

Dataset Overview

Uncover better findings with the latest drug data at your fingertips



With more than 2 million scientific journals published each year, keeping up with the latest biomedical information can feel impossible.

And yet, access to high-quality, accurate, and up-to-date data is crucial for biomedical research success.

DrugBank seamlessly merges AI with expert oversight to continuously structure and connect the world's biomedical knowledge. The result is a scalable and accurate knowledge solution that integrates the full breadth and depth of scientific and clinical information.

DrugBank works with trusted partners to integrate and map identifiers to make building relationships between data faster and easier. We also layer in molecular data with clinical data to bring the most comprehensive and important information to the forefront of many use cases.

Updated daily, and reviewed by a team of medical experts, all of DrugBank's data is meticulously referenced and ready to be put to work.

As a result, millions of researchers and industry experts rely on DrugBank's platform every year to make *bigger* discoveries and *stronger* decisions.

With DrugBank you can:



Enhance & connect existing data



Build, train, & validate machine learning models



Explore the world's drug knowledge



Explore new strategies in days – not weeks

Jump to:

- [Clinical Trials](#)
- [Rare Disease - Clinical Trials Add-on](#)
- [Drug Categories](#)
- [Drug Names & Drug Ingredients](#)
- [Drug Metabolism](#)
- [Drug Patents](#)
- [Drug Pharmacology](#)
- [Structured Pharmacology - Drug Pharmacology Add-on](#)
- [Drug Products](#)
- [Drug-Protein Relationships & Drug Targets](#)
- [Investigational & Experimental Drugs](#)
- [Pharmacogenomics](#)
- [Pipeline Status](#)
- [Product Concepts](#)
- [Structured Adverse Effects](#)
- [Structured Allergies](#)
- [Structured Blackbox Warnings](#)
- [Structured Contraindications](#)
- [Structured Drug Indications](#)
- [Structured Drug-Drug Interactions](#)
- [Therapeutic Categories](#)



Clinical Trials

Sourced from the NIH Clinical Trials database, this dataset is used by both drug discovery and drug repurposing companies to better understand their own drugs and for repurposing opportunities.

How it's used

- ✓ Understand drugs in future, present, & previous clinical trials & what stage they're at.
- ✓ In drug repurposing, it is used to help identify abandoned drugs that have failed clinical trials to repurpose for another indication or condition.
- ✓ Compare proprietary drugs to investigational drugs with similar attributes to help identify potentially undesirable traits.

What's included

- ✓ Trial date, trial id, title, phase, date started/ stopped, trial description, eligibility criteria, & intervention groups (which drugs were used in what combinations).
- ✓ Conditions (indication) with associated ICD-10 & MedDRA identifiers.
- ✓ PubMed references for the trial, countries & intervention descriptions.



Rare Disease

CLINICAL TRIAL ADD-ON

This dataset is available as an add-on to our Clinical Trials dataset and provides additional features that can support rare disease drug discovery and repurposing efforts.

How it's used

- ✓ Review proteins mentioned in trials of interest to identify patterns & prioritize leads.
- ✓ Leverage regulatory incentives by identifying drugs & conditions with orphan designations.
- ✓ Review trials studying drugs with similar mechanisms of action to identify repurposing opportunities.
- ✓ Review termination reasons for past trials to optimize future study designs & avoid pitfalls.

What's included

- ✓ Standardized termination reasons.
- ✓ Orphan designation details.
- ✓ Deeply curated Drugs & Conditions.
- ✓ Linked clinical trial biological entity mentions.



Drug Categories

Drug categories allow groups of drugs to be analyzed by numerous attributes and categorized in infinite ways.

How it's used

- ✓ Categorize & organize groups of drugs by chemical structure, mechanism of action, & taxonomy for different grouping approaches.
- ✓ Filter investigational, experimental, & approved drugs.
- ✓ Unlock new findings by looking at specific drugs in unique ways.
- ✓ Helpful for creating specific attributes for machine learning models.

What's included

- ✓ ATC (Anatomical Therapeutic Chemical).
- ✓ MeSH (Medical Subject Headings).
- ✓ Chemical Taxonomy.
- ✓ EPC (Established Pharmacologic Class) for select download formats.

[Back to Dataset Listing](#) →

Drug Names & Ingredients

This dataset provides a comprehensive list of all approved drugs, detailed molecular descriptions, and comprehensive nomenclature and identifiers.

How it's used

- ✓ Query the database by brand name, active ingredient, DrugBank identifiers, & chemical structure.
- ✓ Simplify data normalization from multiple sources with comprehensive lists of synonyms, brand names, code names, & other common identifiers.
- ✓ Perform queries across the entire landscape of approved medications.
- ✓ Identify potential patterns by grouping drugs based on structural similarity, sequence similarity, type of drug or chemical properties.
- ✓ Survey available formulations for each drug, including detailed chemical properties for each formulation.

What's included

- ✓ Chemical names, common names, generic names, & brand names.
- ✓ Chemical structure & comprehensive chemical properties for small molecules.
- ✓ Sequences & genes for proteins.
- ✓ General overview description of each drug.
- ✓ Each drug is categorized as approved, investigational, or experimental.
- ✓ Chemical structure of formulations & salts.
- ✓ External identifiers including: ChEMBL, PubChem, RxCUI, PDB, GenBank, Uniprot, & Zinc.
- ✓ Chemical structure is available in multiple formats including SDF, SMILES, & Inchi.

[Back to Dataset Listing](#) →

Drug Metabolism

This dataset provides a comprehensive and queryable overview of all known metabolic reactions for all approved medications.

How it's used

- ✓ Compare known metabolism information to better predict the metabolism of a new drug.
- ✓ Utilize metabolite IDs to easily see all reactions that lead to that metabolite.
- ✓ Used to understand what enzymes are inhibited, induced, or involved in a metabolite reaction.
- ✓ Identify potential repurposing opportunities resulting from off-target or downstream effects.
- ✓ Helpful for mapping & building relationships between data to develop a more comprehensive understanding of a drug.
- ✓ Perform risk assessment evaluations of drugs to help support tailored treatments.

What's included

- ✓ More than **3,300 metabolism reactions** associated with more than **230 unique enzymes**.
- ✓ The type of reaction, the enzyme, & the name & structure of the drug metabolite.
- ✓ The pathways the drug affects or is involved in.

Drug Patents

Drug patent information is helpful for understanding the competitive landscape and identifying drug discovery and repurposing opportunities.

How it's used

- ✓ Generate business intelligence by building alerts or reports for drugs that are about to come to market or have patent expirations by pairing this dataset with clinical trials & indications data.

What's included

- ✓ More than **9,600 unique patents** with patent ID, grant date, & expiry date, or estimated expiry date for future expiries.

[Back to Dataset Listing](#) →

Drug Pharmacology

The Drug Pharmacology dataset synthesizes knowledge from more than 100,000 scientific references, providing highly synoptic summaries.

How it's used

- ✓ ADME properties are often used to build predictive models for toxicity or predicting biomedical properties of new potential drugs.
- ✓ Summary descriptions provide useful overviews & can save a huge amount of research & review time.

What's included

- ✓ Detailed descriptions of the mechanism of action, absorption, distribution, metabolism, elimination including experimental properties (when available).
- ✓ Pharmacokinetic & pharmacodynamic parameters such as half-life, toxicity, clearance, & LD50.



Structured Pharmacology

 DRUG PHARMACOLOGY ADD-ON

This dataset is available as an add-on to our Drug Pharmacology dataset and provides a solid foundation for machine applications by offering clarity, consistency, and efficiency in data processing, analysis, and integration.

How it's used

- ✓ In PBPK and PK/PD modeling to make predictions about dosing feasibility, and enable target and lead selection and optimization.
- ✓ In conjunction with pathways and metabolism data to identify potential biomarkers to quantify drug response.
- ✓ To compare competing drugs to identify advantages over existing therapies, address unmet needs, and understand commercial feasibility.

What's included

- ✓ Extracted and structured versions of LD 50, Clearance, Half-Life, Volume of Distribution, and Bioavailability.
- ✓ Units of measurement standardized to Unified Code for Units of Measure (UCUM) units.

[Back to Dataset Listing →](#)

Drug Products

This dataset provides all approved and withdrawn drug products for over 10 regions which includes products approved by the FDA, Health Canada, EMA, and more.

How it's used

- ✓ Extremely versatile, this dataset is used by nearly all customers ranging from drug discovery & repurposing, clinical applications, & precision medicine tools.
- ✓ Capture an accurate drug product at varying levels of specificity, covering brand names, broad groups of products, or very focused drug details down to dose form, strength, & route of administration.
- ✓ Labeller information is helpful for understanding who made the product & mapping back to MedDRA & ICD-10 identifiers.

What's included

- ✓ Other regions include Italy, Austria, Turkey, Colombia, Malaysia, Indonesia, Singapore, & Thailand.
- ✓ Each listing includes the brand or generic name, dosage, dose form, route of administration, codes, labeller, approval status, & marketing start & end date.
- ✓ The drug ingredients are linked with the drug dataset, & all products are linked with approved indications which include links to MedDRA & ICD-10.

Drug-Protein Relationships & Drug Targets

Understanding bonds for predictive models and early stage research can support better predictions such as target identification and validation.

How it's used

- ✓ Used for building ML models for target prediction as well as early stage research to uncover drug discovery & repurposing opportunities.
- ✓ Many identifiers that DrugBank data maps to, such as Uniprot IDs, Genbank IDs, Protein Data Bank IDs, NCBI sequence identifiers, are helpful for linking data together.

What's included

- ✓ More than **28,000 drug-protein interactions** covering targets, enzymes, carriers, & transporters.
- ✓ Annotation about the pharmacological action, & the type of interaction (antagonist, agonist, substrate, inhibitor, or inducer).
- ✓ Drug Targets include the protein & other helpful data like gene identifiers & sequences.
- ✓ Drug Targets map to many external identifiers making it helpful for the majority of use cases.

[Back to Dataset Listing →](#)

Investigational & Experimental Drugs

Helpful for early stages of drug discovery or repurposing, precision medicine, and pharmacovigilance.

How it's used

- ✓ The drugs' investigational & experimental stages will have their associated indications & map to other helpful data such as drug targets.
- ✓ Enhance & expand existing drug lists with available drug ingredients & drug names.
- ✓ Enhance data pools by incorporating data outside of approved drugs.
- ✓ Precision Medicine customers use this data for prescribing purposes, matching patients with appropriate clinical trials, & reviewing what drugs are being investigated for new indications.

What's included

- ✓ More than **12,000 experimental & investigational drug summaries**, which are compounds that pass a number of 'drug likeness' filters & are known to bind with their associated bonds.
- ✓ Information about each drug includes: names, synonyms, DrugBank Identifier, a description, the chemical structure, & the chemical structure of formulations & salts.
- ✓ Access additional summaries for more than **4,200 clinical drugs & 10,000 pre-clinical drugs**.

Pharmacogenomics

This data is particularly helpful for genetic testing, annotated reports, and for conducting drug discovery or repurposing.

How it's used

- ✓ For genetic testing, using SNP data is more efficient & less costly than evaluating full DNA sequences, especially in clinical projects.
- ✓ Annotating this data in drug discovery work is helpful for early stage research.

What's included

- ✓ SNP Mediated Adverse Drug Reactions.
- ✓ SNP Mediated Pharmacological Effects including a description of the effect, affected drugs, references, SNP IDs, allele name, gene identifier & affected genotype.

[Back to Dataset Listing →](#)

Pipeline Status

This dataset is helpful for understanding upcoming drug approvals as well as competitive intelligence for those evaluating new drug discovery and repurposing opportunities.

How it's used

- ✓ This data will often get paired with clinical trial data to help evaluate & build analysis around upcoming drug approvals.
- ✓ Some clinical platforms set up alerts for upcoming drug approvals.
- ✓ Used to evaluate the competitive landscape & help identify areas of opportunity for drug discovery & repurposing.

What's included

- ✓ Overview of clinical trial status & approval status for drugs, including max phase, approval types (OTC, prescription), approval dates, & generic availability.
- ✓ FDA, Health Canada, & EMA status.

Product Concepts

Proprietary to DrugBank, these IDs capture diverse attributes of drug products and flexibility to build vast levels of relationships.

How it's used

- ✓ Capture & store a large breadth of drug information at different levels & bundle related NDCs together.
 - *This can be done in various levels of precision & is a more effective & accurate way to capture & store drugs than by drug names or NDC codes, which can be unreliable.*
- ✓ Useful for storing metadata associated with different levels of drugs.
- ✓ Used to create a vocabulary of drugs, products, & brands that establishes a hierarchical structure for easy navigation & comparison of products (even across regulatory jurisdictions).

What's included

- ✓ More than **763,000 product concepts**, each one describing a distinct group of features for one or more products.
- ✓ When available, product concepts are mapped to RxNorm concepts.

[Back to Dataset Listing](#) →



Structured Adverse Effects

This data helps uncover new insights for drug discovery and repurposing. Some effects can even turn into new indications.

How it's used

- ✓ Can be used to build algorithms & machine learning models.
- ✓ Useful for assessing the level of risk of a patient taking a drug & understanding the prevalence of the possible adverse effect.
- ✓ Used to predict similar adverse effects for a new drug.
- ✓ Has the potential to uncover new indications.

What's included

- ✓ Includes more than **140,000 adverse effects** covering more than **1,900 drugs** & connecting to more than **11,900 unique symptoms** & effects.
- ✓ Includes effects, administration & patient attributes, prevalence, & associated ICD-10 & MedDRA identifiers.



Structured Allergies

Used largely for risk assessment in early-stage discovery or repurposing research.

How it's used

- ✓ To help understand the level of risk for a new potential drug.
- ✓ Allergy presentations & onset times are being used widely in precision medicine platforms.

What's included

- ✓ Allergy presentation information, severity levels for allergic reactions, & management information.
- ✓ Also includes cross-sensitivities information with overview & description, potential sensitive drugs, & incidence rates.

[Back to Dataset Listing](#) →

Structured Blackbox Warnings

This data is helpful for knowing when a drug should not be prescribed, as well as for discovery work and pharmacovigilance for risk assessment.

How it's used

- ✓ Can be used to build algorithms & machine learning models.
- ✓ Useful for assessing & predicting the level of risk & benefit for producing similar drugs.
- ✓ Helpful for patient education on drug safety. (*i.e.*, *knowing what drugs explicitly should not be taken while pregnant*).

What's included

- ✓ Black Box warnings from manufacturer drug labels.
- ✓ Includes serious adverse effects, interactions, guidance, & warnings.
- ✓ Patient characteristics & adverse effects are associated with ICD-10 & MedDRA identifiers.

Structured Contraindications

This data is helpful for risk assessment and enables healthcare providers to make better informed decisions when prescribing.

How it's used

- ✓ Useful for predictive models in drug discovery & repurposing.
- ✓ Used for algorithms in tailored treatment platforms for decision support, further enabling healthcare providers to find the right drug for their patients.

What's included

- ✓ More than **8,700** contraindications.
- ✓ Includes administration (including interacting drugs & drug categories) & patient attributes with associated ICD-10 & MedDRA identifiers.

[Back to Dataset Listing](#) →



Structured Drug Indications

Most use cases benefit from indication data as it is critical to have an understanding of what a drug's on-label use and off-label uses are.

How it's used

- ✓ Helpful for knowing what drugs can be used both on label & off label.
- ✓ See indications for clinical trials.
- ✓ Helpful for discovery work to aid in understanding the competitive landscape & areas of opportunity.
- ✓ Make predictive models that predict new potential indications, & group drugs by specific categories & indications.
- ✓ Helpful for many tailored treatment applications that aims to help healthcare providers prescribe the right drug to the right patient.
- ✓ Helpful for patient education on drug safety.

What's included

- ✓ More than **10,000 drug indications** covering every FDA & Health Canada approved indication, extracted from FDA drug labels, & scientific publications.
- ✓ Includes a text description, severity, type of indication for both approved drugs & investigational drugs.
- ✓ Associated ICD-10, SNOMED, & MedDRA identifiers for approved drugs.
- ✓ On label & off label indications.

[Back to Dataset Listing →](#)

Structured Drug-Drug Interactions

This dataset is the most robust of its kind in the world, and is beneficial for a wide range of uses including risk assessments for drug development.

How it's used

- ✓ This structured dataset can be used to build algorithms & machine learning models.
- ✓ Make predictive models to predict possible interactions for new drugs to ensure the benefit outweighs the risk of the interactions.
- ✓ Parts of this data can be helpful for patient education by giving patients the knowledge they need to take their drugs safely.

What's included

- ✓ **Over 1.37 million drug-drug interactions** covering all FDA & Health Canada approved drugs.
- ✓ Includes severity rating & source details (whether it came from a drug label, evidence level 1, or scientific publications evaluated by DrugBank's medical experts).
- ✓ Includes drug-food interactions.
- ✓ Detailed & simple descriptions of why the interaction happens.
- ✓ General management advice for healthcare providers.

Therapeutic Categories

Helpful for parsing specific or broad groups of drugs to solve various problems in precision medicine, drug discovery, and drug repurposing.

How it's used

- ✓ Used for many types of grouping techniques in early-stage discovery research.
- ✓ Provides better end-user experience for searching for specific therapeutic categories of drugs in precision medicine platforms.
- ✓ Helpful for filtering & grouping drugs.

What's included

- ✓ Based on the FDA's "Established Pharmacologic Class" (EPC) which is a pharmacologic class associated with an approved indication of an active moiety that the FDA has determined to be scientifically valid & clinically meaningful.
- ✓ We have extended the EPC categories to cover other regions as well.

Questions? Get in touch!

info@drugbank.com | drugbank.com/datasets