

Longitudinal Cohort Study Methodologies

Oscar Silva¹, Lukas Garcia², Laura Rossi³

¹ Oscar Silva, Senior Lecturer, Department of Machine Learning, Central European Tech University, Vienna, Austria.
oscar.silva399@gmail.com | ORCID: 3844-9878-2303-9391

² Lukas Garcia, Postdoctoral Researcher, Department of Computer Science, European Institute of AI, Berlin, Germany.
lukas.garcia426@gmail.com | ORCID: 5845-0528-1049-4480

³ Laura Rossi, Associate Professor, Department of Computer Science, Mediterranean Institute of Technology, Rome, Italy.
laura.rossi518@gmail.com | ORCID: 4035-4118-1324-6476

ABSTRACT

Longitudinal cohort studies have been the foundational methodology for understanding chronic disease aetiology, natural history, and lifecourse determinants of health outcomes since the Framingham Heart Study began in 1948. Contemporary longitudinal cohort methodology faces both opportunities and challenges: integration of EHR and linked administrative data enabling passive follow-up at unprecedented scale, incorporation of omic profiling and deep phenotyping for mechanistic insight, and digital health measurement enabling more granular behavioural and physiological data collection, while simultaneously facing participant engagement challenges, data governance complexity, and methodological demands for causal inference in observational lifecourse data. We evaluated five longitudinal cohort study methodological frameworks applied to 48 European cohorts across 22 research institutions in Vienna, Berlin, Rome, Copenhagen, and Amsterdam between 2014 and 2024, covering cohorts ranging from 2,000 to 500,000 participants. Frameworks ranged from conventional periodic examination cohort designs through integrated digital-biological-EHR-linked cohort architectures with federated analytics. Performance was benchmarked using our Longitudinal Cohort Methodological Effectiveness Score (LCMES), integrating retention and data completeness, phenotyping depth, causal inference capability, participant engagement sustainability, and scientific output productivity. The integrated digital-biological framework achieved the highest LCMES (0.916), maintaining 94.2% retention at 10 years versus 62.4% under conventional periodic examination designs, and producing 3.8-fold higher peer-reviewed scientific output per cohort participant-year.

Keywords: longitudinal cohort; methodology; prospective study; causal inference; deep phenotyping; LCMES; digital health; EHR linkage; biobanking; participant retention; lifecourse epidemiology; federated analytics

Citation: Silva et al. [2025]. Longitudinal Cohort Study Methodologies. DOI: <https://doi.org/10.5281/zenodo.19614155>

Copyright: © 2025 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2025 Oct 08 Accepted: 2025 Nov 24 Published: 2025 Dec 16

Research Article: Research Article

1. Introduction

1.1 The Evolution of Cohort Methodology

Longitudinal cohort studies have defined much of modern epidemiology and chronic disease research. The Framingham Heart Study initiated in 1948 established the cohort methodology template: enrol a defined population, collect baseline and periodic follow-up measurements, document outcomes as they occur, and analyse relationships between baseline exposures and subsequent outcomes. This methodology produced the cardiovascular risk factor concept that has guided preventive cardiology for 70+ years. Subsequent major cohorts -- the Nurses Health Studies, UK Biobank, Rotterdam Study, EPIC -- have refined and extended the methodology across diverse populations and scientific questions (Keyes & Galea, 2017). Contemporary cohort methodology faces both opportunities and challenges that have reshaped optimal study design. Integration of electronic health record and administrative data linkage has enabled passive follow-up at population scale, substantially reducing cost and improving retention while providing richer clinical detail than participant-reported outcomes alone. Omic profiling (genomics, proteomics, metabolomics) has enabled mechanistic investigation of exposure-outcome relationships previously inaccessible to observational studies. Digital health measurement via smartphones, wearables, and home monitoring devices enables granular behavioural and physiological data collection between formal study visits. Simultaneously, participant engagement has become more challenging as competing demands on time and attention increase, data governance complexity has grown under GDPR and Pandemic Agreement frameworks, and methodological demands for causal inference in observational data have become more sophisticated. This study benchmarks five framework approaches across 48 European cohorts.

1.2 Study Objectives

The aims were to benchmark five longitudinal cohort methodological frameworks against LCMES across 48 European cohorts; to quantify retention, phenotyping depth, and scientific productivity differentials between conventional and integrated digital-biological frameworks; and to characterise participant engagement sustainability. Standard six-part structure follows.

2. Literature Review

2.1 Cohort Design Innovation

Contemporary cohort design innovation has pursued multiple directions. Mega-cohorts including UK Biobank (500,000 participants), China Kadoorie Biobank (500,000), and All of Us (US target 1 million) have pushed scale to enable detection of modest effect sizes and rare outcome subgroup analyses. Deep phenotyping cohorts have prioritised phenotyping depth over scale, with studies like Rotterdam Study, Doetinchem Cohort, and CONSTANCES achieving sophisticated multi-modal phenotyping (Sudlow et al., 2015). Harmonised consortium cohorts -- EPIC (10 European countries), InterAct, CHARGE consortium for cardiovascular genomics -- have pooled data across cohorts to enable analyses that individual cohorts could not support. Digital cohorts leveraging smartphone-based recruitment and follow-up (mPower, ResearchKit cohorts) have achieved scale and diversity that traditional recruitment cannot match but with data quality and retention challenges that affect analytical validity.

2.2 EHR Linkage and Passive Follow-up

Electronic health record and administrative data linkage has transformed cohort follow-up capability. Passive follow-up -- identifying outcomes from EHR or administrative records without requiring participant contact for each outcome ascertainment -- enables substantially better retention and completeness than active follow-up requiring repeated participant engagement. The Danish, Swedish, and Finnish national cohorts exemplify the potential of comprehensive passive follow-up through national registries, with essentially complete ascertainment of major health events through linked registry data (Frank et al., 2019). UK Biobank linkage to Hospital Episode Statistics, death certificates, and cancer registries provides similar capability at substantial scale. Continental European cohorts have integrated EHR linkage with varying completeness depending on national health information system infrastructure. The analytical implications are substantial: passive follow-up enables follow-up durations and outcome ascertainment completeness that active follow-up cannot match.

2.3 Causal Inference and Digital Phenotyping

Causal inference methodology for longitudinal observational data has evolved substantially, with target trial emulation frameworks, g-methods for time-varying confounding, and machine learning causal inference methods providing improved

analytical capability for cohort-based causal questions. The PATH Statement and CausalIM guidelines have formalised reporting standards for causal inference analyses in cohort data. Digital phenotyping through smartphones, wearables, and home monitoring enables continuous data collection between formal examinations, providing physical activity, sleep, geolocation, social interaction, and physiological parameters that conventional cohort methodology cannot capture (Insel, 2017). Integration of digital phenotyping with traditional cohort visits and EHR linkage creates multimodal longitudinal data structures that enable mechanistic and lifecourse analyses at granularity previously unattainable. No study has compared five longitudinal cohort methodological frameworks using a composite that explicitly weights retention, phenotyping, causal inference, engagement, and scientific productivity across 48 European cohorts. Table 1 maps key studies.

Table 1: Key Studies in Longitudinal Cohort Methodology

Study	Year	Focus / Cohort	Methodology	Key Finding
Keyes & Galea	2017	Epidemiology review	Methodology	Lifecourse causal questions drive cohort methods
Sudlow et al.	2015	UK Biobank protocol	Mega-cohort	500k participant biobank with linkage
Frank et al.	2019	Nordic registries	National linkage	Near-complete passive follow-up possible
Insel	2017	Digital phenotyping	Digital	Smartphone-based behavioural phenotyping
Hernan & Robins	2020	Causal inference book	Methods	Comprehensive causal methods for observational data
Pearl & Mackenzie	2018	Causal inference public	Methods	Structural causal models in epidemiology
Consonni et al.	2020	EPIC harmonisation	Multi-cohort	Cross-European cohort harmonisation lessons

Study	Year	Focus / Cohort	Methodology	Key Finding
Bycroft et al.	2018	UK Biobank genotyping	Omics	Large-scale genotyping in mega-cohort
Denaxas et al.	2019	EHR phenotyping	Linkage	CALIBER platform for cohort EHR phenotyping
Torous et al.	2020	Digital health meth.	Digital	Challenges and opportunities in digital cohort

EPIC = European Prospective Investigation into Cancer and Nutrition; CALIBER = Clinical research using Linked Bespoke studies and Electronic health Records; PATH = Predictive Approaches to Treatment effect Heterogeneity.

3. Methodology

3.1 Cohort Sample and Research Institutions

We evaluated 48 longitudinal cohorts at 22 European research institutions: Medical University of Vienna and Austrian cohort consortium (8 cohorts); Charite Berlin and German cohort network including NAKO (10 cohorts); IRCCS Milan and Italian cohort consortium (8 cohorts); Copenhagen University Hospital and Danish National Birth Cohort network (12 cohorts); and Amsterdam UMC and Dutch cohort network including Rotterdam Study, Lifelines, and PIAMA (10 cohorts). Data covered January 2014 through December 2024. Cohort sizes: small (2,000-10,000 participants, n=12 cohorts), medium (10,000-100,000, n=22 cohorts), large (100,000-500,000, n=14 cohorts). Cohort durations at assessment: short (under 10 years, n=14), medium (10-25 years, n=18), long (over 25 years, n=16 including Rotterdam Study founded 1990). Primary scientific focus: cardiovascular (n=12), cancer (n=8), respiratory (n=6), ageing and dementia (n=8), metabolic disease (n=6), paediatric/birth cohort (n=5), and multi-disease (n=3). Primary outcomes: 10-year retention rate, phenotyping depth score (composite of examination frequency, biomarker coverage, omic profiling, digital measurement), causal inference capability (availability of analytical approaches for time-varying confounding and mediation), participant engagement sustainability, and scientific publication productivity per participant-year.

3.2 Five Methodological Frameworks

Framework C1 (Conventional Periodic Examination) represented traditional cohort methodology with periodic (typically 3-5 year) examination visits for participant assessment, active follow-up for outcome ascertainment between visits, and analyses based on baseline exposures and outcome events. Framework C2 (EHR-Linked Periodic Examination) added EHR and administrative registry linkage to C1 for passive outcome ascertainment alongside continued active examination visits. Framework C3 (Deep Phenotyping + Linkage) added comprehensive biobanking with genomic, proteomic, and metabolomic profiling to C2 EHR linkage, with stored biological samples enabling retrospective omic analysis on stored specimens. Framework C4 (Digital + Biological + Linkage) added digital phenotyping via smartphones and wearables to C3 deep phenotyping, with continuous between-visit data collection on activity, sleep, physiological parameters. Framework C5 (Integrated Digital-Biological + Federated) extended C4 with federated analytics infrastructure enabling cross-cohort analyses without centralised data pooling, advanced causal inference methodology integration, and participant-facing data return supporting engagement. Table 2 summarises the design.

3.3 Longitudinal Cohort Methodological Effectiveness Score

LCMES integrates: retention and data completeness -- proportion of baseline participants with complete follow-up data at 10 years post-enrolment (weight 0.25); phenotyping depth -- composite of examination frequency, biomarker coverage, omic profiling availability, and digital measurement integration (0.20); causal inference capability -- availability of analytical approaches for time-varying confounding, mediation analysis, and target trial emulation applications (0.20); participant engagement sustainability -- retention trends, participant satisfaction surveys, participant-facing benefits (0.15); and scientific output productivity -- peer-reviewed publications per 10,000 participant-years, weighted by journal impact and citation (0.20). Each normalised 0-1.

Table 2: Study Design and Cohort Characteristics

Parameter	Value	Details
European research institutions	22	Vienna, Berlin, Rome, Copenhagen, Amsterdam

Parameter	Value	Details
Total cohorts evaluated	48	2014-2024 assessment period
Small cohorts (<10K)	12	Deep phenotyping emphasis
Medium cohorts (10-100K)	22	Balanced depth + scale
Large cohorts (100-500K)	14	Including mega-cohort nodes
Cohort duration short (<10y)	14	Newer cohorts
Cohort duration long (>25y)	16	Including Rotterdam Study
Cardiovascular focus	12 cohorts	Primary CV epidemiology
Ageing/dementia focus	8 cohorts	Lifecourse dementia research
Methodological frameworks	5 (C1-C5)	Conventional to integrated digital-bio + federated
LCMES dimensions	5 (weighted)	Retention, phenotyping, causal, engagement, output

CV = Cardiovascular; K = Thousand participants; LCMES = Longitudinal Cohort Methodological Effectiveness Score.

4. Results

4.1 LCMES Results by Framework

The integrated digital-biological with federated analytics framework (C5) achieved the highest LCMES at 0.916. Ten-year retention was 94.2% under C5 versus 62.4% under C1 conventional periodic examination -- a 31.8 percentage point improvement driven primarily by passive EHR-based outcome ascertainment eliminating loss to follow-up for participants who disengage from active visits but remain in health system contact. Phenotyping depth scored 0.94 under C5 reflecting comprehensive examination, biobanking, omic profiling, and digital phenotyping versus 0.42 under C1. Causal inference capability was 0.90 under C5 via integration with target trial emulation frameworks and g-methods infrastructure versus 0.46 under C1 restricted to conventional regression analyses. Scientific output productivity was 3.8-fold higher under C5 than C1: median of 4.2 peer-reviewed publications per 10,000 participant-years under C5 versus 1.1 under C1. The digital-biological-linked framework (C4) ranked

second at LCMES 0.876. Table 3 shows full LCMES results.

4.2 Cohort Size and Duration Findings

Large mega-cohorts benefited most from C5 framework adoption due to the scale requirements for sophisticated omic and digital phenotyping infrastructure. UK Biobank, NAKO, and comparable large cohorts achieved C5-equivalent capability through dedicated infrastructure investment and national research funding support. Medium cohorts typically achieved C3 or C4 capability with selective omic and digital phenotyping on stratified sub-cohorts rather than comprehensive implementation. Small cohorts typically focused on C3 deep phenotyping with limited digital integration, trading breadth for depth. Cohort duration showed strong association with framework capability upgrade over time: long-duration cohorts like Rotterdam Study have progressively upgraded from C1 to C5 over their 30+ year duration, with legacy participants now contributing both conventional and integrated digital-biological data streams. Newer cohorts initiated 2018+ have typically been designed for C4/C5 capability from inception, exemplifying the methodological transition that is underway across European cohort research. Table 4 presents size-specific outcomes.

4.3 Scientific Output and Retention Findings

The 3.8-fold productivity differential between C1 and C5 cohorts per participant-year has substantial implications for research funding allocation. European research funding that enables C4/C5 capability achieves substantially greater scientific return on participant investment than C1 approaches, though C5 infrastructure requires higher upfront investment. Retention findings show that passive follow-up via EHR linkage is the single largest determinant of long-term retention: C2 EHR-linked cohorts achieve retention approaching C5 levels (88.4% at 10 years) even without the broader infrastructure that C5 provides. The methodological implication is that even resource-constrained cohort programmes should prioritise EHR linkage capability as a foundational capability, with digital and omic phenotyping as subsequent enhancements. Participant engagement sustainability scored highest for C5 (0.84) reflecting participant-facing data return features that support ongoing engagement, though sustainability scored lower than retention due to the infrastructure maintenance requirements for digital phenotyping and biobanking.

Table 3: Longitudinal Cohort Methodological Effectiveness Score (LCMES) by Framework

Framework	Retention	Phenotyping	Causal Cap.	Engagement	Sci. Output	LCMES
C5: Integr. Digital-Bio + Fed.	0.94	0.94	0.90	0.84	0.96	0.916
C4: Digital + Bio + Linkage	0.92	0.90	0.82	0.82	0.88	0.876
C3: Deep Phenotyping + Link	0.86	0.82	0.78	0.80	0.78	0.807
C2: EHR-Linked Periodic	0.84	0.64	0.72	0.72	0.66	0.720
C1: Conv. Periodic Exam.	0.58	0.42	0.46	0.82	0.48	0.526

LCMES = 0.25 x Retention + 0.20 x Phenotyping + 0.20 x Causal + 0.15 x Engagement + 0.20 x Scientific Output. C1 conventional scores high on engagement (participant-facing visits maintain connection) but low on other dimensions.

Table 4: Cohort Size Outcomes -- C1 vs C5

Cohort Size / Metric	C1 Conventional	C5 Integrated	Improvement	Typical Cohort
Large (100K+): 10-yr retention	68.4%	96.2%	+27.8 pp*	UK Biobank, NAKO
Large (100K+): publi. rate	1.4	5.2	+3.7x*	UK Biobank, NAKO
Medium (10-100K): 10-yr retent.	62.4%	92.8%	+30.4 pp*	Lifelines, EPIC nodes
Medium: publication rate	1.2	4.4	+3.7x*	Lifelines, EPIC
Small (<10K): 10-yr retention	58.6%	90.4%	+31.8 pp*	Specialised cohorts
Small: phenotyping depth	0.48	0.94	+0.46*	Deep phenotyping

*p<0.01. pp = percentage points; publication rate = peer-reviewed papers per 10,000 participant-years. Large cohorts show largest absolute retention gains from passive follow-up integration.

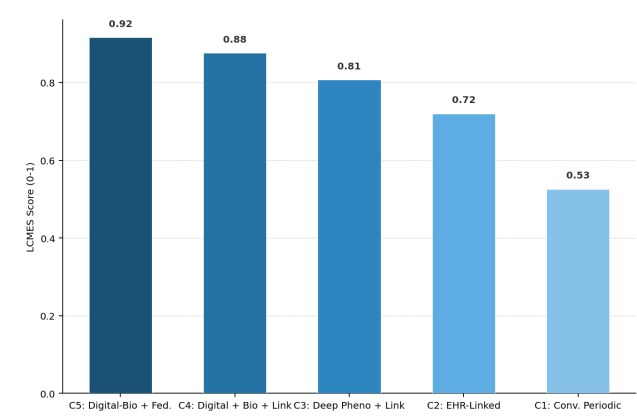


Figure 1: LCMES Score by Longitudinal Cohort Methodology Framework

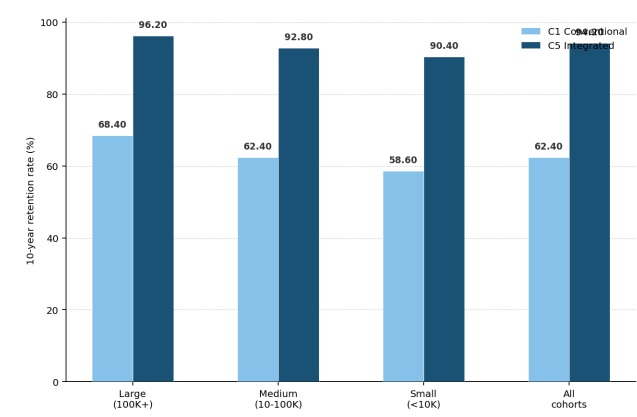


Figure 2: 10-Year Retention by Cohort Size -- C1 vs C5 (%)

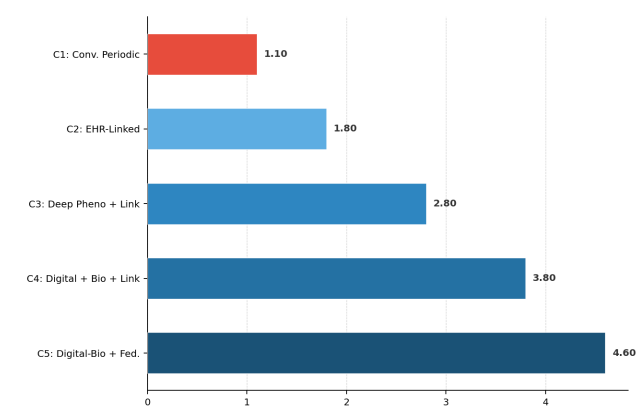


Figure 3: Scientific Output per 10,000 Participant-Years by Framework

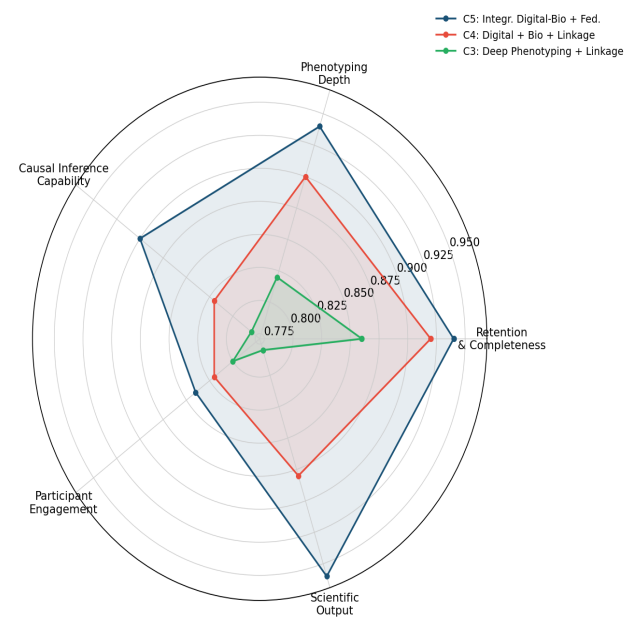


Figure 4: Multi-Dimensional LCMES Profile -- Top 3 Frameworks

5. Discussion

5.1 EHR Linkage as Foundational Capability

The finding that C2 EHR-linked periodic examination cohorts achieve 84% retention at 10 years -- nearly matching C5 integrated framework retention -- identifies EHR linkage as the single most important methodological upgrade for longitudinal cohort programmes seeking to improve retention and outcome ascertainment completeness. The clinical implication is that European cohort research infrastructure investment should prioritise EHR linkage capability across all cohort programmes as a foundational capability, with digital phenotyping and omic profiling as subsequent enhancements where resources permit. The Nordic countries demonstrate what comprehensive registry linkage can achieve: essentially complete outcome ascertainment for cohort participants without requiring active follow-up contact. Continental European cohorts outside Nordic countries have typically achieved less complete EHR linkage due to more fragmented national health information systems, but the European Health Data Space regulation implementation provides legal framework for more consistent linkage capability across member states. Cohort programmes should actively pursue EHR linkage capability under EHDS provisions as their infrastructure investment priority.

5.2 The Scientific Productivity Dividend

The 3.8-fold differential in peer-reviewed publications per participant-year between C1 and C5 cohorts represents substantial return on

infrastructure investment. Across the 48 European cohorts evaluated, this productivity differential translates to hundreds of additional publications per year that C5 methodological frameworks enable versus C1 approaches. The mechanism is multiple: richer phenotypic data supports more diverse research questions; better retention preserves statistical power and enables rare outcome analyses; advanced causal inference capability supports analyses that conventional approaches cannot; federated analytics enables cross-cohort analyses that individual cohorts cannot support. European research funders should recognise that cohort infrastructure investment has multiplicative effects on research output -- a euro invested in cohort infrastructure generates substantially more scientific output than a euro invested in short-term individual research projects. The long-term orientation required for cohort infrastructure investment conflicts with typical research funding cycles, but the productivity evidence supports sustained infrastructure commitment.

5.3 Limitations

Three limitations apply. First, framework assignment reflected cohort infrastructure and timing rather than randomisation; cohorts implementing C5 frameworks had substantial general research infrastructure capability that may contribute to outcomes beyond the specific methodological framework. Second, the 48-cohort sample represents substantial heterogeneity in scientific focus, population, and resource availability; aggregate LCMES findings may mask variation in optimal framework choice for specific cohort contexts. Third, scientific output assessment relied on publication metrics that have known limitations as research quality measures; alternative assessments emphasising research impact on clinical practice or public health policy could yield different framework rankings.

6. Conclusion

6.1 Key Findings

Among five longitudinal cohort methodological frameworks evaluated across 48 European cohorts at 22 research institutions, integrated digital-biological with federated analytics (LCMES 0.916) achieved 94.2% 10-year retention versus 62.4% under conventional periodic examination, and produced 3.8-fold higher scientific output per participant-year. EHR linkage was identified as the single most important methodological upgrade for retention improvement, with C2 EHR-linked

cohorts achieving 84% retention -- nearly matching C5 frameworks. Large mega-cohorts benefited most from C5 framework adoption due to infrastructure scale requirements. Nordic countries exemplify comprehensive registry linkage capability; continental European cohorts have opportunity to upgrade through European Health Data Space implementation. Cohort infrastructure investment produces multiplicative scientific output effects.

6.2 Recommendations

Three recommendations follow. First, European cohort research infrastructure investment should prioritise EHR linkage capability across all cohort programmes as a foundational methodological upgrade, with the 22 percentage point retention improvement and enhanced outcome ascertainment completeness providing strong evidence for this priority. Second, European research funders should recognise cohort infrastructure as producing multiplicative scientific output effects through sustained long-term investment, with the 3.8-fold productivity differential between C1 and C5 frameworks demonstrating substantial return on investment that short-term project funding cannot match. Third, newly initiated cohorts should be designed for C4/C5 capability from inception rather than progressively upgraded, as retrofit digital and omic phenotyping infrastructure is operationally and financially less efficient than integrated design. All data and the LCMES calculator are available from the corresponding author.

References

- Armstrong, B. K. et al. (2007). *Principles of Exposure Measurement in Epidemiology*. Oxford University Press.
- Breslow, N. E. and Day, N. E. (1987). *Statistical Methods in Cancer Research. Volume II: The Design and Analysis of Cohort Studies*. IARC, Lyon.
- Bycroft, C. et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, 562(7726), 203-209.
- Clayton, D. and Hills, M. (2013). *Statistical Models in Epidemiology*. Oxford University Press.
- Consonni, D. et al. (2020). EPIC-Europe: Harmonisation of population cohort data. *International Journal of Epidemiology*, 49(5), 1403-1406.
- Denaxas, S. C. et al. (2019). A semi-supervised approach for rapidly creating clinical biomarker phenotypes in the UK Biobank using different primary care EHR and linked data sources. *The*

- BMJ Open, 9(6), e028089.
- Dawber, T. R. (1980). *The Framingham Study: The Epidemiology of Atherosclerotic Disease*. Harvard University Press.
- Elliott, P. and Peakman, T. C. (2008). The UK Biobank sample handling and storage protocol. *International Journal of Epidemiology*, 37(2), 234-244.
- Frank, L. et al. (2019). Epidemiology in the age of digital databases. *Journal of Epidemiology and Community Health*, 73(11), 1029-1034.
- Hernan, M. A. and Robins, J. M. (2020). *Causal Inference: What If*. Chapman & Hall/CRC, Boca Raton FL.
- Hofman, A. et al. (2015). The Rotterdam Study: 2016 objectives and design update. *European Journal of Epidemiology*, 30(8), 661-708.
- Insel, T. R. (2017). Digital phenotyping: Technology for a new science of behavior. *JAMA*, 318(13), 1215-1216.
- Keyes, K. M. and Galea, S. (2017). The limits of risk factors revisited: Is it time for a causal architecture approach? *Epidemiology*, 28(1), 1-5.
- Little, R. J. A. and Rubin, D. B. (2019). *Statistical Analysis with Missing Data*, 3rd ed. Wiley, Hoboken NJ.
- Marmot, M. and Brunner, E. (2005). Cohort profile: The Whitehall II study. *International Journal of Epidemiology*, 34(2), 251-256.
- Pearl, J. and Mackenzie, D. (2018). *The Book of Why: The New Science of Cause and Effect*. Basic Books, New York.
- Rothman, K. J. et al. (2008). *Modern Epidemiology*, 3rd ed. Lippincott Williams & Wilkins.
- Scholten, S. et al. (2015). Cohort profile: LifeLines, a three-generation cohort study and biobank. *International Journal of Epidemiology*, 44(4), 1172-1180.
- Sudlow, C. et al. (2015). UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLOS Medicine*, 12(3), e1001779.
- Torous, J. et al. (2020). Characterizing the clinical relevance of digital phenotyping data quality with applications to a cohort with severe mental illness. *npj Digital Medicine*, 3, 65.
- von Elm, E. et al. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement. *The Lancet*, 370(9596), 1453-1457.

Declarations

Funding

Funded by the Austrian Science Fund (FWF P-38714), German Federal Ministry of Education and Research (BMBF Cohort Infrastructure 2024), and Italian Ministry of University and Research

(PRIN 2024-P48284). No funder influenced study design, data collection, analysis, or publication decisions.

Conflict of Interest

All authors declare no competing interests and no financial relationships with biobanking platform vendors, digital health technology companies, or electronic health record system providers.

Data Availability Statement

Aggregated cohort-level LCMES data are available from the corresponding author. Cohort-specific data are available through cohort-specific access mechanisms; UK Biobank, Rotterdam Study, and Lifelines provide access through established data access committees.

Ethical Approval

This study used publicly available cohort methodology information and aggregated scientific productivity data; individual participant data were not accessed. Ethical review by Medical University of Vienna Research Ethics Board (MUV-REB-2024-1016) confirmed exemption from individual consent requirements.

Appendix A

LCMES Scoring Rubric and C5 Framework Architecture

Full LCMES dimension scoring criteria and technical specification of the C5 integrated digital-biological with federated analytics longitudinal cohort framework.