

Cohort Profile

Cohort Profile Update: Lifelines, a three-generation cohort study and biobank

Anna Sijtsma,^{1*} Johanna Rienks ,¹ Pim van der Harst,^{2,3}
Gerjan Navis,⁴ Judith GM Rosmalen⁵ and Aafje Dotinga¹

¹Lifelines Cohort Study, Roden, The Netherlands, ²Department of Cardiology, University Medical Centre Utrecht, Utrecht, The Netherlands, ³Department of Cardiology, University Medical Center Groningen, Groningen, The Netherlands, ⁴Department of Nephrology, University Medical Center Groningen, Groningen, The Netherlands and ⁵Departments of Psychiatry and Internal Medicine, University Medical Center Groningen, Groningen, The Netherlands

*Corresponding author. Lifelines Cohort Study, PO Box 50, 9300 AB Roden, The Netherlands. E-mail a.sijtsma@lifelines.nl

Received 20 April 2021; Editorial decision 8 November 2021; Accepted 25 November 2021

Key Features

- The Lifelines Cohort (Lifelines) is a large population-based multi-disciplinary dynamic cohort and biobank which facilitates FAIR (findable, accessible, interoperable and reusable) research on healthy ageing. Over 167 000 participants were included at baseline, with the aim to follow up for at least 30 years.
- To questionnaires, new dimensions in work, health status, lifestyle, stress and psychosocial parameters were added. New collected samples included hair, feces, cord blood, placental tissue and breast milk. New measurements for cognition, eye function and vitamin status were added. Brain MRI (magnetic resonance imaging) and lung CT (computed tomography) were performed. Linkage with noise and air pollution exposure and various other registries is made possible. Finally, the number of participants with GWAS (genome-wide association studies) data is largely extended.
- More than 70% participated in both the baseline and the second assessment.
- Lifelines is open to researchers worldwide and provides easy access to data and samples [www.lifelines.nl/researcher].

The original cohort

The Lifelines Cohort (Lifelines) is a large population-based multidisciplinary dynamic cohort and biobank designed to unravel the determinants of healthy ageing. Lifelines continuously generates deep pheno- and genotypic data to enable research on the development of chronic and rare diseases and their risk factors with the ultimate goal to contribute to healthy ageing, public health and precision

medicine. Between 2006 and 2013, 10% (around 167 000) of the northern population of The Netherlands aged 0 to 93 years was enrolled using a three-generation recruitment design, as described previously.¹ Data are collected through questionnaires (all ages) and measurements (age 8 years and older) on physical and mental health, lifestyle and the exposome. Furthermore, the Lifelines biobank has samples (blood, plasma, serum, hair, urine and feces)

stored in a freezer warehouse (from age 8 years and older). Lifelines allows analyses at multiple levels, ranging from genetic and molecular domains through clinical, psychosocial and environmental determinants of health and disease, which allows for analyses within specific domains and also inter- and multidisciplinary analyses. All participants are invited for follow-up assessments every 5 years including a comprehensive questionnaire, and will be followed for at least 30 years (Figure 1). In between assessments, follow-up questionnaires are completed approximately once every 1.5–2.5 years.

What is the reason for the new data collection?

The protocol of the study waves is largely fixed to enable longitudinal analyses into changes in health over the life course. However, there is flexibility in additions to the protocol to adapt to societal developments of public health importance, which resulted in new data on, for example, vaccinations, prevention programmes and COVID-19.

What will be the new areas of research?

Lifelines additionally started collecting hair and fecal samples to research the influence of cortisol and the microbiome, respectively, on healthy ageing. Furthermore Cogstate,² a test to monitor cognitive change, was introduced and a jump height measurement was added to study physical strength (Table 1). Furthermore, new topics were added to the questionnaires over time; more details can be found in Table 2. Finally, the number of participants with available genome-wide data increased from 15 000 to 51 000 within the UMCG (University Medical Center Groningen) Genetics Lifelines Initiative (UGLI), and we are under way to increase this number to 110 000 participants.

Beside the regular study waves, Lifelines allows researchers to collect additional in-depth data in a

subcohort of) the Lifelines population. A complete overview of these embedded data collection projects can be found on the Lifelines wiki page [https://wiki-lifelines.web.rug.nl/doku.php?id=additional_assessments]. Five examples are highlighted as follows.

- i. Lifelines NEXT, a birth cohort; 1500 pregnant women; 2016—still running; the origin of chronic diseases at a very early stage of life can be studied. Samples and information (from mother, father and baby) are collected.³
- ii. VESPER: diversity in sperm quality; 2019–21; 200 men; analyses on sperm and data from questionnaires are collected; generates control values in fertility research, for example on cancer-related infertility.
- iii. Three subcohorts focused on gastrointestinal health ('DArm Gezondheid'): Lifelines DEEP (DAG1), DAG2 and DAG3; 1500, 100 and 9500 participants, respectively; 2012–16; gaining insight into the relationship between microbiome, genetic and metabolic variation and phenotype variation. Samples and information from questionnaires were collected. Genome-wide transcriptomics and methylation data were generated, metabolites and biomarkers have been measured and the gut microbiome has been assessed.⁴
- iv. ImaLife; 12 000 participants; 2017—still running; low-dose CT; very early indicators of chronic obstructive lung diseases, coronary artery diseases and lung cancer can be studied.⁵
- v. COVID-19 project; 77 000 participants with at least one questionnaire; 2020–21; a series of at least 25 questionnaires; risk factors of COVID-19 and psychological and societal impacts can be studied.⁶

Record linkage

Air quality,⁷ noise registries⁷ and distance to local sports facilities and retail facilities (LISA) are examples of available environmental data linked to participants' residential geocodes. Lifelines is also a member of the Geoscience and

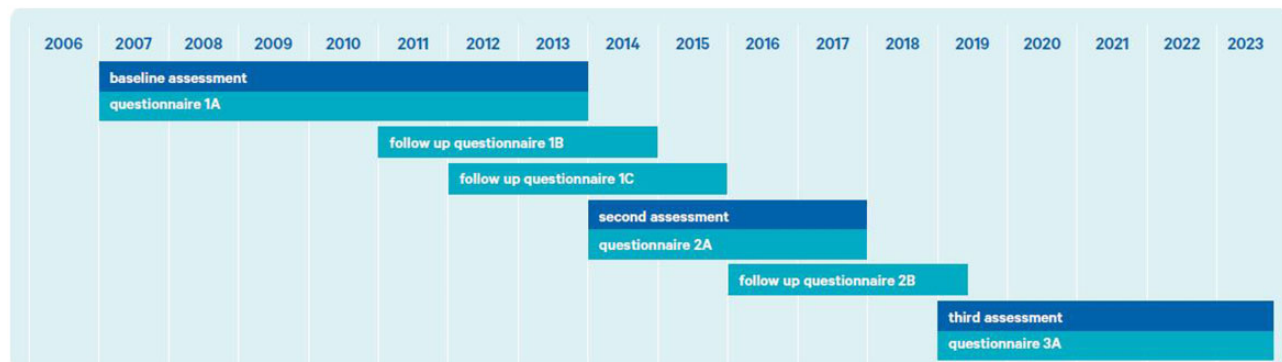


Figure 1 Timeline of Lifelines data collections

Table 1 Number of participants with specific examination data from general Lifelines visits

Year	Baseline		2nd assessment		3rd assessment
	2006–13		2014–17		2019–23
Age	8–17	18+	8–17	18+	
Total <i>n</i> (visit)	9500	152 500	6500	112 000	Ongoing
Anthropometry ^a	9500	152 500	6500	112 000	Ongoing
Blood pressure	9500	152 500	6500	112 000	Ongoing
Lung function	9500	113 000	–	32 000	Ongoing
ECG	1500	152 500	–	112 000	Ongoing
RFFT	–	87 500	–	–	Ongoing
MINI (depression/anxiety)	–	149 000	–	91 500	Ongoing
MMSE	–	13 000 ^b	–	–	Ongoing
AGE reader	–	83 000	–	–	–
Cogstate	–	–	2000 ^c	66 500	Ongoing
Jump height	–	–	5500	26 500	–

Numbers are rounded down to the nearest 500.

AGE, advanced glycation end products; ECG, electrocardiogram; MINI, mini international neuropsychiatric interview (anxiety and depression); MMSE, mini mental state examination; RFFT, Ruff figural fluency test.

^aHeight, weight, waist and hip circumference.

^bAge 65+.

^cAge 13–17 years.

health cohort consortium (GECCO).⁸ Other registries that were linked at an individual level, are: CBS (Statistics Netherlands) with cause of death and sociodemographic factors; IADB (Dutch prescription medication database); PALGA (Dutch pathology/anatomy archive); PERINED (Dutch perinatal registry); and NCR (Netherlands Cancer Registry). Linkage with general practitioners and COVID-19 PCR testing are in preparation.

Who is in the cohort?

At baseline 167 729 participants were enrolled into the cohort, aged 0 to 93 years, with a mean age of 41 years. During follow-up, mean age of the population increased to 49 years at the start of the third research round. Lifelines has a stable group of active participants: approximately 71% took part in both the baseline and the second assessment (Figure 2). Over a decade later, 120 000 participants are expected to participate in the third wave for study centre visits and completing questionnaires.

In between waves (1A and 2A in Figure 2), participants received follow-up questionnaires. The response rates were 84%, 68% and 58% for the follow-up questionnaires 1B, 1C and 2B, respectively (Figure 2).

Finally, participants were invited to participate in additional studies initiated by external researchers. Response rates on additional studies varied between 10% and 50%, depending on the topic.

Table 3 shows the baseline characteristics of participants. In total, 8900 participants (5.9%) withdrew from participating and 2500 (1.7%) died. On average, this

group of withdrawn participants were older, more often female and overweight, had more often a low educational level, no paid job and a medium-to-bad perceived health and were more likely to smoke.

Reasons for withdrawal were diverse: participating is too time consuming, lost interest, lost trust in big data/negative media attention, moved outside the region or entered regular health care programme based on health-related complaints. Because Lifelines aims to retain as many participants as possible, the following options for withdrawal are presented to the participants (percentages of total Lifelines population): (i) no invitations for study centre visits, only for questionnaires (1.8%); (ii) no invitations for questionnaires, only for visits (0.8%); (iii) withdrawal, with permission to continue using previously collected data and biomaterial (5.7%); (iv) complete withdrawal, all collected materials will be destroyed (0.1%).

What has been measured?

Data and samples

Data and samples from the general assessments

Measurements taken during the general study visits are shown in Table 1. In addition to the measurements, participants collected urine, fasting blood samples were taken and questionnaires were filled out. Immediately after blood draw, one of the blood and one of the urine samples were sent to the laboratory for analysis of a selected list of parameters. Parameters measured from EDTA Sysmex and NaF (sodium fluoride) tubes were

Table 2 Overview of the content of the questionnaires for adults

Topic	Original dimensions	New dimensions
General information		
Demographics	Marital status, household composition, nationality	Ethnicity, twins, parental divorce, language
Family composition	Date of birth, date of death father and mother, children	
Work	Employment, function, income, absences	Postal code, work history, work functioning, work environment, informal care, voluntary work, shift work
Education	Highest level of education	
Health		
Health status	History and prevalence of diseases, disabilities and symptoms	Allergies, hearing, joint problems, bodily pain, fatigue, gastrointestinal conditions, skin conditions, oral health, ocular health, family health, acne, attention-deficit hyperactivity disorder, COVID-19, autistic traits, depression, psychosis
Medication use	Current prescription, doses	
Health care use	Contact with health care professionals, receiving informal care	Functionality quantified self-devices
Questions for females	Number of pregnancies, age at menopause and age at menarche	
Birth and development	Birthweight, birth defects, breastfeeding	Early life factors
Body weight	Highest and lowest body weight in past 5 years, unwanted weight loss	
Lifestyle and environment		
Physical activity	Physical activity	
Nutrition	Nutrition intake, alcohol use	Alcohol dependence
Smoking	Current and past active and passive smoking behaviour	Nicotine dependence
Daytime	Sleeping, activities, volunteer work	
Physical environment	Type of floor covering at home, domestic animals	Perceived environment, local earthquakes
Sedentary behaviour		Sedentary behaviour
Participation in prevention programmes		Participation in vaccinations, (combined) personal lifestyle programmes, national health screenings
Screen time/internet use		Internet use, blue light
Psychosocial parameters		
Quality of life, well-being	Quality of life, emotional affect	Life satisfaction
Personality	Neuroticism, extroversion, openness, personality, anxiety	
Stress	Threatening experiences, long-term difficulties	Childhood trauma
Social support, independence	Social well-being, frailty, health care needs	Self-management ability, negative social interactions, independence, social patterns
Health anxiety		Health anxiety

comparable between all three assessments.¹ From heparin gel tubes: sodium, potassium, urea (only third assessment), creatinine, eGFR (estimated glomerular filtration rate), triglyceride, cholesterol, HDL (high-density lipoprotein)-cholesterol, LDL (low-density lipoprotein)-cholesterol were repeated and chloride was additionally measured compared with baseline assessment. From urine

only creatinine from 24-h urine was repeated compared with baseline assessment.

Plasma, serum, buffy coat and lymphocytes were stored. From adults, 24-h urine and morning urine were stored; from children (8–17 years), timed overnight urine was stored. During the second assessment, in addition to blood and urine, also scalp hair and fecal samples were collected

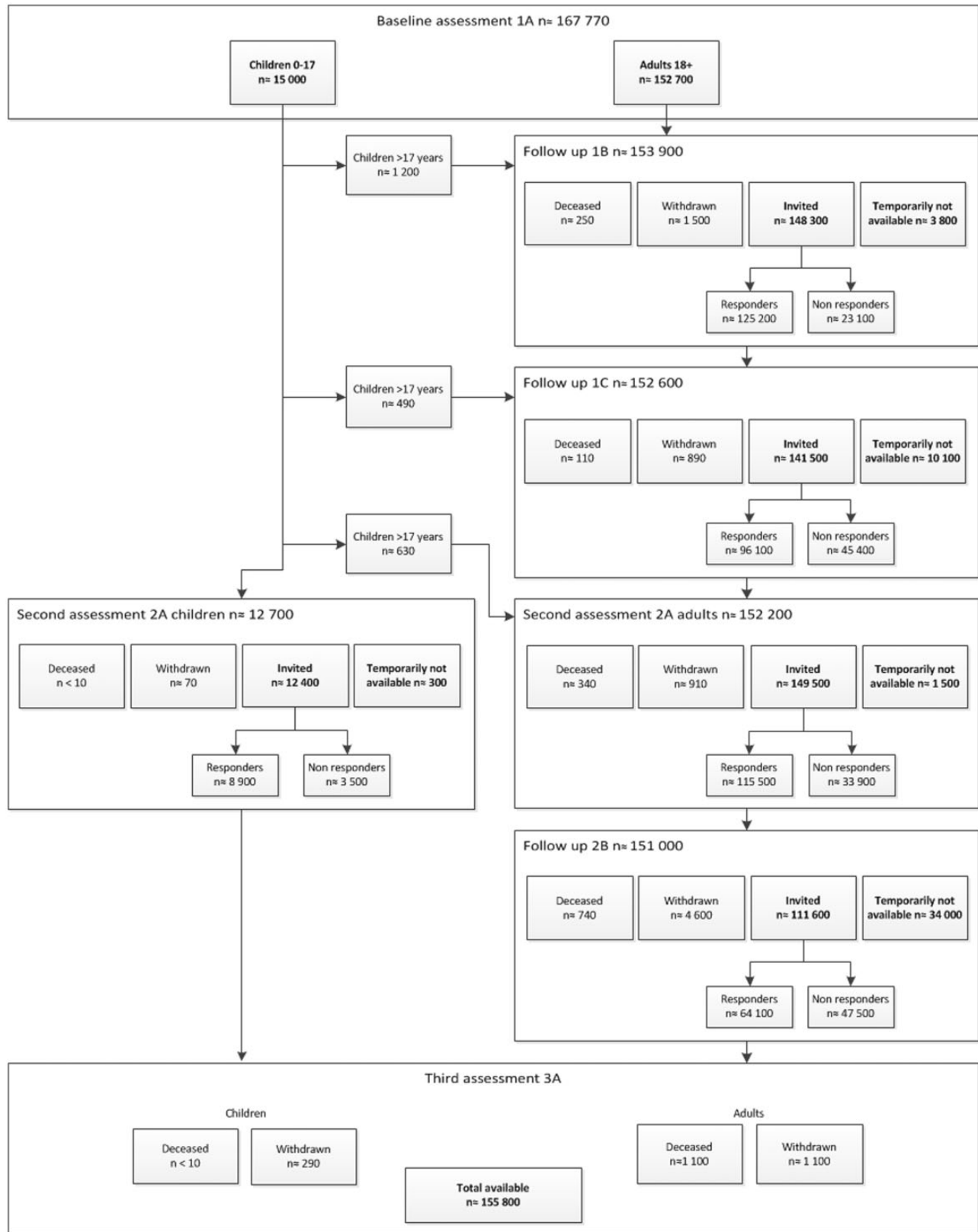


Figure 2 Flowchart of number of participants until the start of the third assessment

Table 3 Baseline characteristics of adult participants according to different participation categories

	All included at baseline	Active ^a	Withdrawn ^a	Deceased ^a
<i>n</i> ^b	152 700	141 200	8900	2500
Age (mean ± SD)	44.6 ± 13.1	44.1 ± 12.8	47.9 ± 14.0	61.6 ± 13.8
% Male	41.5	41.4	39.2	55.9
% Low educational level	16.6	15.7	27.2	32.0
% No paid job ^c	10.7	10.2	16.6	17.6
% Medium/bad perceived health	9.7	9.3	14.0	18.1
% Smoking	21.4	21.3	23.4	25.1
% Overweight (BMI ≥25)	55.3	54.6	61.0	68.6

SD, standard deviation; BMI, body mass index.

^aParticipation status at the end of 2020.

^bRounded to the nearest hundred.

^cExcluding students and retirement.

in a subcohort of the participants. Stored samples are available for researchers worldwide.

Data and samples from subcohorts

Next to the general Lifelines measurements, in subcohorts of Lifelines additional data and samples were collected or analysed. Samples of cord blood, placental tissue and breast milk were collected and stored (NEXT).³ And samples from nose (nasal brush), breath⁴ and semen (VESPER) were analysed. Furthermore plasma testosterone (VESPER), vascular function (ImaLife⁵), folic acid and the vitamins B6, B12, D, A, E and K (MINUTHE⁹) were or are going to be determined. Eye measurements including refraction, intraocular pressure, visual field, structural layers of the eye, fundus photograph and eye dryness are being collected (EyeLife). MRI (magnetic resonance imaging) of the heart and of the brain and a CT of the lungs were or will be available (ImaLife). Finally measurement of social interaction by smartphone is planned.

Questionnaire data

The questionnaires completed during the regular study waves generated around 8000 variables and cover a broad range of topics (Table 2). For detailed information about the available data, please visit our Lifelines Wiki [http://wiki-lifelines.web.rug.nl] or catalogue [https://data-catalogue.lifelines.nl].

Genomic data

Genomic data have been collected in subcohorts of the Lifelines cohort population. Whole genome sequencing is available for 165 participants. The Infinium Global Screening Array[®] (GSA) MultiEthnic Disease Version was used for SNP (single nucleotide polymorphism) genotyping of 38 000 participants. In combination with 15 400

samples already genotyped on the Cytochip GWAS (genome-wide association study), Illumina CytoSNP-12v2), a single quality controlled GWAS dataset of $n \approx 50\,000$ subjects is available. In the near future the availability of GWAS data of all participants is planned.

What has it found? Key findings and publications

Lifelines data are used in diverse research disciplines, from social sciences to biomedical sciences, and have led to a considerable scientific contribution. To date, a total of about 600 proposals have been approved: over 500 papers are published; and more than 125 researchers currently work with Lifelines data from all over the world. A complete overview of the publications can be found at the Lifelines website [www.lifelines.nl/researcher/publications].

Some highlights

- Almost 25% of the participants who had electrocardiographic evidence of a myocardial infarction (MI) where unaware of it (silent MI).¹⁰ These participants do not have the increased risk of anxiety disorders that was found in participants with a recognized MI.¹¹
- Food patterns were strongly regionally defined.¹²
- A diet with relative high percentage of proteins, and high part plant-based proteins compared with animal-based protein, was associated with a lower HbA1 score and a better kidney function.¹³
- A combination of 16 volatile organic compounds from breath samples can accurately distinguish irritable bowel syndrome patients from healthy controls.¹⁴
- The gut microbiome is related to the risk of cardiovascular diseases, independent of dietary patterns or inflammatory markers.¹⁵

- Occupational exposure to pesticides is genome-wide associated with differential DNA methylation.¹⁶
- Use of corticosteroids was associated with metabolic syndrome,¹⁷ reduced executive cognitive functioning and a higher likelihood of mood and anxiety disorders.¹⁸
- Family structures have no influence on the effects of body mass index on physical and mental quality of life.¹⁹
- It has been claimed that functional somatic syndromes, like irritable bowel syndrome, chronic fatigue syndrome and fibromyalgia, share a common aetiology, but predictors of these new-onset syndromes were more often syndrome-specific.²⁰
- Low social economic status is a risk factor for vitamin deficiencies (B6, B12 and folic acid), partly explained by diet quality (Lifelines-MINUTHE study⁹).

What are the main strengths and weaknesses?

Strengths

The main strengths of the Lifelines cohort and biobank are its longitudinal study design, with repeated measures on a wide range of multidisciplinary topics along with the collection of a large variety of biological samples. Furthermore, the three-generation design (increasing to the fourth generation in the additional birth cohort Lifelines NEXT) is a unique feature providing excellent opportunities for genetic research and the influence of family culture and social environment on health. Another strength of the Lifelines cohort is that it is being conducted in a large and stable cohort in the sense that relatively few people move out of the area. Moreover, the response rate of participants to re-invitation and questionnaires is promising for future research rounds. Finally, several opportunities exist to obtain more detailed research data, by the collection of additional data in (sub)cohorts or by data linkage.

Weaknesses

Next to the strengths, also some weaknesses are present in Lifelines. Almost 30% of the population did not participate in the second research round and attrition was selective. For the third assessment, a clear goal is set to welcome at least 120 000 participants to the study centre. The study visits and completed questionnaires will be closely monitored to retain as many participants as possible in the cohort. In case of an unexpected drop in number, timely appropriate action will be taken, dependent on the reasons for this drop. Second, inherent in the population composition of the three northern provinces in The Netherlands, relatively few ethnic minorities participate in Lifelines.

Can I get hold of the data? Where can I find out more?

Lifelines provides researchers worldwide with easy and fast access to FAIR (findable, accessible, interoperable and reusable) data and samples for research on healthy ageing. A submitted application [www.lifelines.nl/researcher] will be reviewed within 2 weeks. After signing the offer and the DMTA (Data and/or Material Transfer Agreement), in general, data will be available within 2 weeks and samples within 6–12 weeks. For detailed information about the available data please take a visit to our Lifelines Wiki [https://wiki-lifelines.web.rug.nl] or catalogue [https://data-catalogue.lifelines.nl/]. Because Lifelines is a not-for-profit cohort and biobank, we charge only the cost price of release and linking materials for researchers of scientific and educational institutes. Lifelines can always be contacted for any information requests [research@lifelines.nl].

Ethics approval

The Lifelines study was approved by the ethics committee of the University Medical Center Groningen, document number METC UMCG METc 2007/152. Informed consent was obtained from all individuals included in the study.

Funding

The Lifelines initiative has been made possible by subsidy from the Dutch Ministry of Health, Welfare and Sport, the Dutch Ministry of Economic Affairs, the University Medical Center Groningen, Groningen University and the provinces in the north of the Netherlands (Drenthe, Friesland, Groningen)

Author Contributions

A.S., J.R. and A.D. drafted the paper and were involved in reviewing the paper in detail. A.S. finalized writing the paper. P.H., G.N. and J.G.M.R. were involved as members of the Scientific Advisory Committee of Lifelines, and were involved in reviewing this paper in detail.

Conflict of interest

None declared.

References

1. Scholtens S, Smidt N, Swertz MA *et al.* Cohort Profile: LifeLines, a three-generation cohort study and biobank. *Int J Epidemiol* 2015;44:1172–80.
2. Fredrickson J, Maruff P, Woodward M *et al.* Evaluation of the usability of a brief computerized cognitive screening test in older people for epidemiological studies. *Neuroepidemiology* 2010;34:65–75.

3. Warmink-Perdijk WDB, Peters LL, Tigchelaar EF *et al.* Lifelines NEXT: a prospective birth cohort adding the next generation to the three-generation Lifelines cohort study. *Eur J Epidemiol* 2020;**35**:157–68.
4. Tigchelaar EF, Zhernakova A, Dekens JA *et al.* Cohort Profile: LifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: study design and baseline characteristics. *BMJ Open* 2015;**5**:e006772.
5. Xia C, Rook M, Pelgrim GJ *et al.* Early imaging biomarkers of lung cancer, COPD and coronary artery disease in the general population: rationale and design of the ImaLife (Imaging in Lifelines) study. *Eur J Epidemiol* 2020;**35**:75–86.
6. Mc Intyre K, Lanting P, Deelen P *et al.* Lifelines COVID-19 cohort: investigating COVID-19 infection and its health and societal impacts in a Dutch population-based cohort. *BMJ Open* 2021;**11**:e044474044474.
7. Zijlema WL, Smidt N, Klijs B *et al.* The LifeLines Cohort Study: a resource providing new opportunities for environmental epidemiology. *Arch Public Health* 2016;**74**:32.
8. Timmermans EJ, Lakerveld J, Beulens JWJ *et al.* Cohort Profile: The Geoscience and Health Cohort Consortium (GECCO) in the Netherlands. *BMJ Open* 2018;**8**:e021597.
9. Zhu Y, Minovic I, Dekker LH *et al.* Vitamin status and diet in elderly with low and high socioeconomic status: The Lifelines-MINUTHE study. *Nutrients* 2020;**12**:2659. doi: 10.3390/nu12092659.
10. van der Ende MY, Hartman MHT, Schurer RAJ *et al.* Prevalence of electrocardiographic unrecognized myocardial infarction and its association with mortality. *Int J Cardiol* 2017;**243**:34–39.
11. Iozzia G, de Miranda Azevedo R, van der Harst P, Rosmalen JGM, de Jonge P, Roest AM. Association of recognized and unrecognized myocardial infarction with depressive and anxiety disorders in 125,988 individuals: a report of the Lifelines cohort study. *Psychosom Med* 2020;**82**:736–43.
12. Dekker LH, Rijnks RH, Strijker D, Navis GJ. A spatial analysis of dietary patterns in a large representative population in the north of The Netherlands – the Lifelines cohort study. *Int J Behav Nutr Phys Act* 2017;**14**:166–68.
13. Moller G, Sluik D, Ritz C *et al.* A protein diet score, including plant and animal protein, investigating the association with HbA1c and eGFR – The PREVIEW Project. *Nutrients* 2017;**9**:763.
14. Baranska A, Mujagic Z, Smolinska A *et al.* Volatile organic compounds in breath as markers for irritable bowel syndrome: a metabolomic approach. *Aliment Pharmacol Ther* 2016;**44**:45–56.
15. Kurilshikov A, van den Munckhof ICL, Chen L *et al.*; LifeLines DEEP Cohort Study, BBMRI Metabolomics Consortium. Gut microbial associations to plasma metabolites linked to cardiovascular phenotypes and risk. *Circ Res* 2019;**124**:1808–20.
16. van der Plaats DA, de Jong K, de Vries M *et al.*; Biobank-based Integrative Omics Study Consortium. Occupational exposure to pesticides is associated with differential DNA methylation. *Occup Environ Med* 2018;**75**:427–35.
17. Savas M, Muka T, Wester VL *et al.* Associations between systemic and local corticosteroid use with metabolic syndrome and body mass index. *J Clin Endocrinol Metab* 2017;**102**:3765–74.
18. Savas M, Vinkers CH, Rosmalen JGM *et al.* Systemic and local corticosteroid use is associated with reduced executive cognition, and mood and anxiety disorders. *Neuroendocrinology* 2020;**110**:282–91.
19. Demetrashvili N, Smidt N, Snieder H, van den Heuvel ER, Wit EC. Variance components models for analysis of big family data of health outcomes in the Lifelines cohort study. *Twin Res Hum Genet* 2019;**22**:4–13.
20. Monden R, Rosmalen JGM, Wardenaar KJ, Creed F. Predictors of new onsets of irritable bowel syndrome, chronic fatigue syndrome and fibromyalgia: the Lifelines study. *Psychol Med* 2020. doi: 10.1017/S0033291720001774. Online ahead of print.