BCADD 2026 / WAY2DRUG KEYNOTE

Dataset Constraints: A Case of Pharmacologically Stratified Multi-Target QSAR Benchmark for Opioid Receptor Functional Activity

A data contract, tested in opioid receptor modeling

Khaled M. Elokely

Division of Pharmaceutical Sciences · University of Wyoming

What does an R^2 of 0.825 validate?

0.825

Locked external R^2

DOR functional IC50 label
330 external compounds
RMSE 0.664

Which biological quantity did the model learn?

The anatomy of a QSAR claim

Molecular identity

What counts as the same chemical?

Measured quantity

Which experimental outcome is the target?

Prediction scenario

What new chemistry or future setting is tested?

Decision and metric

Which decision does the reported performance support?

Chemical identity

1	**As deposited**	Salts, mixtures, charge and stereochemistry
2	**Standardized**	Declared transformations and tool versions
3	**Parent representation**	A stated identity for duplicates and split groups

The identity rule changes duplicates, neighbors and the evaluation.

Endpoint meaning

Binding affinity

Competition or occupancy
How strongly does it bind?

Functional potency

Position of a response curve
At what concentration?

Efficacy

Response relative to a reference
How much effect can it produce?

An IC50 label alone does not specify the biological event.

The prediction scenario

In-distribution

Interpolation among represented chemistry

Scaffold or cluster

Transfer across separated cores or groups

Temporal

Prediction into later chemistry

A split tests a scenario. Similarity diagnostics test what still leaks.

Average metrics and activity cliffs

Dataset-wide score

Average predictive fit

Lead optimization decision

Local activity changes

Activity cliffs can remain difficult even when the global metric looks strong.

The seven-part dataset contract

1 **Executable manifest**
Exact rules, versions and reconstruction

2 **Chemical identity mappings**
Submitted, standardized and parent identities

3 **Endpoint provenance**
Assays, units, aggregation and censoring

4 **Split suite**
Assignments labeled by intended use

5 **Leakage diagnostics**
Similarity and dependency pathways

6 **Decision alignment**
Metrics matched to deployment

7 **Manual curation log**
Structured records of expert decisions

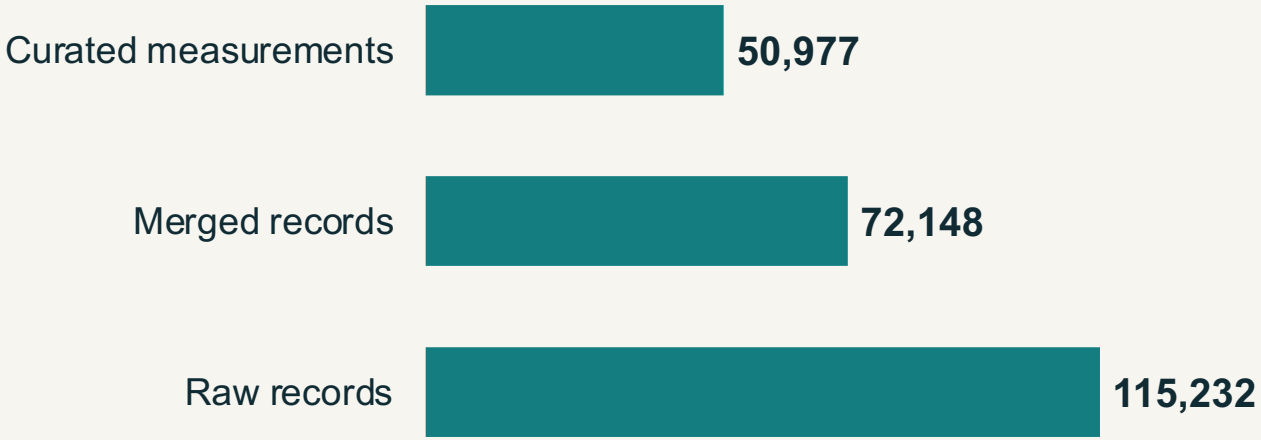
One receptor family, different biological events

Four opioid receptors: MOR, DOR, KOR and NOP

cAMP	Regulation of adenylyl cyclase
GTPγS	Proximal G-protein activation
β-arrestin	Receptor–arrestin engagement
Radioligand displacement	Binding competition

Mechanism and readout operationalize endpoint provenance.

Four sources reconciled into a benchmark



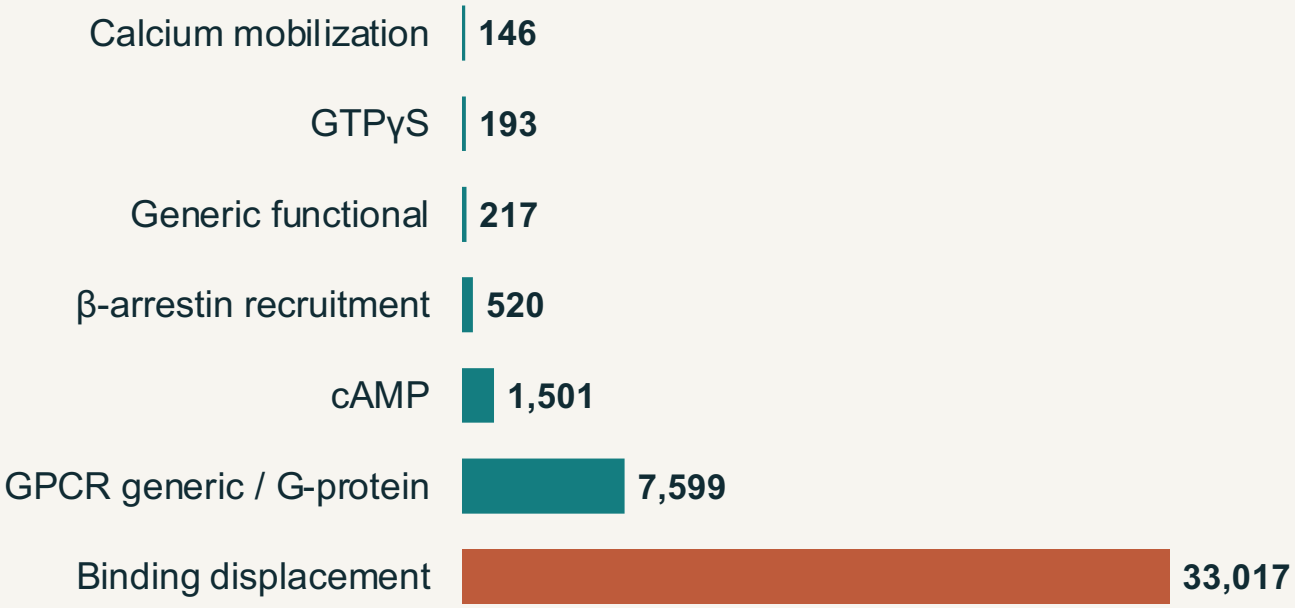
19,585

curated compounds

ChEMBL · BindingDB · PubChem · IUPHAR/BPS

Gold + silver: 97% of curated measurements. Exclusions were logged.

The pre-routing readout census



3.24:1

displacement to all genuine functional readouts combined

43,193 assigned readouts in 43,197 development measurements; counts precede routing.

A task with a biological interpretation

Receptor	Scale	Class	Readout
DOR	**pEC50**	**Agonist**	**GPCR generic**

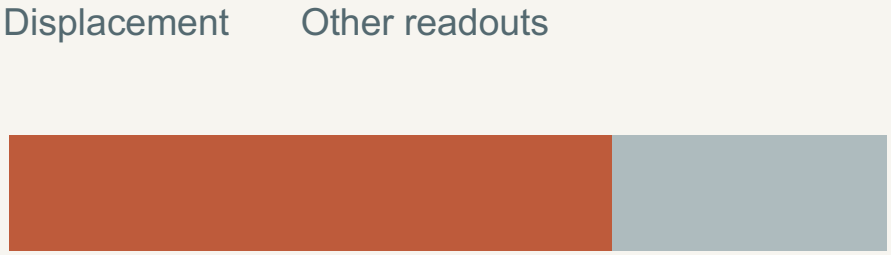
DOR agonist potency in a GPCR-generic assay

This is one endpoint definition within the broader dataset contract.

A high score can carry the wrong claim

0.825

external R^2



1,539 / 2,240 pooled records

69% measured displacement

The pool does not define a pure functional-inhibition endpoint.

EC50 scores remain similar

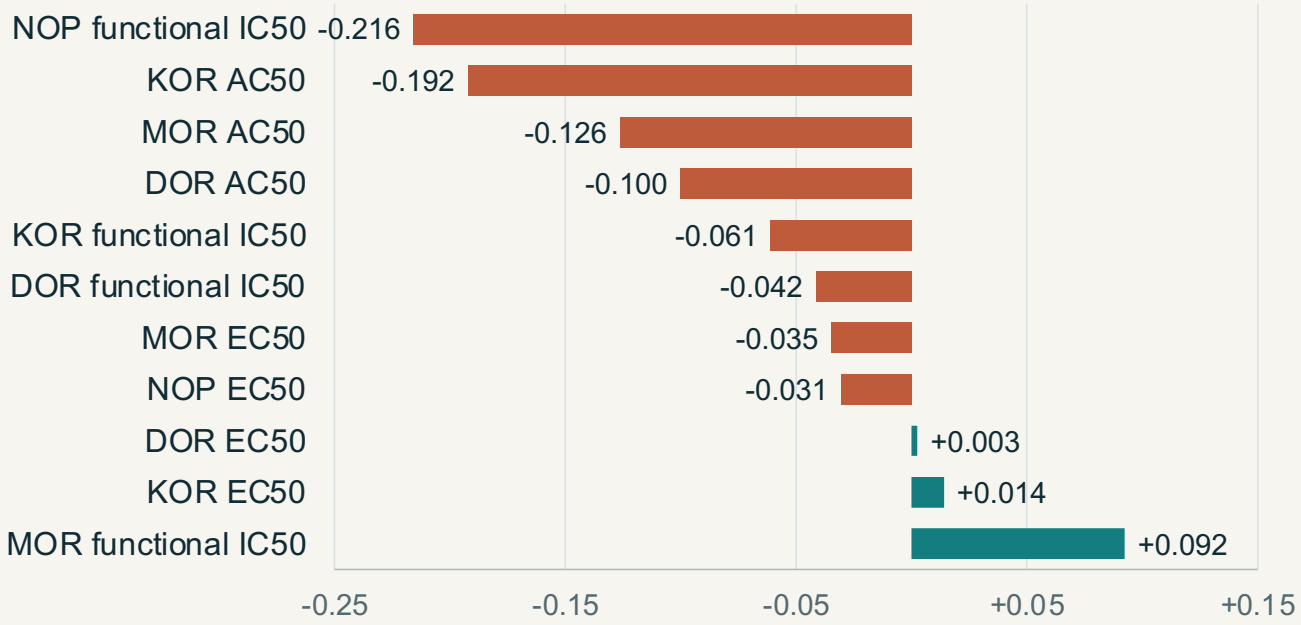


≤0.035

absolute R^2 difference on each receptor

Comparisons are not sample-matched. KOR's selected stratum has only 15 external compounds.

Pooled R^2 is higher in 8 of 11 comparisons



Positive Δ favors stratified Negative Δ favors pooled

Endpoint meaning is the contribution.

Different sample sizes.
A maximum over strata.
No causal estimate of why R^2 changes.

Nael et al. | Digital Discovery 2026 | 10.1039/D6DD00392C / Released comparison; Δ = best stratified − pooled

A contract the audience can audit

1 **Define the prediction target**

Identity rules and endpoint provenance

2 **Test the intended use**

Splits, similarity diagnostics and decision metrics

3 **Release the contract artifacts**

Executable rules, assignments and curation decisions

The dataset defines what the score can support

Every performance number needs an explicit endpoint and a prediction scenario.

The contract makes that claim auditable.

Acknowledgments and discussion

Manal A. Nael, Laxman M. Alakonda, Abhik Ghosh, Sara Jane Ward, Lee-Yuan Liu Chen, Anjali M. Rajadhyaksha, Magid Abou-Gharbia and Khaled M. Elokely

ChEMBL, BindingDB, PubChem and IUPHAR/BPS Guide to Pharmacology

NIH/NIGMS Institutional Development Award 2P20GM103432

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

JCIM 2026: 10.1021/acs.jcim.6c00514
Digital Discovery 2026: 10.1039/D6DD00392C

github.com/kelokely/CURATO-OPIOID
kelokely@uwyo.edu