

D360 customisation at AZ

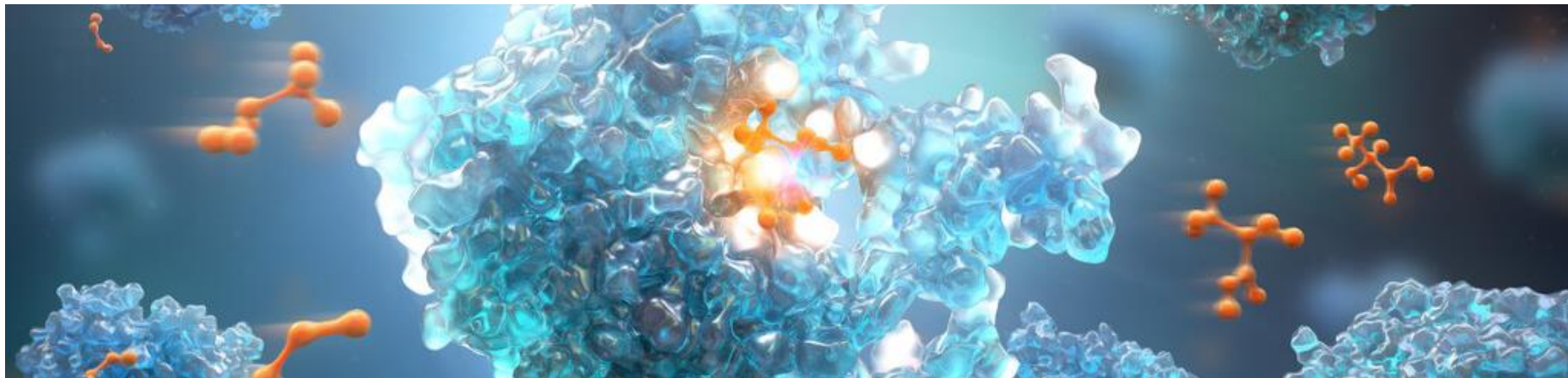
Gostar patent analysis

Curve viewer

Wolfgang Klute, Chemoinformatics Data Scientist, Science IT

Certara D360 UGM Boston 2019

09 Oct 2019

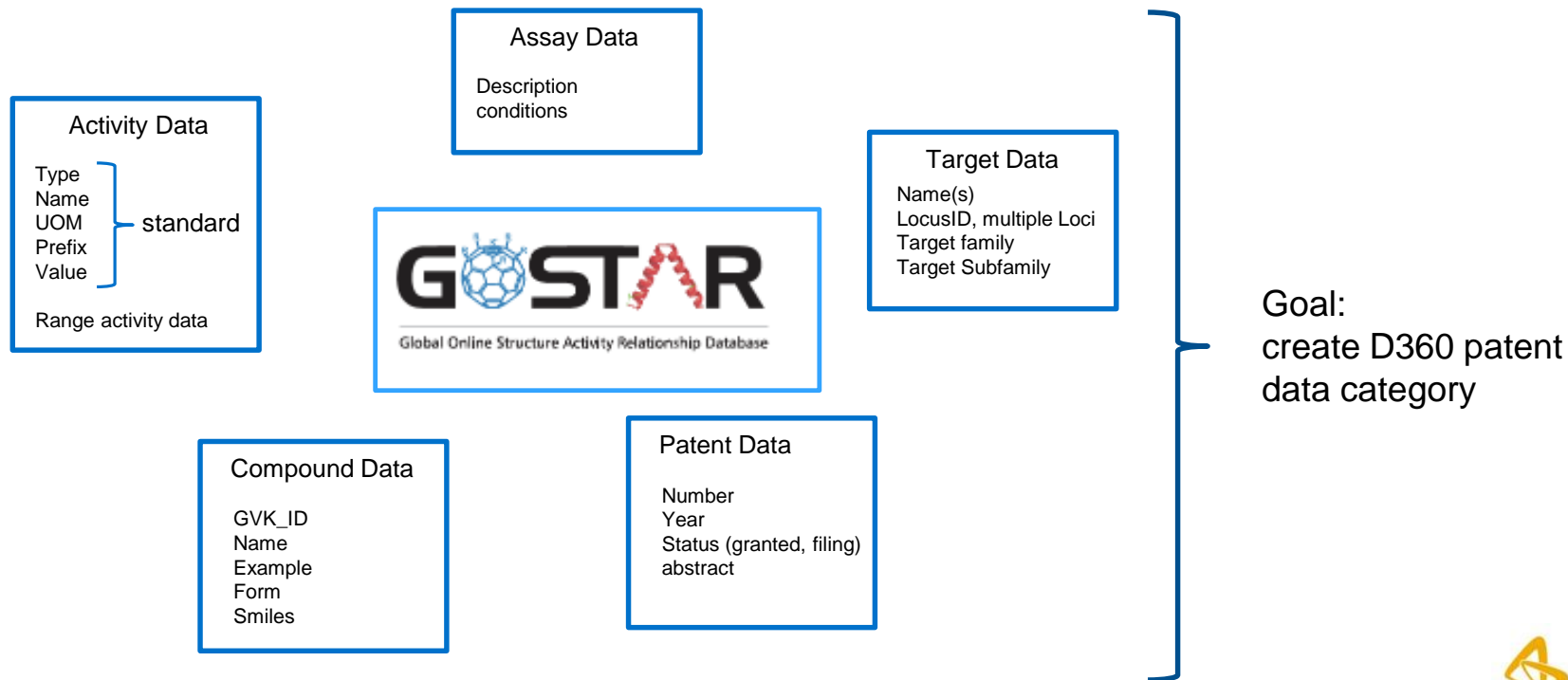


Gostar patent analysis

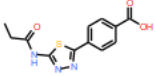
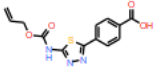
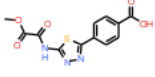
Gostar patent data



- Database distributed by Excelra with SAR data from patents and MedChem lit
- 80 K patents, > 600 K assays, > 8 M compounds, 16 M activities



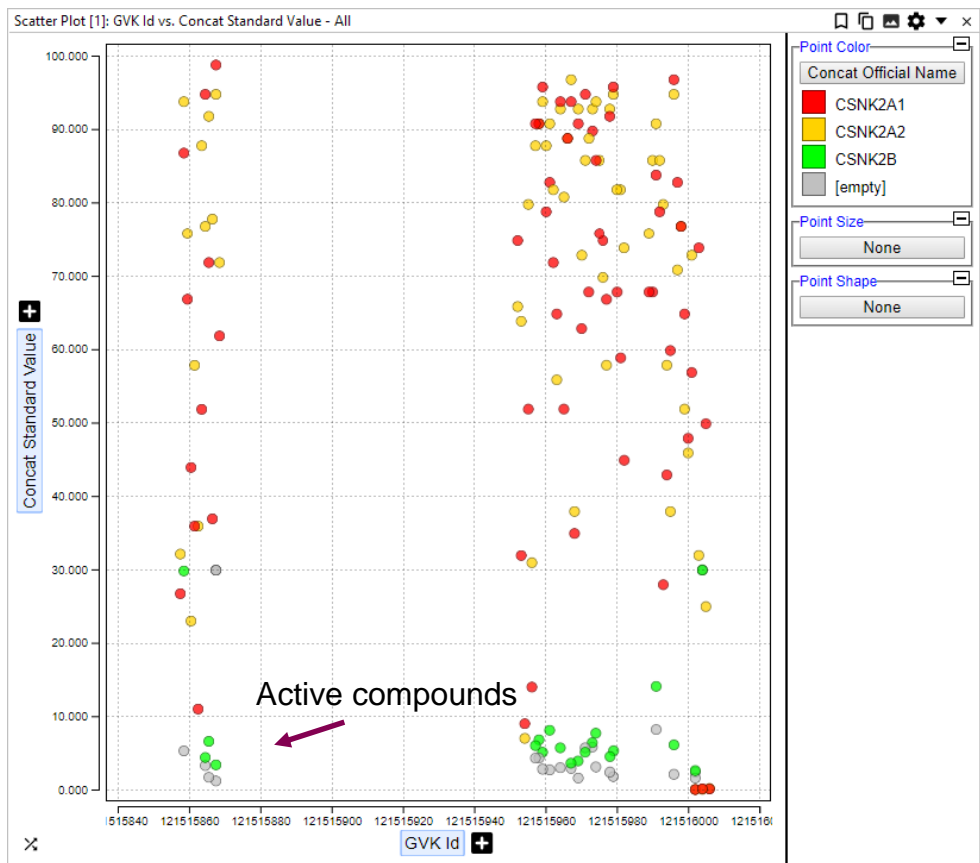
Conventional and easy: unpivoted

GVK Id	GVK Structure	Concat Normalized Journ...	Concat Assay Id	Concat Official Name	Concat Assay Type	Concat Std Activity Type	Concat Standard Uom	Concat Std Act Prefix	Concat Standard Value
121515960		WO2012121168	10998749	CSNK2A1	B	INHIBITION	%	=	79.000
			11000058	CSNK2A2	B	INHIBITION	%	=	88.000
			11000059	CSNK2B	Q	Qualitative	none	=	0.000
activity									
121515961		WO2012121168	10998749	target	B	INHIBITION	%	=	83.000
			10999825	CSNK2A1	B	INHIBITION	%	=	91.000
			10999826	CSNK2A2	B	IC50	uM	=	8.100
			11000058	CSNK2B	B	IC50	uM	=	2.700
			11000059	CSNK2B	Q	Qualitative	none	=	0.000
assay									
121515963		WO2012121168	10998749	CSNK2A1	B	INHIBITION	%	=	65.000
			11000058	CSNK2A2	B	INHIBITION	%	=	56.000
			11000059	CSNK2B	Q	Qualitative	none	=	0.000

Easy to create
Had these views for many years



Analysis of un-pivoted data in D360



D360 offers some utility to analyse non-pivoted data

can see which compounds are active

can't correlate activity

- assay 1 vs assay2
- assay 1 vs clogP
- etc

Could we create pivoted view?

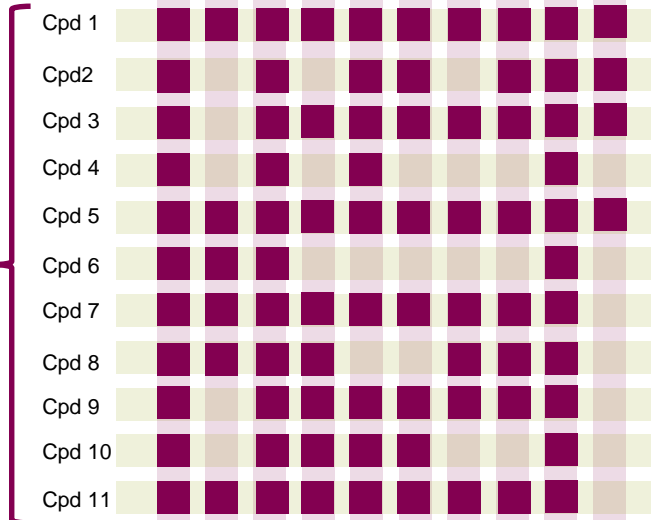


Corporate data vs public data

ibis Curated test shared across projects
sharing discovery data



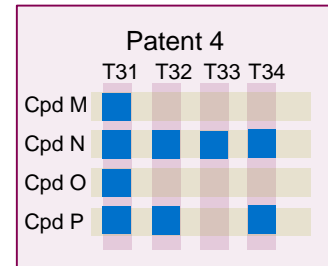
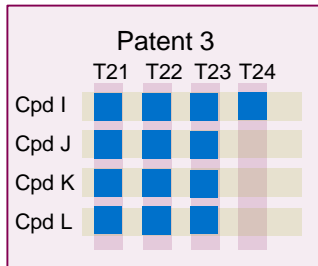
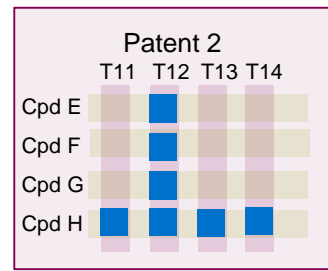
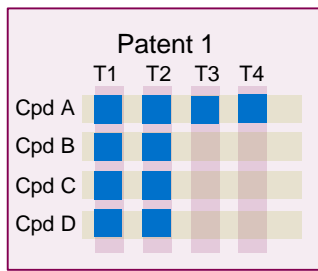
T1 T2 T3 T4 T5 T6 T7 T8 T9 T10



registered compounds shared across projects

GOSTAR
Global Online Structure Activity Relationship Database

Document centric
Fragmented data



M's compounds, 10 000s tests, 10 M's activities (number results)
 Activities / compound = 15 : 1
 Activities / assay = 1800 : 1

- 80 K patents, > 600 K assays, > 8 M compounds, 16 M activities
- Activities / compound = 2 : 1
- Activities / assay = 25 : 1



Gostar patents: user friendly query interface

Query: GVK Patent Structure 16-sep-2019 21:14 [1]
File Edit Info Quick Search Window Help

Data Category: GVK Patent Structure
Filter:

- IRON ION BINDING
- ISOMERASE
- KINASE
 - TYR PROTEIN KINASE FAMILY
 - ABI FAMILY
 - ADENYLATE KINASE FAMILY
 - AGC SER/THR PROTEIN KINASE FAMILY
 - ALPHA-TYPE PROTEIN KINASE FAMILY
 - ATP:GUANIDO PHOSPHOTRANSFERASE FAMILY
 - CAMK SER/THR PROTEIN KINASE FAMILY
 - CARBOHYDRATE KINASE PFKB FAMILY
 - CHOLINE/ETHANOLAMINE KINASE FAMILY
 - CK1 SER/THR PROTEIN KINASE FAMILY
 - CASEIN KINASE
 - US20090093465
 - WO2005070180
 - WO2012121168
 - CASEIN KINASE 1
 - CASEIN KINASE 1 ALPHA 1
 - CASEIN KINASE 1 DELTA
 - CASEIN KINASE 1 DELTA|CASEIN KINASE 1 EPSILON
 - CASEIN KINASE 1 EPSILON
 - CASEIN KINASE 1 EPSILON|CASEIN KINASE 1 DELTA
 - TAU TUBULIN KINASE
 - CMGC SER/THR PROTEIN KINASE FAMILY
 - CMGC SER/THR PROTEIN KINASE FAMILY - CYCLIN FAMILY
 - CMGC SER/THR PROTEIN KINASE FAMILY|AGC SER/THR PROTEIN KINASE FAMILY
 - CMGC SER/THR PROTEIN KINASE FAMILY|PI3/P14-KINASE FAMILY
 - CMGC SER/THR PROTEIN KINASE FAMILY|SER/THR PROTEIN KINASE FAMILY
 - CMGC SER/THR PROTEIN KINASE FAMILY|TYR PROTEIN KINASE FAMILY

Document centric query tree, target classification

Query Bands

- GVK Id
- GVK Structure
- Chem Connect ID(GVK AZID Map) (Concat)
- AZ Number(GVK AZID Map) (Concat)
- Journal Patent Name(GVK Patent) (Concat)
- Normalized Journal Patent Name(GVK Patent) (Concat)
- EntrezGene ID(GVK Target) (Concat)
- Common Name(GVK Target) (Concat)
- Standard Name(GVK Target) (Concat)
- Official Name(GVK Target) (Concat)
- GVK Compound Name

Normalized Journal Patent Name (Concat)

Select Specific Filter:

- AT 205169
- AT 366032
- AT 598897
- AU2007200354
- BE 610030
- BE 618968
- BE 628441
- BE 632770
- BE 644682
- BE 659636

Pre-selected query bands

Pick lists for patent names and target information

WO2012121168 (GVK Results)

Results & Conditions Analysis Information

Results

- IC50 (uM) (GMean)
- Qualitative (none)
- INHIBITION (%)

Conditions

Conditions affecting Multiple Results

- Assay Type
- Assay Id
- Assay Method Name
- Temperature
- Protein

Conditions affecting IC50

- Cells Celline Organ
- Incubation Time

Conditions affecting INHIBITION

- Buffer
- Dose Concentration
- In Presence Of
- In Presence Of Conc
- Substrate

Result types and conditions

Pivoting options, column header

Pivoted views

Patent information Patent compounds

Ref Id: 01 jan 2012

Reference:

Title: Flag:

Summary:

Abstract:

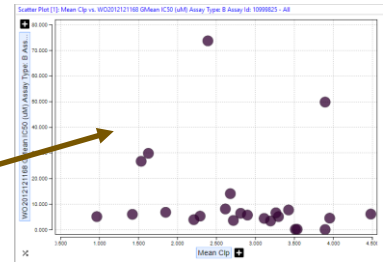
Patent URL:

Pubmed Id:

Patent No: Year:

Patent Specification: Volume:

FOQ on patent number



One column per assay, useful for scatter plots

GVK Id	Mean Chemistry Connect ID	GVK Structure	Concat Normalized Journ...	WO2012121168 GMean IC50 (uM) Assay Type: B Assay Id: 10999825	WO2012121168 GMean IC50 (uM) Assay Type: B Assay Id: 10999826	WO2012121168 GMean IC50 (uM) Assay Type: FC Assay Id: 11000927	WO2012121168 GMean IC50 (uM) Assay Type: FC Assay Id: 11000928	WO2012121168 GMean IC50 (uM) Assay Type: FC Assay Id: 11000940	WO2012121168 Mean INHIBITION (%) Protein: Casein kinase 2 alpha Assay Type: B Assay Id: 11000058	WO2012121168 Mean INHIBITION (%) Protein: Casein kinase 2, alpha prime Assay Type: B Assay Id: 11000059
<input type="checkbox"/> 121516002	285286146		WO2012121168	0.032	0.046	2.600	1.600	2.400		
<input type="checkbox"/> 121515867	285286145		Synonym for Chemistry Connect ID					>30.000	99.000	95.000
<input type="checkbox"/> 121516004	32778689									
<input type="checkbox"/> 121515857	221306189									

Database Name

ID	Database Name	Primary Compound ID	Name
1	ACD	MFCD28127268	MFCD28127268
2	chembl	CHEMBL2087023	CHEMBL2087023
3	ChemSpider	28523182	28523182
4	GOSTAR	GVK120991876	GVK120991876
5	GOSTAR	GVK121516002	GVK121516002
6	GOSTAR	GVK123093966	GVK123093966
7	GOSTAR	GVK123193077	GVK123193077

FOQ on CC_ID

Most potent compound appears in 4 patents + other refs



learnings

- D360 useful tool to analyse complex deep data
- Pivoted patent analysis useful for one patent at a time
- Use follow-on queries to provide customisable views of meta data
- Key is user-friendly implementation (many supporting custom tables...)
- ChemistryConnect integration essential to pick up relationships between patents
- Gaps:
 - ❖ Range activity data
 - ❖ Column aggregation with BAO
 - ❖ useful visualisations for un-pivoted data (heat map)



D360 curve viewer

AZ curve viewer in D360

Why do we need a curve viewer in D360?

- Need reliable decision making on complex assays where numerical end points are not sufficient
- Need to be able to visually spot outliers
- Need to identify effects not captured by numerical end points (e.g. cytotoxic events at high concentration in phenotypical assays)

AZ built interactive curve viewer extension to D360 (D3, HTML)
Data consumed from a datamart with normalised curve parameters (I/EC50, S0, Sinf, slope) and point data

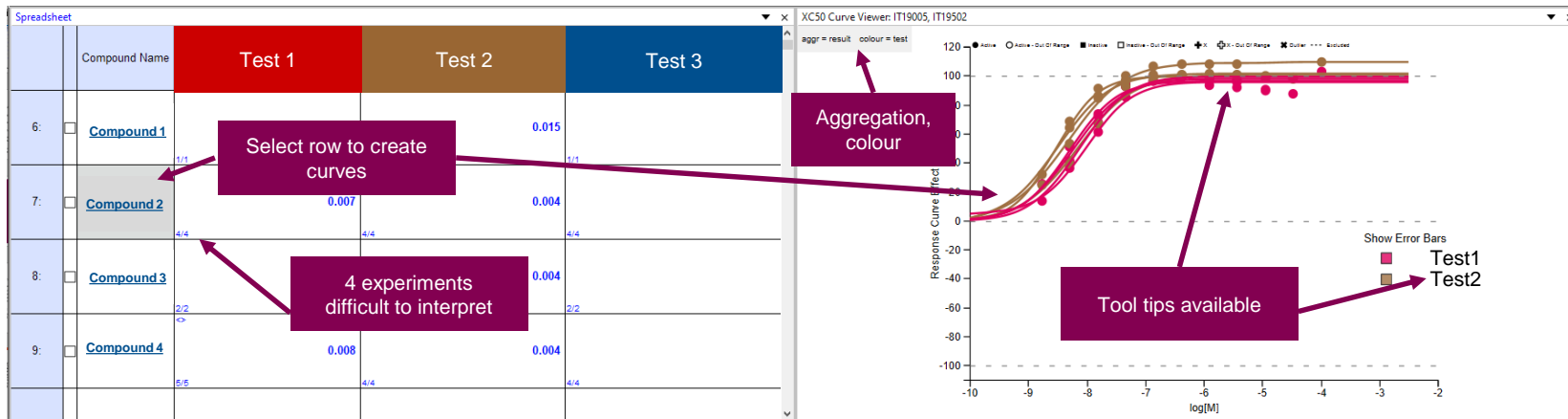
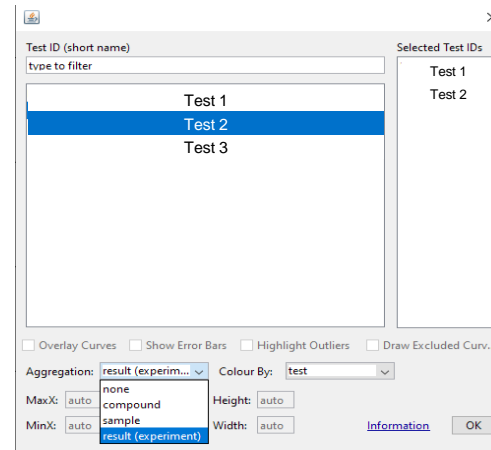


functionality

- Curves shown in a separate interactive window
- Curves from a single compound, multiple tests
- aggregation levels:
none, experiment, sample, compound
- Colour code by experiment or test
- Shapes according to AZ activity flag
+ out of range information (> or <)

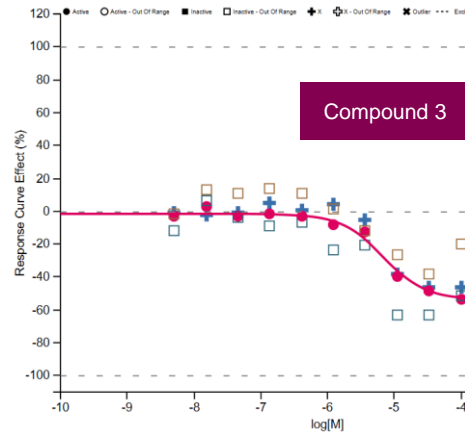
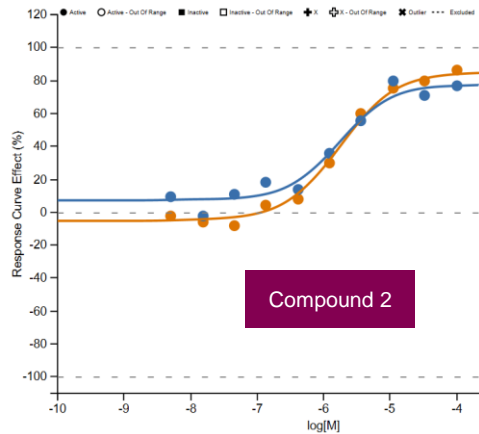
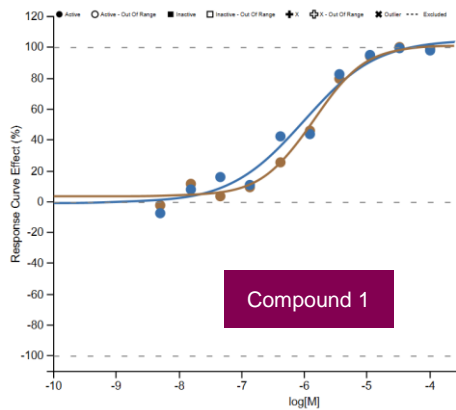
Active
 Active - Out Of Range
 Inactive
 Inactive - Out Of Range
 + X
 + X - Out Of Range

- Handling of point outliers, error bars, tool tips



Example 1: detecting inverse agonists

3 compounds in an agonist assay, project table contained many more columns



Exp1: excluded
Exp2: inverse agonist
Exp3: inactive
Exp4: inactive

Easy to miss outliers in project tables where few end points are reported



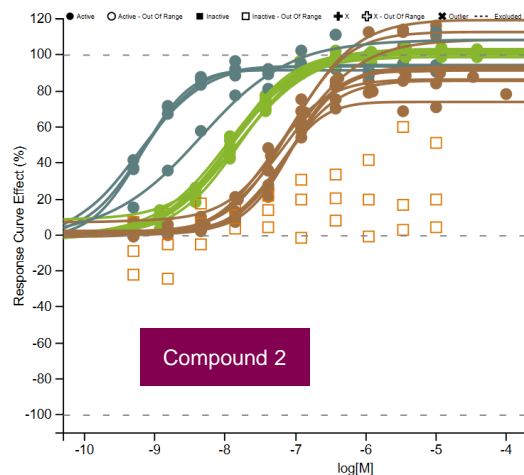
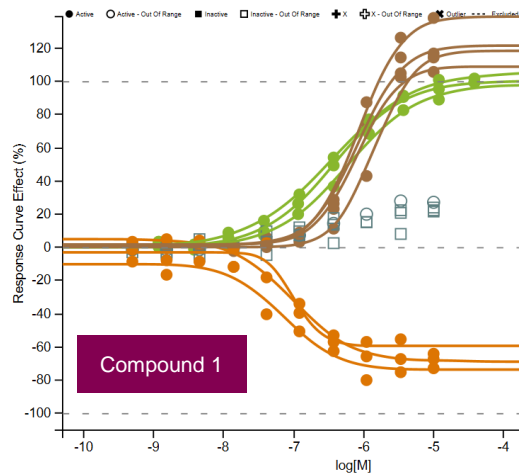
Curve viewing required to make correct decisions

	Compound Name	GMean EC50 (µM)	Concat Distinct Act Flag
1:	Compound 1	1.155	Active
2:	Compound 2	1.740	Active
3:	Compound 3	>40.817	Irregular Active Not Active



Example 2: complex decision making in NHR

Challenge: assay variability, complex decision criteria involving 4 assays
 -> project can't rely on aggregated number results



Individual experiments shown
 Colour coded by test ID

	Compound Name	Assay 1	Assay 2	Assay 3	Assay 4
1:	Compound 1	0.418 3/3	0.091 3/3	>10.000 4/4	0.937 4/4
2:	Compound 2	0.013 8/8	>10.000 3/3	0.001 4/4	0.074 9/9

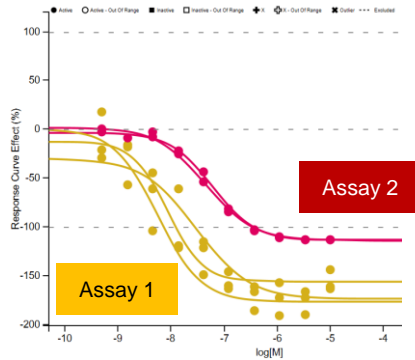
Assay 4: needs full antagonism
 -> compound 1 better
 assay 2: no negative or positive response
 -> can't conclude this from data



Curve aggregation

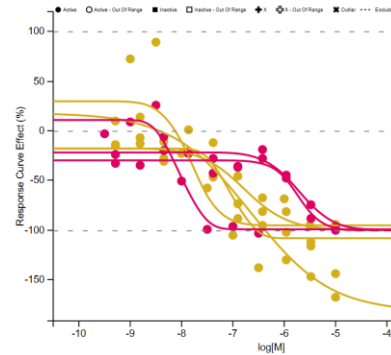
Useful when comparing several tests. However dangerous....

Compound 1

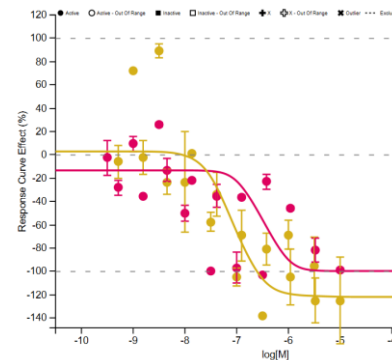
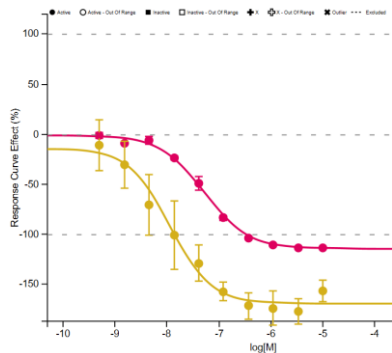


Experiment level
each curve is shown

Compound 2



Compound level
each curve is shown



Even at higher aggregation
sufficient detail needs to be
shown to avoid mis-interpretation



Conclusion

- In projects using complex assays curve viewer is used as a critical tool for decision making and as a complement to numerical end points.
- Opportunities for further improvements
- At AZ dependency on curve datamart, however offers opportunity to use same data for generic data science purposes



acknowledgements

Gostar patent viewer

Dennis Powell (Certara)
Alex Hird
Magnus Polla
Scott Throner
Lars Brive
Rahul Singh
Kalyan Ponamalli

D360 curve viewer

Plamen Petrov
Frank Janssen
David Murray
Kenneth Granberg
Ashvin Dookun
Ana Teixeira

The D360 leadership group:

Andrew Poirrette, Nick Tomkinson, Bill McCoull, Emma Evertsson, Alex Hird

Confidentiality Notice

This file is private and may contain confidential and proprietary information. If you have received this file in error, please notify us and remove it from your system and note that you must not copy, distribute or take any action in reliance on it. Any unauthorized use or disclosure of the contents of this file is not permitted and may be unlawful. AstraZeneca PLC, 1 Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge, CB2 0AA, UK, T: +44(0)203 749 5000, www.astrazeneca.com

