale de voix disponible pour la recherche dans l'équipe DDP. Il est notamment prévu d'étudier l'évolution intra-individuelle des caractéristiques de la voix, indépendamment de l'évolution des symptômes, puis d'étudier comment certaines caractéristiques de la voix permettraient un suivi longitudinal de symptômes, en particulier les problèmes cognitifs et le stress qui sont également des symptômes très présents chez les personnes touchées par le CL.

Evolution de l'application Long COVID Companion

Des recherches de financement sont en cours afin de pérenniser l'application et de la faire évoluer par l'ajout de fonctionnalités attendues par les utilisateurs et qui n'ont pas pu être intégrées dans la première version en raison du budget et de temps limités.

Des études complémentaires seront également nécessaires afin d'évaluer si l'application pourrait être qualifiée de logiciel dispositif médical, ce qui favoriserait son déploiement et son intégration à part entière dans le circuit de prise en charge des personnes avec un CL.

Enfin, Long COVID Companion pourra servir d'outil de collecte de données dans le cadre d'études cliniques d'évaluation d'interventions médicamenteuses ou non médicamenteuses dans le cadre du CL.

Applications à d'autres domaines de recherche

Au-delà de la problématique du CL, l'application Long COVID Companion pourra être adaptée pour d'autres pathologies ou problématiques de santé publique et servir d'outil de collecte de données dans le cadre d'études cliniques. Un projet est actuellement en cours de préparation sur la santé des femmes, s'intéressant en particulier à la santé reproductive et maternelle, la santé mentale, le bien-être, et la santé cardiovasculaire. Il est prévu d'utiliser une version modifiée de Long COVID Companion, avec des modules supplémentaires permettant une collecte longitudinale de multiples enregistrements de voix, de selfies et de vidéos.

7. Conclusion générale

L'ensemble de ce travail de thèse présente un caractère multidisciplinaire intégrant la recherche clinique, l'épidémiologie, les méthodes mixtes qualitatives et quantitatives, l'analyse de données, l'intelligence artificielle et la santé digitale. La collaboration mise en place avec une association de patients CL et l'implication de patients et de professionnels de santé a servi de levier à la mise en place d'une recherche plus pertinente, centrée sur le patient et ses besoins réels. Ainsi, l'application Long COVID Companion a été développée dans le cadre de cette thèse pour répondre aux besoins des patients dans une démarche de recherche participative pour s'assurer de maximiser son impact.

Ce travail de thèse s'intègre dans le développement de la santé digitale et propose des réflexions sur le potentiel d'intégration d'une nouvelle technologie qu'est l'analyse de la voix au travers des biomarqueurs vocaux dans la prise en charge d'une maladie chronique comme le CL. Ce projet a montré l'acceptabilité de cette technologie par les personnes touchées par le CL mais aussi par les professionnels de santé.

Cette thèse contribue également à une meilleure compréhension du CL et met en évidence le besoin d'une prise en charge adaptée et personnalisée des personnes touchées afin de les soulager du fardeau de cette maladie au quotidien.

Cependant, malgré les efforts de recherche déjà entrepris, quatre ans après le début de la pandémie de COVID-19, le CL reste toujours un problème majeur de santé publique sous étudié. Il est donc crucial de continuer les recherches pour une meilleure compréhension de cette maladie, trouver des traitements et mieux anticiper les impacts à long terme sur la santé des populations.

Références

1. Why is COVID-19 data being presented as weekly statistics? In: datadot [Internet]. [cited 17 Apr 2024]. Available: https://data.who.int/dashboards/covid19/cases

2. Msemburi W, Karlinsky A, Knutson V, Aleshin-Guendel S, Chatterji S, Wakefield J. The WHO estimates of excess mortality associated with the COVID-19 pandemic. Nature. 2023;613: 130–137.

3. One year since the emergence of COVID-19 virus variant Omicron. [cited 30 May 2024]. Available: https://www.who.int/news-room/feature-stories/detail/one-year-since-the-emergence-of-omicron

4. Konings F, Perkins MD, Kuhn JH, Pallen MJ, Alm EJ, Archer BN, et al. SARS-CoV-2 Variants of Interest and Concern naming scheme conducive for global discourse. Nat Microbiol. 2021;6: 821–823.

5. Bedford J, Enria D, Giesecke J, Heymann DL, Ihekweazu C, Kobinger G, et al. COVID-19: towards controlling of a pandemic. Lancet. 2020;395: 1015–1018.

6. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020;395: 497–506.

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

8. Website. Available: https://iris.who.int/bitstream/handle/10665/347764/WHO-2019-nCoV-Post-COVID-19-condition-Clinical-case-definition-2021.1-fre.pdf?sequence=1&isAllowed=y

9. Website. doi:10.17226/27768.

10. Al-Aly Z, Topol E. Solving the puzzle of Long Covid. Science. 2024 [cited 18 Apr 2024]. doi:10.1126/science.adl0867

11. Woodrow M, Carey C, Ziauddeen N, Thomas R, Akrami A, Lutje V, et al. Systematic Review of the Prevalence of Long COVID. Open Forum Infect Dis. 2023;10: ofad233.

12. O'Mahoney LL, Routen A, Gillies C, Ekezie W, Welford A, Zhang A, et al. The prevalence and long-term health effects of Long Covid among hospitalised and non-hospitalised populations: A systematic review and meta-analysis. EClinicalMedicine. 2023;55: 101762.

13. Coronavirus disease (COVID-19): Post COVID-19 condition. [cited 18 Apr 2024]. Available: https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-(covid-19)-post-covid-19-condition

14. Davis HE, McCorkell L, Vogel JM, Topol EJ. Long COVID: major findings, mechanisms and recommendations. Nat Rev Microbiol. 2023;21: 133–146.

15. Fang Z, Ahrnsbrak R, Rekito A. Evidence Mounts That About 7% of US Adults Have Had Long COVID. JAMA. 2024. doi:10.1001/jama.2024.11370

16. Coste J, Delpierre C, Richard J-B, Alleaume C, Gallay A, Tebeka S, et al. Prevalence of long COVID

in the general adult population according to different definitions and sociodemographic and infection characteristics. A nationwide random sampling survey in France in autumn 2022. Clin Microbiol Infect. 2024;30: 924–929.

17. Al-Aly Z, Davis H, McCorkell L, Soares L, Wulf-Hanson S, Iwasaki A, et al. Long COVID science, research and policy. Nat Med. 2024. doi:10.1038/s41591-024-03173-6

18. Salmon D, Slama D, Linard F, Dumesges N, Le Baut V, Hakim F, et al. Patients with Long COVID continue to experience significant symptoms at 12 months and factors associated with improvement: A prospective cohort study in France (PERSICOR). Int J Infect Dis. 2024;140: 9–16.

19. Cutler DM. The Costs of Long COVID. JAMA Health Forum. 2022;3: e221809.

20. Perumal R, Shunmugam L, Naidoo K, Abdool Karim SS, Wilkins D, Garzino-Demo A, et al. Long COVID: a review and proposed visualization of the complexity of long COVID. Front Immunol. 2023;14: 1117464.

21. 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 Intern Med. 2023;183: 566–580.

22. Prevalence of post-acute coronavirus disease 2019 symptoms twelve months after hospitalization in participants retained in follow-up: analyses stratified by gender from a large prospective cohort. Clin Microbiol Infect. 2023;29: 254.e7–254.e13.

23. Bowe B, Xie Y, Al-Aly Z. Acute and postacute sequelae associated with SARS-CoV-2 reinfection. Nat Med. 2022;28: 2398–2405.

24. Al-Aly Z, Bowe B, Xie Y. Long COVID after breakthrough SARS-CoV-2 infection. Nat Med. 2022;28: 1461–1467.

25. Romero-Ibarguengoitia ME, Rodríguez-Torres JF, Garza-Silva A, Rivera-Cavazos A, Morales-Rodriguez DP, Hurtado-Cabrera M, et al. Association of vaccine status, reinfections, and risk factors with Long COVID syndrome. Sci Rep. 2024;14: 2817.

26. Brannock MD, Chew RF, Preiss AJ, Hadley EC, Redfield S, McMurry JA, et al. Long COVID risk and pre-COVID vaccination in an EHR-based cohort study from the RECOVER program. Nat Commun. 2023;14: 2914.

27. Malden DE, Liu I-LA, Qian L, Sy LS, Lewin BJ, Asamura DT, et al. Post-COVID conditions following COVID-19 vaccination: a retrospective matched cohort study of patients with SARS-CoV-2 infection. Nat Commun. 2024;15: 4101.

28. The impact of COVID-19 vaccination prior to SARS-CoV-2 infection on prevalence of long COVID among a population-based probability sample of Michiganders, 2020-2022. Ann Epidemiol. 2024;92: 17–24.

29. Choi Y, Kim HJ, Park J, Lee M, Kim S, Koyanagi A, et al. Acute and post-acute respiratory complications of SARS-CoV-2 infection: population-based cohort study in South Korea and Japan. Nat Commun. 2024;15: 4499.

30. Kim S, Lee H, Lee J, Lee SW, Kwon R, Kim MS, et al. Short- and long-term neuropsychiatric outcomes in long COVID in South Korea and Japan. Nat Hum Behav. 2024. doi:10.1038/s41562-024-01895-8

31. Watanabe A, Iwagami M, Yasuhara J, Takagi H, Kuno T. Protective effect of COVID-19 vaccination against long COVID syndrome: A systematic review and meta-analysis. Vaccine. 2023;41: 1783–1790.

32. Notarte KI, Catahay JA, Velasco JV, Pastrana A, Ver AT, Pangilinan FC, et al. Impact of COVID-19 vaccination on the risk of developing long-COVID and on existing long-COVID symptoms: A systematic review. EClinicalMedicine. 2022;53: 101624.

33. Nayyerabadi M, Fourcade L, Joshi SA, Chandrasekaran P, Chakravarti A, Massé C, et al. Vaccination after developing long COVID: Impact on clinical presentation, viral persistence, and immune responses. Int J Infect Dis. 2023;136: 136–145.

34. Tsuchida T, Hirose M, Inoue Y, Kunishima H, Otsubo T, Matsuda T. Relationship between changes in symptoms and antibody titers after a single vaccination in patients with Long COVID. J Med Virol. 2022;94: 3416–3420.

35. Kahlert CR, Strahm C, Güsewell S, Cusini A, Brucher A, Goppel S, et al. Post-Acute Sequelae After Severe Acute Respiratory Syndrome Coronavirus 2 Infection by Viral Variant and Vaccination Status: A Multicenter Cross-Sectional Study. Clin Infect Dis. 2023;77: 194–202.

36. Antonelli M, Pujol JC, Spector TD, Ourselin S, Steves CJ. Risk of long COVID associated with delta versus omicron variants of SARS-CoV-2. The Lancet. Elsevier BV; 2022. pp. 2263–2264.

37. Perlis RH, Santillana M, Ognyanova K, Safarpour A, Lunz Trujillo K, Simonson MD, et al. Prevalence and Correlates of Long COVID Symptoms Among US Adults. JAMA Netw Open. 2022;5: e2238804.

38. Padilla S, Ledesma C, García-Abellán J, García JA, Fernández-González M, de la Rica A, et al. Long COVID across SARS-CoV-2 variants, lineages, and sublineages. iScience. 2024;27: 109536.

39. Carfì A, Bernabei R, Landi F, Gemelli Against COVID-19 Post-Acute Care Study Group. Persistent Symptoms in Patients After Acute COVID-19. JAMA. 2020;324: 603–605.

40. Salmon-Ceron D, Slama D, De Broucker T, Karmochkine M, Pavie J, Sorbets E, et al. Clinical, virological and imaging profile in patients with prolonged forms of COVID-19: A cross-sectional study. J Infect. 2021;82: e1–e4.

41. Vernon SD, Hartle M, Sullivan K, Bell J, Abbaszadeh S, Unutmaz D, et al. Post-exertional malaise among people with long COVID compared to myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). Work. 2023;74: 1179–1186.

42. Servier C, Porcher R, Pane I, Ravaud P, Tran V-T. Trajectories of the evolution of post-COVID-19 condition, up to two years after symptoms onset. Int J Infect Dis. 2023;133: 67–74.

43. Giussani G, Westenberg E, Garcia-Azorin D, Bianchi E, Khan Y, Khan AH, et al. Prevalence and Trajectories of Post-COVID-19 Neurological Manifestations: A Systematic Review and Meta-Analysis. Neuroepidemiology. 2024;58: 120–133.

44. 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.

45. Spudich S, Nath A. Nervous system consequences of COVID-19. Science. 2022;375: 267–269.

46. Hampshire A, Azor A, Atchison C, Trender W, Hellyer PJ, Giunchiglia V, et al. Cognition and Memory after Covid-19 in a Large Community Sample. N Engl J Med. 2024;390: 806–818.

47. Boscolo-Rizzo P, Spinato G, Hopkins C, Marzolino R, Cavicchia A, Zucchini S, et al. Evaluating long-term smell or taste dysfunction in mildly symptomatic COVID-19 patients: a 3-year follow-up study. Eur Arch Otorhinolaryngol. 2023;280: 5625–5630.

48. Natarajan A, Shetty A, Delanerolle G, Zeng Y, Zhang Y, Raymont V, et al. A systematic review and meta-analysis of long COVID symptoms. Syst Rev. 2023;12: 88.

49. Gheorghita R, Soldanescu I, Lobiuc A, Caliman Sturdza OA, Filip R, Constantinescu-Bercu A, et al. The knowns and unknowns of long COVID-19: from mechanisms to therapeutical approaches. Front Immunol. 2024;15: 1344086.

50. Kim Y, Bae S, Chang H-H, Kim S-W. Characteristics of long COVID and the impact of COVID-19 vaccination on long COVID 2 years following COVID-19 infection: prospective cohort study. Sci Rep. 2024;14: 854.

51. Linh TTD, Ho DKN, Nguyen NN, Hu C-J, Yang C-H, Wu D. Global prevalence of post-COVID-19 sleep disturbances in adults at different follow-up time points: A systematic review and meta-analysis. Sleep Med Rev. 2023;71: 101833.

52. McMahon DE, Gallman AE, Hruza GJ, Rosenbach M, Lipoff JB, Desai SR, et al. Long COVID in the skin: a registry analysis of COVID-19 dermatological duration. Lancet Infect Dis. 2021;21: 313–314.

53. Li J, Zhou Y, Ma J, Zhang Q, Shao J, Liang S, et al. The long-term health outcomes, pathophysiological mechanisms and multidisciplinary management of long COVID. Signal Transduct Target Ther. 2023;8: 416.

54. Lamb LE, Timar R, Wills M, Dhar S, Lucas SM, Komnenov D, et al. Long COVID and COVID-19-associated cystitis (CAC). Int Urol Nephrol. 2022;54: 17–21.

55. Cairns R, Hotopf M. A systematic review describing the prognosis of chronic fatigue syndrome. Occup Med . 2005;55: 20–31.

56. Taquet M, Sillett R, Zhu L, Mendel J, Camplisson I, Dercon Q, et al. Neurological and psychiatric risk trajectories after SARS-CoV-2 infection: an analysis of 2-year retrospective cohort studies including 1 284 437 patients. Lancet Psychiatry. 2022;9: 815–827.

57. Tran V-T, Porcher R, Pane I, Ravaud P. Course of post COVID-19 disease symptoms over time in the ComPaRe long COVID prospective e-cohort. Nat Commun. 2022;13: 1812.

58. Greenhalgh T, Sivan M, Perlowski A, Nikolich JŽ. Long COVID: a clinical update. Lancet. 2024. doi:10.1016/S0140-6736(24)01136-X

59. van den Houdt SCM, Slurink IAL, Mertens G. Long COVID is not a uniform syndrome: Evidence from person-level symptom clusters using latent class analysis. J Infect Public Health. 2024;17: 321–328.

60. Reese J, Blau H, Bergquist T, Loomba JJ, Callahan T, Laraway B, et al. Generalizable Long COVID Subtypes: Findings from the NIH N3C and RECOVER Programs. medRxiv. 2022. doi:10.1101/2022.05.24.22275398

61. Kenny G, McCann K, O'Brien C, Savinelli S, Tinago W, Yousif O, et al. Identification of Distinct Long COVID Clinical Phenotypes Through Cluster Analysis of Self-Reported Symptoms. Open Forum Infect Dis. 2022;9: ofac060.

62. Fernández-de-Las-Peñas C, Martín-Guerrero JD, Florencio LL, Navarro-Pardo E, Rodríguez-Jiménez J, Torres-Macho J, et al. Clustering analysis reveals different profiles associating long-term post-COVID symptoms, COVID-19 symptoms at hospital admission and previous medical co-morbidities in previously hospitalized COVID-19 survivors. Infection. 2022. doi:10.1007/s15010-022-01822-x

63. Ayadi H, Bour C, Fischer A, Ghoniem M, Fagherazzi G. The Long COVID experience from a patient's perspective: a clustering analysis of 27,216 Reddit posts. Front Public Health. 2023;11: 1227807.

64. Iversen KK, Ronit A, Ahlström MG, Nordestgaard BG, Afzal S, Benfield T. Lung Function Trajectories in Mild COVID-19 With 2-year Follow-up. J Infect Dis. 2024; jiae037.

65. Wynberg E, Verveen A, van Willigen HDG, Nieuwkerk P, Davidovich U, Lok A, et al. Two-year trajectories of COVID-19 symptoms and their association with illness perception: A prospective cohort study in Amsterdam, the Netherlands. Influenza Other Respi Viruses. 2023;17: e13190.

66. Ramakrishnan RK, Kashour T, Hamid Q, Halwani R, Tleyjeh IM. Unraveling the Mystery Surrounding Post-Acute Sequelae of COVID-19. Front Immunol. 2021;12: 686029.

67. Tejerina F, Catalan P, Rodriguez-Grande C, Adan J, Rodriguez-Gonzalez C, Muñoz P, et al. Post-COVID-19 syndrome. SARS-CoV-2 RNA detection in plasma, stool, and urine in patients with persistent symptoms after COVID-19. BMC Infect Dis. 2022;22: 211.

68. Ceulemans LJ, Khan M, Yoo S-J, Zapiec B, Van Gerven L, Van Slambrouck J, et al. Persistence of SARS-CoV-2 RNA in lung tissue after mild COVID-19. Lancet Respir Med. 2021;9: e78–e79.

69. Bernal KDE, Whitehurst CB. Incidence of Epstein-Barr virus reactivation is elevated in COVID-19 patients. Virus Res. 2023;334: 199157.

70. Peluso MJ, Deveau T-M, Munter SE, Ryder D, Buck A, Beck-Engeser G, et al. Chronic viral coinfections differentially affect the likelihood of developing long COVID. J Clin Invest. 2023;133. doi:10.1172/JCI163669

71. Su Y, Yuan D, Chen DG, Ng RH, Wang K, Choi J, et al. Multiple early factors anticipate post-acute COVID-19 sequelae. Cell. 2022;185: 881–895.e20.

72. Arthur JM, Forrest JC, Boehme KW, Kennedy JL, Owens S, Herzog C, et al. Development of ACE2 autoantibodies after SARS-CoV-2 infection. PLoS One. 2021;16: e0257016.

73. Klein J, Wood J, Jaycox J, Lu P, Dhodapkar RM, Gehlhausen JR, et al. Distinguishing features of Long COVID identified through immune profiling. medRxiv. 2022. doi:10.1101/2022.08.09.22278592

74. Turner S, Khan MA, Putrino D, Woodcock A, Kell DB, Pretorius E. Long COVID: pathophysiological factors and abnormalities of coagulation. Trends Endocrinol Metab. 2023;34: 321–344.

75. Kell DB, Khan MA, Kane B, Lip GYH, Pretorius E. Possible Role of Fibrinaloid Microclots in Postural Orthostatic Tachycardia Syndrome (POTS): Focus on Long COVID. J Pers Med. 2024;14. doi:10.3390/jpm14020170

76. Álvarez-Santacruz C, Tyrkalska SD, Candel S. The Microbiota in Long COVID. Int J Mol Sci. 2024;25. doi:10.3390/ijms25021330

77. Annesley SJ, Missailidis D, Heng B, Josev EK, Armstrong CW. Unravelling shared mechanisms: insights from recent ME/CFS research to illuminate long COVID pathologies. Trends Mol Med. 2024;30: 443–458.

78. Zhao Y, Liang Q, Jiang Z, Mei H, Zeng N, Su S, et al. Brain abnormalities in survivors of COVID-19 after 2-year recovery: a functional MRI study. Lancet Reg Health West Pac. 2024;47: 101086.

79. Horowitz T, Pellerin L, Zimmer ER, Guedj E. Brain fog in long COVID: A glutamatergic hypothesis with astrocyte dysfunction accounting for brain PET glucose hypometabolism. Med Hypotheses. 2023;180: 111186.

80. NICE Rapid Guideline on Long Covid. In: The ME Association [Internet]. 10 Jan 2024 [cited 1 Jul 2024]. Available: https://meassociation.org.uk/nice-rapid-guideline-long-covid/

81. Seo J-W, Kim SE, Kim Y, Kim EJ, Kim T, Kim T, et al. Updated Clinical Practice Guidelines for the Diagnosis and Management of Long COVID. Infect Chemother. 2024;56: 122–157.

82. Lai Y-J, Liu S-H, Manachevakul S, Lee T-A, Kuo C-T, Bello D. Biomarkers in long COVID-19: A systematic review. Front Med. 2023;10: 1085988.

83. MacEwan SR, Rahurkar S, Tarver WL, Forward C, Eramo JL, Teuschler L, et al. Patient Experiences Navigating Care Coordination For Long COVID: A Qualitative Study. J Gen Intern Med. 2024;39: 1294–1300.

84. Wolf S, Zechmeister-Koss I, Erdös J. Possible long COVID healthcare pathways: a scoping review. BMC Health Serv Res. 2022;22: 1076.

85. Parker M, Sawant HB, Flannery T, Tarrant R, Shardha J, Bannister R, et al. Effect of using a structured pacing protocol on post-exertional symptom exacerbation and health status in a longitudinal cohort with the post-COVID-19 syndrome. J Med Virol. 2023;95: e28373.

86. Bramante CT, Buse JB, Liebovitz DM, Nicklas JM, Puskarich MA, Cohen K, et al. Outpatient treatment of COVID-19 and incidence of post-COVID-19 condition over 10 months (COVID-OUT): a multicentre, randomised, quadruple-blind, parallel-group, phase 3 trial. Lancet Infect Dis. 2023;23: 1119–1129.

87. Choi YJ, Seo YB, Seo J-W, Lee J, Nham E, Seong H, et al. Effectiveness of Antiviral Therapy on Long COVID: A Systematic Review and Meta-Analysis. J Clin Med Res. 2023;12. doi:10.3390/jcm12237375

88. Ioannou GN, Berry K, Rajeevan N, Li Y, Mutalik P, Yan L, et al. Effectiveness of Nirmatrelvir-Ritonavir Against the Development of Post-COVID-19 Conditions Among U.S. Veterans : A Target Trial Emulation. Ann Intern Med. 2023;176: 1486–1497.

89. Momtazmanesh S, Ansari S, Izadi Z, Shobeiri P, Vatankhah V, Seifi A, et al. Effect of famotidine on cognitive and behavioral dysfunctions induced in post-COVID-19 infection: A randomized, double-blind, and placebo-controlled study. J Psychosom Res. 2023;172: 111389.

90. Patterson BK, Yogendra R, Guevara-Coto J, Mora-Rodriguez RA, Osgood E, Bream J, et al. Case series: Maraviroc and pravastatin as a therapeutic option to treat long COVID/Post-acute sequelae of COVID (PASC). Front Med. 2023;10: 1122529.

91. Scheppke KA, Pepe PE, Jui J, Crowe RP, Scheppke EK, Klimas NG, et al. Remission of severe forms of long COVID following monoclonal antibody (MCA) infusions: A report of signal index cases and call for targeted research. Am J Emerg Med. 2024;75: 122–127.

92. Hadanny A, Zilberman-Itskovich S, Catalogna M, Elman-Shina K, Lang E, Finci S, et al. Long term outcomes of hyperbaric oxygen therapy in post covid condition: longitudinal follow-up of a randomized controlled trial. Sci Rep. 2024;14: 3604.

93. Laubscher GJ, Khan MA, Venter C, Pretorius JH, Kell DB, Pretorius E. Treatment of Long COVID symptoms with triple anticoagulant therapy. Research Square. 2023. doi:10.21203/rs.3.rs-2697680/v1

94. Kim Y, Bae S, Chang H-H, Kim S-W. Long COVID prevalence and impact on quality of life 2 years after acute COVID-19. Sci Rep. 2023;13: 11207.

95. Vélez-Santamaría R, Fernández-Solana J, Méndez-López F, Domínguez-García M, González-Bernal JJ, Magallón-Botaya R, et al. Functionality, physical activity, fatigue and quality of life in patients with acute COVID-19 and Long COVID infection. Sci Rep. 2023;13: 19907.

96. Kitsios GD, Blacka S, Jacobs JJ, Mirza T, Naqvi A, Gentry H, et al. Subphenotypes of self-reported symptoms and outcomes in long COVID: a prospective cohort study with latent class analysis. BMJ Open. 2024;14: e077869.

97. Walker S, Goodfellow H, Pookarnjanamorakot P, Murray E, Bindman J, Blandford A, et al. Impact of fatigue as the primary determinant of functional limitations among patients with post-COVID-19 syndrome: a cross-sectional observational study. BMJ Open. 2023;13: e069217.

98. Hawke LD, Nguyen ATP, Sheikhan NY, Strudwick G, Rossell SL, Soklaridis S, et al. Swept under the carpet: a qualitative study of patient perspectives on Long COVID, treatments, services, and mental health. BMC Health Serv Res. 2023;23: 1088.

99. Website. Available: https://iris.who.int/bitstream/handle/10665/344250/9789240027558-fre.pdf

100. Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. Scientific Data. 2016;3: 1–9.

101. Orri M, Lipset CH, Jacobs BP, Costello AJ, Cummings SR. Web-based trial to evaluate the efficacy and safety of tolterodine ER 4 mg in participants with overactive bladder: REMOTE trial. Contemp Clin Trials. 2014;38: 190–197.

102. Martos Pérez F, Dolores Martín Escalante M, Olalla Sierra J, Luis Prada Pardal J, Dolores García de Lucas M, González Vega R, et al. The value of telephone consultations during COVID-19 pandemic. An observational study. QJM. 2021. doi:10.1093/qjmed/hcab024

103. Patra R, Das NC, Bhattacharya M, Shit PK, Patra BC, Mukherjee S. Applications of Artificial Intelligence (AI) Protecting from COVID-19 Pandemic: A Clinical and Socioeconomic Perspective. EAI/Springer Innovations in Communication and Computing. 2021. pp. 45–60. doi:10.1007/978-3-030-68936-0_3

104. Wang Q, Su M, Zhang M, Li R. Integrating Digital Technologies and Public Health to Fight Covid-19 Pandemic: Key Technologies, Applications, Challenges and Outlook of Digital Healthcare. Int J Environ Res Public Health. 2021;18. doi:10.3390/ijerph18116053

105. Whitelaw S, Mamas MA, Topol E, Van Spall HGC. Applications of digital technology in COVID-19 pandemic planning and response. Lancet Digit Health. 2020;2: e435–e440.

106. Shah W, Hillman T, Playford ED, Hishmeh L. Managing the long term effects of covid-19: summary of NICE, SIGN, and RCGP rapid guideline. BMJ. 2021;372: n136.

107. COVID-19 rapid guideline: managing the long-term effects of COVID-19. London: National Institute for Health and Care Excellence (UK); 2021.

108. Bouabida K, Malas K, Talbot A, Desrosiers M-È, Lavoie F, Lebouché B, et al. Remote Patient Monitoring Program for COVID-19 Patients Following Hospital Discharge: A Cross-Sectional Study. Front Digit Health. 2021;3: 721044.

109. Visible - Activity tracking for Long Covid and ME/CFS. [cited 27 May 2024]. Available: https://www.makevisible.com/

110. Covid Recovery - Living With. In: Living With - Smart condition management and medical health apps [Internet]. Living With; 31 Oct 2022 [cited 18 Oct 2023]. Available: https://livingwith.health/covid-recovery/

111. Smuck M, Odonkor CA, Wilt JK, Schmidt N, Swiernik MA. The emerging clinical role of wearables: factors for successful implementation in healthcare. NPJ Digit Med. 2021;4: 45.

112. Meskó B, deBronkart D. Patient Design: The Importance of Including Patients in Designing Health Care. J Med Internet Res. 2022;24: e39178.

113. Grosjean S, Ciocca J-L, Gauthier-Beaupré A, Poitras E, Grimes D, Mestre T. Co-designing a digital companion with people living with Parkinson's to support self-care in a personalized way: The eCARE-PD Study. Digit Health. 2022;8: 20552076221081695.

114. Eysenbach G. The law of attrition. J Med Internet Res. 2005;7: e11.

115. Noorbergen TJ, Adam MTP, Teubner T, Collins CE. Using Co-design in Mobile Health System Development: A Qualitative Study With Experts in Co-design and Mobile Health System Development. JMIR Mhealth Uhealth. 2021;9: e27896.

116. [No title]. [cited 9 Jul 2024]. Available: https://www.ema.europa.eu/en/documents/other/questions-and-answers-qualification-digital-technology-based-methodologies-support-approval-medicinal-products_en.pdf

117. Manta C, Patrick-Lake B, Goldsack JC. Digital Measures That Matter to Patients: A Framework to Guide the Selection and Development of Digital Measures of Health. Digit Biomark. 2020;4: 69–77.

118. Fagherazzi G, Fischer A, Ismael M, Despotovic V. Voice for Health: The Use of Vocal Biomarkers from Research to Clinical Practice. Digit Biomark. 2021;5: 78–88.

119. Tsanas A, Little MA, McSharry PE, Ramig LO. Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. IEEE Trans Biomed Eng. 2010;57: 884–893.

120. Tracy JM, Özkanca Y, Atkins DC, Hosseini Ghomi R. Investigating voice as a biomarker: Deep phenotyping methods for early detection of Parkinson's disease. J Biomed Inform. 2020;104: 103362.

121. Dashtipour K, Tafreshi A, Lee J, Crawley B. Speech disorders in Parkinson's disease: pathophysiology, medical management and surgical approaches. Neurodegener Dis Manag. 2018;8: 337–348.

122. Home. In: Real Time Voice Analyzer [Internet]. 23 Apr 2023 [cited 21 Jun 2024]. Available: https://realtimevoiceanalyzer.com/

123. Vocal Biomarkers for Mental Fitness Scoring and Tracking. In: Sonde Health [Internet]. [cited 3 Jun 2024]. Available: https://www.sondehealth.com/vocal-biomarkers-for-mental-fitness-scoring-and-tracking

124. Website. Available: https://eur-lex.europa.eu/eli/ reg/2016/679/oj

125. Office of the Privacy Commissioner of Canada. The Personal Information Protection and Electronic Documents Act (PIPEDA). 8 Dec 2021 [cited 25 Jun 2024]. Available: https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/

126. Elbéji A, Zhang L, Higa E, Fischer A, Despotovic V, Nazarov PV, et al. Vocal biomarker predicts fatigue in people with COVID-19: results from the prospective Predi-COVID cohort study. BMJ Open. 2022;12: e062463.

127. Higa E, Elbéji A, Zhang L, Fischer A, Aguayo GA, Nazarov PV, et al. Discovery and Analytical Validation of a Vocal Biomarker to Monitor Anosmia and Ageusia in Patients With COVID-19: Cross-sectional Study. JMIR Med Inform. 2022;10: e35622.

128. Fagherazzi G, Zhang L, Elbéji A, Higa E, Despotovic V, Ollert M, et al. A voice-based biomarker for monitoring symptom resolution in adults with COVID-19: Findings from the prospective Predi-COVID cohort study. PLOS Digit Health. 2022;1: e0000112.

129. Fischer A, Elbeji A, Aguayo G, Fagherazzi G. Recommendations for Successful Implementation of the Use of Vocal Biomarkers for Remote Monitoring of COVID-19 and Long COVID in Clinical Practice and Research. Interact J Med Res. 2022;11: e40655.

130. Fiorella ML, Cavallaro G, Di Nicola V, Quaranta N. Voice Differences When Wearing and Not Wearing a Surgical Mask. J Voice. 2021. doi:10.1016/j.jvoice.2021.01.026

131. Davis KD, Aghaeepour N, Ahn AH, Angst MS, Borsook D, Brenton A, et al. Discovery and validation of biomarkers to aid the development of safe and effective pain therapeutics: challenges and opportunities. Nat Rev Neurol. 2020;16: 381–400.

132. Predi-Covid. In: Luxembourg Institute of Health [Internet]. 3 Apr 2023 [cited 16 Apr 2024]. Available: https://www.lih.lu/en/predi-covid-home/

133. Long COVID - CoVaLux. In: researchluxembourg [Internet]. 10 Jan 2022 [cited 16 Apr 2024]. Available: https://www.researchluxembourg.org/en/covid-19-task-force/projects/covalux/

134. Fagherazzi G, Fischer A, Betsou F, Vaillant M, Ernens I, Masi S, et al. Protocol for a prospective, longitudinal cohort of people with COVID-19 and their household members to study factors associated with disease severity: the Predi-COVID study. BMJ Open. 2020;10: e041834.

135. COVID-19 CRF. In: ISARIC [Internet]. 13 Oct 2020 [cited 22 Mar 2024]. Available: https://isaric.org/research/covid-19-clinical-research-resources/covid-19-crf/

136. Tran V-T, Riveros C, Clepier B, Desvarieux M, Collet C, Yordanov Y, et al. Development and Validation of the Long Coronavirus Disease (COVID) Symptom and Impact Tools: A Set of Patient-Reported Instruments Constructed From Patients' Lived Experience. Clin Infect Dis. 2022;74: 278–287.

137. Malik AO, Peri-Okonny P, Gosch K, Thomas M, Mena C, Hiatt WR, et al. Association of Perceived Stress Levels With Long-term Mortality in Patients With Peripheral Artery Disease. JAMA Netw Open. 2020;3: e208741.

138. Naik H, Shao S, Tran KC, Wong AW, Russell JA, Khor E, et al. Evaluating fatigue in patients recovering from COVID-19: validation of the fatigue severity scale and single item screening questions. Health Qual Life Outcomes. 2022;20: 170.

139. Krupp LB, LaRocca NG, Muir-Nash J, Steinberg AD. The fatigue severity scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. Arch Neurol. 1989;46: 1121–1123.

140. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med. 2006;166: 1092–1097.

141. Patient & Public. [cited 22 May 2024]. Available: https://www.spcr.nihr.ac.uk/PPI/patient-public-involvement-ppi

142. Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gülmezoglu AM, et al. How to increase value and reduce waste when research priorities are set. Lancet. 2014;383: 156–165.

143. Szekely GJ, Rizzo ML. Hierarchical clustering via joint between-within distances: Extending ward's minimum variance method. J Classification. 2005;22: 151–183.

144. Agglomerative Hierarchical Clustering Algorithm- A Review. [cited 18 Jul 2024]. Available: https://www.ijsrp.org/research-paper-0313.php?rp=P15831

145. Lin H, McCulloch CE, Rosenheck RA. Latent pattern mixture models for informative intermittent missing data in longitudinal studies. Biometrics. 2004;60: 295–305.

146. Proust-Lima C, Philipps V, Liquet B. Estimation of extended mixed models using latent classes and latent processes: The R package lcmm. J Stat Softw. 2017;78. doi:10.18637/jss.v078.i02

147. Rana J, Gutierrez PL, Oldroyd JC. Quantitative Methods. Global Encyclopedia of Public Administration, Public Policy, and Governance. Cham: Springer International Publishing; 2021. pp. 1–6.

148. [No title]. [cited 9 Aug 2024]. Available: https://www.ifiasa.com/_files/ugd/378e6f_009a90538b664a35b2314bb4f91e6d99.pdf

149. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant. 2018;52: 1893–1907.

150. Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: A systematic review of empirical tests. Soc Sci Med. 2022;292: 114523.

151. Glaser BG, Strauss AL. The Discovery of Grounded Theory: Strategies for Qualitative Research. Transaction Publishers; 2009.

152. Braun V, Clarke V. Thematic analysis. APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological. Washington: American Psychological Association; 2012. pp. 57–71.

153. Creswell JW. A Concise Introduction to Mixed Methods Research. SAGE Publications; 2014.

154. O'Cathain A, Murphy E, Nicholl J. Three techniques for integrating data in mixed methods studies. BMJ. 2010;341: c4587.

155. Skamagki G, King A, Carpenter C, Wåhlin C. The concept of integration in mixed methods research: a step-by-step guide using an example study in physiotherapy. Physiother Theory Pract. 2024;40: 197–204.

156. Guetterman TC, Fàbregues S, Sakakibara R. Visuals in joint displays to represent integration in mixed methods research: A methodological review. Meth Psychol. 2021;5: 100080.

157. WHO Coronavirus (COVID-19) Dashboard. [cited 17 Jan 2022]. Available: https://covid19.who.int

158. A clinical case definition of post COVID-19 condition by a Delphi consensus, 6 October 2021. World Health Organization; 6 Oct 2021 [cited 15 Jul 2022]. Available: https://www.who.int/publications/i/item/WHO-2019-nCoV-Post_COVID-19_condition-Clinical_case_definition-2021.1

159. Rubin R. As Their Numbers Grow, COVID-19 "Long Haulers" Stump Experts. JAMA. 2020. p. 1381. doi:10.1001/jama.2020.17709

160. Crook H, Raza S, Nowell J, Young M, Edison P. Long covid-mechanisms, risk factors, and management. BMJ. 2021;374: n1648.

161. 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; 101019.

162. Chen C, Haupert SR, Zimmermann L, Shi X, Fritsche LG, Mukherjee B. Global Prevalence of Post COVID-19 Condition or Long COVID: A Meta-Analysis and Systematic Review. J Infect Dis. 2022. doi:10.1093/infdis/jiac136

163. Townsend L, Dyer AH, Jones K, Dunne J, Mooney A, Gaffney F, et al. Persistent fatigue following SARS-CoV-2 infection is common and independent of severity of initial infection. PLoS One. 2020;15: e0240784.

164. Fernández-de-Las-Peñas C, Pellicer-Valero OJ, Navarro-Pardo E, Palacios-Ceña D, Florencio LL, Guijarro C, et al. Symptoms Experienced at the Acute Phase of SARS-CoV-2 Infection as Risk Factor of Long-term Post-COVID Symptoms: The LONG-COVID-EXP-CM Multicenter Study. Int J Infect Dis. 2022. doi:10.1016/j.ijid.2022.01.007

165. Liu T, Wu D, Yan W, Wang X, Zhang X, Ma K, et al. Twelve-month systemic consequences of COVID-19 in patients discharged from hospital: a prospective cohort study in Wuhan, China. Clin Infect Dis. 2021. doi:10.1093/cid/ciab703

166. Huang L, Yao Q, Gu X, Wang Q, Ren L, Wang Y, et al. 1-year outcomes in hospital survivors with COVID-19: a longitudinal cohort study. Lancet. 2021;398: 747–758.

167. Lemhöfer C, Sturm C, Loudovici-Krug D, Best N, Gutenbrunner C. The impact of Post-COVID-Syndrome on functioning - results from a community survey in patients after mild and moderate SARS-CoV-2-infections in Germany. J Occup Med Toxicol. 2021;16: 45.

168. Augustin M, Schommers P, Stecher M, Dewald F, Gieselmann L, Gruell H, et al. Post-COVID syndrome in non-hospitalised patients with COVID-19: a longitudinal prospective cohort study. Lancet Reg Health Eur. 2021;6: 100122.

169. 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: 1–12.

170. Determinants of persistence of symptoms and impact on physical and mental wellbeing in Long COVID: A prospective cohort study. J Infect. 2022;84: 566–572.

171. Kondratiuk AL, Pillay TD, Kon OM, Lalvani A. A conceptual framework to accelerate the clinical impact of evolving research into long COVID. Lancet Infect Dis. 2021;21: 756–757.

172. Hoffmann W, Latza U, Baumeister SE, Brünger M, Buttmann-Schweiger N, Hardt J, et al. Guidelines and recommendations for ensuring Good Epidemiological Practice (GEP): a guideline developed by the German Society for Epidemiology. Eur J Epidemiol. 2019;34: 301–317.

173. Buysse DJ, Reynolds CF 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res. 1989;28: 193–213.

174. Ninot G, Soyez F, Préfaut C. A short questionnaire for the assessment of quality of life in patients with chronic obstructive pulmonary disease: psychometric properties of VQ11. Health Qual Life Outcomes. 2013;11: 179.

175. [No title]. [cited 4 Apr 2022]. Available: https://www.has-sante.fr/upload/docs/application/pdf/2021-07/iqss_guide_proms_specifiques_bpco_2021.pdf

176. Clinical Spectrum. [cited 20 Dec 2021]. Available: https://www.covid19treatmentguidelines.nih.gov/overview/clinical-spectrum/

177. Capelle CM, Ciré S, Domingues O, Ernens I, Hedin F, Fischer A, et al. Combinatorial analysis reveals highly coordinated early-stage immune reactions that predict later antiviral immunity in mild COVID-19 patients. Cell Reports Medicine. 2022; 100600.

178. Murtagh F, Legendre P. Ward's hierarchical agglomerative clustering method: Which algorithms implement ward's criterion? J Classification. 2014;31: 274–295.

179. Website. Available: R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

180. Gómez-Rubio V. ggplot2 - Elegant Graphics for Data Analysis (2nd Edition). Journal of Statistical Software. 2017. doi:10.18637/jss.v077.b02

181. Seeßle J, Waterboer T, Hippchen T, Simon J, Kirchner M, Lim A, et al. Persistent Symptoms in Adult Patients 1 Year After Coronavirus Disease 2019 (COVID-19): A Prospective Cohort Study. Clin Infect Dis. 2022;74: 1191–1198.

182. Wynberg E, van Willigen HDG, Dijkstra M, Boyd A, Kootstra NA, van den Aardweg JG, et al. Evolution of COVID-19 symptoms during the first 12 months after illness onset. Clin Infect Dis. 2021. doi:10.1093/cid/ciab759

183. Robineau O, Wiernik E, Lemogne C, de Lamballerie X, Ninove L, Blanché H, et al. Persistent Symptoms After the First Wave of COVID-19 in Relation to SARS-CoV-2 Serology and Experience of Acute Symptoms: A Nested Survey in a Population-Based Cohort. SSRN Electronic Journal. doi:10.2139/ssrn.4016549

184. Han Q, Zheng B, Daines L, Sheikh A. Long-Term Sequelae of COVID-19: A Systematic Review and Meta-Analysis of One-Year Follow-Up Studies on Post-COVID Symptoms. Pathogens. 2022;11: 269.

185. Altmann DM, Boyton RJ. Decoding the unknowns in long covid. BMJ. 2021;372: n132.

186. Yong SJ. Long COVID or post-COVID-19 syndrome: putative pathophysiology, risk factors, and treatments. Infect Dis. 2021;53: 737–754.

187. Taquet M, Dercon Q, Luciano S, Geddes JR, Husain M, Harrison PJ. Incidence, co-occurrence, and evolution of long-COVID features: A 6-month retrospective cohort study of 273,618 survivors of COVID-19. PLoS Med. 2021;18: e1003773.

188. Al-Aly Z, Xie Y, Bowe B. High-dimensional characterization of post-acute sequelae of COVID-19. Nature. 2021;594: 259–264.

189. Holmes E, Wist J, Masuda R, Lodge S, Nitschke P, Kimhofer T, et al. Incomplete Systemic Recovery and Metabolic Phenoreversion in Post-Acute-Phase Nonhospitalized COVID-19 Patients: Implications for Assessment of Post-Acute COVID-19 Syndrome. J Proteome Res. 2021;20: 3315–3329.

190. Doykov I, Hällqvist J, Gilmour KC, Grandjean L, Mills K, Heywood WE. “The long tail of Covid-19” - The detection of a prolonged inflammatory response after a SARS-CoV-2 infection in asymptomatic and mildly affected patients. F1000Res. 2020;9: 1349.

191. Malkova A, Kudryavtsev I, Starshinova A, Kudlay D, Zinchenko Y, Glushkova A, et al. Post COVID-19 Syndrome in Patients with Asymptomatic/Mild Form. Pathogens. 2021;10: 1408.

192. Pfaff E, Madlock-Brown C, Baratta JM, Bhatia A, Davis H, Girvin AT, et al. Coding Long COVID: Characterizing a new disease through an ICD-10 lens. medRxiv. 2022; 2022.04.18.22273968.

193. Moreno-Pérez O, Merino E, Leon-Ramirez JM, Andres M, Ramos JM, Arenas-Jiménez J, et al. Post-acute COVID-19 syndrome. Incidence and risk factors: A Mediterranean cohort study. J Infect. 2021;82. doi:10.1016/j.jinf.2021.01.004

194. Sudre CH, Murray B, Varsavsky T, Graham MS, Penfold RS, Bowyer RC, et al. Attributes and predictors of long COVID. Nat Med. 2021;27: 626–631.

195. Blackett JW, Li J, Jodorkovsky D, Freedberg DE. Prevalence and risk factors for gastrointestinal symptoms after recovery from COVID-19. Neurogastroenterol Motil. 2022;34: e14251.

196. Meringer H, Mehandru S. Gastrointestinal post-acute COVID-19 syndrome. Nat Rev Gastroenterol Hepatol. 2022;19: 345–346.

197. Kuodi P, Gorelik Y, Zayyad H, Wertheim O, Wiegler KB, Jabal KA, et al. Association between

vaccination status and reported incidence of post-acute COVID-19 symptoms in Israel: a cross-sectional study of patients infected between March 2020 and November 2021. doi:10.1101/2022.01.05.22268800

198. Taquet M, Dercon Q, Harrison PJ. Six-month sequelae of post-vaccination SARS-CoV-2 infection: a retrospective cohort study of 10,024 breakthrough infections. medRxiv. 2021; 2021.10.26.21265508.

199. Ballering AV, van Zon SKR, Olde Hartman TC, Rosmalen JGM, Lifelines Corona Research Initiative. Persistence of somatic symptoms after COVID-19 in the Netherlands: an observational cohort study. Lancet. 2022;400: 452–461.

200. Coronavirus disease (COVID-19): Post COVID-19 condition. [cited 21 Oct 2022]. Available: https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-(covid-19)-post-covid-19-condition

201. A clinical case definition of post COVID-19 condition by a Delphi consensus, 6 October 2021. World Health Organization; 6 Oct 2021 [cited 29 Sep 2022]. Available: https://www.who.int/publications/i/item/WHO-2019-nCoV-Post_COVID-19_condition-Clinical_case_definition-2021.1

202. Halpin SJ, McIvor C, Whyatt G, Adams A, Harvey O, McLean L, et al. Postdischarge symptoms and rehabilitation needs in survivors of COVID-19 infection: A cross-sectional evaluation. J Med Virol. 2021;93. doi:10.1002/jmv.26368

203. Garrigues E, Janvier P, Kherabi Y, Le Bot A, Hamon A, Gouze H, et al. Post-discharge persistent symptoms and health-related quality of life after hospitalization for COVID-19. J Infect. 2020;81. doi:10.1016/j.jinf.2020.08.029

204. Tenforde MW, Kim SS, Lindsell CJ, Billig RE, Shapiro NI, Files DC, et al. Symptom Duration and Risk Factors for Delayed Return to Usual Health Among Outpatients with COVID-19 in a Multistate Health Care Systems Network - United States, March-June 2020. MMWR Morb Mortal Wkly Rep. 2020;69. doi:10.15585/mmwr.mm6930e1

205. Ziauddeen N, Gurdasani D, O'Hara ME, Hastie C, Roderick P, Yao G, et al. Characteristics and impact of Long Covid: Findings from an online survey. PLoS One. 2022;17. doi:10.1371/journal.pone.0264331

206. A clinical case definition of post COVID-19 condition by a Delphi consensus, 6 October 2021. World Health Organization; 6 Oct 2021 [cited 2 Aug 2022]. Available: https://www.who.int/publications/i/item/WHO-2019-nCoV-Post_COVID-19_condition-Clinical_case_definition-2021.1

207. Amenta EM, Spallone A, Rodriguez-Barradas MC, El Sahly HM, Atmar RL, Kulkarni PA. Postacute COVID-19: An Overview and Approach to Classification. Open Forum Infectious Diseases. 2020;7. doi:10.1093/ofid/ofaa509

208. Fischer A, Zhang L, Elbéji A, Wilmes P, Oustric P, Staub T, et al. Long COVID Symptomatology After 12 Months and Its Impact on Quality of Life According to Initial Coronavirus Disease 2019 Disease Severity. Open Forum Infect Dis. 2022;9: ofac397.

209. Clinical Spectrum. In: COVID-19 Treatment Guidelines [Internet]. [cited 3 Aug 2022]. Available: https://www.covid19treatmentguidelines.nih.gov/overview/clinical-spectrum/

210. Website. Available: R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

211. Morioka S, Tsuzuki S, Suzuki M, Terada M, Akashi M, Osanai Y, et al. Post COVID-19 condition of the Omicron variant of SARS-CoV-2. J Infect Chemother. 2022;28: 1546–1551.

212. The Lancet. Long COVID: 3 years in. Lancet. 2023;401: 795.

213. Impact of long COVID-19 on work: a co-produced survey. Lancet. 2023;402: S98.

214. [No title]. [cited 19 Apr 2024]. Available: https://scholar.harvard.edu/files/cutler/files/long_covid_update_7-22.pdf

215. Website. Available: https://world.physio/sites/default/files/2021-07/Briefing-Paper-9-Long-Covid-FINAL-English-2021_0

216. COVID-19 vaccination for the prevention and treatment of long COVID: A systematic review and meta-analysis. Brain Behav Immun. 2023;111: 211–229.

217. Zhang H, Zang C, Xu Z, Zhang Y, Xu J, Bian J, et al. Data-driven identification of post-acute SARS-CoV-2 infection subphenotypes. Nat Med. 2023;29: 226–235.

218. Fischer A, Badier N, Zhang L, Elbéji A, Wilmes P, Oustric P, et al. Long COVID Classification: Findings from a Clustering Analysis in the Predi-COVID Cohort Study. Int J Environ Res Public Health. 2022;19. doi:10.3390/ijerph192316018

219. Website. Available: http://www.r-project.org/index.htm

220. [No title]. [cited 4 Apr 2022]. Available: https://www.has-sante.fr/upload/docs/application/pdf/2021-07/iqss_guide_proms_specifiques_bpco_2021.pdf

221. Merkes CM. Generic qPCR limit of detection (LOD) / limit of quantification (LOQ) calculator. U.S. Geological Survey; 2019. doi:10.5066/P9GT00GB

222. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem. 2009;55: 611–622.

223. Bowe B, Xie Y, Al-Aly Z. Postacute sequelae of COVID-19 at 2 years. Nat Med. 2023;29: 2347–2357.

224. Harris E. Some People Still Have Long COVID Symptoms After 2 Years. JAMA. 2023;330: 1127–1127.

225. Mazza MG, Palladini M, Villa G, De Lorenzo R, Rovere Querini P, Benedetti F. Prevalence, trajectory over time, and risk factor of post-COVID-19 fatigue. J Psychiatr Res. 2022;155: 112–119.

226. Mateu L, Tebe C, Loste C, Santos JR, Lladós G, López C, et al. Determinants of the onset and prognosis of the post-COVID-19 condition: a 2-year prospective observational cohort study. Lancet Reg Health Eur. 2023;33: 100724.

227. Smith P, De Pauw R, Van Cauteren D, Demarest S, Drieskens S, Cornelissen L, et al. Post COVID-19 condition and health-related quality of life: a longitudinal cohort study in the Belgian adult population. BMC Public Health. 2023;23: 1433.

228. Kuri-Ayache M, Rivera-Cavazos A, Pérez-Castillo MF, Santos-Macías JE, González-Cantú A, Luviano-García JA, et al. Viral load and its relationship with the inflammatory response and clinical outcomes in hospitalization of patients with COVID-19. Front Immunol. 2022;13: 1060840.

229. Abdulrahman A, Mallah SI, Alqahtani M. COVID-19 viral load not associated with disease severity: findings from a retrospective cohort study. BMC Infect Dis. 2021;21: 688.

230. Girón Pérez DA, Fonseca-Agüero A, Toledo-Ibarra GA, Gomez-Valdivia J de J, Díaz-Resendiz KJG, Benitez-Trinidad AB, et al. Post-COVID-19 Syndrome in Outpatients and Its Association with Viral Load. Int J Environ Res Public Health. 2022;19. doi:10.3390/ijerph192215145

231. Fischer A, Aguayo GA, Oustric P, Morin L, Larche J, Benoy C, et al. Co-Design of a Voice-Based Digital Health Solution to Monitor Persisting Symptoms Related to COVID-19 (UpcomingVoice Study): Protocol for a Mixed Methods Study. JMIR Res Protoc. 2023;12: e46103.

232. Zhang H, Huang C, Gu X, Wang Y, Li X, Liu M, et al. 3-year outcomes of discharged survivors of COVID-19 following the SARS-CoV-2 omicron (B.1.1.529) wave in 2022 in China: a longitudinal cohort study. Lancet Respir Med. 2024;12: 55–66.

233. Ballering AV, van Zon SKR, Olde Hartman TC, Rosmalen JGM, Lifelines Corona Research Initiative. Persistence of somatic symptoms after COVID-19 in the Netherlands: an observational cohort study. Lancet. 2022;400: 452–461.

234. Davis HE, McCorkell L, Vogel JM, Topol EJ. Long COVID: major findings, mechanisms and recommendations. Nat Rev Microbiol. 2023;21: 133–146.

235. A clinical case definition of post COVID-19 condition by a Delphi consensus, 6 October 2021. World Health Organization; 6 Oct 2021 [cited 15 Jul 2022]. Available: https://www.who.int/publications/i/item/WHO-2019-nCoV-Post_COVID-19_condition-Clinical_case_definition-2021.1

236. [No title]. [cited 21 Mar 2024]. Available: https://www.gov.wales/sites/default/files/publications/2023-07/science-evidence-advice-the-social-cost-of-a-covid-infection-october-2022.pdf

237. Hung P, Brehon K, Miciak M, Brown DA, Bostick G, Brown C, et al. “I had to know about it, I had to find it, I had to know how to access it”: Experiences of access to rehabilitation services among people living with Long COVID. Physiother Can. 2024;76: 8–24.

238. Adams N, MacIver E, Grant A, Torrance N, Douglas F, Skåtun D, et al. OP32 *Constrained Candidacy*:Exploring different barriers to attaining healthcare access and treatment for long COVID illness by NHS workers in Scotland. SSM Annual Scientific Meeting. BMJ Publishing Group Ltd; 2023. doi:10.1136/jech-2023-ssmabstracts.32

239. Brehon K, Miciak M, Hung P, Chen S-P, Perreault K, Hudon A, et al. “None of us are lying”: an interpretive description of the search for legitimacy and the journey to access quality health services by individuals living with Long COVID. BMC Health Serv Res. 2023;23: 1396.

240. [No title]. [cited 3 Nov 2023]. Available:

https://world.physio/sites/default/files/2021-07/Briefing-Paper-9-Long-Covid-FINAL-English-2021_0.pdf

241. Kumah-Crystal YA, Pirtle CJ, Whyte HM, Goode ES, Anders SH, Lehmann CU. Electronic Health Record Interactions through Voice: A Review. Appl Clin Inform. 2018;9: 541–552.

242. Bérubé C, Schachner T, Keller R, Fleisch E, Wangenheim F V, Barata F, et al. Voice-Based Conversational Agents for the Prevention and Management of Chronic and Mental Health Conditions: Systematic Literature Review. J Med Internet Res. 2021;23: e25933.

243. Nayak A, Vakili S, Nayak K, Nikolov M, Chiu M, Sosseinheimer P, et al. Use of Voice-Based Conversational Artificial Intelligence for Basal Insulin Prescription Management Among Patients With Type 2 Diabetes: A Randomized Clinical Trial. JAMA Netw Open. 2023;6: e2340232.

244. Utilizing speech analysis to differentiate progressive supranuclear palsy from Parkinson's disease. Parkinsonism Relat Disord. 2023;115: 105835.

245. Home. In: Real Time Voice Analyzer [Internet]. 23 Apr 2023 [cited 3 Jun 2024]. Available: https://realtimevoiceanalyzer.com/

246. Larsen E, Murton O, Song X, Joachim D, Watts D, Kapczinski F, et al. Validating the efficacy and value proposition of mental fitness vocal biomarkers in a psychiatric population: prospective cohort study. Front Psychiatry. 2024;15: 1342835.

247. Website. Available: Fagherazzi G, Zhang L, Elbéji A, Higa E, Despotovic V, Ollert M, et al. (2022) A voice-based biomarker for monitoring symptom resolution in adults with COVID-19: Findings from the prospective Predi-COVID cohort study. PLOS Digit Health 1(10): e0000112. https://doi.org/10.1371/journal.pdig.0000112

248. Visible - Activity tracking for Long Covid and ME/CFS. [cited 18 Oct 2023]. Available: https://www.makevisible.com/

249. Covid Recovery - Living With. In: Living With - Smart condition management and medical health apps [Internet]. Living With; 31 Oct 2022 [cited 18 Oct 2023]. Available: https://livingwith.health/covid-recovery/

250. Fareed N, Swoboda C, Singh P, Boettcher E, Wang Y, Venkatesh K, et al. Developing and testing an integrated patient mHealth and provider dashboard application system for type 2 diabetes management among Medicaid-enrolled pregnant individuals based on a user-centered approach: Mixed-methods study. Digit Health. 2023;9: 20552076221144181.

251. Lazarevic N, Pizzuti C, Rosic G, Bœhm C, Williams K, Caillaud C. A mixed-methods study exploring women's perceptions and recommendations for a pregnancy app with monitoring tools. NPJ Digit Med. 2023;6: 50.

252. Adu MD, Malabu UH, Malau-Aduli AEO, Malau-Aduli BS. Users' preferences and design recommendations to promote engagements with mobile apps for diabetes self-management: Multi-national perspectives. PLoS One. 2018;13: e0208942.

253. Søholm U, Broadley M, Zaremba N, Divilly P, Nefs G, Mahmoudi Z, et al. Investigating the day-to-day impact of hypoglycaemia in adults with type 1 or type 2 diabetes: design and validation protocol of the Hypo-METRICS application. BMJ Open. 2022;12: e051651.

254. Slattery P, Saeri AK, Bragge P. Research co-design in health: a rapid overview of reviews.

Health Res Policy Syst. 2020;18: 1–13.

255. Aguayo GA, Goetzinger C, Scibilia R, Fischer A, Seuring T, Tran VT, et al. Methods to Generate Innovative Research Ideas and Improve Patient and Public Involvement in Modern Epidemiological Research: Review, Patient Viewpoint, and Guidelines for Implementation of a Digital Cohort Study. J Med Internet Res. 2021;23. doi:10.2196/25743

256. Apresj20 : Association Covid Long France. In: apresj20 [Internet]. [cited 10 Mar 2024]. Available: www.apresj20.fr

257. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and Validation of the User Version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth. 2016;4: e72.

258. Hajesmaeel-Gohari S, Khordastan F, Fatehi F, Samzadeh H, Bahaadinbeigy K. The most used questionnaires for evaluating satisfaction, usability, acceptance, and quality outcomes of mobile health. BMC Med Inform Decis Mak. 2022;22: 22.

259. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform. 2019;95: 103208.

260. [No title]. [cited 3 Mar 2024]. Available: https://www.ifiasa.com/_files/ugd/378e6f_009a90538b664a35b2314bb4f91e6d99.pdf

261. [No title]. [cited 3 Mar 2024]. Available: https://www.ifiasa.com/_files/ugd/378e6f_009a90538b664a35b2314bb4f91e6d99.pdf

262. Website. Available: R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

263. Creswell JW, Plano Clark VL. Designing and Conducting Mixed Methods Research. SAGE Publications; 2017.

264. Aguirre Vergara F, Fischer A, Seuring T, de Beaufort C, Fagherazzi G, Aguayo GA. Mixed-methods study protocol to identify expectations of people with type 1 diabetes and their caregivers about voice-based digital health solutions to support the management of diabetes distress: the PsyVoice study. BMJ Open. 2023;13: e068264.

265. Choi BCK, Pak AWP. A catalog of biases in questionnaires. Prev Chronic Dis. 2005;2: A13.

266. Six SG, Byrne KA, Aly H, Harris MW. The Effect of Mental Health App Customization on Depressive Symptoms in College Students: Randomized Controlled Trial. JMIR Ment Health. 2022;9: e39516.

267. Adams JR, Ray D, Willmon R, Pulgar S, Dasari A. Living With Neuroendocrine Tumors: Assessment of Quality of Life Through a Mobile Application. JCO Clin Cancer Inform. 2019;3: 1–10.

268. Evald L, Graarup J, Højskov IE. Diary for self-observation: A self-management tool for recipients of lung transplantation-A pilot study. Nurs Open. 2020;7: 1766–1773.

269. Coghlan D, Brydon-Miller M. The SAGE encyclopedia of action research. 2455 Teller Road, Thousand Oaks, California 91320: SAGE Publications Ltd; 2014. doi:10.4135/9781446294406

270. Schellekens MPJ, Wolvers MDJ, Schroevers MJ, Bootsma TI, Cramer AOJ, van der Lee ML. Exploring the interconnectedness of fatigue, depression, anxiety and potential risk and protective factors in cancer patients: a network approach. J Behav Med. 2020;43: 553–563.

271. Goërtz YMJ, Braamse AMJ, Spruit MA, Janssen DJA, Ebadi Z, Van Herck M, et al. Fatigue in patients with chronic disease: results from the population-based Lifelines Cohort Study. Sci Rep. 2021;11: 20977.

272. Minton O, Berger A, Barsevick A, Cramp F, Goedendorp M, Mitchell SA, et al. Cancer-related fatigue and its impact on functioning. Cancer. 2013;119 Suppl 11: 2124–2130.

273. Skinner TC, Joensen L, Parkin T. Twenty-five years of diabetes distress research. Diabet Med. 2020;37: 393–400.

274. The European Union Medical Device Regulation. [cited 1 Mar 2024]. Available: https://eumdr.com/

275. Sherman KA, Pehlivan MJ, Singleton A, Hawkey A, Redfern J, Armour M, et al. Co-design and Development of EndoSMS, a Supportive Text Message Intervention for Individuals Living With Endometriosis: Mixed Methods Study. JMIR Form Res. 2022;6: e40837.

276. Becker R. Gender and survey participation an event history analysis of the gender effects of survey participation in a probability-based multi-wave panel study with a sequential mixed-mode design. 2017. doi:10.12758/MDA.2021.08

277. [No title]. [cited 29 Apr 2024]. Available: https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32017R0745

Annexes

Annexe 1 : Article au format point de vue sur l'utilisation des biomarqueurs vocaux pour le suivi à distance de symptômes liés à COVID-19 et au COVID Long

Recommendations for Successful Implementation of the Use of Vocal Biomarkers for Remote Monitoring of COVID-19 and Long COVID in Clinical Practice and Research

Aurelie Fischer[1] ; Abir Elbeji[1] ; Gloria Aguayo[1] ; Guy Fagherazzi[1]

Abstract

The COVID-19 pandemic accelerated the use of remote patient monitoring in clinical practice or research for safety and emergency reasons, justifying the need for innovative digital health solutions to monitor key parameters or symptoms related to COVID-19 or Long COVID. The use of voice-based technologies, and in particular vocal biomarkers, is a promising approach, voice being a rich, easy-to-collect medium with numerous potential applications for health care, from diagnosis to monitoring. In this viewpoint, we provide an overview of the potential benefits and limitations of using voice to monitor COVID-19, Long COVID, and related symptoms. We then describe an optimal pipeline to bring a vocal biomarker candidate from research to clinical practice and discuss recommendations to achieve such a clinical implementation successfully.

Interact J Med Res 2022;11(2):e40655

doi:10.2196/40655

Keywords

vocal biomarker (3); COVID-19 symptoms (3); digital health (1695); remote monitoring (134); artificial intelligence (922); voice (16); COVID-19 (2776); Long COVID (30); digital health solution (8); voice-based technology (1); health technology (117); health monitoring (25); digital health monitoring (1); health care application (4); remote patient monitoring (46)

Introduction

Management of COVID-19 and Long COVID

Since February 2020, the COVID-19 pandemic has mobilized the entire research, medical, and pharmaceutical community for treatment and vaccine development as well as care and disease management of patients with COVID-19. Many countries adopted strategic measures to fight the pandemic, such as lockdowns, social distancing, or contact tracing, to reduce transmission [

1]. In parallel, we observed rapid a development of digital tools to support the health care system and an increased use of telemedicine and teleconsultations. COVID-19 infection is caused by the SARS-CoV-2 virus and can take several forms, from asymptomatic to moderate or severe illness. At the onset of the illness, common symptoms are fever, cough, dyspnea, myalgia fatigue, loss of taste or smell, and sometimes gastrointestinal symptoms. Complications include acute respiratory distress syndrome, anemia, or acute cardiac injury [2].

After the acute phase, it has been estimated that up to 68% of patients with COVID-19 will have persisting symptoms after 6 months, and 49% after a year [3]. Long COVID syndrome is defined as a set of symptoms persisting and fluctuating beyond 4 weeks after infection, leading to long-term support and care needs [4]. Long COVID is multiorgan and can affect lungs, heart, brain, kidney, and blood vessels. Long COVID also affects cognitive functions and mental health with diverse symptoms such as brain fog [5], confusion, memory deficiencies, and even posttraumatic stress disorder [6]. The most frequently reported symptoms are fatigue, abnormalities of smell and taste, anxiety, sleeplessness, and dyspnea [4]. The severity of the acute illness seems not to be directly related to the development of Long COVID, and many people with Long COVID did not return to the same level of work and quality of life as before COVID-19 infection [7]. The absence of treatment and the specificities of Long COVID in terms of symptom types and fluctuation over time justifies the need to develop solutions for objective and qualitative symptom monitoring rapidly. As such, a panel of experts from the National Institute for Health and Care Excellence recently recommended developing telemonitoring and encouraging self-management of acute and Long COVID symptoms in a tailored and accessible way for each patient [8]. In particular, at-home self-monitoring is encouraged [9].

From Traditional to Enhanced Approaches of Patient Monitoring

The medical care of patients in real-life situations has been rapidly evolving since the pandemic started. Video or phone consultations have largely replaced traditional visits to the family doctor or specialist and have shown the utility [10] of these formats to follow up patients who are not able to travel or are located in geographic regions lacking medical doctors. In parallel, the use of medical devices for self-monitoring of physiological parameters such as blood pressure or blood glucose level has also increased. The next development step would be to unify and standardize

telemedicine solutions with enhanced teleconsultations, including a real-time assessment of key physiological parameters. This would optimize the consultation time and enable a personalized follow-up with increased communication between the health care professionals and the patients.

Regarding clinical research, there was already, before the pandemic, a global trend toward completely remote, decentralized clinical trials [11,12]. The pandemic has now accelerated such digitization. Participant monitoring in clinical trials is progressively moving away from a paper-based approach (ie, paper informed consent, questionnaires, and medical records) to fully remote or hybrid setups, where patients' visits on-site and remote, 'in-between visits,' and follow-up in real-life are combined. Digitization and decentralization of clinical trials encompass three axes: digital participant recruitment and retention, digital data collection, and digital analytics [13], based on electronic documents (eg, e-consent and electronic case report forms), virtual study visits, and physical self-measurements. Digital data collection facilitates and standardizes data quality, whereas digital analytics using artificial intelligence techniques like machine and deep learning methods allow for deep phenotyping of participants.

Digital Technologies for the Clinical Management of Patients With COVID-19

Technologies already exist to facilitate remote patient monitoring in clinical practice or research, such as electronic patient-reported outcomes, sensors or devices to measure physiological parameters at home (eg, Holter for electrocardiograms, blood pressure, and heart rate), video-based methods, or mobile phone–based remote symptom monitoring systems [14]. In the context of COVID-19, hundreds of contact tracing mobile apps and artificial intelligence–based radiological technologies to facilitate early detection of COVID-19 emerged early during the pandemic [15]. In parallel, several digital technologies have been developed to respond to different patient needs, including diagnosis, prevention, treatment, adherence, lifestyle, or patient engagement [16]. For example, a smartwatch application has been developed in Germany to help COVID-19 diagnosis based on a few vital signs [17], and 2 remote monitoring systems—Telecare-COVID (based on phone calls) and CareSimple-COVID (a telemonitoring app)—are used in Canada and are well accepted by the users [18].

Voice assistants have also been identified as an innovative tool for health care services in the context of a pandemic, as a tool for health information exchange, for remote monitoring or to maintain continuous care with teleconsultations [19].

Voice is an easy-to-collect source of information, requiring less time than completing a questionnaire, being noninvasive, and inducing less burden for patients or study participants. The first use of voice dates back to 2003 in clinical studies and health care services with interactive voice response system [20], a phone-based system mainly used for patient randomization.

Interactive voice response system was also used for symptom monitoring by calling participants and encouraging them to answer questions on their health status and in return, patients could receive specific recommendations to manage their treatment. However, voice is in itself a rich medium providing information on health status and emotions, allowing for a richer characterization of patients through the use of so-called vocal biomarkers. This opens many perspectives of using voice besides the practical aspect of using it as a collection tool.

As we are now at a turning point in telemedicine, we believe that the use of vocal biomarkers is among the most promising approaches to improve patient monitoring of COVID-19–related symptoms.

In the following sections, we provide an extensive description of the potential benefit and limitations as well as recommendations for the development of a digital health solution based on vocal biomarkers. Since the pandemic revealed specific needs, these recommendations are elaborated in the COVID-19 context but can be easily generalized to other diseases or symptoms.

Using Vocal Biomarkers for Remote Symptom Monitoring in the Future

What is a Vocal Biomarker?

As mentioned before, voice is a rich medium, characterized by thousands of different features, potentially affected by our health status. Thus, a vocal biomarker is an extension of a classical biomarker, a factor objectively measured and evaluated representing a biological or pathogenic process or a pharmacological response to a therapeutic intervention [21]. It can also be used as a surrogate marker of a clinical end point. A vocal biomarker can therefore be defined as a signature, a feature, or a combination of features from the audio signal of the voice associated with a clinical outcome. It must have all the properties of a traditional biomarker, needing to be analytically validated and qualified using an evidentiary assessment. A vocal biomarker can be used to monitor patients, diagnose a condition, or grade the severity of a disease [21].

Vocal biomarkers have already been described in pathologies such as Parkinson disease, depression, and cardiovascular diseases with the potential of early diagnosis or disease progression markers [21], but none of them is used in clinical practice yet. Since voice features can be specifically associated with these different pathologies, one can extrapolate that similar voice features could be associated with a COVID-19 infection or with a consequence of COVID-19. COVID-19 infection or complications can affect voice through different mechanisms. For example, respiratory insufficiency can lead to reduced airflow, and therefore, changes in voice

parameters [22]. Other studies showed that voice quality was reduced in patients with COVID-19 due to repeated cough, laryngeal or pharyngeal erythema, or sore throat [22-24].

Potential Benefits and Limitations of Using Vocal Biomarkers in the Context of COVID-19

The first developments based on the use of vocal biomarkers in the context of COVID-19 were meant to enable the detection of COVID-19 infection. Vocal biomarkers for COVID-19 detection in cough and voice have been developed and could one day serve as a screening tool on a very large scale and in a short period of time, for example, at airports or border controls, leading to a direct benefit for the pandemic management [25,26].

Another benefit of vocal biomarkers to monitor COVID-19–related symptoms remotely is a reduced burden for the user by limiting the clinical visits on-site, replacing tedious questionnaires and physical examinations, and facilitating the reporting of symptoms or adverse events [27]. It could also allow for simultaneous monitoring of several COVID-19–related symptoms, with early detection of a worsening in the health or mental condition, or on the other hand, serve as a proxy to assess treatment or rehabilitation effectiveness. All of these benefits combined could in turn lead to reduced risk of hospitalization and increased quality of life.

From a clinician or a researcher's perspective, the use of vocal biomarkers can facilitate and objectivize patient evaluation, in particular when the results are transmitted by a visualization tool that might be easier to interpret than questionnaires. Vocal biomarkers could also serve as a proxy to assess the benefits of a rehabilitation program for people with Long COVID. As the collection of voice recordings is fast, fun, and limits the burden for patients, it could also reduce attrition in clinical trials. Lastly, voice collection reduces costs and allows for fast and high-volume recruitment in trials by helping the inclusion of patients unable to travel or with mobility issues, improving participant representativeness, and reducing selection bias. Participants' engagement should also be increased and limit attrition in the studies by involving them in the management of their pathology.

Integrating vocal biomarkers in care would also facilitate communication between patients and medical teams thanks to better follow-up and medical care; particularly in a pandemic, it would limit contacts and infection risks. Coupling vocal biomarkers with alert systems could improve patient care and safety. The inclusion of voice analysis in health calls or emergency centers would enable augmented consultations, more accurate caller authentication, and real-time analysis of important health-related features [21].

We believe that increased use of voice in the future will maximize the benefits for both the investigators and the study participants, thanks to the combination of the best of both traditional

and digital approaches; it will ultimately increase the quality of the studies and saves time and costs. Besides, both for clinical practice and clinical research, there is a need to avoid the 'in-between clinical visits black hole' and to describe better what is happening to a patient between two follow-up visits. Telemonitoring solutions based on vocal biomarker monitoring could allow for a more accurate follow-up and complement on-site evaluations [28].

However, some researchers have challenged the relevance of a vocal biomarker for COVID-19 detection [29] and raised the issue of whether it is an actual marker of the disease or a proxy of the general health status, or worse, a proxy of the context of recording of the individuals.

Other limitations to the use of vocal biomarkers for remote monitoring of COVID-19–related symptoms include patients' acceptability and readiness of the health care system [19] for this new technology. The health status could also be a limitation, in the way that persons experiencing severe symptoms could be too affected to be willing to do the voice recording regularly, and therefore, affect adherence to the digital solution.

Development of a Digital Health Solution Based on Vocal Biomarkers to Monitor COVID-19–Related Symptoms

Identification of the Outcome to Be Monitored

Many previously cited COVID-19–related symptoms could theoretically be monitored using voice, including fatigue, dyspnea, loss of taste or smell, disease severity, presence or absence of symptoms, as well as impact on mental health (eg, stress, anxiety, and depression). However, the choice of monitoring one of these symptoms should be discussed with health care professionals to ensure its clinical relevance.

Potential clinical applications of vocal biomarkers for these symptoms are presented in Figure 1.

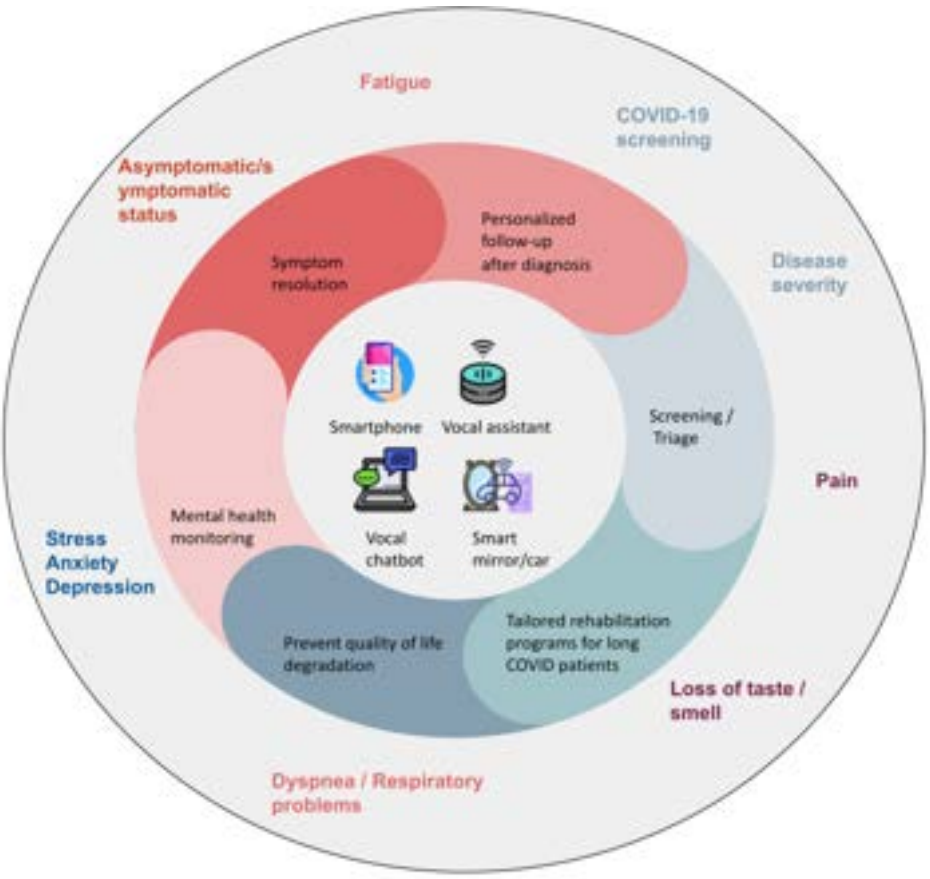


Figure 1. Vocal biomarkers could be implemented in devices, such as smartphone apps, chatbots, smart mirrors or cars, and vocal assistants, to monitor symptom resolution (with a vocal biomarkers of symptomatic or asymptomatic status), mental health, and quality of life degradation (with vocal biomarkers of stress, anxiety, depression, chronic fatigue, or dyspnea); to propose personalized follow-up after diagnosis (with vocal biomarkers of disease severity); to perform a screening and a triage of patients at hospital (with vocal biomarkers of disease severity, pain, or loss of taste and smell); and to propose tailored rehabilitation programs for patients with long COVID (using vocal biomarkers of loss of taste and smell, pain, dyspnea, and fatigue).

Data Collection

This step consists of voice data collection coupled with well-documented clinical data in screening platforms such as Colive Voice [30] or large prospective cohort studies [2,31]. The collected data have to be diverse enough and should represent the target population in terms of languages, accents, and socioeconomic backgrounds to decrease the risk of systemic biases and the risk of increasing a potential preexisting digital and socioeconomic divide in the population.

Different types of voice records, such as vowel phonation, reading a predefined text, counting, or semispontaneous voice tasks, as well as nonverbal vocalization (eg, coughing and breathing) can be collected depending on the future foreseen clinical application. The environment and conditions of voice recordings are critical, in particular in COVID-19 settings. Indeed, background noises and audio features may differ between COVID-19–positive and control participants due to the isolation of COVID-19–positive persons. For this reason, clear instructions to perform the voice recordings have to be provided before data collection.

Depending on the final use case, at home or at the hospital, recordings can be performed under either controlled conditions with high-quality microphones and standard processes or in real-life conditions with the patients' smartphones. As mentioned before, the recording situation may impact audio quality but can also allow for the training of more relevant algorithms based on more diverse data sets. Wearing a surgical mask has been shown to have no impact on vocal parameters, such as vocal intensity, jitter, shimmer, and harmonics-to-noise ratio [32]; voice collection can thus be performed safely at the hospital or during clinical visits. The validity of the data set can also be a concern, as the COVID-19 status may be self-reported [33] and may induce some mislabeling of the audio.

From Voice Recording Toward Vocal Biomarkers

After several preprocessing steps on the raw audio signals, the identification of vocal biomarkers candidates is based on machine and deep learning methods, such as support machine vectors, random forest, and visual geometry group, among many others, which can be supervised or unsupervised. When a limited data set is available, an alternative is to use transfer-learning methods. The interest of this method is to take advantage of a pretraining of the algorithm on a large data set from another domain and to fine-tune it for the defined target. Internal and external validations are then required in other settings and using other data sets.

At this stage, the vocal biomarker candidate has also to be clinically validated in one or several clinical studies in comparison to the gold standard measurement of the outcome of interest (eg, validated scales to assess fatigue or stress or, if available, established physiological parameters such as blood pressure or glycemia). The design of the studies has to be chosen carefully, going from a very standardized double-blind study to a real-life prospective study.

Recommendations for the 3 steps for the evaluation of vocal biomarkers (ie, verification of audio quality, analytical validation, and clinical validation) have been provided by Robin et al [34].

Embedding the Vocal Biomarker in a Digital Health Solution

The next step after identification and validation of a promising vocal biomarker candidate is to design a digital health solution to embed it. Several digital devices can be imagined, such as smartphone apps, chatbots, smart mirrors, or voice assistants. The future end users of the device, namely the patients and the health care professionals, should be involved in the co-design of the final solution [35,36] to ensure it meets their needs and expectations. This is particularly important for voice-based technologies to ensure their future acceptability. Finally, feedback loops should be implemented to improve both the solution and the algorithm through lessons learned in population studies.

Particular caution should be taken when collecting voice; indeed, voice is considered as identifying and sensitive data, and its collection falls under different regulations or laws, such as General Data Protection Regulations [37] in Europe and Personal Information Protection and Electronic Documents Act in Canada. In the United States, there is no single data protection law but rather multiple laws enacted at a federal or state level. These different laws do not protect individuals at the same level, and to minimize future risks for the use of the digital health solution, it is highly recommended to obtain explicit consent prior to collection. Measures have to be implemented by design and by default to securely process voice without privacy leakage (eg, encryption of voice data, splitting data into random components, or using data representations from which sensitive identifiable information is removed).

Once the digital health solution is developed, an additional validation step is mandatory to prove clinical benefit, effectiveness, and security in clinical trials. Indeed, most digital health solutions fall under the new Medical Device Regulation [38]. CE marking or Food and Drug Administration (FDA) certification will be mandatory to bring the solution to the market, and requirements for clinical evaluation are diverse, depending on the final device. The definition of the ‘intend-to-use’ (article 2 of Medical Device Regulation) of the device should be done early in the development process to define clinical evaluation requirements. It is highly recommended to consult guidance documents for digital health interventions (eg, MEDDEV 2.1 [39], FDA’s benefit-risk framework for medical devices [40], and the World Health Organization’s monitoring and evaluating digital health interventions [41]) and to take advice from regulatory authorities or a notified body.

Requests for reimbursement of the device or solution can be made to national health insurance funds after the clinical and economic interest of the new digital system is proved.

An overview of the pipeline to develop a digital remote monitoring solution based on vocal biomarkers is presented in

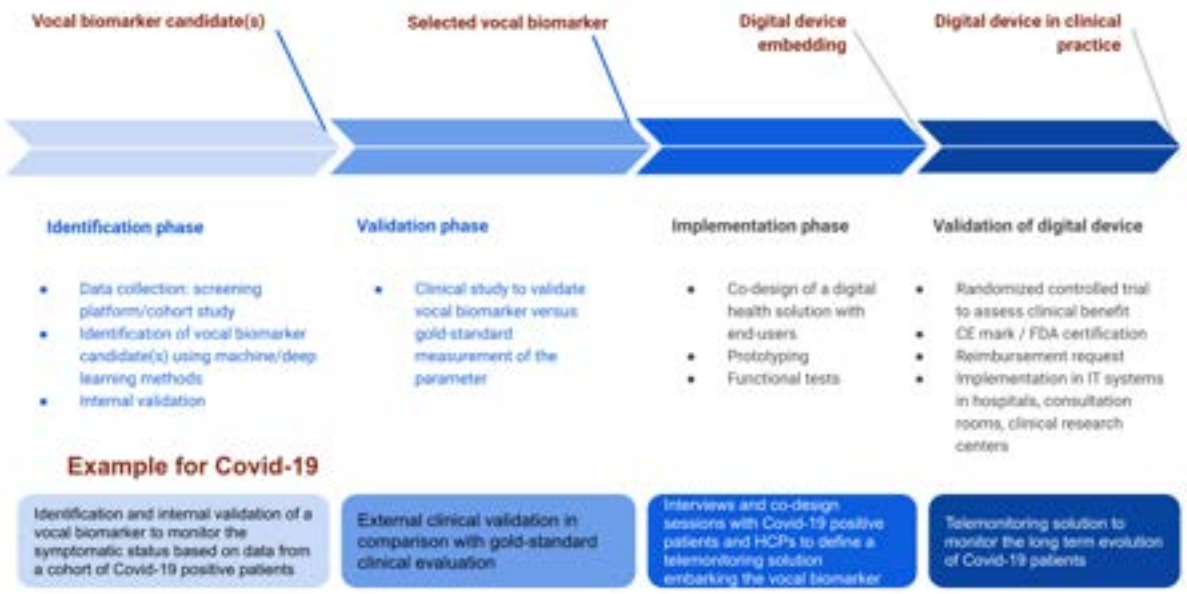


Figure 2.

Finally, as mentioned above, some vocal biomarkers have been identified in several pathologies, but none of them are currently used in clinical or real-life practice; indeed, the field of vocal biomarkers is recent, and the way is still long until a health solution based on them can be commercialized. Companies are currently in the process of requesting FDA authorization or CE mark but are facing challenges related to data security, ethical issues, as well as reliability and reproducibility of the algorithms.

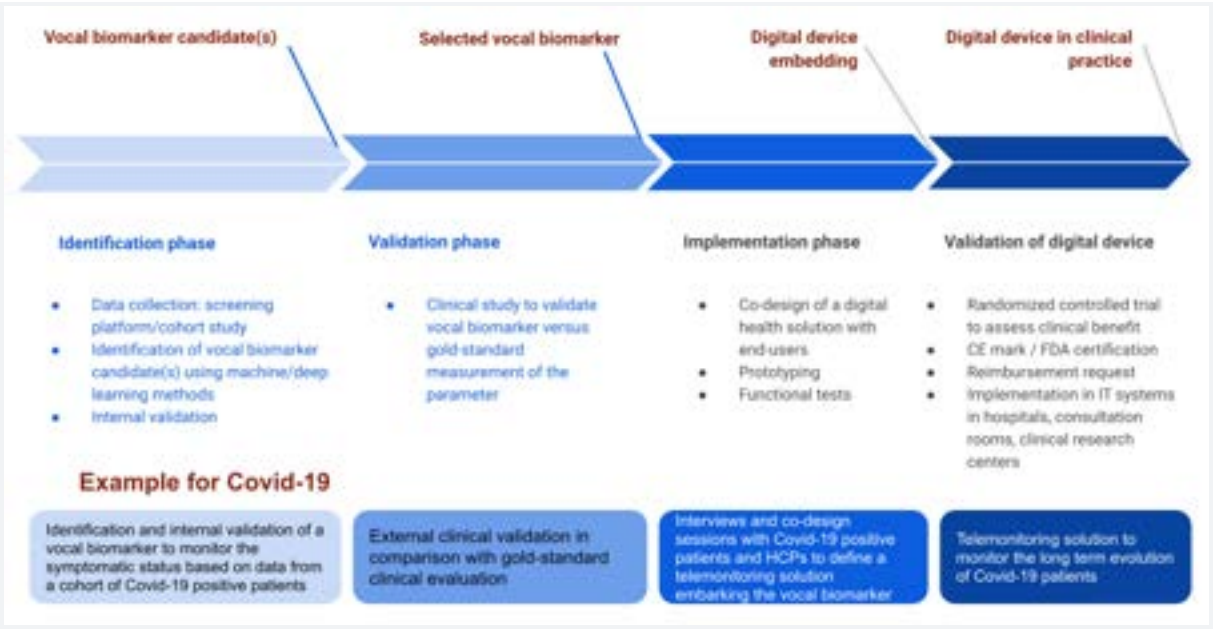


Figure 2. Pipeline from identification to implementation in clinical practice of a vocal biomarker.

CE: conformite europeenne; FDA: Food and Drug Administration; HCP: health care provider; IT: information technology.

Conclusions

This viewpoint presents the need for new digital remote monitoring technologies in the context of COVID-19 and the potential benefit of using vocal biomarkers for this purpose. We also propose a pathway and recommendations for a successful implementation in clinical practice of a digital health solution based on vocal biomarkers.

Implementation of vocal biomarkers in a digital solution for remote patient monitoring of frequently reported symptoms of COVID-19 is of high interest. Its full potential can be achieved in the short term but still includes challenging steps and hurdles to overcome before launching reliable solutions in practice. Vocal biomarker acceptability remains to be properly evaluated, as the use of voice is a rather new technique and needs to be integrated into existing health information technology systems. The future of digital solutions embedding such vocal biomarkers will be diverse and will probably evolve toward multitechnologies solutions combining voice, video, and sensors to offer the most comprehensive view of a patient's health status.

Acknowledgments

The Predi-COVID study is supported by the Luxembourg National Research Fund (FNR) (Predi-COVID, grant 14716273), the André Losch Foundation, and the Luxembourg Institute of Health.

Authors' Contributions

AF wrote the first draft of the manuscript. All other authors critically revised the manuscript and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Bedford J, Enria D, Giesecke J, Heymann DL, Ihekweazu C, Kobinger G, et al. COVID-19: towards controlling of a pandemic. The Lancet 2020 Mar;395(10229):1015-1018.
2. Fagherazzi G, Fischer A, Betsou F, Vaillant M, Ernens I, Masi S, et al. Protocol for a prospective, longitudinal cohort of people with COVID-19 and their household members

to study factors associated with disease severity: the Predi-COVID study. BMJ Open 2020 Nov 23;10(11):e041834
3. Huang L, Yao Q, Gu X, Wang Q, Ren L, Wang Y, et al. 1-year outcomes in hospital survivors with COVID-19: a longitudinal cohort study. The Lancet 2021 Aug;398(10302):747-758.
4. Crook H, Raza S, Nowell J, Young M, Edison P. Long COVID-mechanisms, risk factors, and management. BMJ 2021 Jul 26;374:n1648.
5. Hugon J, Msika E, Queneau M, Farid K, Paquet C. Long COVID: cognitive complaints (brain fog) and dysfunction of the cingulate cortex. J Neurol 2022 Jan 18;269(1):44-46
6. Bellan M, Soddu D, Balbo PE, Baricich A, Zeppegno P, Avanzi GC, et al. Respiratory and psychophysical sequelae among patients with COVID-19 four months after hospital discharge. JAMA Netw Open 2021 Jan 04;4(1):e2036142
7. 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 Aug;38:101019
8. Shah W, Hillman T, Playford ED, Hishmeh L. Managing the long term effects of covid-19: summary of NICE, SIGN, and RCGP rapid guideline. BMJ 2021 Jan 22;372:n136.
9. National Institute for Health and Care Excellence (NICE) , Scottish Intercollegiate Guidelines Network (SIGN), Royal College of General Practitioners (RCGP). COVID-19 Rapid Guideline: Managing the Long-Term Effects of COVID-19. London, UK: National Institute for Health and Care Excellence; 2022. URL: https://www.nice.org.uk/guidance/ng188/resources/covid19-rapid-guideline-managing-the-longterm-effects-of-covid19-pdf-51035515742
10. Martos-Pérez F, Martín-Escalante MD, Olalla-Sierra J, Prada-Pardal JL, García-de-Lucas MD, González-Vega R, et al. The value of telephone consultations during COVID-19 pandemic. an observational study. QJM 2021 Dec 20;114(10):715-720
11. Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. Sci Data 2016 Mar 03;3(1):160011
12. Orri M, Lipset CH, Jacobs BP, Costello AJ, Cummings SR. Web-based trial to evaluate the efficacy and safety of tolterodine ER 4 mg in participants with overactive bladder: REMOTE trial. Contemp Clin Trials 2014 Jul;38(2):190-197.
13. Inan OT, Tenaerts P, Prindiville SA, Reynolds HR, Dizon DS, Cooper-Arnold K, et al. Digitizing clinical trials. NPJ Digit Med 2020 Jul 31;3(1):101
14. Furlong E, Darley A, Fox P, Buick A, Kotronoulas G, Miller M, et al. Adaptation and implementation of a mobile phone-based remote symptom monitoring system for people with cancer in Europe. JMIR Cancer 2019 Mar 14;5(1):e10813
15. Patra R, Das N, Bhattacharya M, Shit P, Patra B, Mukherjee S. Applications of artificial intelligence (AI) protecting from COVID-19 pandemic: a clinical and socioeconomic

perspective. In: Computational Intelligence Techniques for Combating COVID-19. New York, NY: Springer; 2021:45-60.
16. Wang Q, Su M, Zhang M, Li R. Integrating digital technologies and public health to fight COVID-19 pandemic: key technologies, applications, challenges and outlook of digital healthcare. Int J Environ Res Public Health 2021 Jun 04;18(11):6053
17. Whitelaw S, Mamas MA, Topol E, Van Spall HGC. Applications of digital technology in COVID-19 pandemic planning and response. Lancet Digital Health 2020 Aug;2(8):e435-e440.
18. Bouabida K, Malas K, Talbot A, Desrosiers M, Lavoie F, Lebouché B, et al. Remote patient monitoring program for COVID-19 patients following hospital discharge: a cross-sectional study. Front Digit Health 2021 Nov 8;3:721044
19. Sezgin E, Huang Y, Ramtekkar U, Lin S. Readiness for voice assistants to support healthcare delivery during a health crisis and pandemic. NPJ Digit Med 2020 Sep 16;3(1):122
20. Lee H, Friedman ME, Cukor P, Ahern D. Interactive voice response system (IVRS) in health care services. Nursing Outlook 2003 Dec;51(6):277-283.
21. Fagherazzi G, Fischer A, Ismael M, Despotovic V. Voice for health: the use of vocal biomarkers from research to clinical practice. Digit Biomark 2021 Apr 16;5(1):78-88
22. Asiaee M, Vahedian-Azimi A, Atashi SS, Keramatfar A, Nourbakhsh M. Voice quality evaluation in patients with COVID-19: an acoustic analysis. J Voice 2020 Oct 01:1-7
23. Dassie-Leite AP, Gueths TP, Ribeiro VV, Pereira EC, Martins PDN, Daniel CR. Vocal signs and symptoms related to COVID-19 and risk factors for their persistence. J Voice 2021 Aug 11:1-6
24. Tohidast SA, Mansuri B, Memarian M, Ghobakhloo AH, Scherer RC. Voice quality and vocal tract discomfort symptoms in patients With COVID-19. J Voice 2021 Oct 13:1-5
25. Laguarta J, Hueto F, Subirana B. COVID-19 artificial intelligence diagnosis using only cough recordings. IEEE Open J Eng Med Biol 2020;1:275-281
26. Despotovic V, Ismael M, Cornil M, Call RM, Fagherazzi G. Detection of COVID-19 from voice, cough and breathing patterns: dataset and preliminary results. Comput Biol Med 2021 Nov;138:104944
27. Walker RC, Tong A, Howard K, Palmer SC. Patient expectations and experiences of remote monitoring for chronic diseases: systematic review and thematic synthesis of qualitative studies. Int J Med Inform 2019 Apr;124:78-85
28. Izmailova ES, Ellis R, Benko C. Remote monitoring in clinical trials during the COVID-19 pandemic. Clin Transl Sci 2020 Sep 08;13(5):838-841
29. Coppock H, Jones L, Kiskin I, Schuller B. COVID-19 detection from audio: seven grains of salt. Lancet Digital Health 2021 Sep;3(9):e537-e538.

30. ColiveVoice. Luxembourg Institute of Health. URL: https://www.colivevoice.org/en/ [accessed 2021-10-15]
31. Schwoebel J, Schwartz J, Warrenburg L, Brown R, Awasthi A. A longitudinal normative dataset and protocol for speech and language biomarker research. medrxiv Preprint posted online August 24, 2021.
32. Fiorella ML, Cavallaro G, Di Nicola V, Quaranta N. Voice differences when wearing and not wearing a surgical mask. J Voice 2021 Mar 09:1-7
33. Brown C, Chauhan J, Grammenos A, Han J, Hasthanasombat A, Spathis D. Exploring automatic diagnosis of COVID-19 from crowdsourced respiratory sound data. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2020 Presented at: KDD '20; July 6-10; California, US.
34. Robin J, Harrison J, Kaufman L, Rudzicz F, Simpson W, Yancheva M. Evaluation of speech-based digital biomarkers: review and recommendations. Digit Biomark 2020 Oct 19;4(3):99-108
35. Smuck M, Odonkor CA, Wilt JK, Schmidt N, Swiernik MA. The emerging clinical role of wearables: factors for successful implementation in healthcare. NPJ Digit Med 2021 Mar 10;4(1):45
36. Meskó B, deBronkart D. Patient design: the importance of including patients in designing health care. J Med Internet Res 2022 Aug 31;24(8):e39178
37. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). EUR-Lex. URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj [accessed 2022-11-11]
38. O'Brien D. Design Control, Medical Device Risk and Medical Device Regulation (MDR 2017/745): An Integrated Approach for Medical Devices. United States: Independently published; 2020.
39. Guidance document medical devices - Scope, field of application, definition - qualification and classification of stand alone software - MEDDEV 2.1/6. European Commission. URL: https://ec.europa.eu/docsroom/documents/17921/attachments/1/translations [accessed 2021-10-05]
40. Factors to consider regarding benefit-risk in medical device product availability, compliance, and enforcement decisions. Food and Drug Administration. URL: https://www.fda.gov/media/98657/download [accessed 2022-11-11]
41. Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment. World Health Organization. URL: https://apps.who.int/iris/handle/10665/252183 [accessed 2022-11-11]

Abbreviations

FDA: Food and Drug Administration

Edited by A Mavragani; submitted 30.06.22; peer-reviewed by E Sezgin, S D'Arcy; comments to author 01.09.22; revised version received 30.09.22; accepted 18.10.22; published 15.11.22

Annexe 2 : Questionnaire long terme de l'étude Predi-COVID

Après la phase aiguë de l'infection à Covid-19, certaines personnes développent des formes prolongées de la maladie avec une multitude de symptômes. Cependant il reste de nombreuses questions non résolues sur les conséquences à long terme (plus de 3 semaines après le diagnostic) du Covid-19.

C'est pourquoi nous vous demandons de remplir ce dernier questionnaire. Il est très important pour les résultats de l'étude qui seront utiles pour d'autres personnes touchées par cette maladie.

Merci d'indiquer votre état de santé actuel ainsi que les symptômes que vous ressentez de manière continue ou fréquente depuis votre diagnostic de Covid-19.

Avez-vous noté depuis votre diagnostic Covid-19 les symptômes ou maladies suivants :

Modalités de réponse :

- oui et je le ressens encore aujourd'hui
- oui, je l'ai eu mais je ne l'ai plus
  - question conditionnelle : "Ce symptôme a disparu _ _ semaines après le diagnostic
- non, je n'ai jamais eu ce symptôme

Symptômes ORL :

1) Perte de goût
2) Perte d'odorat
3) Écoulement nasal, rhume ou rhinite
4) Douleurs dans les sinus
5) Douleurs dans les oreilles
6) Maux de gorge, sensation de constriction dans la gorge ou douleur lorsque vous avalez
7) Altération voix

Symptômes neurologiques et oculaires :

8) Tremblements des mains ou des membres
9) Maux de tête
10) Migraines

11) Confusion mentale (ralentissement de la pensée et/ou du raisonnement)

12) Sensation de malaise ou étourdissements (vertiges, « tomber dans les pommes »)

13) Crises d'épilepsie/convulsions

14) Troubles de l'équilibre

15) Pertes de mémoire / troubles de la mémoire à court terme

16) Fatigue au niveau de vos yeux

17) Hallucinations

18) Sensibilité à la lumière (photophobie)

19) Inflammation conjonctivale, les yeux qui grattent ou les yeux rouges/rougis

20) Troubles du langage

Symptômes généraux :

21) Fatigue

22) Irritabilité, nervosité exacerbée

23) Anxiété

24) Dépression

25) Sueurs / frissons

26) Fièvre (>=38 °C)

27) Perte d'appétit

28) Perte de poids involontaire (plus de 3 kg)

29) Sensation de soif récurrente

30) Sensations de picotements / fourmillements / engourdissements dans les membres ou sur la peau

31) Douleurs musculaires ou articulaires dans les membres supérieurs

32) Douleurs musculaires ou articulaires dans les membres inférieurs

33) Maux de dos

34) Allergie nouvellement apparue (si oui : quel type d'allergie/ champ libre ou liste)

35) Perte de cheveux

36) Difficultés à marcher

Symptômes ou maladies cardio-respiratoires :

37) Souffle court / essoufflement

38) Sensation d'oppression dans la poitrine

39) Toux sèche

40) Toux grasse

41) Tachycardie (fréquence cardiaque anormalement haute au repos ou durant des activités de faible intensité)

42) Arythmie / Palpitations (pulsations cardiaques irrégulières ou sensation que le cœur bat très fort)

43) Myocardite

44) Péricardite

45) Insuffisance cardiaque

46) Infarctus du myocarde

47) Sensation de brûlures dans la poitrine

48) Douleurs thoraciques

49) Respiration sifflante

50) Crachats sanglants

51) Thrombose

Symptômes gastrointestinaux :

52) Nausées

53) Vomissements

54) Diarrhée (survenue de selles liquides)

55) Constipation

56) Brûlures d'estomac

57) Douleurs abdominales autres

Symptômes ou maladies vasculaires/ganglionnaires :

58) Hypertension

59) Hypotension

60) Accident Vasculaire Cérébral

- Si oui : Ischémique ou hémorragique?

61) Adénopathies (ganglions gonflés, douloureux ou inflammés)

62) Troubles de la circulation (veines gonflées, jambes lourdes..)

63) Hématomes spontanés

Symptômes urinaires

64) Douleurs urinaires

65) Infections urinaires récurrentes

66) Nécessité de dialyse

Symptômes cutanés

67) Eruptions / lésions cutanées

68) Peau sèche

69) Doigts ou orteils bleus/violacés/blancs ou gonflés

70) Autres symptômes : Champ libre

Questions générales

71) Comment vous sentez-vous aujourd'hui? (Je me sens bien / Je me sens fatigué(e) / Je me sens mal)

72) En tenant compte de tous les symptômes pouvant être attribués au Covid-19 que vous avez ressentis au cours des 30 derniers jours (fréquence, intensité, impact sur votre vie), diriez-vous que vous pourriez vivre au long court dans votre état de santé actuel? (Oui/Non)

73) La survenue des symptômes que vous avez cités se fait-elle à l'occasion de "crises", avec des phases d'amélioration ou d'aggravation?

- ○ Oui, avec des crises quotidiennes
- ○ Oui, avec au moins 1 crise par semaine
- ○ Oui, mais moins d'une crise par semaine
- ○ Non, mes symptômes sont constants au cours du temps
- ○ Je n'ai pas de symptômes

74) Combien de temps ces crises durent-elles en général?

- ○ Moins d'une journée
- ○ Plus d'une journée mais moins d'une semaine
- ○ Plus d'une semaine mais moins d'un mois
- ○ Je n'ai pas de crises

75) Quel est votre niveau de douleur actuel ?

(sur une échelle de 1 à 10, 10 étant la pire douleur imaginable)

76) Depuis le remplissage de votre dernier questionnaire, avez-vous consulté pour une raison en lien avec Covid-19 ?

Oui/Non

-> si oui, pour quel motif ? (Champ libre)

77) Avez-vous repris un rythme de vie comparable à celui que vous aviez avant la survenue des premiers symptômes associés à Covid-19 ?

Oui/Non

-> Si non, pourquoi ? Champ Libre

78) Avez-vous repris votre activité professionnelle normalement ?

Oui/Non/Je suis à la retraite ou sans emploi

-> Si non, avez-vous été arrêté des suites de complications du Covid-19 ? Oui/Non

79) Depuis le diagnostic de Covid-19 ou la survenue de symptômes associés à Covid-19, est-ce que vos relations avec votre entourage (famille, amis)

Se sont dégradées / Sont restées les mêmes / Se sont améliorées

-> si dégradées ou améliorées : Pourquoi ? Champ Libre

80) Depuis le diagnostic de Covid-19 ou la survenue de symptômes associés à Covid-19, diriez-vous que : Votre appétit/Votre activité physique/Votre sommeil (a diminué / est identique ou presque / a augmenté)

81) Bénéficiez-vous actuellement ou avez-vous bénéficié d'une consultation en lien avec le syndrome Long Covid ? Oui/Non/ne sais pas

- Si oui vous a-t-on proposé une rééducation ? Oui/Non/ne sais pas
- Si oui, dans quelle structure? Rehazenter/Mondorf/CHNP/Autre préciser

82) Depuis le diagnostic de Covid-19 ou la survenue de symptômes associés à Covid-19, avez-vous noté des changements dans vos cycles menstruels? (Oui/Non/Ne sais pas/ Non applicable)

Si oui, règles plus abondantes / règles irrégulières / ménopause précoce / ne sais pas

83) Sommeil, échelle PSQI

Les questions suivantes ont trait à vos habitudes de sommeil pendant le dernier mois seulement. Vos réponses doivent indiquer ce qui correspond aux expériences que vous avez eues pendant la majorité des jours et des nuits au cours du dernier mois.
Répondez à toutes les questions.

1/ Au cours du mois dernier, quand êtes-vous habituellement allé vous coucher le soir ? Heure habituelle du coucher : .....

2/ Au cours du mois dernier, combien vous a-t-il habituellement fallu de temps (en minutes) pour vous endormir chaque soir ? Nombre de minutes : .....

3/ Au cours du mois dernier, quand vous êtes-vous habituellement levé le matin ? Heure habituelle du lever : .....

4/ Au cours du mois dernier, combien d'heures de sommeil effectif avez-vous eu chaque nuit ? (Ce nombre peut être différent du nombre d'heures que vous avez passé au lit) Heures de sommeil par nuit : ..... Pour chacune des questions suivantes, indiquez la meilleure réponse. Répondez à toutes les questions.

5/ Au cours du mois dernier, avec quelle fréquence avez-vous eu des troubles du sommeil car .

	Pas au cours du dernier mois	Moins d'une fois par semaine	Une ou deux fois par semaine	Trois ou quatre fois par semaine
a) vous n'avez pas pu vous endormir en moins de 30 mn				
b) vous vous êtes réveillé au milieu de la nuit ou précocement le matin				
c) vous avez dû vous lever pour aller aux toilettes				

d) vous n'avez pas pu respirer correctement				
e) vous avez toussé ou ronflé bruyamment				
f) vous avez eu trop froid				
g) vous avez eu trop chaud				
h) vous avez eu de mauvais rêves				
i) vous avez eu des douleurs				
j) pour d'autre(s) raison(s). Donnez une description :				
Indiquez la fréquence des troubles du sommeil pour ces raisons				

6/ Au cours du mois dernier, comment évalueriez-vous globalement la qualité de votre sommeil ? Très bonne Assez bonne Assez mauvaise Très mauvaise

7/ Au cours du mois dernier, combien de fois avez-vous pris des médicaments (prescrits par votre médecin ou achetés sans ordonnance) pour faciliter votre sommeil ? Pas au cours Moins d'une fois Une ou deux fois Trois ou quatre fois du dernier mois par semaine par semaine par semaine

8/ Au cours du mois dernier, combien de fois avez-vous eu des difficultés à demeurer éveillé(e) pendant que vous conduisiez, preniez vos repas, étiez occupé(e) dans une activité sociale ? Pas au cours Moins d'une fois Une ou deux fois Trois ou quatre fois du dernier mois par semaine par semaine par semaine

9/ Au cours du mois dernier, à quel degré cela a-t-il représenté un problème pour vous d'avoir assez d'enthousiasme pour faire ce que vous aviez à faire ? Pas du tout Seulement un Un certain problème Un très gros un problème tout petit problème problème

10/ Avez-vous un conjoint ou un camarade de chambre ? Ni l'un, ni l'autre. Oui, mais dans une chambre différente. Oui, dans la même chambre mais pas dans le même lit. Oui, dans le même lit.

11/ Si vous avez un camarade de chambre ou un conjoint, demandez-lui combien de fois le mois dernier vous avez présenté :

	Pas au cours du dernier mois	Moins d'une fois par semaine	Une ou deux fois par semaine	Trois ou quatre fois par semaine
a) un ronflement fort				
b) de longues pauses respiratoires pendant votre sommeil				
c) des saccades ou des secousses des jambes pendant que vous dormiez				
d) des épisodes de désorientation ou de confusion pendant le sommeil				
e) d'autres motifs d'agitation pendant le sommeil				

84) Qualité de vie (SF12)

1. Dans l'ensemble, pensez-vous que votre santé est :
☐ 1 Excellente ☐ 2 Très bonne ☐ 3 Bonne ☐ 4 Médiocre ☐ 5 Mauvaise

2. En raison de votre état de santé actuel, êtes-vous limité pour :

- des efforts physiques modérés (déplacer une table, passer l'aspirateur, jouer aux boules...) ?
  ☐ 1 Oui, beaucoup limité ☐ 2 Oui, un peu limité ☐ 3 Non, pas du tout limité

- monter plusieurs étages par l'escalier ?
  ☐ 1 Oui, beaucoup limité ☐ 2 Oui, un peu limité ☐ 3 Non, pas du tout limité

3. Au cours de ces 4 dernières semaines, et en raison de votre état physique :

- avez-vous accompli moins de choses que vous auriez souhaité ?
  ☐ 1 Toujours ☐ 2 La plupart du temps ☐ 3 Souvent ☐ 4 Parfois ☐ 5 Jamais

- avez-vous été limité pour faire certaines choses ?
  ☐ 1 Toujours ☐ 2 La plupart du temps ☐ 3 Souvent ☐ 4 Parfois ☐ 5 Jamais

4. Au cours de ces 4 dernières semaines, et en raison de votre état émotionnel (comme vous sentir triste, nerveux ou déprimé) :

- avez-vous accompli moins de choses que vous auriez souhaité ?
  ☐ 1 Toujours ☐ 2 La plupart du temps ☐ 3 Souvent ☐ 4 Parfois ☐ 5 Jamais

- avez-vous eu des difficultés à faire ce que vous aviez à faire avec autant de soin et d'attention que d'habitude ?
  ☐ 1 Toujours ☐ 2 La plupart du temps ☐ 3 Souvent ☐ 4 Parfois ☐ 5 Jamais

5. Au cours de ces 4 dernières semaines, dans quelle mesure vos douleurs physiques vous ont-elles limité dans votre travail ou vos activités domestiques ?
☐ 1 Pas du tout ☐ 2 Un petit peu ☐ 3 Moyennement ☐ 4 Beaucoup ☐ 5 Enormément

6. Les questions qui suivent portent sur comment vous vous êtes senti au cours de ces 4 dernières semaines. Pour chaque question, indiquez la réponse qui vous semble la plus appropriée.

- y a t-il eu des moments où vous vous êtes senti calme et détendu ?
  ☐ 1 Toujours ☐ 2 La plupart du temps ☐ 3 Souvent ☐ 4 Parfois ☐ 5 Jamais

- y a t-il eu des moments où vous vous êtes senti débordant d'énergie ?
  ☐ 1 Toujours ☐ 2 La plupart du temps ☐ 3 Souvent ☐ 4 Parfois ☐ 5 Jamais

- y a t-il eu des moments où vous vous êtes senti triste et abattu ?
  ☐ 1 Toujours ☐ 2 La plupart du temps ☐ 3 Souvent ☐ 4 Parfois ☐ 5 Jamais

7. Au cours de ces 4 dernières semaines, y a t-il eu des moments où votre état de santé physique ou émotionnel vous a gêné dans votre vie sociale et vos relations avec les autres, votre famille, vos amis, vos connaissances ?
☐ 1 Toujours ☐ 2 La plupart du temps ☐ 3 Souvent ☐ 4 Parfois ☐ 5 Jamais

85) Qualité de vie respiratoire (Échelle VQ11)

		Pas du tout	Un peu	Moyennement	Beaucoup	Extrêmement
1	Je souffre de mon essoufflement	❑	❑	❑	❑	❑
2	Je me fais du souci pour mon état respiratoire	❑	❑	❑	❑	❑
3	Je me sens incompris(e) par mon entourage	❑	❑	❑	❑	❑
4	Mon état respiratoire m'empêche de me déplacer comme je le voudrais	❑	❑	❑	❑	❑
5	Je suis somnolent(e) dans la journée	❑	❑	❑	❑	❑
6	Je me sens incapable de réaliser mes projets	❑	❑	❑	❑	❑
7	Je me fatigue rapidement dans les activités de la vie quotidienne	❑	❑	❑	❑	❑
8	Physiquement, je suis insatisfait(e) de ce que je peux faire	❑	❑	❑	❑	❑
9	Ma maladie respiratoire perturbe ma vie sociale	❑	❑	❑	❑	❑
10	Je me sens triste	❑	❑	❑	❑	❑
11	Mon état respiratoire limite ma vie affective	❑	❑	❑	❑	❑

86/ GAD7	
Au cours des 14 derniers jours, à quelle fréquence avez-vous été dérangé(e) par les problèmes suivants?	
Sentiment de nervosité, d'anxiété ou de tension	Jamais / Plusieurs jours / Plus de la moitié des jours / Presque tous les jours
Incapable d'arrêter de vous inquiéter ou de contrôler vos inquiétudes	Jamais / Plusieurs jours / Plus de la moitié des jours / Presque tous les jours
Inquiétudes excessives à propos de tout et de rien	Jamais / Plusieurs jours / Plus de la moitié des jours / Presque tous les jours
Difficulté à se détendre	Jamais / Plusieurs jours / Plus de la moitié des jours / Presque tous les jours
Agitation telle qu'il est difficile de rester tranquille	Jamais / Plusieurs jours / Plus de la moitié des jours / Presque tous les jours
Devenir facilement contrarié(e) ou irritable	Jamais / Plusieurs jours / Plus de la moitié des jours / Presque tous les jours
Avoir peur que quelque chose d'épouvantable puisse arriver	Jamais / Plusieurs jours / Plus de la moitié des jours / Presque tous les jours
87/ PSS4	
Ce questionnaire porte sur votre vécu durant le mois passé	
Durant le mois passé, combien de fois avez-vous eu le sentiment de ne pas pouvoir contrôler les aspects importants de votre vie ?	Jamais [ ] Presque jamais [ ] Parfois [ ] Assez souvent [ ] Très souvent
Durant le mois passé, combien de fois avez-vous eu confiance en votre capacité à gérer vos problèmes personnels ?	Jamais [ ] Presque jamais [ ] Parfois [ ] Assez souvent [ ] Très souvent

Durant le mois passé, combien de fois avez-vous eu le sentiment les choses allaient comme vous le vouliez?	Jamais [ ] Presque jamais [ ] Parfois [ ] Assez souvent [ ] Très souvent
Durant le mois passé, combien de fois avez-vous eu le sentiment que les difficultés s'accumulaient tellement que vous ne pourriez pas les surmonter ?	Jamais [ ] Presque jamais [ ] Parfois [ ] Assez souvent [ ] Très souvent
88/ FSS9	
Je me sens moins motivé du fait de la fatigue.	Pas du tout d'accord/Pas d'accord/Pas vraiment d'accord/Ni d'accord ni en désaccord/ Plutôt d'accord/ D'accord/ Tout à fait d'accord
L'exercice physique est pour moi source de fatigue	Pas du tout d'accord/Pas d'accord/Pas vraiment d'accord/Ni d'accord ni en désaccord/ Plutôt d'accord/ D'accord/ Tout à fait d'accord
Je suis facilement fatigué(e)	Pas du tout d'accord/Pas d'accord/Pas vraiment d'accord/Ni d'accord ni en désaccord/ Plutôt d'accord/ D'accord/ Tout à fait d'accord
La fatigue interfère avec mon activité physique	Pas du tout d'accord/Pas d'accord/Pas vraiment d'accord/Ni d'accord ni en désaccord/ Plutôt d'accord/ D'accord/ Tout à fait d'accord
La fatigue est souvent un problème pour moi	Pas du tout d'accord/Pas d'accord/Pas vraiment d'accord/Ni d'accord ni en désaccord/ Plutôt d'accord/ D'accord/ Tout à fait d'accord
Ma fatigue m'empêche de réaliser des tâches physiques soutenues et prolongées	Pas du tout d'accord/Pas d'accord/Pas vraiment d'accord/Ni d'accord ni en désaccord/ Plutôt d'accord/ D'accord/ Tout à fait d'accord

Ma fatigue interfère avec mes facultés pour la réalisation de certaines activités et responsabilités	Pas du tout d'accord/Pas d'accord/Pas vraiment d'accord/Ni d'accord ni en désaccord/ Plutôt d'accord/ D'accord/ Tout à fait d'accord
La fatigue fait partie de mes trois symptômes les plus gênants	Pas du tout d'accord/Pas d'accord/Pas vraiment d'accord/Ni d'accord ni en désaccord/ Plutôt d'accord/ D'accord/ Tout à fait d'accord
Ma fatigue interfère avec mon travail, ma famille ou ma vie sociale	Pas du tout d'accord/Pas d'accord/Pas vraiment d'accord/Ni d'accord ni en désaccord/ Plutôt d'accord/ D'accord/ Tout à fait d'accord

Annexe 3 : Matériel supplémentaires de l'article 1

Supplementary Table 1: Comparison of baseline characteristics of respondents versus non-respondents to the 12-month questionnaire

Variable	Overall (N = 539)	Non-respondent (N = 250)	Respondent (N = 289)	P-value*
Woman N (%)	243(45.68)	99(40.41)	144(50.17)	0.029
Age (Years)	39.5±12.4	38.6±12.2	40.1±12.5	0.162
Body Mass Index (kg/m2)	25.7±4.7	25.9±4.6	25.6±4.8	0.475
Hypertension N (%)	46(9.13)	20(8.77)	26(9.42)	0.877
Cardiovascular disease N (%)	17(3.37)	8(3.51)	9(3.26)	1
Asthma N (%)	21(4.17)	13(5.70)	8(2.90)	0.124
Diabetes N (%)	14(2.78)	6(2.63)	8(2.91)	1
Current smoker N (%)	96(19.05)	48(21.05)	48(17.39)	0.307
Former smoker N (%)	99(19.64)	44(19.30)	55(19.93)	0.911
Mild illness N (%)	323(64.09)	149(65.35)	174(63.04)	0.901
Moderate/severe illness N (%)	101(20.04)	43(18.86)	58(21.01)	0.765

*P-values were calculated using Fisher exact for categorical variables, and Student's t test for continuous variables

Supplementary Table 2: Distribution of participants experiencing symptoms at 12 months, according to the disease severity during acute infection

	Overall population	Disease severity at inclusion			
		Asymptomatic	Mild	Moderate/Severe	P-value*
Having at least 1 symptom present at 1 year N(%)	172(59.5)	17(38.6)	98(56.3)	48(82.8)	<0.001
Median total number of symptoms N [Q1,Q3] in people with at least 1 reported symptom	6.0[2.0,11.3]	2.0[1.0,4.0]	6.0[3.0,10.0]	8.0[3.0,18.3]	<0.001
Fatigue N(%)	99(34.3)	7(15.9)	53(30.5)	34(58.6)	<0.001
Eyes tiredness N(%)	65(22.5)	3(6.8)	40(23.0)	20(34.5)	0,003
Shortness of breath N(%)	54(18.7)	3(6.8)	26(14.9)	23(39.7)	<0.001
Irritability N(%)	52(18.0)	2(4.6)	30(17.2)	17(29.3)	0,004
Anxiety N(%)	46(15.9)	2(4.6)	26(14.9)	17(29.3)	0,003

Mental confusion N(%)	45(15.6)	2(4.6)	22(12.6)	20(34.5)	<0.001
Memory loss N(%)	45(15.6)	4(9.1)	23(13.2)	17(29.3)	0,009
Muscle or joint pain in the lower limbs N(%)	45(15.6)	3(6.8)	24(13.8)	18(31.0)	0,002
Headaches N(%)	44(15.2)	2(4.6)	28(16.1)	13(22.4)	0,036
Back pain N(%)	43(14.9)	1(2.3)	24(13.8)	16(27.6)	0,001
Feeling of tightness in the chest N(%)	39(13.5)	0(0.0)	19(10.9)	17(29.3)	<0.001
Tingling sensations in the limbs or on the skin N(%)	38(13.1)	3(6.8)	21(12.1)	14(24.1)	0,032
Muscle or joint pain in the upper limbs N(%)	38(13.1)	0(0.0)	22(12.6)	16(27.6)	<0.001
Arrhythmia N(%)	38(13.1)	1(2.3)	20(11.5)	16(27.6)	0,001
Migraines N(%)	37(12.8)	1(2.3)	20(11.5)	14(24.1)	0,004
Dry skin N(%)	34(11.8)	1(2.3)	22(12.6)	11(19.0)	0,025
Circulation disorders N(%)	30(10.4)	1(2.3)	15(8.6)	14(24.1)	0,001
Feeling sick N(%)	29(10.0)	1(2.3)	13(7.5)	14(24.1)	<0.001
Sweats / chills N(%)	29(10.0)	2(4.6)	18(10.3)	8(13.8)	0,313
Loss of smell N(%)	28(9.7)	3(6.8)	17(9.8)	7(12.1)	0,672
Depression N(%)	27(9.3)	2(4.6)	15(8.6)	9(15.5)	0,155
Recurrent feeling of thirst N(%)	26(9.0)	1(2.3)	14(8.1)	11(19.0)	0,012
Tachycardia N(%)	24(8.3)	1(2.3)	13(7.5)	10(17.2)	0,023
Tremor of limbs N(%)	23(8.0)	1(2.3)	12(6.9)	10(17.2)	0,020
Hair loss N(%)	23(8.0)	0(0.0)	17(9.8)	4(6.9)	0,062
Balance disorders N(%)	22(7.6)	2(4.6)	12(6.9)	8(13.8)	0,184
Stomach burn N(%)	22(7.6)	1(2.3)	14(8.1)	7(12.1)	0,206
Loss of taste N(%)	21(7.3)	0(0.0)	16(9.2)	5(8.6)	0,079
Conjunctival inflammation N(%)	21(7.3)	2(4.6)	11(6.3)	7(12.1)	0,274
Runny nose N(%)	20(6.9)	2(4.6)	11(6.3)	7(12.1)	0,274
Photophobia N(%)	20(6.9)	1(2.3)	12(6.9)	6(10.3)	0,289
Chest pain N(%)	17(5.9)	0(0.0)	8(4.6)	9(15.5)	0,003
Other abdominal pain N(%)	17(5.9)	0(0.0)	12(6.9)	5(8.6)	0,143
Pain in the sinuses N(%)	15(5.2)	1(2.3)	8(4.6)	6(10.3)	0,168
Nausea N(%)	14(4.8)	1(2.3)	7(4.0)	6(10.3)	0,145
Diarrhea N(%)	14(4.8)	1(2.3)	7(4.0)	6(10.3)	0,145
Hypertension N(%)	14(4.8)	1(2.3)	6(3.5)	7(12.1)	0,034
Weight loss N(%)	13(4.5)	0(0.0)	9(5.2)	4(6.9)	0,202
Wheezing N(%)	13(4.5)	1(2.3)	6(3.5)	6(10.3)	0,083
Dry cough N(%)	12(4.2)	1(2.3)	6(3.5)	5(8.6)	0,241
Burning sensation in the chest N(%)	12(4.2)	0(0.0)	4(2.3)	8(13.8)	0,001

Pain in the ears N(%)	11(3.8)	0(0.0)	6(3.5)	4(6.9)	0,204
Hematomas N(%)	11(3.8)	0(0.0)	7(4.0)	4(6.9)	0,256
Sore throat N(%)	10(3.5)	1(2.3)	6(3.5)	3(5.2)	0,805
Loss of appetite N(%)	10(3.5)	0(0.0)	7(4.0)	3(5.2)	0,371
Walking difficulties N(%)	8(2.8)	1(2.3)	4(2.3)	3(5.2)	0,514
Skin rashes N(%)	8(2.8)	0(0.0)	6(3.5)	2(3.5)	0,668
Hypotension N(%)	7(2.4)	0(0.0)	2(1.2)	4(6.9)	0,042
Newly appeared allergy N(%)	6(2.1)	0(0.0)	3(1.7)	3(5.2)	0,203
Lymphadenopathy N(%)	6(2.1)	0(0.0)	2(1.2)	4(6.9)	0,042
Fatty cough N(%)	5(1.7)	2(4.6)	0(0.0)	3(5.2)	0,008
Urinary tract infections N(%)	5(1.7)	0(0.0)	4(2.3)	1(1.7)	0,829
Blue / purple / white or swollen fingers or toes N(%)	4(1.4)	0(0.0)	3(1.7)	1(1.7)	1,000
Heart failure N(%)	2(0.7)	0(0.0)	0(0.0)	2(3.5)	0,069
Urinary pain N(%)	2(0.7)	0(0.0)	2(1.2)	0(0.0)	1,000
Epileptic attacks N(%)	1(0.3)	0(0.0)	0(0.0)	1(1.7)	0,370
Fever N(%)	1(0.3)	0(0.0)	0(0.0)	1(1.7)	0,370
Myocarditis N(%)	1(0.3)	0(0.0)	0(0.0)	1(1.7)	0,370
Bloody sputum N(%)	1(0.3)	0(0.0)	0(0.0)	1(1.7)	0,370
Vomiting N(%)	1(0.3)	0(0.0)	1(0.6)	0(0.0)	1,000

*Overall P-values are determined using the Fisher's exact test for each symptom and Kruskal–Wallis test for the total number of symptoms variable.

Supplementary Table 3: Comparison of baseline characteristics of participants who declared not being able coping with their symptoms long term with those declaring to be able coping with their symptoms.

Variable	Overall study population (N = 289)	Able coping with symptoms long term (N = 253)	Not being able coping with symptoms long term (N = 36)	P-value*
Woman N (%)	144(50.17)	128(51.00)	16(44.44)	0,482
Age (Years)	40.1±12.5	40.6±12.8	36.9±10.0	0,054
Body Mass Index (kg/m2)	25.6±4.8	25.5±4.7	26.5±5.8	0,337
Hypertension N (%)	26(9.42)	23(9.54)	3(8.57)	1,000
Cardiovascular disease N (%)	9(3.26)	8(3.32)	1(2.86)	1,000
Asthma N (%)	8(2.90)	8(3.32)	0(0.00)	0,602
Diabetes N (%)	8(2.91)	7(2.92)	1(2.86)	1,000
Current smoker N (%)	48(17.39)	39(16.18)	9(25.71)	0,226
Former smoker N (%)	55(19.93)	49(20.33)	6(17.14)	1,000

Mild illness N (%)	174(63.04)	152(63.07)	22(62.86)	0,612
Moderate/severe illness N (%)	58(21.01)	49(20.33)	9(25.71)	0,385

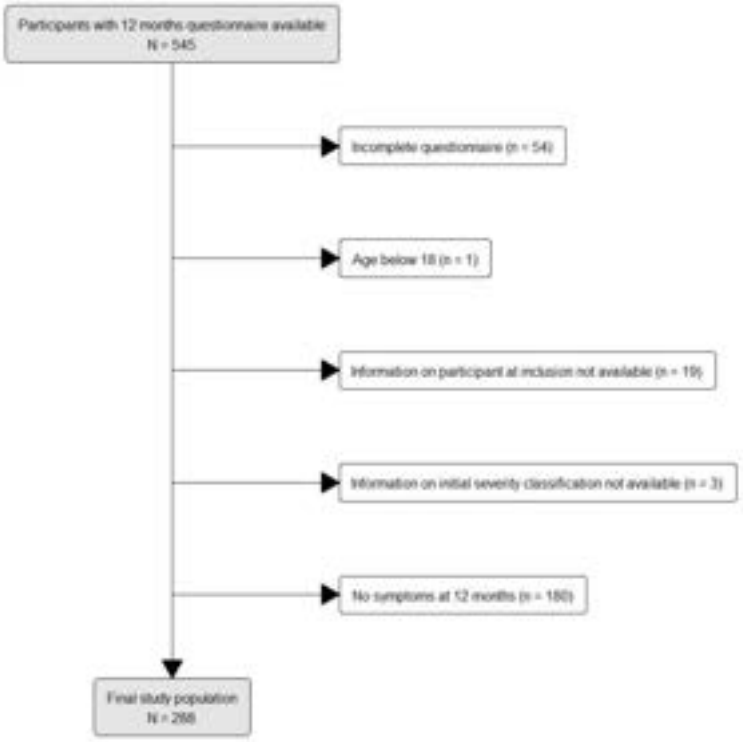
* P-values were calculated using Fisher exact for categorical variables, and Student's t test for continuous variables

Supplementary table 4: Most important symptoms impacting long-term quality of life (Predi-COVID cohort study, N = 289)

	Participants still having the symptom at 1 year	
Symptom (ordered by decreasing Risk Ratio)	**Overall (%)**	**Those who declared not being able to live like this in the long run (%)**
Fatigue	34,3	50,0
Shortness of breath	18,7	36,1
Eyes tiredness	22,5	36,1
Anxiety	15,9	33,3
Feeling of tightness in the chest	13,5	30,6
Irritability	18,0	30,6
Migraines	12,8	27,8
Muscle or joint pain in the lower limbs	15,6	27,8
Sweats / chills	10,0	25,0
Headaches	15,2	25,0
Tremor of limbs	8,0	22,2
Dry skin	11,8	22,2
Arrhythmia	13,1	22,2
Back pain	14,9	22,2
Mental confusion	15,6	22,2
Stomach burn	7,6	19,4
Recurrent feeling of thirst	9,0	19,4
Depression	9,3	19,4
Circulation disorders	10,4	19,4
Tingling sensations in the limbs or on the skin	13,1	19,4
Other abdominal pain	5,9	16,7
Tachycardia	8,3	16,7
Muscle or joint pain in the upper limbs	13,1	16,7
Nausea	4,8	13,9
Chest pain	5,9	13,9
Balance disorders	7,6	13,9
Loss of smell	9,7	13,9
Feeling sick	10,0	13,9

Memory loss	15,6	13,9
Diarrhea	4,8	11,1
Runny nose	6,9	11,1
Photophobia	6,9	11,1
Dry cough	4,2	8,3
Burning sensation in the chest	4,2	8,3
Hypertension	4,8	8,3
Pain in the sinuses	5,2	8,3
Loss of taste	7,3	8,3
Hair loss	8,0	8,3
Sore throat	3,5	5,6
Hematomas	3,8	5,6

Annexe 4 : Matériel supplémentaire de l'article 2



Supplementary Figure A1. Flowchart of participants included in the analyses (N = 288).

Supplementary Table A1: Full list of persisting symptoms considered in the 12-months questionnaire

Ear/Nose/Throat Symptoms	Neurological and ocular symptoms	General symptoms	Cardio-respiratory symptoms or diseases	Gastrointestinal symptoms	Vascular and ganglionic symptoms or diseases	Urinary symptoms	Skin symptoms
Loss of taste	Tremors	Fatigue	Shortness breath	Nausea	Hypertension	Urinary pain	Skin rashes
Loss of smell	Headaches	Irritability	Chest tightness	Vomiting	Hypotension	Urinary infections	Dry skin
Runny nose, cold or rhinitis	Migraines	Anxiety	Dry cough	Diarrhea	Adenopathies	Dialysis	Blue fingers
Sinus pain	Mental confusion	Depression	Fatty cough	Heartburn	Circulation disorders		
Ear pain	Malaise	Sweating	Tachycardia	Abdominal pain	Hematoma		

Sore throat	Convulsions	Fever	Arrhythmia
	Balance	Loss of appetite	Myocarditis
	Memory	Loss weight	Heart failure
	Fatigue in eyes	Thirst	Burning chest
	Hallucinations	Ants	Chest pain
	Sensitivity to light	Muscle pain (upper limbs)	Wheezing
	Conjunctivitis	Muscle pain (lower limbs)	Coughing blood
		Back pain	
		Allergy	
		Loss hair	
		Difficulty walking	

Supplementary Table A2: Full list of features included in the clustering

Sociodemographic Characteristics and Initial Severity Classification	Inclusion at home or at hospital Gender Age BMI Weight loss in last 6 months Smoking status Classification severity initial illness Blood type
Comorbidities	Hypertension Cardiac diseases Pulmonary diseases Asthma Renal diseases Hepatic diseases Neurological diseases Cancer Hematological diseases Obesity Diabetes Rheumatological diseases Malnutrition COPD
Symptoms at inclusion	Fever Cough Cough_sputum Cough hemoptysis

	Sore throat Rhinorrhea Earache Wheezing Chest_pain Myalgia Arthralgia Fatigue Dyspnea Chest tightness Cephalea
	Confusion Abdominal pain Nausea Diarrhea Conjunctivitis Skin rash Lymphadenopathy Fall Hemorrhage
Persisting symptoms at 12 months by categories	Ear Nose Throat (ENT) symptoms Neurological symptoms General symptoms Cardio-respiratory symptoms Gastrointestinal symptoms Vascular symptoms Urinary symptoms Skin symptoms
Quality of life	Could not envisage coping with symptoms long term N(%) PSQI score VQ11 global score VQ11 Functional component score VQ11 Psychological component score VQ11 Relational component score

Annexe 5 : Matériel supplémentaire de l'article 3

Additional file 1: Predicovid (adult participants)- Questionnaire M12-15-24

Have you noticed the following symptoms or illnesses since your COVID-19 diagnosis?

Responses :

● yes and I still feel it today

● yes, I had it but I no longer have it o conditional question: "This symptom disappeared _ _ weeks after diagnosis

● no, I have never had this symptom

ENT symptoms :

1) loss of taste

2) loss of smell

3) Runny nose, cold or rhinitis

4) Pain in the sinuses

5) Pain in the ears

6) Sore throat, feeling of constriction in the throat or pain when you swallow

7) Voice alteration

Neurological and ocular symptoms:

8) Tremor of the hands or limbs

9) Headache

10) Migraines

11) Mental confusion (slowing down of thought and / or reasoning)

12) Feeling sick or dizzy (dizziness, fainting)

13) Epileptic attacks / convulsions

14) Balance disorders

15) Memory loss / short-term memory impairment

16) Tiredness in your eyes

17) Hallucinations

18) Sensitivity to light (photophobia)

19) Conjunctival inflammation, itchy eyes or red / flushed eyes

20) Speech disorders

General symptoms:

21) Fatigue

22) Irritability, exacerbated nervousness

23) Anxiety

24) Depression

25) Sweats / chills

26) Fever (> = 38 ° C)

27) loss of appetite

28) Unintentional weight loss (over 3 kg)

29) recurrent feeling of thirst

30) Tingling / pins and needles / numbness sensations in the limbs or on the skin

31) Muscle or joint pain in the upper limbs

32) Muscle or joint pain in the lower limbs

33) back pain

34) Newly appeared allergy (if yes: what type of allergy / free field or list)

35) hair loss

36) Difficulty walking

Cardio-respiratory symptoms or diseases:

37) Shortness of breath / shortness of breath

38) Feeling of tightness in the chest

39) Dry cough

40) Fatty cough

41) Tachycardia (abnormally high heart rate while resting or during low-intensity activities)

42) Arrhythmia / Palpitations (irregular heartbeat or feeling that the heart is beating very hard)

43) Myocarditis

44) Pericarditis

45) Heart failure

46) Myocardial infarction

47) Burning sensation in the chest

48) Chest pain

49) Wheezing

50) Bloody sputum

51) Thrombosis

Gastrointestinal symptoms:

52) Nausea

53) Vomiting

54) Diarrhea (occurrence of watery stools)

55) Constipation

56) Stomach burn

57) Other abdominal pain

Vascular / lymph node symptoms or diseases:

58) Hypertension

59) Hypotension

60) Stroke

○ If yes: Ischemic or hemorrhagic?

61) Lymphadenopathy (swollen, painful or inflamed lymph nodes)

62) Circulation disorders (swollen veins, heavy legs, etc.)

63) Spontaneous hematomas

Urinary symptoms:

64) Urinary pain

65) Recurrent urinary tract infections

66) Need for dialysis

Skin symptoms:

67) Skin rashes / lesions

68) Dry skin

69) Blue / purple / white or swollen fingers or toes

70) Other symptoms:

General questions

71) How do you feel today? (I feel good / I feel tired) / I feel bad)

72) Taking into account all the symptoms that can be attributed to COVID-19 that you have experienced in the last 30 days (frequency, intensity, impact on your life), would you say that you could live long term in your current state of health? (Yes/No)

73) Do the symptoms you have mentioned occur during crisis, with phases of improvement or worsening?

○ Yes, with daily crises

○ Yes, with at least 1 crisis per week

○ Yes, but less than one crisis per week

○ No, my symptoms are constant over time

○ I have no symptoms

74) How long do these crisis usually last?

○ Less than a day

○ More than a day but less than a week

○ More than a week but less than a month

○ I don't have crisis

75) What is your current pain level? (Rate from 1 to 10)

76) Since filling out your last questionnaire, have you consulted for a reason related to COVID-19? Yes No ->

If yes, for what reason?

77) Have you taken a lifestyle comparable to the one you had before the onset of symptoms associated with COVID-19? Yes/No

If no, why?

78) Did you return to your normal occupation? Yes / No / I am retired or unemployed

If no, have you been on sick leave following complications from COVID-19?

79) Since the diagnosis of COVID-19 or the onset of symptoms associated with COVID-19, have your relationships with those around you (family, friends): Worsened / Stayed the same / Improved

if worsened or improved: Why?

80) Since the diagnosis of COVID-19 or the onset of symptoms associated with COVID-19, would you say that your : appetite/physical activity/sleep: Worsened / Stayed the same / Improved

81) Do you currently benefit or have you benefited from a consultation dedicated to the Long COVID syndrom? Yes/no/Don’t know

- If yes: have you been offered rehabilitation? Yes/no/Don’t know
- If yes: at which institution? Rehazenter/Mondorf/CHNP/Other, precise

82) Since the diagnosis of COVID-19 or the onset of symptoms associated with COVID-19, have you noticed any changes in your menstrual cycles? Yes/No/Don’t know/ Not applicable

- If yes: heavier periods / irregular periods / early menopause/ Don’t know

83) Sleep, PSQI scale

1. During the past month, when have you usually gone to bed at night?
   USUAL BED TIME _____

2. During the past month, how long (in minutes) has it usually take you to fall asleep each night?
   NUMBER OF MINUTES _____

3. During the past month, when have you usually gotten up in the morning?
   USUAL GETTING UP TIME _____

4. During the past month, how many hours of actual sleep did you get at night? (This may be different than the number of hours you spend in bed.)
   HOURS OF SLEEP PER NIGHT _____

5. During the past month, how often have you had trouble sleeping because you...

	Not during the past month	Less than once a week	Once or twice a week	Three or more times a week
(a) ...cannot get to sleep within 30 minutes	☐	☐	☐	☐
(b) ...wake up in the middle of the night or early morning	☐	☐	☐	☐
(c) ...have to get up to use the bathroom	☐	☐	☐	☐
(d ...cannot breathe comfortably	☐	☐	☐	☐
(e) ...cough or snore loudly	☐	☐	☐	☐
(f) ...feel too cold	☐	☐	☐	☐
(g) ...feel too hot	☐	☐	☐	☐
(h) ...had bad dreams	☐	☐	☐	☐
(i) ...have pain	☐	☐	☐	☐
(j) Other reason(s), please describe				
How often during the past month have you had trouble sleeping because of this?	☐	☐	☐	☐

	Very good	Fairly good	Fairly bad	very bad
6. During the past month, how would you rate your sleep quality overall?	☐	☐	☐	☐

	Not during the past month	Less than once a week	Once or twice a week	Three or more times a week
7. During the past month, how often have you taken medicine (prescribed or "over the counter") to help you sleep?	☐	☐	☐	☐
8. During the past month, how often have you had trouble staying awake while driving, eating meals, or engaging in social activity?	☐	☐	☐	☐

	No problem at all	Only a very slight problem	Somewhat of a problem	A very big problem
9. During the past month, how much of a problem has it been for you to keep up enough enthusiasm to get things done?	☐	☐	☐	☐

	No bed partner or roommate	Partner/ roommate in other room	Partner in same room, but not same bed	Partner in same bed
10. During the past month, how much of a problem has it been for you to keep up enough enthusiasm to get things done?	☐	☐	☐	☐

If you have a roommate or bed partner, ask him/her how often in the past month you have had...

	Not during the past month	Less than once a week	Once or twice a week	Three or more times a week
(a) ...loud snoring	☐	☐	☐	☐
(b) ...long pauses between breaths while asleep	☐	☐	☐	☐
(c) ...legs twitching or jerking while you sleep	☐	☐	☐	☐
(d) ...episodes of disorientation or confusion during sleep	☐	☐	☐	☐
(e) Other restlessness while you sleep; please describe	☐	☐	☐	☐

84) Quality of life (SF12)

1. **In general, would you say your health is:**

☐ Excellent ☐ Very good ☐ Good ☐ Fair ☐ Poor

The following questions are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?

	YES, limited a lot	YES, limited a little	NO, not limited at all
2. **Moderate activities** such as moving a table, pushing a vacuum cleaner, bowling, or playing golf.	☐	☐	☐
3. Climbing **several** flights of stairs.	☐	☐	☐

During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

	YES	NO
4. **Accomplished less** than you would like.	☐	☐
5. Were limited in the **kind** of work or other activities.	☐	☐

During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

	YES	NO
6. **Accomplished less** than you would like.	☐	☐
7. Did work or activities **less carefully than usual**.	☐	☐

8. **During the past 4 weeks, how much did pain interfere with your normal work (including work outside the home and housework)?**

☐ Not at all ☐ A little bit ☐ Moderately ☐ Quite a bit ☐ Extremely

These questions are about how you have been feeling during the past 4 weeks.
For each question, please give the one answer that comes closest to the way you have been feeling.

How much of the time during the past 4 weeks...

	All of the time	Most of the time	A good bit of the time	Some of the time	A little of the time	None of the time
9. Have you felt calm & peaceful?	☐	☐	☐	☐	☐	☐
10. Did you have a lot of energy?	☐	☐	☐	☐	☐	☐
11. Have you felt down-hearted and blue?	☐	☐	☐	☐	☐	☐

12. **During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting friends, relatives, etc.)?**

☐ All of the time ☐ Most of the time ☐ Some of the time ☐ A little of the time ☐ None of the time

85) Respiratory quality of life (VQ11)

I suffer from breathlessness

I am worried about my respiratory condition

I feel my entourage (family, friends, etc.) misunderstands me

My respiratory condition prevents me from moving about as easily as I would like

I feel sleepy during the day

I feel unable to achieve my objectives

I quickly get tired when doing day-to-day activities

Physically, I am dissatisfied with what I can do

My respiratory disease disrupts my social life

I feel sad

My respiratory condition restricts my emotional life

86) Generalized Anxiety Disorder (GAD7)

1. Over the last two weeks how often have you been bothered by any of the following problems?

	Not at all (0)	**Several days** (1)	**More than half the days** (2)	**Nearly every day** (3)
a. Feeling nervous, anxious or on edge	☐	☐	☐	☐
b. Not being able to stop or control worrying	☐	☐	☐	☐
c. Worrying too much about different things	☐	☐	☐	☐
d. Trouble relaxing	☐	☐	☐	☐
e. Being so restless that is hard to sit still.	☐	☐	☐	☐
f. Becoming easily annoyed or irritable	☐	☐	☐	☐
g. Feeling afraid as if something awful might happen	☐	☐	☐	☐

Total Score: _____

87) Perceived stress scale (PSS4)

The questions in this scale ask you about your feelings and thoughts during THE LAST MONTH. In each case, please indicate your response by placing an "X" over the square representing HOW OFTEN you felt or thought a certain way.

	Never 0	Almost Never 1	Sometimes 2	Fairly Often 3	Very Often 4
1. In the last month, how often have you felt that you were unable to control the important things in your life?	☐	☐	☐	☐	☐
2. In the last month, how often have you felt confident about your ability to handle your personal problems?	☐	☐	☐	☐	☐
3. In the last month, how often have you felt that things were going your way?	☐	☐	☐	☐	☐
4. In the last month, how often have you felt difficulties were piling up so high that you could not overcome them?	☐	☐	☐	☐	☐

88) Fatigue severity scale (FSS9)

During the past week, I have found that:	Disagree ← → Agree						
1. My motivation is lower when I am fatigued.	1	2	3	4	5	6	7
2. Exercise brings on my fatigue.	1	2	3	4	5	6	7
3. I am easily fatigued.	1	2	3	4	5	6	7
4. Fatigue interferes with my physical functioning.	1	2	3	4	5	6	7
5. Fatigue causes frequent problems for me.	1	2	3	4	5	6	7
6. My fatigue prevents sustained physical functioning.	1	2	3	4	5	6	7
7. Fatigue interferes with carrying out certain duties and responsibilities.	1	2	3	4	5	6	7
8. Fatigue is among my three most disabling symptoms.	1	2	3	4	5	6	7
9. Fatigue interferes with my work, family, or social life.	1	2	3	4	5	6	7

Additional file 2:

Supplementary table 1: Determination of the optimal class number

	BIC	conv	loglik	Entropy
nclass_2	3877.439	1	-1900.806	0.6610324
nclass_3	3896.396	1	-1900.806	0.3558408
nclass_4	3915.353	1	-1900.806	0.2913690

The optimal number of classes is determined by the lowest BIC and the highest entropy.

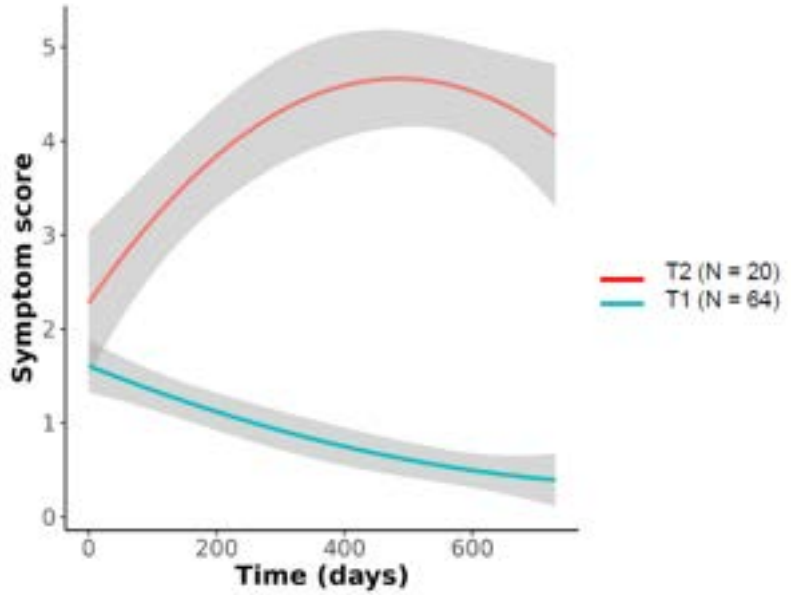
Additional file 3:

Supplementary table 2: Participant's individual characteristics in fatigue trajectories.

Variable Mean (±SD) for normally distributed continuous variables ; Median [min;max] for not ormally distributed continuous variables ; N (%) for categorical variables	**Class 1 (N=71)**	**Class 2 (N=110)**	**Class 3 (N=116)**	**Class 4 (N=258)**	**P-Value**
BMI (kg/m^2)	27[18,41.2]	24.5[18.6,39.6]	25.7[18.3,48.9]	24.7[16.7,55.1]	0.003
Female (yes)	36(50.70)	58(52.73)	78(67.24)	114(44.19)	<0.001
Age (years)	43.7±13.7	40.2±11.4	42.6±11.5	41.2±13.1	0.216
At least 2 comorbidities (yes)	7(9.86)	4(3.64)	11(9.48)	13(5.04)	0.131
Current smoker (yes)	16(22.54)	13(11.82)	20(17.24)	41(15.95)	0.219
Total symptom score initial	0[1,4]	2[1,6]	2[1,8]	0[1,3]	<0.001
Total symptom score M12	3[0,8]	0[0,5]	3[0,9]	0[0,4]	<0.001
Total symptom score M15	3[1,8]	0[0.4]	3[1,10]	0[0,3]	<0.001
Total symptom score M24	3[1,6]	0[0,4]	3[0,9]	0[0,6]	<0.001
Participants with at least 1 medication before COVID-19	36(50.7)	33(30)	46(39.66)	64(24.81)	<0.001
Sleep aids	2(2.82)	0(0.00)	4(3.48)	0(0.00)	0.004
Anti hypertensive	15(21.13)	9(8.18)	14(12.07)	20(7.75)	0.015
Anti pain/inflammation	5(7.04)	3(2.73)	3(2.61)	3(1.16)	0.051
Anti cholesterol	10(14.08)	5(4.55)	8(6.96)	18(6.98)	0.14
Diabetes treatment	5(7.04)	3(2.73)	3(2.59)	4(1.55)	0.109
Treatment for anxiety	3(4.23)	5(4.55)	12(10.34)	3(1.16)	<0.001
Antibiotic in the 2 months before COVID-19 initial	7(10.00)	16(14.55)	14(12.28)	20(7.81)	0.211
Anti coagulant	6(8.45)	2(1.82)	4(3.48)	6(2.33)	0.09
Anti depressant	2(2.82)	7(6.36)	10(8.62)	3(1.16)	<0.001
Anti convulsivant	1(1.41)	2(1.82)	1(0.86)	2(0.78)	0.731

P-values between each class were calculated using the student t-test for normally distributed continuous variables, the Wilcoxon test for non normally distributed continuous variables and the Fisher's exact test for categorical variables

Additional file 4:



Supplementary figure 1: Complete case analysis: total symptom score evolution in T1 and T2 from baseline up to 24 months after (in days) for 84 participants who completed the 4 timepoints

Annexe 6 : Protocole de l'étude UpcomingVoice

Protocol

Co-Design of a Voice-Based Digital Health Solution to Monitor Persisting Symptoms Related to COVID-19 (UpcomingVoice Study): Protocol for a Mixed Methods Study

Aurelie Fischer[1,2], MSc; Gloria A Aguayo[1], MD, PhD; Pauline Oustric[3], PhD; Laurent Morin[3], PhD; Jerome Larche[4], MD; Charles Benoy[5,6], PhD; Guy Fagherazzi[1], PhD

[1]Deep Digital Phenotyping Research Unit, Department of Precision Health, Luxembourg Institute of Health, Strassen, Luxembourg
[2]Université de Lorraine, Nancy, France
[3]Association Apres J20 COVID Long France, Luce, France
[4]Fédération des Acteurs de la Coordination en Santé-Occitanie, Hôpital La Grave, Toulouse, France
[5]Centre Hospitalier Neuro-Psychiatrique, Ettelbruck, Luxembourg
[6]Universitäre Psychiatrische Kliniken Basel, Basel, Switzerland

Abstract

Background: Between 10% and 20% of people with a COVID-19 infection will develop the so-called long COVID syndrome, which is characterized by fluctuating symptoms. Long COVID has a high impact on the quality of life of affected people, who often feel abandoned by the healthcare system and are demanding new tools to help them manage their symptoms. New digital monitoring solutions could allow them to visualize the evolution of their symptoms and could be tools to communicate with healthcare professionals (HCPs). The use of voice and vocal biomarkers could facilitate the accurate and objective monitoring of persisting and fluctuating symptoms. However, to assess the needs and ensure acceptance of this innovative approach by its potential users—people with persisting COVID-19–related symptoms, with or without a long COVID diagnosis, and HCPs involved in long COVID care—it is crucial to include them in the entire development process.

Objective: In the UpcomingVoice study, we aimed to define the most relevant aspects of daily life that people with long COVID would like to be improved, assess how the use of voice and vocal biomarkers could be a potential solution to help them, and determine the general specifications and specific items of a digital health solution to monitor long COVID symptoms using vocal biomarkers with its end users.

Methods: UpcomingVoice is a cross-sectional mixed methods study and consists of a quantitative web-based survey followed by a qualitative phase based on semistructured individual interviews and focus groups. People with long COVID and HCPs in charge of patients with long COVID will be invited to participate in this fully web-based study. The quantitative data collected from the survey will be analyzed using descriptive statistics. Qualitative data from the individual interviews and the focus groups will be transcribed and analyzed using a thematic analysis approach.

Results: The study was approved by the National Research Ethics Committee of Luxembourg (number 202208/04) in August 2022 and started in October 2022 with the launch of the web-based survey. Data collection will be completed in September 2023, and the results will be published in 2024.

Conclusions: This mixed methods study will identify the needs of people affected by long COVID in their daily lives and describe the main symptoms or problems that would need to be monitored and improved. We will determine how using voice and vocal biomarkers could meet these needs and codevelop a tailored voice-based digital health solution with its future end users. This project will contribute to improving the quality of life and care of people with long COVID. The potential transferability to other diseases will be explored, which will contribute to the deployment of vocal biomarkers in general.

Trial Registration: ClinicalTrials.gov NCT05546918; https://clinicaltrials.gov/ct2/show/NCT05546918

Introduction

Context

COVID-19 and Long COVID

The novel SARS-CoV-2 infected >532 million people and caused >6.3 million deaths worldwide as of June 2022 [1].

COVID-19 initial infection can take several forms, from asymptomatic to moderate or severe illness. Common symptoms are fever, cough, dyspnea, fatigue, loss of taste or smell, and gastrointestinal symptoms. Complications include acute respiratory distress syndrome, anemia, or acute cardiac injury [2].

After the acute phase, some people with COVID-19 develop a so-called *long COVID* or *post-COVID condition*. The World Health Organization's definition of this condition is as

follows: “Post-COVID-19 condition occurs in individuals with a history of probable or confirmed SARS-CoV-2 infection, usually 3 months from the onset of COVID-19 with symptoms that last for at least 2 months and cannot be explained by an alternative diagnosis. Common symptoms include fatigue, shortness of breath, and cognitive dysfunction but also others that generally have an impact on everyday functioning. Symptoms may be new onset, following initial recovery from an acute COVID-19 episode, or persist from the initial illness. Symptoms may also fluctuate or relapse over time” [3,4].

It has been estimated that a mean of 10% to 20% of patients with COVID-19, representing millions of people worldwide, will have persisting or relapsing symptoms >12 weeks after acute infection [5] with complaints such as tachycardia, extreme fatigue, or dyspnea as observed by Rubin [6]. In Luxembourg, we showed that 59% of COVID-19–infected people included in the Predi-COVID cohort study, who completed a detailed 1-year questionnaire, still declared 1 or more persisting symptoms. The number of persisting symptoms increased with the initial disease severity, and the quality of life of these participants was highly affected by sleep disorders (54%) and altered respiratory quality of life (12.9%) [7].

The COVID-19 pandemic accelerated the use of remote patient monitoring in clinical practice or research for safety and emergency reasons, justifying the need for innovative digital health solutions to monitor key parameters or symptoms related to COVID-19 or long COVID. A panel of experts from the National Institute for Health and Care Excellence recommended the development of telemonitoring and encouraged self-management of acute and long COVID symptoms in a tailored and accessible way for each patient [8].

Although long COVID is now a recognized illness in many countries, few dedicated consultations exist, and the awareness and scientific knowledge of this disease are still poor among not only the general population but also general practitioners. For these reasons, many people experiencing persisting symptoms after a COVID-19 infection (confirmed or not) may feel abandoned by or lost in the health care system, with difficulties in obtaining a long COVID diagnosis or obtaining specific support, and there is a demand among them for new monitoring solutions to objectify the symptom evolution and identify causes of symptom degradation. Moreover, people with long COVID experience difficulties in performing daily physical tasks, and many of them cannot engage in the same levels of activity or work as before [9].

Vocal Biomarkers of COVID-19 Symptoms

Using voice is an interesting approach for telemonitoring, as it is easy to collect, quick, and energy efficient, inducing less fatigue on the patients. Moreover, the use of voice-based

technologies is increasing swiftly; in 2019, 31% of smartphone users worldwide used voice technology at least once a week, and 20% of queries on Google's mobile app and Android devices were voice searches [10].

Voice is a rich source of health-related information, as many voice features can be related to symptoms or health status [11-14]. In addition, because voice analysis involves highly complex methods for processing audio features, this kind of development might also be capable of detecting subtle changes associated with COVID-19 symptoms [10].

For example, voice features have previously been associated with a COVID-19 infection or a consequence of COVID-19 infection [15,16]. COVID-19 infection or complications can affect the voice through different mechanisms. Indeed, respiratory insufficiency can lead to reduced airflow and thus to changes in voice parameters. Other studies showed that voice quality was reduced in patients with COVID-19 owing to repeated cough, laryngeal or pharyngeal erythema, or sore throat [17-19].

In Luxembourg, participants from the Predi-COVID hybrid prospective cohort study were invited to perform voice recordings at the same time as they completed web-based questionnaires regarding their symptoms and health status. To date, almost 6000 voice recordings from more than >500 patients with COVID-19 have already been collected in the Predi-COVID study [2]. These voice recordings have been analyzed, and promising vocal biomarker candidates for fatigue, loss of taste and smell, and symptomatic status in people with a COVID-19 infection have already been identified [21-23].

The question now is to determine how these vocal biomarkers could help and improve the quality of life of people with long COVID and how they could be implemented in a digital health solution such as a smartphone app. For example, people with COVID-19 or long COVID symptoms could record their voice in real-life settings by following basic instructions such as "read a pre-defined text, count from 1 to 20, or say a vowel as long as possible," and the app would analyze voice features and give them back a result for one or several vocal biomarkers. Importantly, to ensure that the developed digital solution using vocal biomarkers is meaningful and useful for the patients, our study involves them at an early stage together with health care professionals (HCPs).

Objectives

The UpcomingVoice project aims to co-design a voice-based digital health solution for screening and self-monitoring of frequently reported COVID-19–related symptoms with its end users, namely, (1) people with persisting COVID-19–related symptoms, with or without a long COVID diagnosis confirmed by an HCP, and (2) HCPs involved in the care of people

with long COVID. The goal of this study is to develop a digital health solution that really matters and helps people with long COVID in their daily lives and could be a complementary support in addition to the medical care that they could benefit from.

To achieve this main objective, the UpcomingVoice study will (1) explore the impact of long COVID in the everyday lives of people with long COVID and the needs of people with long COVID to improve their quality of life and manage their symptoms; (2) explore the extent to which the use of voice and, in particular, vocal biomarkers could propose a solution to the patient's needs (ie, we will investigate the expectancies, acceptability, fears, barriers, and leverages regarding using voice to self-monitor or screen for long COVID symptoms); and (3) define the specifications that such a mobile app should meet so that it could be recommended to patients by HCPs and be considered acceptable and useful by its intended users in terms of technological aspects (ie, type of device, type of voice recordings, etc), frequency of use, design, etc.

As a secondary aim, we will also assess the transferability potential of the results of this study in the context of COVID-19 to other diseases (eg, monitoring fatigue could also be used in the context of cancer).

Methods

Study Design

This cross-sectional study is based on a mixed method sequential explanatory design. It consists of 2 successive phases: a quantitative phase (descriptive approach), followed by a qualitative phase (inductive pragmatic approach). The study design is summarized in Figure 1. The research questions addressed in the 2 phases of the study are listed in Table 1.

Table 1. Research questions (RQs) addressed in phase 1 and phase 2.

Study phase and objectives	Measure instrument	RQs	Questions of HCP[a] survey	Questions of PWLC[b] survey
Phase 1 (quantitative phase)				

	To understand the needs of PWLC and to assess acceptability, expectancies, and fears regarding the use of voice to meet these needs	Web-based anonymous survey (different questions for PWLC and HCPs)	· 1. What are the demographic and sociodemographic characteristics of the participants?	· 1-3	· 1-3
	To define the general specifications of the digital health solution based on vocal biomarkers	Web-based anonymous survey (different questions for PWLC and HCPs)	· 2. What is the acceptance rate of the use of voice for symptom monitoring? · 3. What proportion of participants already heard about the notion of vocal biomarker? · 4. What are the most important LC[c] symptoms to be monitored? · 5. What are the needs, fears, expectations, and use cases for PWLCs and HCPs regarding the use of voice for symptom monitoring? · 6. What are the main specifications of a smartphone app based on VB[d] (engagement, functionalities, esthetics, information content, and subjective items) · 7. What are the most important aspects of the smartphone app? · 8. Should such an app be reimbursed or be paid by the patient? What cost would be acceptable? · 9. To what extent do COVID-19 infection and vaccination have an impact on the opinions and needs regarding the use of voice for symptom monitoring?	· 5[e] · 4[f] · 11-12[g] · 6-10; 14[h] · 19-35[i] · 13[j] · 15-18[k]	· 17[e] · 16[f] · 24-25[g] · 18-27[h] · 32-57[i] · 26[j] · 28-31[k] · 4-15[l]

Phase 2 (qualitative phase)					
	To further clarify PWLC's needs and their expectations and fears regarding voice-based technologies	Individual semistructured interviews with PWLC and HCPs	· 10. What are the needs and problems of PWLC in their everyday lives and how could the use of vocal biomarkers bring a solution to these problems?	N/A[m]	N/A
	To define the specifications of the voice-based digital health solution	Individual semistructured interviews with PWLC and HCPs	· 11. What are the detailed specifications of a smartphone app based on VB (engagement, functionalities, esthetics, information content, and subjective items)? · 12. What are the best app items and response options?	N/A	N/A
	To develop a smartphone app prototype	2 focus groups with a panel of 8-10 PWLC and HCPs together	· 13. How is the smartphone app prototype perceived? · 14. What are the improvements suggested by PWLC and HCPs?	N/A	N/A

[a]HCP: health care professional.
[b]PWLC: people with long COVID.
[c]LC: long COVID.
[d]VB: vocal biomarkers.
[e]Survey questions related to RQ2.
[f]Survey questions related to RQ3.
[g]Survey questions related to RQ4.
[h]Survey questions related to RQ5.
[i]Survey questions related to RQ6.
[j]Survey questions related to RQ7.
[k]Survey questions related to RQ8.
[l]Survey questions related to RQ9.
[m]N/A: not applicable.

A conceptual framework of the key areas of daily functioning will be elaborated based on the literature review, the survey, and individual interview results as previously described [24], and the framework will be used to identify the most relevant use cases of the digital health solution. The general design of the digital health solution will be defined using the survey results, and its specific items and response modalities will be further elaborated during the qualitative phase.

To study the transferability potential of the digital health solution that will be developed, we will provide an extensive description of our study population and their daily life concerns to identify other populations, diseases, or conditions with similar characteristics (in terms of impact on quality of life, main symptoms, and intensity of symptoms) that could also benefit from a voice-based technology. Subsequently, we will use the APEASE (Affordability, Practicability, Effectiveness, Acceptability, Side Effects, and Equity) criteria and other relevant checklists previously described for the evaluation of transferability of health technology assessments [25,26] to assess the potential transferability of each use case or the main characteristics of the solution identified in the 2 phases of the study to the other fields.

Phase 1 (Quantitative)

The first phase, the quantitative phase, consists of a web-based anonymous survey.

The objective of this survey is to define the main aspects of daily life impacted by long COVID and the needs of people with long COVID and to assess the acceptability and expectancy toward the use of vocal biomarkers for the symptom monitoring of long COVID. The general outline of the digital health solution based on vocal biomarkers will be defined based on these results.

The following dimensions adapted from the user version of the Mobile Application Rating Scale (uMARS) [27] will be covered by different questions: engagement, functionality, aesthetics, information, and subjective items (“Would you recommend,” “Would you be interested in,” etc).

People With Long COVID

In addition to the above-mentioned items, the survey for people with long COVID has been designed to further identify the most impairing symptoms for their quality of life, to assess whether the severity of their initial infection or of their long COVID symptoms or their vaccination status has an impact on their opinions and needs regarding a new health technology based on voice, and to assess how the use of vocal biomarkers could be of interest for the management of their symptoms. The detailed questionnaire for people with long COVID is provided in Multimedia Appendix 1.

HCP Survey

The HCP survey is focused on assessing the needs of their patients with long COVID in their daily lives and how they foresee the use of vocal biomarkers of long COVID symptoms in the

management of their patients and in the health care system. The detailed HCP questionnaire is provided in Multimedia Appendix 2. Completing the survey will take a maximum of 20 minutes to ensure a high acceptability rate, particularly in people with long COVID who frequently experience fatigue and trouble concentrating. The participants are expected to complete the survey once.

Web-based survey data will be collected and managed using REDCap (*Research Electronic Data Capture*; *Vanderbilt University)* tools hosted at the Luxembourg Institute of Health (LIH) [28].

Phase 2 (Qualitative)

The second phase is based on qualitative methods and consists of semistructured individual interviews and focus groups. Participants in this phase will have either participated in the web-based survey or not.

Individual interviews will be conducted to deeply understand the impact of long COVID on daily life, the needs of people with long COVID, and the potential interest in using vocal biomarkers to meet these needs, and define use cases for the digital health solution. The maximum duration for the interview will be 60 minutes.

Following the individual interviews, 2 successive focus groups (maximum duration of 90 minutes) will be organized with a panel comprising people with long COVID and HCPs. During the first focus group, the objective is to present 1 or 2 prototypes of the digital health solution and to collect users' comments and suggestions. The prototype or prototypes will be improved according to the collected comments and validated during the interaction with the second focus group.

The individual interview and focus group guides will be elaborated based on the insights obtained from the survey in phase 1. The questions will be asked in a general manner to ensure the identification of the most relevant aspects for people with long COVID in their daily lives and to minimize the expectancy and confirmatory bias.

Using teleconferencing software, interviews and focus groups will be conducted by a qualified interviewer and a moderator, respectively, both trained in qualitative methods.

Patient Public Involvement

The INVOLVE recommendations [29] will be followed to involve people with long COVID and HCPs in the most efficient way during the entire course of the project. First, we contacted a

representative of a long COVID association and an HCP in charge of people with long COVID and presented them with the study aims and the participation procedures. They were then invited to review the study protocol, participant documents, and questionnaires before the study implementation.

Second, we will also follow the framework of core principles for selecting and developing measurements in digital health developed by Manta et al [30]. The 4 levels of this framework to evaluate the meaningfulness of digital biomarkers for patients are the Meaningful Aspect of Health, the Concept of Interest, the Outcome to be measured, and the End point.

In this perspective, the UpcomingVoice study design is a co-design process that will continuously involve patients with long COVID and HCPs.

At each step, the participants will receive feedback on the results from the previous step: during individual interviews, feedback on survey results will be given, and during the first focus group, results from the individual interviews will be provided and discussed.

Finally, participants with long COVID and HCPs willing to be even more deeply involved in our research will be invited to coauthor the scientific articles presenting the results of this research and to promote the results of the study to the lay audience.

Study Participants

Inclusion Criteria

Participation in the study will be proposed to adults (men and women) with persisting COVID-19–related symptoms, with or without a long COVID diagnosis, and to HCPs in charge of patients with long COVID (doctors, nurses, psychologists, physiotherapists, etc).

All participants must be aged >18 years, have a sufficient level of speaking and comprehension in French, and have internet access to participate in the web-based survey and the web-based individual interviews and focus groups.

Seeking a representative sample of people is not part of qualitative research aims [31]; however, to increase the variations in our study population, the inclusions in the qualitative phase will be monitored to obtain an equal representation of different age and gender categories.

Recruitment and Enrollment

Before starting each part of the study, pilot phases will be performed to correct potential technical issues and ensure that the time for the completion of the survey and the interviews

and focus groups do not exceed the expected duration (20 minutes for the survey and 60 and 90 minutes for individual interviews and focus groups, respectively).

The recruitment of participants will be based on electronic and paper flyers disseminated via social media, long COVID consultations, and long COVID patient associations in Luxembourg and neighboring countries (eg, France). Participation will also be proposed to participants from the Predi-COVID cohort study [2] who declared persisting symptoms at 1 year.

Flyers present information about the study and contain a direct link to the web-based survey so that interested persons can directly participate. The contact details of the study assistant at the LIH are also provided so that people can contact her directly if they are interested in participating in the qualitative part of the study.

People who expressed their interest in participating in the second part of the study, either after the survey completion or by contacting us directly, will be contacted by the study assistant. The study assistant will provide additional information regarding the second part of the project, and if the person confirms his or her interest in participating, an electronic ethical information notice and consent form will be sent to the participant. After signing the electronic consent form, the participant will be able to download a copy of the document. The study assistant will contact him or her to organize the individual interview. Electronic informed consent will be obtained before any data collection for the second part of the study.

Once the study assistant has received the notification that the participant has a valid signed consent, they will contact her or him to organize the web-based individual interview.

When all the individual interviews are completed, a panel comprising 7 to 8 people with long COVID and 1 to 2 HCPs will be constituted among the participants who expressed their willingness to participate in the focus group sessions.

When all individual interviews and focus groups are conducted, a restitution meeting will be organized. Everyone who has expressed an interest in this project will be invited.

Sample Size Calculation

For the quantitative part of the study, a formal size calculation was not feasible, given the exploratory nature of the study. The quantitative data from the survey will be exclusively used for descriptive purposes and to evaluate the different topics' importance for the participants.

For the qualitative part, no gold standard for the “right” sample size exists and depends closely on the nature and aims of the study [32]. Qualitative studies do not aim to estimate magnitudes or generalize the results to a larger population but evaluate patterns in a data set. Participants will be sampled until data saturation is reached, which can be formalized as the point at which new data repeat what were already collected [33,34]. A recent systematic review showed that data saturation was achieved with a mean of 9 to 17 interviews but with high disparities between studies [35]. The ideal size for focus groups is estimated to be between 6 and 10 persons, particularly when including participants with high knowledge of the problem, which is the case for people with long COVID and HCPs [36].

To achieve data saturation, at least 15 people with long COVID and 5 HCPs involved in the management of patients with long COVID will participate in the individual semistructured interviews. However, these are estimations, and we will recruit participants until data saturation is reached.

A maximum of 7 to 8 people with long COVID and 1 to 2 HCPs will participate in 2 successive focus groups with an iterative co-design objective. Participants in the focus groups may be the same as those in the individual part, but this is not mandatory.

Data Analysis

Phase 1: Quantitative Part

Quantitative data from the web-based survey will be analyzed using descriptive statistics methods. Continuous variables will be presented as means with SDs. Discrete numeric variables and ordinal variables will be presented as medians with IQRs. Nominal and categorical variables will be presented as proportions. Standard statistical methods will be followed.

Phase 2: Qualitative Part

Qualitative data from the semistructured interviews and focus groups will be audio recorded and transcribed verbatim. Transcripts will be checked for accuracy to ensure data quality and analyzed by using the thematic analysis approach [37]. This is the most commonly used method in qualitative research and allows one to identify, analyze, and report patterns within the data through the following five major steps:

1. Compiling: during this step, the recorded interviews are transcribed and organized.

2. Disassembling to allow for the creation of meaningful codes by identifying themes, concepts, or ideas. The coding process will be iterative and based on grounded theory [38].

3. Reassembling to further dive into the themes and creating subthemes. The coding process will be iterative until saturation.

4. Interpreting: thematic maps or code clusters will be created to analyze the relationship between the themes and thus answer the research question.

5. Concluding: conclusions are not generalizable to the population; they are merely an analytical generalization—a guide to assess how the findings can be transferred and applied. Indeed, qualitative research typically aims for in-depth understanding rather than statistical generalizability.

The MAXQDA software (VERBI Software GmbH) will be used for the transcription of the audio files, for coding and qualitative analysis.

Data analysis will be performed by the study investigators. In particular, coding will be performed by 2 researchers, including an experienced qualitative researcher.

On the basis of the data, the codes and the overall coding structure will be created by the researchers themselves, which is justified because the analysis will be based on grounded theory. Common themes and subthemes from the qualitative data will be identified using a thematic analysis. The first step will be performing coding from scratch. To facilitate coding, we will also create memos. We will clearly define each code that both researchers should approve. The 2 researchers will also have regular meetings to decide whether to merge or exclude codes (data reduction). The themes will be created by grouping similar codes. Subsequently, each theme will be revised to ensure a good definition that aligns with the required codes and to avoid redundancy or missing themes. In addition, we will define themes automatically, using nonsupervised machine learning algorithms (natural language processing using topic modelling R package [R Foundation for Statistical Computing]).

We will assess the intercoder reliability by double-blinded coding performed by 2 independent researchers. The Intercoder Agreement function from the MAXQDA software will be used to compare the coding of the 2 researchers involved in the coding process.

Mixed Methods Integration and Analysis

The integration of the quantitative and qualitative data will be done twice. First, in the design of the qualitative phase of the study: the findings from the quantitative phase will be used to

develop the semistructured interview guide that is to be used for the qualitative phase. In particular, the most important topics raised by the survey will be addressed first in the discussions.

Second, after the end of the qualitative phase, the results from the 2 phases will be integrated to triangulate the findings.

Ethics Approval, Informed Consent, and Participation

This study was approved by the National Research Ethics Committee of Luxembourg (study number 202208/04) in August 2022. Interested persons can participate either in the 2 parts of the project or only in 1 part.

In the first part (quantitative web-based survey), each person who receives the link can participate. Detailed information on the survey is provided before acceding to the survey. Informed consent is not required, as this part is anonymous.

An extensive electronic information and informed consent form will be sent to people interested in participating in the qualitative part of the study. Electronically signed informed consent will be collected before any data collection for the second part of the study. REDCap will be used to manage the electronic consent forms.

Participants enrolled in the qualitative part may terminate their participation at any time by contacting the study assistant, and they will not be required to provide reasons for the decision. The assistant will complete the withdrawal form for our own records to ensure that the participant has declared whether LIH may continue to use the information that was already collected or if it should be destroyed (if applicable).

Participation in the individual interviews and the focus groups may tire some participants, particularly people with long COVID. To prevent this risk, regular pauses will be proposed during the interviews and focus groups, and the maximum duration of 60 and 90 minutes, respectively, will be strictly respected. The interviewer and the moderator will be the investigators trained in qualitative research methods and confidentiality.

There will be no financial incentive for participation in one or both parts of the study. Participants will be informed that they will not have any direct benefit but will contribute to a research effort that could make an impact not only in the care of patients with long COVID but also that of patients with other chronic diseases in the following years.

Data Management—Confidentiality

The personal data of participants will be protected according to the *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*. The participants will receive detailed information on data privacy and protection in the information sheets from both parts of the study. In particular, they will be informed of how the data will be collected and stored, what security measures will be in place, and who will be authorized to access the data. They will also be asked to give their consent for the potential secondary use of their data for other research projects and will be informed that in case they agree, only their anonymized data will be shared with other researchers. Participants in the anonymous survey will be informed that their data will potentially be used for other research studies and shared with other researchers. The participants will be informed that they can contact the institution's data protection officer if they have questions related to data privacy and protection. The contact details of the data protection officer are provided in the electronic information and consent form.

The survey will be anonymous and accessible by following a direct link to the web-based survey platforms (no application to download).

No account, log-in, or password will have to be created by the participants. There will be no link between the email addresses collected after the survey completion from people interested in the second part of the study and the survey answers.

The data collected from the qualitative part will be pseudonymized. Participants who express their interest in participating in the second part of the project will receive a unique link to access the electronic consent form. After the completion of the electronic consent form, a unique study number will be assigned to them.

Data collected during the individual interviews and focus groups will never be linked to a participant's name or email address. A nickname will be used to refer to each participant during the exchanges. Interviews and focus groups will be recorded and transcribed using a transcription tool. All records will be destroyed directly after the transcription. The interviewer will be responsible for audio transcription and destruction.

The nominative data collected via the consent form will be stored on LIH internal servers, separated from the pseudonymized data from the interviews and focus group notes. All data stored on the LIH servers will be encrypted using the Transport Layer Security (TLS) encryption protocol. Access to the data will be restricted by user accounts and will be

provided by the principal investigator of the study to a limited number of people. Access logs will be implemented to control data accesses and ensure high data security.

The contact data (email address and identification data) of the participants will be deleted 2 years after the individual interview. The pseudonymized data will be anonymized after 2 years by deleting the correspondence table and deleted after 10 years according to the LIH standard procedure for archiving.

LIH internal procedures will be followed to ensure data destruction after the retention period ends.

Results

The web-based survey was launched in October 2022. The individual interviews and focus groups will start after the end of the survey in March 2023. The end of data collection is expected in September 2023, and the results will be published at the end of 2023 or in 2024.

Discussion

This paper presents a protocol for a mixed methods study that aims to codevelop a digital health solution based on vocal biomarkers to monitor long COVID symptoms and conditions with its potential end users, namely, people with long COVID and HCPs in charge of patients with long COVID.

Expected Findings

Long COVID is a complex disease with a high impact on the daily life and quality of life of affected people [39,40]. Patients often feel abandoned by the health care system and demand for support and new tools to track their symptoms and relapses. This could be helpful to identify the potential causal factors of symptom exacerbation and relapses to improve pacing and prevention. In particular, postexertional symptom exacerbation is a well-known problem in long COVID and can be defined as the worsening of symptoms following a physical or mental activity [41]. Postexertional symptom exacerbation needs to be carefully monitored, for example, to ensure a safe rehabilitation program or to personalize the daily activities to prevent symptom degradation.

HCPs also recognize that there is a need for more integrated health care structures to manage patients with long COVID. Additional tools to ensure an efficient and safe rehabilitation would also be helpful for both people with long COVID and HCPs, as some patients with long COVID experience postexertional malaise [41] and may need a tailored

rehabilitation program. Moreover, with new therapies and drugs currently in development for long COVID treatment, this tool might also be used as a companion tool to monitor the health status of people with long COVID during clinical trials.

The use of vocal biomarkers to monitor long COVID symptoms seems of high interest because voice samples are noninvasively collected and are easy to collect compared with the completion of tedious self-reported questionnaires. They could also objectify a subjective condition such as fatigue or mental health and facilitate the communication between patients and HCPs [10,42]. Moreover, audio recordings may be more adapted and accurate than traditional paper questionnaires for people with neuropsychological disorders or other patients who are chronically ill with reduced functioning [43,44].

Vocal biomarkers of several symptoms related to COVID-19 and long COVID have been identified previously and need to be integrated into a digital health solution before being used in daily life or clinical practice. We recently proposed an optimal pipeline and recommendations to achieve the successful implementation of vocal biomarkers in practice, with their potential benefits and limitations in the context of long COVID [42]. The identification of the problem and the symptoms to be monitored and the involvement of the end users as soon as possible in the development process are key points. The protocol of the study described in this paper is based on these recommendations.

The findings of this study will have several potential implications. First, we will take advantage of combining quantitative and qualitative data in a mixed methods design to describe the needs of people with long COVID in depth and understand how the use of vocal biomarkers in a voice-based digital health solution could solve their problems. The visualization of the results in the digital health solution will also be discussed and tailored to best fit its end users. The tool will be developed with an interoperability objective to ensure its future dissemination to a maximum number of people.

Second, the findings could potentially be transferred to other diseases, and this is of the highest importance for the future of vocal biomarkers and voice-based health technologies to understand the extent of transferability.

Finally, the concrete output of the study will be a voice-based digital health solution that will provide support to people with long COVID in their daily lives and improve their quality of life.

Strengths and Limitations

One of the strengths of this study is that patients with long COVID and HCPs in charge of patients with long COVID have been involved in the study design and will be involved in the entire study course.

The use of a mixed methods design, using quantitative and qualitative approaches, will provide an extensive description of the future user's expectancies, barriers, and leverages regarding the use of vocal biomarkers in a digital health solution and will provide the specifications of a digital health solution optimally designed with and for its end users.

The results will reveal the transferability potential of the use of vocal biomarkers for other diseases and contribute to the expansion of this recent technology.

This study has some limitations. First, the fully web-based format of the study may limit the participation of some population categories (such as people with low eHealth literacy) but by contrast, may also facilitate the participation of people who are geographically distant or experience more impacting symptoms.

Another potential limitation is that even if the chosen sample size should ensure achieving data saturation, it may probably not represent all different subpopulations with long COVID. However, this is not the objective of this mixed methods study. The opinions collected should be varied enough to ensure the development of a digital solution that provides help and support to all patients with long COVID.

Finally, mixing people with long COVID and HCPs in the focus groups may induce conformity, dominance, and shyness biases. However, people with long COVID are highly involved in the management of their disease and generally interact easily with HCPs. The moderator's role will also be crucial in minimizing the risks of bias during focus groups by communicating from the start the objectives of the discussions and by ensuring that each participant receives the same amount of speaking time. The cultural bias will also be reduced by the inclusion of participants from different countries.

Conclusions

This mixed methods study will identify the needs of people affected by long COVID in their daily lives and describe the main symptoms or problems that would need to be monitored and improved. We will determine how the use of vocal biomarkers in a digital health solution could meet these needs and further describe the end user's expectations and barriers

regarding the use of this new digital technology to monitor COVID-19–related symptoms. Finally, our study will ensure the development of a digital health solution that will match the actual needs of people with long COVID by involving them in the design process from the beginning. The study will also provide important insights into the transferability of the results to other diseases and contribute to the deployment of vocal biomarkers in general.

Publication Plan

The study results will be disseminated at the end of 2023 in a scientific article submitted to an international peer-reviewed journal, to long COVID patient associations, and through social media communications to the lay audience.

Acknowledgments

This study is funded by the Luxembourg National Research Fund (FNR; Predi-COVID, grant 14716273) and the Luxembourg Institute of Health. The Predi-COVID study was supported by the FNR (Predi-COVID, grant 14716273), the André Losch Foundation, and the European Regional Development Fund (FEDER, convention 2018-04-026-21). The authors are thankful to all the participants of the Predi-COVID study. The authors also acknowledge the involvement of the interdisciplinary and interinstitutional study team that contributed to Predi-COVID. The full list of the Predi-COVID team can be found [45].

Data Availability

The data sets generated during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

AF, GA, and GF contributed to the study design, literature review, and drafting of the manuscript. PO, LM, and CB contributed to the study design and reviewed the study documents. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Abbreviations

APEASE: Affordability, Practicability, Effectiveness, Acceptability, Side Effects, and Equity

HCP: health care professional

LIH: Luxembourg Institute of Health

REDCap: Research Electronic Data Capture

TLS: Transport Layer Security

uMARS: user version of the Mobile Application Rating Scale

Annexe 7 : Matériel supplémentaire de l'article 4

Supplementary file 1

Semi-structured interview guide

1/ Introduction

Introduce the interviewer, the study objectives and the interview procedure.

Remind participants that they have given their consent to participate, and that the interview will be audio recorded. Check that the participant still agrees.

2/ Interview

Collect socio-demographic data:

- participant category: PWLC or HCP
- age
- education level

Following questions are used to guide the interview but are not exhaustive and are adaptable to the participant's answers and reactions.

Impact of Long COVID

- What are your (your patients) most impacting symptoms?
- What difficulties do you (your patients) encounter in the daily life
- Would you be interested in a symptom monitoring app for you or your patients, and why? (do you already track your symptoms? In which manner?)
- How could such an app help you/your patients on a daily basis?
- What advantages do you see in using a symptom monitoring application?
- Did Long COVID have an impact on return to work or leisure activities (for you/your patients)? Could the app provide any help in this area?

Voice use

- Imagine that you (or your patients) use an app to monitor symptoms. How do you think voice could be used in this way?

- Do you think it's possible to measure symptoms in the voice using voice biomarkers? What advantages do you see?

- Do you have any fears about such a voice-based application? If so, what are they? What do you see as the drawbacks or barriers?

App characteristics

- What frequency of use do you consider acceptable for a symptom monitoring app? for voice recordings? for questionnaires?

- What should be included in addition to symptom monitoring?

- What do you think of the possibility of integrating a module offering rehabilitation exercises? How do you see it?

- What do you think should be implemented in the application if a worsening of symptoms is detected?

- What do you think about integrating personalized advice into the application? Is it important to you?

- Should the application include information on COVID Long? What do you think?

- How do you imagine the application in terms of interactions with other people and the sharing of results?

- What features could help you/your patients use the application over the long term?

- Would you be willing to make several successive voice recordings? If so, what kind of recordings (standardized, free, etc..)

Focus group guide

1/ Introduction

Presentation of the focus group moderator, reminder of study aims and status.

2/ Long COVID App demo

3/ Discussion of different app functionalities

- Self-assessment of overall health
- Self-assessment of symptom intensity
- User journey for daily health assessment
- In-depth symptom questionnaires with scores
- Audio recordings for research
- Voice diary
- Calendar view
- Graphical display of results
- Life journal
- Resources displayed on home page
- Integration of chatbot

4/ Open discussion - feedbacks

Stimulating questions:

- What is your first impression of the application?
- Does the application correspond to the idea you had of it? Why or why not?
- Do you think this application meets the expectations of people affected by Long COVID?
- Do you see any negative points or potential barriers to using the app?
- What do you think of the content we integrated into the prototype?

Supplementary table 1: Long COVID impact on the lives of affected people

	Survey results	Interviews themes	Subthemes	Verbatim
Medical aspect / Care	60% of the participants declared that access to care was difficult or very difficult.	Access to care	Complicated, inequities, care unavailability	"Some people don't have access to the same care as the treatments I had access to." (PWLC 19) "We're in a very, very unequal situation" (PWLC10) "They told me that they had already enough patients, so they didn't take me" (PWLC 43). "I was lucky, I found what I needed so for the moment I'm under care." (PWLC 32) "(There is) not always a well-defined care pathway:" (HCP12) "Patients have great difficulty finding a speech therapist for cognitive rehabilitation." (PWLC16) "Long COVID consultation exist, but sometimes it's discovered completely by chance, and that's a pity because, as people say, "I've wasted an incredible amount of time, if I'd known I'd have gone quicker". (HCP 12)
	64% of participants with LC already benefited from a specialized LC consultation.	Disease recognition	Long-term illness recognition, diagnosis, Poorly known disease, relations PWLC-HCPs, no treatment	"My doctor don't believe in my symptoms", "Long-term illnesses are laughable. If you don't have an attending physician who sticks by you and fights for you, you'll never get over it." (PWLC 10) "Some people don't know they're affected by Long COVID, so they wonder what's going on and don't really understand." "I made my diagnosis all by myself " " it's not recognized as a long-term illness, but some people are able to obtain it all the same." (PWLC 55) "We even have the possibility of evolving into a Long-Term Illness, so it can be recognized as such. It's difficult to get it, but you have to do it." (HCP 37)
				"Even neuropsychologists are sometimes disarmed regarding neurocognitives troubles" (HCP 7) "General practitioners not always know the LC specialized care" (HCP), "Long Covid is not very well recognized or known." (PWLC 51) "In the beginning, it was complicated because the treating doctors didn't know much about the disease. So it was almost up to us to explain the disease to the doctor." (PWLC 32) "Physicians are overwhelmed" (PWLC), "My physiotherapist is good, but he lacks of experience with LC" (PWLC 43) "They don't know how to manage that pain." (PWLC 51) "Are these patients being ignored? That's the question." (HCP 37) "Health professionals are overworked, it's difficult to get appointments. And when they're overwhelmed like that, they become

				maltreating." (PWLC 10) "The medical desert, the lack of understanding we can experience with family and friends" (PWLC55)
		Rehabilitation	Pacing, lots of visits, loss of taste and smell, physical activity, neurocognitive, not always efficient or adapted	"I'm practicing pacing. So I get up in the morning, I rest for 1/4 hour, I do half an hour of activity, I rest for 1/4 hour." (PWLC 44) "We tell them to take breaks, not to do 2 things at the same time, we tell them these pacing measures, to prioritize activities " (HCP 66) "(About rehabilitation) it's going to involve appointments. I'm not sure I'll be able to go if they tell me at 3pm to be there and all that." (PWLC 32) "You can tell people to do sudoku or memorys, but it won't necessarily help them with their shopping list or their job." (HCP 7) "During the sessions I keep yawning, because it tires me out." (PWLC 23) "I believed in exercise rehabilitation, and then I had a relapse:" (PWLC 43)
Daily life impact	88% of the participants declared that they did not recover the same level of professional and leisure activity as before COVID-19 infection	Feelings	Anxiety	"It's a source of psychological difficulties for them, as they find themselves in a completely unexpected state." (HCP12) "I'm going to lose what little independence I have today. And that's a huge stress, I admit." (PWLC57) "I see, in discussion groups, there are a lot of fears about the future." (PWLC16)
			Feel alone or misunderstood	“Honestly, I'm lost. I feel completely alone and powerless” (PWLC 57). “The difficulty is feeling alone in the face of this disease and all its unknowns” (PWLC58). "we've all had doctors tell us "it's all in your head, you need to see a psychiatrist, you're not ill". And for me, that was the hardest part of my illness. Apart from the physical suffering, to feel diminished, to feel rejected by the medical profession, for me, that was something terrible to live with." (PWLC 55)
		Financial issues	Economic insecurity, cost treatments	"Economically I have this insecurity and precariousness, both medical and economic." (PWLC 23) "There's always a cost involved, since aromatherapy is... they're going to spend a fortune buying a kit with fragrances, and if they use it incorrectly, there are also risks". (HCP 7) " I'm a bit depressed and I'm anxious about the future, because if things go wrong, I'll have to sell my house (PWLC 23)"

Work issues	Unability to work, work time adaptation, risk of loosing his/her position	"I had one lady, for example, who described that she'd had 2 warnings at work because she'd made mistakes. Mistakes where she simply forgot she'd already done something, or forgot to do something she'd been asked to do simply because she didn't remember it." (HCP 66) "As I do intellectual work, I have problems concentrating, I have to make enormous efforts to stay focused and in the end I leave my working days completely exhausted." (PWLC 61) "I'm retired now, but I don't think I could have worked if I had to" (PWLC 43) "On a day-to-day basis, cognitive problems and chronic fatigue very often lead to sick leave or part-time work." (HCP 12) "Fatigue and neurocognitive disorders generally prevent them from carrying out activities and working." (HCP 9) "I've never been on sick leave in my life, I've always worked, and now I haven't worked in the last 2 1/2 years." (PWLC 23)
Social life	Isolation, less pleasure or inability to participate in social activities	" I can't be in a noisy environment, so I can't go to restaurants with friends. I can invite people at home, but not many." (PWLC 16) "There's less pleasure in eating" (PWLC 4) "I limit myself on a social level. I can't really go out anymore, I can't enjoy the things I like, I used to play sports, now I can't ride my bike. " (PWLC 44) "The social repercussions are sometimes quite significant." (HCP 37)
Daily functioning	Fatigue, loss of energy, less capacity to do several tasks at the same time, need to reduce day-to-day activities	"I live in slow motion" (PWLC 21), "Life has changed completely. It's complicated to do the things we used to do, (so) we don't do them anymore" (PWLC 32), "I do the bare minimum on a day-to-day basis", "I've got a high chair in my kitchen, so I can do little things otherwise I can't. I can't wait at a supermarket checkout, so there are things, lots of things I can't do anymore." (PWLC 16) "It's also going to be the everyday activities, whether it's cleaning, shopping, paying bills, administrative documents, it really affects all aspects of life." (HCP 7) "On a day-to-day basis, I have to arbitrate between taking a shower and taking out the rubbish." (PWLC 62) "When I walked, I was out of breath, when I talked, I was out of breath. Climbing stairs was complicated too." (PWLC 19)

Supplementary table 2: survey results voice-related questions

Question	Answer	Total PWLC N(%) N=201	Total HCP N(%) N=15
Did you already heard about vocal biomarkers?	No	183 (91%)	10 (66.7%)
To which extend are you convinced or not convinced that symptoms can be measured by using vocal biomarkers?	Convinced + totally convinced	77 (38%)	10 (66.7%)
To which extend do you think that this kind of app would be useful for people living with Long COVID and already engaged in a LC care?	Somewhat useful		3 (20%)
	Very useful		5 (33%)
	Useful		7 (47%)
To which extend do you think that this kind of app would be useful for people living with Long COVID and NOT engaged in a LC care?	Somewhat useful		3 (20%)
	Very useful		5 (33%)
	Useful		7 (47%)
Would you recommend the app use to your LC patients?	Yes		11 (73%)
Would you be interested in a smartphone app based on vocal biomarkers to monitor your health?	Very interested	58 (29%)	
	Interested	98 (49%)	
	Somewhat interested	30 (15%)	
	Weakly interested	9 (4.5%)	
	Not at all interested	6 (3.0%)	
Should the app be based on voice recordings only or also on questionnaires?	Voice recordings completed by regular questionnaires	140 (70%)	9 (60%)
	Only voice recordings, for ease of use	30 (15%)	6 (40%)
	Don't know	30 (15%)	0 (0%)
Acceptable frequency of recordings	As needed	56 (28%)	2 (13%)

	Once a week	67 (33%)	7 (47%)
	Twice a week	26 (13%)	3 (20%)
	Three times a week	11 (5.5%)	2 (13%)
	Everyday	28 (14%)	1 (6.7%)
Type of recordings	Vowel phonation	98 (49%)	5 (33%)
	Counting	124 (62%)	7 (47%)
	Breathing	96 (48%)	5 (33%)
	Coughing	67 (33%)	4 (27%)
	Reading a text	142 (71%)	13 (87%)
	Answering a question	112 (56%)	9 (60%)
	Free speech on health	57 (28%)	6 (40%)
	Fully free speech	29 (14%)	3 (20%)
Willingness to use a voice-based app	Yes	156 (78%)	
If no, why	Don't like his voice	2 (1.0%)	
	Fear of data leak	1 (0.5%)	
	Doubt in the fiability of this technology	4 (2.0%)	
In which context would you consider making the voice recordings?	Anywhere (work, home, outdoors)	32 (16%)	
	Alone or with my family (at home)	52 (26%)	
	Only if I'm sure I'm alone	100 (50%)	

Additional file 4

Supplementary table 3

Supplementary table 3: Expectancies-fears-barriers-facilitators. *answer modality only proposed to HCP

	Survey results	Interviews themes	Interviews subthemes	Verbatim	Inference
Expectancies	Symptom monitoring: PWLC N=143 (71%) / HCP N=12 (80%) Objective measurement of symptoms: PWLC N=112 (56%) ; HCP N=8 (53%)	Care	Symptom monitoring/identification	"I have a diary to write down my symptoms. So, I write them down sometimes, but it's true that it's not super practical." (PWLC 44) "I was invaded by hundreds of symptoms, but it took me a year and a half in fact, to put the right words on what was going wrong." "It can also monitor evolution and progress." (HCP 12) "Symptoms and their monitoring are the priority." (PWLC 62). "Having a calendar of symptoms like this can be very therapeutic,for psycho-education, to show the fluctuation of symptoms, to show and visualize the course of certain symptoms" (HCP 66) "It's interesting to have the subjective feeling, as well as something more objective. Because it's possible that, at certain times, we find a symptom very disabling, whereas objectively speaking, things may have improved considerably." (PWLC 21)	Convergence, expansion
	Evaluation of a rehabilitation program: PWLC N=96 (48%) ; HCP N=11 (73%)		Improve communication PWLC-HCP	"Something more objective than the patient's feelings, which are not necessarily heard by doctors." (PWLC 61) "It could be a benefit for the hospital to see when you had a symptom, why, how? The follow-up can be improved" (PWLC 51) "To facilitate caregiver-patient dialogue" (PWLC 10)	
	Improve adherence to treatment or rehabilitation program: HCP 8 (53%)*		Improve care and access to care	"A digital application can help to adapt and change behavior." (HCP 66) " It can also help you better manage yourself" (PWLC 62) "Do things a little more tailored to their needs" (HCP 9) "It could be interesting to make the link between symptom monitoring and care." (HCP 9) "If at the very beginning I'd had a to-do list of just which doctor I should see, how to name my symptoms, that it would have saved me 2 years in my care path." (PWLC 62) "It could have helped in the follow-up or in implementing solutions." (PWLC 4) "If they discover information in this app that can help them improve their medical and paramedical care, that's great!" (HCP 12)	Convergence, expansion
	Reduce medical visits: PWLC 62 (31%) ; HCP 7 (47%)		Reduce in-person visits	"This could allow us, for example, not to see them systematically every month" (HCP 9) "one interest I see in the application is that it could perhaps limit the need to visit doctors or other health professionals" (PWLC 32)	Convergence
	Diagnosis: PWLC 87 (43%) ; HCP 5 (33%)				Divergence, diagnosis was not a theme that emerged during interviews
	Companion tool: HCP 8 (53%)*	Daily life	Psychological support	"We're starting from such a long way off that any help, even the smallest, will make a difference." (PWLC 62) "Psychological help, I would like that. Honestly, I would like that. (...) Centralizing all this data on a specialized application would help me, relieve me, reassure me and reduce my stress." (PWLC 57) "Measuring and sending back information in relation to what we have can be very interesting and reassuring too." (PWLC 44) "To feel less alone, to feel understood, to not feel abandoned" (HCP 7) "Give a bit of encouragement" (PWLC 21) "It has the advantage of at least informing them and explaining to them that they're not mad, that these are symptoms that we come across and that we often see in Long COVID" (HCP 37). "I don't know where or how it can be set up, but here it is, psychological help." (PWLC 55)	Convergence, expansion
			Administrative support	"It's very difficult at first to understand everything that's going on at the administrative level."(PWLC 55) "A section on "Going back to work: how is it going?" (PWLC 44). "Inform them that there are existing Long Covid associations" (PWLC 51)	Expansion: this new theme emerged during individual interviews
Facilitators	Ease of use: PWLC N = 144 (72%) ; HCP N = 11 (73%)	User engagement	Adaptability/personalisation	"It has to be general, but I'd like to adapt it to my own needs." (PWLC 57) "The more customized, the better" (PWLC 19) "Everything individualized, especially for patients who have been in medical wandering for a long time, who are sometimes stigmatized, who are sometimes not believed, this idea of being interested in them personally." (PWLC 16) "Adaptability to each patient is the key word in care. For an application, I think it could be good" (HCP 7)	Convergence, expansion
			Easy to use conception	"It has to be super-intuitive" (HCP 7) "It has to be easy, it has to be simple." (PWLC 14) "I'd like something very simple." (PWLC 16) "In terms of access and interactivity, it has to be simple." (PWLC 44)	Convergence, expansion
			Voice dictation	"Registering oneself is much easier." (PWLC 19) "I do almost all my stuff orally because I'm tired of writing," (PWLC 14) "it might help some people who are less at ease with the written word." (PWLC 4)	Expansion
	Visualization of VB results: PWLC N = 123 (61%) ; HCP N = 7 (47%)		Results visualization	"Something quite visual and not too complicated, something like green-orange-red perhaps" (PWLC 4) "It could be something really appreciated by patients to have regular feedback with nice graphics." (HCP 66) "when you see a curve that goes up and down, it's worth a thousand words." (HCP 9)	Expansion
	Reliability: PWLC N = 108 (54%) ; HCP N = 12 (80%)	App content	Transparency	"It has to be very clear that it's not a substitute for medical care, but rather that it's an additional aid." (PWLC 14) "The person has to be informed and agree or disagree (about linkage with other apps)" (PWLC 23). Trust has to be established. But then, once he's got the trust, it's good for me (PWLC 55). "We have to be careful about what we offer and what we promise them and what we sell them, exactly. If the proposal is completely humble and based on the current state of knowledge, everything's fine." (HCP 37	Expansion
			Validated content	"With validated medical and management advice or recommendations" (HCP 9) "It's really a question of establishing trust and proving the reliability of this technology." (PWLC 19) "The application will then have to meet all the security criteria required for any healthcare application." (HCP 9)	Expansion
			Targeted content	"Provide links for more information." (HCP 9) "If they can be targeted, we're more inclined to pay attention to them. (...) A "search" area where we can type in the keyword" (PWLC 4)"	Expansion
	Incorrect result interpretation: PWLC N = 136 (68%) ; HCP N = 10 (67%)		Disabilities, language, vision problems	"It may exclude people who are less at ease with reading, or people who don't speak the language, or people with a disability." (PWLC 62) "Reading is always a problem because you don't know whether people are literate or not." (HCP 12). "It's not accessible to patients who don't speak the language" " you have to be careful with colors and sounds if there are alerts, it must not be blinding or irritating. (...) On my screen, in fact, the colors are reversed. The writing is always on a black background. "(PWLC 16)	Convergence

Supplementary table 3: Expectancies-fears-barriers-facilitators. *answer modality only proposed to HCP

	Survey results	Interviews themes	Interviews subthemes	Verbatim	Inference
Expectancies	Symptom monitoring: PWLC N=143 (71%) / HCP N=12 (80%)	Care	Symptom monitoring/identification	"I have a diary to write down my symptoms. So, I write them down sometimes, but it's true that it's not super practical." (PWLC 44) "I was invaded by hundreds of symptoms, but it took me a year and a half in fact, to put the right words on what was going wrong." "It can also monitor evolution and progress." (HCP 12) "Symptoms and their monitoring are the priority." (PWLC 62). "Having a calendar of symptoms like this can be very therapeutic,for psycho-education, to show the fluctuation of symptoms, to show and visualize the course of certain symptoms" (HCP 66) "It's interesting to have the subjective feeling, as well as something more objective. Because it's possible that, at certain times, we find a symptom very disabling, whereas objectively speaking, things may have improved considerably." (PWLC 21)	Convergence, expansion
	Objective measurement of symptoms: PWLC N=112 (56%) ; HCP N=8 (53%)				
	Evaluation of a rehabilitation program: PWLC N=96 (48%) ; HCP N=11 (73%)		Improve communication PWLC-HCP	"Something more objective than the patient's feelings, which are not necessarily heard by doctors." (PWLC 61) "It could be a benefit for the hospital to see when you had a symptom, why, how? The follow-up can be improved" (PWLC 51) "To facilitate caregiver-patient dialogue" (PWLC 10)	Convergence, expansion
				"A digital application can help to adapt and change behavior." (HCP 66) " It can also help you better manage yourself" (PWLC 62) "Do things a little more tailored to their needs" (HCP 9) "It could be interesting to make the link between symptom monitoring and care." (HCP 9)	
	Improve adherence to treatment or rehabilitation program: HCP 8 (53%)*		Improve care and access to care	"If at the very beginning I'd had a to-do list of just which doctor I should see, how to name my symptoms, that it would have saved me 2 years in my care path." (PWLC 62) "It could have helped in the follow-up or in implementing solutions." (PWLC 4) "If they discover information in this app that can help them improve their medical and paramedical care, that's great!" (HCP 12)	Convergence, expansion
	Reduce medical visits: PWLC 62 (31%) ; HCP 7 (47%)		Reduce in-person visits	"This could allow us, for example, not to see them systematically every month" (HCP 9) "one interest I see in the application is that it could perhaps limit the need to visit doctors or other health professionals" (PWLC 32)	Convergence
	Diagnosis: PWLC 87 (43%) ; HCP 5 (33%)				Divergence, diagnosis was not a theme that emerged during interviews
	Companion tool: HCP 8 (53%)*	Daily life	Psychological support	"We're starting from such a long way off that any help, even the smallest, will make a difference." (PWLC 62) "Psychological help, I would like that. Honestly, I would like that. (...) Centralizing all this data on a specialized application would help me, relieve me, reassure me and reduce my stress." (PWLC 57) "Measuring and sending back information in relation to what we have can be very interesting and reassuring too." (PWLC 44) "To feel less alone, to feel understood, to not feel abandoned" (HCP 7) "Give a bit of encouragement" (PWLC 21) "It has the advantage of at least informing them and explaining to them that they're not mad, that these are symptoms that we come across and that we often see in Long COVID" (HCP 37). "I don't know where or how it can be set up, but here it is, psychological help." (PWLC 55)	Convergence, expansion
			Administrative support	"It's very difficult at first to understand everything that's going on at the administrative level."(PWLC 55) "A section on "Going back to work: how is it going?" (PWLC 44). "Inform them that there are existing Long Covid associations" (PWLC 51)	Expansion: this new theme emerged during individual interviews
Facilitators	Ease of use: PWLC N = 144 (72%) ; HCP N = 11 (73%)	User engagement	Adaptability/personalisation	"It has to be general, but I'd like to adapt it to my own needs." (PWLC 57) "The more customized, the better" (PWLC 19) "Everything individualized, especially for patients who have been in medical wandering for a long time, who are sometimes stigmatized, who are sometimes not believed, this idea of being interested in them personally." (PWLC 16) "Adaptability to each patient is the key word in care. For an application, I think it could be good" (HCP 7)	Convergence, expansion
			Easy to use conception	"It has to be super-intuitive" (HCP 7) "It has to be easy, it has to be simple." (PWLC 14) "I'd like something very simple." (PWLC 16) "In terms of access and interactivity, it has to be simple." (PWLC 44)	Convergence, expansion
			Voice dictation	"Registering oneself is much easier." (PWLC 19) "I do almost all my stuff orally because I'm tired of writing." (PWLC 14) "It might help some people who are less at ease with the written word." (PWLC 4)	Expansion
	Visualization of VB results: PWLC N = 123 (61%) ; HCP N = 7 (47%)		Results visualization	"Something quite visual and not too complicated, something like green-orange-red perhaps" (PWLC 4) "It could be something really appreciated by patients to have regular feedback with nice graphics." (HCP 66) "when you see a curve that goes up and down, it's worth a thousand words." (HCP 9)	Expansion
	Reliability: PWLC N = 108 (54%) ; HCP N = 12 (80%)	App content	Transparency	"It has to be very clear that it's not a substitute for medical care, but rather that it's an additional aid." (PWLC 14) "The person has to be informed and agree or disagree (about linkage with other apps)" (PWLC 23). Trust has to be established. But then, once he's got the trust, it's good for me (PWLC 55). "We have to be careful about what we offer and what we promise them and what we sell them, exactly. If the proposal is completely humble and based on the current state of knowledge, everything's fine." (HCP 37	Expansion
			Validated content	"With validated medical and management advice or recommendations" (HCP 9) "It's really a question of establishing trust and proving the reliability of this technology." (PWLC 19) "The application will then have to meet all the security criteria required for any healthcare application." (HCP 9)	Expansion
			Targeted content	"Provide links for more information." (HCP 9) "If they can be targeted, we're more inclined to pay attention to them. (...) A "search" area where we can type in the keyword" (PWLC 4)"	Expansion
	Incorrect result interpretation: PWLC N = 136 (68%) ; HCP N = 10 (67%)		Disabilities, language, vision problems	"It may exclude people who are less at ease with reading, or people who don't speak the language, or people with a disability." (PWLC 62) "Reading is always a problem because you don't know whether people are literate or not." (HCP 12). "It's not accessible to patients who don't speak the language" " you have to be careful with colors and sounds if there are alerts, it must not be blinding or irritating. (...) On my screen, in fact, the colors are reversed. The writing is always on a black background. "(PWLC 16)	Convergence

Concerns	Symptom intensity: PWLC N = 75 (37%) ; HCP N = 1 (6,7%)	Accessibility	Lack of energy - time needed	"I only have a certain amount of time I can concentrate on each day. So if I've already given my all by entering my symptoms in the app, I might not have the strength to do more." (PWLC 62) "If it takes me more than 2 minutes a day, I won't do it." (PWLC 14) "I don't see any constraints. The only thing is, does it take a lot of time?" (PWLC 6)	Convergence
			Too much information	"The risk is that there's too much information, that people get overwhelmed and that it becomes too complex." (PWLC 61) "If you give advice for all the symptoms that exist, it can be too much information." (PWLC 19) "If we start to get into the explanations of Long Covid, then we're getting into something really huge." (PWLC 16)	
	Data protection: PWLC N = 73 (36%) ; HCP N = 8 (53%)	User confidence	Confidentiality issues	"I'm always a bit scuspicious, I prefer to keep the collection of personal and health data to the strict minimum." (PWLC 62) "We're perhaps a little afraid, a little suspicious, so I think it's really a question of developing trust, of saying that this is really for the user, that there's something serious behind it, ... what usefulness it can bring us, and also to show that RGPD is respected." (PWLC 19) "It has to be properly protected, you need very strong data protection." (PWLC 14) "everyone around you hears what you're saying, and I think that can be a bit of a barrier for some people, who will wait until they're alone, until they're well isolated, to use the app, because they have to use their voice." (HCP 7)	Convergence
	Age: PWLC N = 49 (24%) ; HCP N = 4 (27%)		Vocal biomarkers are a new technology	"It remains to be seen whether it really works" (PWLC 51) "Since I don't know anything about it, I'm wondering about its reliability." (PWLC 16)	Convergence (age was not an essential barrier)

Supplementary table 4: App characteristics

	Survey question	Answers	PWLC survey results	HCP survey results	Verbatim	Inference
Engagement	Personalization options	Type of result display	129 (64%)			
		Informative content	124 (62%)			
		Notifications	93 (46%)			
		Sounds	68 (34%)		"No, I don't think it's important for me to personalize it." (PWLC 61) "If you want to see the other things, you always have to change the parameters" (PWLC 21) "I'd say everyone could decide for themselves what to use and what to leave out." (PWLC 6)	Convergence
	Interactions with other app users	Yes	68 (34%)		"You need a community manager" (HCP 37) "as soon as there's an exchange platform, there has to be moderation. And there are already a lot of virtual places for exchange. I think at some point, we're getting a bit bored of that. We need more real, physical meetings." (PWLC 16) "You can come across someone who is toxic" (HCP 12)	Divergence
	Possibility to give feedback on the app	Yes	161 (80%)			Divergence
	Is fun aspect important for you? Yes	Yes	85 (42%)			Divergence
	Reminders - notifications	Yes	140 (70%)		"it has to be programmable, not 10 a day." (PWLC 16) "Push-ups that come, if they come too often you're used to just releasing them " (HCP 66) "There have to be reminders adapted to our need" (PWLC 14) "Small encouraging messages when you log on or record your voice" (PWLC 61) "We need to find something that helps us to remember, but at the same time doesn't generate pressure or stress." (PWLC 16) "A kind of widget on the cell phone, so you can easily click on it and receive notifications" (PWLC 19) "Notifications in the form of reminders could be in." (HCP 7)	Convergence- Expansion
	Gamification	Yes	55 (27%)		"The gaming aspect could keep a lot of people motivated." (PWLC 62)	Convergence
Functionalities	Alert system	Integration of alert system? - Yes	127 (63%)	11 (73%)	"Maybe just indicate that this week is not so good and that you need to check with your doctor." (PWLC 66) "Adding an alert if there's a deterioration can also scare people a little, make them a little anxious" (PWLC 61) "To alert without worrying. But yes, we need to be told, "Get some rest."" (PWLC 44) "In a scenario where caregivers are also involved in the app, I'd tend to say that it's better to send the alert to the caregiver." (PWLC 10) " Maybe you just need to put in a piece of advice like "Have your fatigue level assessed, so that your GP can be aware of it when you next see him or her, or call him or her if you feel it's necessary"." (HCP 9) "I think the simplest way to make it accessible to everyone is to generate a PDF report of these alerts, of what the application has identified, so that it can be easily transmitted to any healthcare professional." (HCP 7) "You'll have to do it very carefully, so as not to frighten the person"	Expansion
		Alert recipient:				
		App users	114 (57%)	7 (47%)		
		Designated person	29 (14%)	2 (13%)		
		Physician	60 (30%)	7 (47%)		
		Other caregivers	26 (13%)	7 (47%)		
	Rehabilitation	Possibility of integration of an external rehabilitation program in the calendar	143 (71%)	13 (87%)	"to be careful not to push yourself too hard, or restrict yourself too much. And that's not easy. So, with the right explanations about post-exertion malaise, yes." (PWLC 10) "I had included the exercises in my proposals. But But since then, because I didn't actually think that what I was doing was harmful to my health. This contributed to me having another PEM crash. " (PWLC 44) "Exercises that people can do at home on their own, yes and no.." (PWLC 16) "I don't see any problem in integrating this, even if it's only beneficial for 5% of patients...Mindfulness exercises, neurocognitive training exercises, relaxation exercises, breathing exercises." (HCP 66) "What I think would be very useful would be breathing exercises. With explanations, of course...physical rehabilitation, cardio or endurance exercises" (PWLC 21) "Small cognitive exercises" (PWLC62)	Expansion, divergence
		Rehabilitation exercices proposed in the app	155 (77%)	12 (80%)		
		App should propose exercices depending on the symptoms reported by the user	117 (58%)	8 (53%)		
	Symptom monitoring	Monitoring several symptoms	173 (86%)	14 (93%)	"it would be interesting to have the questionnaires as well. It would make it easier to correlate them." (PWLC 12) "To refine the voice biomarker, you can associate it with a scale" (HCP 9) "Everything has to be prepared so that you can click and set the intensity. " (PWLC 14) "To monitor symptoms on a daily basis." (PWLC 16)	Confirmation, expansion
		Monitoring should combine VB and questionnaires	140 (70%)	9 (60%)		
		Symptoms to monitor:				
		Cardiorespiratory sypmtoms	**119 (59%)**	**8 (80%)**	"Cognitive problems, with fatigue" (PWLC 44) "Fatigue, which is the first and most important symptom." (PWLC 62) "it seems that neurocognitive problems are much more chronic than fatigue" (HCP 66) "Cognitive problems are a huge handicap for me" (PWLC 57) "Cognitive problems, often accompanied by fatigue" (HCP 37) "Fatigue, neurocognitive problems, shortness of breath or palpitations." (HCP 9) "Brain fog, memory loss, difficulty finding words." (PWLC 55) "Respiratory discomfort that recurs regularly." (PWLC 61) "Cardio-respiratory disorders in the broadest sense, i.e. shortness of breath, loss of endurance" (HCP 7)	Confirmation (confirmation for fatigue, cardio and neuro symptoms, anxiety not that much mentioned)
		Fatigue	**178 (88%)**	**11 (73%)**		
		Symptomatic/asymptomatic status	38 (19%)	1 (6.7%)		
		Loss taste or smell	36 (18%)	4 (27%)		
		Anxiety	**91 (45%)**	**7 (47%)**		
		Gastrointestinal symptoms	70 (35%)	2 (13%)		
		Neurological symptoms	**158 (78%)**	6 (40%)		
	Interaction with HCPs through app	Possibility of sharing data/results with HCP or family? - Yes	153 (76%)		"A report showing how my fatigue has evolved over the last month...and so that the doctor has a visual record of what has happened" (PWLC 32) "I'm wondering about feasibility." (PWLC 16) "Something that seems interesting to me is that the doctor who followed me for the COVID, whom I would have designated, (..) could also receive, if he accepts, to receive a small alert if something goes wrong" (PWLC 23) "the risk is that healthcare professionals may not necessarily have the capacity to manage" (HCP 9) "The easiest way to do this is to generate a PDF report of what the application has recorded, so that it can be easily transmitted to any professional. " (HCP 7)	Confirmation, expansion
		Emails to HCP	68 (34%)			
	Diary	Calendar with dates of recordings	183 (91%)		"Yes, to be able to enter their treatment. And if someone's going to do meditation or breathing exercises, for example, they can enter it into the application. " (PWLC 19) "I'd love to have an application where I could put my symptoms, my treatments, all my Long Covid follow-up in one place, so that when I see a doctor, I can show him everything." (PLWC 14) "The diary, yes. But that's something that requires a lot of investment on the part of patients." (HCP 66) "It could be interesting, yes, it could be interesting" (PWLC 23) "The diary could be accessed at any time, like a notebook." (PWLC 44)	Confirmation - Expansion
		Integration of informative contents? - Yes	174 (87%)	13 (87%)	"Having the list of other people's symptoms would have helped me identify mine sooner." (PWLC 62) "To be able to go and get additional information yes, that could be good." (PWLC 61) "Tips on symptoms" (PWLC 44) "Provide links to patient associations that can offer support groups, answers to questions,	
		Voice recordings explanations	103 (51%)	7 (47%)		

Information content	Information	Information on vocal biomarkers	147 (73%)	11 (73%)	etc." (PWLC 14) "you could list a number of sites where people can connect, but I don't think the application should become a Long Covid encyclopedia. " (HCP 9) "There are quite a number of things available that are pretty well done. " (HCP 7) "There's a whole social network to know about, and that's disability care." (HCP 37) "How, when and why to apply for recognition as a disabled worker" (PWLC 10) "Personally, I don't need anyone to tell me what I should eat and what I should do." (PWLC 57) "Suggest, for example, different types of activities, relaxation strategies." (PWLC 19)	Convergence, expansion
		Information on Long COVID	159 (79%)	12 (80%)		
		Rehabilitation exercices	134 (67%)	8 (53%)		
		Advices	167 (83%)	15 (100%)		
	Advices	Dietary advices	95 (47%)	7 (47%)	"Pacing and restarting physical activity will apply to the vast majority of people." (HCP 7) "Advice on matters that concern me" (PWLC 4) "(Advices) on diet, sleep and physical activity" (HCP 12) "Advices linked to pacing because we know nowadays that it's almost the only thing that really works." (PWLC 44) "What we can eat, what we can't eat, and what we can do to improve our diet to get a little better."	Convergence, expansion
		Physical activity advices	138 (69%)	14 (93%)		
		Sleep advices	115 (57%)	10 (67%)		
		Psychological advices	123 (61%)	12 (80%)		
Aesthetic		Aesthetic	2 (1.0%)		"It's not a decisive field." (HCP 66) "I like colors. So no matter what the colors are." (PWLC 55) "I don't like flashy colors" (PWLC 43) "About the colors, no, I must say I don't have any preferences." (PWLC 61)	Convergence
		Visualization of VB results	123 (61%)		"Something graphic to see how it improves or doesn't improve" (PWLC 21) "In terms of visualization, I think it would be really cool to be able to track symptoms over time." (PWLC 14) "a brief summary of the days when I was very bad, and the days when I was good." (PWLC 16) "a curve to track changes over time, so we can see whether or not there has been any improvement." (PWLC 61)	Convergence, expansion
		Scores represented as colors corresponding to broad categories (e.g. green for a normal result, orange for an intermediate result and red for an abnormal result).	103 (51.2%)			
Generalities	App usage	Maximum app usage time acceptable per week < 1h	148 (74%)		"Once a week, I would have a sort of weekly follow-up" (PWLC 57) "It depends on where you are in the disease. I think that someone who's just getting into it (the Long COVID), and depending on his or her symptomatology, someone who's very affected will probably need to go every day, while someone who's less affected may need to go twice a week." (PWLC 10) "I'd say one recording a week would be a good frequency." (PWLC 61) "if it takes me more than 2 minutes a day, I won't do it" (PWLC 14)	Expansion
	App reimbursement	App should be reimbursed	166 (83%)	11 (73%)		
	Reason	Quality of life improvement	122 (61%)	4 (27%)		Divergence, theme did not emerged during interviews
		Early care	80 (40%)	7 (47%)		
		Cost reduction of care	92 (46%)	10 (67%)		
	Accessibility				"For ease of use I'd prefer to use the computer because I have a computer at home. Then, if you're working, staying in a hotel every night and don't have a computer, you'd rather use a cell phone. " (PWLC 23) " I'd prefer on the computer because my smartphone has a lot of applications for which it's no longer compatible." (PWLC 43) "The most user-friendly and easiest would be smartphone." (PWLC 4)	Expansion

Résumé - Quatre ans après le début de la pandémie à COVID-19, il est estimé que près de 100 millions de personnes dans le monde souffrent de conséquences à long terme, regroupées sous le terme de COVID Long. Le développement de solutions de suivi à distance a été recommandé par le *National Institute for Health Excellence* (NICE) et l'utilisation de la voix et des biomarqueurs vocaux s'avère particulièrement pertinente dans ce cadre. Les travaux de cette thèse s'articulent en deux axes principaux pour d'une part améliorer les connaissances sur le COVID sur base des données de la cohorte Predi-COVID au Luxembourg et d'autre part développer une solution digitale de santé innovante intégrant la voix pour le suivi de la santé des personnes touchées par le COVID Long. Nous avons montré que le COVID Long n'est pas une entité unique mais est constitué de sous-phénotypes de sévérité variable et que les symptômes avaient tendance à se présenter par groupes. Des trajectoires d'évolution des symptômes depuis l'infection initiale jusqu'à 2 ans après ont été modélisées, montrant que 30% des participants de l'étude Predi-COVID voyaient leurs symptômes persister, avec un fort impact sur leur qualité de vie à 2 ans et une prédominance de la fatigue et de l'anxiété. Ces données soulignent le besoin d'outils innovants pour permettre une prise en charge personnalisée. L'étude UpcomingVoice a permis de décrire les besoins et les attentes des personnes touchées par le COVID Long en termes de solution digitale de suivi de symptômes intégrant la voix, de définir les spécifications attendues et de développer une première version de l'application Long COVID Companion. Cette application, créée pour et avec des personnes touchées par le COVID Long apporte une réponse concrète à leurs besoins, en permettant un suivi au quotidien de leur santé et de leurs symptômes, une visualisation graphique pouvant servir de support de discussion avec les professionnels de santé et en proposant des ressources validées sur le COVID Long. Long COVID Companion est la première application de ce type accessible à tous, gratuitement et en 3 langues. Ce projet de thèse a ainsi un impact direct pour les personnes avec un COVID Long en proposant un outil de support au quotidien. L'application Long COVID Companion représente également une cohorte virtuelle de personnes avec un COVID Long, ce qui en fait une ressource supplémentaire unique pour la recherche sur le COVID Long et les biomarqueurs vocaux. Des collaborations ont été mises en place avec l'association #ApresJ20 Covid long France ainsi qu'avec les réseaux de prise en charge spécialisés COVID Long au Luxembourg et en France et contribueront à améliorer les connaissances sur le COVID Long. Enfin des études complémentaires pour évaluer le bénéfice clinique éventuel de Long COVID Companion et des améliorations pour l'obtention de la certification dispositif médical sont envisagées pour que cet outil soit pleinement intégré dans la prise en charge des personnes vivant avec un COVID Long.

Mots-clés : COVID long ; suivi des symptômes ; santé digitale ; voix ; co-design

Abstract - Four years after the start of the COVID-19 pandemic, it is estimated that almost 100 million people worldwide are suffering from long-term consequences, defined under the umbrella term of Long Covid. The development of remote monitoring solutions has been recommended by the National Institute for Health Excellence (NICE), and the use of voice and vocal biomarkers is particularly relevant in this context. This thesis project consists of two axes. The first one aims to improve knowledge on Long COVID based on data from the Predi-COVID cohort in Luxembourg, and on the other, to develop an innovative voice-based digital health solution to monitor the health of people affected by Long Covid. We have shown that Long Covid is not a single entity, but appears to be made up of sub-phenotypes of varying severity, and that symptoms tend to occur in clusters. Trajectories of symptom evolution from initial infection to 2 years later were modeled, showing that 30% of Predi-COVID participants experienced high levels of persisting symptoms, with a strong impact on their quality of life at 2 years, and a predominance of fatigue and anxiety. These data underline the need for innovative tools to enable personalized management. The UpcomingVoice study described the needs and expectations of people affected by Long Covid in terms of a digital, voice-based symptom-tracking solution, defined the expected specifications and allowed the development of a first version of the Long Covid Companion application. This application, created for and with people affected by Long COVID, provides a concrete response to their needs, enabling them to monitor their health and symptoms on a daily basis, providing a graphic visualization that can be used as a basis for discussion with healthcare professionals, and offering validated resources on Long COVID. Long COVID Companion is the first application of its kind accessible to all, free of charge and in 3 languages. This thesis project thus has a direct impact on people with Long COVID, providing them with a tool to support them in their daily lives. The Long COVID Companion application also represents a virtual cohort of people with Long Covid, making it an additional unique resource for research on Long COVID and vocal biomarkers. Collaborations have been set up with the #ApresJ20 Covid long France association, as well as with specialized COVID Long care networks in Luxembourg and France, and will contribute to improving knowledge on Long COVID. Finally, further studies to assess the potential clinical benefit of Long COVID Companion and improvements to obtain medical device certification are envisaged to ensure that this tool is fully integrated into the care of people living with Long COVID.

Key words : Long COVID ; symptom monitoring ; digital health ; voice ; co-design