

**A RICH AND OPEN DATASET THAT ENABLES RESEARCH ON A VARIETY OF QUESTIONS,
EVEN BEYOND THE FIELD OF BEHAVIORAL GENETICS.**

WHAT IS TWINLIFE?

The German TwinLife study is a longitudinal twin-family design study that collects data from about 4,000 families. It covers a vast number of multidisciplinary topics (education, personality, employment, social integration, psychological and physical health, deviant and delinquent behavior, COVID-related developments) in order to discover how inter-individual differences and social inequality arise. In addition to using the data for twin analyses, the wide range of topics in TwinLife offers a variety of possibilities for other analyses, e.g., in the fields of sociology, psychology, political science, or public health. The data of the study is available to the research community **free of charge**.

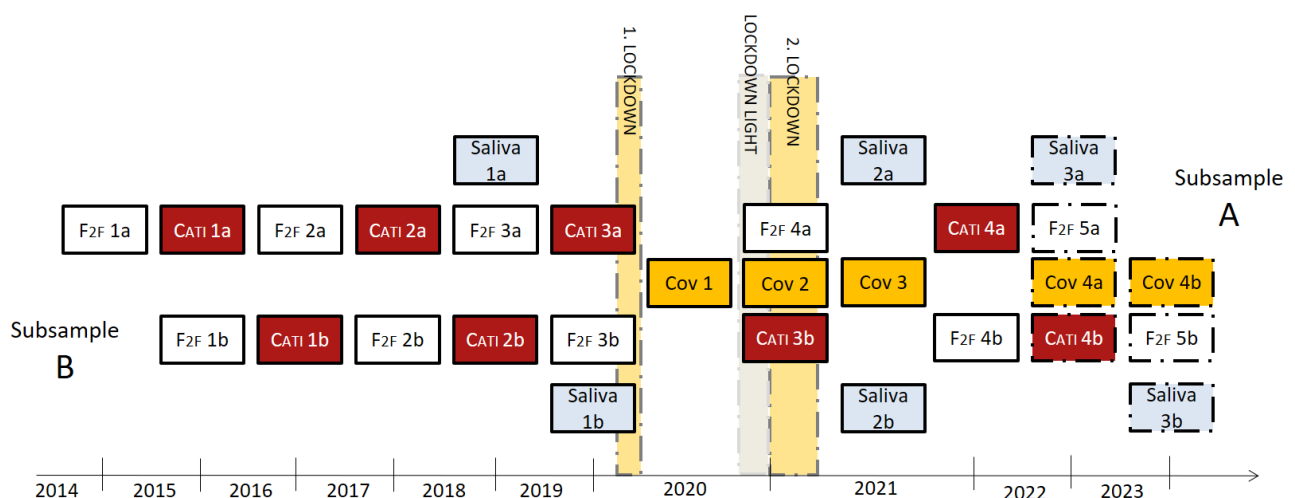
STUDY DESIGN

Data collection began in 2014 with a population-based sample of **4,097 twin families**. The cross-sequential survey design contains **four twin birth cohorts** with ~1,000 same-sex (both monozygotic and dizygotic) twin pairs in each cohort. **Face-to-face interviews (F2F)** within the households take place every other year, and **telephone interviews (CATI)** are conducted in the consecutive years (see figure below). In the face-to-face interviews, data was collected using a **mixed-mode design**: Participants were surveyed by an interviewer (computer-assisted personal interview; CAPI), by means of questionnaires on a tablet or laptop (computer-assisted self-interview; CASI), via a paper-and-pencil interview (PAPI), and/or via online questionnaires (computer-assisted web interview; CAWI). The survey was supplemented with **additional COVID-related questionnaires (COV)** and multiple saliva samples.

	Year of birth	Starting age
Cohort 1	2009/10	~5
Cohort 2	2003/04	~11
Cohort 3	1997/98	~17
Cohort 4	1990-93	~23

FOR WHICH PERSONS COULD THE TWINLIFE DATA BE RELEVANT?

Scientists from a broad disciplinary range (sociology, psychology, economics, political science, medicine) **interested in using openly available datasets** and performing secondary analyses. Students searching for data to use for their thesis are also welcome.



A RICH AND OPEN DATASET THAT ENABLES RESEARCH ON A VARIETY OF QUESTIONS, EVEN BEYOND THE FIELD OF BEHAVIORAL GENETICS.

WHAT KIND OF DATA IS AVAILABLE?

So far, **five panel data collections** and **two COVID-related data collections** have been made available to the scientific community, covering a broad range of information of up to **16,000 individuals** (including parents of the twins, siblings and (new) partners of the twins). Here is an excerpt of the data available:

- 01 Skill formation and education**
E.g., photos of school (leaving) certificates, assessment of cognitive abilities (via CFT), motivational aspects, school context and feeling of burden
- 06 Quality of life**
E.g., global and domain-specific life satisfaction, experiences of bullying
- 02 Personality and other traits**
E.g., big Five personality, self-esteem, self-efficacy, locus of control, self-regulation, narcissism
- 07 Deviant & problematic behavior**
E.g., internalizing, externalizing, delinquent behavior
- 03 Employment**
E.g., detailed information on employment, assets and income, commuting, shift-work
- 08 COVID-related information**
E.g., health during the pandemic, economic changes, behavioral changes, coping/resilience, threat, burden and stress during the pandemic
- 04 Social and political integration**
E.g., political attitudes, participation in political events, social networks and engagement in social associations, discrimination experiences.
- 09 Molecular genetic and epigenetic data**
E.g., polygenic and methylation-based risk scores (these are currently being collected and evaluated in separate satellite projects, and will be shared with the scientific community after the projects have ended)
- 05 Psychological and physical health**
E.g., height/weight, medical diagnoses, depressive-ness, impairment caused by diseases

FOR WHICH ANALYSES CAN THE DATA BE USED?

Here, you find a short overview on models that can be calculated using the TwinLife data. Also, you find some fields of research that could benefit from using the TwinLife data.

Twin studies

E.g., ACE decompositions, bivariate ACE models, GxE interaction models

Family and partnership research

E.g., extended twin family design, assortative mating

Longitudinal modelling

E.g., cross-lagged panel investigations, latent growth models, life course research

Educational research

E.g., coherence of educational attainment, grades of specific school subjects, connections with cognitive abilities or the social environment, educational inequalities

Sociological research

E.g., social inequality, distribution of resources, transgenerational transmission of values and resources

Psychological research

E.g., developmental psychology and changes across the life span, person-situation interactions, development of psychological traits and diseases

Public health studies

E.g., co-twin control models, origins of discordances in family members

Survey methodology

E.g., survey mode effects, construction of scales and assessment of their quality

Criminology and victimization

E.g., persistence and development of delinquent behavior, antecedents of victimization

**A RICH AND OPEN DATASET THAT ENABLES RESEARCH ON A VARIETY OF QUESTIONS,
EVEN BEYOND THE FIELD OF BEHAVIORAL GENETICS.**

WHERE DO I FIND MORE INFORMATION?

For more information on the TwinLife study, please visit our documentation website: <https://www.twin-life.de/documentation/>

HOW DO I GET STARTED?

Design

An overview of the theoretical and empirical background, the study design and the content as well as the implementation of the study can be found in the [TwinLife reference paper \(Hahn et al., 2016\)](#).

Questionnaires

The questionnaire files include the implemented questionnaires. These files are only available in the original study language (German). Original questionnaires are available on the [download section](#) on the TwinLife data documentation website.

The [Technical Reports Series](#) are scientific contributions dealing with technical aspects of data presentation (e.g. sampling design and data collection, data documentations, field reports, etc.) as well as methodical questions. Furthermore, they contain various reports on data handling for a selection of constructs and variables.

Citation

Please acknowledge the use of the TwinLife data in your work by citing both the [dataset](#) (Diewald et al., 2022) and the [TwinLife reference paper \(Hahn et al., 2016\)](#).

Data

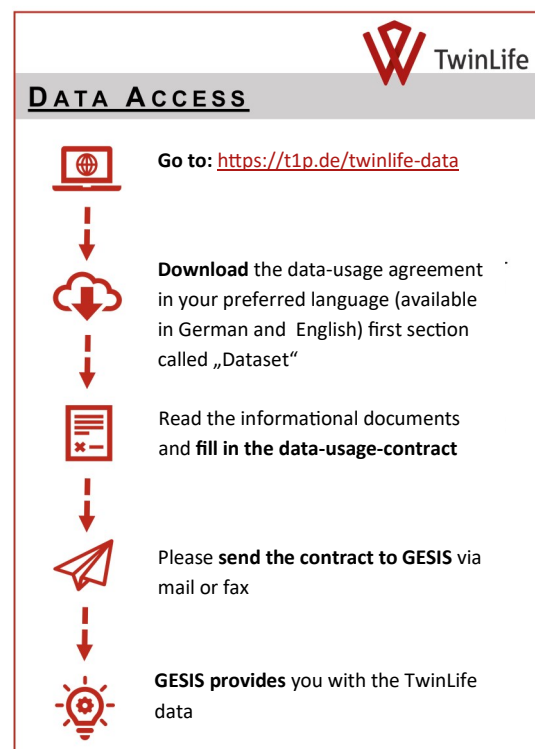
The TwinLife data are described and archived in the [GESIS data catalogue](#).

To get access to the TwinLife data, please fill in the Data Use Agreement which you can find under 'Actions' in the GESIS data catalogue. All data sets are available with English and German labels in SPSS and Stata formats. You can also follow the direct link to the [Data Use Agreement](#).

Variables

In the Codebooks for each data collection (ZA6701_cod_wid\$.pdf), you can find a list of the complete set of variables with names, variable and value labels as well as the distributions of frequencies. Furthermore, it documents question texts, filter conditions, and references of all variables. You can find the codebooks at the [TwinLife documentation website](#).

A list of all variables (including their frequencies), their corresponding question text, answer options, and filter conditions can be found in the Datasets and Instrument files at the metadata documentation platform [paneldata.org](#).



If you have any questions, please feel free to contact us at the following email address: info@twin-life.de