

## EL7: Medical coding

### ICD, ATC, KVÅ, regex

#### Reading

- [1, ch. 3] on standardized vocabularies. You may skip the sections “CPT”, “LOINC”, “RxNorm” and “Using the Unified Medical Language System” (not examined within the course). Even if you skip those section, remember to read the conclusions in the end of the chapter!
- M. A. Alharbi, G. Isouard, and B. Tolchard [2] provides an historical expose of the development of medical coding, with focus on the International Classificatin of Diseases (ICD).
- S. J. Nelson *et al.* [3] argue (based on a statistical analysis) that we should not put to much trust in the coded data (you may skip the methods section).
- L. J. Bindel and R. Seifert [4] introduces the Anatomical Therapeutic Chemical (ATC) and some associated problems. Focus on the introduction and conclusion sections (results and discussion may be skipped).
- [5, ch. 15]: Using `{stringr}` for regular expressions in R including exercises

#### Recommended references:

- [History of the Statistical Classification of DiseasesandCauses of Death](#)

#### Recommended practice:

- [Regular expressins](#): But please not that this is general practice and deviations may exist between those exercises and R.

#### Overview

- Standardized vocabularies, controlled vocabularies, terminologies and ontologies ...
- This is a field of its own (health informatics)
- Let’s just call it “medical coding” for now.

#### Relevance

- Imaganine you are diagnosed with “cancer” (hope not ...)
- Your doctor writes that you have “kräfte” in your medical records
  - “**kräfte**” (Swedish) = **cancer** (latin), although the astrological sign “cancer” is a “crab” (latin does not distinguish the two)
- She might as well write:
  - The patient was diagnosed with a **malignant neoplasm** of the colon.
  - Histology confirms **invasive adenocarcinoma**.
  - Evidence of **metastatic disease** to the liver.
- Natural languages (English/Swedish/Latin etc) are not well suited for statistical analysis
  - Natural language processing (NLP) is nice but outside the scope of the course

- Statisticians need clear definitions of diagnoses, procedures, medications etc.
- Therefore, such information is encoded in a standardized way

### **Granularity/reliability**

- Cancer might be coded by an ICD-10 code (International Classification of Diseases v. 10) as “C” (or possibly “D”)
- Cancer, however, is a very general term. Is it lung cancer, brain cancer, skin cancer etc (those are very different)
- The more we learn about a disease, the more granularity we expect from the coding
- The coding systems therefore tend to be quite complex, evolve over time and often have regional differences
- Even though the intention of the coding system might be granular and precise, the data quality often relies on different coding practices in different hospitals etc.
- The codes might also be misused for re-imbursment practices
  - There was a regional scandal in Western Sweden not so many years ago ([you may read about int in Swedish](#))
- In practice, the medical doctor might dictate a diagnosis, which then needs to be translated to a code by administrative staff

### **Example**

- The Swedish Hip Arthroplasty Register identified that one hospital appeared to have an unusually high number of patients recorded with severe respiratory problems
- At first glance, this raised a clinical question: could hip problems somehow lead to serious breathing problems?
- However, hip surgery is often performed under general anesthesia. During general anesthesia, patients are intubated and mechanically ventilated, which involves procedures related to the respiratory system.
- It was eventually discovered that a procedural code related to anesthesia and airway management had been incorrectly registered as a severe respiratory diagnosis.
- The apparent “complication” was therefore not a real clinical problem, but a coding error.

**Lesson:** Register data reflect coding practices. Without understanding how variables are defined and recorded, one may draw incorrect conclusions.

### **ICD – International Classification of Diseases**

- Maintained by the World Health Organization (WHO)
- Global standard for coding diseases and causes of death
- Used for:
  - Clinical documentation
  - Mortality statistics
  - Epidemiological research
  - Health system planning and monitoring

## Historical Background

- First version: 1893 (International List of Causes of Death)
- WHO assumed responsibility in 1948 (ICD-6)
- Major revisions approximately every 10–20 years
- Each revision reflects:
  - Advances in medical knowledge
  - Changes in disease concepts
  - Administrative and reporting needs

ICD has evolved from a mortality list to a comprehensive disease classification.

## Major ICD Versions

- ICD-7 (used in many countries in the 1950s–1970s)
  - Still used in the Swedish cancer register for backward compability
- ICD-8 (used in many countries in the 1960s–1980s)
- ICD-9 (widely used until the early 2000s)
  - Also still used in the Swedish cancer register
- ICD-10 (introduced in the 1990s; still dominant in many countries)
  - From 1997 in Sweden. What we currently most care about
- ICD-11 (adopted in 2019; gradually being implemented)

Different countries adopted versions at different times, creating challenges for international comparisons.

## National Modifications

Several countries use national adaptations:

- ICD-10-CM (USA; Clinical Modification)
- ICD-10-CA (Canada)
- ICD-10-SE (Sweden)
  - A fifth position (ignoring the dot) sometimes used for more granularity
  - ICD-10: S72.0 Fracture of neck of femur
  - ICD-10-SE: S72.00 Fracture of neck of femur, closed; S72.01 Fracture of neck of femur, open; S72.10 Pertrochanteric fracture, closed; S72.11 Pertrochanteric fracture, open, ...

---

Feature	WHO ICD-10	ICD-10-SE (Sweden)	ICD-10-CM (USA)
Maintained by	WHO	Swedish National Board of Health and Welfare (Socialstyrelsen)	U.S. National Center for Health Statistics (NCHS)
Primary purpose	Global disease classification	National clinical and statistical reporting	Clinical documentation and reimbursement

Feature	WHO ICD-10	ICD-10-SE (Sweden)	ICD-10-CM (USA)
Level of detail	Moderate	More detailed than WHO ICD-10	Much more detailed than WHO ICD-10
Additional digits	Typically 3–4 characters	Often includes 5th character extensions	Up to 7 characters
Laterality (right/left)	Usually not specified	Limited	Frequently specified
Encounter type (initial, follow-up, sequela)	Not included	Not included	Explicitly coded
Administrative focus	Epidemiology and mortality statistics	Clinical and national register reporting	Strongly tied to billing and reimbursement
International comparability	High (reference standard)	High within Nordic context, requires mapping internationally	Requires crosswalk to WHO ICD-10 for comparison

## Structure of ICD-10

Typical format:

- One letter (chapter)
- Two digits (category)
- Optional dot
- additional digit(s) (subcategory)

Example:

- **I21** – Acute myocardial infarction (AMI)
- **I21.0** – AMI of anterior wall
- **I21.9** – AMI, unspecified

## Hierarchy

- Chapter → Block → Category (3-digit) → Subcategory (4-digit+)

Researchers must decide:

- Analyse at 3-digit level?
- Or at more detailed subcategory level?

There is a trade-off between specificity and statistical power.

## What Does an ICD Code Represent?

An ICD code reflects:

- Clinical documentation

- Coding rules
- Administrative structure
- Local practice

It does not necessarily reflect:

- Biological mechanism
- Diagnostic certainty
- Uniform clinical interpretation

Register data therefore reflect both medicine and administration.

## Changes Over Time

Between ICD versions:

- Codes may be split (1-to-many) into more detailed categories (common)
- Codes may be merged (many-to-one, although uncommon)
- Codes may move between chapters (affects the aggregated chapter counts/incidence/prevalence)
- Definitions may change

Example: A condition classified under one chapter in ICD-9 may appear elsewhere in ICD-10.

Implication: Observed changes in incidence may reflect coding changes rather than true epidemiology.

## Crosswalks Between Versions

When analyzing long time series:

- Mapping tables (“crosswalks”) are often used
- Mapping may be:
  - One-to-one
  - One-to-many
  - Many-to-one

Crosswalks are rarely exact. Information loss or ambiguity is common.

Aggregation to broader diagnostic groups is often necessary.

## Crosswalk Patterns (WHO ICD-9 → WHO ICD-10)

When mapping between ICD versions, different structural relationships may occur.

Mapping type	ICD-9 (WHO)	ICD-10 (WHO)	Interpretation
One → Many (Split)	250 – Diabetes mellitus	E10–E14	One broad ICD-9 category split into multiple etiological types
Many → One (Merge)	038 – Septicaemia; 790.7 – Bacteraemia	A41 – Other sepsis	Separate ICD-9 concepts consolidated

Mapping type	ICD-9 (WHO)	ICD-10 (WHO)	Interpretation
			into broader ICD-10 category
Many ↔ Many (Re-organisation)	296 – Affective psychoses; 300 – Neurotic disorders	F30–F39 (Mood disorders); F40–F48 (Neurotic disorders)	Structural reorganisation and conceptual reclassification across chapters
Chapter relocation	011 – Pulmonary tuberculosis	A15 – Respiratory tuberculosis	Infectious diseases reorganised under new chapter structure

### Crosswalk before or after

- The Swedish cancer register has an internal “crosswalk” applied uniformly to the register itself
  - Updated regularly, [2026 version with 360 pages \(in Swedish\)](#)
- Other registers typically only records the current version in use
- If so, you might perform the crosswalk yourself after receiving the data
  - Applies if you want to look at longer time trends etc or combine data from different periods

### ICD-O

ICD-O (International Classification of Diseases for Oncology):

- Used mainly in cancer registries
- Combines:
  - Topography (tumour site)
  - Morphology (histology and behaviour)
- Current commonly used version: - ICD-O-3
- Earlier versions: ICD-O, ICD-O-2
- ICD-O is more detailed than ICD-10 for cancer incidence studies.

### Relationship Between ICD-10 and ICD-O

- ICD-10 commonly used for mortality and hospital discharge diagnoses
- ICD-O primarily used for cancer registry incidence data

A cancer case may have:

- An ICD-10 code in hospital data
- An ICD-O morphology and topography code in a cancer registry

Researchers must understand which system underlies their dataset.

## Läpp C00

Kliniskt läge C00	ICD-O/3.2	ICD-O/2	ICD10	ICD9	ICD7
Övertäppens yttersida (läppröda)	C00.0	C00.0	C00.0	140.0	140.0
Underläppens yttersida (läppröda)	C00.1	C00.1	C00.1	140.1	140.1
Läppens yttersida (läppröda), läpp ej angiven	C00.2	C00.2	C00.2	140.9	140.9
Övertäppens insida (slemhinna)	C00.3	C00.3	C00.3	140.3	140.0
Underläppens insida (slemhinna)	C00.4	C00.4	C00.4	140.4	140.1
Läppens insida (slemhinna), läpp ej angiven	C00.5	C00.5	C00.5	140.5	140.9
Läppkommissur eller mungipa	C00.6	C00.6	C00.6	140.6	140.9
Överväxt till/från angränsande sublokal inom läpp (C00) med okänt ursprung	C00.8	C00.9	C00.9	140.9	140.9
Läpp UNS	C00.9	C00.9	C00.9	140.9	140.9
Vid flera samtida tumörer inom läge C00 ska aktuell(a) ICD-O/3.2-kod(er) registreras, men med följande översättning		C00.8	C00.8	140.8	140.8

Se lägeskommentarer inklusive kommentarer om multifokalt läge s 18.

Morfologisk typ C00	ICD-O/3.2	ICD-O/2	C24/hist
<b>Skivepiteltumörer</b>			
Skivepitelcarcinom	80703	80703	146
Skivepitelcarcinom hornbildande/keratiniserande	80713	80703	146
Skivepitelcarcinom icke-hornbildande/icke-keratiniserande	80723	80703	146
Skivepitelcarcinom HPV-pos	80853	80703	146
Skivepitelcarcinom HPV-neg	80863	80703	146
Skivepitelcarcinom basaloitt	80833	80703	146
Skivepitelcarcinom papillärt	80523	80703	146
Skivepitelcarcinom spolcelligt	80743	80703	146
Skivepitelcarcinom verruköst	80513	80703	146
Skivepiteldysplasi höggradig HSIL	80772/b	80702/b	144/b
Adenoskvamöst carcinom	85603	85603	196

Tabellen fortsätter på följande sida.

Se morfologiska kommentarer för ÖNH-lägena på s 20.

## Variation in Coding

Differences between hospitals or regions may arise due to:

- Coding training
  - A primary health care unit may encounter all possible diagnosis (wide but shallow knowledge) while a very specialized unit might have routines for a very narrow but detailed coding
- Local guidelines
  - Regions are independent in Sweden
- Administrative incentives
- Reimbursement systems
  - public and private health care providers may have different incentives
- Electronic health record design
  - National registers often relies on combining multiple different sources
- Somatic vs psychiatric care
  - In psychiatric care, diagnoses may sometimes be recorded with less specificity, potentially due to concerns about stigma or the sensitive nature of certain conditions

Registers capture both clinical events and coding behaviour.

## Validity of ICD Codes

Important research question:

- Does the code correspond to the true disease?

Validation studies compare ICD codes with a reference standard (e.g., chart review, clinical registry, laboratory confirmation).

## Key Measures of Validity

- **Sensitivity**

Among patients who truly have the disease,  
how many receive the correct ICD code?  
→ Measures undercoding (missed cases).

- **Specificity**

Among patients who do not have the disease,  
how many are correctly not assigned the code?  
→ Measures overcoding (false positives).

- **Positive Predictive Value (PPV)**

Among patients assigned the ICD code,  
how many truly have the disease?  
→ Measures how reliable the code is for identifying true cases.

## Conceptual 2×2 Table

	True disease	No disease
ICD code present	True positive	False positive
ICD code absent	False negative	True negative

- Sensitivity = True positives / (True positives + False negatives)
- Specificity = True negatives / (True negatives + False positives)
- PPV = True positives / (True positives + False positives)

## Why It Matters

Low sensitivity → underestimated incidence

Low PPV → inflated case counts

Variation in validity may depend on:

- Diagnosis (e.g., myocardial infarction vs mild depression)
- Care setting (inpatient vs primary care)
- Time period (coding changes)
- ICD version and national modification

Not all ICD codes are equally reliable for research.

## Practical Implications for Statisticians

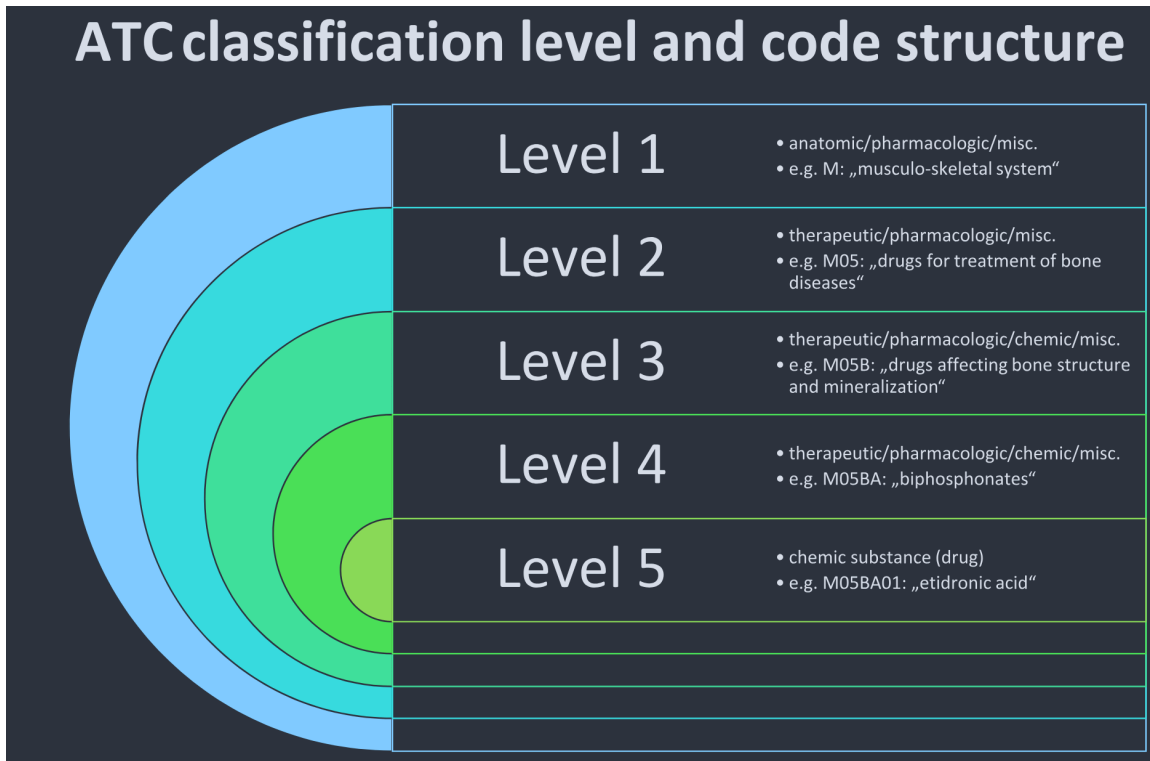
Before analysis, always clarify:

- Which ICD version?
- Which national modification?
- Which coding level (3-digit vs 4-digit)?
- Has coding practice changed over time?
- Are crosswalks required?
- Is there validation evidence for the diagnosis?

ICD is a classification system. It is not identical to clinical truth. Transparent documentation of code selection is essential for reproducible research.

## ATC for drugs

- Anatomical Therapeutic Chemical (ATC) classification
- categorizing therapeutic drugs,
- structured into 14 main groups and 5 levels, with a disease-oriented focus
- introduced in the 1960s
- In 1980, the World Health Organization (WHO) recommended the ATC system as the “state of the art”
- follow a hierarchical structure: 1 letter, 2 digits, 2 letters, 2 digits
  - Example: C09AA05
- Several problems exist [4] but it is nevertheless widely used



## Sweden

- We have ATC codes in the national prescription register
- When a doctor prescribes a medication, the ATC code will follow automatically
- Less room for interpretation when coding
- New medications are introduced over time
- The Swedish Medical Products Agency (MPA; Läkemedelsverket) make such decisions
- Daily updates to the [National Substance Register](#)

## Procedure codes

- We use ICD for diagnoses and medical condition
- But how are patients with such diagnosis treated?
- What actions (in addition to the prescription of medicines) do we have?
- USA has a special version of ICD for this: ICD-10-PCS
  - PCS = Procedure Coding System

## NOMESCO

In Sweden, medical procedures are coded using the **NOMESCO Classification of Surgical Procedures (NCSP)**.

- Developed by the Nordic Medico-Statistical Committee (NOMESCO)
- Used in Sweden, Denmark, Finland, Norway, and Iceland
- Primarily for on surgical procedures

## **KVÅ**

- In Sweden implemented through **KVÅ** (Klassifikation av vårdåtgärder)
- Maintained nationally by Socialstyrelsen
- Includes the NOMESCO-NCSP codes for surgery
- Also includes additional codes for non-surgical treatments and activities
  - Administration of chemotherapy (cytostatic treatment)
  - Radiotherapy sessions
  - Dialysis treatment (hemodialysis, peritoneal dialysis)
  - Blood transfusion
  - Vaccination
  - Advanced wound care (non-surgical)
  - Multidisciplinary team conference (MDT conference)
  - Smoking cessation counselling
  - Nutritional counselling
  - Physiotherapy interventions
  - Occupational therapy interventions
  - Psychotherapeutic treatment sessions
  - Structured patient education programmes
  - Palliative care planning

## **Structure**

- Alphanumeric codes (typically 5 characters)
- First letter indicates anatomical or procedural group
- Subsequent characters specify procedure type and detail

Example:

- **NFB49** – Primary total hip replacement
- **JKA20** – Appendectomy

## **Purpose**

- Record surgical and certain non-surgical interventions
- Used in:
  - National Patient Register
  - Quality registers
  - Reimbursement and administrative reporting
  - Health services research

## **Important Distinction**

- **ICD-10-SE** → Diagnosis codes
- **NOMESCO/KVÅ** → Procedure codes

A patient record may therefore contain:

- An ICD diagnosis (e.g., hip fracture)

- A NOMESCO procedure code (e.g., hip replacement surgery)

### **Implications for Research**

- Diagnosis and procedure must not be confused
- Trends in procedures may reflect:
  - Clinical practice changes
  - Technology changes
  - Policy and reimbursement incentives
- International comparisons require awareness that other countries (e.g., USA) use different coding systems

NOMESCO codes capture *what was done*, not *what disease the patient had*.

### **DRG**

**DRG (Diagnosis-Related Groups)** is a classification system used to group hospital cases into categories expected to require similar levels of resources.

In Sweden:

- Based on the **NordDRG** system
- Used for:
  - Hospital reimbursement
  - Resource allocation
  - Health care management
  - Productivity and efficiency analyses

### **How Is a DRG Determined?**

A DRG is assigned based on a combination of:

- Primary diagnosis (ICD-10-SE)
- Secondary diagnoses
- Procedure codes (NOMESCO/KVÅ)
- Age
- Sex
- Discharge status
- Presence of complications or comorbidities

DRG codes are therefore derived classifications, not primary clinical codes.

### **Implications for Research**

- DRG reflects resource use, not disease incidence.
- Changes in reimbursement rules may influence coding behavior.
- Regional comparisons must consider administrative incentives.
- DRG is suitable for health services and economic analyses, but less appropriate for etiological research.

## SNOMED CT – What Is It?

SNOMED CT (**S**ystematized **N**omenclature of **M**edicine – **C**linical **T**erms) is a large clinical terminology system.

- Maintained by **SNOMED International**
- Contains **hundreds of thousands of clinical concepts**
- Designed for **structured documentation in electronic health records**

Unlike ICD or ATC, SNOMED CT is primarily a **terminology**, not a statistical classification.

## Terminology vs Classification

System	Type	Purpose
ICD	Classification	Epidemiology and health statistics
ATC	Classification	Drug classification
KVÅ / NOMESCO	Classification	Medical procedures
<b>SNOMED CT</b>	Terminology	Detailed clinical documentation

Classification systems simplify reality for **statistics and reporting**, while terminologies allow **very detailed clinical descriptions**.

## Why SNOMED CT Is Not Widely Used in Registers

Despite its strengths, SNOMED CT is rarely used directly in:

- national health registers
- epidemiological statistics

Main reasons:

- **Too detailed** for statistical aggregation
- Harder to ensure consistent coding
- Statistical reporting systems are built around **ICD**

## Regular expressions

### Regular Expressions (Regex)

- A way to describe **patterns in text**
- Used to:
  - Identify diagnosis codes (ICD-10)
  - Identify drug codes (ATC)
  - Clean register data
  - Validate variables
- In R:
  - Base R: `grep()`, `sub()`, `gsub()`
  - Tidyverse: `stringr::str_detect()`, `str_extract()`, `str_replace()`

- More efficient for big data: [stringfish](#)

## Relevance

Typical use cases:

- Select all ICD-10 codes starting with "I21" (acute myocardial infarction)
  - Identify all ATC codes beginning with "C09" (antihypertensives)
  - Check for malformed codes (quality control)
  - Extract codes embedded in free text (e.g., notes, text fields)
- 

## Basic Building Blocks

Symbol	Meaning
<code>^</code>	Start of string
<code>\$</code>	End of string
<code>.</code>	Any character
<code>*</code>	0 or more repetitions
<code>+</code>	1 or more repetitions
<code>?</code>	0 or 1 repetition
<code>{m, n}</code>	Between m and n repetitions
<code>[ABC]</code>	Any of A, B, or C
<code>[0-9]</code>	Any digit
<code>\\d</code>	Any digit (PCRE)

### ! Different versions

There are different implementations of regular expressions! The implementation in base R is described by `?base::regex` in R. Perl-like Regular Expressions (PCRE) is a commonly used alternative requiring `(perl = TRUE` as argument) for the base functions.

## Example: ICD-10 Structure

ICD-10 codes typically follow:

- One letter
- Two digits
- Optional dot and additional digit

Regex pattern: `^[A-Z][0-9]{2}(\\. [0-9])?$`

where `\\` is used to remove the special meaning of `.` as described above. Hence, in this case `\\.`  is interpreted as a literal `.` as to be found in the character string. In R:

```

grepl("^[A-Z][0-9]{2}(\\. [0-9])?$", icd) # base
stringr::str_detect(icd, "^[A-Z][0-9]{2}(\\. [0-9])?$")
# Faster and using 10 CPU cores in parallel (only relevant for "big enough
data"):
stringfish::sf_grepl(icd, "^[A-Z][0-9]{2}(\\. [0-9])?$", nthreads = 10L)

```

## Example: Select a Diagnosis Group

- All acute myocardial infarction codes: ^I21
- In R: `stringr::str_detect(icd, "^I21")`

This selects:

- I21
- I21.0
- I21.9

But not:

- I20
- I22

## ATC Code Structure

- Example: C09AA05
- Regexp pattern: `^[A-Z][0-9]{2}[A-Z]{2}[0-9]{2}$`
- In R: `stringr::str_detect(atc, "^[A-Z][0-9]{2}[A-Z]{2}[0-9]{2}$")`
- This will find any ATC code.
- You might receive data with a variable supposed to contain only ATC codes
- It might as well contain other information such as `??`, `don't know`, `XXXXXXX` etc
- You might replace such character strings by `<NA>`

## Implementations

- base R and `{stringr}` both use the same underlying regex engine (PCRE)
  - but `{stringr}` is more “user friendly”.
- Stringfish seems technically superior but is less maintained (more of a hobby project).

## Common Mistakes

- Forgetting `^` when matching prefixes
  - This is problematic even in the `stringfish::sf_starts()` implementation! [See bug report.](#)
- Forgetting to escape `.`
- Not validating full string with `$`
- Overmatching (e.g., `I2` instead of `^I21`)

## Standardised Groupings

To account for overall disease burden, researchers often use established grouping systems, such as:

- **Charlson Comorbidity Index** (ICD)
- **Elixhauser Comorbidity Index** (ICD)
- Similar groupings of ATC-codes
- Combinations of those

These indices:

- Aggregate multiple ICD codes into clinically meaningful comorbidity categories
- Are commonly used for:
  - Risk adjustment
  - Prognostic modelling
  - Confounding control in observational studies

### {decoder}

- The R package {decoder} provides descriptions for many commonly used coding systems.
- In register data, you often only have the raw codes (e.g., ICD, ATC), without textual labels.
- {decoder} allows you to translate codes into meaningful descriptions (in Swedish or English), making interpretation easier and more transparent.

**i** Up-to-date?

I am the maintainer of {decoder} and {coder} but I have not had the time or energy to update them for a couple of years. There are some reported issues.

### {coder}

- The R package {coder} can be used to aggregate individual diagnosis codes into broader clinical categories.
- Common applications include:
  - **Charlson Comorbidity Index**
  - **Elixhauser Comorbidity Index**
  - Other diagnosis-based groupings

This allows:

- Standardised comorbidity adjustment
- (Sort of/relatively ...) transparent and reproducible case definitions
- Consistent grouping across studies (hopefully)

---

## Bibliography

- [1] A. Nguyen, *Hands-on healthcare data: taming the complexity of real-world data*, First edition. Beijing Boston Farnham Sebastopol Tokyo: O'Reilly, 2022.

- [2] M. A. Alharbi, G. Isouard, and B. Tolchard, "Historical development of the statistical classification of causes of death and diseases," *Cogent Medicine*, vol. 8, no. 1, p. 1893422, Jan. 2021, doi: [10.1080/2331205X.2021.1893422](https://doi.org/10.1080/2331205X.2021.1893422).
- [3] S. J. Nelson *et al.*, "Are ICD codes reliable for observational studies? Assessing coding consistency for data quality," *DIGITAL HEALTH*, vol. 10, p. 20552076241297056, Sep. 2024, doi: [10.1177/20552076241297056](https://doi.org/10.1177/20552076241297056).
- [4] L. J. Bindel and R. Seifert, "Problems associated with the ATC system of drug classification," *Naunyn-Schmiedeberg's Archives of Pharmacology*, Dec. 2025, doi: [10.1007/s00210-025-04833-1](https://doi.org/10.1007/s00210-025-04833-1).
- [5] H. Wickham, M. Çetinkaya-Rundel, and G. Grolemund, "R for Data Science (2e)." [Online]. Available: <https://r4ds.hadley.nz/>