University of Strathclyde Science

Accuracy of drug-related information on Wikipedia

Luca Bartek

18th February, 2016

Why are we interested?

- Free, well-known, multilingual resource
- Wikipedia has over 10,000 drug-related pages
- Wikipedia is on the first page of drug-related Google search results ~80% of the time[1]
- 51% of students of Medicine use Wikipedia as source[2]
- Has been researched previously – but only small-scale[3]

1. Laurent et al., *Seeking Health Information Online: Does Wikipedia Matter?* J.Am Med Inf Assc, 2009. **16**(4): p. 471-479.
2. Judd et al., *Expediency-based practice? Medical students' reliance on Google and Wikipedia for biomedical inquiries.* British J. Ed. Tech., 2011. **42**(2): p. 351-360.
3. Kraenbring et al., *Accuracy and completeness of drug information in Wikipedia: a comparison with standard textbooks of pharmacology.* PLoS One, 2014. **9**(9): p. e106930.

What kind of information is available on Wikipedia?

Not logged in Talk Contributions Create account Log in

Article Talk | Read Edit View history | Search

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

Print/export
Create a book
Download as PDF
Printable version

Languages
Afrikaans
Alemannisch

Aspirin

Title. Main content. Table of contents. Infobox.

From Wikipedia, the free encyclopedia

Aspirin, also known as **acetylsalicylic acid** (**ASA**), is a medication, often used to treat pain, fever, and inflammation.[2] Aspirin is also used long-term, at low doses, to help prevent heart attacks, strokes, and blood clot formation in people at high risk of developing blood clots.[3] Low doses of aspirin may be given immediately after a heart attack to reduce the risk of another heart attack or the death of heart tissue.[4][5] Aspirin may be effective at preventing certain types of cancer, particularly colorectal cancer.[6][7][8]

The main side effects of aspirin are gastric ulcers, stomach bleeding, and ringing in the ears, especially with higher doses. While daily aspirin can help prevent a clot-related stroke, it may increase risk of a bleeding stroke (hemorrhagic stroke).[9] In children and adolescents, aspirin is not recommended for flu-like symptoms or viral illnesses, because of the risk of Reye's syndrome.[10]

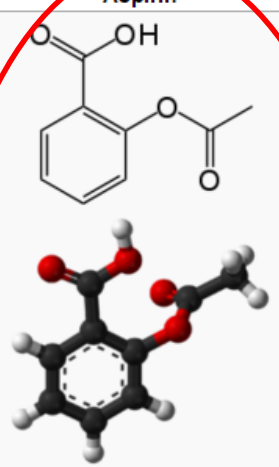
Aspirin is part of a group of medications called nonsteroidal anti-inflammatory drugs (NSAIDs), but differs from most other NSAIDs in the mechanism of action. The salicylates have similar effects (antipyretic, anti-inflammatory, analgesic) to the other NSAIDs and inhibit the same enzyme cyclooxygenase (COX), but aspirin does so in an irreversible manner and, unlike others, affects the COX-1 variant more than the COX-2 variant of the enzyme.[11] Aspirin also has an antiplatelet effect by stopping the binding together of platelets.

The therapeutic properties of willow tree bark have been known for at least 2,400 years, with Hippocrates prescribing it for headaches.[12] Salicylic acid, the active ingredient of aspirin, was first isolated from the bark of the willow tree in 1763 by Edward Stone of Wadham College, University of Oxford.[13] Felix Hoffmann, a chemist at Bayer, is credited with the synthesis of aspirin in 1897, though whether this was of his own initiative or under the direction of Arthur Eichengrün is controversial.[14][15] Aspirin is one of the most widely used medications in the world with an estimated 40,000 tonnes of it being consumed each year.[16] In countries where "Aspirin" is a registered trademark owned by Bayer, the generic term is acetylsalicylic acid (ASA).[17] It is on the WHO Model List of Essential Medicines, the most important medications needed in a basic health system.[18]

Contents [hide]

1 Medical use
- 1.1 Pain
- 1.2 Fever
- 1.3 Inflammation
- 1.4 Heart attacks and strokes
- 1.5 Cancer prevention

Aspirin



Systematic (IUPAC) name	
2-(acetoxy)benzoic acid	
Clinical data	
Pronunciation	acetylsalicylic acid /əˌsiːtəlˌsælɪˈsɪlɪk/
AHFS/Drugs.com	monograph
MedlinePlus	a682878
Pregnancy category	AU: C US: C (Risk not ruled out) D in the 3rd trimester
Legal status	AU: S2 (Pharmacy only) except when given intravenously (in which case

Infobox (drugbox). Contains specific information for each type of article, has a template.

What data was retrieved and how?

- Wikipedia API (mediawiki.org)
- Get a list of pages in which an “Infobox drug” Template is embedded
- Use list to retrieve timestamps for all revisions, and the content of section 0 for the latest revision
- Extract physical, chemical properties and database ID’s including those which can be confirmed by OPS

What data was retrieved?

Physical & Chemical Properties

- Smiles
- Standard & Non-Standard InChIs
- Standard & Non-Standard InChIKeys
- Molecular weight
- Molecular formula

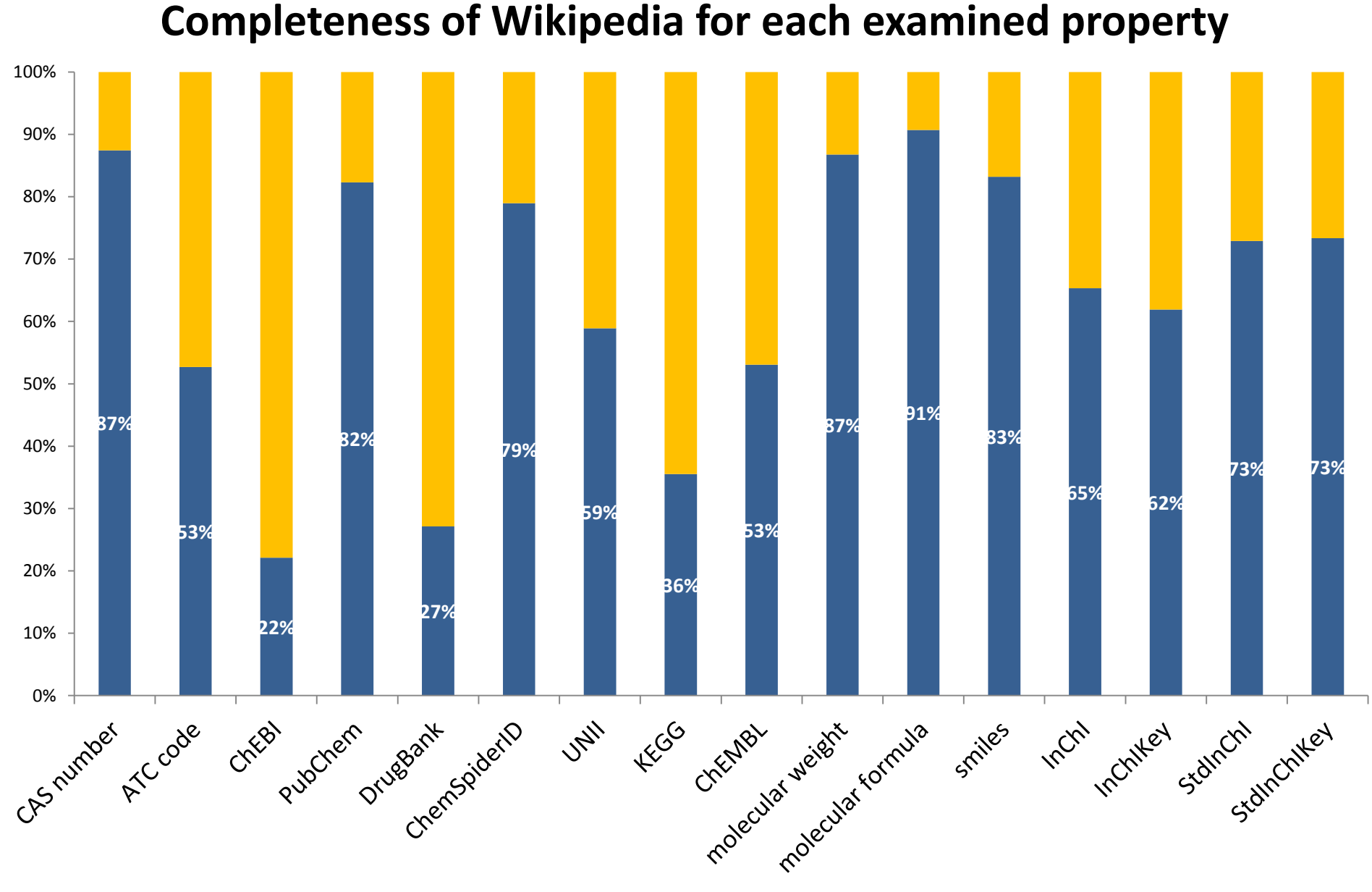
Database ID's

- CAS number
- DrugBank ID
- ChEMBL ID
- ChEBI ID
- PubChem ID
- KEGG ID
- UNII ID
- ATC code

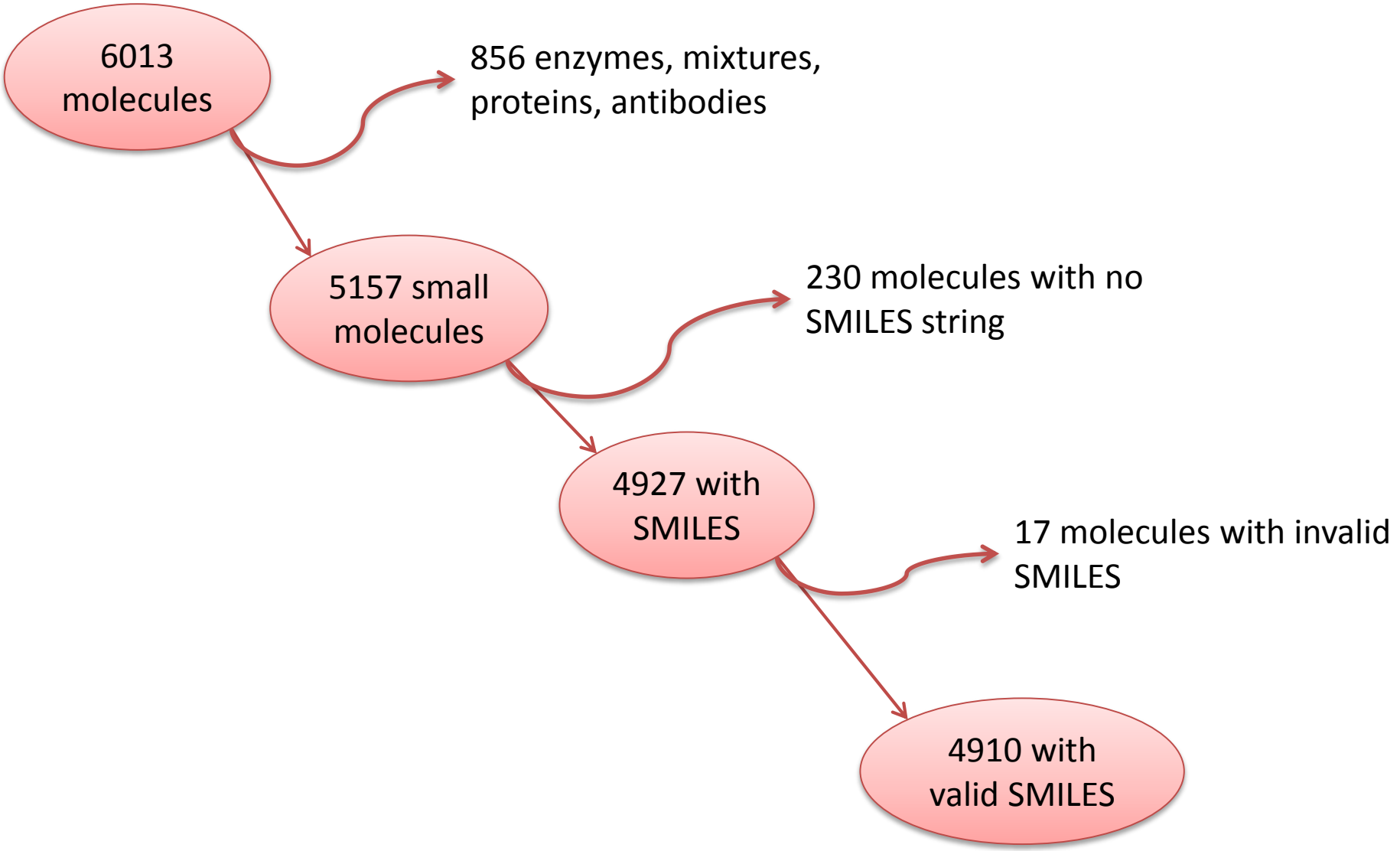
How to verify all this data?

- Does all the information match with each other?
  - Can check for physical/chemical data
- Does this data actually correspond to given drug? Do the ID's also match?
  - A reliable "gold standard" is needed – enter Open PHACTS
- Does the number of edits / age of the article have an effect on accuracy?

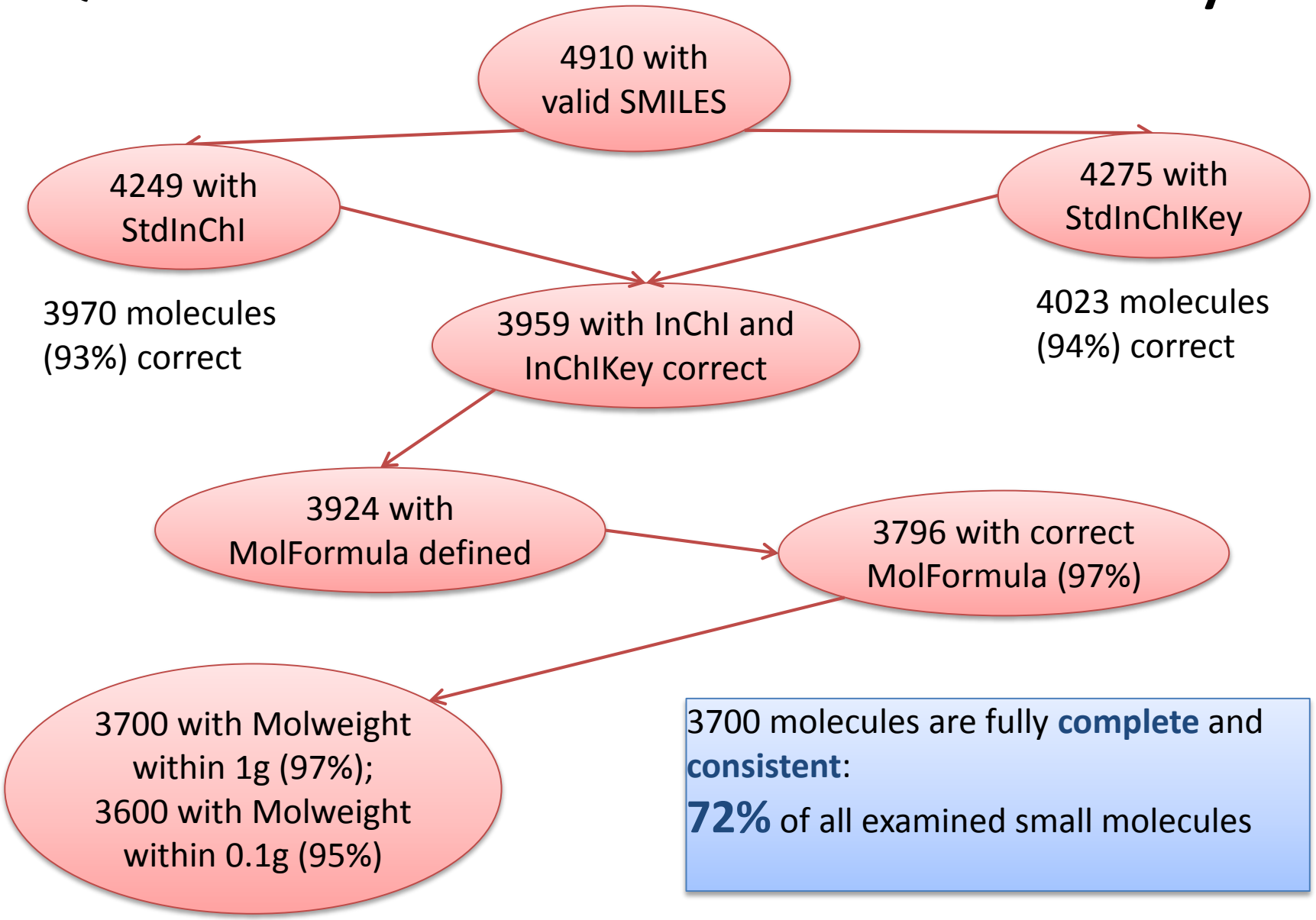
Question 1 – Internal consistency



Question 1 – Internal consistency



Question 1 – Internal consistency



3700 molecules are fully **complete** and **consistent**:

72% of all examined small molecules

Question 2 – External verification

Input for Open PHACTS?
Need a URI.

Drugs matched via identifiers.org → HMDB

- Identifiers.org URI's created in the format *http://info.identifiers.org/wikipedia.en/drug*
- Used as an input to Map URI API call
- Retrieve HMDB URI
- Successful for 1158 drugs

Total of **4031** drugs matched between Open PHACTS and Wikipedia.

Drugs matched via Concept search

- Compound names used as an input for Get concept description API call, using the UUID for Chemical viewed structurally
- Exact matches are annotated:

Drug

- 2609 drugs successfully matched with this method
- Another 264 matched manually (more than 1)

Question 2 – External verification

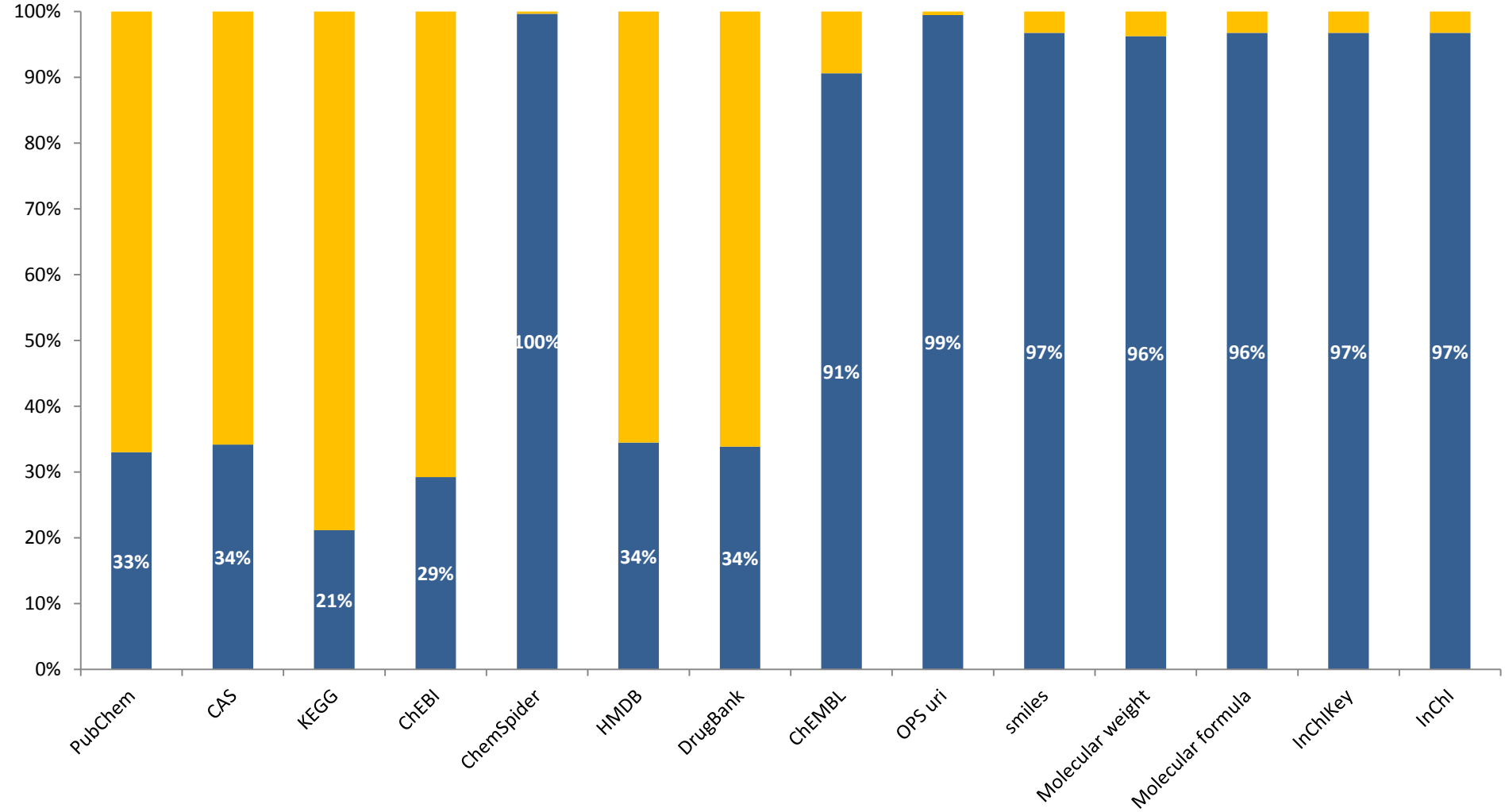
- Previously acquired URI's used as input
- Retrieving ID's – Using the Map URI Call
- targetUriPattern parameter used to filter results
- The identifiers were taken from the result URI's via regular expressions
- Acquired identifiers:
  - -PubChem
  - -CAS
  - -ChEBI
  - -ChemSpider
  - -DrugBank
  - -ChEMBL
  - -OPS
  - -(KEGG)
  - -(HMDB)

Question 2 – External verification

- Retrieving physical/chemical data using the Compound information API call
- OPS URI is used as an input (no OPS URI = no compound information)
- The following properties were extracted:
  - -Smiles
  - -Molecular weight
  - -Molecular formula
  - -InChIKey
  - -InChI

Question 2 – External verification

Completeness of Open PHACTS for each examined property

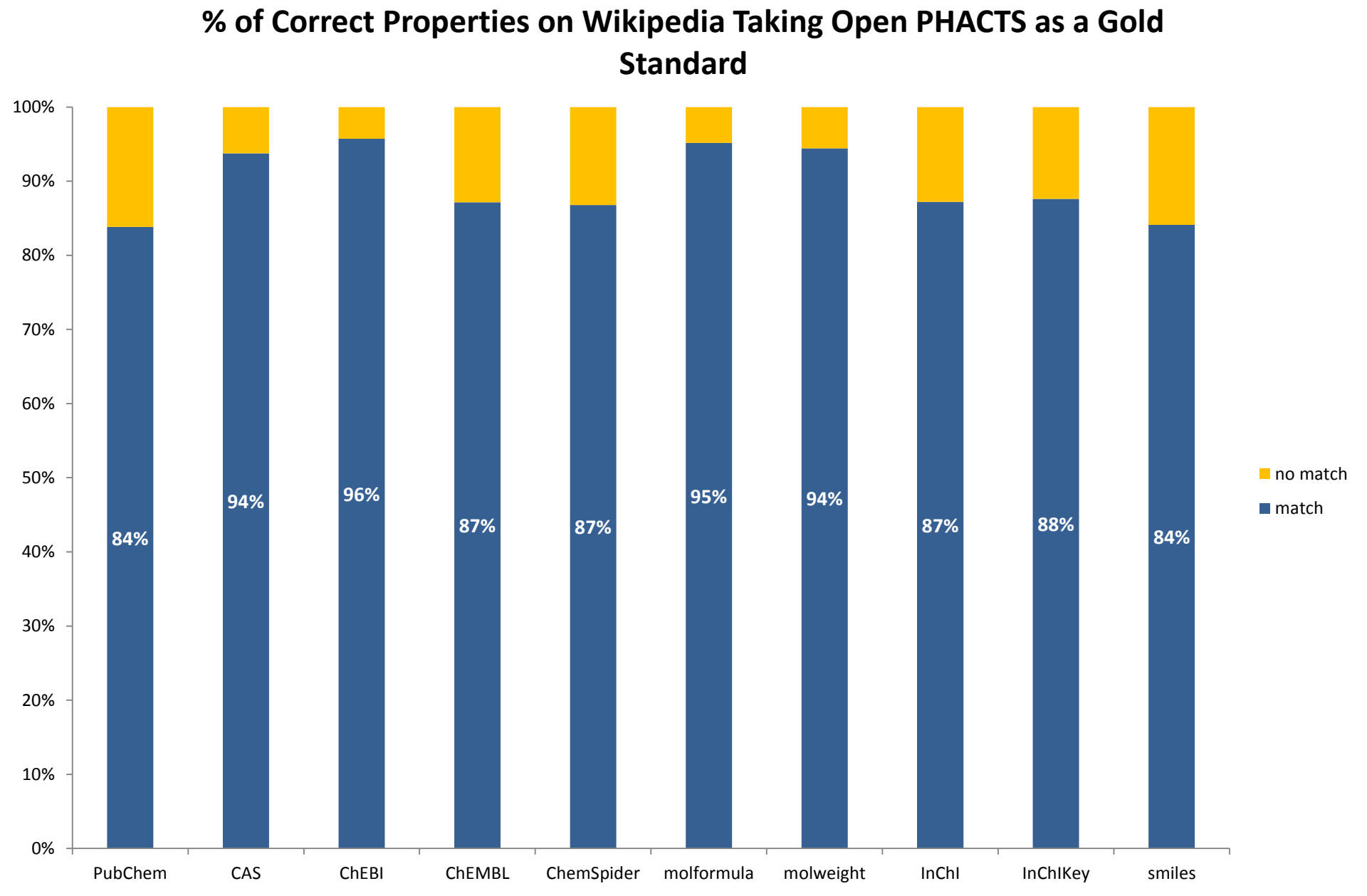


Question 2 – External verification

Method:

- For ID's: string matching
- SMILES:
  - A molecule is created from both smiles
  - It is re-converted into standard InChIs
- InChI & InChIKey: string matching
- Molecular formula: an array with the atom types and the numbers in each element
- Molecular weight: rounded to whole numbers

Question 2 – External verification



Question 2 – External verification

- Lowest accuracy: SMILES
  - 64% of cases: stereochemistry
  - Salt vs. parent drug e.g. Edrophonium vs Edrophonium Chloride
  - © vs @ in smiles
- ID with lowest accuracy:
  - PubChem

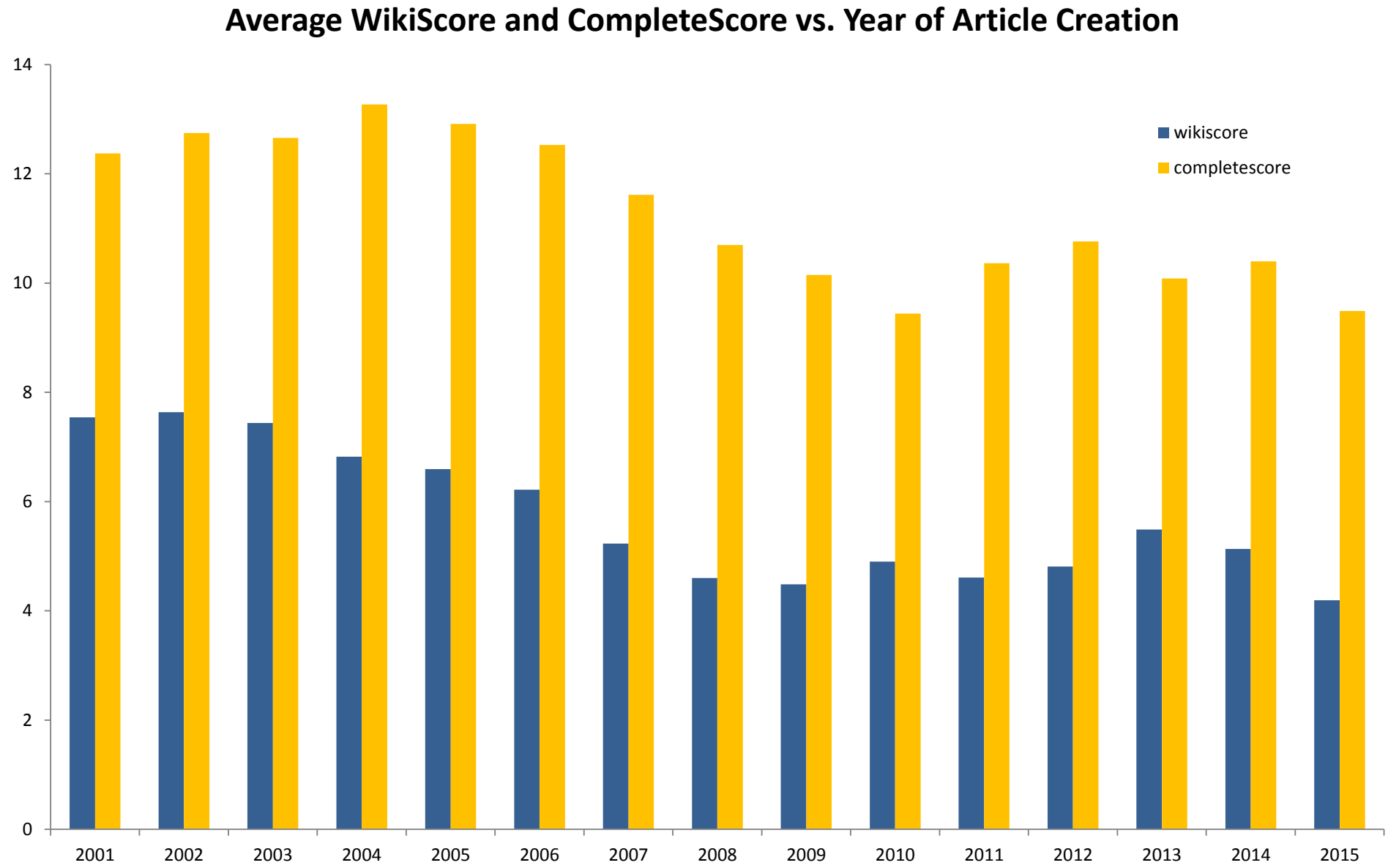
Question 3 – What influences completeness/correctness on Wikipedia?

- Two scores created
- CompleteScore:
  - Measure of how complete data is

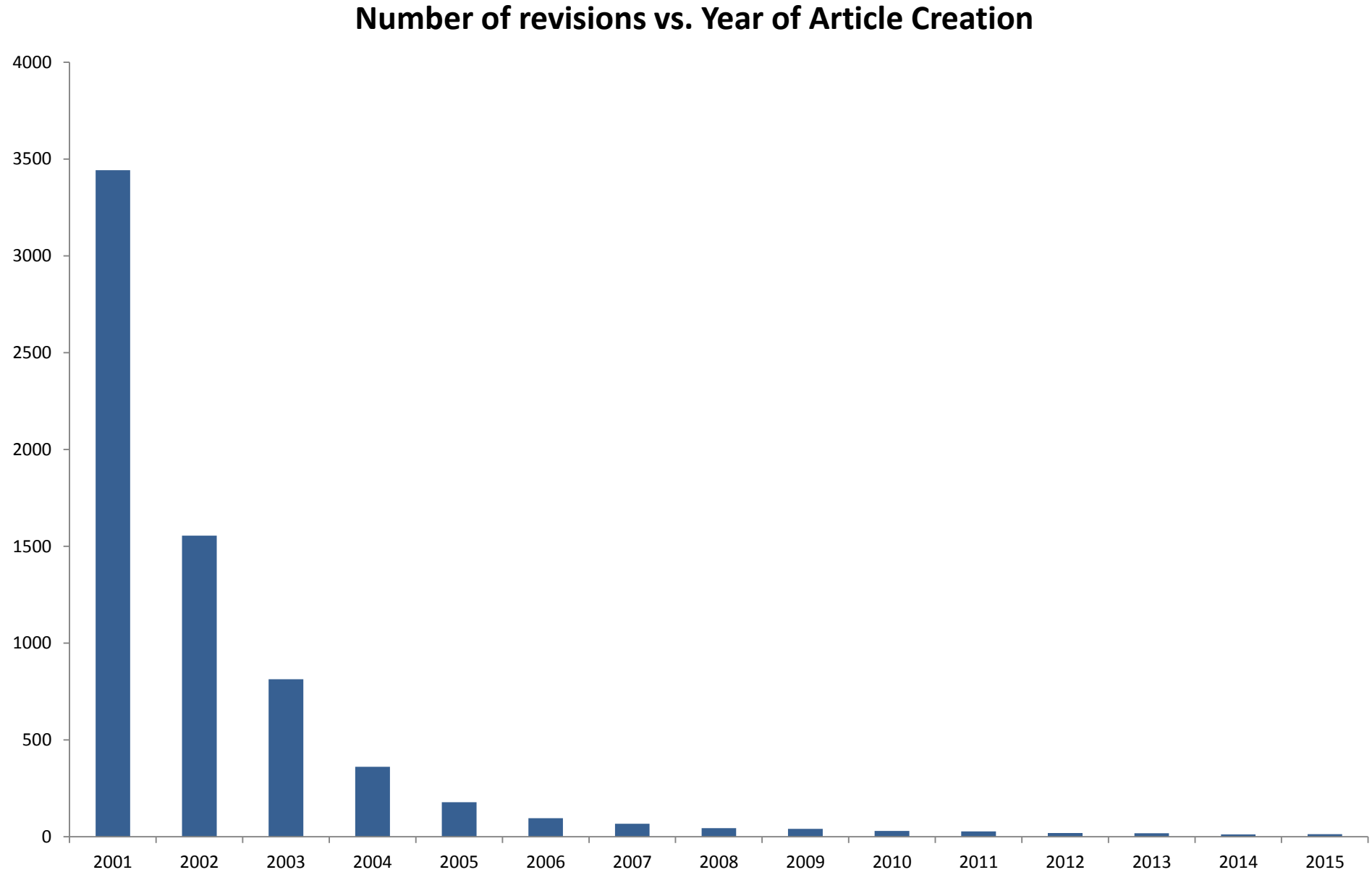
    CompleteScore = # data defined
- WikiScore
  - Measure of both quality and completeness of data

    WikiScore = (# correct)x1 + (# incorrect)x(-1)

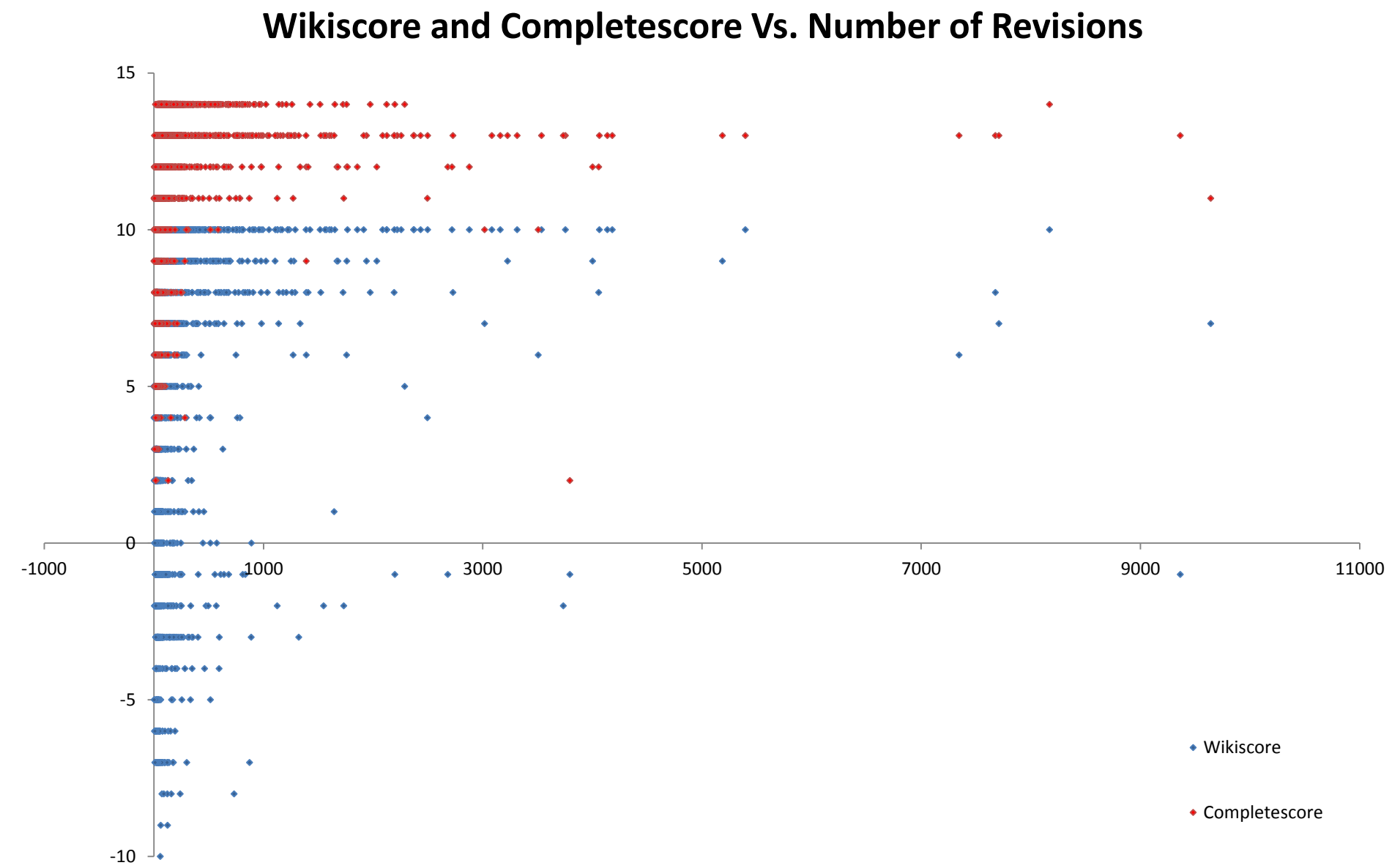
Question 3 – What influences completeness/correctness on Wikipedia?



Question 3 – What influences completeness/correctness on Wikipedia?



Question 3 – What influences completeness/correctness on Wikipedia?



Main outcome & What's next

- Confidence in the accuracy of Drug related information on Wikipedia
- Mapping between OPS and Wikipedia for over 4000 drugs
  - can be added to OPS mapping
  - OPS number can be added to Infoboxes
- Flagged some typical issues with Wikipedia data
  - update Wikipedia via API or publishing results
- Identified some issues where OPS mapping is imperfect

Acknowledgements

- Stefan Senger (external advisor)
- Tell Tuttle (project supervisor)
- Biovia (provided Pipeline Pilot license)

Thank you for your attention.