Existing Data Sources for Clinical Epidemiology: The PHARMO Database Network

Josephina G Kuiper [ID][1,2]
Marina Bakker[1]
Fernie JA Penning-van Beest[1]
Ron MC Herings [ID][1,3]

[1]Department Research, PHARMO Institute for Drug Outcomes Research, Utrecht, the Netherlands; [2]Department of Public Health, Erasmus University Medical Centre, Rotterdam, the Netherlands; [3]Amsterdam UMC, Vrije Universiteit Amsterdam, Epidemiology and Biostatistics, Amsterdam Public Health Research Institute, Amsterdam, the Netherlands

Abstract: The PHARMO Database Network provides a unique opportunity to gain insight in the complete patient journey and healthcare in the Netherlands. The PHARMO Database Network is a population-based network of electronic healthcare databases and combines anonymous data from different primary and secondary healthcare settings in the Netherlands. Healthcare settings include general practitioners, out-patient and in-patient pharmacies, hospitals and clinical laboratories. Furthermore, databases are linked with external registries such as the Cancer Registry, Pathology Registry and Perinatal Registry. The different data sources are linked on a patient level through probabilistic linkage based on validated algorithms. The longitudinal and ongoing nature of the PHARMO Database Network system enables to follow up more than 10 million residents of the Netherlands for an average of 12 years. Data collection period, catchment area and overlap between data sources differ. Access to the PHARMO Database Network is, by governance regulations of the data collection, restricted to researchers of the PHARMO Institute and academic affiliates. Each data request is checked against privacy and company policies, and requires approval of the privacy and governance board. The terms and conditions and a data application form are available on the PHARMO website (www.pharmo.com).

Keywords: database, record linkage, pharmacoepidemiology, observational research, electronic health record, Netherlands

Introduction

Medicines play an important role in treating many conditions and diseases. However, in daily practice medicines do not always work and some patients are faced within unintended side effects. Longitudinal, multidisciplinary data are needed to study medication use and its challenges in a real-world setting. The PHARMO Database Network was developed to study the safety and effectiveness of medication in daily clinical practice by means of (pharmaco-)epidemiological research with real-world data (RWD) from the Netherlands. Starting in 1990 as the PHARMO Record Linkage System, dispensing records from out-patient pharmacies (ie community pharmacies) were linked to hospital admission records.[1] Drug utilization could thus be linked to clinical outcomes. Over the years, the PHARMO Database Network has been extended with data from other healthcare providers, including general practitioners (GPs), clinical laboratories and in-patient pharmacies, resulting in a patient-centric population-based network of electronic healthcare record (EHR) databases that combines ongoing data collections from primary and secondary care in the Netherlands (Table 1). The PHARMO Database Network provides a unique opportunity to gain insight in the complete patient journey and healthcare in the Netherlands.

Correspondence: Josephina G Kuiper
Tel +31 30 7440 800
Email josine.kuiper@pharmo.nl

Table 1 Databases Included in the PHARMO Database Network

Database	Data Source	Catchment Area[a]	Data Collected	Coding System
General Practitioner (GP) Database (in-house)	Data from electronic patient records registered by GPs of all patients enrolled at the GP	Subnational (~20% of the Dutch population)	• Diagnoses • Symptoms • Laboratory test results • Referrals to specialists • Healthcare product/drug prescriptions	Diagnoses and symptoms: ICPC Drug prescriptions: WHO ATC Classification System
Out-patient Pharmacy Database (in-house)	Data on all GP or specialist prescribed healthcare products dispensed by the community pharmacies	Subnational (~25% of the Dutch population)	• Type of product • Date • Strength • Dosage regimen • Quantity • Route of administration • Prescriber specialty • Costs	Dispensings: National product classification WHO ATC Classification System
In-patient Pharmacy Database (in-house)	Data on all drug dispensing from the hospital pharmacy, given during hospitalization	Subnational (~10% of the Dutch population)	• Type of product • Start and end date of use • Strength • Dosage regimen • Route of administration • Prescriber specialty	Dispensings: National product classification WHO ATC Classification System
Clinical Laboratory Database (in-house)	Results of tests performed on clinical specimens, requested by GPs or specialists	Subnational (~5% of the Dutch population)	• Date and time of testing • Test result • Unit of measurement • Type of clinical specimen	WCIA Coding System LOINC (partly)
Hospital Database (external)	Data on all hospitalizations for more than 24 hours or for which a bed is required, out-patient visits and high budget impact medication. Data are obtained from the Dutch Hospital Data Foundation.[b]	Subnational (80% of hospitals in the Netherlands)	• Diagnoses • In- and out-patient procedures • High budget impact medication • Admission, discharge and visit dates	Diagnoses: WHO ICD Procedures: DHD registration system for procedures Medication: Dutch classification system
Perinatal Registry (external)	Data on pregnancies, birth and neonatal outcomes. Data are obtained from Perined.[c]	National	Information on mothers, eg: • Maternal age • Obstetric history • Parity Information on pregnancies, eg: • Mode of conception • Mode of delivery Information on children, eg: • Birth weight • Gestational age • Apgar score	
Pathology Registry (external)	Excerpts of histological, cytological and autopsy examinations. Data are obtained from PALGA.[d]	National	• Summary of pathology report • PALGA diagnosis, structured along five classification axes: 1. Topography 2. Morphology 3. Function 4. Procedure 5. Diseases	Diagnosis codes are a combination of diagnostic terms (localization, acquisition technique, abnormality) and related to the SNOMED coding system.

(Continued)

Table 1 (Continued).

Database	Data Source	Catchment Area[a]	Data Collected	Coding System
Cancer Registry (external)	Data on all newly diagnosed cancer cases. Data are obtained from the Dutch Comprehensive Cancer Organization.[e]	National	• Cancer diagnosis • Tumor staging • Tumor site • Morphology • Comorbidity at diagnosis • Initial treatment	Tumor staging: TNM-classification Tumor site and morphology: WHO ICD-O

Notes: [a]as determined September 2016; [b]www.dhd.nl; [c]www.perined.nl; [d]www.palga.nl; [e]www.iknl.nl.
Abbreviations: ATC, Anatomical Therapeutic Chemical; DHD, Dutch Hospital Data Foundation; GP, general practitioner; ICD, International Classification of Diseases; ICD-O, International Classification of Diseases for Oncology; ICPC, International Classification of Primary Care; LOINC, Logical Observation, Identifiers, Names and Codes; PALGA, Pathologisch-Anatomisch Landelijk Geautomatiseerd Archief (ie nationwide network and registry of histo- and cytopathology in the Netherlands); SNOMED, Systematized Nomenclature of Medicine; TNM, Tumor Nodes Metastasis Classification of Malignant Tumors; WCIA, Werkgroep Coordinatie Informatie Automatisering (ie Dutch coding system for lab tests); WHO, World Health Organization.

Study Population and Follow-Up

The longitudinal nature of the PHARMO Database Network system enables to follow up more than 10 million persons over the years, or about 7 million active persons (January 1, 2018), regardless of age and gender, of a well-defined population in the Netherlands with a follow-up of 10 to 30 years.

The databases in the PHARMO Database Network are supplemented by linking external, national databases and registries in order to fill gaps in the patient journey. For example, pharmacy, GP, laboratory and hospital records contain limited details on cancer characteristics and outcomes. Linkages with the Netherlands Cancer Registry[2] (maintained by the Netherlands Comprehensive Cancer Organization, since 1989) and the Pathology Registry[3] (maintained by the PALGA Foundation, since 1991) were established to add detailed information on, among other, staging and morphology and access to tissue samples. Likewise, linkage with the national Perinatal Registry[4] (maintained by Perined, since 1999) facilitates studies on pregnancy and outcomes. Because mother-child pairs are identified, exposure or events during pregnancy can be studied in relation to long-term follow-up of both mother and child.[5] Permission to use these external databases is requested from the database holders on a project basis.

Study populations are created from the PHARMO Database Network using the data sources needed to address the study objectives. Each database had specific limitations regarding the study population that should be taken into consideration when selecting the study population. In general, the population included in the PHARMO Database Network is representative of the Dutch population with respect to age and gender (see Figure 1). However, the population >80 years old is relatively large and the population <20 years old is relatively small in PHARMO compared to the general Dutch population.[6] This is most likely the result of the differences in healthcare utilization by age. In addition, data collection periods and catchment areas vary between the databases. Therefore, the size of the source population for any study depends on the databases included.

Main Variables

Data are retrieved directly from the source, ie the electronic medical records of the healthcare providers who agree to collaborate with "Stichting Informatievoorziening voor Zorg en Onderzoek" (STIZON) and contribute to the anonymized datasets used by the PHARMO Institute for research purposes. Healthcare providers remain owner of the data and contribute to the PHARMO Database Network voluntarily. Since 2012, the data is managed and collected at least once a year by STIZON, an ISO/IEC 27001 and NEN 7510 certified foundation, compliant to data protection regulations. STIZON acts as a Trusted Third Party between the data sources and users of the anonymized data. STIZON has data processing and service agreements with all participating healthcare providers and acts as a processor in terms of the General Data Protection Regulation (GDPR). STIZON collects identifiable patient data from the contributing healthcare providers through a technology-assisted extract, transform, load (ETL) process that maps data from disparate sources to a common data model.

STIZON also links the different sources on a patient level using a combination of deterministic linkage on a
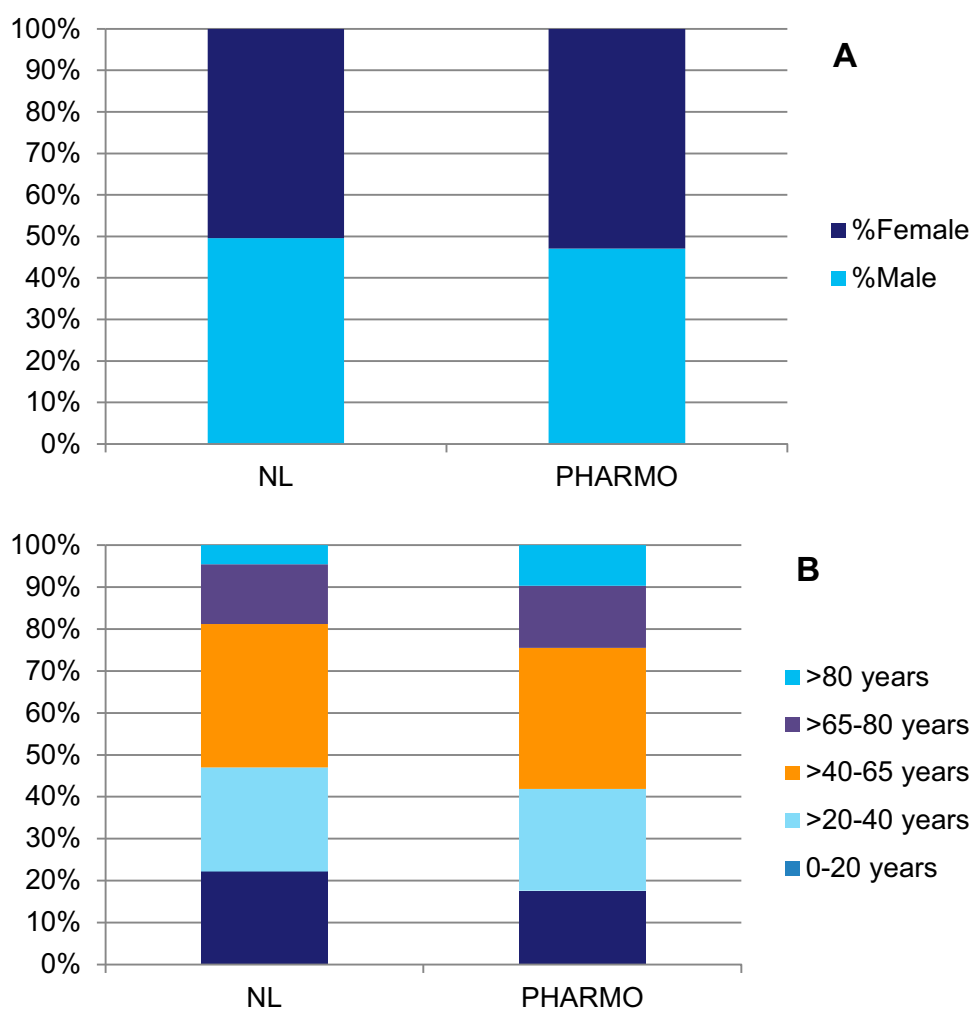
Figure 1 Sex (**A**) and age (**B**) distribution of population in the PHARMO Database Network compared to the Dutch population.
Notes: (**A**) dark blue: female, bright blue: male; (**B**) bright blue: >80 years, purple: >65–80 years, orange: >40–65 years, light blue: >20–40 years, dark blue: 0–20 years.

unique patient identification number, when available, and probabilistic linkage on patient characteristics.[1] Validation of the linkage against name and address information for a sample of the patients resulted in a sensitivity and specificity of 0.98.[7] The linkage is updated on a yearly basis. After linkage, study-specific datasets can be requested by the PHARMO Institute or academic affiliates by submitting a data request accompanied by a study protocol to STIZON. Upon approval by the Compliance Committee, described in more detail below, the study-specific datasets are depleted of patient identifiable information by STIZON and made available to the PHARMO Institute and academic affiliates for research purposes. The use of anonymized, proportional datasets via a remote server is in accordance with the GDPR. Records are anonymized by replacing identifiers with randomly generated ones that only have meaning within the dataset. While this approach

provides direct non-identifiability, pseudonymized records may still be indirectly identifiable through probabilistic matching between datasets with quasi-identifiers (eg postal codes). To mitigate this issue, all identifiers that have external relevance are removed (eg social security numbers are cleared), and common quasi-identifiers are aggregated into categories (ie postal codes are shortened to four-digit numbers) before the data are made available for research purposes.

All patients registered at the contributing healthcare providers are included, unless the patient requested to opt out. Healthcare providers are advised to inform their patients regarding participation in the PHARMO Database Network and information materials are made available by STIZON. The frequency of data collection varies per healthcare provider but is at least on an annual basis. Linkage of the different data sources is yearly; the

lag time is about 1 year. Once the linked data are made available to PHARMO, the quality of the data is assessed by data acceptance tests. The contents of these yearly tests differ per database but include quality indicators such as level of missing data, values within a reasonable range and appropriate coding used.

The type of data collected per data source is described in Table 1. The GP Database contains detailed information on patient characteristics, symptoms, diagnosis, prescribed medication and requested laboratory tests, including test results. The GP is the gatekeeper of the Dutch healthcare system and all patients, regardless of age or insurance, are uniquely registered by a single GP. Access to secondary care is only through referral by the GP. Due to the central position of the GP in the healthcare system, they play a key role in the patient journey. The GPs manage specific care programs for chronic diseases such as type 2 diabetes mellitus, chronic obstructive pulmonary disease (COPD)/ asthma and cardiovascular risk management. In these programs, patients are monitored and treated in accordance with predefined, nationwide implemented standardized rules (eg specific tests and frequency of testing). GPs are reimbursed based on their compliance with these predefined rules. In order to support GPs with being compliant to these care programs, feedback information is provided to the GP based on the data collected by STIZON. As a result of this feedback loop, the disease cohorts based on these care programs are validated and contain frequently monitored patients with a very high level of detail available.

The Out-patient Pharmacy Database and the In-patient Pharmacy Database include information on the type of medication dispensed, including Anatomical Therapeutic Chemical (ATC) code and brand names, dosage, duration of dispensing and information on the prescriber. The Out-patient Pharmacy Database includes all medication dispensed in community- and hospital-based out-patient pharmacies, regardless of insurance, employment status, prescriber, age or gender. The In-patient Pharmacy Database includes all medication dispensed during a hospitalization. The Out-patient Pharmacy and In-patient Pharmacy Databases do not include information on the indication of treatment, due to local laws and regulations. Linkage with another database, such as the GP Database is needed to infer the indication, by concurrent diagnoses.

The Clinical Laboratory Database includes the type of laboratory test, by whom the test was requested and the results of the laboratory tests.

The Hospital Database comprises datasets containing data on hospital admissions, ambulatory consultations and high budget impact medication. The datasets on hospital admissions and ambulatory consultations include information on admission, discharge and consultation dates, primary and secondary diagnoses and procedures. The dataset on high budget impact medication includes information on type of drug dispensed, date of dispensing, prescriber and indication of use. Hospital data are maintained and processed by the Dutch Hospital Data foundation (DHD), STIZON is authorized by the individual hospitals to obtain this data from the DHD and use it for research purposes. Each hospital is responsible for data collection according to the common data model developed by the DHD in collaboration with the Dutch hospital associations.[8]

As with any database, identification of medical events is limited to data that are captured as part of the medical records or other linked data sources in daily practice. These data are not primarily collected for research purposes and rely on appropriate diagnostic coding to detect these events. Furthermore, only the complications that require medical attention are included in the PHARMO Database Network, which mostly concerns the more severe and acute events.

Through the collaboration with STIZON, in its role as Trusted Third Party, it is possible to go back to the healthcare provider for validation or extension of the retrospectively collected data by chart review or additional data collection.

Research Examples

The use of the study-specific datasets from the PHARMO Database Network is controlled by an independent privacy and governance board, the Compliance Committee. The Compliance Committee consists of representatives of all different healthcare providers and a separate privacy expert, including the GDPR data protection officer of STIZON. All studies require permission of the Compliance Committee and its decisions are based on the applicable Dutch and European legislations, including the Medical Treatment Contracts Act (WGBO). The majority of studies using the anonymized data from the PHARMO Database Network are not subject to ethics review according to the Medical Research Involving Human Subjects Act (WMO). In the cases where ethics approval is required, the study protocol is submitted to an accredited medical research ethics committee in the Netherlands for review.

Data from the PHARMO Database Network have been used in over 1000 studies and 500 publications in peer-reviewed journals, both national and international. These studies and publications cover all major therapeutic areas, including oncology and metabolic, cardiovascular and respiratory diseases. Research topics include drug safety (including authority driven post-authorization safety studies), healthcare resource utilization, relative effectiveness, drug utilization, disease surveillance, burden of illness and treatment patterns.

Besides studies focusing on the Netherlands, the PHARMO Institute coordinates and participates in international database studies. In these studies, PHARMO collaborates with other expertise centers to benefit from each partner's local expertise of the data and the healthcare system; the context of the data puts the study results in perspective. The majority of the international database studies are registered in the ENCePP register.[9] In the context of a multi-country study, PHARMO can share data on an aggregated level.

The list of publications based on the PHARMO Database Network can be obtained online at http://www.pharmo.nl/. Three examples of publications showing the research possibilities with data from the PHARMO Database Network are described below.

For a study on benign prostatic hyperplasia (BPH) a combination of three databases was used.[10] The objective of the study was to compare the rates of BPH-related prostate surgery and acute urinary retention between dutasteride and finasteride alone or in combination with an alpha-adrenoreceptor antagonist. The Out-patient Pharmacy Database was used to identify the study population, ie men aged 50 years or older who were treated with dutasteride or finasteride. The Hospitalization Database was used to determine study outcomes, ie BPH-related prostate surgery and acute urinary retention. The GP Database was used to determine complaints and covariates, eg comorbidities and number of GP visits. The study showed that the risk of BPH-related prostate surgery was significantly lower among men using dutasteride than among men using finasteride.

Another publication described musculoskeletal hospital admissions and surgical procedure rates among patients treated for rheumatoid arthritis (RA) during the introduction of biological disease-modifying anti-rheumatic drugs (DMARDs) in the Netherlands.[11] The rates were compared with population references. RA patients were identified by dispensing of any relevant DMARD and absence of alternative indications. Relevant DMARDs were based on Dutch treatment guidelines for RA and were categorized as conventional or biological DMARD. These selection criteria may have resulted in an overestimation of patients with RA due to the limited specificity in the group without an admission for RA. Patients with an admission for RA most likely represent patients with more advanced disease. This study showed that the rates of musculoskeletal surgical procedures among patients treated for RA decreased between 1999 and 2012. The largest reduction occurred during the period when bio-DMARDs were introduced. However, the surgical procedure rate remained more than twice as high for patients with RA as compared to reference subjects and was almost four times higher among patients receiving bio-DMARDs. It was hypothesized that these increased rates are most likely explained by the fact that bio-DMARDs are indicated for patients with more advanced disease. Severity of disease indicators could not be retrieved from the databases included in this study.

In a study on the association between expression of HLA class I antigen, aspirin use and survival in patients diagnosed with colon cancer, tissue samples available through the PALGA Foundation were used in combination with data from the PHARMO Database Network.[12] Information on diagnosis and medication use were retrieved from the PHARMO Database Network and was linked to the HLA class I antigen results as determined in the tissue samples retrieved through the PALGA Foundation. By retrieving data on medication use from dispensing records rather than patients, recall bias was avoided. The study showed that in tumors expressing HLA class I antigen, aspirin use after colon cancer diagnosis was associated with improved survival. Upon confirmation in other data sets, the results of this study could impact timing and dose of aspirin in clinical practice.

Data Access and Funding

Access to the PHARMO Database Network is, by governance regulations of the data collection, restricted to researchers of the PHARMO Institute and academic affiliates. Academic affiliates from universities or research institutes can apply for access to the data of the PHARMO Database Network for scientific studies. This endeavor is in line with the policy and mission of the PHARMO Institute to contribute to a better understanding of the use, safety, effectiveness and cost of pharmaceuticals as used in real-life. The application form can be found

on the PHARMO website (www.pharmo.com) and should be submitted together with a study protocol. Applications are checked against the policies that apply for use of data from the PHARMO Database Network and as agreed upon with the contributing healthcare providers. Funding for academic research is not provided by the PHARMO Institute and should be obtained by the researcher taking into account the data access fees for use of the data. Upon approval of the data application by the CC, researchers are provided access to the data at the PHARMO Institute offices. In 2019, almost one-third of all projects using data from the PHARMO Database Network were performed by academic affiliates. The remaining studies were mainly commissioned by pharmaceutical companies (43%) and contract research organizations (18%). Other funding sources include the European Medicines Agency (EMA) and funding programs, eg FP6/7.

Conclusion

The PHARMO Database Network provides a unique opportunity to conduct studies on the safety and effectiveness of medication use in the Netherlands. The PHARMO Database Network covers a well-defined population-based sample of patients with detailed information from multiple healthcare settings. The longitudinal nature of the data allows for long-term follow-up of patients and surveillance of long-term outcomes, beyond the scope of most prospective studies.

Disclosure

JK, MB, FP and RH are employees of the PHARMO Institute for Drug Outcomes Research. This independent research institute performs financially supported studies for government and related healthcare authorities and several pharmaceutical companies. As this paper uses data from existing publications without any direct enrolment of subjects, ethical approval or informed consent is not necessary according to the Dutch law regarding medical research involving human subjects (WMO), which is enforced by the Central Committee on Research involving Human Subjects (Centrale Commissie Mensgebonden Onderzoek [CCMO]). The authors report no other conflicts of interest in this work.

References

1. Herings RMC, Pedersen L. Pharmacy -based medical record linkage systems. In: Strom BL, Kimmel SE, Hennessy S, editors. *Pharmacoepidemiology, 5th Edition*. 5th ed. John Wiley & Sons, Ltd.; 2012:270–286.
2. Netherlands Comprehensive Cancer Organization. Netherlands Cancer Registry. 2018 Available from: www.iknl.nl. Accessed April 17, 2018.
3. Foundation P The nationwide network and registry of histo- and cytopathology in the Netherlands. Available from: www.palga.nl. Accessed 17 April 2018.
4. Perined. 2018. Netherlands Perinatal Registry. Available from: www. perined.nl. Accessed 17 April 2018.
5. Houweling LM, Bezemer ID, Penning-van Beest FJ, Meijer WM, van Lingen RA, Herings RM. First year of life medication use and hospital admission rates: premature compared with term infants. *J Pediatr*. 2013;163(1):61–66 e61. doi:10.1016/j.jpeds.2012.12.014
6. Statistics Netherlands. Bevolking naar leeftijd (op 1 januari). Available from: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/37296ned/table?ts=1578480210688. Accessed 23 January, 2020.
7. van Herk-sukel MP, van de Poll-franse LV, Lemmens VE, et al. New opportunities for drug outcomes research in cancer patients: the linkage of the Eindhoven Cancer Registry and the PHARMO record linkage system. *Eur J Cancer*. 2010;46(2):395–404. doi:10.1016/j.ejca.2009.09.010
8. DHD Foundation. Gebruikershandleiding Landelijke Basisregistratie Ziekenhuiszorg (LBZ). 12 December 2019; version 1.2.: Available from: https://www.dhd.nl/klantenservice/handleidingen_formulieren/documents/Gebruikershandleiding%20Landelijke%20Basisregistratie%20Ziekenhuiszorg%20%28LBZ%29.pdf. Accessed February 28, 2020.
9. European Network of Centers for Pharmacoepidemiology and Pharmacovigilance. Resources Database. 2018. Available from: http://www.encepp.eu/encepp/links.htm?id=17375&resourceType=ResearchCentre. Accessed 2018.
10. Kuiper JG, Bezemer ID, Driessen MT, et al. Rates of prostate surgery and acute urinary retention for benign prostatic hyperplasia in men treated with dutasteride or finasteride. *BMC Urol*. 2016;16(1):53. doi:10.1186/s12894-016-0170-6
11. Bezemer ID, Kool-Houweling LMA, Alemao E, Penning-van Beest FJA, Herings RMC. Musculoskeletal Hospital admission and surgical procedure rates among patients treated for rheumatoid arthritis in the Netherlands 1999–2012. *J Orthopedics Rheumatol*. 2016;3:1.
12. Reimers MS, Bastiaannet E, Langley RE, et al. Expression of HLA Class I antigen, aspirin use, and survival after a diagnosis of colon cancer. *JAMA Intern Med*. 2014;174(5):732–739. doi:10.1001/jamainternmed.2014.511