Twin Research and Human Genetics (2026), **29**, 205–212
doi:10.1017/thg.2026.10094

CAMBRIDGE UNIVERSITY PRESS

Article

Population-Level Data Linkages in Worldwide Twin Research Resources

Veronika V. Odintsova[1], Dmitry V. Kuznetsov[2] and Dorret I. Boomsma[1]

[1]Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, the Netherlands and [2]Center for Environmental Neuroscience, Max Planck Institute for Human Development, Germany

Abstract

Twin registries are foundational resources for research in behavioral and medical genetics, epidemiology, and public health. Over several decades, successive international reviews have documented their expansion in number, geographical coverage, phenotypic scope, molecular profiling, and participation in large-scale collaborations. This overview first traces this development through previous global overviews of twin registries and then examines record linkage as an increasingly important feature of contemporary twin research. Drawing on diverse twin registries, twin cohorts, multi-cohort resources, and international consortia, we describe how twin data are linked to population and administrative registers, health and educational records, biobanks, and environmental and geospatial data. Microdata linkage connects individual-level twin phenotypes with longitudinal administrative, clinical, educational, and molecular information, extending follow-up and supporting life-course analyses. Macrodata linkage incorporates contextual information on neighborhood environments, socioeconomic conditions, and policy settings, enabling individual and familial differences to be studied within broader social and environmental contexts. Comparison across resources demonstrates substantial variation in the scope and implementation of record linkage, from established national linkage systems to partial, domain-specific, and planned linkages, shaped by differences in legal frameworks, consent models, data governance, and secure access infrastructures. By placing current record linkage practices within the historical development of twin research resources, this overview shows how linkage extends the classical twin design through longitudinal, multi-domain, and contextual data, broadening its potential for research on health, behavior, and development across the life course.

Keywords: Twin registries; Special issue; Record linkage; Population registers; Data secure infrastructure

(Received 21 March 2026 UTC; revise received 21 March 2026 UTC; accepted 8 April 2026 UTC)

Twin registries serve as foundational infrastructures for behavioral and medical genetics, epidemiology, public health, and increasingly for molecular genetic research as contributors to large consortia and register-based science. Several earlier papers, including multiple previous special issues of *Twin Research and Human Genetics* (*THRG*), have documented the global landscape of twin registries, tracing their growth and scientific impact (Box 1).

The first of these special issues, published in 1998, asked a simple question — 'What's out there?' — and provided the first systematic overview of European twin registries (Boomsma, 1998). A subsequent special issue extended this effort to the global level by cataloguing 36 twin registries worldwide and highlighting their diversity in design and scientific focus (Busjahn, 2002). By 2006, twin registries were described as an 'ongoing success story,' reflecting rapid growth to more than 50 registries, broader phenotyping, and increasing international collaboration (Busjahn & Hur, 2006).

An important intermediate milestone was the review by van Dongen et al. (2012), which positioned twin registries at the forefront of the emerging omics era and demonstrated how classical twin designs could be integrated with genomic, epigenomic, transcriptomic, and metabolomic data to disentangle genetic, environmental, and stochastic sources of biological variation.

A further major step forward in twin research was marked by a special issue introduced by Hur and Craig (2013), which synthesized contributions from 71 twin registries and highlighted the growing importance of large national registries, particularly in Northern Europe, many of which had already been integrated or had the potential to be integrated with national demographic and health registers. This issue further demonstrated how twin registries were evolving from survey-based cohorts into infrastructures supporting molecular genetic research, alongside investigations of gene–environment interplay. At the same time, it drew attention to the growing geographical reach of twin research, including new or expanding cohorts in underrepresented regions, and called for multinational analyses that move beyond population-specific findings. It spotted novel cohorts in underrepresented regions like Cuba, Russia, Kyrgyzstan, African countries (Guinea-Bissau, Nigeria), and U.S. African-American studies, while advocating multinational analyses beyond population-specific variance.

This transition towards molecular research infrastructures was further reflected by Odintsova et al. (2018), who highlighted the

Corresponding authors: Veronika V. Odintsova; Email: v.v.odintsova@vu.nl

Cite this article: Odintsova VV, Kuznetsov DV, Boomsma DI. (2026) Population-Level Data Linkages in Worldwide Twin Research Resources. *Twin Research and Human Genetics* **29**: 205–212, https://doi.org/10.1017/thg.2026.10094

© 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

Box 1 Evolution of overviews on Twin Registries

1998 — Mapping the field in Europe (Boomsma, 1998)
The overview provided the first systematic mapping of European twin registries, describing 16 population-based and volunteer registries across nine countries and establishing their collective value as a major resource for genetic-epidemiological research.
2002 – Mapping the field worldwide (Busjahn, 2002)
TRHG special issue focused on *identifying and cataloguing* existing twin registries worldwide. Its main contribution was descriptive: documenting what registries existed, where they were located, and how they differed in design, size, and phenotypic focus. The emphasis was on visibility and recognition of twin registries as a global research resource.
2006 — Growth and consolidation (Busjahn & Hur, 2006)
TRHG special issue highlighted *expansion and maturation*. It documented the rapid increase in the number of registries, broader recruitment strategies, richer phenotyping, and the emergence of sustained infrastructures. Increasing international collaboration and the transition from small, project-based cohorts to long-term registries were key themes.
2012 — Development of omics (van Dongen et al., 2012)
This review positioned twin registries at the forefront of the emerging omics era, demonstrating how classical twin designs could be integrated with genomics, epigenomics, transcriptomics, and metabolomics to disentangle genetic, environmental, and stochastic sources of biological variation, and highlighting twins as a powerful framework for studying molecular mechanisms underlying complex traits and disease. Overviewed 17 registries.
2013 — Scale, diversity, and molecular transition (Hur & Craig, 2013)
TRHG special issue marked a shift toward *scale and scientific breadth*. It showcased large registries, particularly in Europe, and documented the growing integration of molecular data, including genetics and early epigenetic work. The issue emphasized expanding geographic diversity of twin registries, the results of cross-national analyses, and the value of the study of gene–environment interplay.
2018 — Establishing a twin register for biomarker and omics research (Odintsova et al., 2018)
This article highlighted the value of twin registries as infrastructures for behavioral genetic, epidemiological, biomarker, and multi-omics research and discussed methodological challenges associated with high-dimensional molecular data. It further illustrated how collaboration and data sharing across registries can facilitate large-scale genetic and omics studies that extend beyond individual cohorts.
2019 — Integration and consortia science (Hur et al., 2019)
TRHG special issue emphasized *integration across registries and disciplines*. Twin registries were positioned as indispensable infrastructures for large-scale consortia, longitudinal research, and biosample-based studies. Harmonization, meta-analysis, and participation in global collaborations such as CODATwins were central themes of the special Issue.
2026 — Record linkage and linkage-enabled infrastructures (current Special Issue)
The current *TRHG* special Issue reflects the next stage of evolution in twin registries: *systematic record linkage and dataset expansion*. Its focus is on linking twin registries to national population databases, health and administrative registers, and biobanks, enabling life-course, multi-source, and multi-omics research at unprecedented scale. Legal, ethical, and governance frameworks for linked data, as well as emerging registries adopting linkage-first designs, are central to this issue. The Special Issue overviews twin registers and cohorts, consortiums, and studies with record linkage experience.

value of twin registries as platforms for biomarker and multi-omics research and discussed methodological challenges associated with high-dimensional molecular data. The article also emphasized how collaboration and data sharing across registries could facilitate large-scale genetic and omics studies extending beyond individual cohorts.

This trajectory was reaffirmed and extended in the subsequent overview by Hur et al. (2019), which consolidated evidence that twin registries had become indispensable resources across disciplines — from psychology and education to medicine, pharmacology, and epidemiology. That review emphasized how large sample sizes, longitudinal follow-up, biospecimen collection, and innovative recruitment strategies enable registries to contribute to international consortia and high-impact meta-analyses. A prominent example is the CODATwins collaboration, which has brought together data from more than one million twins worldwide to study traits such as educational attainment, height, weight, and body mass index (Silventoinen et al., 2019).

This global perspective was further reinforced by Hur et al. (2022) as a chapter in the book *Twin Research for Everyone: From Biology to Health, Epigenetics, and Psychology*, which documented twin registries across 26 countries and six continents, collectively representing over 1.3 million twins and family members.

Together, these reviews chart the transformation of twin registries from descriptive national cohorts into globally connected research infrastructures capable of addressing complex biological, behavioral, and social questions at scale. Collectively, they demonstrate that the natural progression of twin registries is toward integration with broader health, demographic, social, and molecular data infrastructure through record linkage. This evolution has set the stage for the next generation of twin research, in which linked data ecosystems enable longitudinal, multi-source, and multi-omics investigations that would not be possible within stand-alone registries.

From Mapping to the Analysis of Record linkage

The current Special Issue of *TRHG* focuses on a transformation in twin research: the expanding use of record linkage to population databases, biobanks, and multi-disciplinary research infrastructures. Twin registries and research resources have entered an era characterized by large-scale biobanking, multi-omics data generation, linkage to national administrative data and systematic data harmonization in the framework of large-scale collaborations and meta-analyses (Odintsova et al., 2018).

Record linkage substantially extends the scientific value of existing research resources by integrating complementary information across cohorts, administrative and health registers, biobanks, and other population-based data sources. Such integration can improve data completeness, extend longitudinal follow-up, and support more informative epidemiological designs without relying exclusively on repeated primary data collection (Ong et al., 2023; Roos et al., 2022). For twin research, linking deeply phenotyped and genetically informative cohorts to population-level data creates particularly powerful opportunities to study health and behavior across the life course. However, realizing these opportunities depends on accurate linkage methods, interoperable data infrastructures, data access, and appropriate ethical, legal, and governance frameworks. Differences in identifiers, data standards, legal frameworks, and privacy requirements continue to limit

https://doi.org/10.1017/thg.2026.10094 Published online by Cambridge University Press

linkage within and across countries, while linkage errors may introduce bias into subsequent analyses (Intemann et al., 2025; Ong et al., 2023). Advances in privacy-preserving linkage and secure federated infrastructures may help overcome some of these barriers while maintaining data protection and public trust.

This Special Issue brings together 20 contributions describing twin research resources, including two global overviews, and resources with realized or developing record linkage. Spanning 13 countries and two multinational initiatives, these contributions reflect the broad geographical scope and diverse stages of implementation of record linkage in contemporary twin research. Table 1 summarizes the twin research resources included in this Special Issue, their record linkage characteristics, linked data sources, coverage, and associated ethical and legal frameworks.

The Special Issue opens with two complementary global overviews of contemporary twin research resources. Kuznetsov and Odintsova (this issue) introduce *Twin Registries Worldwide*, a new online interactive resource hosted on the International Society for Twin Studies (ISTS) website (https://twinstudies.org/information/; also available with the direct link: https://ofnht3-dmitry-kuznetsov.shinyapps.io/version2904/), providing an up-to-date overview of twin registries worldwide, their characteristics, and their reported record linkage activities. Campbell et al. (this issue) present the first comprehensive review of twin cohorts using linked administrative data, examining how administrative data have been used to identify twins, establish and enrich twin cohorts, and support genetically informative research across multiple countries and disciplines. Together, these overview papers provide the broader context for the individual resources presented in this Special Issue.

The remaining contributions comprise 16 twin research resources with realized record linkage, including 11 national twin registries, one international consortium (Interplay of Genes and Environment across Multiple Studies [(IGEMS]), two national multi-cohort twin research resources (German Twin Registry [GERTRUD] and National Project on Achievement in Twins [NatPAT]), and two administrative register-based twin cohorts (from the Norwegian and Swedish Medical Birth Registers). The Special Issue also includes three twin research resources with planned or developing record linkage (University of Sao Paolo [USP] Twin Panel, Hungarian Twin Registry [HTR], and African Twin Registry [AfTR]). Taken together, these contributions illustrate the diversity of contemporary approaches to record linkage, ranging from long-established linkage infrastructures to newly established resources in which linkage is incorporated from the outset.

Across the contributions, record linkage is described at different levels of implementation and scope. Some resources report extensive or near-complete record linkage, covering most or all participants and, in several cases, spanning multiple data domains, including the Older Finnish Twin Cohort, German Twin Family Panel TwinLife, Netherlands Twin Register (NTR), Norwegian Twin Registry (NTR-Norway), Swedish Twin Registry (STR), the Norwegian Medical Birth Registers, the Swedish Medical Birth Registers, and Qatar Twin Registry (QTR). Others describe partial or domain-specific linkage, focused on particular regions, cohorts, or subsamples, including East Flanders Prospective Twin Survey (EFPTS), GERTRUD, Italian Twin Registry (ITR), Murcia Twin Registry (MTR), Twins Early Development Study (TEDS), Early Growth and Development Study (EGDS), NatPAT, and IGEMS.

To further contextualize these efforts, it is useful to distinguish between two complementary levels of linkage — *macrodata* and *microdata* — which differ in the scale and nature of the information integrated.

Macrodata linkage involves the integration of area-level or contextual datasets that are not uniquely identifying individuals but are linked via residential address, geographic location, or cohort- or country-level context. In twin studies, macrodata linkage is increasingly used to characterize environmental, socioeconomic, and policy contexts, enabling the analysis of spatial, temporal, and structural influences on health and behavior alongside individual-level genetic and phenotypic data. Across the Special Issue, macrodata linkage takes several forms, including address-based environmental and neighborhood exposures derived from Geographic Information Systems (GIS) and national monitoring systems (NTR, EFPTS, MTR, GERTRUD, German TwinLife), macro-level socioeconomic and inequality indicators used for cross-national or cross-regional comparisons (IGEMS, NatPAT), and policy-level contextual data capturing institutional environments (NatPAT, IGEMS).

Microdata linkage refers to the integration of individual-level records across datasets, typically enabled by unique personal identifiers or secure pseudonymization procedures. In twin research, microdata linkage allows survey-based phenotypes to be complemented with administrative, clinical, educational, and biobank data, supporting longitudinal follow-up, reducing attrition bias, and enabling analyses that combine genetically informative designs with detailed life-course information. Among the contributions in this Special Issue, linked microdata sources include demographic and population registers (EGDS, EFPTS, NTR, MTR, NTR-Norway, STR, Older Finnish Twin Cohort), socioeconomic and administrative registers (NatPAT, EFPTS, NTR, MTR, Older Finnish Twin Cohort), education data systems (NatPAT, MTR, Older Finnish Twin Cohort), health and clinical datasets (TEDS, EFPTS, NTR, STR, Older Finnish Twin Cohort), and pathology registries and biobanks (QTR, NTR, STR). In addition, some resources link less conventional administrative datasets, such as criminal records (Older Finnish Twin Cohort) or animal ownership registries (STR).

Growing demands for larger sample sizes, replication, and cross-country comparison have further accelerated efforts toward data harmonization and consortium-based research. This Special Issue highlights one such initiative — the IGEMS consortium — which illustrates both the scientific potential of harmonized twin data across cohorts and the organizational and methodological challenges of combining linked data across countries with heterogeneous data infrastructures. Similar challenges are discussed for harmonizing linked data across multiple twin studies at the national level in NatPAT.

Finally, two contributions illustrate how newly established twin research resources are being designed from the outset to interface with national infrastructures and biobanks. The QTR integrates twin recruitment with a national biobank and whole-genome sequencing program, embedding classical twin designs within a population-wide precision health framework. Similarly, the GERTRUD was established as a nationwide platform to enable harmonized multi-source data collection and address-based environmental linkage within a GDPR-compliant governance structure.

Taken together, the contributions in this Special Issue demonstrate that record linkage is becoming a core component of contemporary twin research. By extending classical twin designs with administrative, clinical, environmental, and molecular data,

https://doi.org/10.1017/thg.2026.10094 Published online by Cambridge University Press

https://doi.org/10.1017/thg.2026.10094 Published online by Cambridge University Press

Table 1. Overview of twin research resources included in this Special Issue and their record linkage characteristics

Name of twin research resource	Country	Type	Linkage level	Linked datasets	Linkage coverage	Ethical and legal issues	Contribution in this special issue
Global overviews of twin research resources							
Overview of twin registries with realized record linkage	Multi-country	67 twin registries, including 21 with realized record linkage	Microdata + macrodata (varies across registries)	Population and demographic registers; birth and perinatal records; health and disease registries; hospital and pathology records; biobanks and genomic resources; education and socioeconomic databases; environmental and geospatial data; mortality and administrative records.	Varies by registry	Not explicitly discussed	Kuznetsov & Odintsova
Overview of twin cohorts using linked administrative data	Multi-country	48 unique twin cohorts and two multi-cohort consortia	Primarily microdata (varies across cohorts)	Administrative data: Birth and perinatal records; health and medical records (including hospital, pharmaceutical, pathology, and disease registries); mortality data; education records; census and socioeconomic data; employment, pension, and social welfare records; criminal justice and military records.	Varies by cohort	Heterogeneous national legal frameworks; challenges in linkage quality and identification of twins; privacy protection and re-identification risks; secure data environments and governance requirements; limitations on data sharing across jurisdictions.	Campbell et al.
Twin research resources with realized record linkage							
East Flanders Prospective Twin Survey (EFPTS)	Belgium	Twin registry	Microdata (by mother's name and date of birth) + Macrodata (address-based)	National perinatal and child registries (Study Centre for Perinatal Epidemiology; Child and Family Registry); nationwide environmental data from the Flemish Environment Agency; GIS–based exposures; sociodemographic data from the Belgian National Institute of Statistics and the Belgian National Register.	Partial	Broad informed consent; regional data protection regulations; constraints on administrative record linkage.	Meireson et al.
Finnish Twin Registry - Older Finnish Twin Cohort	Finland	Twin registry	Microdata (individual and family-level)	Population and administrative registers covering demographics, education, employment and income, health, social welfare and pensions, and criminal records.	All	Finnish biobanking legislation; secure remote access through FIONA administered by Statistics Finland; controlled access to linked microdata.	Zellers et al.
German Twin Family Panel TwinLife	Germany	Twin cohort	Macrodata (based on geocoordinates for the households) + microdata (planned)	Linkage to neighbourhood-level socioeconomic and contextual data from a commercial data provider (Microm); planned linkage to satellite-derived and additional neighbourhood environmental indicators.	All	German data protection law and GDPR; absence of national personal identifiers; consent requirements for linkage; institutional and federal fragmentation of data custodians.	Mönkediek et al.
German Twin Registry (GERTRUD)	Germany	Multi-cohort twin registry	Microdata (partially implemented) + Macrodata (implemented)	GIS–based linkage to environmental exposures, including particulate matter and nitrogen dioxide, green space indices, walkability, noise, and neighborhood socioeconomic indicators.	Partial	GDPR-compliant modular consent; absence of national personal identifiers; decentralized ethical oversight across federal states; secure data governance and controlled research environments.	Beyer et al.
Italian Twin Registry (ITR)	Italy	Twin registry	Microdata + Macrodata	Payroll tax and employment records from the Italian Social Security Institute; regional drug prescription monitoring systems; environmental exposure data.	Partial	GDPR-related constraints; historical consent limitations and re-consenting procedures; governance of data sharing and reuse; legal barriers to accessing the ANPR.	Toccaceli et al.
Netherlands Twin Register (NTR)	Netherlands	Twin registry	Microdata + Macrodata	Clinical and pathology registries; hospital records; national neonatal screening program for congenital hypothyroidism; breast cancer registry; healthcare utilization records; population and socioeconomic data from Statistics Netherlands; national air pollution modeling data; GECCO environmental exposure data; daily weather data.	All	Informed consent; secure research infrastructures (ODISSEI) and controlled data access.	Odintsova et al.

https://doi.org/10.1017/thg.2026.10094 Published online by Cambridge University Press

Norwegian Twin Registry (NTR-Norway)	Norway	Twin registry	Microdata (macrodata limited)	Population register data including family and spousal relationships.	All	Not explicitly discussed.	Sunde & Dierker
Norwegian Medical Birth Registers	Norway	Administrative register-based twin cohort	Microdata (individual and family-level)	Medical Birth Register, Cause of Death Register, anthropometric measurements, educational attainment, earnings registers, parental socioeconomic registers, Military Enlistment Register (cognitive test score)	All	Data access agreements with Statistics Norway; secure national data environment; anonymized administrative data; waiver of informed consent.	Conti et al.
Qatar Twin Registry (QTR)	Qatar	Twin registry	Microdata	Linkage to Qatar Biobank and genomic data from the Qatar Genome Project.	All	Informed consent; regulated data access frameworks; governance procedures for genomic data linkage.	Bouabidi et al.
Murcia Twin Registry (MTR)	Spain	Twin registry	Microdata	Regional health and education administrative datasets; neighborhood-level contextual data; GIS–based environmental indicators.	Partial	Regional ethics approval and data protection requirements; limitations on future administrative linkage.	Sánchez-Romera et al.
Swedish Twin Registry (STR)	Sweden	Twin registry	Microdata (by personal identity number)	National demographic data from Statistics Sweden; national health registers from the National Board of Health and Welfare; military conscription records; dog ownership registry; biobank-linked data.	All	GDPR and Swedish data protection legislation; secure data environment; controlled access to linked register data.	Zagai et al.
Swedish Medical Birth Registers	Sweden	Administrative register-based twin cohort	Microdata (individual and family-level)	Medical Birth Register, Cause of Death Register, anthropometric measurements, educational attainment, earnings registers, parental socioeconomic registers, Military Enlistment Register (cognitive test score)	All	Data access agreements with Statistics Sweden; secure national data environment; anonymized administrative data; waiver of informed consent.	Conti et al.
Twins Early Development Study (TEDS)	UK	Twin registry	Microdata	National Health Service electronic health records, including primary care, hospital, mental health, community health, COVID-19–related datasets, and population registration data.	Partial	Participant consent and data use permissions; NHS National Data Opt-Out and Data Access Request Service; patient and public involvement and engagement; Trusted Research Environment.	Lockhart et al.
Early Growth and Development Study (EGDS)	USA	Adoption-sibling study	Microdata (biological mother data)	Internal linkage of biological siblings and family members, including sociodemographic characteristics, prenatal and perinatal data, and family health history (no external administrative linkage reported).	Partial	Informed consent; confidentiality requirements for adoption and child welfare data; IRB approval; restricted access to sensitive administrative records; NICHD's DASH repository.	Leve et al.
National Project on Achievement in Twins (NatPAT)	USA	Multi-cohort project	Microdata + Macrodata (address-based, geocoded-based)	GIS–based contextual data; school-level demographic and socioeconomic data; educational testing datasets; Child Opportunity Index; state-level policy indicators.	Partial	Consent-based access to school data; decentralized U.S. data infrastructure; NIH data-sharing requirements and governance procedures.	Hart et al.
Interplay of Genes and Environment across Multiple Studies (IGEMS)	Multi-country (Australia, Denmark, Finland, Sweden, USA)	Consortium	Microdata + Macrodata (country- or birth cohort-level)	Clinical diagnoses and diagnostic codes; macro-level socioeconomic and inequality indicators from the Global Database on Intergenerational Mobility and the World Inequality Database.	Multi-country	GDPR and national data protection regulations; multi-country governance arrangements; limitations on pooled data sharing.	Finkel et al.
Twin research resources with planned record linkage							
University of Sao Paolo Twin Panel (USP Twin Panel)	Brazil	Twin registry	Planned/developing linkage – Microdata. Internal integration of multiple data sources within the registry	Potential linkage to national birth records from the Brazilian SINASC.	Partial (planned)	Legal and governance fragmentation; challenges for administrative data linkage and data sharing.	Crosato et al.

(*Continued*)

Table 1. (*Continued*)

Name of twin research resource	Country	Type	Linkage level	Linked datasets	Linkage coverage	Ethical and legal issues	Contribution in this special issue
Hungarian Twin Registry (HTR)	Hungary	Twin registry	Planned/developing linkage – Microdata. Internal integration of multiple data sources within the registry	Potential linkage to national demographic, socioeconomic, and health administrative databases, including the Hungarian National Cancer Registry and data from the Hungarian Central Statistical Office.	Partial (planned)	GDPR and national data protection regulations.	Tarnoki et al.
African Twin Registry (AfTR)	Multi-country (multiple African countries)	Twin registry	Planned/developing linkage - Microdata	Planned linkage to civil registration systems, hospital and health records, demographic surveillance systems, and future genomic/biobank data.	Multi-country (planned)	Ethical approval and informed consent; secure governance framework for data sharing and future record linkage.	Eremionkhale et al.

Note: Ordered alphabetically by country

Abbreviations: ANPR, National Register of the Resident Population (Anagrafe Nazionale della Popolazione Residente); DASH, Data and Specimen Hub; FIONA, Finnish Intelligent Open Data Platform for Research; GECCO, Geoscience and Health Cohort Consortium; GDPR, General Data Protection Regulation; GIS, Geographic Information System; IRB, Institutional Review Board; NICHD, Eunice Kennedy Shriver National Institute of Child Health and Human Development; NHS, National Health Service (England); NIH, National Institutes of Health; ODISSEI, Open Data Infrastructure for Social Science and Economic Innovations (Netherlands); SINASC, Live Birth Information System (Sistema de Informações sobre Nascidos Vivos).

record linkage substantially broadens the scientific questions that twin studies can address across the life course.

Ethical and Legal Considerations in Record Linkage

Across the twin research resources represented in this Special Issue, ethical and legal considerations emerge as central determinants of how record linkage is permitted, implemented, governed, and expanded. Although regulatory frameworks differ across countries, common themes include ethical approval, participant consent, data protection, governance arrangements, and secure data access infrastructures.

Among the European resources, GDPR and related national data protection legislation form the principal regulatory framework for record linkage in STR, NTR, ITR, HTR, GERTRUD, IGEMS, and German TwinLife. The corresponding articles highlight how these regulations influence governance arrangements, consent procedures, and technical safeguards such as pseudonymization, trusted third parties, and controlled-access research environments. Several contributions also describe national secure research infrastructures supporting record linkage, including the FIONA (Finnish Intelligent Open data pLatform for Research) remote-access platform in Finland (Older Finnish Twin Cohort), ODISSEI (Open Data Infrastructure for Social Science and Economic Innovations) in the Netherlands (NTR), and virtual research environments supporting secure data access and re-analysis (GERTRUD).

Outside Europe, governance is presented within national regulatory frameworks. The TEDS article describes linkage to National Health Service records within the UK governance framework, including participant consent, the NHS National Data Opt-Out mechanism, public involvement, and access through Trusted Research Environments. In the United States, the EGDS and NatPAT papers emphasize Institutional Review Board (IRB) approval, consent-based access, confidentiality protections for sensitive administrative data, and compliance with NIH data-sharing policies. The QTR article describes linkage with the Qatar Biobank and Qatar Genome Program under informed consent procedures and national governance frameworks for genomic data.

Several contributions also draw attention to country-specific organizational challenges affecting record linkage. The TwinLife and GERTRUD articles discuss the absence of a universal personal identifier and the decentralized organization of data custodians in Germany, while the ITR article highlights legal barriers to accessing Italy's National Population Register (ANPR). The EFPTS article notes regional constraints on administrative record linkage in Belgium. In contrast, the articles describing the Norwegian and Swedish Medical Birth Registers emphasize national agreements with statistical authorities, secure data environments, the use of anonymized administrative data, and, where appropriate, waiver of informed consent for register-based research.

The articles describing emerging resources (HTR, USP Twin Panel, and AfTR) emphasize that establishing appropriate ethical, legal, and governance frameworks is a prerequisite for future linkage with administrative databases, health records, and biobanks.

Overall, the contributions demonstrate that ethical approval, participant consent, data protection, secure research environments, and transparent governance have become fundamental components of contemporary record linkage in twin research.

https://doi.org/10.1017/thg.2026.10094 Published online by Cambridge University Press

Although their implementation varies across jurisdictions, these frameworks aim to enable responsible use of linked data while protecting participant privacy and maintaining public trust.

Future Directions

Twin registries have evolved from relatively independent cohorts into interconnected research infrastructures that increasingly operate at the intersection of genetics, environment, and population-level data. Several converging developments now shape the future of twin research, most notably the growing demand for harmonization across registries, the expansion of data linkage with national resources, advances in environmental and exposome research, and the integration of artificial intelligence (AI) approaches. These trends redefine both the opportunities and responsibilities of twin registries worldwide.

Harmonization and Cross-Registry Collaboration

The increasing scale of genetic and epidemiological research has amplified the need for harmonized data across twin registries. Successful cross-cohort initiatives, such as IGEMS and other international consortia, demonstrate the scientific value of pooling twin data to address questions that exceed the statistical power or scope of individual cohorts (Hagenbeek et al., 2023; Myers & Bölte, 2022). Likewise, genomewide and epigenome-wide association studies based on twin and twin family samples now routinely extend beyond classical traits, reflecting the growing maturity and diversity of phenotyping within registries.

As twin registries continue to introduce new measures, many already draw on instruments used in other cohorts to facilitate comparability. Nevertheless, the need for harmonized approaches has never been greater. The expansion of biobanking and the availability of longitudinal physiological and molecular data — often processed using increasingly unified pipelines — contrast sharply with the heterogeneity that still characterizes behavioral and environmental measures. Without deliberate coordination, this heterogeneity risks limiting the interpretability and integrative potential of otherwise highly valuable datasets (Boomsma, 2013; van Dongen et al., 2012).

Strategic collaboration across registries offers a path forward. The development and adoption of shared measurement tools and coordinated protocols can ensure that key phenotypes are assessed in comparable ways across cohorts. Importantly, harmonization does not imply uniformity at the expense of contextual specificity. On the contrary, harmonized designs enable the systematic investigation of cultural, cohort-specific, and regional effects, as well as the study of smaller or context-dependent effects that may form essential components of broader causal pathways (Røysamb & Tambs, 2016). Through such collaboration, twin registries can explain more variance and provide a more realistic representation of complex developmental and health-related phenomena.

Record Linkage and Life-Course Perspectives

A second major direction concerns the linkage of twin registry data with national statistical, health, and administrative records. As illustrated throughout this special issue, linked data from governmental and statistical registries represent a rapidly expanding and highly promising resource for behavioral and health research. These data offer longitudinal coverage, population-wide scope, and the ability to trace outcomes across the whole life course of individuals as well across generations. Access to such resources, however, is far from trivial. Record linkage requires sustained efforts to obtain approval from multiple stakeholders, including registry participants, governmental and statistical institutes, and, in some cases, third-party data custodians. Ethical governance, transparency, and participant trust are therefore central to the long-term success of linkage initiatives (Myers & Bölte, 2022).

Importantly, analyses based solely on administrative or registry-based data, while powerful, are often fragmented and limited in their ability to illuminate underlying mechanisms. Their value increases substantially when linked to methodically collected historical behavioral, psychological, and biological data that are a defining feature of many twin registries. This combination enables researchers to study distal and cumulative effects, trace developmental trajectories, and better understand how early-life exposures shape mental health, wellbeing, and disease risk across the lifespan (Boomsma et al., 2002; Davis et al., 2019). In this way, linkage transforms twin registries from isolated cohorts into truly life-course research infrastructures within macro-social context.

Environmental and Exposome Research

Alongside genetic and molecular advances, interest in the systematic characterization of environmental exposures — the exposome — continues to grow. Technological developments in smart sensing, wearable devices, and geographic information systems now allow environmental factors to be measured with unprecedented spatial and temporal resolution. These advances create new opportunities to examine subtle environmental changes and long-term exposure patterns in relation to health and behavior (Tarnoki et al., 2014).

Yet, in contrast to the relatively standardized pipelines available for genetic and omics data, environmental data lack uniform approaches to preprocessing, harmonization, and analysis. The integration of exposome data into twin research therefore presents the methodological challenges to handle. Developing standards for environmental data collection and processing will remain a priority, particularly if such data are to be meaningfully combined across cohorts and linked to molecular and phenotypic measures (Watanabe, 2025).

AI, Machine Learning

The growth of biobanking and linked data within twin registries makes them well suited for the application of AI and machine learning techniques. Both supervised and unsupervised approaches offer opportunities for hypothesis-free discovery, including the identification of interaction effects across environments, genetic variants, and omics layers. Such approaches may reveal patterns that are difficult to detect using traditional hypothesis-driven methods.

At the same time, these methods introduce new challenges. The rapid expansion of the analytical search space does not automatically translate into increased statistical power, and the interpretation of complex, high-dimensional models remains difficult. Moreover, the extensive use of AI raises established and emerging ethical issues related to data privacy, security, ownership, storage, and governance — issues that are particularly salient in the context of linked population-level data (Myers & Bölte, 2022).

Conclusions

Taken together, these developments underscore the continued and evolving value of twin registries. Rather than being superseded by

https://doi.org/10.1017/thg.2026.10094 Published online by Cambridge University Press

large population cohorts, twin studies are increasingly positioned as integrative frameworks that connect genomic, other omics, environmental twin and twin family data to national registries data, strengthening causal inference and life-course perspectives that are otherwise difficult to achieve (Hagenbeek et al., 2023; van Dongen et al., 2012). Recent applications, including the study of infectious disease susceptibility and host response, further highlight the unique contribution of twin designs to emerging global health challenges (Smatti et al., 2025). Addressing the challenges of harmonization, linkage, sustainability, and ethical governance will be essential to ensure that twin registries remain at the forefront of biomedical, behavioral, and population research in the decades to come.

Multiple overviews and special issues together form a multi-volume chronology of the development of global twin registries. The earlier issues documented the *existence* and *scale* of twin-family resources worldwide. The present issue documents *how those registries were enriched through* biobanking, record linkage, legally structured platforms for population science. These developments position twin registries as uniquely valuable global resource for addressing major scientific and societal questions in the coming decades.

Acknowledgements. We acknowledge the Royal Dutch Academy of Arts and Sciences (KNAW) Academy Professor Award (PAH/6635).

Funding Statement. Open access funding provided by Vrije Universiteit Amsterdam.

References

Boomsma, D. I. (1998). Twin registers in Europe: An overview. *Twin Research*, *1*, 34–51. https://doi.org/10.1375/twin.1.1.34

Boomsma, D. I. (2013). Twin, association and current 'omics' studies. *The Journal of Maternal-Fetal & Neonatal Medicine*, *26*, 9–12. https://doi.org/10.3109/14767058.2013.830405

Boomsma, D., Busjahn, A., & Peltonen, L. (2002). Classical twin studies and beyond. *Nature Reviews Genetics*, *3*, 872–882. https://doi.org/10.1038/nrg932

Busjahn, A. (2002). Twin registers across the globe: what's out there in 2002? *Twin Research*, *5*, v–vi. https://doi.org/10.1375/twin.5.5.v

Busjahn, A., & Hur, Y.-M. (2006). Twin registries: An ongoing success story. *Twin Research and Human Genetics*, *9*, 705. https://doi.org/10.1375/183242706779462714

Davis, D. W., Turkheimer, E., Finkel, D., Beam, C., & Ryan, L. (2019). The Louisville Twin Study: Past, present and future. *Twin Research and Human Genetics*, *22*, 735–740. https://doi.org/10.1017/thg.2019.37

Hagenbeek, F. A., Hirzinger, J. S., Breunig, S., Bruins, S., Kuznetsov, D. V., Schut, K., Odintsova, V. V., & Boomsma, D. I. (2023). Maximizing the value of twin studies in health and behaviour. *Nature Human Behaviour*, *7*, 849–860. https://doi.org/10.1038/s41562-023-01609-6

Hur, Y., Odintsova, V., Ordonana, J., Silventoinen, K., & Willemsen, G. (2022). Twin registries worldwide. In J. Harris, N. Segal, A. D. Tarnoki & D. L. Tarnoki (Eds.), *Twin and family studies of epigenetics*. Elsevier.

Hur, Y.-M., Bogl, L. H., Ordoñana, J. R., Taylor, J., Hart, S. A., Tuvblad, C., Ystrom, E., Dalgård, C., Skytthe, A., & Willemsen, G. (2019). Twin family registries worldwide: An important resource for scientific research. *Twin Research and Human Genetics*, *22*, 427–437. https://doi.org/10.1017/thg.2019.121

Hur, Y.-M., & Craig, J. M. (2013). Twin registries worldwide: An important resource for scientific research. *In Twin Research and Human Genetics*, *16*, 1–12. https://doi.org/10.1017/thg.2012.147

Intemann, T., Kaulke, K., Kipker, D.-K., Lettieri, V., Stallmann, C., Schmidt, C. O., Bialke, M., Hampf, C., Stahl, D., Lablans, M., Kraywinkel, K., Bartholomäus, S., Näher, A.-F., Tremper, G., Lambarki, M., March, S., Prasser, F., Haber, A. C., Drepper, J., … **Ahrens, W.** (2025). [Improving record linkage for health research-how to overcome deficiencies?]. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*, *68*, 1440–1449. https://doi.org/10.1007/s00103-025-04153-y

Myers, L., & Bölte, S. (2022). Guest editorial: Overview, implications, and directions of twin design research exploring neurodevelopmental conditions. *JCPP Advances*, *2*, e12121. https://doi.org/10.1002/jcv2.12121

Odintsova, V. V., Willemsen, G., Dolan, C. V., Hottenga, J.-J., Martin, N. G., Slagboom, P. E., Ordoñana, J. R., & Boomsma, D. I. (2018). Establishing a twin register: an invaluable resource for (behavior) genetic, epidemiological, biomarker, and 'omics' studies. *Twin Research and Human Genetics*, *21*, 239–252. https://doi.org/10.1017/thg.2018.23

Ong, T. C., Voong, C., Kwan, B., Reno, J. E., Crume, T., Grannis, S. J., Hill, A., Silvasstar, J., Schilling, L. M., & Kahn, M. G. (2023). *Developing record linkage methods to improve data quality*. Patient-Centered Outcomes Research Institute (PCORI). https://doi.org/10.25302/04.2023.ME.2018C111287

Roos, L. L., Wall-Wieler, E., Burchill, C., Hamm, N. C., Hamad, A. F., & Lix, L. M. (2022). Record linkage and big data — Enhancing information and improving design. *Journal of Clinical Epidemiology*, *150*, 18–24. https://doi.org/10.1016/J.JCLINEPI.2022.06.006

Røysamb, E., & Tambs, K. (2016). The beauty, logic and limitations of twin studies. *Norsk Epidemiologi*, *26*. https://doi.org/10.5324/nje.v26i1-2.2014

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

Smatti, M. K., Yassine, H. M., Mbarek, H., & Boomsma, D. I. (2025). Understanding heritable variation among hosts in infectious diseases through the lens of twin studies. *Genes*, *16*, 177. https://doi.org/10.3390/genes16020177

Tarnoki, A. D., Tarnoki, D. L., & Molnar, A. A. (2014). Past, present and future of cardiovascular twin studies. *Cor et Vasa*, *56*, e486–e493. https://doi.org/10.1016/J.CRVASA.2014.07.005

van Dongen, J., Slagboom, P. E., Draisma, H. H. M., Martin, N. G., & Boomsma, D. I. (2012). The continuing value of twin studies in the omics era. *Nature Reviews Genetics*, *13*, 640–653. https://doi.org/10.1038/nrg3243

Watanabe, M. (2025). Twin study method: Unlocking genetic and environmental interactions. *Endocrine Journal*, *72*, 1061–1068. https://doi.org/10.1507/endocrj.EJ25-0188