Twin Research and Human Genetics (2026), **29**, 258–268
doi:10.1017/thg.2026.10057

CAMBRIDGE UNIVERSITY PRESS

Article

The German Twin Registry (GERTRUD): Towards Harmonized, Multisource Twin Research

Moana Beyer[1,2], Dmitry Kuznetsov[1] and Simone Kühn[1,3,4,5]

[1]Center for Environmental Neuroscience, Max Planck Institute for Human Development, Germany; [2]Department of Psychology, Humboldt-Universitat zu Berlin, Germany; [3]Clinic and Policlinic for Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf, Germany; [4]Psychiatric Environmental Neuroscience (PEN) Clinic for Psychiatry CCM, Charité - Universitätsmedizin Berlin, Germany and [5]Max Planck UCL Centre for Computational Psychiatry and Ageing Research, UK

Abstract

Twin registries worldwide increasingly function as large-scale research infrastructures, enabling standardized phenotyping across the lifespan, integration of biological and environmental data streams, and international cross-cohort collaborative research and replications. This development is also taking place in Germany. The GERman Twin Registry Under Development (GERTRUD; www.gertrud.info) was established in 2022 as the first nationwide research platform for recruiting twins and higher order multiples of all ages within Germany to support the large-scale genetically informative psychological, sociological, health, and neuroscience twin research at national and international levels. GERTRUD is being developed as a modular infrastructure that supports classical and extended twin family designs, combining annual core survey waves with optional embedded modules for intensive phenotyping (e.g., neuroimaging, smartphone-based assessments), biosampling, and linkage of participants' residential context to external geographic datasets via geospatial information systems (GIS). To operate within Germany's stringent data protection landscape, GERTRUD implements project-specific pseudonymisation, role-based access control, and contract-governed remote analysis access. This article describes GERTRUD's governance and legal–technical framework, its multisource data architecture, and the potential for collaboration across Germany and internationally. Examples of early data implementations further illustrate that the continuously collected multimodal twin data constitute a critical asset, essential for successful harmonization, replication, and collaborative and integrative behavioral genetics research.

Keywords: Research infrastructure; Environmental neuroscience; Twin registry; Data linkage; Data governance

(Received 29 November 2025; revise received 7 March 2026; accepted 11 March 2026; First Published online 4 May 2026)

Twin studies have long been a cornerstone of behavioral, medical and developmental research, offering rigorous designs to disentangle genetic and environmental sources of variance in complex traits. The establishment of large twin registries worldwide has made a significant contribution to this. These registries extended the samples by inclusion of twin relatives, and moved from the single phenotype focus to a rather broad range of phenotypes measured in longitudinal designs. Across Europe and internationally, large-scale collaborations have demonstrated how coordinated governance and harmonization can generate scientific value beyond any single cohort, including EU-funded multicohort initiatives such as GenomEUtwin (Peltonen, 2003), registry-linked Nordic collaborations such as NorTwinCan (Harris et al., 2019; Mucci et al., 2016), and large-scale pooled analyses across dozens of registries or twin studies such as CODATwins (Silventoinen et al., 2019) and IGEMS (Pedersen et al., 2019). These efforts underscore a broader shift in twin research towards integrative, multicohort science that supports meta- and mega-analyses, replication, and methodological innovation.

In contrast to this international trend, the German twin research landscape has remained fragmented. While several high-quality twin studies have emerged, they have typically been constrained by specific regions, age groups, or research questions (Busjahn, 2012; Enck et al., 2021; Hahn et al., 2012; Kandler et al., 2019). For example, the nationwide TwinLife panel study has recruited approximately 4000 families with monozygotic (MZ) and same-sex dizygotic (DZ) twins from across Germany to achieve population representativeness, and has followed them longitudinally since 2014 (Diewald et al., 2019; Hahn et al., 2016; Mönkediek et al., 2019). Another completed study, the Study of Personality Architecture and Dynamics (SPeADy), recruited approximately 800 twin families across three waves (2016–2022), using media calls, twin clubs, and registration office contacts — predominantly in North Rhine-Westphalia and Lower Saxony — to study personality and individual differences across the lifespan (Kandler et al., 2019). Additional regional or domain-focused studies have also produced important findings, particularly in clinical and biomedical contexts (Busjahn, 2012; Enck et al., 2021; e.g., Hahn et al., 2012). However, a shared national infrastructure for ongoing

Corresponding author: Moana Beyer; Email: beyer@mpib-berlin.mpg.de
Cite this article: Beyer M, Kuznetsov D, Kühn S. (2026) The German Twin Registry (GERTRUD): Towards Harmonized, Multisource Twin Research. *Twin Research and Human Genetics* **29**: 258–268, https://doi.org/10.1017/thg.2026.10057

© The Author(s), 2026. Published by Cambridge University Press on behalf of International Society for Twin Studies. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Check for updates

recruitment, harmonized data sharing, and long-term follow-up has remained absent. Consequently, German twin data have been less readily positioned for international integration, replication, or systematic cross-cohort harmonization.

The absence of this type of infrastructure is partly the result of scientific history and structural conditions, but it also reflects a set of uniquely German cultural, legal, and systemic barriers. First, twin and genetic research is a particularly sensitive topic in German society, given the historical misuse of science, and twin studies in particular, during National Socialism. Second, Germany has among the strictest data privacy laws in Europe, operative even before the General Data Protection Regulation (GDPR) took effect. The Federal Data Protection Act (*Bundesdatenschutzgesetz*, BDSG) and state-level equivalents impose additional restrictions on top of the GDPR baseline — particularly regarding the collection, storage, and linkage of identifiable or genetic data — resulting in a more restrictive framework than GDPR alone requires (Intemann et al., 2023; March et al., 2018). Third, unlike countries with national personal identifiers (e.g., the Nordic countries), Germany operates without a universal ID number. Record linkage must therefore rely on indirect approaches such as probabilistic matching on combinations of name, date of birth, and address — a technically and legally challenging process that has been characterized as requiring time-consuming individual case solutions for each project (Domhoff et al., 2022; March et al., 2018).

Despite these challenges, Germany presents an enormous largely untapped potential for twin research in within-country analyses and international collaborations. With a population exceeding 83 million and approximate twinning rates of 1.84% (Torres et al., 2023), the theoretical capacity for large-scale twin recruitment in Germany is at least comparable with that of countries with long-standing registries. To illustrate, co-twin control designs for low-prevalence disorders such as schizophrenia (lifetime prevalence ~0.4%; Saha et al., 2005) require in the order of several hundred discordant pairs to achieve adequate statistical power. The example of the Swedish Twin Registry, with over 200,000 registered individuals (Zagai et al., 2019), demonstrates that this threshold for such traits can be reached. Projections on a German registry of comparable relative coverage suggest a substantially larger absolute pool of pairs. From a research perspective, this would mean the possibility to identify more than 400 discordant twin pairs — a critical threshold for genetically informed causal inference — not only for schizophrenia but also for other traits and diseases of low prevalence.

In response to this scientific opportunity and infrastructural gap, the German Twin Registry (GERTRUD) was established in 2022 as a nationally coordinated framework for research on twins and their families across the lifespan. The registry has been initiated and is hosted by the Centre for Environmental Neuroscience at the Max Planck Institute for Human Development (lead: Prof. Dr Simone Kühn) and was founded as a collaboration across institutions in multiple federal states (Berlin, Baden-Württemberg, North Rhine-Westphalia, Hamburg, Saarland), including Prof. Dr Christian Kandler (University of Bielefeld), Prof. Dr Jan Beucke (Medical School Hamburg), Dr Miriam Mosing (Max Planck Institute for Empirical Aesthetics), Prof. Dr Frank Spinath (Saarland University) and Prof. Dr Paul Enck (University Hospital Tübingen). To support centralized recruitment, governance, and harmonized longitudinal data collection, GERTRUD established a dedicated, GDPR- and BDSG-compliant participant management system.

GERTRUD is designed as a modular, multipurpose infrastructure that facilitates a wide range of scientific aims across behavioral, cognitive, environmental, psychiatric, sociological and neurobiological domains. Rather than replacing existing twin studies, it provides a shared platform they can collaborate with: twins who have participated in earlier German studies are eligible to enrol, and participating studies continue to operate under their own governance structures. The collaborative framework ensures that recruitment and data collection protocols remain flexible and inclusive, supporting both hypothesis-driven and exploratory approaches in a continuously enlarging sample of twins and their families across Germany. Nationally coordinated recruitment, long-term follow-up, openness for collaboration, and the extensive phenotyping efforts with advanced measurement techniques, allow complex research programs to scale in ways that no individual twin study could achieve on its own.

Methods and Materials

Study Design

GERTRUD is an open, nationwide family-based registry for MZ and DZ twins, as well as higher order multiples of all ages, residing in Germany. There are no age restrictions or diagnostic criteria for inclusion: all twins, triplets, and higher order multiples with at least one member residing in Germany, along with their nontwin family members, are eligible to register. The registry is intended to achieve national coverage across all federal states (see Figure 1) and currently enrols participants spanning an age range from birth to over 70 years. The medium-term recruitment target is approximately 1000 complete twin pairs per year, corresponding to roughly 6000 twin pairs registered in GERTRUD over a 5-year period.

Participation is voluntary and contingent upon informed consent, with ethics approval obtained for each study module implemented within the registry. Consent procedures are adapted to participant age. For adults (aged 16 and above), participants provide their own informed consent. For children under 16 years, a parent or legal guardian provides consent on the child's behalf; the current core survey questionnaire is directed at adult participants (16+). Additional questionnaires targeting parents of twin children and age-relevant questionnaires for twins aged 10–15 are forthcoming in the second half of 2026. Beyond classical twin designs, GERTRUD supports extended twin family designs (ETFDs), enabling the analysis of additional family members, namely biological and foster parents, nontwin siblings, spouses of twins, and children of twins. ETFDs leverage systematic differences in genetic relatedness and shared environment across relatives to relax restrictive assumptions of the classical twin design and to support richer models of familial resemblance, including assortative mating and intergenerational transmission, thereby improving interpretability and reducing bias in estimates of genetic and environmental contributions (Keller et al., 2009).

Recruitment Strategies

Recruitment combines multiple strategies intended to support both reach and diversity of ascertainment. Twins and their family members can opt into registry membership and then separately opt into specific modules and/or the core longitudinal survey waves.

First, GERTRUD collaborates with existing German twin studies and cohort infrastructures, using study-specific recontact procedures where ethically and legally permissible. The registry is

https://doi.org/10.1017/thg.2026.10057 Published online by Cambridge University Press

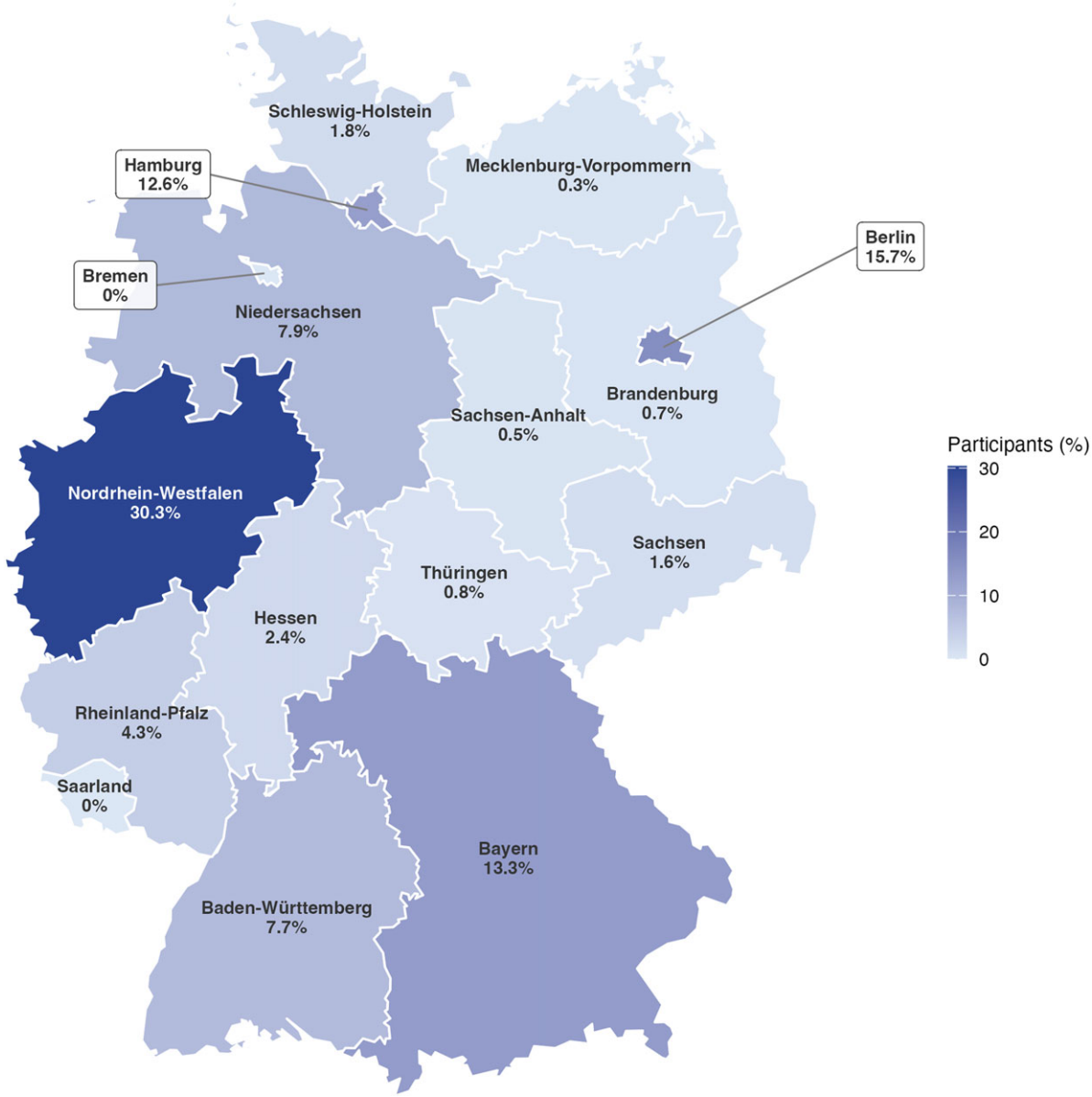


Figure 1. Geographic distribution of GERTRUD participants across German federal states ($n = 2221$ with available postal code). Shading indicates the percentage of participants residing in each state.

explicitly designed to integrate and expand on past and ongoing twin efforts in Germany; the subsets of twins from earlier twin and cohort studies — such as TwinLife (Diewald et al., 2019; Hahn et al., 2016; Mönkediek et al., 2019), SPeADy (Kandler et al., 2019), Hamburg City Health Study (Jagodzinski et al., 2019) and OCDTWIN (Mataix-Cols et al., 2023) — are already enrolled in GERTRUD. Additional collaborations with existing cohorts are underway. To facilitate cohort integration and scalability, partner institutions directly recontact eligible twins within their samples to inform them about GERTRUD and invite participation, ensuring compliance with privacy regulations.

Second, population-oriented recruitment in GERTRUD is tailored to Germany's decentralized administrative structure. To reach twins and their families across German states, the registry collaborates with multiple local registration offices, which are authorised to perform targeted searches based on specific criteria such as date of birth and birth surname. With about 5000 local offices across Germany, GERTRUD is prioritizing those that provide the larger coverage. The search of date of birth and birth surname within local offices enables the identification of potential twin pairs within the general population to increase population coverage and reduce reliance on convenience sampling, while remaining compatible with decentralized administrative procedures across municipalities.

Third, the study-specific channels are actively utilized. To recruit young adult twins, community outreach is conducted via online platforms and via university offices. Public engagement activities (e.g., science fairs and press coverage), depending on the topic, aim to engage twins of all ages or younger twins through their parents. Finally, for prenatal and early life studies, we contact and recruit women with twin pregnancies through obstetrics departments of major hospitals.

https://doi.org/10.1017/thg.2026.10057 Published online by Cambridge University Press

Module Structure

GERTRUD is organized around a modular data architecture designed to support the integration of the multisource data while maintaining clear governance boundaries. The registry distinguishes (1) a harmonized core survey wave repeated once a year. These core surveys provide a common set of measures for longitudinal and cross-cohort comparisons (see description of the core waves below); and (2) optional embedded modules, which can include add-on questionnaires, lab-based experiments, intensive phenotyping (e.g., neuroimaging), smartphone-based assessments, and biosampling. A module is a self-contained data collection effort with its own protocol, ethics approval, and consent form, implemented within the shared GERTRUD infrastructure. Modules differ in how they were initiated: some were developed in collaboration with existing twin studies and may recruit from both the partner study's sample and the GERTRUD registry (e.g., *TwinLife Environment*); others were conceived and designed entirely within GERTRUD and recruit from the GERTRUD registry (e.g., *Envision*, *TwinIDS*). It is important to note that collaboration-based modules do not imply a merger of registries: GERTRUD and partner studies such as TwinLife (Hahn et al., 2016) remain distinct infrastructures with separate datasets stored on independent servers. For example, the *TwinLife Environment* module recruited approximately 56 MZ pairs from the TwinLife panel for neuroimaging, biosampling, and geographic information system (GIS) linkage; participants signed a dedicated consent form, and their data are held independently. The retrospective linkage to the anonymized TwinLife wave data, although not the primary aim, is possible in future under an additional consent component that permits linkage of TwinLife and GERTRUD records for that specific subsample. Each module is led by an individual researcher or team within GERTRUD and GERTRUD collaborations and may use newly and previously developed data collection procedures in line with the planned schedule and timing depending on the protocol (e.g., lab visits, mail-in kits, smartphone-based assessments, or online questionnaires).

The rationale for this structure is to separate the core longitudinal survey waves — which provide harmonized measures across all participants — from optional intensive or domain-specific data collection components appropriate only for subsamples. This avoids overburdening participants and allows each module to obtain targeted ethics approval without requiring a single umbrella consent for all possible future data uses. Integration across modules is implemented in practice by linking module-specific datasets through a shared participant pseudonym within the secure GERTRUD environment, for participants who have consented to participate in multiple components and allowed data linkage across modules.

Registry membership does not automatically imply consent for and participation in all modules; instead, participants provide a single informed consent for registry enrolment and recontact, and separate consent forms for each module they join and for specific linkage components such as GIS linkage or biosample storage and analysis. This preserves legal clarity about purpose limitation and participants' rights to withdraw from individual components independently.

Core Wave

GERTRUD is structured around annual (or otherwise periodic) core survey waves that include repeated, standardized measures of mental health, cognitive functioning, life circumstances, and environmental exposures. These core measures are designed to facilitate harmonization across time points and to support cross-cohort mapping of common data elements. The first core wave includes validated self-report instruments covering depressive symptoms (PHQ-9; Kroenke et al., 2001; generalized anxiety (GAD-7; Spitzer et al., 2006), psychotic-like experiences and associated distress (CAPE-15; Capra et al., 2015), perceived stress (PSS-10; Cohen et al., 1983), loneliness (UCLA Loneliness Scale; Russell, 1996), obsessive–compulsive symptoms (OCI-R; Foa et al., 2002) and adverse childhood experiences (ACE; short-form anchored in the original ACE framework; Felitti et al., 1998). Rich sociodemographic indicators (education, occupation, income, household composition, housing tenure, religious affiliation/commitment, sexual orientation) and targeted twin items (birth order, gestational age, cohabitation duration, relationship quality/conflict) are included to support harmonization and twin-specific modeling. Environmental modules capture childhood urbanicity, time in nature across seasons and nature connectedness (Nature Relatedness Scale; Nisbet et al., 2009). These core measures, administered via a secure online platform, enable twin-only analyses and twin–singleton comparisons through harmonization with large German cohorts (e.g., SOEP, TwinLife). The second core wave of data collection is forthcoming in the second quarter of 2026. Data analyses of the first core wave are currently underway; data are available for 1436 participants (60.4% of registered participants), with full results to be reported in dedicated analysis papers.

Biobanking

Biosample collection within GERTRUD is implemented as a module-based workflow, with study-specific ethics approval, consent, and standardized protocols for collection, processing, and storage. So far, the biosampling procedure has been successfully implemented in two modules: *TwinLife Environment* and *NeoTwins*.

TwinLife Environment. Three biosample types were collected from adult twin participants (all samples collected from the same individuals): saliva, stool, and urine. Saliva was collected on-site directly after MRI scanning (to ensure participants had not eaten or drunk for at least 1.5 hours prior) using the Oragene DNA kit (1 ml), intended for DNA extraction to support zygosity confirmation and downstream genetic and epigenetic analyses. Stool samples were collected via a mail-in protocol: participants received collection kits (Zymo DNA/RNA Shield Fecal Collection Tube) with detailed instructions and were asked to collect a sample 24 hours before their on-site appointment and bring it to the session; samples are intended for future microbiome profiling. Urine was collected on-site (50 ml and 1 ml aliquots), stored at approximately 10°C for up to two hours, then transferred to −80°C storage; samples are intended for planned metabolomics and hormone assays. All samples are stored at the Max Planck Institute for Human Development at −80°C pending analysis.

NeoTwins. The biosampling in this module is performed in newborn twin pairs and their mothers and includes stool, saliva, urine, and hair samples. Stool samples from newborns are collected on day 7 postbirth (by caregivers at home, using the same Zymo collection kit protocol) and additionally on-site at the scheduled study appointment. Saliva from newborns is collected on-site. Urine and hair samples are collected from mothers during pregnancy.

https://doi.org/10.1017/thg.2026.10057 Published online by Cambridge University Press

Planned scale-up across GERTRUD more broadly includes additional biosample types (e.g., dried blood spots) and additional assay targets (e.g., inflammatory markers, epigenotyping via DNA methylation profiling), subject to module-specific ethics approvals. Genetic data are planned to be shared as polygenic indices; epigenetic data as DNA methylation scores; other biological data as raw estimates, provided individuals cannot be re-identified.

Additional Modules

Beyond the core wave, GERTRUD supports additional modules targeting specific research questions. These modules can take various forms — including online surveys, lab-based protocols, and smartphone-based ecological assessments — and may draw on instruments already used within GERTRUD or introduce newly developed measures, allowing cumulative methodological development across studies. Crucially, module submission is open to any researcher regardless of institutional affiliation or prior involvement with GERTRUD. Proposals are evaluated by GERTRUD's internal governance board (comprising the registry founders) for ethical compliance, methodological feasibility, and participant burden. An overview of ongoing core and embedded module data collection efforts is provided in Table 1.

Operationalization of GIS Linkage

With explicit consent, participant residential address information is processed within a secure environment and transformed into geocodes (coordinates) that are stored separately from direct identifiers. Environmental indicators are then linked to geocodes using external geographic datasets, and derived exposure variables (rather than raw address data) are integrated into the research dataset under project-specific pseudonyms. This approach allows genetically informed analyses of environmental exposures while complying with data minimization principles by ensuring that raw address data do not leave the secure processing environment.

Current linkage is based on home addresses; for the *TwinLife Environment* subsample, continuous GPS tracking data from a geographical ecologigal momentary assessment (GEMA) are additionally available. GERTRUD's linkage strategy is designed to support both cross-sectional and longitudinal exposure modeling: participants' addresses are updated at each core annual wave, with duration at each address recorded, allowing longitudinal modelling of residential exposure histories. Additionally, historical home address data — particularly covering sensitive periods within the first 15 years of life — are collected using specifically developed geo-questionnaires.

Environmental indicators are derived from a set of publicly available German and European datasets, extracted at multiple buffer radii (100 m, 200 m, 500 m, 1000 m, and 2000 m) around geocoded residential locations. These include: air pollution ($PM_{2.5}$, PM_{10}, NO_2, ozone; European Environment Agency, 1–2 km resolution; German Environment Agency, 2.4 km resolution), vegetation cover (Normalised Difference Vegetation Index, NDVI; Copernicus High Resolution Layer, 10 m), tree cover density (Copernicus High Resolution Layer, 10–100 m), dominant leaf type (Copernicus, 10 m), land use and land cover (CORINE Land Cover and Urban Atlas; Copernicus Land Monitoring Service, vector), land use intensity (Mendeley Data, 10 m, Germany), building height (Copernicus Urban Atlas, 10 m), terrain elevation (DGM200; Bundesamt für Kartographie und Geodäsie, 200 m), noise pollution (German Environment Agency, vector, EU Environmental Noise Directive categories), a Quietness Suitability Index (European Environment Agency, 100 m), surface soil moisture (Copernicus, ~900 m), population density (WorldPop, ~100 m), night light intensity (VIIRS/Earth Observation Group, ~450 m), and street network connectivity (OpenStreetMap via OSMnx).

Results

Participation in the Registry and Core Survey

As of February 2026, the GERTRUD registry has enrolled over 2379 individual twins, including 884 pairs where both twins registered, spanning the full age range of 0–76 years. Among all individuals with confirmed zygosity, 44.5% are MZ, 44.8% are DZ and 10.8% remain unclassified pending genetic confirmation. Recruitment has successfully covered both urban and rural areas across Germany, enabling broad sociodemographic representation and facilitating regionally stratified analyses (see Figure 1 for the geographic distribution across federal states). Characteristics of the sample are also listed in Table 2. Gender is recorded using the three categories available under German civil law: female, male, and *divers* (the legally recognised third gender category introduced by the Personenstandsgesetz amendment of 2018). Core wave 1 participation (2023–2025) achieved response rates of 60.4%. Moreover, over 2221 participants have already consented to address-based linkage to official environmental datasets, including air pollution ($PM_{2.5}$, NO_2), NDVI-based green space, noise and deprivation indices via the GIS pipeline described earlier.

To support multicohort and multipurpose use, GERTRUD tracks participation at two levels: (1) registry enrolment and (2) module participation. At enrolment, participants complete a brief registration form collecting basic demographic information, including address and twin status, and provide consent for recontact. Completion of the core wave questionnaire is voluntary and constitutes a separate, optional step. Address information and twin status are therefore available for all registered participants, irrespective of core wave participation.

Participation in Additional Modules

To evaluate the feasibility of high-dimensional, multisource data integration, several pilot studies were implemented in parallel with the first core wave. Dedicated publications for each module are in preparation; data analyses are currently ongoing. One such study, the *TwinLife Environment* module, initiated in 2023, involves a sample of MZ twin pairs ($n = 310$ twins) and integrates structural and functional neuroimaging, real-time GPS tracking from which we derive GIS-based environmental indicators and GEMA, neurocognitive testing. Biosample collection and structured clinical interviews (Mini-DIPS; Margraf & Cwik, 2017) were also piloted in this cohort, laying the groundwork for future biobanking and genetic data integration. Geographic information was linked to individual-level data using the GIS pipeline described in the Methods section. This subsample is currently being followed up for the second MRI measurement within the timeframe of 2025−2026.

Other substudies included additional modules on parenting (conducted in 2025, $n = 340$), which investigated strategies parents use to foster individuality or similarity among twins, and additional modules of a perceptual experiment (ongoing 2025−2026, $n > 220$) examining how visual processing differs in relation to the degree of urbanicity and nature in the living environment. Further experimental protocols are currently under development, including targeting the performance of episodic memory in both

https://doi.org/10.1017/thg.2026.10057 Published online by Cambridge University Press

https://doi.org/10.1017/thg.2026.10057 Published online by Cambridge University Press

Table 1. Description of different data types collected within GERTRUD and its core and additional embedded modules (status: February 2026)

Data modality	Description and Key Variables	Module	Sample	Sample Size	Status
Core survey data	Demographics, mental health (e.g., PHQ-9, GAD-7), wellbeing, life events, physical health	GERTRUD core	All consenting multiples	1436	Wave 1 (since 2023, ongoing); Wave 2 starting in spring 2026
GIS-linked data	Address-based linkage to air pollution (PM2.5, NO2), green space (NDVI), noise, neighborhood socioeconomic status	GERTRUD core	All consenting multiples	2221	Functional and ongoing (since 2023)
Neuroimaging (MRI) data	Structural (T1), resting-state functional (rs-fMRI) scans, two task-related fMRI scans	TwinLife Environment; NeoTwins	MZ twins; newborn twins	310 individuals (155 MZ pairs); 6 twin pairs	Wave 1 complete, follow-up ongoing; Ongoing
Mobile sensing (GEMA)	1-week protocol: GPS tracking, ecological momentary assessment of mood and context, physiological parameters	TwinLife Environment; NeoTwins	MZ and DZ twins	~380 individuals (goal: 600)	Ongoing (2023–2027)
Biobank	Saliva (DNA for genotyping and epigenotyping/DNA methylation profiling; ~415 samples); Stool (microbiome profiling; 295 samples); Urine (metabolomics, hormone assays e.g., cortisol; ~180 samples)	TwinLife Environment; NeoTwins	MZ twins; Newborn twins and their biological mothers	~540 samples total	Saliva: wave 1 completed (2023–2025), wave 2 ongoing; Stool: pilot completed (2023–2025); Urine: wave 1 completed, wave 2 started
Neurocognitive data	Verbal memory, working memory, processing speed, executive function, semantic verbal fluency, crystallized intelligence	TwinLife Environment	MZ twins	310 individuals (~110 with follow-up)	Wave 1 completed (2023–2025), wave 2 ongoing (2025–2026)
Clinical data	Semistructured clinical interview (Mini-DIPS)	TwinLife Environment	MZ twins	310 individuals (~110 with follow-up)	Wave 1 completed, wave 2 ongoing
Experimental data: Visual perception strategies	Visual tasks (illusions, field dependence/independence, peripheral vision, global-local bias), autism traits, psychotic experiences, environmental sensitivity, stress, OCD	Envision	MZ and DZ twins	$n > 220$	Ongoing (2025–2026)
Survey data: Parenting practices	Retrospective and current parenting practices (individuality/similarity emphasis), twin relationship quality	TwIDS	MZ and DZ twins and their parents	340	Completed (2024–2025)

Note: Sample sizes in this table are not additive and do not all correspond to the registered GERTRUD sample reported in Table 2: some module participants (e.g., within TwinLife Environment, NeoTwins, Envision, TwIDS) are affiliated with GERTRUD and contribute data to specific modules but are not formally registered in the GERTRUD registry. MZ, monozygotic; DZ, dizygotic; PHQ-9, Patient Health Questionnaire-9; GAD-7, Generalized Anxiety Disorder-7; $PM_{2.5}$, particulate matter ≤2.5 µm; NO_2, nitrogen dioxide; NDVI, Normalized Difference Vegetation Index; GIS, geographic information systems; T1, T1-weighted structural MRI; rs-fMRI, resting-state functional MRI; GPS, global positioning system; GEMA, geographical ecological momentary assessment; OCD, obsessive compulsive disorder; PIAP, Peas-in-a-Pod questionnaire.

Table 2. Characteristics of GERTRUD-registered participants (status: February 2026)

Section	Variable	Data
Participants	Participants, total	2379
	Completed baseline questionnaire	1436 (60.4%)
Addresses	Participants with address available	2221 (93.4%)
Matched Multiples	Twins	1768 (74.3%); 884 pairs
	Unmatched singles	587 (24.7%)
	Triplets	18 (0.8%); 6 sets
	Quadruplets	4 (0.2%); 1 set
Zygosity	Monozygotic	1057 (44.5%)
	Dizygotic	1064 (44.8%)
	Zygosity unknown	256 (10.8%)
	Determined by questionnaire (PIAP/similarity)	762
	Determined by blood test (self-report)	929
	Determined by DNA genotyping	0 (in planning)
Gender	Female	1528 (64.2%)
	Male	832 (35%)
	Divers	9 (0.4%)
	Unknown/missing	6 (0.3%)
Age	Mean (*SD*), years	27.9 (15.2)
	Range, years	0–76
	Adults (≥16 years): *n*	1971
	Adults (≥16 years): mean (*SD*)	31.8 (13.3)
	Adults (≥16 years): range	16–76
	Children (<16 years): *n*	389
	Children (<16 years): mean (*SD*)	8.2 (5.1)
	Children (<16 years): range	0–15

Note: Some participants contribute data to embedded modules (Table 1) but are not formally registered in GERTRUD; these individuals are not included in this table. MZ, monozygotic; DZ, dizygotic.

urban and natural settings, investigating working memory, and studying how the brains of newborn twins differ in relation to environmental influences. The data of participants who consent will be linked via pseudonyms within GERTRUD.

High-Resolution Environmental Assessment

In addition to the self-reported data on neighbourhood features and the role of the particular environmental factors in their lives, in a GEMA substudy (goal: 600 MZ and DZ twins), GPS-based ecological momentary assessment captures dynamic exposures and affective states in daily life. Currently, approximately, 380 individuals took part in this module.

Neuroimaging

GERTRUD incorporates large-scale neuroimaging with complete structural and functional MRI datasets from 155 MZ twin pairs (310 individuals, aged 18–60 years), with a follow-up wave currently underway (2025–2026), totalling 420 scans to date. Imaging protocols include anatomical, resting-state and task-based scans targeting cognitive control and emotion regulation. All imaging data follow BIDS-compliant pipelines, facilitating national and international cooperation. Together, the two waves will enable analyses of change over time and provide insights into longitudinal trends in brain structure and function across adulthood.

Biobanking

Following the collection and storage procedures described in the Methods section, the pilot biobank within the *TwinLife Environment* module comprises saliva ($n \approx 415$), stool ($n = 295$), and urine ($n \approx 180$) samples from roughly 320 individuals. Analyses are currently underway. Planned scale-up includes additional biosample types (dried blood spots, hair, deciduous teeth), which — combined with address history — would provide time-resolved indicators of environmental burden across the life course.

GIS Linkage

GERTRUD's GIS linkage is designed not merely to append contextual covariates to phenotypic data, but to enable genetically informed environmental epidemiology. In conventional environmental epidemiology, associations between neighborhood exposures (e.g., air pollution, green space, walkability, deprivation) and health outcomes may be biased by residential selection and other unmeasured familial factors that shape both where people live and their risk for mental or physical health outcomes. Twin designs provide a principled way to address these limitations. In particular, within-pair analyses among MZ twins control for all genetic factors and much of the early shared environment, thereby reducing confounding and strengthening causal inference about environmental exposures that differ between cotwins. More broadly, combining GIS-derived exposures with twin and family models allows GERTRUD to quantify the extent to which apparent 'environmental effects' reflect shared familial background or gene–environment correlation, and to test whether environmental associations differ as a function of genetic liability or family context, providing novel insights into mechanisms of behavioral and health phenomena. As a result, the analyses in twin cohorts consistently demonstrate the added value of genetically informed approaches for neighborhood and built-environment questions (Cohen-Cline et al., 2017; e.g., Duncan et al., 2014; Wang et al., 2025). In GERTRUD, the same logic motivates the GIS pipeline: environmental indicators derived from geocoded residential histories are integrated with harmonized phenotypes and twin-family structure to support both population-level associations and within-family comparisons that go beyond standard observational designs.

Legal-Technical Infrastructure

A key output of GERTRUD's early implementation is the operational legal–technical infrastructure that enables multisource integration under German constraints. The registry's architecture supports project-specific pseudonymization, modular consent, and contract-based access pathways that make it possible to integrate behavioral, environmental, imaging, and biosample-derived variables without undermining purpose limitation and participant rights. This infrastructure directly addresses barriers

https://doi.org/10.1017/thg.2026.10057 Published online by Cambridge University Press

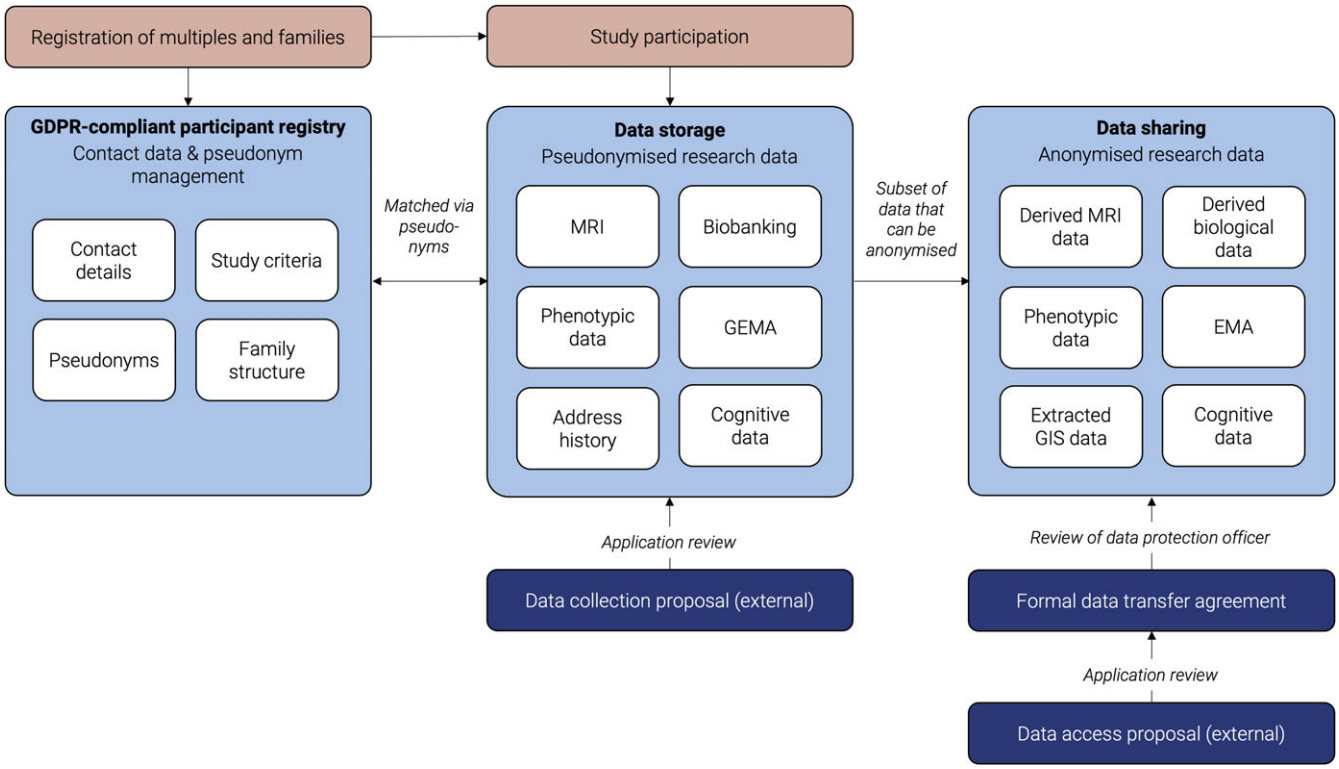


Figure 2. Schematic of GERTRUD's infrastructure and data flow. The figure illustrates (1) separation of identifiers from research data; (2) project-specific pseudonymization; (3) integration of core wave and embedded modules; and (4) controlled researcher access via contracts and secure analysis environment.
Note: GDPR, General Data Protection Regulation; Castellum, participant management and pseudonymization system; GIS, geographic information systems; EMA, ecological momentary assessment; GEMA, geographical ecological momentary assessment; MRI, magnetic resonance imaging.

that have historically limited German participation in multicohort twin research, providing a governance template for harmonized research use while maintaining compliance with stringent data protection expectations. A schematic overview of the infrastructure is provided in Figure 2.

Storage of Contact Data

The registry is hosted at the Max Planck Institute for Human Development and uses *Castellum* (Bengfort et al., 2022), a secure data governance platform designed for GDPR-compliant pseudonymization and controlled data access. *Castellum* ensures a strict separation between personally identifying information (PII) and research data by assigning study- and project-specific pseudonyms. Role-based access permissions, encryption protocols and audit logs guarantee that identifiable information remains inaccessible to researchers and never leaves the secure server environment without explicit authorization.

All personal and sensitive data are collected only after explicit informed consent. Each participating study module operates under project-specific data use agreements, which outline the legal responsibilities, purpose and scope of data access.

Data Storage and Linkage

GERTRUD employs a modular and scalable approach to data linkage; collection, geocoding, and environmental linkage procedures are described in the Operationalization of GIS Linkage section above.

Data Sharing

External researchers, as well as the collaborators involved in the additional study modules, can request access to the data collected within longitudinal core modules and to the data of the additional research modules. To enable secure and compliant research collaborations, GERTRUD is currently testing virtual research environments (VRE) that provide encrypted, remote access to de-identified datasets. In this model, raw data would never leave the institutional infrastructure. This model could allow external researchers to conduct analyses without direct access to personally identifying information, offering a potential pathway to balance open science with data protection requirements.

Cross-Cohort Harmonization and Multisource Data Integration

GERTRUD is being developed with the explicit aim of enabling harmonization across modules and future cross-cohort collaboration. Initial steps toward this goal include ongoing methodological exchanges and data-harmonization work with collaborators at the Karolinska Institutet (Stockholm, Sweden) and the University of Oslo (Norway). At the infrastructure level, GERTRUD adopts design principles intended to facilitate cross-cohort integration, including structured wave identifiers, modular metadata schemas, and standardized data formats for high-dimensional modalities (e.g., BIDS for neuroimaging).

Harmonization within GERTRUD is approached at two levels. First, across age groups: where developmentally appropriate, similar or identical questionnaire instruments are used across

https://doi.org/10.1017/thg.2026.10057 Published online by Cambridge University Press

cohorts and age ranges to enable standardization, pooling and comparison. Second, across modules: the core survey instruments (e.g., PHQ-9, GAD-7, PSS-10) are selected to overlap with measures used in embedded modules and with large German population surveys (e.g., SOEP, TwinLife), facilitating cross-cohort comparisons. Currently, harmonization relies primarily on choosing widely used validated instruments; a GERTRUD data dictionary documenting available variables and their cross-cohort mapping is planned.

The *TwinLife Environment* module provides the most concrete current example of multimodal data integration. All 310 participants (155 MZ pairs) in this module have neuroimaging (structural and functional MRI), GIS-linked environmental exposure, neurocognitive, and biosample data. Among these, many twins also completed the first core survey questionnaire, constituting a subset with overlapping self-report phenotypes, objective environmental measures, neuroimaging, and biosample data. A key limitation is that not all participants contribute data to all modalities, and overlap will vary across research questions. Tracking and reporting modality overlap is therefore part of GERTRUD's ongoing data documentation strategy.

Beyond technical alignment, GERTRUD's governance model is designed to support collaboration under GDPR constraints. Project-specific pseudonymization, module-based consent, and remote-access analysis pathways provide a practical framework for multisite research in settings where data export and linkage are highly restricted. In this sense, GERTRUD contributes not only a new national recruitment resource, but also a test case for how legally robust infrastructures can enable harmonized, multicohort twin science. As the cohort grows and overlap across modalities increases, GERTRUD is positioned to contribute German twin data to international initiatives focused on replication and pooled analyses, particularly in areas where combining genetically informative designs with objective environmental measurement is scientifically critical.

Discussion

GERTRUD represents a shift in the German twin research landscape from project-specific, siloed studies toward a shared infrastructure designed for harmonized longitudinal phenotyping and multisource data integration. By focusing on integration across data sources, research domains, and institutions, GERTRUD addresses long-standing limitations in German twin research — particularly limited opportunities for coordinated recruitment, replication, and cross-study harmonization — and aligns directly with the special issue's emphasis on infrastructure enabling cross-cohort collaboration.

A central contribution of GERTRUD is its modular architecture for integrating multisource data within a single governance framework. Rather than accumulating isolated datasets around narrow research aims, GERTRUD combines a harmonized core wave with embedded modules that add intensive phenotyping and objective exposures. This approach supports replication and harmonization in two ways: first, core measures provide common anchors that can be mapped to other cohorts; second, embedded modules allow targeted deep phenotyping to be integrated with the shared core wave in subsets of participants.

GERTRUD also illustrates how legally robust infrastructures can enable multicohort twin research in highly regulated environments. Germany's lack of a national personal identifier and decentralized oversight constrain traditional forms of health-record linkage and long-term administrative follow-up. Within these constraints, GERTRUD operationalizes a model that relies on modular consent, project-specific pseudonymization, and controlled remote access to support collaborative research while maintaining purpose limitation and participant rights. This governance solution is not merely administrative: it determines which forms of integration are feasible, which linkages can be implemented ethically, and which collaborative models are realistic for German twin research. Linkage to national administrative datasets — such as health insurance claims, hospital discharge records, or mortality registries — could substantially enrich a twin registry's scientific value, as demonstrated by Nordic examples (Harris et al., 2019; Mucci et al., 2016). In Germany, however, such linkages face the same structural barriers described above: the absence of a universal identifier and site-based data storage require project-specific legal agreements for each linkage. GERTRUD's framework is designed to accommodate such linkages where they become legally feasible, but this remains a future aspiration.

The integration of biosampling and objectively linked environmental exposures further positions GERTRUD for research on complex behavioral and health phenotypes in which biological and contextual pathways are intertwined. While biobanking in twin cohorts is well established internationally, integrating biosamples with longitudinal phenotyping, repeated high-dimensional measures (e.g., neuroimaging), and objective environmental indicators remains challenging in practice, particularly under strict privacy constraints. GERTRUD's early biobanking pilot demonstrates feasibility of implementing end-to-end workflows that can be scaled as the cohort grows. This will enable future large-scale analyses that combine genetic liability, biomarkers, and environmental exposures to elucidate fine-grained mechanistic pathways underlying complex phenotypes.

Looking ahead, GERTRUD's design choices — harmonized core assessments, modular embedded studies, standardized high-dimensional data formats, and governance for secure collaboration — are intended to facilitate large-scale genetically informed research in the era of collaboration and international consortia. As overlap across modalities increases and harmonization procedures mature, GERTRUD can support cross-cohort meta- and mega-analyses and replication studies, particularly in research areas that benefit from genetically informative designs paired with objective measurement of environments. In this way, GERTRUD aims to position German twin research as an active contributor to the next generation of collaborative behavioral genetic studies.

Acknowledgments. We thank participating twins and their families, and our colleagues at collaborating institutions. We also want to thank everyone who has helped with the establishment of GERTRUD: Kathi Schmalen, Tobias Bengfort, Laura Scherliess, Janek Stahlberg, Daniela Petrosino, Susanne Kassung, Tina Plokarz, Maria Einhorn, Berit Styll, Sarah Otterstetter, Nicole Siller, Nadine Taube, Anja Stanojlovic, Nele Behrens, Elena Isenberg, Kim Pfirschke, Helena Ackermann, Sandra Solaja, Annegret Hartmann and Akane Lang. Special thanks to Laura Scherliess, Kathi Schmalen and Daniela Petrosino for proofreading this manuscript.

Financial support. GERTRUD and the TwinLife Environment study were mainly funded by the Max Planck Society. The work has been partially funded by the European Union (ERC-2022-CoG-BrainScape-101086188 awarded to SK). Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency (ERCEA). Neither the European Union nor the granting authority can be held responsible for them. During the work on her dissertation, MB was a predoctoral fellow of the International Max Planck Research School on Computational Methods in Psychiatry and Ageing Research

https://doi.org/10.1017/thg.2026.10057 Published online by Cambridge University Press

(COMP2PSYCH, https://www.mps-ucl-centre.mpg.de/comp2psych; participating institutions: Max Planck Institute for Human Development, University College London). DK was funded by the Centre for Environmental Neuroscience, Max Planck Institute for Human Development.

Competing interests. None.

Funding Statement. Open access funding provided by Max Planck Society.

Ethical standards. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

AI use declaration. Large language model-based AI tools (Claude, Anthropic) were used to assist with grammar checking and rephrasing of certain passages of this manuscript. All scientific content was developed, verified, and approved by the authors.

References

Bengfort, T., Hayat, T., & Göttel, T. (2022). Castellum: A participant management tool for scientific studies. *Journal of Open Source Software*, *7*, 4600. https://doi.org/10.21105/joss.04600

Busjahn, A. (2012). HealthTwiSt: The Berlin Twin Registry for Health Research. *Twin Research and Human Genetics*, *16*, 163–166. https://doi.org/10.1017/thg.2012.95

Capra, C., Kavanagh, D. J., Hides, L., & Scott, J. G. (2015). Current CAPE-15: a measure of recent psychotic-like experiences and associated distress. *Early Intervention in Psychiatry*, *11*, 411–417. https://doi.org/10.1111/eip.12245

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, *24*, 385–396. https://doi.org/10.2307/2136404

Cohen-Cline, H., Beresford, S. A. A., Barrington, W. E., Matsueda, R. L., Wakefield, J., & Duncan, G. E. (2017). Associations between neighbourhood characteristics and depression: A twin study. *Journal of Epidemiology and Community Health*, *72*, 202–207. https://doi.org/10.1136/jech-2017-209453

Diewald, M., Riemann, R., Spinath, F. M., Gottschling, J., Hahn, E., Kornadt, A. E., Kottwitz, A., Mönkediek, B., Schulz, W., Schunck, R., Baier, T., Bartling, A., Baum, M. A., Eifler, E. F., Hufer, A., Kaempfert, M., Klatzka, C. H., Krell, K., Lang, V., … Weigel, L. (2019). *TwinLife. GESIS Data Archive.* https://doi.org/10.4232/1.13208

Domhoff, D., Seibert, K., Stiefler, S., Wolf-Ostermann, K., & Peschke, D. (2022). Data linkage of German statutory health insurance claims data and care needs assessments preceding a population-based cohort study on nursing home admission. *BMJ Open*, *12*, e063475. https://doi.org/10.1136/bmjopen-2022-063475

Duncan, G. E., Mills, B., Strachan, E., Hurvitz, P., Huang, R., Moudon, A. V., & Turkheimer, E. (2014). Stepping towards causation in studies of neighborhood and environmental effects: How twin research can overcome problems of selection and reverse causation. *Health & Place*, *27*, 106–111. https://doi.org/10.1016/j.healthplace.2014.02.008

Enck, P., Goebel-Stengel, M., Rieß, O., Hübener-Schmid, J., Kagan, K. O., Nieß, A. M., Tümmers, H., Wiesing, U., Zipfel, S., Stengel, A., Dufke, A., Brucker, S. Y., Linneweh, M., Fischer, K., Renner, T., Frick, J.-S., Martus, P., Nahnsen, S., & Weimer, K. (2021). Medizinische Zwillingsforschung in Deutschland. *Bundesgesundheitsblatt ¾ Gesundheitsforschung ¾ Gesundheitsschutz*, *64*, 1298–1306. https://doi.org/10.1007/s00103-021-03400-2

Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., Koss, M. P., & Marks, J. S. (1998). Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults: The Adverse Childhood Experiences (ACE) study. *American Journal of Preventive Medicine*, *14*, 245–258. https://doi.org/10.1016/S0749-3797(98)00017-8

Foa, E. B., Huppert, J. D., Leiberg, S., Langner, R. R., Kichic, R., Hajcak, G., & Salkovskis, P. M. (2002). The obsessive–compulsive inventory: Development and validation of the OCI-R. *Psychological Assessment*, *14*, 485–496. https://doi.org/10.1037/1040-3590.14.4.485

Hahn, E., Gottschling, J., Bleidorn, W., Kandler, C., Spengler, M., Kornadt, A. E., Schulz, W., Schunck, R., Baier, T., Krell, K., Lang, V., Lenau, F., Peters, A.-L., Diewald, M., Riemann, R., & Spinath, F. M. (2016). What drives the development of social inequality over the life course? The German TwinLife Study. *Twin Research and Human Genetics*, *19*, 659–672. https://doi.org/10.1017/thg.2016.76

Hahn, E., Gottschling, J., & Spinath, F. M. (2012). Current Twin Studies in Germany: Report on CoSMoS, SOEP, and ChronoS. *Twin Research and Human Genetics*, *16*, 173–178. https://doi.org/10.1017/thg.2012.106

Harris, J. R., Hjelmborg, J., Adami, H.-O., Czene, K., Mucci, L., & Kaprio, J. (2019). The Nordic Twin Study on Cancer: NorTwinCan. *Twin Research and Human Genetics*, *22*, 817–823. https://doi.org/10.1017/thg.2019.71

Intemann, T., Kaulke, K., Kipker, D.-K., Lettieri, V., Stallmann, C., Schmidt, C. O., Geidel, L., Bialke, M., Hampf, C., Stahl, D., Lablans, M., Rohde, F., Franke, M., Kraywinkel, K., Kieschke, J., Bartholomäus, S., Näher, A.-F., Tremper, G., Lambarki, M., … Semler, S. C. (2023). *Verbesserung des Record Linkage für die Gesundheitsforschung in Deutschland. arXiv*. https://doi.org/10.48550/arXiv.2312.10093

Jagodzinski, A., Johansen, C., Koch-Gromus, U., Aarabi, G., Adam, G., Anders, S., Augustin, M., Kellen, R. B. der, Beikler, T., Behrendt, C.-A., Betz, C. S., Bokemeyer, C., Borof, K., Briken, P., Busch, C.-J., Büchel, C., Brassen, S., Debus, E. S., Eggers, L., … Blankenberg, S. (2019). Rationale and Design of the Hamburg City Health Study. *European Journal of Epidemiology*, *35*, 169–181. https://doi.org/10.1007/s10654-019-00577-4

Kandler, C., Penner, A., Richter, J., & Zapko-Willmes, A. (2019). The Study of Personality Architecture and Dynamics (SPeADy): A longitudinal and extended twin family study. *Twin Research and Human Genetics*, *22*, 548–553. https://doi.org/10.1017/thg.2019.62

Kandler, C., Riemann, R., Spinath, F. M., Bleidorn, W., Thiel, W., & Angleitner, A. (2012). The Bielefeld Longitudinal Study of Adult Twins (BiLSAT). *Twin Research and Human Genetics*, *16*, 167–172. https://doi.org/10.1017/thg.2012.67

Keller, M. C., Medland, S. E., & Duncan, L. E. (2009). Are extended twin family designs worth the trouble? A comparison of the bias, precision, and accuracy of parameters estimated in four twin family models. *Behavior Genetics*, *40*, 377–393. https://doi.org/10.1007/s10519-009-9320-x

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*, 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

March, S., Antoni, M., Kieschke, J., Kollhorst, B., Maier, B., Müller, G., Sariyar, M., Schulz, M., Enno, S., Zeidler, J., & Hoffmann, F. (2018). Quo vadis Datenlinkage in Deutschland? Eine erste Bestandsaufnahme. *Das Gesundheitswesen*, *57*, e20–e31. https://doi.org/10.1055/s-0043-125070

Margraf, J., & Cwik, J. C. (2017). Mini-DIPS open access. *Aufl*, *7*, 590–592.

Mataix-Cols, D., Fernández de la Cruz, L., De Schipper, E., Kuja-Halkola, R., Bulik, C. M., Crowley, J. J., Neufeld, J., Rück, C., Tammimies, K., Lichtenstein, P., Bölte, S., & Beucke, J. C. (2023). In search of environmental risk factors for obsessive-compulsive disorder: study protocol for the OCDTWIN project. *BMC Psychiatry*, *23*. https://doi.org/10.1186/s12888-023-04897-4

Mönkediek, B., Lang, V., Weigel, L., Baum, M. A., Eifler, E. F., Hahn, E., Hufer, A., Klatzka, C. H., Kottwitz, A., Krell, K., Nikstat, A., Diewald, M., Riemann, R., & Spinath, F. M. (2019). The German Twin Family Panel (TwinLife). *Twin Research and Human Genetics*, *22*, 540–547. https://doi.org/10.1017/thg.2019.63

Mucci, L. A., Hjelmborg, J. B., Harris, J. R., Czene, K., Havelick, D. J., Scheike, T., Graff, R. E., Holst, K., Möller, S., Unger, R. H., McIntosh, C., Nuttall, E., Brandt, I., Penney, K. L., Hartman, M., Kraft, P., Parmigiani, G., Christensen, K., Koskenvuo, M., … Kaprio, J. (2016). Familial risk and heritability of cancer among twins in Nordic countries. *JAMA*, *315*, 68. https://doi.org/10.1001/jama.2015.17703

Nisbet, E. K., Zelenski, J. M., & Murphy, S. A. (2009). The nature relatedness scale: Linking individuals' connection with nature to environmental concern and behavior. *Journal of Environmental Psychology*, *29*, 315–327. https://doi.org/10.1016/j.jenvp.2009.03.003

Pedersen, N. L., Gatz, M., Finch, B. K., Finkel, D., Butler, D. A., Dahl Aslan, A., Franz, C. E., Kaprio, J., Lapham, S., McGue, M., Mosing, M. A., Neiderhiser, J., Nygaard, M., Panizzon, M., Prescott, C. A., Reynolds, C.

A., Sachdev, P., & Whitfield, K. E. (2019). IGEMS: The Consortium on Interplay of Genes and Environment Across Multiple Studies ¾ An Update. *Twin Research and Human Genetics*, *22*, 809–816. https://doi.org/10.1017/thg.2019.76

Peltonen, L. (2003). GenomEUtwin: A strategy to identify genetic influences on health and disease. *Twin Research*, *6*, 354–360. https://doi.org/10.1375/136905203770326358

Russell, D. W. (1996). UCLA loneliness scale (version 3): Reliability, validity, and factor structure. *Journal of Personality Assessment*, *66*, 20–40. https://doi.org/10.1207/s15327752jpa6601_2

Saha, S., Chant, D., Welham, J., & McGrath, J. (2005). A Systematic Review of the Prevalence of Schizophrenia. *PLoS Medicine*, *2*, e141. https://doi.org/10.1371/journal.pmed.0020141

Silventoinen, K., Jelenkovic, A., Yokoyama, Y., Sund, R., Sugawara, M., Tanaka, M., Matsumoto, S., Bogl, L. H., Freitas, D. L., Maia, J. A., Hjelmborg, J. v. B., Aaltonen, S., Piirtola, M., Latvala, A., Calais-Ferreira, L., Oliveira, V. C., Ferreira, P. H., Ji, F., Ning, F., … **Kaprio, J.** (2019). The CODATwins Project: The current status and recent findings of COllaborative Project of Development of Anthropometrical Measures in Twins. *Twin Research and Human Genetics*, *22*, 800–808. https://doi.org/10.1017/thg.2019.35

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Torres, C., Caporali, A., & Pison, G. (2023). The Human Multiple Births Database (HMBD). *Demographic Research*, *48*, 89–106. https://doi.org/10.4054/demres.2023.48.4

Wang, Z., Zellers, S., Piirtola, M., Aaltonen, S., Salvatore, J., Dick, D., Kühn, S., & Kaprio, J. (2025). Gene-environment interplay in the relationship between the visibility of the environment and self-reported depression in early midlife: A Finnish twin cohort study. *medRvix*. https://doi.org/10.1101/2025.05.23.25328215

Zagai, U., Lichtenstein, P., Pedersen, N. L., & Magnusson, P. K. E. (2019). The Swedish Twin Registry: Content and management as a research infrastructure. *Twin Research and Human Genetics*, *22*, 672–680. https://doi.org/10.1017/thg.2019.99

https://doi.org/10.1017/thg.2026.10057 Published online by Cambridge University Press