International Journal of Population Data Science

Journal Website: www.ijpds.org

Data resource profile: the ORIGINS project databank: a collaborative data resource for investigating the developmental origins of health and disease

Belinda C. Davey[1,*], Wesley Billingham[1], Jacqueline A. Davis[1,2,3], Lisa Gibson[1,2], Nina D'Vaz[1,2], Susan L. Prescott[1,2,4,5,6], Desiree T. Silva[1,2,3,4,7], and Sarah Whalan[1,*]

Submission History

Submitted:	24/01/2024
Accepted:	19/06/2024
Published:	30/09/2024

[1]Telethon Kids Institute, North Entrance Perth Children's Hospital, 15 Hospital Ave, Nedlands, WA 6009, Australia
[2]Edith Cowan University, School of Medical and Health Sciences, Edith Cowan University, Perth, WA 6027, Australia
[3]Curtin University, School of Population Health, Faculty of Health Sciences, Curtin University, Perth, WA 6102, Australia
[4]The University of Western Australia, Medical School, University of Western Australia, Nedlands, WA 6009, Australia
[5]Nova Institute for Health, Scholars Program, Nova Institute for Health, Baltimore, MD 21231, USA
[6]Perth Children's Hospital, Department of Immunology, Perth Children's Hospital, Nedlands, WA 6009, Australia
[7]Joondalup Health Campus, Department of Paediatrics and Neonatology, Joondalup Health Campus, Perth, WA 6027, Australia

Abstract

Introduction
The ORIGINS Project ("ORIGINS") is a longitudinal, population-level birth cohort with data and biosample collections that aim to facilitate research to reduce non-communicable diseases (NCDs) and encourage 'a healthy start to life'. ORIGINS has gathered millions of datapoints and over 400,000 biosamples over 15 timepoints, antenatally through to five years of age, from mothers, non-birthing partners and the child, across four health and wellness domains: 'Growth and development', 'Medical, biological and genetic', 'Biopsychosocial and cognitive', 'Lifestyle, environment and nutrition'.

Methods
Mothers, non-birthing partners and their offspring were recruited antenatally (between 18 and 38 weeks' gestation) from the Joondalup and Wanneroo communities of Perth, Western Australia from 2017 to 2024. Data come from several sources, including routine hospital antenatal and birthing data, ORIGINS clinical appointments, and online self-completed surveys comprising several standardised measures. Data are merged using the Medical Record Number (MRN), the ORIGINS Unique Identifier and the ORIGINS Pregnancy Number, as well as additional demographic data (e.g. name and date of birth) when necessary.

Results
The data are held on an integrated data platform that extracts, links, ingests, integrates and stores ORIGINS' data on an Amazon Web Services (AWS) cloud-based data warehouse. Data are linked, transformed for cleaning and coding, and catalogued, ready to provide to sub-projects (independent researchers that apply to use ORIGINS data) to prepare for their own analyses. ORIGINS maximises data quality by checking and replacing missing and erroneous data across the various data sources.

Conclusion
As a wide array of data across several different domains and timepoints has been collected, the options for future research and utilisation of the data and biosamples are broad. As ORIGINS aims to extend into middle childhood, researchers can examine which antenatal and early childhood factors predict middle childhood outcomes. ORIGINS also aims to link to State and Commonwealth data sets (e.g. Medicare, the National Assessment Program – Literacy and Numeracy, the Pharmaceutical Benefits Scheme) which will cater to a wide array of research questions.

Keywords
databank; ORIGINS; DoHAD; birth cohort; non-communicable disease; microbiome; longitudinal; population-level; childhood; development

*Corresponding Author:
Email Address: m.rees-roberts@kent.ac.uk (Melanie Rees-Roberts)

https://doi.org/10.23889/ijpds.v8i6.2388

Key features

What is unique about the dataset

- The ORIGINS Project ("ORIGINS") is a data and biobank that provides access to comprehensive longitudinal data and biosamples collected antenatally through to early childhood from mothers, non-birthing partners and children, as well as some environmental household and neighbourhood data. ORIGINS sub-projects (independent researchers that apply to use ORIGINS data) can add their own measures and actively implement interventions, the results of which are made available to other researchers.

The dataset

Why the dataset was created

- ORIGINS was created to identify factors that contribute to 'a healthy start to life' and to implement interventions to reduce the rising epidemic of non-communicable diseases (NCDs). ORIGINS allows researchers to examine the complex interactions between the 'exposome' (environmental factors throughout life) and epigenetics, proteomics, metabolomics and the microbiome.

Details about the dataset: location, size, composition of the population

- ORIGINS comprises non-active participants (n=4,457 mothers, 5,227 children and 1,117 non-birthing partners) and active participants (3,448 mothers, 3,806 children and 1,403 non-birthing partners), who are distinguished by the data provided by each. Both groups provide ORIGINS data routinely collected antenatally and during birth (that is, via data linkage), while ORIGINS also collects additional data from active participants via online surveys and face-to-face clinical appointments. For active participants, additional data are collected over several timepoints, antenatally to five years of age, covering four health and wellness domains: 'Growth and Development', 'Medical, Biological and Genetic', 'Biopsychosocial and Cognitive', 'Lifestyle, environment and nutrition', as well as baseline demographic data and biosamples (e.g. blood, urine, meconium, colostrum, breast milk, hair, umbilical cord gases).

Description of any data linkage

- As just noted, ORIGINS links to routine antenatal and birthing data for active and non-active participants collected by hospitals and midwives. ORIGINS is currently applying to link to State, Commonwealth and cross-jurisdictional health, demographic and educational datasets such as the Australian Immunisation Register (AIR), the National Assessment Program – Literacy and Numeracy (NAPLAN), Medicare, the Pharmaceutical Benefits Scheme, and the Registry of Births, Deaths and Marriages.

Main categories of data

- Data are categorised in several different ways: by participant (mother, non-birthing partner, child), study timepoint (15 in total, at time of writing), consent level (active versus non-active), data source (collected by ORIGINS versus data linkage), and domain (mentioned above).

How to collaborate and access the dataset [contact details]

- To become an ORIGINS sub-project, and/or to access ORIGINS' data, a research proposal needs to be reviewed and approved by the ORIGINS Project Management and Scientific Committee and then an agreement is signed, with cost-recovery fees negotiated according to the size and complexity of data and sample requests. Application and cost details can be found on the ORIGINS website, The ORIGINS Project Subsite (telethonkids.org.au), and real-time data and sample collection details can be found at the ORIGINS Data Catalogue, https://bitly.ws/34uGB.

Background

The ORIGINS Project ("ORIGINS") is a collaboration between Telethon Kids Institute and the Joondalup Health Campus in Perth, the state capital of Western Australia. ORIGINS is a longitudinal, population-level birth cohort with data and biosample collections that aim to facilitate research to reduce non-communicable diseases (NCDs) and encourage 'a healthy start to life'. According to the Developmental Origins of Health and Disease (DoHAD), the environmental context in the early years influences lifelong health, including microbial diversity, nutrition, nature, and social interactions [1, 2]. Maternal and paternal health at preconception, and antenatally, as well as during early childhood, influence the multifaceted interactions between the child's physical, structural, immune, metabolic and emotional behaviour and development, which shape health outcomes and disease susceptibility throughout life. Much of the impact may be subtle, so that effects may not become evident until much later in life [3, 4].

Globally, NCDs impact individuals across the life-course, and are responsible for poor life-quality, substantial disease burden, premature death, and excessive costs to governments and societies [5, 6]. NCDs such as obesity, heart disease, and allergy share underlying causes, such as immune dysfunction, chronic low-grade inflammation, metabolic dysregulation and alterations of the human microbiome (dysbiosis), which suggest similar causal pathways [7, 8]. Adverse events and exposures during early-life development, including pre-conception and in-utero, can have profound effects on structures, functions and behaviours that contribute to causes of NCDs [3].

Given this, typical single-intervention, single timepoint, discrete research studies will not be adequate to overcome these challenges. Rather, a more integrated systems approach that considers the entirety of the exposome, and extends beyond the impact of individual risk factors in individual body systems, is required [9, 10]. Large, longitudinal, birth cohort studies covering several domains allow a transgenerational,

multifaceted holistic approach to examining the complex interactions that impact upon the developmental trajectory and shape lifelong health outcomes [11].

ORIGINS has developed a world-class platform, collecting several measures and biosamples across multiple domains to capture outcomes and exposure data across several timepoints, antenatally to early childhood. These collections cover four health and wellness domains: 'Growth and development', 'Medical, biological and genetic', 'Biopsychosocial and cognitive', 'Lifestyle, environment and nutrition', as well as baseline demographic data and biosamples (e.g. blood, urine, meconium, colostrum, breast milk, hair, umbilical cord gases). This comprehensive sample and data collection allows examination of the multifaceted interactions between biopsychosocial, immune, metabolic and epigenetic influences upon and emotional, behavioural and physical development.

An additional key aspect of ORIGINS is the nesting of research projects (referred to as "sub-projects") to harmonise recruitment processes as well as data and sample collection, and the ability of projects to add their own data and sample collection, which is then returned to ORIGINS to share with other researchers. These projects can be both observational and/or interventions, which allows ORIGINS to be an integrated data and biobank platform, as well as a prospective birth cohort. This also means that researchers can design studies that support causational modelling by nesting experimental or intervention studies, or they can design studies that allow predictive modelling using data already collected, or by adding their own measures and/or samples.

Sub-projects are approved research projects conducted by independent researchers who are using ORIGINS data and/or biosamples. ORIGINS undertakes a rigorous governance process to approve and then subsequently facilitate on boarded research projects, referred to as 'sub-projects' from here-in.

Methods

The study population, data sources and data linkage methods are described in this section.

Study population

Pregnant women, and their non-birthing partners, who were due to give birth at Joondalup Health Campus (JHC), in the northern corridor of Perth, Western Australia, were invited to participate in ORIGINS during their first or subsequent antenatal clinic appointments from 2017 onwards. Parents then consented their newborn babies at the time of birth.

Active (2017-2023) vs. non-active enrolment (2017-2024)

ORIGINS comprises non-active participants (n=4,457 mothers, 5,227 children and 1,117 non-birthing partners) and active participants (3,448 mothers, 3,806 children and 1,403 non-birthing partners), who are distinguished by the data provided by each. Non-active participants are those who were willing to be involved in the study, but did not want to commit to the extra time and effort required to be an active participant. Instead, they consented to ORIGINS linking to data routinely collected by hospital and midwives, antenatally and during birth. For active participants, as well as linking to this routine data, additional data are collected by ORIGINS over several timepoints, antenatally to five years of age, via online surveys, face-to-face clinical appointments and the collection of biosamples. Both active and non-active participants could then consent to ORIGINS linking to additional State and Commonwealth Government data.

Hence, ORIGINS consent forms were broken into three parts:

1. Full participation: consent obtained from active participants to complete surveys, provide samples and attend clinic visits.
2. Collection of routine data and routine biological samples: consent obtained from both active and non-active participants to collect data from JHC medical files and databases.
3. Data linkage to collections held by State and Commonwealth Government: optional additional consent by active and non-active participants to collect data held by Australian Government, such as the Australian Immunisation Register (AIR).

Note that, unless otherwise specified, we hereafter use the term 'participant' to refer to all types of participants, both active and non-active, and mothers, non-birthing partners and children. While all active participants have now been recruited (active participants were recruited from 2017 to 2023), recruitment of non-active participants is ongoing, and should be complete by 2024, having started in 2017 as well. In addition, ORIGINS aims to enhance the databank through external linkage to available State and Commonwealth datasets (e.g. Medicare, the National Assessment Program – Literacy and Numeracy, the Pharmaceutical Benefits Scheme, and the Australian Immunisation
Register).

Once consented into The ORIGINS Project, each individual is issued an ORIGINS Unique Identification Number (OUID), including the mother, non-birthing partner and children. If the mother or non-birthing partner have previously consented into the study with a different pregnancy, they will retain the previously issued OUID. This allows ORIGINS to identify siblings within the databank. An ORIGINS Pregnancy Number (OPN) is generated for each pregnancy, which is especially important for multiple births (i.e. twins and triplets). The JHC Medical Reference Number (MRN) is collected from Meditech (the database used by JHC) and entered into The ORIGINS Project database. Figure 1 provides an overview of participant recruitment and retention across the antenatal and birth timepoints. The top half of the figure shows the number of pregnancies consented to the study antenatally, whereas the bottom half of the figure shows the mothers and children who remain consented in the study after birth. Of the 17,403 women approached across one or more pregnancies antenatally or at birth, 8,043 (46.2%) declined, 3,448 (19.8%) remained consented as active participants at birth (with 3,806 children),

Figure 1: Overview of recruitment and retention of mothers and babies antenatally and at birth*

*Note that the number of mothers, pregnancies and children differ slightly due to pregnancies with multiple births (i.e. twins and triplets), and enrolment of siblings.

and 4,457 (25.2%) consented to be non-active participants at birth (with 5,227 children).

Comparing target population to Western Australia and Australia

Women giving birth at JHC tend to be primarily from Joondalup, north of Perth, the capital city of Western Australia. Table 1 shows the demographic characteristics of Joondalup city compared to Western Australia and Australia [12]; a more detailed comparison can be found at 2021 Joondalup, Census All persons QuickStats | Australian Bureau of Statistics (abs.gov.au). Note that, these data were not available for mothers in Joondalup specifically, and ORIGINS collects data about families, so the demographic characteristics of Joondalup generally have been provided. The table shows that, compared to Western Australia and Australia, fewer people in Joondalup identify as Aboriginal

Table 1: A comparison of the demographic characteristics of Joondalup compared to Western Australia and Australia [12]

Demographic Characteristics	Joondalup (%)	Western Australia (%)	Australia (%)
Aboriginal and/or Torres Strait Islander	0.9	3.3	3.2
Highest level of education			
Bachelor degree or above	27.0	23.8	26.3
Advanced Diploma and Diploma level	11.2	9.3	9.4
Certificate level IV	4.1	3.9	3.5
Certificate level III	14.5	13.9	12.6
Year 12	16.2	5.0	14.9
Marital status			
Married	54.6	47.3	46.5
Separated	2.7	3.3	3.2
Divorced	7.8	8.8	8.8
Widowed	4.1	4.4	5.0
Never married	30.8	36.1	36.5
Ancestry top responses			
English	46.2	37.6	33.0
Australian	30.3	29.7	29.9
Irish	12.1	8.8	9.5
Scottish	11.0	8.7	8.6
Italian	4.7	5.2	4.4
Religious affiliation			
No religion	44.4	42.5	38.4
Catholic, Anglican, or 'mild' Christian	37.6	32.2	32.5
Not stated	4.4	7.5	6.9
Occupation			
Professionals	25.3	22.0	24.0
Technicians and trades workers	15.2	15.3	12.9
Clerical and administrative workers	13.9	12.1	12.7
Managers	13.6	12.3	13.7
Labourers	6.7	9.4	9.0
Machinery operators and drivers	4.3	7.7	6.3

and/or Torres Strait Islanders, are more likely to be married, and tend to be slightly better educated, are more likely to be in "white-collar" professions, from North Western Europe and identify as Christian or with no religion.

Comparing demographics of active versus non-active participants

Table 2 provides a comparison of key demographic variables among active versus non-active mothers. The table shows that, compared to non-active mothers, active mothers are, on average, slightly older (less than half a year older), are more likely to be having their first baby, have a higher mean socioeconomic status (0.2 deciles, or 2%, greater) as indicated by their higher average Index of Relative Socioeconomic Advantage and Disadvantage, and are less likely to identify as Aboriginal and/or Torres Strait Islander ($p < 0.001$). These differences are in line with the well-known observation that people from more socioeconomically advantaged backgrounds are more likely to be actively involved in research, who in turn are more likely to be older when they have their first child [13]. In addition, those with only one child are more likely to feel they have the time to be involved in research. These differences are very small, as, given the large sample size, even very small differences are statistically significant. A distinction needs to be made between statistical significance and practical or clinical significance. Researchers will need to make their own judgements regarding the impact of these differences on their specific research questions.

Comparing ORIGINS-enrolled to non-enrolled JHC birthing mothers

ORIGINS is fortunate to have access to additional routine data (from the Midwives Notification System) collected antenatally from some birthing mothers not enrolled at the Joondalup Health Centre ($n = 4{,}131$) during the study period. This allows a comparison of those enrolled in ORIGINS to those not enrolled in ORIGINS. These data are available to ORIGINS as part of the sub-project, "A Family Journey at JHC: analyses of routinely collected data", which utilises waiver of consent that allows all routinely collected data at JHC

Table 2: Demographic differences between active and non-active mothers with a child consented into ORIGINS[#] (significant differences bolded; $p < 0.001$)

Characteristic†	Overall N = 7,905	Active N = 3,448	Non-active N = 4,457
Age at Birth (years)* Mean (SD)	31.86 (4.88)	**32.20 (4.71)**	**31.60 (4.99)**
Pre-pregnancy weight (kg) Mean (SD)	70.10 (15.27)	70.44 (15.46)	69.75 (15.08)
Pre-pregnancy BMI N(%)			
Healthy weight (18.5 to <25)	3,222 (51%)	1,632 (51%)	1,588 (51%)
Underweight (<18.5)	196 (3%)	99 (3%)	97 (3%)
Overweight (25 to <30)	1,700 (27%)	863 (27%)	837 (27%)
Obese (30+)	1,217 (19%)	637 (20%)	580 (19%)
Previous Pregnancies (Gravidity)* N(%)			
0	2,675 (37%)	**1,358 (42%)**	**1,317 (33%)**
1	2,100 (29%)	**884 (28%)**	**1,216 (30%)**
2	1,244 (17%)	**511 (16%)**	**733 (18%)**
3	593 (8%)	**237 (7%)**	**356 (9%)**
4 or more	617 (9%)	**221 (7%)**	**396 (10%)**
Index of Relative Socioeconomic Advantage and Disadvantage Decile Mean (SD)	7.70 (1.82)	**7.81 (1.71)**	**7.61 (1.89)**
Aboriginal and/or Torres Strait Islander N(%)	57 (0.8%)	**8 (0.23%)**	**49 (1.1%)**

BMI = Body Mass Index.
Index of Relative Socioeconomic Advantage and Disadvantage Decile: scored from 1 to 10, where 10 is more socially advantaged.
[#]Metric variables examined via Wilcoxon rank sum test; 2 x 2 categorical variables examined by Fisher's exact test; > 2x2 examined by Pearson's chi-square test of independence; where mothers have more than one child enrolled in ORIGINS (n= 291 active mothers and 667 non-active mothers with multiple children in ORIGINS) demographic data in the table are provided at the time of their first ORIGINS child.
†The total n for each characteristic may differ from that provided overall in the table.

to be available for analyses by ORIGINS. This allows a comparison of those enrolled in ORIGINS (n = 9,360 active and non-active mothers) to those not enrolled in ORIGIN (n = 4,131 mothers). Table 3 provides such a comparison on key demographic variables. Such a comparison allows researchers to be aware of any biases within the sample given the population ORIGINS aims to represent (that is, primarily those giving birth at the Joondalup Health Campus from 2017 to 2023). Note also that Table 3 comprises different comparisons to Table 2 as different data were available for this comparison.

The table shows that compared to non-ORIGINS birthing patients from JHC, ORIGINS participants are, on average, approximately one year older, over 2kg heavier, have slightly higher BMI (by 0.54 points), and are slightly more advantaged (scoring 0.37 decile points higher on the IRSAD) ($p < 0.001$, Wilcoxon rank sum test). Typically, those of higher socioeconomic status weigh less, on average, than those of a lower socioeconomic status [14], which contrasts with findings here. However, weight tends to increase with age [14], which may account for the heavier weight of ORIGINS versus non-ORIGINS patients, despite their average higher socioeconomic status. Again, these differences are small, so may have minimal clinical or practical impact [15], but the large sample size allows such differences to be detected. Further information about the ORIGINS cohort can be found in Talati et al (manuscript in preparation).

Data sources

ORIGINS comprises data from several different sources, including routine data collected primarily at JHC, data collected at ORIGINS clinical appointments, and data collected online for ORIGINS primarily via the 'ORIGINS Core Questionnaire'. Finally, ORIGINS also receives any additional data or samples collected by nested sub-projects. These are described below. For each, a note is added in brackets after the source heading to indicate whether the data are collected for active participants only, or for both active and non-active participants. The various data sources are summarised in Table 4 and Figure 2.

Before describing these sources in more details, note that data are collected at multiple timepoints, antenatally to when the child is five-years. These timepoints include: registration, 18 weeks gestation, 28 weeks gestation, 36 weeks gestation (all summarised as 'antenatal' in the figure below), birth, 2 months, 4 months, 6 months, 9 months, 12 months, 18 months, 24 months, 30 months, 36 months, 48 months, 60 months. These timepoints were subsequently updated to remove 4, 9 and 30 months, other than the Ages and

Table 3: Mean (SD) demographic differences between patients enrolled versus not enrolled in ORIGINS (significant differences bolded)

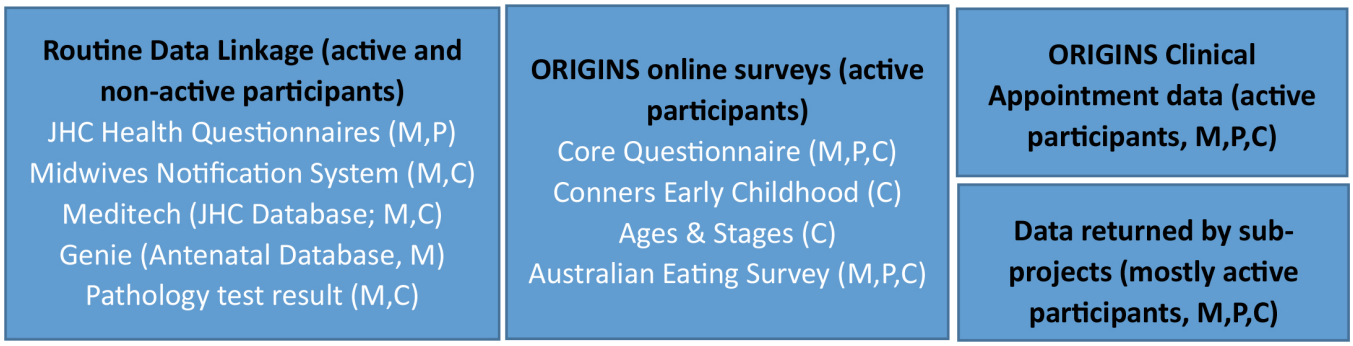
Characteristic	Overall (n = 13,518)	Not enrolled (n = 4,131)	Enrolled (n = 9,360)
Age*	31.44 (4.91)	**30.66 (4.97)**	**31.80 (4.84)**
Weight*	69.26 (15.21)	**67.76 (14.99)**	**70.06 (15.27)**
BMI*	25.42 (5.21)	**25.04 (5.18)**	**25.60 (5.21)**
IRSAD Decile*	6.95 (1.94)	**6.77 (2.01)**	**7.04 (1.89)**
Aboriginal and/or Torres Strait Islander	94 (0.7%)	37 (0.90%)	57 (0.61%)

*Significant at $p < 0.001$, examined by Wilcoxon rank sum test; BMI = Body Mass Index; IRSAD = Index of Relative Social Disadvantage and Advantage.

Table 4: Routine data collected from mothers, non-birthing partners and their children by timepoint (M = mother, P = non-birthing partner, C = child)

Data sources	Antenatal	Birth	Postnatal
JHC Mother's Health Questionnaire	M		
JHC Partner's Health Questionnaire	P		
Genie (Antenatal Database)	M		
Meditech (JHC Database)	M, P		
Midwives' Notification System (MNS)		M, C	
Paper-based medical files	M	M,C	M,C

Figure 2: Summary of data sources currently utilised by ORIGINS (M = mother, P = non-birthing partner, C = child)



Stages Questionnaire (ASQ) which is still administered at 4 and 9 months. A summary of the timeline is shown in Figure 3.

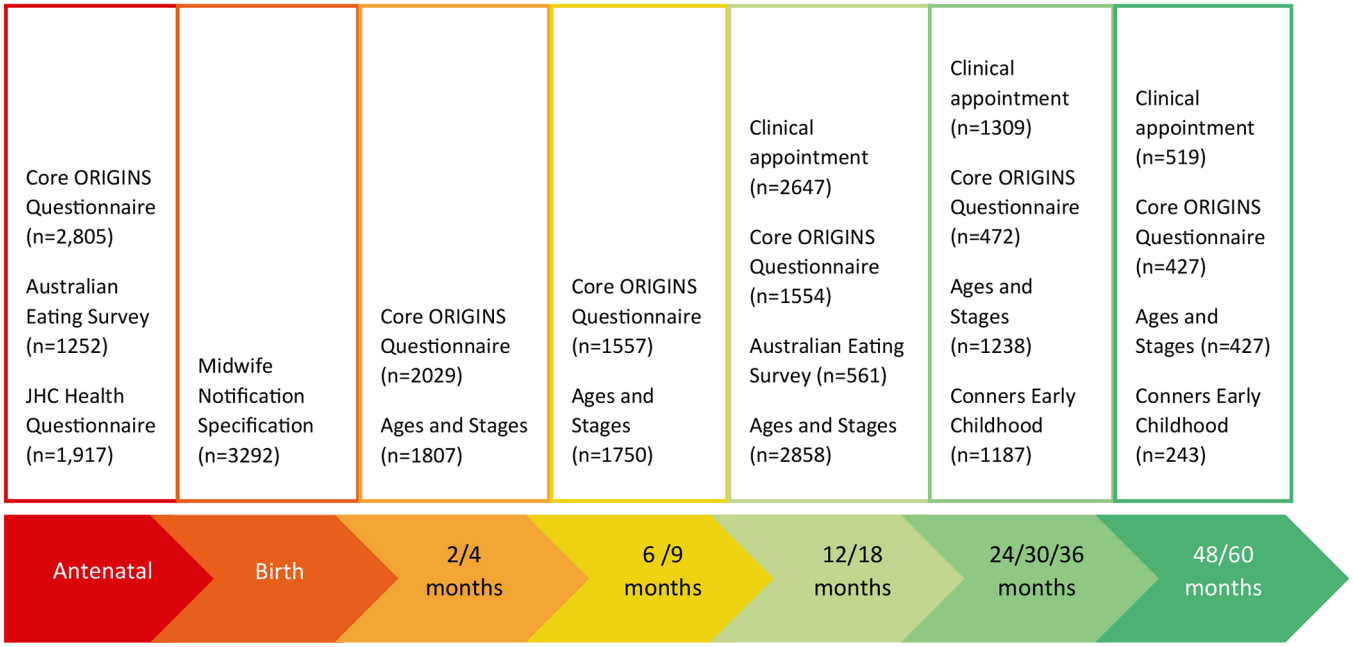
Routine data linkage (for active and non-active participants)

As noted above, ORIGINS has access to all data routinely collected at antenatal appointments, and at birth, for both non-active and active participants. All women giving birth at JHC, and their non-birthing partners, complete the Joondalup Health Questionnaire (JHQ). ORIGINS is fortunate to have access to all of these data during the study period, for both active and non-active participants, as well as for those not enrolled in ORIGINS. Hence, analyses can be performed to compare ORIGINS to non-ORIGINS patients, as well as non-active versus active ORIGINS participants with available JHC data.

The maternal digital JHQ collects information on participant characteristics and demographics, birth and pregnancy details, and information about fertility, mental health conditions and allergies. The paternal digital JHQ is similar but instead of pregnancy information, it collects information about marital status, smoking, drinking, drug-use and sexually transmitted diseases. These can be found in the Supplementary Appendix.

Other medical data collected via JHC are also shared with ORIGINS; that is, medical data collected antenatally via the Genie desktop digital patient data collection system at JHC, birthing and postnatal data collected via the Meditech hospital computer system, and data from the Midwives Notification System (MNS), a statutory data collection from the Manager of Administration and Health Information that collects data at birth. The MNS form undergoes frequent changes so the questions change; details of what questions are asked when can be found at Midwives Notification System on the Western Australian Department of Health website: (health.wa.gov.au). Participants also give ORIGINS permission to access their physical hospital files to check and supplement digital hospital data. Again, ORIGINS is fortunate to currently

Figure 3: Data sources by time-point



have access to MNS data for all mothers birthing at JHC during the study period, so comparisons can be made between ORIGINS and non-ORIGINS patients, as well as between non-active and active participants using available MNS data.

ORIGINS also has access to ultrasounds of ORIGINS participants (both non-active and active) during their pregnancy from the Perth Radiological Clinic. A legal and data transfer agreement was signed by both Perth Radiological Clinic and Telethon Kids Institute outlining the details of this agreement. Participant MRNs are used to match their data across both sites.

Similarly, ORIGINS has access to some limited data from participants' routine blood tests during pregnancy (mother) and childbirth (mother and baby) through Western Diagnostics (a pathology provider that has partnered with ORIGINS). This allows ORIGINS access to the mother's ferritin, haemoglobin levels, white blood cell count, platelets, HbA1C, Vitamin D, Glucose Tolerance Test (GTT), Thyroid-stimulating hormone (TSH) test, Group B Streptococcus (GBS) swab results (positive or negative) and BGN blood gases, and to the newborn infants' full blood count, white blood cell count, platelets, serum bilirubin, vitamin D, PGL (to detect genes associated with hereditary paraganglioma or pheochromocytoma), c-reactive protein, cord Ph, lactate (BGN), and ferritin if available. The data in the pdf files is parsed using R-coding into csv format, but this is imperfect, and so needs to be manually checked and corrected by The ORIGINS Team.

ORIGINS clinical appointment database (for active participants only)

All data collected at clinical appointments is stored in ORIGINS' instance of the Research Electronic Data Capture (REDCap) database, such as anthropometry, morphology, dental, allergies (e.g. food, asthma, eczema, skin-prick test results), Nevisense [16] (uses electrical impedance to detect histological changes in the skin), ear otoscopy, optometry, joints, burns, medication use (including antibiotics), blood tests (e.g. serum ferritin levels, haemoglobin, HBA1c, and c-reactive protein), hospital admissions and referrals, family history of mental health issues, sleep and the cardiovascular, respiratory and neurological systems. Anthropometric measures resulting from the Bod Pod and the Paediatric Enclosed Air (PEA) Pod are also recorded. These use new technology that measures air displacement via plethysmography to determine body composition for children or adults and infants, respectively. These are now the gold standard for safe, repeatable and non-invasive assessment of body composition [17]. In addition, the novel Veggie Meter® was used to detect skin carotenoids using refraction spectrometry as a proxy for recent fruit and vegetable intake [18]; this can be used to validate self-report dietary intake data. Finally, during clinical appointments, patients provide their immunisation records and their maternal and child health care 'Purple Book' (a logbook used by midwives and maternal child health nurses to track a newborn's development and vaccinations into early childhood) to be photocopied for later manual entry into ORIGINS' databases.

All of this data were collected at the 1, 3 and 5-year clinical appointments. These are shown in Table 5 below. Much of this data was also self-reported online via the online ORIGINS Core Questionnaire at other timepoints, so will appear in both Tables 5 and 7, but at different time-points.

Key administrative data are also collected, such as participant type (mother, non-birthing partner, child), participant status (withdrawn or lost-to-follow-up), consent level (active or non-active), contact details, sub-projects, and

Table 5: Measures collected from mothers and non-birthing partners by timepoint (M = mother, P = non-birthing partner, C = child)

Measures	(Childhood years)		
	1	3	5
Respiratory conditions, bronchiolitis or asthma, cold, flu, fever, medical conditions	C	C	C
Allergic reactions, eczema, hay fever, seasonal allergies	C	C	C
Oral health	C		C
Paediatric appointments developmental review and assessment: cardiovascular, respiratory and neurological systems	C	C	C
Skin-Prick Test (S), Nevisense [16]$^{+}$	C	C	C
Ear otoscopy, optometry, joints, burns, medication use (including antibiotics)	C	C	C
Anthropometry$^{\#}$	C,M,P	C,M,P	C,M,P
Blood test results (serum ferritin levels, haemoglobin, HBA1c, and c-reactive protein)	C	C	C
Hospital admissions	C	C	C
Family history of mental health issues*	C	C	C

(S) = standardised measure; $^{+}$Nevisense provides a measure of cellular changes to skin cells; $^{\#}$anthropometry includes BOD and PEA Pod body composition measurement, as well as height, weight, dysmorphia and body proportions, and infant anthropometry is more detailed than parental anthropometry; *all measures are from the Medical, Biological and Genetic domain except this one, which is from the Biopsychosocial and Cognitive domain.

status of some measures and biosamples collected. Biosamples are tracked and managed via the online platform Open Specimen. A detailed overview is provided in D'Vaz et al [19] but a summary of collected biosamples is provided in Table 6.

ORIGINS online survey data (for active participants only)

ORIGINS online survey data primarily comprises the Core Questionnaire, administered at several timepoints throughout the ORIGINS journey. The survey comprises several items and measures administered at several different timepoints; not all measures were administered at all timepoints. What is administered when, and to whom, is summarised in Tables 7 and 8. Note that these tables only include measures collected at more than one timepoint; a full list of all measures is provided in Tables 1 to 5 in the Supplementary Appendix, for each domain. Incentives for Core Questionnaire completion are provided throughout the study period, primarily $20 gift cards for completion of the initial antenatal questionnaire and the completion of the 5-year Core Questionnaire, as well as the chance to go into a monthly draw for a $100 gift card.

Due to licensing arrangements, there are also three additional standardised measures collected by others on behalf of ORIGINS, the Conner's Early Childhood Questionnaire, the Strengths and Difficulties Questionnaire and the Australia Eating Survey [20, 21]. The child version of the Australian Eating Survey was originally administered at 1 and 3-years postnatally. However, low completion rates, and issues with the birthing mother incorrectly entering her own data instead of the child's, led the child version to be replaced by the 68-item Food Frequency Questionnaire [22]. This update resulted in greatly improved completion rates, from 20% to 58% completion per month.

Sub-project data (could be either active, non-active, or both, depending on the sub-project)

ORIGINS has incorporated 51 sub-projects to date, many of which have collected their own data over different timepoints, while others have analysed existing data collected by ORIGINS. Sub-projects cover nine major research areas: Allergy, Inflammation and Immunity; Brain and Behaviour; Growth and Development; Nutrition and Metabolism; Environment and Lifestyle; Mental Health and Wellbeing; Microbiome; Parenting; and Oral Health. Nested sub-projects are required to return additional data they collect to The ORIGINS Project, so these data can be made available to other researchers. Further details about each sub-project can be found at The ORIGINS Project: Sub-projects website, https://originsproject.telethonkids.org.au/about-the-origins-project/.

ORIGINS can add measures to REDCap on behalf of sub-projects collecting additional data. For sub-projects conducting interventions, contact details of interested ORIGINS participants are passed to the sub-projects to organise data collection directly from participants.

Data security

Data collected through these various sources are collated and stored in a private, firewall-protected Telethon Kids Institute network that is backed up at regular intervals. Within these databases and services, access is limited to role-specific permissions and monitored by the ORIGINS Data Manager. Access requires staff to sign Telethon Kids Institute and JHC confidentiality agreements.

The primary database is REDCap, a service maintained by Telethon Kids Institute information technology staff within the secure Telethon Kids Institute data network. All electronic

Table 6: Biobank samples collected from mothers (M), non-birthing partners (P) and children (C) at each timepoint

	Antenatal (weeks' gestation)		Postnatal (months)			Childhood (years)		
Measures	**20**	**36**	**Birth**	**2**	**6**	**9**	**2**	**5**
Blood	M, P	M				C	C	C
Urine	M	M		M, C	M, C	C	C	C
Buccal swab, saliva	M, P	M	P			M, C	C	C
Stool	M	M		M, C	M, C	C	C	C
House dust		M				M	M	M
Meconium, cord blood and tissue, placenta, colostrum			M,C					
Breast milk				M	M	M		
Hair		M				M		

Table 7: All items collected online by ORIGINS at more than one timepoint for mothers (M), non-birthing partners (P) and children (C) in the Biopsychosocial and Cognitive, and Lifestyle and Environment domains (all standardised scales are referenced)

Domain and measures	**Antenatal**	**Postnatal**			**Childhood**				
Biopsychosocial and Cognitive		**Birth**	**2/4**	**6/9**	**1**	**2**	**3**	**4**	**5**
Strengths and Difficulties Questionnaire [23]							C		C
Conner's Early Childhood [24]							C		C
Connor-Davidson Resilience Scale-Short Form [25]	M,P					M,P			M,P
DASS-21 [26]	M,P	M,P		M,P	M,P	M,P		M,P	M,P
Attachment Scale [27, 28]	M			M,P	M,P				
Perceived Social Support Scale [29, 30]	M	M			M,P	M,P			P
Stressful Life Events [31]		M,P			M,P	M,P			
Mental Health Continuum [32]	M,P			M,P			M,P		M,P
Trait-Hope Scale [33]							M		M
Lifestyle and Environment	**Antenatal**	**Birth**	**2/4**	**6/9**	**1**	**2**	**3**	**4**	**5**
Drinking water		M			M	M			M
Cooking		M		M	M				
Buckner's Neighbourhood Cohesion [34]						M			M
Neighbourhood Environment Walkability Scale [35]	M,P					M			M
Electronic devices in the household	M,P	M		M	M	M	M	M	M
Alcohol, smoking, drug use [36]	M	M,P	M,P	M,P	M,P	M,P		M,P	M,P
Physical Activity In The Last 7 Days (IPAQ) [37]	M	M,P		M,P	M,P				
Godin Leisure Time [38]		M,P		M,P	M,P	M,P	M		M,P
Health Related Quality of Life (EQ-5D-5L) [39]	M,P	M		M		M			M
Nature [40, 41]*					M,P,C	M,P	M		C
Time in the sun	**AN**	M,P	C	C	C	C	C	C	C
Diet [20–22, 42]*	M	M	M,P	M,P	M,P,C	M,P,C	M,P,C	M,P	M,P,C
Childcare, playgroup, play time	M	M	C	C	C	C	C	C	C
Technology or internet use	M			C	C	C	C	M,C	M,C
Pittsburgh Sleep Index [43]	M	M	M	M	M	M			M

*Nature = Connectedness to Nature Index [41], Nature Play and/or Nature-Relatedness [40]; Diet = Australian Eating Survey [20, 44], Food Frequency Questionnaire [22] and/or Mediterranean Diet Index [45].

communication between REDCap and users, either within the Telethon Kids Institute intranet or the public internet, is encrypted and requires multi-factor authentication to access. Participant questionnaires that are generated through REDCap require participants to log in using a username and password.

An Amazon Web Services (AWS) storage platform, held on a secure Telethon Kids Institute data network, is utilised to integrate and link data from all of these different data sources into one central location. The data within AWS are de-identified and AWS is used primarily for reporting purposes, as well as to integrate The ORIGINS Project Smart App. This Smart App doesn't store any data itself, as it is used primarily to communicate with participants rather than for data collection, and requires participants to log in with a username and password.

Table 8: All items collected online by ORIGINS at more than one timepoint for mothers (M), non-birthing partners (P) and children (C) in the Medical and Genetic, and Growth and Development domains (all standardised scales are referenced)

Measures	Antenatal (weeks' gestation)		Postnatal (months)		Childhood (years)				
Medical and Genetic	**20**	**36**	**2**	**6**	**1**	**2**	**3**	**4**	**5**
Continence	M		M	M	M	M	M	M	M
COVID diagnosis and vaccinations	M	M	M	M	M	M	M	M	M
Musculoskeletal pain	M	M		M	M	M	M	M	M
Prescription medication use	M	M	M	M	M	M	M	M	M
Daytime naps			M	M	M	M			M
Respiratory conditions, bronchiolitis or asthma, cold, flu, fever, medical conditions			C	C	CA	C	CA	C	CA
Allergic reactions, eczema, hay fever, seasonal allergies			C	C	CA	C	CA	C	CA
Oral health		M		C	CA	C			CA
Child's health			C	C	CA	C			
Growth and Development	**20**	**36**	**2**	**6**	**1**	**2**	**3**	**4**	**5**
Breastfeeding, formula, solids	M	M	C	C					
Brief-Infant Sleep Questionnaire [46]			C	C	C				C
Ages & Stages Questionnaire [47]			C	C	C	C	C	C	C

CA = was asked at the Clinical Appointment, rather than via the online ORIGINS Core Questionnaire; COVID data collection started in April 2020.

Data merging and sharing

As ORIGINS collects data from several different sources, data are received in different formats. When a sub-project requests particular data, ORIGINS will identify the relevant data sources and then merge the data from these sources into one file to be securely shared with the particular sub-project. When combining data across the various data sources, OUIDs are used (one per person and one per pregnancy) as well as the hospital MRN when available. When participant data are provided to sub-projects conducting observational research, each participant is provided a unique de-identified participant identification number. ORIGINS keeps a copy of this unique number as well as each participant's original OUID, so participants can be re-identified and merged when nested sub-projects return their data to ORIGINS. For sub-projects that are collecting additional data from participants.

ORIGINS has worked in partnership with an independent data analytics consultancy to develop an integrated databank which extracts, links, ingests, integrates and stores complex ORIGINS data on an AWS cloud-based data platform. The data platform provides a solution to overcome issues in using and interpreting the large volume of data. Key data are linked, transformed for cleaning and coding, and catalogued, ready to provide to sub-projects to prepare for their own analyses. The platform merges with other applications, such as Microsoft's Power BI, to permit ad-hoc requests for static and dynamic data, provide a 360° view of a participant, and full extraction of specific fields.

Examples of the end applications that rely on the AWS data platform are an ORIGINS Smart App; a Power BI self-service tool to assist with data queries and data extraction as well as live Power BI quality and progress monitoring graphs and reports; automated email reminders to assist with participant engagement and milestone completion; and a publicly available data visualisation catalogue so anyone can use the Power BI data to determine the availability of data or biosamples for participants with particular characteristics and timepoints.

Results

In this section, the structure of the data set, the completion rates of some standardised measures for completed timepoints (antenatal and birth) and a discussion of data quality are provided.

Structure

As noted above, The ORIGINS Project is an integrated data platform that combines data from various sources. There is no single file where all data are kept. Rather, data are prepared in response to sub-project data requests. The characteristics of the dataset are outlined in the Methods section above, when providing a detailed description of the various data sources. Data are categorised by domain, as discussed in the Methods section.

Compliance characteristics

Compliance for some standardised measures can be seen in Table 9 for active mothers, and Table 10 for active non-birthing partners, including an inferential comparison across a few demographic characteristics between those who completed a given measure (completers) and those who did not (non-completers). Demographic characteristics were chosen for comparison. Data are provided antenatally only, as no other

timepoints are yet complete except for birth, which comprises no standardised measures.

Table 9 shows that completion rates across all measures could be improved. The measures are listed in descending order according to proportion completed, which happens to be the same order as they are presented to participants, indicating that participant fatigue is likely to largely be responsible for the lower levels of completion for the later measures. One exception is The Mental Health Continuum, which was added after the study had started, accounting for the relatively low completion rates. Nonetheless, mean age and IRSAD are similar between completers and non-completers for all measures, even where significantly different ($p < 0.001$); that is, all significant mean differences in age are within 1 year difference, and the one significant mean difference in IRSAD Decile is only 0.22 (out of 10 points, so equivalent to 2.2% difference). The largest significant difference between completers and non-completers is the percentage of Caucasians. Compared to non-completers, the percentage of Caucasians is at least 5% greater for completers of the IPAQ, the Pittsburgh Sleep Quality, the Australian Eating Survey, and the Connor-Davidson Resilience Scale ($p < 0.001$). Hence, there is minimal evidence of any difference between completers and non-completers for active mothers across measures, except on the percentage of Caucasians. Sub-project researchers will need to be aware of these differences, and consider the impact on their findings.

The rate of completion of some non-birthing partner measures is low. Again, the proportion completed aligns with the order the measures were presented to participants, indicating participant drop-off during study completion. In addition, the Connor-Davidson Resilience Scale Short Form was added later for non-birthing partners. The rate of completion for non-birthing partners in general is lower than for mothers, which is to be expected as some mothers don't have a partner, or their partner is not available to complete the survey. The only significant difference between completers and non-completers for non-birthing partners was that those who completed that DASS-21 had higher sociodemographic status as indicated by their IRSAD decile, on average ($p < 0.001$).

Fortunately, the frequency of completion is still greater than 300 for all measures other than the Australian Eating Survey, which allows some insights into the cohort, though this is obviously far from the full population. In addition, there is no evidence of a difference in either mean age or mean IRSAD Decile between completers and non-completers of any measure (utilising a significance level of .001, due to the large sample size).

Data quality

The databank team within ORIGINS have maximised the quality of the various sources of data by conducting extensive checks of missing or erroneous data, and have used alternative sources to check and replace missing or contradictory information. Moreover, ORIGINS engages in considerable consumer involvement and feedback via dedicated Participant Reference and Community Reference Groups, to check the validity and relevance of measures, biological samples and project plans, and ensure that questions are clear and logical to participants, and the workflow logic makes sense. They provide guidance and inform the research questions of importance to families, and are represented in the ORIGINS governance structure, including the Biobank Governance Committee, to ensure the community is fully engaged, informed and has the opportunity to contribute meaningfully to ORIGINS. Hence, ORIGINS utilises strategies to maximise the completeness, correctness and validity of the data, and to minimise measurement error and bias [48].

In terms of reliability, several standardised measures are used. At times, simpler items were used in place of full standardised measures, in order to prevent the study from becoming too long. Sub-projects are able to add their own measures to explore particular aspects in more detail should they wish to.

As ORIGINS data have been collected specifically as a research platform, rather than being administrative population-level data that are subsequently used for research, the data do not suffer from issues associated with some large population-level datasets, such as duplicated participants, lack of clarity as to whether each participant represents a real person, and massaging data to be used for research purposes [49]. That said, as noted above, data collected by ORIGINS is merged with other datasets not collected primarily for research purposes (e.g. the JHC Questionnaire and MNS data) and so may suffer from some of these issues. The ORIGINS Databank team carefully identify and remove duplicates and only include JHC and MNS data in datasets that match an ORIGINS participant, so there are no issues with duplicates or entries from participants who are not actual patients.

One issue to be aware of, however, is that while ORIGINS participants were rigorously followed-up, the rate of completion of particular measures varies, particularly on the self-completed ORIGINS Core Questionnaire. For ethical reasons, ORIGINS did not apply force-completion rules on any measures in the Core Questionnaire, so participants could decide whether or not to complete particular measures or items. This is particularly a problem given that all participants were exposed to measures in roughly the same order at each timepoint, and so measures that appear earlier in the survey have higher completion rates than those that appear later. Since then, the number of measures asked of participants has been reduced to encourage more thorough completion of all measures. In addition, ORIGINS has now launched a Smart App for mobile phones to allow participants to book in their own clinical appointments, as well as to provide automatic notifications when timepoints are due, providing links to their online questionnaires in the one central location. The antenatal rates of completion of the ORIGINS Core Questionnaire have been provided below; postnatal completion rates are not provided, as these timepoints are not yet complete.

In addition, some strategies to reduce measurement error were not utilised by ORIGINS. For example, ORIGINS did not enforce integrity constraints, and tended to allow text entry instead of selection of pre-set set options (e.g. from a drop-down list). This allowed different clinicians to complete data fields differently. For example, when completing months since an event, clinicians tended to represent weeks differently, such as 2.25 months, versus 2 months and 2 weeks or 2 months and 14 days. Regular databank team meetings were held to try to gain consistency across such issues. Another issue to

Table 9: Completion rate for active mothers antenatally for all standardised measures and the Joondalup Health Questionnaire, and comparison between completer (C) and non-completer (NC) active mothers (significant differences bolded)

Measure	% complete	Mean age (SD)		Mean IRSAD Decile (SD)		Caucasian N (%)	
		C	NC	C	NC	C	NC
Mediterranean Diet Index [42]	71.07	32.36 (4.53)	31.85 (4.94)	7.24 (1.78)	7.05 (1.89	2170 (83.7%)	758 (79.4%)
DASS-21 [26]	70.02	**32.38 (4.54)**	**31.82 (4.89)**	7.24 (1.78)	7.05 (1.89)	2142 (83.7%)	786 (79.4%)
International Physical Activity Questionnaire [37]	63.78	32.38 (4.52)	31.92 (4.87)	7.26 (1.77)	7.05 (1.88)	**1978 (84.9%)**	**950 (77.9%)**
Godin-Leisure Time [38]	63.56	**32.39 (4.52)**	**31.92 (4.86)**	**7.26 (1.77)**	**7.04 (1.88)**	1972 (84.9%)	956 (78.0%)
Pittsburgh Sleep Quality [43]	62.93	32.38 (4.52)	31.93 (4.85)	7.27 (1.77)	7.04 (1.88)	**1954 (84.9%)**	**974 (78.1%)**
Australian Eating Survey [44]	45.27	**32.49 (4.51)**	**32.00 (4.75)**	7.30 (1.78)	7.09 (1.83)	**1406 (86.0%)**	**1522 (79.5%)**
JHC Health Questionnaire	42.68	32.08 (4.63)	32.36 (4.65)	7.17 (1.84)	7.21 (1.78)	1419 (83.1%)	1509 (82.0%)
Connor-Davidson Resilience Scale – Short Form [25]	42.58	32.42 (4.49)	32.08 (4.76)	7.30 (1.76)	7.10 (1.84)	**1338 (86.7%)**	**1590 (79.3%)**
Stressful Life Events [31]	39.16	**32.50 (4.49)**	**32.03 (4.74)**	7.28 (1.75)	7.13 (1.85)	1236 (84.3%)	1692 (81.3%)
Perceived Social Support Scale [29, 30]	36.87	**32.56 (4.49)**	**32.02 (4.73)**	7.27 (1.75)	7.14 (1.85)	1159 (84.0%)	1769 (81.5%)
Maternal Attachment Scale [27]	36.84	**32.56 (4.48)**	**32.02 (4.74)**	7.26 (1.75)	7.14 (1.85)	1159 (84.0%)	1769 (81.5%)
The Mental Health Continuum [32]	28.51	32.60 (4.50)	32.07 (4.70)	7.21 (1.75)	7.18 (1.84)	855 (80.6%)	2073 (83.3%)
Neighbourhood Environment Walkability Scale [35]	9.30	32.68 (4.58)	32.18 (4.65)	7.22 (1.65)	7.18 (1.83)	286 (85.9%)	2642 (82.2%)

IRSAD = Index of Relative Social Disadvantage and Advantage.

be aware of in regard to the ORIGINS Core Questionnaire is that, in order to improve completion rates for non-birthing partner data, mothers completed the questionnaire on behalf of non-birthing partners in 44% of cases, which would be less accurate than if the non-birthing partners had completed the data themselves. As providers of these data, ORIGINS leaves it up to individual researchers how they choose to manage this.

Discussion

Over more than five years, ORIGINS has collected detailed information about the early environment's influence on a broad range of non-communicable diseases from almost 8,000 mother-child dyads from pregnancy through early childhood. The databank has several strengths, and some areas for improvement.

Strengths

Extensive, unique, multi-faceted, longitudinal data

The main strength of the ORIGINS data and biobank are the extent of the data collected and the novelty of much of this data. Since 2017, ORIGINS has collected over 300,000 biological samples and millions of datapoints to create an integrated data and biobank platform for researchers worldwide. ORIGINS has tracked mothers, non-birthing partners and children longitudinally to enable examination of multiple influences on child development from pre-conception, conception and early childhood across the various domains examined. This allows insights into the complex interactions between multifaceted exposures in the total lived environment (the 'exposome'), and biological influence at the microscale, measured via multiomic analysis, including genomic, metabolomic and microbiomic techniques. Such analyses have not been possible in previous birth cohorts [11].

Table 10: Completion rate for active participants antenatally for some standardised measures, and comparison between completers and non-completers for non-birthing partners (significant differences bolded)

Non-birthing partner	N	% complete	Mean age (SD)		Mean IRSAD Decile (SD)	
			C	NC	C	NC
Mediterranean Diet Index [42]	2652	66.13	34.24 (5.39)	33.86 (5.65)	7.26 (1.77)	7.04 (1.89)
DASS-21 [26]	2431	60.62	34.32 (5.34)	33.79 (5.68)	**7.27 (1.76)**	**7.04 (1.88)**
International Physical Activity Questionnaire [37]	2297	57.28	34.30 (5.40)	33.86 (5.59)	7.27 (1.78)	7.07 (1.85)
Godin-Leisure Time [38]	1046	26.08	34.33 (5.56)	34.04 (5.44)	7.21 (1.75)	7.18 (1.83)
The Mental Health Continuum [32]	1007	25.11	34.38 (5.43)	34.03 (5.49)	7.22 (1.75)	7.18 (1.83)
Stressful Life Events [31]	980	24.44	34.33 (5.55)	34.05 (5.45)	7.21 (1.75)	7.18 (1.83)
Connor-Davidson Resilience Scale Short Form [25]	331	8.25	34.59 (5.78)	34.08 (5.45)	7.25 (1.63)	7.18 (1.83)

IRSAD = Index of Relative Social Disadvantage and Advantage.

A further strength is that the extensive data collected by ORIGINS is supplemented by regularly merging with routine antenatal care data, as well as various other sources. In addition, as noted above, ORIGINS has utilised some recent and relatively novel data collection technologies not utilised by other birth cohorts, such as the BOD and PEA Pods, Nevisense [16], and the Veggie Meter. This allows novel insights into contemporary development.

ORIGINS is also unique in that it permits nesting of sub-projects to maximise the benefits of harmonised recruitment, observation and measurement. Moreover, ORIGINS allows sub-projects to test interventions, rather than just observing outcomes over time. Hence, ORIGINS provides a cost-effective way for researchers to conduct high-quality longitudinal observational and interventional studies using data already collected by ORIGINS, as well as adding their own. All data collected by ORIGINS sub-projects is returned to ORIGINS so other researchers can access these data. Further detail about this is provided below, when discussing how the ORIGINS data have been used.

Timeliness of data

A strength of the study is the timeliness of data availability. Often population datasets are several years old by the time they are cleaned, prepared and ready for public use [48]. By contrast, ORIGINS is able to clean and prepare data for use by researchers within months of its collection, so that the data are timely and relevant to researchers when they receive it.

Different levels of consent allow insight to bias

In addition, ORIGINS' different levels of consent provide a unique ability to compare those willing to actively participate in research (active participants) to those who are only willing to share their routinely collected data (non-active participants). ORIGINS also has ethics approval to access some data (JHC and MNS data) for antenatal hospital patients who did not enrol in ORIGINS. This allows ORIGINS to estimate bias inherent in their dataset by comparing these groups, such as that shown in Table 3.

Stakeholder consultation improves data quality

Finally, ORIGINS has engaged community stakeholders throughout the development of the research platform in order to ensure the research design reflects community's needs. Decision-making is shared with the community, and community feedback is utilised to check and then improve the research design. This has enhanced the design of the research in terms of acceptability and validity. In addition, ORIGINS has conducted several community events and has dedicated marketing personnel, so the project is well known to the community, particularly within the northern corridor of Perth where the Joondalup Health Campus is located. The ORIGINS Team actively follow up participants to reduce attrition, as well as ensuring the study provides tangible benefits to participants, such as extra paediatric check-ups and referrals. All of these strategies encourage maximum participation and retention, as well as enhancing the validity and quality of the data collected [50].

Limitations

Population-level databases may not contain the entire population

While The ORIGINS Project comprises over 50% of the population it aims to represent (i.e. those giving birth at JHC), and so can be regarded as a population-level cohort, a large proportion of the population did not consent to participate (47%). Younger and more affluent individuals are more likely to be included in digital data collections compared to older people, migrants, or those with a lower socio-economic status [51]. Although the recruitment of culturally and linguistically diverse (CALD) and Indigenous participants would have been beneficial, these particular groups weren't prioritised. The recruitment material and questionnaires are in English and no translation services are offered, therefore participants must be sufficiently fluent in English to be able to participate. ORIGINS is in a unique position of accessing some data for those not enrolled in the study, so can do some comparisons of enrolled versus non-enrolled participants. Regardless, the latter are not

included in the active cohort who complete the ORIGINS' Core Questionnaire.

Observational data affected by interventions

An additional potential issue is that data collected from some active participants is not purely observational, as is usually the case for population-level data, due to the impact of interventions with this group. Rather, true observational data are only available for non-active participants. This was a deliberate decision on behalf of ORIGINS as it was felt that the time for observation, or 'watch and wait' is over, and the opportunities to intervene in these crucial early years should not be squandered [11, 52]. Again, antenatal observational data are available for non-active participants, and for active participants not yet exposed to any interventions. Moreover, once ORIGINS links to other datasets, further observational data are available for non-active participants.

Missing data and attrition

As noted above, although ORIGINS attempts to complete missing data, there is inevitably missing data and attrition across the timepoints. Some strategies have been put in place to reduce this as much as possible, but this is unavoidable, and is a well-known issue in longitudinal studies [53]. As The ORIGINS Project itself does not analyse data, but rather, provides data to sub-projects, researchers need to examine the impact of the missing data on their particular research (for example, whether the data are missing completely at random or not).

Relatedly, as ORIGINS collects information from a range of sources, there will be different levels of completion of such data across these sources for each participant. A comparison has been provided of any differences on some key demographic characteristics between those who completed each standardised measure for completed timepoints (antenatal and birth) but not for non-completed timepoints, nor for non-ORIGINS data sources. However, researchers accessing merged datasets provided by ORIGINS can conduct such comparisons.

Another potential issue is that ORIGINS integrates data that are routinely collected at JHC by health professionals for administrative purposes and for the patient's health records. Hence, the primary aim of such data collection is not necessarily for research purposes, so the collected data may not meet the same quality criteria as data intentionally collected by ORIGINS, as discussed above.

How has the data been used so far?

To date, ORIGINS has nested 51 sub-projects. Some of these sub-projects have analysed existing data and/or samples, whereas others have also collected and returned additional data, while others still also implement interventions ($n = 17$).

Sub-projects are categorised into various themes. Examples of sub-projects are shown in tables below categorised into similar themes, covering Allergy, Inflammation and Immunity,

Table 11: ORIGINS sub-projects in the domains of Allergy, Inflammation and Immunity, Microbiome, and Oral Health

Themes and studies	Aim(s) and findings if available
Allergy, Inflammation and Immunity	
PrEggNut	To determine the effectiveness of higher maternal food allergen consumption during pregnancy and lactation on infant food allergy outcomes [54]
SYMBA	To reduce the risk of allergic disease in children by improving the balance of 'healthy bacteria' in the gut, using a high fibre prebiotic supplement in pregnancy and while breastfeeding [54].
AERIAL	To determine gene signature patterns in epithelial cells that may predict the development of wheeze, allergy and asthma later in childhood.
BENEFIT	To determine whether the amount of eggs and peanuts a mother eats during breastfeeding influences development of baby food allergy [55].
The Mast Cell Study	To compare how mast cells, part of the body's immune response, are 'programmed' in allergic and non-allergic children as they migrate through the body, as they may be a suitable target for new allergy drugs.
Cashew study	To pilot regular cashew nut spread intake by infants from 6 months to 1 year of age to determine dosage recommendations prior to a larger RCT [56].
Microbiome	
TUMS	To examine the effects of untreated tap versus filtered water on the development of the gut microbiome in infants via a randomised-controlled trial to see if exposure to chlorine, heavy metals and pesticides in tap water is safe for microorganisms that colonise the gut and if gut dysbiosis leads to chronic disease.
ADAPTs	To test whether exposure to antibiotics in the neonatal unit at JHC has long term health impacts, ADAPTS uses supplemental probiotics to promote the normal development of gut flora.
Oral Health	
The Dental Screening Study	To evaluate the feasibility of tele-dental screening for the identification of early childhood caries (ECC) in pre-schoolers using an app operated by their parents with remote review by oral-health therapists.

Microbiome, and Oral Health (shown in Table 11), Brain and Behaviour, Growth and Development, Nutrition and Metabolism (shown in Table 12), and Environment and Lifestyle, Mental Health and Wellbeing, and Parenting (shown in Table 13).

Future use of data

ORIGINS collects a wide array of data across several domains, which allows researchers several options in terms of future research. A key goal of ORIGINS is to identify underlying causes of chronic inflammation, immune dysfunction, metabolic dysregulation and dysbiosis that then lead to chronic non-communicable diseases, which research suggests may start early in life, including in-utero [3]. A comprehensive biobank and databank have been established to assist with this goal. Due to the wide array of data that ORIGINS has collected, studies can examine complex interactions between the 'exposome' and epigenetics, metabolomics, proteomics and the microbiome [11, 52].

ORIGINS is currently seeking funding to continue tracking participants into middle childhood, with the eventual aim of tracking even further. Hence, more unique opportunities are provided to link complex factors within antenatal and early childhood to middle primary and eventually adolescence using ORIGINS data and biosamples. Researchers wishing to join the project are encouraged to think about the later years when developing research ideas, as well as the early childhood years.

In addition, as ORIGINS allows interventions, not just observation, future sub-projects are likely to implement interventions now and then examine their impact on later outcomes. This is a unique opportunity provided by ORIGINS. Given the wide range of data and biosamples collected by ORIGINS, such interventions may be from any of the domains covered by ORIGINS. As ORIGINS tracks parents as well as children, sub-projects may examine interactions between parenting or family structure upon later outcomes. ORIGINS is also currently geocoding participant addresses, so the interaction of various measures with aspects of the local environment, such as the number of local parks, highways, schools and advertisements, and associated elements such as level of pollution, can also be examined.

Given that many mental health disorders have their origins in childhood [70], it may also be beneficial for future sub-projects to add additional measures in early childhood that allow prospective tracking of early indicators, to identify endophenotypes and biomarkers of later conditions, such as mental health issues like depression, anxiety, OCD and schizophrenia [71, 72]. This can greatly assist with untangling the complex aetiologies of these conditions, as well as allowing earlier diagnosis and intervention, thereby potentially improving later outcomes.

Given the increasingly widespread use of artificial intelligence (AI) within large datasets, and the ability of AI to detect patterns by analysing multiple relationships simultaneously, AI may provide previously undetected insights into disease progression, risk factors, and developmental trajectories. Hence, it is expected that ORIGINS will increasingly be used for AI and machine learning sub-projects in the future. AI algorithms can also group individuals with similar health profiles, potentially uncovering subpopulations with unique conditions or responses to interventions.

Machine learning models or simulations can also assist with this. Machine learning is being utilised in some new ORIGINS studies currently being onboarded, to identify predictors of asthma before the symptoms develop, to diagnose various diseases from facial images, and to assist with anatomical measurement of foetal ultrasounds.

Data linkage

ORIGINS aims to link with external State and Commonwealth datasets, such as Births, Deaths and Marriages, Medicare,

Table 12: ORIGINS sub-projects in the domains of Brain and Behaviour, Growth and Development, and Nutrition and Metabolism

Themes and studies	Aim(s) and findings if available
Brain and Behaviour	
Early Moves	To investigate whether a baby's early movements can predict learning difficulties later in childhood via short (3 minute) videos of babies at several timepoints using a smartphone app [67].
Baby AICS/CUBs	To test a new program designed to support baby brain development by providing parents/carers with information and skills to optimise 'back and forth' interactions.
TALK	To understand how testosterone exposure in the womb may relate to brain growth before birth and language development after birth.
Growth and Development	
Kindy Readiness	To assess the development and wellbeing of all non-active participants enrolled in The ORIGINS Project to enable early identification, timely feedback and early intervention to vulnerable children prior to commencing preschool, kindergarten and/or an alternate early learning environment.
School Readiness	To identify factors that influence successful development in the early years, in order to promote the health and wellbeing of the population.
Nutrition and Metabolism	
ACE Feeding	To determine whether breastfeeding outcomes can be improved by teaching pregnant women how to hand express colostrum using a novel online instructional video.
PLAN	To examine whether a lifestyle intervention in early pregnancy can reduce offspring adiposity [68, 69].

NAPLAN, the Australian Early Development Census (AEDC), the WA Register of Developmental Abnormalities, and the Australian Immunisation Register (AIR). This will allow researchers to link antenatal and early childhood health and development information to educational and cognitive outcomes, to demographic outcomes (via Births, Deaths and Marriages) and to link the in-depth data of ORIGINS to other health outcomes, such as the impact of being vaccinated.

Data linkage will also allow greater comparison of non-active and active participants. Comparing non-active to active participants will allow researchers to ascertain the benefits of active participation in ORIGINS, such as extra paediatric check-ups and referrals, albeit, with the limitation that participants were not randomised to active versus non-active conditions.

Data access

The ORIGINS research platform offers researchers the opportunity to access multiple longitudinal data sets and the ability to embed interventions and clinical trials in this cohort with existing infrastructure and resources, thereby maximising harmonisation of recruitment and data collection processes. Researchers can still add measures or additional sample collection to timepoints beyond one year as the children develop, and can conduct predictive analyses using antenatal and birth data.

Researchers can access databank metrics and real-time cohort, data and sample completion rates at various timepoints in the ORIGINS Data Catalogue, https://bitly.ws/34uGB. Researchers interested in accessing the ORIGINS cohort, database or biological samples must undergo a process of scientific committee and ethics review and approval. Once the project is approved, the researchers will need to pay cost recovery fees for access to the data, samples and/or cohort. An accurate quote will be provided during the application process based on the unique needs of the project. To begin the application process, contact origins.research@telethonkids.org.au or Jacqueline Davis (Co-Director) at Jackie.Davis@telethonkids.org.au. Further information can also be found in The ORIGINS Project Collaboration Policy available at https://originsproject.telethonkids.org.au/for-collaborators/useful-documents-collaborators/.

Conclusions

ORIGINS provides a comprehensive longitudinal collection of millions of data-points and over 400,000 samples collected both antenatally and postnatally from mothers, non-birthing partners, and the child, as well as some household and

Table 13: ORIGINS sub-projects in the domains of Environment and Lifestyle, Mental Health and Wellbeing, and Parenting

Themes and studies	Aim(s) and findings if available
Environment and Lifestyle	
Screen ORIGINS	To understand family screen technology use, particularly mobile touchscreen devices (i.e. tablet computers and smartphones), including what influences family screen technology use and potential implications for child health and development [57–60]
Nature, Play and Grow	To assess whether connecting families to nature has a positive influence on physical activity, diet, sleep and emotional well-being in young children to then inform a future randomised controlled trial (RCT) of the intervention.
Community Wellbeing	To investigate how ORIGINS families coped during the COVID-19 pandemic, and their experiences at this time [61–63].
PLANET	To evaluate the appropriateness of already-collected ORIGINS samples for plastic-related research, and to optimise prospective sample collections.
Mental Health and Wellbeing	
Mum's Minds Matter	To improve the mental-health of pregnant women and measure the effects of different types of home-based interventions (mindfulness-based training, self-compassion-based training, and a relaxation intervention) on maternal distress, self-compassion, mindfulness and emotion regulation [61, 64].
CARE Dads	To assess the health of expectant fathers by providing a health check-up, as a healthy Dad is an important part of a nurturing early environment [65, 66].
Parenting	
Happy Parenting in Childhood	To determine whether infant/toddler group classes and parent discussion sessions affect parent confidence and stress levels.
Flourishing in Fatherhood	To follow up with the expectant fathers from an earlier study (CARE-Dads) to assess mental and physical health of fathers, as well as the impact on well-being of partners and growth and developmental milestones in children.
Positive Family Foundations	To conduct a pilot evaluation of the Family Foundations project in Australia, with the addition of content on parental reflective functioning, with the aim of eventually becoming a randomised controlled trial.

environmental data (e.g. household dust, and information about each family's neighbourhood). While this in itself is relatively unique, ORIGINS also allows researchers to apply to actively implement interventions or add particular measures, thereby nesting their projects within the main ORIGINS cohort, and making maximum use of harmonised recruitment and data and sample collection processes in order to implement their own longitudinal interventions to derive causational conclusions across several domains. ORIGINS is planning to extend the platform into middle childhood and is currently applying to link existing data with other government-collected data, allowing the data to be even more comprehensive. ORIGINS implements strategies to enhance completeness, accuracy and validity, and to minimise bias and measurement error, as well as engaging community stakeholders, actively following up participants over time, and preparing data rapidly for researchers to maximise data quality. Overall, the ORIGINS databank is a world-class data platform that provides a rare opportunity for researchers to not only access complex data with biosamples from mother, non-birthing partner and child antenatally and into childhood, it also provides researchers the opportunity to embed their own measures and conduct interventions.

Ethics approval

Ramsay SA/WA HREC.

Author contributions

BD and SW prepared the manuscript with input from EF. WB performed the data analysis. All authors reviewed the manuscript prior to publication.

Funding

ORIGINS has received core funding support from the Telethon Perth Children's Hospital Research Fund, Joondalup Health Campus, the Paul Ramsay Foundation and the Commonwealth Government of Australia through the Channel 7 Telethon Trust. ORIGINS has received core funding support from several government and philanthropic organisations, as well as substantial in-kind support from many collaborators.

Acknowledgements

Telethon Kids Institute and Joondalup Health Campus

Conflict of Interest

None to report.

Publication consent

The authors consent to publish this paper. Data relevant to this paper are available via the ORIGINS Data Catalogue mentioned in the paper.

Data availability

Data are available from ORIGINS via an application process, including Scientific Committee and ORIGINS Director approval, Ramsay Human Research Ethics approval, and payment of an access fee. This application process is in place as the data collected are sensitive health data related to pregnancy, birth and children.

References

1. Barker DJ. The fetal and infant origins of adult disease. Bmj. 1990;301(6761):1111. https://doi.org/10.1136/bmj.301.6761.1111. Available from: https://www.ncbi.nlm.nih.gov/pubmed/2252919.

2. Barker DJ, Osmond C. Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales. Lancet. 1986;1(8489):1077-81. https://doi.org/10.1016/s0140-6736(86)91340-1 Available from: https://www.ncbi.nlm.nih.gov/pubmed/2871345.

3. Gluckman PD, Hanson MA, Cooper C, Thornburg KL. Effect of in utero and early-life conditions on adult health and disease. N Engl J Med. 2008;359(1)61-73. https://doi.org/10.1056/NEJMra0708473. Available from: https://www.ncbi.nlm.nih.gov/pubmed/18596274.

4. Silva DT, Lehmann D, Tennant MT, Jacoby P, Wright H, Stanley FJ. Effect of swimming pools on antibiotic use and clinic attendance for infections in two Aboriginal communities in Western Australia. Med J Aust. 2008;188(10)594-8. https://doi.org/10.5694/j.1326-5377.2008.tb01800.x. Available from: https://www.ncbi.nlm.nih.gov/pubmed/18484935.

5. Muka T, Imo D, Jaspers L, Colpani V, Chaker L, van der Lee SJ, et al. The global impact of non-communicable diseases on healthcare spending and national income: a systematic review. Eur J Epidemiol. 2015;30(4)251-77. https://doi.org/10.1007/s10654-014-9984-2. Available from: https://www.ncbi.nlm.nih.gov/pubmed/25595318.

6. Magnusson R, Patterson D. Global action, but national results: strengthening pathways towards better health outcomes for non-communicable diseases. Critical Public Health. 2019;31(4)464-76. https://doi.org/10.1080/09581596.2019.1693029

7. Logan AC, Jacka FN, Prescott SL. Immune-Microbiota Interactions: Dysbiosis as a Global Health Issue. Curr Allergy Asthma Rep. 2016;16(2)13. https://doi.org/10.1007/s11882-015-0590-5. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26768621.

8. Prescott SL, Wegienka G, Logan AC, Katz DL. Dysbiotic drift and biopsychosocial medicine: how the microbiome links personal, public and planetary health. Biopsychosoc Med. 2018;12(7)7. https://doi.org/10.1186/s13030-018-0126-z. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29743938.

9. Renz H, Holt PG, Inouye M, Logan AC, Prescott SL, Sly PD. An exposome perspective: Early-life events and immune development in a changing world. J Allergy Clin Immunol. 2017;140(1)24-40. https://doi.org/10.1016/j.jaci.2017.05.015. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28673401.

10. Logan AC, Prescott SL, Haahtela T, Katz DL. The importance of the exposome and allostatic load in the planetary health paradigm. J Physiol Anthropol. 2018;37(1)15. https://doi.org/10.1186/s40101-018-0176-8. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29866162.

11. Silva DT, Hagemann E, Davis JA, Gibson LY, Srinivasjois R, Palmer DJ, et al. Introducing the ORIGINS project: a community-based interventional birth cohort. Reviews on environmental health. 2020;35(3)281-93. https://doi.org/10.1515/reveh-2020-0057. Available from: https://www.ncbi.nlm.nih.gov/pubmed/32853171.

12. Region summary: Joondalup, Region code 54170 [Internet]. Australian Bureau of Statistics. 2023 [cited 18 April 2024]. Available from: https://abs.gov.au/census/find-census-data/quickstats/2021/LGA54170.

13. Vasireddy S, Berrington A, Kuang B, Kulu H. Education and fertility in Europe in the last decade: A review of the literature. . Comparative Population Studies. 2023;48:553-88. https://doi.org/10.12765/CPoS-2023-21

14. Ball K, Crawford D. Socioeconomic status and weight change in adults: A review. . Social Science & Medicine 2005;60(9)1987-2010. https://doi.org/10.1016/j.socscimed.2004.08.056

15. Sedgwick P. Clinical significance versus statistical significance. British Medical Journal. 2014;348:1-2. https://doi.org/10.1136/bmj.g2130

16. Mohr P, Birgersson U, Berking C, Henderson C, Trefzer U, Kemeny L, et al. Electrical impedance spectroscopy as a potential adjunct diagnostic tool for cutaneous melanoma. Skin Res Technol. 2013;19(2)75-83. https://doi.org/10.1111/srt.12008. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23350668.

17. Bailey BW, LeCheminant G, Hope T, Bell M, Tucker LA. A comparison of the agreement, internal consistency, and 2-day test stability of the InBody 720, GE iDXA, and BOD POD®gold standard for assessing body composition. Measurement in Physical Education and Exercise Science. 2018;22(3)231-8. https://doi.org/10.1080/1091367x.2017.1422129

18. Obana A, Asaoka R, Takayanagi Y, Gohto Y. Inter-device concordance of Veggie Meter-A reflection spectroscopy to measure skin carotenoids. J Biophotonics. 2023;16(8)e202300071. https://doi.org/10.1002/jbio.202300071. Available from: https://www.ncbi.nlm.nih.gov/pubmed/37072378.

19. D'Vaz N, Kidd C, Miller S, Amin M, Davis JA, Talati Z, et al. The ORIGINS Project Biobank: A Collaborative Bio Resource for Investigating the Developmental Origins of Health and Disease. Int J Environ Res Public Health. 2023;20(13)6297. https://doi.org/10.3390/ijerph20136297. Available from: https://www.ncbi.nlm.nih.gov/pubmed/37444144.

20. Collins C, Watson J, Burrows T, Guest M, Pezdirc K. Validation of an Adult Food Frequency Questionnaire and Development of a Diet Quality Score for Children and Adults: University of Newcastle; 2011.

21. Collins CE, Burrows TL, Truby H, Morgan PJ, Wright IMR, Davies PSW, et al. Comparison of energy intake in toddlers assessed by food frequency questionnaire and total energy expenditure measured by the doubly labeled water method. J Acad Nutr Diet. 2013;113(3)459-63. https://doi.org/10.1016/j.jand.2012.09.021. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23317500.

22. Zheng M, Campbell KJ, Scanlan E, McNaughton SA. Development and evaluation of a food frequency questionnaire for use among young children. PLoS One. 2020;15(3)e0230669. https://doi.org/10.1371/journal.pone.0230669. Available from: https://www.ncbi.nlm.nih.gov/pubmed/32210467.

23. Goodman A, Lamping DL, Ploubidis GB. When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): Data from British parents, teachers and children. . J Abnorm Child Psychol. 2010;38:1179-91. https://doi.org/10.1007/s10802-010-9434-x

24. Conners CK. Conners Early Childhood™. 2009.

25. Connor KM, Davidson JR. Development of a new resilience scale: the Connor-Davidson Resilience Scale (CD-RISC). Depress Anxiety. 2003;18(2)76-82. https://doi.org/10.1002/da.10113. Available from: https://www.ncbi.nlm.nih.gov/pubmed/12964174.

26. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther. 1995;33(3)335-43. https://doi.org/10.1016/0005-7967(94)00075-u. Available from: https://www.ncbi.nlm.nih.gov/pubmed/7726811.

27. Condon JT. The assessment of antenatal emotional attachment: development of a questionnaire instrument. Br J Med Psychol. 1993;66 (Pt 2)(2)167-83. https://doi.org/10.1111/j.2044-8341.1993.tb01739.x. Available from: https://www.ncbi.nlm.nih.gov/pubmed/8353110.

28. Condon JT, Corkindale CJ, Boyce P. Assessment of postnatal paternal–infant attachment: development of a questionnaire instrument. Journal of Reproductive and Infant Psychology. 2008;26(3)195-210. https://doi.org/10.1080/02646830701691335.

29. Zimet GD, Dahlem NW, Zimet SG, Farley GK. The Multidimensional Scale of Perceived Social Support. Journal of Personality Assessment. 1988;52(1)30-41. https://doi.org/10.1207/s15327752jpa5201_2.

30. Zimet GD, Powell SS, Farley GK, Werkman S, Berkoff KA. Psychometric characteristics of the Multidimensional Scale of Perceived Social Support. J Pers Assess. 1990;55(3-4):610-7. https://doi.org/10.1080/00223891.1990.9674095. Available from: https://www.ncbi.nlm.nih.gov/pubmed/2280326.

31. Holmes TH, Rahe RH. The Social Readjustment Rating Scale. J Psychosom Res. 1967;11(2)213-8. https://doi.org/10.1016/0022-3999(67)90010-4. Available from: https://www.ncbi.nlm.nih.gov/pubmed/6059863.

32. Keyes CL. Brief description of the Mental Health Continuum Short Form (MHC-SF). 2009 [Available from: http://www.sociology.emory.edu/ckeyes/.

33. Snyder CR, Sympson SC, Ybasco FC, Borders TF, Babyak MA, Higgins RL. Development and validation of the State Hope Scale. J Pers Soc Psychol. 1996;70(2)321-35. https://doi.org/10.1037//0022-3514.70.2.321. Available from: https://www.ncbi.nlm.nih.gov/pubmed/8636885.

34. Buckner JC. The development of an instrument to measure neighborhood cohesion. Am J Commun Psychol. 1988;16(6)771-91. https://doi.org/10.1007/bf00930892.

35. Saelens BE, Sallis JF, Black JB, Chen D. Neighborhood-based differences in physical activity: an environment scale evaluation. Am J Public Health. 2003;93(9)1552-8. https://doi.org/10.2105/ajph.93.9.1552. Available from: https://www.ncbi.nlm.nih.gov/pubmed/12948979.

36. 2011 AIoHaW. National Drug Strategy Household Survey report. Canberra: AIHW; 2010 Contract No.: Cat. no. PHE 145.

37. Craig CL, Marshall AL, Sjostrom M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc. 2003;35(8)1381-95. https://doi.org/10.1249/01.MSS.0000078924.61453.FB. Available from: https://www.ncbi.nlm.nih.gov/pubmed/12900694.

38. Godin G, Shephard RJ. Godin Leisure-Time Exercise Questionnaire. Medicine and Science in Sports and Exercise A Collection of Physical Activity Questionnaires for Health-Related Research. 2015;29(6 Suppl):36-8. https://doi.org/10.1037/t31334-000

39. Stolk E, Ludwig K, Rand K, van Hout B, Ramos-Goni JM. Overview, Update, and Lessons Learned From the International EQ-5D-5L Valuation Work: Version 2 of the EQ-5D-5L Valuation Protocol. Value Health. 2019;22(1)23-30. https://doi.org/10.1016/j.jval.2018.05.010. Available from: https://www.ncbi.nlm.nih.gov/pubmed/30661630.

40. Nisbet EK, Zelenski JM. The NR-6: a new brief measure of nature relatedness. Front Psychol. 2013;4:813. https://doi.org/10.3389/fpsyg.2013.00813. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24198806.

41. Sobko T, Jia Z, Brown G. Measuring connectedness to nature in preschool children in an urban setting and its relation to psychological functioning. PLoS One. 2018;13(11)e0207057. https://doi.org/10.1371/journal.pone.0207057. Available from: https://www.ncbi.nlm.nih.gov/pubmed/30496300.

42. Schroder H, Fito M, Estruch R, Martinez-Gonzalez MA, Corella D, Salas-Salvado J, et al. A short screener is valid for assessing Mediterranean diet adherence among older Spanish men and women. J Nutr. 2011;141(6)1140-5. https://doi.org/10.3945/jn.110.135566. Available from: https://www.ncbi.nlm.nih.gov/pubmed/21508208.

43. 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(2)193-213. https://doi.org/10.1016/0165-1781(89)90047-4 Available from: https://www.ncbi.nlm.nih.gov/pubmed/2748771.

44. Collins CE, Boggess MM, Watson JF, Guest M, Duncanson K, Pezdirc K, et al. Reproducibility and comparative validity of a food frequency questionnaire for Australian adults. Clin Nutr. 2014;33(5)906-14. https://doi.org/10.1016/j.clnu.2013.09.015. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24144913.

45. Silva-del Valle MA, Sánchez-Villegas A, Serra-Majem L. Association between the adherence to the Mediterranean diet and overweight and obesity in pregnant women in Gran Canaria. Nutricion Hospitalaria. 2013;28(3)654-9.

46. Sadeh A. A brief screening questionnaire for infant sleep problems: validation and findings for an Internet sample. Pediatrics. 2004;113(6)e570-7. https://doi.org/10.1542/peds.113.6.e570. Available from: https://www.ncbi.nlm.nih.gov/pubmed/15173539.

47. Squires J, Twombly E, Bricker D, Potter L. ASQ-3™ User's Guide. Baltimore, MD: Brookes Publishing; 2009.

48. Smith M, Lix LM, Azimaee M, Enns JE, Orr J, Hong S, et al. Assessing the quality of administrative data for research: a framework from the Manitoba Centre for Health Policy. J Am Med Inform Assoc. 2018;25(3)224-9. https://doi.org/10.1093/jamia/ocx078. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29025002.

49. Christen P, Schnell R. Thirty-three myths and misconceptions about population data: From data capture and processing to linkage. International Journal of Population Data Science. 2023;8(1).

50. Hotze T. Identifying the challenges in community-based participatory research collaboration. Virtual Mentor. 2011;13(2)105-8. https://doi.org/

10.1001/virtualmentor.2011.13.2.jdsc2-1102. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23121849.

51. Van Dijk J. The Digital Divide. Cambridge, UK: John Wiley and Sons; 2020.

52. Hagemann E, Colvin LJ, Gibson L, Miller S, Palmer D, Srinivas Jois R, et al. The ORIGINS Project. 2019. In: Pre-emptive Medicine: Public Health Aspects of Developmental Origins of Health and Disease [Internet]. Singapore: Springer Current Topics in Environmental Health and Preventive Medicine, https://doi.org/10.1007/978-981-13-2194-8_6.

53. Gustavson K, von Soest T, Karevold E, Roysamb E. Attrition and generalizability in longitudinal studies: findings from a 15-year population-based study and a Monte Carlo simulation study. BMC Public Health. 2012;12:918. https://doi.org/10.1186/1471-2458-12-918. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23107281.

54. Palmer DJ, Sullivan TR, Campbell DE, Nanan R, Gold MS, Hsu PS, et al. PrEggNut Study: protocol for a randomised controlled trial investigating the effect of a maternal diet rich in eggs and peanuts from <23 weeks' gestation during pregnancy to 4 months' lactation on infant IgE-mediated egg and peanut allergy outcomes. BMJ Open. 2022;12(6).

55. Palmer DJ, Silva DT, Prescott SL. Maternal peanut and egg consumption during breastfeeding randomized pilot trial. Pediatr Allergy Immunol. 2022;33(9)e13845. https://doi.org/10.1111/pai.13845. Available from: https://www.ncbi.nlm.nih.gov/pubmed/36156821.

56. Palmer DJ, Silva DT, Prescott SL. Feasibility and safety of introducing cashew nut spread in infant diets-A randomized trial. Pediatr Allergy Immunol. 2023;34(6)e13969. https://doi.org/10.1111/pai.13969. Available from: https://www.ncbi.nlm.nih.gov/pubmed/37366200.

57. Hood R, Zabatiero J, Silva D, S RZ, Straker L. 'There's good and bad': parent perspectives on the influence of mobile touch screen device use on prenatal attachment. Ergonomics. 2022;65(12)1593-608. https://doi.org/10.1080/00140139.2022.2041734. Available from: https://www.ncbi.nlm.nih.gov/pubmed/35164662.

58. Hood R, Zabatiero J, Silva D, Zubrick SR, Straker L. "Coronavirus Changed the Rules on Everything": Parent Perspectives on How the COVID-19 Pandemic Influenced Family Routines, Relationships and Technology Use in Families with Infants. Int J Environ Res Public Health. 2021;18(23). https://doi.org/10.3390/ijerph182312865. Available from: https://www.ncbi.nlm.nih.gov/pubmed/34886591.

59. Hood R, Zabatiero J, Zubrick SR, Silva D, Straker L. The association of mobile touch screen device use with parent-child attachment: a systematic review. Ergonomics. 2021;64(12)1606-22. https://doi.org/10.1080/00140139.2021.1948617. Available from: https://www.ncbi.nlm.nih.gov/pubmed/34190030.

60. Hood R, Zabatiero J, Silva D, Zubrick SR, Straker L. "It helps and it doesn't help": maternal perspectives on how the use of smartphones and tablet computers influences parent-infant attachment. Ergonomics. 2023:1-20. https://doi.org/10.1080/00140139.2023.2212148. Available from: https://www.ncbi.nlm.nih.gov/pubmed/37154796.

61. Davis JA, Gibson LY, Bear NL, Finlay-Jones AL, Ohan JL, Silva DT, et al. Can Positive Mindsets Be Protective Against Stress and Isolation Experienced during the COVID-19 Pandemic? A Mixed Methods Approach to Understanding Emotional Health and Wellbeing Needs of Perinatal Women. Int J Environ Res Public Health. 2021;18(13). https://doi.org/10.3390/ijerph18136958. Available from: https://www.ncbi.nlm.nih.gov/pubmed/34209693.

62. Gibson LY, Lockyer B, Dickerson J, Endacott C, Bridges S, McEachan RRC, et al. Comparison of Experiences in Two Birth Cohorts Comprising Young Families with Children under Four Years during the Initial COVID-19 Lockdown in Australia and the UK: A Qualitative Study. Int J Environ Res Public Health. 2021;18(17). https://doi.org/10.3390/ijerph18179119. Available from: https://www.ncbi.nlm.nih.gov/pubmed/34501709.

63. Sakalidis VS, Rea A, Perrella SL, McEachran J, Collis G, Miraudo J, et al. Wellbeing of Breastfeeding Women in Australia and New Zealand during the COVID-19 Pandemic: A Cross-Sectional Study. Nutrients. 2021;13(6)1831-46. https://doi.org/10.3390/nu13061831. Available from: https://www.ncbi.nlm.nih.gov/pubmed/34072039.

64. Davis JA, Ohan JL, Gregory S, Kottampally K, Silva D, Prescott SL, et al. Perinatal Women's Perspectives of, and Engagement in, Digital Emotional Well-Being Training: Mixed Methods Study. J Med Internet Res. 2023;25(1)e46852. https://doi.org/10.2196/46852. Available from: https://www.ncbi.nlm.nih.gov/pubmed/37847537.

65. Pettigrew S, Jongenelis MI, Cronin S, Dana LM, Silva D, Prescott SL, et al. Health-related behaviours and weight status of expectant fathers. Aust N Z J Public Health. 2022;46(3)275-80. https://doi.org/10.1111/1753-6405.13216. Available from: https://www.ncbi.nlm.nih.gov/pubmed/35357737.

66. Hadlow NC, Brown SJ, Lim EM, Prentice D, Pettigrew S, Cronin SL, et al. Anti-Müllerian hormone concentration is associated with central adiposity and reproductive hormones in expectant fathers. Clin Endocrinol (Oxf). 2022;97(5)634-42.

67. Elliott C, Alexander C, Salt A, Spittle AJ, Boyd RN, Badawi N, et al. Early Moves: a protocol for

a population-based prospective cohort study to establish general movements as an early biomarker of cognitive impairment in infants. BMJ Open. 2021;11(4)e041695. https://doi.org/10.1136/bmjopen-2020-041695. Available from: https://www.ncbi.nlm.nih.gov/pubmed/33837094.

68. Huang RC, Burke V, Newnham JP, Stanley FJ, Kendall GE, Landau LI, et al. Perinatal and childhood origins of cardiovascular disease. Int J Obes (Lond). 2007;31(2)236-44. https://doi.org/10.1038/sj.ijo.0803394. Available from: https://www.ncbi.nlm.nih.gov/pubmed/16718281.

69. Willcox JC, Chai D, Beilin LJ, Prescott SL, Silva D, Neppe C, et al. Evaluating Engagement in a Digital and Dietetic Intervention Promoting Healthy Weight Gain in Pregnancy: Mixed Methods Study. J Med Internet Res. 2020;22(6)e17845. https://doi.org/10.2196/17845. Available from: https://www.ncbi.nlm.nih.gov/pubmed/32442153.

70. Fryers T, Brugha T. Childhood determinants of adult psychiatric disorder. Clin Pract Epidemiol Ment Health. 2013;9:1-50. https://doi.org/10.2174/1745017901309010001. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23539489.

71. Geller DA, Homayoun S, Johnson G. Developmental Considerations in Obsessive Compulsive Disorder: Comparing Pediatric and Adult-Onset Cases. Front Psychiatry. 2021;12:678538. https://doi.org/10.3389/fpsyt.2021.678538. Available from: https://www.ncbi.nlm.nih.gov/pubmed/34248714.

72. Kumperscak HG. Childhood and adolescent schizophrenia and other early-onset psychoses. Psychiatric Disorders: Trends and Developments. 2011;131.

Supplementary Appendix

A full list of measures by for domain can be seen in the following tables.

Supplementary Table 1: Growth and Development measures collected from mothers (M) and children (C) for each timepoint

Measures	36 weeks antenatal	Postnatal (months)					Childhood (years)				
		Birth	2	4	6	9	1	2	3	4	5
Ages & Stages Questionnaire (S)				C		C	C	C	C	C	C
Infant and child sleep		M#	M#, C*	M#	M#, C*	M#	M#,C*	M#			M#, C
Breastfeeding, formula, solids	M		C		C		C				

(S) = standardised measure; *Brief-Infant Sleep Questionnaire (S), #Pittsburgh Sleep Index (S).

Supplementary Table 2: Medical, Biological and Genetic measures collected from mothers and non-birthing partners by timepoint (M = mother, C = child)

Measures	Antenatal (weeks' gestation)		Postnatal (months)		Childhood (years)				
	20	36	2	6	1	2	3	4	5
Paediatric appointments (developmental review and assessment), Skin-Prick Test (S), TEWL/Nevisense, anthropometry					C		C		C
Respiratory conditions					C	C			
Allergic reactions, bronchiolitis			C	C	C				
Asthma history					C	C	C	C	
Eczema				C	C	C	C	C	
Hay fever and seasonal allergies					C		C		C
Cough, cold, flu, fever, medical conditions				C	C	C			C
Child's health			C	C	C	C			
Oral health		M		C	C	C			C
Continence	M		M	M	M	M	M	M	M
COVID diagnosis and vaccinations	M	M	M	M	M	M	M	M	M
COVID impact									M
Musculoskeletal pain	M	M		M	M	M	M	M	M
Prescription medication use	M	M	M	M	M	M	M	M	M
Vaccinations during pregnancy		M							
Daytime naps			M	M	M	M			M
Eye, Hair and Skin Colour (S)	M, P								
Grandparent medical history	M, P								

(S) = standardised measure.

Supplementary Table 3: Lifestyle, Environment and Nutrition measures collected from mothers (M), non-birthing partners (P) and children (C) at each timepoint

Measures	Antenatal (gestation weeks)		Postnatal (months)		Childhood (years)				
	20	36	2	6	1	2	3	4	5
Alcohol Consumption (S)	M, P	M, P	M, P	M, P	M, P	M, P		M, P	M, P
Smoking (S)	M, P								
Illicit Drug Use (S)	M, P		M	M	M				M
Physical Activity In The Last 7 Days (IPAQ)	P	M, P		M, P	M, P				
Godin Leisure Time (S)	M, P	M, P		M, P	M, P	M, P	M		M, P
Health Related Quality of Life (EQ-5D-5L)		M		M		M			M
Weight perception		M							
Connectedness to Nature (S)					C				C
Engagement in nature									C
Nature Play WA (S)					M		M		
Nature Relatedness (S)	M, P				M, P	M, P			M, P
Time in the sun	M, P	M, P	C	C	C	C	C	C	C
Australian Eating Survey	M, P	M		M	M				
Mediterranean Diet	M, P	M	M, P	M, P	M, P	M, P	M, P	M, P	M, P
Food Frequency Questionnaire (S)					C	C	C		C
Gestational weight gain		M							
Drinking water		M			M	M			M
Childcare	M	M	C	C	C	C	C	C	C
Playgroup				C	C	C	C	C	
Child's play time					C	C	C	C	C
Physical activity	M, P								C
Child's technology use				C	C	C	C	C	C
Internet use	P							M	M

(S) = standardised measure.

Supplementary Table 4: Lifestyle and Environment household measures collected from mothers (M) at each timepoint

Measures	Antenatal (gestation weeks)		Postnatal (months)		Childhood (years)				
	20	36	2	6	1	2	3	4	5
Food Insecurity Screener (S)									M
Drinking water		M			M	M			M
Cooking	M	M		M	M				
Buckner's Neighbourhood Cohesion	M					M			M
Neighbourhood Environment Walkability Scale (S)	M					M			M
Cooking fuel, housing, heating and cooling	M								
Flooring			M						
Refuelling a vehicle, renovations, animals		M							
Electronic devices in the household	M	M		M	M	M	M	M	M
Hobbies and other activities around the home		M							

(S) = standardised measure.

Supplementary Table 5: Biopsychosocial and Cognitive measures collected from mothers (M), non-birthing partners (P) and children (C) for each timepoint

Measures	Antenatal (gestation weeks)		Postnatal (months)		Childhood (years)				
	20	**36**	**2**	**6**	**1**	**2**	**3**	**4**	**5**
Strengths and Difficulties questionnaire (S)							C		C
Conners Early Childhood assessment (S)							C		C
Climate change									M
Connor-Davidson Resilience Scale Short Form (S)	M, P					M, P			M, P
DASS-21 (S)	M, P	M, P		M, P	M, P	M, P		M, P	M, P
Edinburgh Postnatal Depression Scale (S)			M, P						
Attachment Scale (S)		M		M, P	M, P				
Parenting Sense of Competence Scale (S)					M, P				
Perceived Social Support Scale (S)		M			M, P	M, P			P
Stressful Life Events (S)		M, P			M, P	M, P			
Mental Health Continuum (S)	M, P			M, P			M, P		M, P
Trait-Hope Scale (S)							M		M
Flourishing Child									C
McMaster Family Assessment Device (S)									M

(S) = standardised measure.