International Journal of Epidemiology, 2025, **54(6)**, dyaf200
https://doi.org/10.1093/ije/dyaf200
Data Resource Profile

IEA
International Epidemiological Association
OXFORD

Data Resource Profile

Data resource profile: the Intego-II primary care database

Ahmed Medhat Zayed[1,†], Arne Janssens[1,†], Maarten Caspers[1], Pavlos Mamouris[2], Simon Gabriël Beerten[1], Tine De Burghgraeve[1], Pieter J.K. Libin[3,4], Thomas Neyens[4,5], Nicolas Delvaux[1], Gijs Van Pottelbergh[1], Bert Aertgeerts[1], Bert Vaes[1,*]

[1]Academic Centre for General Practice, Department of Public Health and Primary Care, Faculty of Medicine, KU Leuven, Leuven, Belgium
[2]Real-World Data Solutions, Larissa, Greece
[3]Artificial Intelligence Lab, Department of Computer Science, Vrije Universiteit Brussel, Brussels, Belgium
[4]CenStat, Data Science Institute, Hasselt University, Hasselt, Belgium
[5]L-Biostat, Department of Public Health and Primary Care, Faculty of Medicine, KU Leuven, Leuven, Belgium

*Corresponding author. Academic Centre for General Practice, Department of Public Health and Primary Care, Faculty of Medicine, KU Leuven, Kapucijnenvoer 7 blok g—Box 7001, 3000 Leuven, Belgium. E-mail: bert.vaes@kuleuven.be.
[†]= equal contribution.

Keywords: electronic medical records; health data; primary care; research database.

Key Features

- Intego-II builds on three decades of primary care data collection in Flanders, Belgium, incorporating substantial advancements in the database's structure, operations, and accessibility.
- Established in 1994, the database collects pseudonymized electronic medical record (EMR) data from participating general practices. Its integration with Healthdata.be provides scalable linkage to mortality, environmental, and disease-specific datasets at the national level.
- A robust two-step Extract-Transform-Load (ETL) process ensures data security, privacy, tidiness, and quality. To enhance international research interoperability, the database is aligned with the OMOP Common Data Model.
- Intego-II is organized into three key modules: Patient Information, Medical History, and Clinical Encounters, enabling longitudinal analyses across diverse healthcare domains covering, among others, demographic variables, diagnoses, prescriptions, and laboratory test results.
- Structured quarterly releases with unique Digital Object Identifiers (DOIs) and detailed metadata ensure findability and reusability. Researchers can access Intego-II via a secure research environment provided by Healthdata.be, following submission and approval of a study protocol. The Data Access Process is detailed at www.intego.be.
- As of December 2024, 180 practices have contributed data (of which 135 active practices). The data included 973 444 unique individuals (of which 512 093 active patients). Currently, active patients represent 7.46% of the Flemish population.

Data resource basics

The Intego project, initiated by the Academic Centre for General Practice (ACHG) at KU Leuven, was established to create a comprehensive primary care morbidity database in Flanders, Belgium. Starting with semi-automated data collection in 1994, the project evolved into Belgium's first fully automated general practice registration system by 2004 [1].

Coordinated by KU Leuven's ACHG, primarily funded by the Department of Care of the Flemish government and backed by data collected through software provided by Corilus (Corilus NV, Ghent, Belgium) [2], the success and evolution of the Intego project can be attributed to the effective collaboration between governmental services, academic institutions, and software companies. Involvement of these three main actors has been noted as the most effective strategy to build and sustain long-term, wide-ranging data collection projects [3].

In this Data Resource Profile, we describe the release of Intego-II, which builds on three decades of primary care data collection, incorporating substantial advancements in the database's structure, operations, and accessibility.

Primary care in Belgium

In Belgium, nearly the entire population (99%) is covered by government-funded compulsory health insurance [4–6]. Patients have the freedom to choose their general practitioner (GP), who typically act as the first point of contact for healthcare and refer patients to specialists when needed [6].

Patients can request their chosen GP to maintain a global medical file (GMF)—an electronic health record that

Received: 12 March 2025; **Accepted:** 16 October 2025

consolidates their medical history, promoting continuity of care [5]. As of 2022, 89.9% of the Flemish population had a GMF, and 87% had at least one GP consultation [7]. Loyalty to GPs is high, with 80% of patients consistently visiting the same practice, a trend that has remained stable over time [7].

The Intego-II database

The Intego database, initially described by Truyers *et al.* in 2014, has undergone major updates. In 2018, participating GPs transitioned from the obsolete Medidoc® EMR to the more advanced CareConnect® system. Over time, the database has been broadened by adding more practices to improve geographic coverage and enriched by linking patient-level data to other sources.

To reinforce the role of Intego-II as a vital resource for advancing health research, driving innovation in primary care practice, and informing policy development, substantial advancements have been made to ensure full compliance with FAIR data principles—Findable, Accessible, Interoperable, and Reusable [8].

Data resource area and population coverage

Practices are recruited through outreach by the ACHG, with the goal of achieving broad geographic coverage across Flanders, allowing Intego to expand its network effectively. Figure 1 illustrates the database's growth, showcasing the increase in the number of patients who visited participating practices within a given year—referred to as the yearly contact group (YCG). Additionally, Table 1 presents the main characteristics of the Intego-II population of December 2024 [9].

On 31 December 2024, the database included 512 093 active patients (7.46% of the Flemish population). To validate the distribution of key population demographics, Fig. 2A demonstrates that Intego-II is representative of the Flemish population in terms of gender and age distribution. Figure 2B shows geographical variations in the population coverage, with an overrepresentation in Antwerp and Flemish Brabant and an underrepresentation in West Flanders, East Flanders, and Limburg.

Linked external data sources

Intego-II data are enriched through linking patient-level data with additional external sources, providing a more comprehensive resource for research. The currently linked external data sources include:

- Mortality data:
To ensure more comprehensive data and accurate research studies, Intego-II links patient data with mortality records from the National Register of Natural Persons (RRNP) to obtain the dates of death of deceased patients on a weekly basis [10]. This continuous linkage started in 2021 and included historic mortality data.
- Environmental data:
Since 2017, Intego has been commissioned by the Flemish Department of Care to monitor the impact of environmental pollution on public health [11]. Environmental data are provided by the Flemish Institute for Technological Research (VITO), an independent Flemish research organization, and include annual averages of noise, traffic, air quality, and climate indicators, along with daily measurements of temperature and relative humidity. These indicators are linked to patients' residential locations at various spatial resolutions.
- Diabetes Data Cell project:
The Diabetes Data Cell Project seeks to create a comprehensive database to monitor diabetes trends and improve care quality. Launched in 2024, the project received approval from the information security council to link two major databases with Intego; the Initiative for Quality Improvement and Epidemiology in Diabetes (IQED) database, and data from the Inter-Mutualistic Agency (IMA), a non-profit organization founded by Belgian health insurance funds [12–14].

Data collected

The creation of the Intego-II database involved three distinct steps: data acquisition, transformation, and release (Fig. 3).

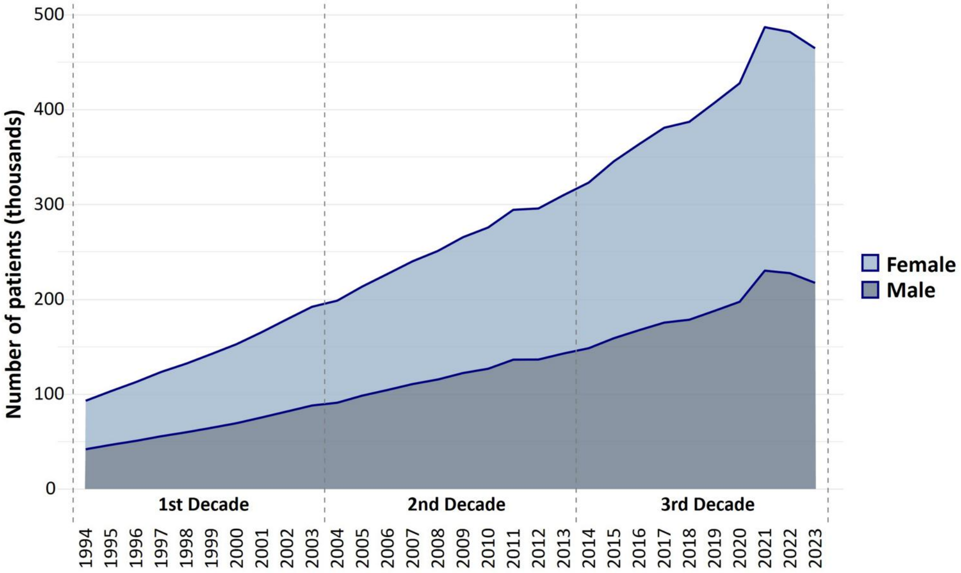


Figure 1. Growth of Intego database over time showing the evolution of the Yearly Contact Group (YCG).

Table 1. Metadata of the Intego-II database as of December 2024

Metadata of Intego-II database—December 2024 ***Version cutoff date: 31 December 2024***			
Total number of contributing practices	180 (*Total*)		135 (*Active*[a])
Total number of patients	973 444 (*Acceptable*[b])		512 093 (*Active*)
Age in years (*Active*)			
Mean (standard deviation)		43.54 (23.91)	
Median (25th and 75th percentile)		42.51 (24.50–62.51)	
Gender (*Active*)			
Female		52.54%	
Male		47.44%	
Unknown		0.02%	
Increased compensation/reimbursement (*Active*)			
No		81.56%	
Yes		14.58%	
Ambiguous or unknown		3.86%	
Geographical distribution of patients (*Active*)			
Flemish Region		97.23%	
Flemish Brabant		18.88%	
Antwerp		36.96%	
Limburg		11.80%	
West Flanders		15.84%	
East Flanders		13.75%	
Brussels-Capital Region		0.27%	
Walloon Region		0.61%	
Ambiguous or unknown		1.89%	
Patient follow-up time[c] in years	*Acceptable*		*Active*
Mean (standard deviation)	10.17 (8.80)		12.44 (9.06)
Median (25th and 75th percentile)	7.81 (2.79–15.87)		10.47 (4.90–18.79)
Encounters per patient per year (*Acceptable*)			
<1 encounters/year		7.40%	
[1–3] encounters/year		33.47%	
[3–5] encounters/year		25.45%	
≥5 encounters/year		33.68%	
Average		4.71 encounters/year	
Clinical data records (*Acceptable)*			
Total number of diagnoses		18 160 392	
Total number of prescriptions		44 755 145	
Total number of lab results		224 227 062	

Adapted from Casper *et al.* (2025).

[a] *Active cohort*: non-deceased patients from the acceptable cohort with at least one registered encounter within 2 years before the version cutoff date.

[b] *Acceptable cohort*: patients with at least two registered encounters and with unambiguous gender and birth date.

[c] *Patient follow-up time*: time between first encounter and death date for deceased patients; time between first encounter and last encounter for non-deceased, non-active patients; time between first encounter and version cutoff date for active patients.

Data acquisition

Data collection for the Intego-II database began in 1994 and has continued uninterrupted. All participating practices (data providers) use the EMR software provided by Corilus. Until 2018, the software used was Medidoc®, which was later replaced by CareConnect®, featuring a more robust technological infrastructure that supports weekly automated data collection and additional coded health data.

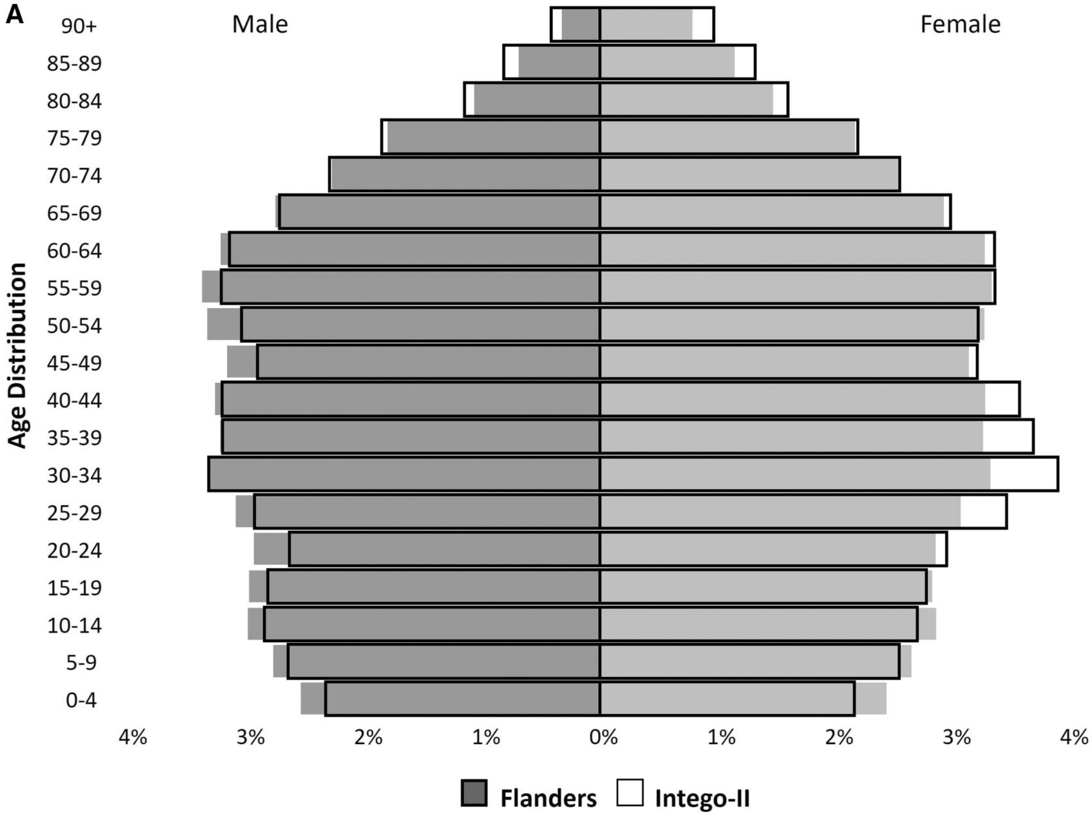
In participating practices, patients are informed of the practice's participation through a document provided in the waiting room. Data collection includes recent weekly data and all historical records stored in the practice's EMR system, but only if the patient has not opted out of data sharing. This approach accounts for the extensive time span of the collected data, with the earliest records in Intego-II dating back to 1924—70 years before the official start of data collection in 1994—and covering nearly a century of data up to the present.

Data transformation

Intego-II undergoes two Extract-Transform-Load (ETL) cycles, each serving a specific purpose: the first ensures data security and privacy, while the second focuses on data tidiness and quality.

First ETL cycle: data security and privacy

The first ETL cycle is managed by Healthdata.be, a Belgian platform developed by Sciensano, a third-party organization responsible for the secure transfer of data from participating practices. The data are loaded into a secure data warehouse, which provides controlled access to researchers via two-factor authentication [15]. The trusted third-party process has been in place since 2012. A key transformation in this cycle is the implementation of end-to-end (E2E) encryption to protect patient information during transmission. To safeguard patient privacy while enabling data linkage across different sources, Healthdata.be employs Belgium's eHealth services to receive pseudonymized national patient identifiers (Social Security Numbers or SSN) [15]. Through the SSN identifier, data from the same patient can be linked across different databases without revealing their identity, thereby maintaining confidentiality. However, in Intego data that were collected using Medidoc software before 2015, SSN identifiers were not included and hence patients in this



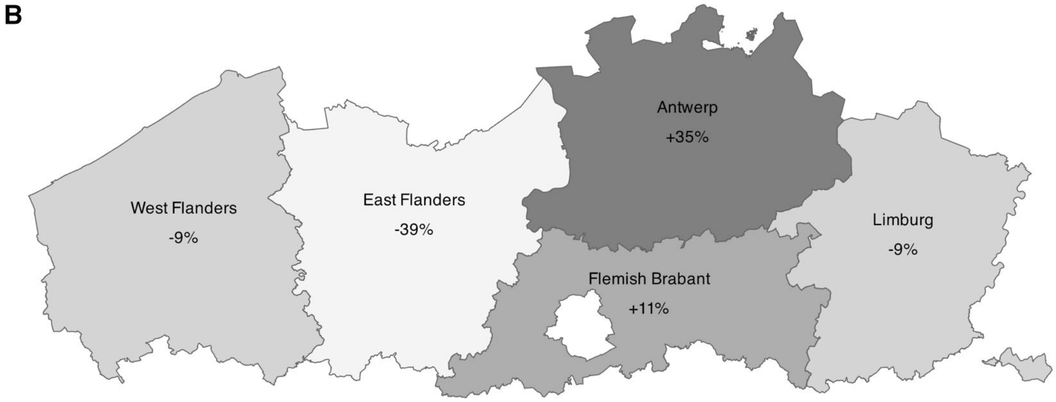


Figure 2. Comparison of the Intego-II population with the general Flemish population. This figure compares the characteristics of the active Intego-II population on 31 December 2023, with the general population of Flanders in 2023, obtained from the Belgian statistical office. (A) Distribution of key demographics, illustrating the representativeness of Intego-II in terms of gender and age distribution. (B) Geographical distribution of the active Intego-II population across the five Flemish provinces, showing deviations from the expected distribution if the Intego-II population mirrored the general population of Flanders. Compared to the expected distribution, there is an overrepresentation in Antwerp and Flemish-Brabant, and an underrepresentation in West-Flanders, East-Flanders, and Limburg.

historic subset are ineligible for linkage to external data sources, such as mortality or environmental data.

In addition to encrypting patient identifiers, a Small Cell Risk Assessment (SCRA) was conducted by a third party to evaluate and mitigate the risk of patient re-identification through indirect identifiers within the database [16]. Measures such as removing and pseudonymizing potential identifiers were implemented, successfully reducing the risk of disclosure while preserving the data's utility for research purposes.

Second ETL cycle: data tidiness and quality

Once the first ETL cycle is completed, the pseudonymized data are loaded into a secure research environment. At this stage, the Intego data management team conducts the second ETL cycle, which focuses on ensuring data tidiness and quality using R scripts.

Data tidiness

The database from Healthdata.be is restructured to improve usability for researchers. Originally designed for clinical consultations, its format is not optimal for research. For example, diagnosis codes, which are collected in two separate tables, are merged into one. The diagnosis data are then divided based on the nature of diagnoses: chronic conditions, which persist over a patient's lifetime, are placed in a dedicated table, while other diagnoses related to short- or medium-term illnesses remain in a separate table. The

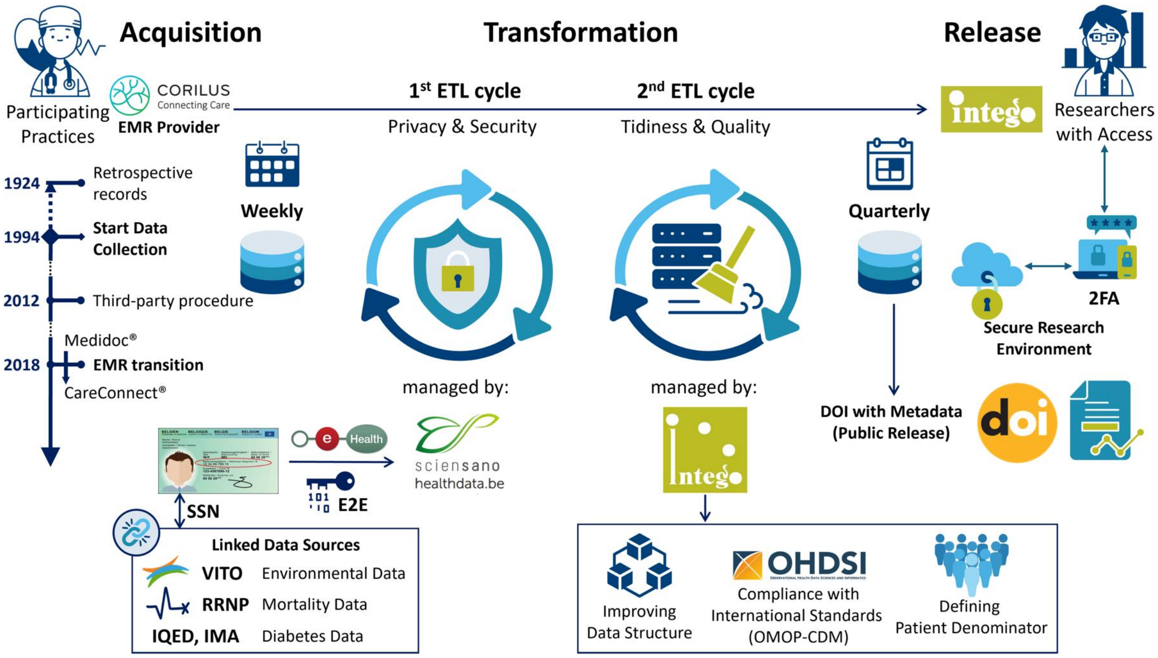


Figure 3. Overview of Intego-II data process. This figure provides an overview of the Intego-II data process, consisting of three main steps. Data acquisition, initiated in 1994 in collaboration with Corilus [the electronic medical record (EMR) software provider], has continued uninterrupted, using Medidoc® until 2018 and CareConnect® thereafter. Practices across Flanders are recruited by KU Leuven's ACHG, with data collection encompassing both historical records—dating back to 1924—and weekly updates. External data sources, including mortality records from the National Register of Natural Persons (RRNP), environmental data from the Flemish Institute for Technological Research (VITO), data related to diabetes from the IQED and the IMA databases, are linked via pseudonymized Social Security Numbers (SSNs) and secured through end-to-end encryption using Belgium's eHealth services. The data undergoes two ETL (Extract, Transform, Load) cycles: the first, managed by Healthdata.be, focuses on encryption, pseudonymization, and integration of external datasets, while the second, managed by the Intego team, emphasizes restructuring, standardization, and adjustments to accurately define the patient population for epidemiological studies. Quarterly data releases are assigned unique DOIs, accompanied by metadata to ensure reproducibility, with access provided to researchers via two-factor authentication in a secure environment. ACHG, Academic Centre for General Practice; DOI, digital object identifier; ETL, extract, transform, load; EMR, electronic medical record; IQED, initiative for quality improvement and epidemiology in diabetes; IMA, InterMutualistic Agency; RRNP, national register of natural persons; SSN, social security number; VITO, Flemish Institute for Technological Research.

classification of chronic diseases follows the methodology of Nielen *et al*. to categorize the International Classification of Primary Care (ICPC-2) codes, adapted for the Belgian healthcare system [17]. This reorganization enhances the database's structure, making analysis more efficient for researchers.

Data quality

To assess the data quality (DQ) of Intego-II, we applied the harmonized data quality assessment framework tailored for the secondary use of EMR data as outlined by Kahn *et al* [18]. A key measure introduced is the concept of "acceptable" patients, which filters data at patient level to improve completeness and accuracy. Acceptable patients must have unique, non-null gender and birth year values. Additionally, acceptable patients must have more than one registered clinical encounter over 2 years before the release date to remove transient cases.

Compliance with international standards

Another key objective of the second ETL cycle is to ensure the database's compliance with international standards for semantic interoperability. To achieve this, we adopted the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) v5.4, an open community data standard designed to harmonize the structure and content of observational healthcare data from diverse sources [19]. By aligning Intego-II with the OMOP CDM, we enable the database's consistent and standardized representation, facilitating interoperability and comparative research across global databases. For data structuring, we developed an ETL script in R to transform Intego-II data into the OMOP CDM format. Additionally, we developed a centralized mapping table to serve as a comprehensive glossary, which aligns both source and standardized codes and vocabularies used across different tables.

Specific considerations for epidemiological studies: defining the patient denominator

For epidemiological studies using EMR data, defining the patient denominator is crucial to ensure accurate surveillance. In the Intego-II database, this involves two steps: selecting eligible practices based on the quality of their morbidity data registration and estimating their population size.

First, practices are evaluated to minimize recording bias by ensuring they meet specific criteria: recording, on average, more than one new diagnosis per patient per year and keeping the proportion of uncoded diagnoses below 20%. This eligibility is reassessed with each data release to update the list of practices eligible for epidemiological studies.

Second, the practice population is estimated for eligible practices. In Belgium, unlike in the Netherlands or the UK, patients are not registered with a specific practice but tend to visit the same practice over time. Therefore, we use a method developed by Bartholomeeusen *et al*., which combines the

Yearly Contact Group (YCG) with correction factors from reimbursement claims data. These factors adjust for individuals who did not visit a GP, stratified by demographics [20]. However, as these corrections have a 2-year delay, researchers should use the "active patient" population—patients with at least one recorded contact in the past 2 years and no death date—for recent data.

Data release

To enhance reproducibility across multiple data extractions, the Intego-II database adopts a structured quarterly release schedule. Each version is assigned a unique Digital Object Identifier (DOI), accompanied by a detailed metadata table that outlines the database's key characteristics and documents its growth across versions. Researchers are required to cite the specific version they accessed, ensuring clarity and consistency in data references. This versioning approach aligns with the findability and reusability principles of the FAIR data framework, promoting transparent and consistent use of the database [8].

Modular structure of data releases

The Intego-II database is organized into three key modules: Patient Information, Medical History, and Clinical Encounters (Fig. 4). These modules are designed to facilitate longitudinal analysis, enabling researchers to link data across all tables using unique patient and practice identifiers, and the date associated with each record.

Data resource use

Over the years, data from Intego have facilitated the production of over 125 research reports to date, contributing valuable insights into multiple facets of primary care research. A significant portion of these studies has focused on epidemiological research, mapping the incidence, prevalence, and trends of diseases ranging from chronic conditions like cardiovascular disorders [21] and diabetes to infectious [22] and mental health issues [23]. Another group of studies has explored healthcare utilization, examining patterns in lab test use [24], vaccine coverage [25], and medication prescribing [26] that inform day-to-day clinical practices. Furthermore, methodological innovations in data cleaning [27], ethical registration, and audit processes [28] have emerged from the Intego network. Finally, a smaller yet impactful group of studies has contributed to public health and health policy research, addressing broader societal and economic outcomes and the effects of major public health events [29]. A full list of publications can be found on the Intego website (www.intego.be). It is important to note that these research domains are not mutually exclusive, and the rich Intego-II dataset holds untapped potential for further exploration in other research domains such as machine learning model development and other emerging domains.

Strengths and weaknesses

Strengths

The Intego-II database draws considerable strength from its breadth of coverage, large sample size, long-term follow-up,

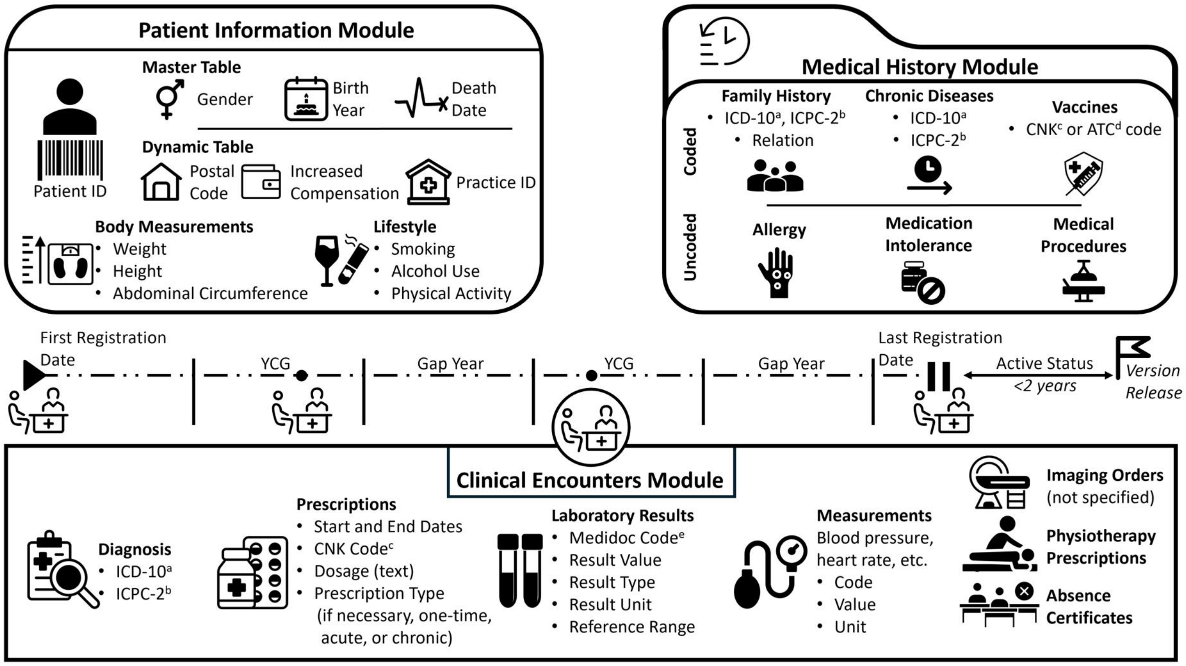


Figure 4. Modular data structure of Intego-II. This figure illustrates the three core modules of the Intego-II database, highlighting their underlying tables (in bold) and key associated variables (column names). The timeline in the middle tracks patient encounters between first and last registration dates illustrating three variables from the Patient Master Table: Yearly Contact Group (YCG), Gap Years, and Active Status (determined based on the duration between last registration date and release date). **The Patient Information Module** serves as the core of the Intego-II database. **The Medical History Module** provides a detailed, long-term view of a patient's medical background, including both coded and uncoded data. Unlike other modules, it allows retrospective entry, meaning recorded dates reflect the actual historical event rather than a clinical encounter. **The Clinical Encounters Module** provides detailed records of patient diagnoses, prescriptions, laboratory test results, measurements, imaging orders, physiotherapy prescriptions, and absence certificates. **Used code systems (vocabularies)**: [a]ICD-10: the International Classification of Diseases, 10th Revision. [b]ICPC-2: the International Classification of Primary Care, 2nd edition. [c]CNK: Medications and vaccines are registered using the Belgian National Code Number (CNK) codes, managed by the Association of Pharmacists of Belgium (APB). [d]ATC: the Anatomical Therapeutic Chemical classification. [e]Medidoc: a coding system for lab tests maintained by Corilus and used in Belgium by many laboratories, EMRs and Healthcare providers.

and representativeness to the Flemish population. It is strengthened by an intensive two-step ETL procedure, which not only enforces strict privacy standards but also enhances the quality, standardization, and usability of the data. Additionally, the robust integration with the Healthdata.be infrastructure provides an efficient, scalable mechanism for enriching Intego-II with further datasets at the Belgian level. This capacity for seamless linkage enables the database to continually expand, offering new insights and supporting in-depth analyses across various domains of population health. Beyond these scalable linkage capabilities, the alignment of Intego-II with the OMOP CDM facilitates interoperability and comparative studies on an international scale, broadening the potential for multidisciplinary collaborations. Taken together, these features underscore the database's value as a reliable resource for advancing health research and informing policy.

Weaknesses

Despite these advantages, several limitations warrant careful consideration. One notable challenge is the reliance on a secure research environment for data access. While this model preserves confidentiality, it can impose computational constraints on researchers, especially those requiring high-performance computing for large-scale or machine-learning–intensive projects. Furthermore, although rigorous quality checks minimize inconsistencies, certain limitations—such as missing values and uncoded entries—remain and may necessitate advanced statistical or imputation techniques. Finally, because Belgium does not mandate strict patient registration with a single general practice, accurately defining denominators and practice populations can be more complex, requiring tailored methodological approaches. In sum, Intego-II offers a rich, multidimensional data source for primary care research, but users must account for these challenges to fully harness its potential.

Data resource access

Researchers interested in accessing the Intego-II database should visit the Intego website (www.intego.be) for comprehensive guidelines on the application process. A preliminary Data Access Request (DAR) must be submitted, detailing the study's objectives, methodology, and authorized users. The Data Access Committee organizes the access process based on the technical and compliance requirements of the secure research environment, ensuring that all users adhere to privacy protocols through a signed Data Access Agreement (DAA) that specifies permitted data uses and terms of compliance. Approved data use is restricted to the specific study described in the DAR, with any additional or modified uses requiring a new request. Any further inquiries can be directed to intego@kuleuven.be.

Ethics approval

The Intego-II database consists of pseudonymized EMR data collected during primary care encounters across participating practices in Flanders. The use of pseudonymized health data for the Intego project was approved by the KU Leuven Ethics Committee (approval number S68557) and the Belgian Information Security Committee, specifically the Social Security and Health Chamber in their decision number 13.026, dated 19 March 2013, with the most recent update in decision number 23.424, dated 5 December 2023 [30]. The latest approval followed a comprehensive review to ensure strict compliance with data protection laws.

Acknowledgements

The authors acknowledge the participating primary care practices for their essential contributions in providing access to data, which constitute the foundation of the Intego-II database. We are also grateful to the patients whose data were included in this study, ensuring adherence to the highest ethical and privacy standards. Finally, we thank all collaborators of the Intego project for their dedication to advancing healthcare research.

Author contributions

AMZ and AJ wrote the initial draft and final manuscript and developed the figures. AMZ designed the initial database structure. AMZ, AJ, MC, and BV collaboratively defined the final Intego-II database structure. AJ, AMZ, MC, and PM contributed to developing the R scripts for data cleaning and organization. MC serves as the data manager of Intego-II and oversees the quarterly data releases. BA is the founder of Intego, and BV is the current principal investigator. TDB manages data acquisition, expanding the Intego database by recruiting GP practices, monitoring contracts, and maintaining network communication. MC, PM, SGB, TBD, PJKL, TN, ND, GVP, BA, and BV reviewed the initial draft and provided input to finalize the manuscript. ND and TN, as supervisors of AMZ and AJ, respectively, provided guidance throughout the project. GVP is the principal investigator working on the Intego environmental project to assess impact of environmental pollution on public health. PJKL contributed to discussions on the current limitations of the database infrastructure and potential future improvements. All authors reviewed and approved the final manuscript.

Supplementary data/code

To maintain consistency and data integrity across updates and transformations, relevant data and code scripts are available online at the Intego-II GitHub repository (https://github.com/Intego-db/Intego-II). The repository includes technical documentation on the secure research environment, as well as a code glossary and the complete mapping table across OMOP-CDM vocabularies with associated frequencies.

Conflict of interest

None declared.

Funding

The Intego project is primarily funded by the Department of Care of the Flemish government. PJKL acknowledges support from the Research council of the Vrije Universiteit Brussel (OZR-VUB) via grant number OZR3863BOF. The funders had no role in study design, data collection, analysis, decision to publish or preparation of the manuscript.

Data availability

See Data resource access above.

Use of artificial intelligence (AI) tools

ChatGPT (OpenAI) was used as a writing assistant to improve the language and clarity of the manuscript. The authors maintained full control over the content, verified the accuracy of all AI-generated suggestions, and take full responsibility for the integrity of the final manuscript.

References

1. Intego. Historiek Intego. *intego.be*. https://www.intego.be/over-intego/historiek (10 March 2025, date last accessed).
2. Corilus NV. https://www.corilus.be/en/ (10 March 2025, date last accessed).
3. Gentil M-L, Cuggia M, Fiquet L *et al.* Factors influencing the development of primary care data collection projects from electronic health records: a systematic review of the literature. *BMC Med Inform Decis Mak* 2017;**17**:139.
4. Cartier T, Ryssaert L, Bourgueil Y. *Building Primary Care in a Changing Europe: Case Studies*. Copenhagen: European Observatory on Health Systems and Policies, 2015.
5. Mathieson A. *Creating 21st Century Primary Care in Flanders and Beyond*. Copenhagen: WHO Regional Office for Europe, 2019, 76.
6. WHO Team European Observatory on Health Systems and Policies. *Belgium: Country Health Profile 2023*. Paris: OECD/European Observatory on Health Systems and Policies, 2023. https://doi.org/10.1787/dd6df7bd-en (10 March 2025, date last accessed).
7. Huisartsencontacten in België. *IMA-AIM*. https://ima-aim.be/Huisartsencontacten-in-Belgie?lang=nl (10 March 2025, date last accessed).
8. Wilkinson MD, Dumontier M, Aalbersberg I *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 2016;**3**:160018.
9. Caspers M, Janssens A, Zayed A, Raat W, Van Pottelbergh G, Vaes B. *Intego-II database December 2024*. KU Leuven RDR, 2025. https://doi.org/10.48804/B1DB2X (10 March 2025, date last accessed).
10. Federale Overheidsdienst Binnenlandse Zaken. *Beslissing nr. 020/2021*. Federale Overheidsdienst Binnenlandse Zaken, 2021. https://www.ibz.rrn.fgov.be/fileadmin/user_upload/nl/rr/machtigingen/2021/Beslissing_020-2021.pdf (10 March 2025, date last accessed).
11. Sectoraal Comité van de Sociale Zekerheid en van de Gezondheid. *SCSZG/18/176 Beraadslaging nr. 18/097*. 2019. https://www.ehealth.fgov.be/ehealthplatform/file/cc73d96153bbd5448a56f19d925d05b1379c7f21/fc311893138543a93a0f877c2fba38772855fcd7/18-097-n098-intego-milieu-gewijzigd-op-7-mei-2019.pdf (10 March 2025, date last accessed).
12. Belgian Information Security Committee, Social Security and Health Chamber. *IVC/KSZG/24/132 Beraadslaging nr. 24/062*. Belgian Information Security Committee, Social Security and Health Chamber, 2024. https://ehealth.fgov.be/ehealthplatform/file/cc73d96153bbd5448a56f19d925d05b1379c7f21/73279f93c3eb849c082c861c6e21c839feaf376b/24-062-n132-diabetes-datacel-project-sciensano.pdf (10 March 2025, date last accessed).
13. Initiative for Quality improvement and Epidemiology in Diabetes. https://www.sciensano.be/en/projects/initiative-quality-improvement-and-epidemiology-diabetes (10 March 2025, date last accessed).
14. IMA-AIM—InterMutualistic Agency. https://www.aim-ima.be/ (10 March 2025, date last accessed).
15. Delvaux N, Aertgeerts B, Bussel JCv, Goderis G, Vaes B, Vermandere M. Health data for research through a nationwide privacy-proof system in Belgium: design and implementation. *JMIR Med Inform* 2018;**6**:e11428.
16. Méroc E. *Small Cell Risk Assessment: Intego*. 2022. https://fair.healthdata.be/sites/default/files/distribution/7bc34c8c-40d1-483f-8da5-2a4b6b0aaace/REPORT_SCRA_INTEGO.pdf (10 March 2025, date last accessed).
17. Nielen MM, Spronk I, Davids R *et al.* Estimating morbidity rates based on routine electronic health records in primary care: observational study. *JMIR Med Inform* 2019;**7**:e11929.
18. Kahn MG, Callahan TJ, Barnard J *et al.* A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. *EGEMS (Wash DC)* 2016;**4**:1244.
19. Hripcsak G, Duke JD, Shah NH *et al.* Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform* 2015;**216**:574–8.
20. Bartholomeeusen S, Kim C-Y, Mertens R, Faes C, Buntinx F. The denominator in general practice, a new approach from the Intego database. *Fam Pract* 2005;**22**:442–7.
21. Smeets M, Vaes B, Mamouris P *et al.* Burden of heart failure in Flemish general practices: a registry-based study in the Intego database. *BMJ Open* 2019;**9**:e022972.
22. Schweikardt C, Goderis G, Elli S, Coppieters Y. Gonorrhoea and syphilis epidemiology in Flemish general practice 2009–2013: results from a registry-based retrospective cohort study compared with mandatory notification. *AIMS Public Health* 2016;**3**:800–22.
23. Walrave R, Beerten SG, Mamouris P *et al.* Trends in the epidemiology of depression and comorbidities from 2000 to 2019 in Belgium. *BMC Prim Care* 2022;**23**:163.
24. Zayed AM, Delvaux N. Analyzing laboratory test utilization trends in belgian primary care: a decade of insights with international perspectives. *BMC Prim Care* 2024;**25**:270.
25. Janssens A, Vaes B, Abels C *et al.* Pneumococcal vaccination coverage and adherence to recommended dosing schedules in adults: a repeated cross-sectional study of the INTEGO morbidity registry. *BMC Public Health* 2023;**23**:1104.
26. Burvenich R, De Boodt S, Lowie L *et al.* Temporal trends in antibiotic prescribing and serious and nonserious infections in children presenting to general practice: a registry-based longitudinal cohort study of 162 507 individuals. *J Antimicrob Chemother* 2024;**79**:1397–406.
27. Mamouris P, Nassiri V, Verbeke G, Janssens A, Vaes B, Molenberghs G. A longitudinal transition imputation model for categorical data applied to a large registry dataset. *Stat Med* 2023;**42**:5405–18.
28. Paridaens R, Vaes B, Van den Bulck S, Soetaert J. Benchmarks for low back pain in general practice in Flanders: electronic audit of INTEGO. *BMC Prim Care* 2024;**25**:431.
29. Petrof O, Neyens T, Vaes B, Janssens A, Faes C. Using a general practice research database to assess the spatio-temporal COVID-19 risk. *BMC Prim Care* 2024;**25**:175.
30. Belgian Information Security Committee, Social Security and Health Chamber. *IVC/KSZG/23/424 Beraadslaging nr. 13/026*. Belgian Information Security Committee, Social Security and Health Chamber, 2023. https://www.ehealth.fgov.be/ehealthplatform/file/cc73d96153bbd5448a56f19d925d05b1379c7f21/9ac9f86db21c28742a16b1481b34ef7b7010a11c/13-026-n424-intego-gewijzigd-op-5-december-2023.pdf (10 March 2025, date last accessed).

International Journal of Epidemiology, 2025, 54, 1–8
https://doi.org/10.1093/ije/dyaf200
Data Resource Profile