A Textbook of Pharmacoepidemiology and Pharmacovigilance for Pharmacy Students

Lesson 5:

Disease and Drug classifications

Fernando Fernandez-Llimos

ISBN-13: 978-84-124195-6-6

Porto, Portugal.

Published by CIPF, S.L.

Suggested citation: Fernandez-Llimos F. Lesson 5: Disease and Drug classifications. In: A Textbook of Pharmacoepidemiology and Pharmacovigilance for Pharmacy Students. Porto: CIPF; 2025.

Available at http://www.cipf-es.org

Lesson 5. Disease and Drug classifications

Contents

5.1. The Pillars of Classification in Pharmacoepidemiology

Disease and drug classification systems form the bedrock of pharmacoepidemiologic research, enabling the transformation of complex clinical realities into **structured**, **analyzable data**. These systems, ranging from the ***International Classification of Diseases*** **(ICD)** to the ***Anatomical Therapeutic Chemical*** **(ATC)** framework, serve as universal lexicons that standardize diagnoses, therapies, and outcomes across diverse populations and healthcare settings. Their role extends far beyond mere administrative convenience: they are critical tools for ensuring study validity, enabling cross-study comparability, and advancing causal inference in real-world drug effect analyses.

At its core, **disease classification** seeks to encapsulate heterogeneous pathophysiological states into discrete, operationalized categories. The ICD-11, for instance, distinguishes between "type 1 diabetes mellitus with ketoacidosis" (5A10.1) and "type 2 diabetes mellitus without complications" (5A11), **granularity** that directly influences drug safety studies. When investigating SGLT-2 inhibitors and diabetic ketoacidosis risk, misclassifying diabetes subtypes could conflate drug effects with inherent disease severity, yielding biased risk estimates. Similarly, **drug classification** systems like the ATC hierarchy (e.g., "A10BA02" for metformin) standardize exposure definitions, distinguishing between chemical entities, formulations, and therapeutic intent. This precision is vital when comparing glucagon-like peptide-1 (GLP-1) receptor agonists (A10BJ) to dipeptidyl peptidase-4 inhibitors (A10BH), as their mechanisms and risk profiles differ substantially despite shared indications.

The integrity of these classifications directly impacts **study validity**. Consider antidepressants: the DSM-5's diagnostic criteria for major depressive disorder (MDD) shape cohort inclusion in pharmacovigilance studies. If MDD severity is misclassified, grouping mild episodic cases with treatment-resistant depression, analyses may falsely attribute hospitalization risks to drug toxicity rather than underlying disease progression. Likewise, ATC codes that conflate

immediate, and extended-release opioid formulations (e.g., oxycodone N02AA05 vs. oxycodone/naloxone N02AA55) obscure formulation-specific overdose risks, compromising internal validity.

Cross-study comparability hinges on classification consistency. The 2022 ICD-11 update introduced "post-COVID-19 condition" (RA02.0), enabling standardized tracking of long COVID complications across vaccine safety studies. Prior inconsistencies, where researchers used ad-hoc definitions like "persistent fatigue U08.9", hampered meta-analyses of antiviral drug efficacy. Similarly, the WHO Drug Dictionary harmonizes global pharmacovigilance data by mapping disparate national drug names (e.g., paracetamol vs. acetaminophen) to unified identifiers, ensuring signal detection algorithms function transnationally.

For **causal inference**, classification systems act as both enablers and obstacles. Well-structured **ontologies**[1] facilitate adjustment for confounders: ICD-coded comorbidities (e.g., hypertension = I10) allow **propensity score matching** in studies of NSAID-associated renal failure. However, structural limitations, such as the ATC system's inability to capture **off-label uses**, introduce residual confounding. A drug like pregabalin (N03AX16), prescribed for both neuropathic pain and generalized anxiety, may appear associated with falls in elderly populations due to channeling to frailer patients, a bias invisible to standard ATC-based analyses.

Emerging challenges, such as **phenotype drift** in chronic diseases, complicate classifications. The 2023 Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria redefined COPD subtypes based on eosinophil counts, altering inclusion thresholds for studies of inhaled corticosteroids. Researchers analyzing legacy data must reconcile outdated ICD-10 codes (J44.9) with new clinical phenotypes, risking temporal confounding. Similarly, precision oncology's shift from organ-based (ICD-O-3) to molecular classifications (e.g., HER2-low breast cancer) demands parallel evolution in drug categorization systems to maintain causal clarity.

Ultimately, these classification frameworks are not passive repositories but active determinants of evidence quality. Their structure dictates which drug-disease relationships become visible and which remain obscured in the vast datasets driving modern pharmacovigilance. As pharmacoepidemiology increasingly informs regulatory decisions, from drug approvals to safety withdrawals, the rigor of these systems underpins the reliability of real-world evidence shaping global therapeutic practice.

5.1.1 What is a Classification?

A classification is the process of **systematically arranging entities** (e.g., objects, concepts, organisms, or data) **into groups or categories** based on **shared characteristics** or criteria. This organizational activity is foundational across scientific disciplines, providing structure to otherwise complex or heterogeneous sets. The essential goal is to **make sense of diversity** by sorting items according to similarities and distinguishing features, thereby enabling clearer understanding, efficient retrieval of information, and meaningful comparison.

In practice, classification may involve grouping diseases by underlying etiology, drugs by their mechanism of action, or data according to measurable attributes. The criteria for classification

[1] Ontology: a set of concepts and categories in a subject area or domain that shows their properties and the relations between them.

are **defined in advance** and **applied consistently** to ensure that each item is placed in a group reflecting its nature or function.

5.1.2 Characteristics of a Good Classification System

To serve its intended purpose effectively, a good classification system should embody several key characteristics:

- A **clear and logical structure**. Categories and subcategories must be well-defined, organized in a manner that reflects relationships among entities, and arranged either hierarchically or categorically for intuitive navigation and retrieval.
- **Mutually exclusive categories**: each item should, ideally, belong to only one category, minimizing ambiguity and overlap. This promotes clarity and prevents confusion during application and interpretation.
- **Comprehensiveness**. An effective system must encompass the full range of items within its scope, ensuring that nothing relevant is excluded and no data are left unclassified.
- **Consistency and standardized terminology** underpin the system's reliability, permitting unambiguous interpretation and facilitating interoperability between systems or users. The same principles and language should be applied across time and circumstances.
- **Adaptability and flexibility** allow the system to evolve in response to advances in knowledge or changes in the landscape—new diseases, treatments, or scientific discoveries. While stability is valuable, a good classification system must be elastic enough to accommodate justified modifications without undue disruption.
- The system should also be **intuitive and easy to use**, enabling users to efficiently organize, locate, and analyze information. Scalability is important as the quantity and complexity of data grow.
- Finally, **compatibility** with other relevant systems enhances its utility, particularly in multi-disciplinary fields such as pharmacoepidemiology, where integration with clinical, regulatory, or surveillance frameworks may be necessary.

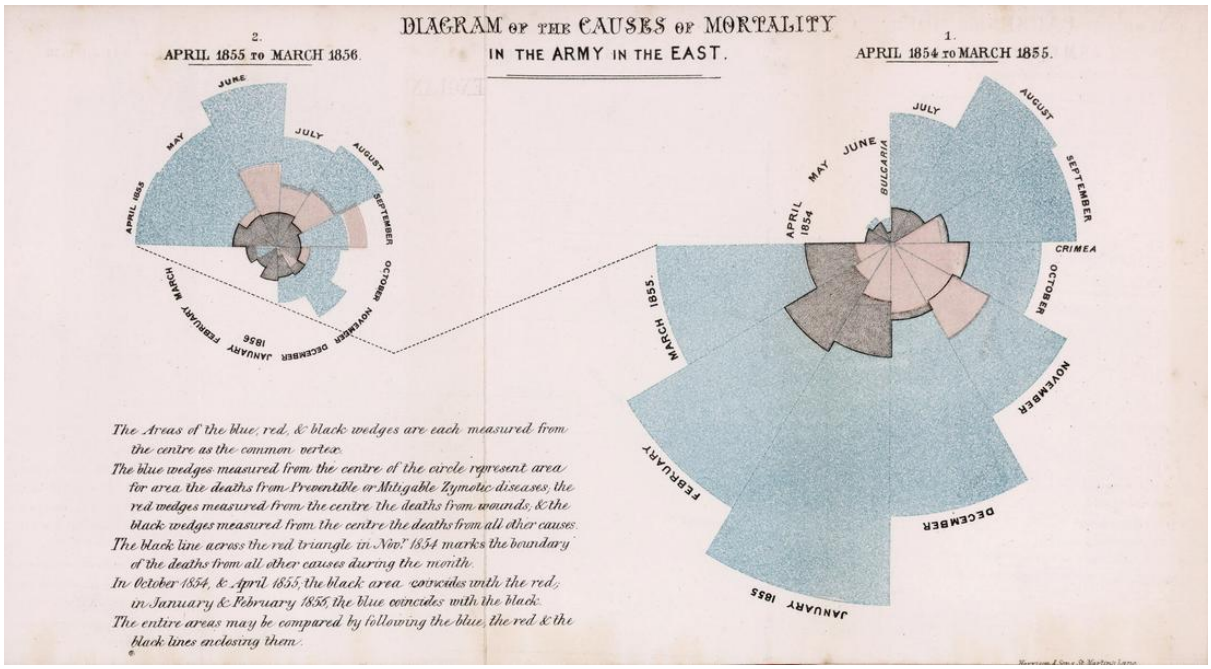
5.2. Disease Classification

5.2.1. Florence Nightingale

Florence Nightingale's pioneering work, acting as a volunteer nurse during the Crimean War (1853–1856), reshaped early disease data collection methods by introducing systematic approaches to public health. Her innovations addressed critical flaws in existing practices and laid the groundwork for modern epidemiology and healthcare statistics. Key influences include:

- **Standardization of data collection**: Nightingale encountered chaotic and inconsistent record-keeping in military hospitals, where soldiers who died between weekly counts were omitted from mortality reports, undercounting deaths by as much as 86%. She advocated for continuous, standardized data collection, insisting on tracking daily admissions, discharges, and causes of death. This replaced *ad hoc* methods that obscured true mortality rates, such as excluding deaths from cholera or accidents to artificially lower statistics. Her collaboration with statistician William Farr led to uniform **protocols for categorizing causes of death**, ensuring comparability across datasets.

- **Focus on preventable causes**: By analyzing mortality data, Nightingale demonstrated that preventable diseases (e.g., typhus, cholera) caused far more deaths than battlefield injuries. In Crimea, she showed that **improving sanitation reduced hospital mortality from 42.7% to 2.2% within a year**. This evidence forced governments to prioritize hygiene in data collection, linking environmental factors (clean water, ventilation) to health outcomes.
- **Visual data communication**: Nightingale's polar area diagrams (or "**coxcomb charts**" [N.B., crista de galo (*Celosia plumosa*)]) revolutionized how data was presented. These visualizations color-coded deaths by cause (blue for disease, red for wounds) and scaled wedges by mortality rates, making trends unmistakable to policymakers. Her 1858 report to the British Parliament used these graphics to prove that 16,000 of 18,000 soldier deaths were preventable, catalyzing reforms in military and civilian hospitals.



- **Advocacy for accuracy and transparency**: She exposed systemic flaws in data integrity, criticizing officials who excluded inconvenient death cause categories (e.g., cholera) to manipulate statistics. In India, she noted that omitting certain diseases from records would "logically lead to zero mortality". Her insistence on comprehensive, unbiased data led to reforms in how hospitals and governments tracked morbidity and mortality, emphasizing accountability.
- **Legacy in public health systems**: Nightingale's methods became foundational for modern epidemiology. Her 1860 push for standardized hospital statistics at the International Statistical Congress enabled cross-institutional comparisons. In India, her recommendations reduced soldier mortality from 69 to 18 per 1,000 through sanitation reforms informed by rigorous data collection.

5.2.2. The International Classification of Diseases (ICD)

The *International Classification of Diseases* (ICD) represents one of the greatest achievements in global health-data standardization. Its evolution from a mortality-tracking tool to a

multidimensional nosological[2] framework mirrors the trajectory of modern medicine itself, reflecting shifts in scientific understanding, clinical priorities, and technological capabilities.

ICD-1 to ICD-6: Foundations in mortality statistics

The ICD originated in 1893 as the *Bertillon Classification of Causes of Death*, developed by French physician Jacques Bertillon (1851-1922). This first version (retroactively termed ICD-1) organized **44 categories of fatal conditions**, prioritizing infectious diseases (e.g., tuberculosis, cholera) and trauma, reflecting 19th-century public health challenges. By 1900, the system had expanded to 161 categories and was adopted by 26 countries, driven by its utility for comparative mortality statistics.

The ICD-6 (1949) marked a paradigm shift under World Health Organization (WHO) stewardship. For the first time, it incorporated **morbidity,** non-fatal conditions, enabling its use beyond death certificates. This revision introduced a dedicated chapter for mental disorders (e.g., schizophrenia, neuroses), acknowledging psychiatry's growing clinical relevance. However, ICD-6 retained a rigid structure ill-suited for chronic diseases or comorbidities, limitations that persisted through ICD-7 (1958) and ICD-8 (1968).

Into effect	Classification
1983	Bertillon Classification of Causes of Death
1910	ICD-2
1921	ICD-3
1930	ICD-4
1939	ICD-5
1949	ICD-6
1958	ICD-7
1968	ICD-8
1979	ICD-9
1993	ICD-10
2022	ICD-11

ICD-9 to ICD-10: The rise of granularity[3]

The 1979 ICD-9 expanded diagnostic specificity with 17,000 codes, including procedural classifications (ICD-9-CM). Its alphanumeric format (e.g., **250.0** for type 1 diabetes) enabled finer distinctions but struggled with emerging conditions like HIV/AIDS. By 1990, ICD-10 addressed these gaps through:

- **Alphanumeric codes**: 14,400 codes across 22 chapters (e.g., **E10** for type 1 diabetes, **E11** for type 2).
- **Multiaxial structure**: Separating diagnoses, disabilities, and contextual factors.
- **Global adoption**: Translated into 43 languages and used in more than 100 countries.

ICD-10's granularity revolutionized pharmacoepidemiology. For example, distinguishing **J45.0 (allergic asthma)** from **J45.1 (non-allergic asthma)** allowed targeted studies of omalizumab efficacy. However, its static framework faltered with genomic advances and digital health integration.

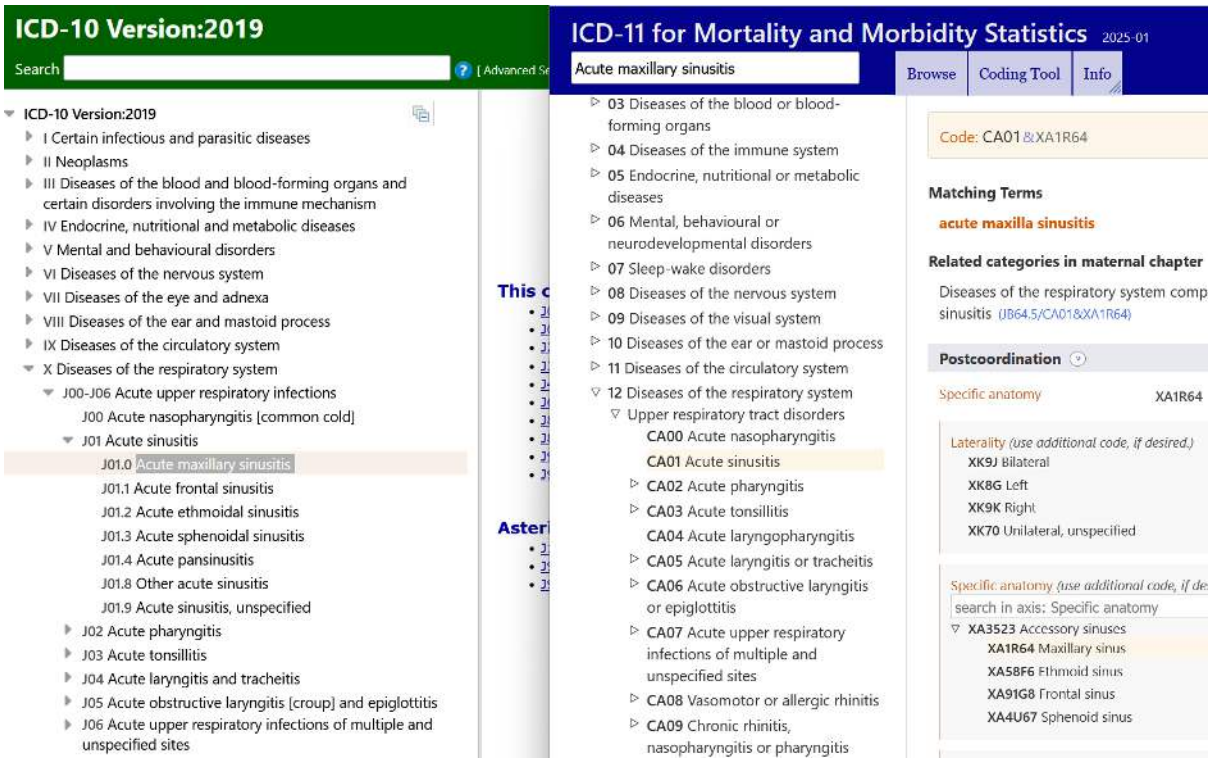
[2] Nosology: the branch of medicine concerned with the classification of diseases
[3] ICD-10: https://icd.who.int/browse10/2019/en

ICD-11: A digital-first paradigm[4]

Approved in 2019 and implemented in 2022, ICD-11 represents a revolution in classification science. Its **Foundation Component** integrates 55,000 entities across 27 chapters, with three transformative features:

1. **Cluster coding**: Combines multiple codes to model complex conditions. For example, **DA63 (duodenal ulcer) + ME24.90 (acute gastrointestinal bleeding)** captures ulcer-related hemorrhage.
2. **Extension codes**: Add temporal, severity, or etiological context (e.g., **XT9T (pandemic origin)** for COVID-19).
3. **Digital tools**: Machine-readable APIs, AI-assisted coding, and automated mapping to legacy systems.

ICD-11's modular design supports precision medicine. The **6A02 (autism spectrum disorder)** code incorporates genetic subtypes (e.g., **6A02.0 with CDKL5 mutation**), while **RA02.0 (post-COVID-19 condition)** standardizes long COVID research. Its **MMS linearization** (Mortality and Morbidity Statistics) simplifies reporting by grouping entities into mutually exclusive categories, though "gray nodes" allow cross-chapter linkages (e.g., *pneumonia* listed under both respiratory infections and infectious diseases).



As of 2025, full ICD-11 implementation remains incomplete in many countries due to costs and training requirements.

Also, residual challenges persist. Off-label drug uses remain invisible in ATC-aligned codes, while diagnostic criteria updates (e.g., COPD's 2023 GOLD criteria) often outpace ICD revisions.

[4] ICD-11: https://icd.who.int/browse/2025-01/mms/en

5.2.3. The International Classification of Primary Care (ICPC-2)

The **International Classification of Primary Care, 2nd edition (ICPC-2)**, developed by the **World Organization of Family Doctors (WONCA)**, is a classification system tailored to the dynamic, patient-centered nature of **primary care**. Unlike the ICD-10, which prioritizes disease-specific granularity, **ICPC-2 captures the full spectrum of primary care encounters, including symptoms, diagnoses, processes of care, and social determinants**. Its design reflects the reality that 30–50% of primary care visits involve **undifferentiated symptoms** (e.g., fatigue, dizziness) rather than formal diagnoses, making it indispensable for pharmacoepidemiologic studies seeking to understand real-world drug use patterns and outcomes.

ICPC-2 employs a **biaxial structure** with **17 chapters** (each denoted by a letter) and **7 components** per chapter, enabling comprehensive documentation of:

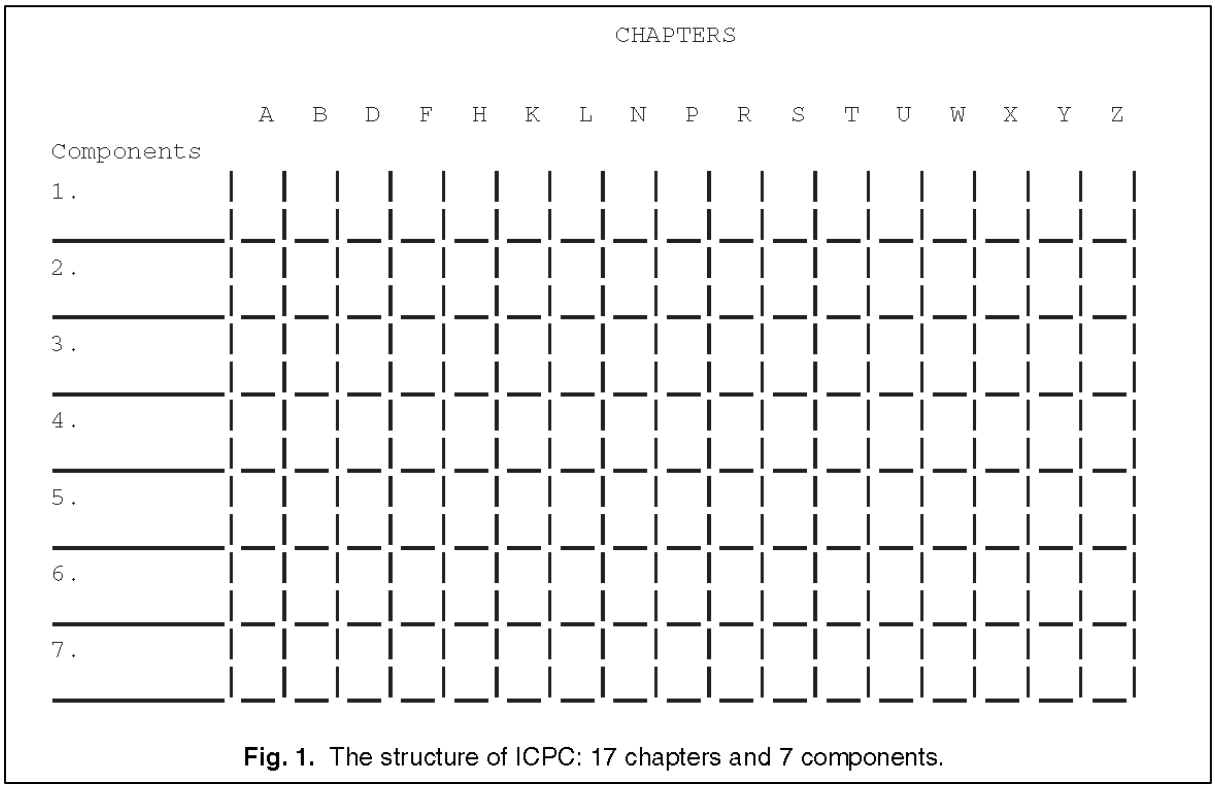


Fig. 1. The structure of ICPC: 17 chapters and 7 components.

The 17 chapters, representing **body systems**, are:

- A General and unspecified
- B Blood, blood-forming organs and immune mechanism (spleen, bone marrow)
- D Digestive
- F Eye
- H Ear (Hearing)
- K Circulatory
- L Musculoskeletal (Locomotion)
- N Neurological
- P Psychological
- R Respiratory
- S Skin
- T Endocrine, metabolic and nutritional

- U Urological
- W Pregnancy, child-bearing, family planning (Women)
- X Female genital (X.-chromosome)
- Y Male genital (Y-chromosome)
- Z Social problems

And the seven components (standard for each of the 7 chapters) are:

1. Complaint and symptom component
2. Diagnostic, screening, and preventive component
3. Medication, treatment, procedures component
4. Test results component
5. Administrative component
6. Referrals and other reasons for encounter
7. Disease component:
    - infectious diseases
    - neoplasms
    - injuries
    - congenital anomalies
    - other

The episode of care is a crucial concept within the ICPC-2 classification. An episode of care refers to the chronological sequence of interactions between a patient and a healthcare provider concerning a single health problem or condition. This concept differs from an episode of disease or illness, as it is defined not just by the condition but by the entire process of care provided. An episode of care is constructed prospectively, beginning when a patient first seeks help for a particular problem (often recorded in their own words as the "reason for encounter").

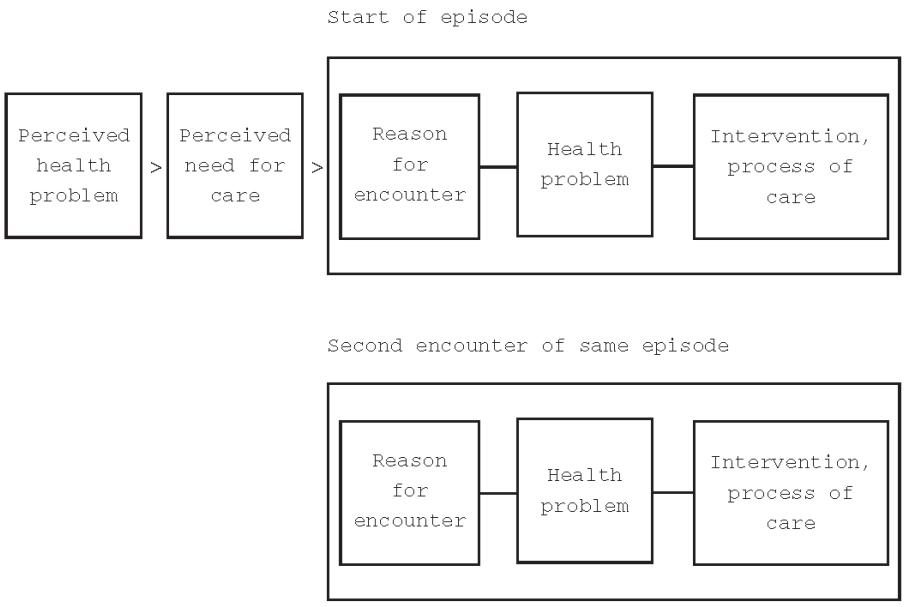


Fig. 2. An episode of care.

ICPC-2 is implemented in Portugal and used for coding in primary care centers [N.B., Unidades de Saúde Familiar (USF)]. Coding with ICPC-2 is complex, and its reliability may not be sufficiently high to be used in pharmacoepidemiologic studies[5].

Rule 1

The reason for the encounter should be coded as specifically as possible and may require some clarification by the provider.

Example

Chest pain can be coded as All (chest pain not otherwise specified (NOS)), or as KOI (pain attributed to heart), or as R01 (pain respiratory), or as L04 (chest symptoms/ complaints). The decision as to the correct selection is not based on the opinion of the provider as to the type of chest pain but, rather, to the manner in which the patient expresses his/her reason for encounter when clarification is sought by the provider.

'Its all over my chest …'	A11
'My chest hurts when I cough'	R01
'I have chest pain … I think its my heart'	KO1
'I have chest pain after falling down stairs'	L04

Several alternatives were created with a limited interest globally. For instance, the ICPC-2 PLUS[6] was created and is maintained and regularly updated by the Classifications and terminologies node, WHO-CC for Strengthening Rehabilitation Capacity in Health Systems, University of Sydney (Australia), and its use is virtually limited to this country.

The 2021 release of **ICPC-3**, poorly implemented, introduces enhanced functionalities[7]:

- **Expanded Domains**: Codes for functioning (e.g., mobility limitations), environmental factors (e.g., air quality), and personal health behaviors (e.g., diet).
- **Digital Integration**: APIs for machine learning applications, enabling predictive analytics (e.g., identifying at-risk patients for polypharmacy).
- **Global Equity**: Incorporates traditional medicine practices (Chapter 26) and indigenous health concepts, addressing disparities in post-colonial healthcare systems.

[5] Frese T, Herrmann K, Bungert-Kahl P, Sandholzer H. Inter-rater reliability of the ICPC-2 in a German general practice setting. Swiss Med Wkly. 2012;142:w13621.
[6] ICPC-2 PLUS: https://www.sydney.edu.au/medicine-health/our-research/research-centres/who-collaborating-centre-for-strengthening-rehabilitation-capacity-in-health-systems/classifications-and-terminologies/icpc-2-plus.html
[7] ICPC-3: https://www.icpc-3.info/

5.3. Drug Classifications

5.3.1. The Anatomical Therapeutic Chemical (ATC)

The **Anatomical Therapeutic Chemical (ATC) Classification**[8] system is the global standard for drug classification, maintained by the WHO Collaborating Centre for Drug Statistics Methodology (WHOCC), which exists under the coordination of the Norwegian Institute of Public Health[9]. The ATC is designed to classify active ingredients of drugs according to the organ or system on which they act, as well as their therapeutic, pharmacological, and chemical properties.

The ATC system was first published by the WHO in 1976, evolving from earlier pharmaceutical classification systems developed for market research. Its primary aim is to facilitate the **monitoring and improvement of medication quality and use at the population level**. Importantly, the **ATC system is not intended to recommend drugs or judge efficacy**; rather, it provides a standardized structure for categorizing pharmaceutical substances.

Key foundational principles include:

- **International Nonproprietary Names (INN):** Drugs are classified by their INN[10] when available, ensuring global consistency. If an INN is not available, USAN (United States Adopted Name) or BAN (British Approved Name) may be used.
- **Main therapeutic use:** Drugs are classified according to their principal therapeutic use, which may differ between countries. The main indication is determined by consensus within the WHO International Working Group for Drug Statistics Methodology.
- **Mechanism of action**: ATC codes are often assigned based on pharmacological mechanism rather than solely on therapeutic use, meaning drugs with similar mechanisms may be grouped together even if their clinical indications differ.
- **Combination products:** Fixed-dose combinations receive unique codes, but the system generally assigns only one code per combination and per route, prioritizing the main therapeutic use.

[8] ATC: https://atcddd.fhi.no/atc_ddd_index/
[9] Norwegian Institute of Public Health: https://www.fhi.no/en/
[10] INN: https://www.who.int/teams/health-product-and-policy-standards/inn/guidance-on-inn

International Nonproprietary Names (INN): WHO Guidance Overview

The **International Nonproprietary Names (INN) system** is a global standard developed and maintained by the **World Health Organization (WHO)** to assign unique, universally recognized generic names to pharmaceutical substances, mainly active ingredients. Established in 1950, the system facilitates safe prescribing, regulatory clarity, and unambiguous communication among scientists, healthcare professionals, and the public, regardless of language or region.

An INN represents the **official, nonproprietary, and public property** name for a specific, well-defined pharmaceutical substance. By convention, INNs cannot be registered as trademarks. This legal and practical safeguard prevents proprietary use of generic drug names, thereby reducing the risk of confusion in medical practice that could compromise patient safety.

The WHO INN Program follows a structured, transparent naming process. Manufacturers, researchers, or national authorities may submit a proposed new name. An international expert group reviews the proposal against strict criteria, such as distinctiveness, linguistic acceptability, and avoidance of any similarity that could cause prescribing errors. The draft name is then published for a public comment period, during which stakeholders, including regulatory bodies and the pharmaceutical industry, can provide feedback. Final naming decisions are made after considering these comments. Importantly, an INN refers to the *active* moiety alone, not specific salts, esters, or dosage forms.

One hallmark of the INN system is the use of **common stems,** specific syllables or endings that indicate pharmacological or chemical relationships among substances. For example, the stem "-pril" denotes angiotensin-converting enzyme (ACE) inhibitors, while "-mab" signals monoclonal antibodies. The use of stems allows healthcare providers to identify drug classes and related actions quickly, even when encountering unfamiliar molecules.

When an existing active substance is presented in the form of a salt, ester, derivative, or other close chemical variant, a **Modified INN** is designated. **INNMs** are built on the original INN, ensuring both continuity and clarity by indicating the modification alongside the parent name.

Many countries historically maintained their own official generic naming systems, such as **British Approved Name (BAN)** in the UK, **United States Adopted Name (USAN)** in the USA, **Dénomination Commune Française (DCF)** in France, and **Japanese Accepted Name (JAN)** in Japan, **Denominação Comum Brasileira (DCB)** in Brazil, where adalimumab is adalimumabe, or **Denominación Oficial Española (DOE) e Denominação comum em português (DCPt)**, where amoxicillin is amoxicilina.

WHO urges national authorities to ensure generic prescribing and labeling prominently display the INN (or harmonized equivalent). Furthermore, WHO advises against allowing trademarks to incorporate parts of INNs, especially established stems, as this may blur the distinction between proprietary and generic naming, creating a risk to patient safety.

DCPt: https://diariodarepublica.pt/dr/detalhe/deliberacao/957-2005-2761064

Structure of the ATC System

The ATC system is a strict hierarchy with **five levels**, each providing increasing specificity:

First Level (Anatomical Main Group): One letter indicating the main anatomical group (14 groups in total). Examples:

- A: Alimentary tract and metabolism
- C: Cardiovascular system
- J: Anti-infectives for systemic use

Second Level (Therapeutic Subgroup): Two digits specifying the main therapeutic subgroup within the anatomical group.

- Example: C03 (Diuretics)

Third Level (Pharmacological Subgroup): One letter indicating the pharmacological or therapeutic subgroup.

- Example: C03C (High-ceiling diuretics)

Fourth Level (Chemical/Therapeutic/Pharmacological Subgroup): One letter indicating the chemical or further therapeutic subgroup.

- Example: C03CA (Sulfonamides)

Fifth Level (Chemical Substance): Two digits specifying the individual chemical substance.

- Example: C03CA01 (Furosemide)

Example of a Full ATC Code: C03CA01

- C: Cardiovascular system
- 03: Diuretics
- C: High-ceiling diuretics
- A: Sulfonamides
- 01: Furosemide

Each bottom-level ATC code represents either a single pharmaceutically used substance or a specific combination of substances for an indication. This means that a drug may have more than one ATC code if it is used for different indications or routes of administration. For example, acetylsalicylic acid has different codes for use as a local oral treatment (A01AD05), as a platelet inhibitor (B01AC06), and as an analgesic/antipyretic (N02BA01), or the 21 ATC codes for prednisolone:

A01AC04	prednisolone	R01AD02	prednisolone
A01AC54	prednisolone, combinations	S01BA04	prednisolone
A07EA01	prednisolone	S01BB02	prednisolone and mydriatics
C05AA04	prednisolone	S01CA02	prednisolone and antiinfectives
D07AA03	prednisolone	S01CB02	prednisolone
D07BA01	prednisolone and antiseptics	S02BA03	prednisolone
D07CA03	prednisolone and antibiotics	S02CA01	prednisolone and antiinfectives
D07XA02	prednisolone	S03BA02	prednisolone
H02AB06	prednisolone	S03CA02	prednisolone and antiinfectives
H02AB06	prednisolone	V03AB05	prednisolone and promethazine

Some important semantic and hierarchical features about the ATC codes:

- **Strict Hierarchy:** Each ATC code has only one parent code, ensuring a clear lineage and unambiguous classification. This strict hierarchy supports robust data aggregation and analysis.
- **Semantic Identifiers:** ATC codes are semantic, meaning the code itself conveys information about the drug's classification at each level (e.g., anatomical, therapeutic, pharmacological, chemical).
- **Annual Updates**: The ATC/DDD index is updated annually, reflecting new substances, changes in clinical practice, and evolving therapeutic landscapes. The guidelines and index are available in multiple languages and formats from the WHOCC.

The WHO Collaborating Centre for Drug Statistics Methodology (WHOCC) has a rigorous procedure to create new ATC codes, based on:

- **Assignment Process:** New entries in the ATC system are established upon request by manufacturers, regulatory agencies, or researchers. The WHOCC oversees this process, ensuring that each substance is classified according to its main therapeutic use and route of administration.
- **Inclusion Criteria:** Substances are typically included if they fulfill criteria such as widespread use, regulatory approval, or significant clinical relevance. Herbal medicinal products and certain rare disease treatments may not receive **DDDs** (Defined Daily Doses) or may be listed separately.
- **Multiple ATC Codes**: A drug may receive **more than one ATC code** if it is available in different strengths or routes with clearly distinct therapeutic uses.
- **Combination Products**: Fixed-dose combinations are assigned a unique code at the fifth level. For example, C09BB04 is the code for the combination of perindopril and amlodipine, while each component also has its own code when prescribed alone.

C CARDIOVASCULAR SYSTEM
C09 AGENTS ACTING ON THE RENIN-ANGIOTENSIN SYSTEM
C09A ACE INHIBITORS, PLAIN
C09B ACE INHIBITORS, COMBINATIONS
C09C ANGIOTENSIN II RECEPTOR BLOCKERS (ARBs), PLAIN
C09D ANGIOTENSIN II RECEPTOR BLOCKERS (ARBs), COMBINATIONS
C09X OTHER AGENTS ACTING ON THE RENIN-ANGIOTENSIN SYSTEM

5.3.2. Classificação farmacoterapêutica de medicamentos (Infarmed)

The **Classificação farmacoterapêutica de medicamentos** (CFM) is the authoritative Portuguese system established and maintained by **INFARMED**, I.P. (Autoridade Nacional do Medicamento e Produtos de Saúde) for the categorization of all medicines authorized and marketed in Portugal. Its formal adoption, which dates to Despacho n.º 21 844/2004, de 12 de Outubro, marked the official integration of a pharmacotherapeutic framework tailored to the

clinical and regulatory needs of the Portuguese health system. It is periodically updated and published in the Diário da Republica[11].

The CFM is structured hierarchically to reflect both clinical reality and regulatory mandates in Portugal. Medicines are organized into:

- Major therapeutic groups:

Grupo 1 - Medicamentos anti-infecciosos
Grupo 2 - Sistema nervoso central
Grupo 3 - Aparelho cardiovascular
Grupo 4 - Sangue
Grupo 5 - Aparelho respiratório
Grupo 6 - Aparelho digestivo
Grupo 7 - Aparelho geniturinário
Grupo 8 - Hormonas e medicamentos
Grupo 9 - Aparelho locomotor
Grupo 10 - Medicação antialérgica
Grupo 11 - Nutrição
Grupo 12 - Correctivos da volémia e das alterações electrolíticas
Grupo 13 - Medicamentos usados em afecções cutâneas
Grupo 14 - Medicamentos usados em afecções otorrinolaringológicas:
Grupo 15 - Medicamentos usados em afecções oculares: S01
Grupo 16 -Medicamentos antineoplásicos e imunomoduladores
Grupo 17 - Medicamentos usados no tratamento de intoxicações: V03A B; V03A C; V03A F
Grupo 18 - Vacinas e imunoglobulinas: J07; J06
Grupo 19 - Meios de diagnóstico: V04
Grupo 20 - Material de penso, hemostáticos locais, gases medicinais e outros produtos

- Subgroups that represent finer subdivisions based on therapeutic rationale and pharmacological mechanisms
- Pharmacological classes and chemical groups, prioritizing recognized indication(s) and mechanisms

1.1 - Antibacterianos:
1.1.1 - Penicilinas:
1.1.1.1 - Benzilpenicilinas e fenoximetilpenicilina;
1.1.1.2 - Aminopenicilinas;
1.1.1.3 - Isoxazolilpenicilinas;
1.1.1.4 - Penicilinas antipseudomonas;
1.1.1.5 - Amidinopenicilinas.
1.1.2 - Cefalosporinas:
1.1.2.1 - Cefalosporinas de l.ª geração;
1.1.2.2 - Cefalosporinas de 2.ª geração;
1.1.2.3 - Cefalosporinas de 3.ª geração;
1.1.2.4 - Cefalosporinas de 4.ª geração;
1.1.3 - Monobactamos;
1.1.4 - Carbapenemes;
1.1.5 - Associações de penicilinas com inibidores das lactamases beta;
1.1.6 - Cloranfenicol e tetraciclinas;
1.1.7 - Aminoglicosídeos;
1.1.8 - Macrólidos;
1.1.9 - Sulfonamidas e suas associações;
...

[11] Classificação farmacoterapêutica de medicamentos: https://www.infarmed.pt/documents/15786/1072289/110-AB6_Desp_4742_2014_VF.pdf

Each medicine is assigned to a group and subgroup according to its approved national indications, mechanism of action, and, where relevant, chemical structure. This approach ensures alignment with Portuguese prescribing practices, formulary management, and reimbursement policies, as well as facilitating epidemiological and utilization studies.

A notable feature of the CFM is its systematic mapping to the ATC (Anatomical Therapeutic Chemical) Classification designed by the WHO. Official tables, published as annexes to regulatory decrees, present direct correspondences between each CFM code and its ATC counterpart.

Classificação Farmacoterapêutica	Códigos ATC
1.1 — Antibacterianos:	
1.1.1 — Penicilinas	J01C
1.1.1.1 — Benzilpenicilinas e fenoximetilpenicilina	J01CE
1.1.1.2 — Aminopenicilinas	J01CA
1.1.1.3 — Isoxazolilpenicilinas	J01CF
1.1.1.4 — Penicilinas anti-Pseudomonas	J01CA
1.1.1.5 — Amidinopenicilinas	J01CA
1.1.2 — Cefalosporinas	J01DA
1.1.2.1 — Cefalosporinas de 1ª. Geração	
1.1.2.2 — Cefalosporinas de 2ª. Geração	
1.1.2.3 — Cefalosporinas de 3ª. Geração	
1.1.2.4 — Cefalosporinas de 4ª. Geração	
1.1.3 — Monobactamos	J01DF
1.1.4 — Carbapenemes	J01DH
1.1.5 — Associações de penicilinas com inibidores das beta lactamases	J01CR
1.1.6 —Cloranfenicol e tetraciclinas	J01BA; J01AA
1.1.7 — Aminoglicosídeos	J01G
1.1.8 — Macrólidos	J01FA
1.1.9 —Sulfonamidas e suas associações	J01E
1.1.10 — Quinolonas	J01M
1.1.11 — Outros antibacterianos	J01X
1.1.12 — Antituberculosos	J04A
1.1.13 — Antilepróticos	J04B
1.2 — Antifúngicos	J02; D01BA
1.3 — Antivíricos	J05
1.3.1 — Anti-retrovirais:	

This dual coding facilitates international harmonization and enables researchers, clinicians, and regulatory authorities to conduct comparative analyses using ATC-based global datasets, while retaining the granularity and national specificity of the Portuguese system.

The necessity of maintaining a dedicated national system like the CFM is often questioned considering the widespread adoption and international harmonization provided by the WHO's ATC classification. On one hand, the CFM allows Portugal to tailor drug categorization specifically to national clinical priorities, regulatory authorizations, and reimbursement policies. This can streamline health system management, providing more granular control over prescription guidance, drug financing, and surveillance that reflects domestic therapeutic realities. Such national customization may foster clinical relevance and administrative clarity for Portuguese stakeholders, addressing needs not always prioritized by the global ATC consensus.

However, the existence of the CFM also presents considerable challenges and inefficiencies. Having two parallel systems forces regulatory authorities, healthcare providers, and researchers to engage in duplicative coding, cross-mapping, and administrative conversion whenever drugs are tracked, funded, or studied. This not only increases the risk of errors but also hinders

international comparability, potentially obscuring Portugal's data in multinational pharmacoepidemiologic studies and meta-analyses that rely on standardized ATC codes. Furthermore, when the CFM diverges from the ATC, for example, in assigning "main" therapeutic use or grouping multi-indication drugs, statistical analyses and funding decisions may misalign with global best practices, creating access or equity concerns. For a modern, interconnected research and regulatory environment, such duplication might represent an unnecessary legacy, best addressed by integrating national requirements directly into the evolving ATC framework, thus minimizing inefficiency and maximizing comparability.

5. Other Standardized Systems

SNOMED-CT: Granular Phenotyping in Electronic Health Records

SNOMED Clinical Terms (SNOMED-CT)[12] is a comprehensive, multilingual clinical healthcare terminology that provides a standardized way to represent clinical phrases captured by clinicians. It enables granular phenotyping in **electronic health records** (EHR) by covering a vast range of clinical concepts, including **diseases**, **findings**, **procedures**, **microorganisms**, **pharmaceuticals**, and **body structures**. SNOMED-CT's hierarchical and polyhierarchical structure allows for detailed and precise coding, supporting complex queries and clinical decision support.

Granular phenotyping with SNOMED-CT involves the detailed characterization of patient features, comorbidities, and disease subtypes, which is essential for advanced pharmacoepidemiologic analyses. Researchers can identify specific patient subpopulations, track disease progression, and evaluate drug safety and effectiveness with high precision. SNOMED-CT's integration into EHRs enhances interoperability across healthcare systems and improves data quality by reducing ambiguity and variability in clinical documentation. It also facilitates linkage with other classification systems such as ICD-10/11 and LOINC, supporting comprehensive and multi-dimensional data analysis. SNOMED-CT's role is especially prominent in precision medicine, where detailed phenotypic data is crucial for tailoring therapies and understanding heterogeneous treatment responses.

SNOMED CT is poly-hierarchical, meaning concepts can be connected to multiple parent categories. This enables rich relationships between concepts and supports detailed data capture. SNOMED CT includes over 100,000 unique concepts and extensive synonyms, enabling clinicians to document with high specificity, e.g., distinguishing a left ankle fracture from a right ankle fracture, which in ICD might only be represented as a generic "ankle fracture".

SNOMED CT, in contrast, represents clinical concepts with unique numeric identifiers, supporting compositional and post-coordinated expressions and granularity beyond disease diagnosis, such as findings, procedures, allergies, and body structure relationships.

Post-coordinated expressions of SNOMED CT allow detailed data capture beyond individual concepts using compositional grammar. For instance: Post-coordinated representation of "Intolerance to histamine", but SNOMED CT does not contain a concept that represents this

[12] SNOMED-CT: https://www.nlm.nih.gov/healthit/snomedct/international.html

clinical idea. However, it is possible to represent it using the following post-coordinated expression:

782197009 |intolerance to substance| : 47429007 |associated with| = 54235008 |histamine|

WHODrug

The WHODrug[13], maintained by the Uppsala Monitoring Centre under the World Health Organization, is a standardized global pharmaceutical reference designed to support **pharmacovigilance activities**. It provides a comprehensive, structured, and harmonized database of **trade name**, **active ingredients**, **pharmaceutical form**, **strength**, **country of sales**, and **marketing authorization holder** (MAH), facilitating the coding and analysis of adverse drug reaction (ADR) reports submitted to global databases such as **VigiBase**.

By standardizing drug names, formulations, and ingredients, the WHO Drug Dictionary enables consistent aggregation and comparison of safety data across countries and regulatory jurisdictions. Its hierarchical structure organizes drugs by active substance, pharmaceutical form, and brand names, allowing for detailed identification and grouping. The dictionary supports multiple languages and is updated quarterly to incorporate new drugs and regulatory changes. Each drug entry is linked to its corresponding ATC code, facilitating integration with drug utilization and epidemiological data.

Asking an LMM about the WHODrug Drug Code and Medicinal Product ID (MPID):

Regarding the **unique identifier in the UMC's Drug Dictionary (WHODrug Global)**: The WHODrug Drug Code and Medicinal Product ID (MPID) specific to "Xanax 1 mg 60 comprimidos" in Portugal are proprietary, accessible only by licensed users of WHODrug Global.

CITS database (Cedência de Informação de Tecnologias da Saúde)

The **CITS** database (Cedência de Informação de Tecnologias da Saúde) is a comprehensive platform developed and administered by INFARMED, Portugal's National Authority of Medicines and Health Products. Its primary function is to centralize and disseminate up-to-date information about medicines, medical devices, and other health technologies that are regulated within the Portuguese health system.

CITS emerged from the integration of previous databases, such as **CHNM (Código Hospitalar Nacional do Medicamento)** and **Infomed**, aiming to enhance accessibility, data quality, and interoperability for healthcare institutions, regulatory bodies, software providers, and market authorization holders (MAHs). This unified portal allows authorized users to verify, validate, and export information related to medicinal products, including **pricing**, **reimbursement status**, **packaging**, **market authorization details**, and **commercialization status**. The system updates daily and is accessed via INFARMED's Online Services area, using specific credentials for secure data management.

[13] Lagerlund O, Strese S, Fladvad M, Lindquist M. WHODrug: A Global, Validated and Updated Dictionary for Medicinal Information. Ther Innov Regul Sci. 2020;54(5):1116-1122. doi:10.1007/s43441-020-00130-6

For MAHs, CITS provides restricted access to validate the product data relevant to current and upcoming market periods, with functionalities to filter product lists, export to Excel, and trace regulatory changes. Health institutions, both public and private, use CITS to support clinical practice, electronic prescribing, and pharmacy management software. The portal supports integration with platforms like SGICM (hospital medication management) and PEM (primary care prescribing), thereby streamlining data flows and compliance in the Portuguese healthcare sector.