Predictive Toxicology – Day Two

In Silico Models

Wednesday, January 12, 2005

8:40 **Opening remarks** –

Dr. Steven Muskal

8:45 **A Model for Herg K Channel Binding** –

Dr. Kalyanasundaram Subramanian

9:15 **Predictive GenoToxicity, there is Light at the End of the Tunnel** –

Dr. Joseph Votano

9:45 **Predictive, Structure Based, In Silico P450 Metabolism: Principle, Problems, and Progress** —

Dr. Dan L. Harris

10:15 **Coffee Break**

10:45 **Critical Pathway Identification: Integrated Transcriptional Profiling and Bioinformatics to Predict Toxicity** -

Dr. Gordon Kennovin

11:15 **Leveraging Reference Similarity as a Means to Rapidly Gauge Compound Toxicity -**

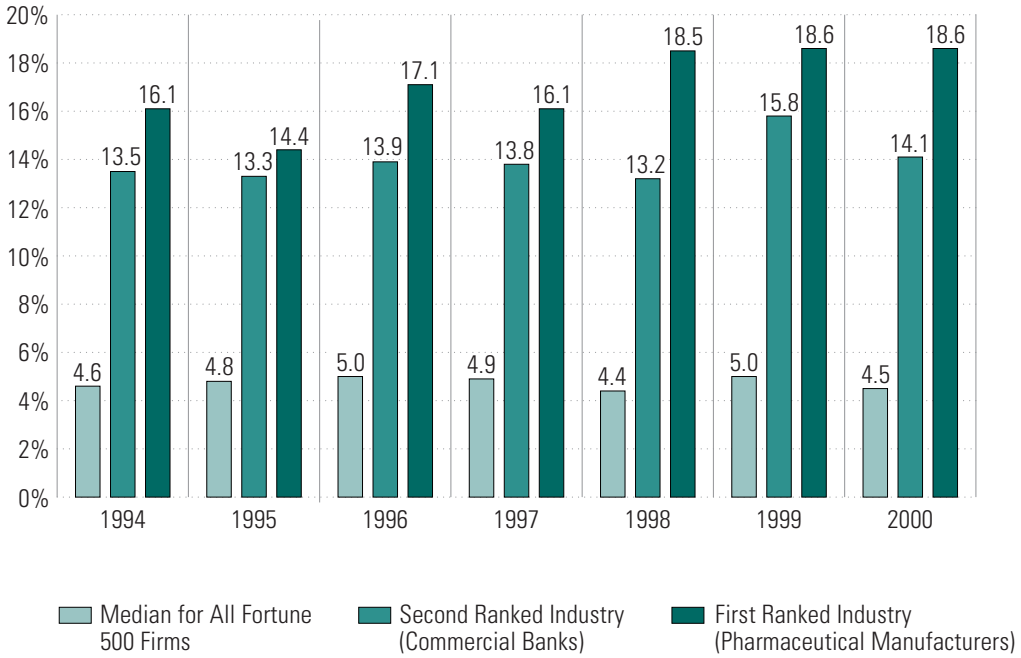
Dr. Steven Muskal

11:45 **Lunch on your own**

Profitability Among Pharmaceutical Manufacturers Compared to Other Industries, 1994–2000

exhibit

32



note

Percent shown is the median percent net profit after taxes as a percent of firm revenues for all firms in the industry. The second ranked industry each year was commercial banks.

source

Fortune 500 Industry Rankings, *Fortune*, April issues, various years.

Projected Prescription Drug Spending By and For the Medicare Population, 2001–2011

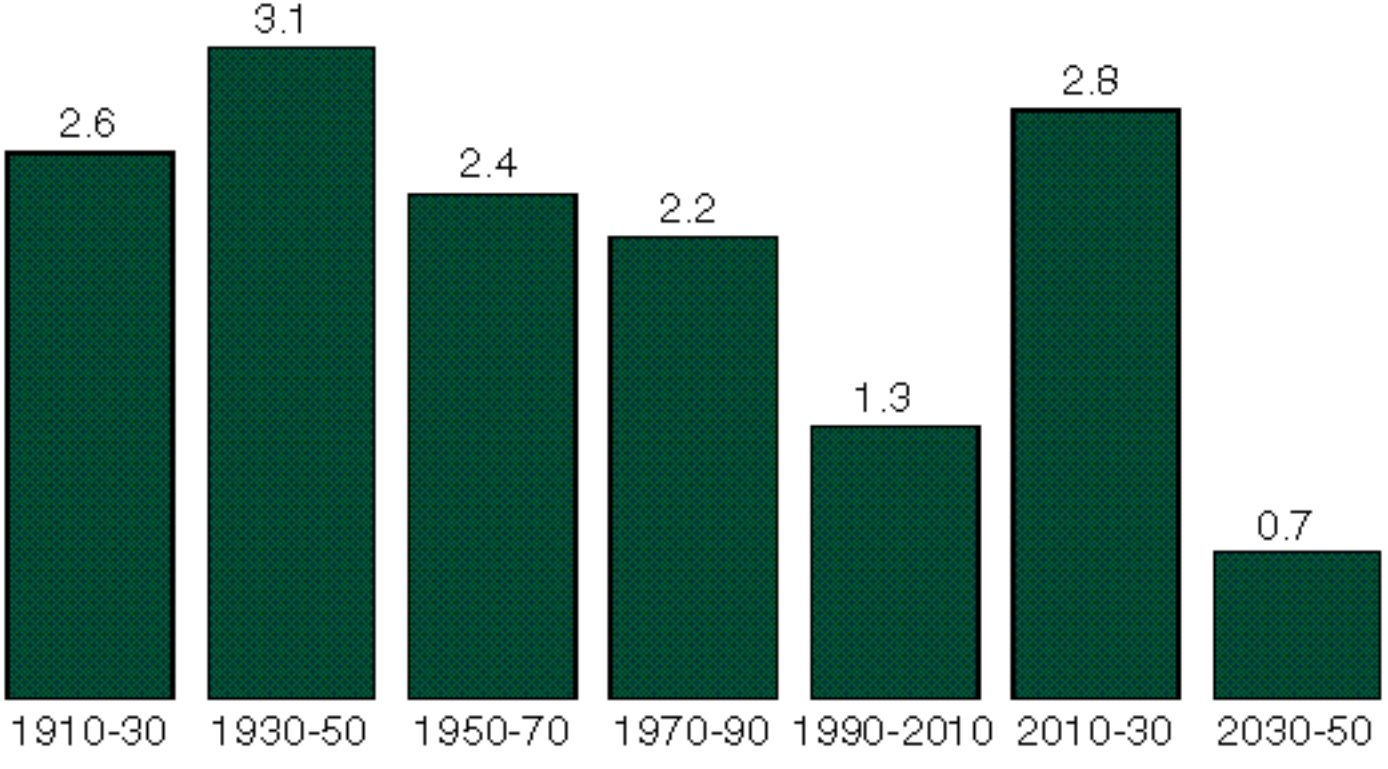
exhibit 10

note
Projections based on Congressional Budget Office adjustments to data from the Medicare Current Beneficiary Survey. Includes spending by and for both the non-institutionalized and institutionalized Medicare population for prescription drugs not currently covered by the Medicare program (e.g., most outpatient prescription drugs). Includes spending from all payment sources, including out-of-pocket spending and private and public health coverage payments.

source
Congressional Budget Office, January 2001 baseline.

Fifteen Years From Now, Elderly Population Growth Will Explode

Average annual growth rate (in percent) of the elderly population: 1910-30 to 2030-50



http://www.census.gov/population/socdemo/statbriefs/agebrief.html

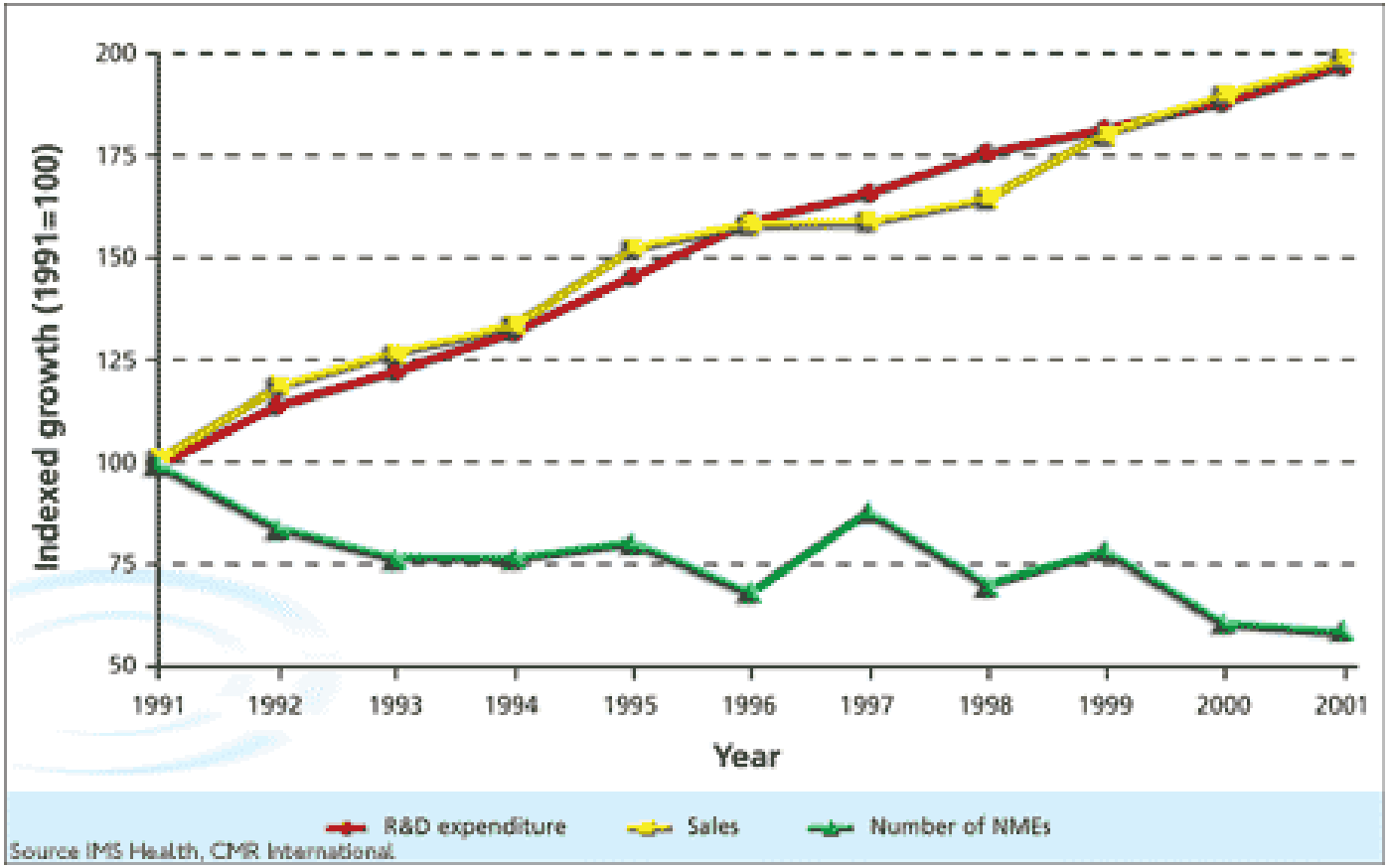
ADRs are a Major Cause of Death!!!

Deaths Per Year	Cause
710,000	Heart Disease
550,000	Cancer
170,000	Stroke
120,000	Pulmonary
100,000	***Adverse Drug Reactions***
98,000	Accidents
69,000	Diabetes
65,000	Pneumonia/Flu
50,000	Alzheimers
37,000	Nephritis
31,000	Septicemia

Refs: CDC Fastats estimated 2000 causes of death(http://www.cdc.gov/nchs/fastats/lcod.htm);
To Err Human, National Institute of Medicine, 1999; Bates et al., Incidence of adverse drug events and potential adverse drug events. JAMA 274:29, 1995;
Porter & Jick, Drug-related deaths among medical inpatients. JAMA 237:879-281, 1977.

See: http://www.drugintel.com/pharma/cause_of_death.htm

Current cost of bringing a new medicine to market can be as high as $0.8 to 1.7 billion*



See: http://www.cmr.org/pdfs/springnews2002.pdf

* Tufts Center for the Study of Drug Development,Backgrounder: How New Drugs Move Through the Development and Approval Process, Boston: November 2001; and Gilbert J, P Henske, and A Singh, "Rebuilding Big Pharma's Business Model," In Vivo, the Business & Medicine Report,Windhover Information,Vol. 21, No. 10,November 2003.

Attrition

Reasons for Failure	Reasons for Slow Down
Poor biopharmaceutical properties, 41%	Poor biopharmaceutical properties
Lack of efficacy, 31%	Low potency
Toxicity, 22%	Ambiguous toxicity finding
Market reasons, 6%	Inherently time-intensive target indication
	Synthetic complexity

• Prentis, R. A., Lis, Y. & Walker, S. R. Pharmaceutical innovation by the seven UK-owned pharmaceutical companies (1964-1985). *Br J Clin Pharmacol* 25, 387-96. (1988)

• Lipper, R.A. How can we optimize selection of drug development candidates from many compounds at the discovery stage? *Modern Drug Discovery,* 1999, 2 *(1)*, 55-60.

Challenge and Opportunity

- **“New compounds entering Phase-I testing after almost a decade of preclinical research have only an 8% chance of reaching the market.”**
- **“One pharmaceutical company estimates that clinical failures based on liver toxicity alone have cost them more than $2 billion in the last decade.”**
- **“Improving the prediction of failures before clinical trials by only 10% could save $100 million in development costs per drug.”**
- **“Extensive use of in silico technologies could reduce the overall cost of drug development by as much as 50 percent.”**

- **Challenge and Opportunity on the Critical Path to New Medical Products. FDA, 2004 (http://www.virtualscopics.com/pdf/Critical_Path_to_New_Medical_Products.pdf)**
- **Rotman,D,"Can Pfizer Deliver?" Technology Review, February 2004**
- **PricewaterhouseCoopers,"Pharma 2005 Silicon Rally:The Race to e-R&D" Paraxel's Pharmaceutical R&D Statistical Sourcebook 2002/2003**

Predictive Toxicology – Day Two

In Silico Models

Wednesday, January 12, 2005

8:40 **Opening remarks** –

Dr. Steven Muskal

8:45 **A Model for Herg K Channel Binding** –

Dr. Kalyanasundaram Subramanian

9:15 **Predictive GenoToxicity, there is Light at the End of the Tunnel** –

Dr. Joseph Votano

9:45 **Predictive, Structure Based, In Silico P450 Metabolism: Principle, Problems, and Progress** —

Dr. Dan L. Harris

10:15 **Coffee Break**

10:45 **Critical Pathway Identification: Integrated Transcriptional Profiling and Bioinformatics to Predict Toxicity** -

Dr. Gordon Kennovin

11:15 **Leveraging Reference Similarity as a Means to Rapidly Gauge Compound Toxicity -**

Dr. Steven Muskal

11:45 **Lunch on your own**

Leveraging Reference Similarity as a Means to Rapidly Gauge Compound Toxicity

Steven Muskal, Ph.D.
Chief Executive Officer
Sertanty, Inc
1/12/2005

Did you know?

Compound	$LD50_{(oral/rat)}$	Reference
H2O	>90 mL/kg	Food Research 21,348,1956
NaCl	3000 mg/kg	Toxicology and Applied Pharmacology 20,57,1971
Sucrose	29700 mg/kg	Toxicology and Applied Pharmacology 7,609,1965
Caffeine	192 mg/kg	Journal of New Drugs 5,252,1965

Note: For a 170 Lb human (77.1 Kg):
LD50 Water: 6.9 L -> 234.7 fl-oz -> 29 x 8oz cups!
LD50 Caffeine: 14.8 g -> 110 strong expresso shots! A cup of coffee contains between 60-135 mg caffeine
LD50 Sucrose: 2289.9 g ->10x8oz cups of sugar!

Toxicity Prediction Software Packages are Available

- DEREK - (J. E. Ridings *et al.*, *Toxicology* 106, 267-79. (1996)
- CASETOX/MULTICASE (G. Klopman, *J Chem Inf Comput Sci* 38, 78-81. (1998)
- TOPKAT (http://www.accelrys.com/products/topkat/)
- CSGenoTox (http://www.chemsilico.com/)
- TOXSYS (http://www.scivision.com/ToxSys.html)
- HazardExpert/ToxAlert (http://www.compudrug.com/)
- OncoLogic – (http://www.logichem.com/)

A Simple Toxicity Assessment Strategy – "Like Behaves Like"

- **Premise**: A compound's toxicity (e.g. pLD50) can be gauged based on the toxicities of other structurally similar compounds.
- **Algorithm**:
  - For each query-mol identify "similar" reference-mol(s).
  - Calculate average reference property ("ARP") from similar reference-mol(s).
  - Assign "ARP" to query-mol, provided a minimum number of similar reference-mol(s) exist.
- **Requirement**: A reasonably large, descriptive reference set

Defining Molecule Similarity – Structural Keys v. Fingerprints

A *structural key* is typically a bitmap in which each bit represents the presence (TRUE) or absence (FALSE) of a specific structural feature or pattern.

Some 2D structural key examples:

- The presence/absence of an element, or "at least 1 N", "at least 2 N", etc.
- Unusual or important electronic configurations, e.g "sp3 carbon" or "triple-bonded nitrogen."
- Rings and ring systems, such as cyclohexane, pyridine, or napthalene.
- Common functional groups, such as alcohols, amines, hydrocarbons, etc.
- Functional groups of special importance in a particular database. E.g. a database of organo-metallic molecules might have bits assigned for metal-containing functional groups; in a drug database one might have bits for specific skeletal features such as steroids and barbiturates.
- "Disjunctions" of unusual features.

Defining Molecule Similarity – Structural keys v. Daylight Fingerprints

A Daylight *Fingerprint* is a boolean array, or bitmap. Unlike a structural key with its pre-defined patterns, the patterns for a molecule's fingerprint are generated from the molecule itself.

The Daylight fingerprinting algorithm examines the molecule and generates the following

- a pattern for each atom
- a pattern representing each atom and its nearest neighbors (plus the bonds that join them)
- a pattern representing each group of atoms and bonds connected by paths up to 2 bonds long
- ... atoms and bonds connected by paths up to 3 bonds long
- ... continuing, with paths up to 4, 5, 6, and 7 bonds long.

A Common Structural Similarity Metric

Tanimoto Coefficient = BC / (B1 + B2 - BC)

Symbol	Definition	Description
bits(F)		A function that returns the number of "1" bits in a bitmap
BT		The total number of bits (the fingerprint's size); a constant
B1 =	**bits(F1)**	**The number of 1's in F1**
B2 =	**bits(F2)**	**The number of 1's in F2**
BC =	**bits(F1 AND F2)**	**The number of 1's in common between F1 and F2**
BI =	bits(F1 XOR (NOT F2))	The number of identical bits (1's and 0's) between F1 and F2
BU1 =	bits(F1 AND (NOT F2))	The number of unique bits (1's) in F1
BU2 =	bits(F2 AND (NOT F1))	The number of unique bits (1's) in F2

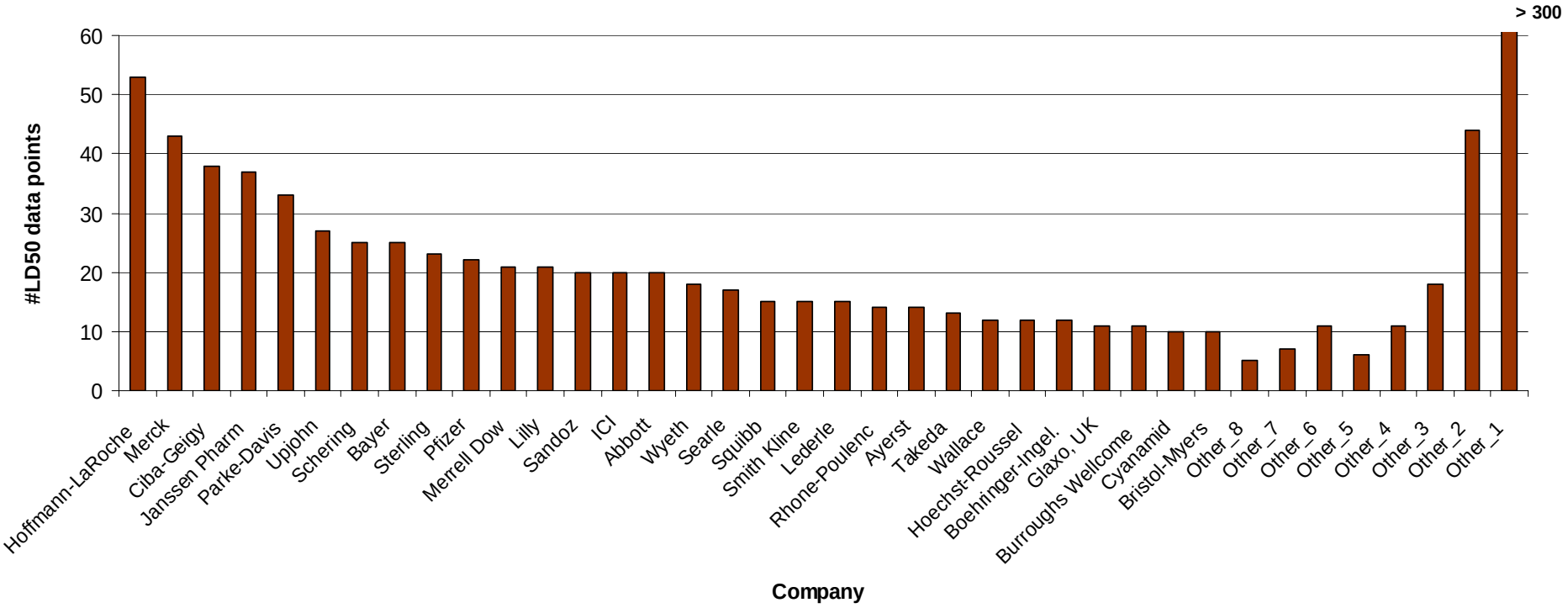
A Starter Reference Set – “RefSet”

- **Starter Source**: RTECS – over 133K compounds
  - http://www.nisc.com/factsheets/qrtc.htm
- **Subset** - 13645 examples
  - Route: Oral Species: Rat EndPoint: LD50
- **MWT**: Avg: 304.09 Std: 183.78
- **ClogP**: Avg: 2.05 Std: 2.53
- **QPlogS**: Avg: -2.95 Std: 2.37
- **TPSA**: Avg: 65.28 Std: 70.39
- **RotBonds**: Avg: 5.32 Std: 5.5

Drug Subset- “DrugSet”

- Comprehensive Medicinal Chemistry – CMC 2002.1 (MDL)
- Approx. 8500 compounds tested in/on man
- “DrugSet” subset: 1781 “RefSet” compounds found in CMC
- **MWT**: Avg: 367.4 Std: 95.5
- **ClogP**: Avg: 2.87 Std: 2.7
- **QPlogS** Avg: -3.97 Std: 2.1
- **TPSA**: Avg: 72.24 Std: 48.3
- **RotBonds**: Avg: 5.96 Std: 4.2

DrugSet Source Companies

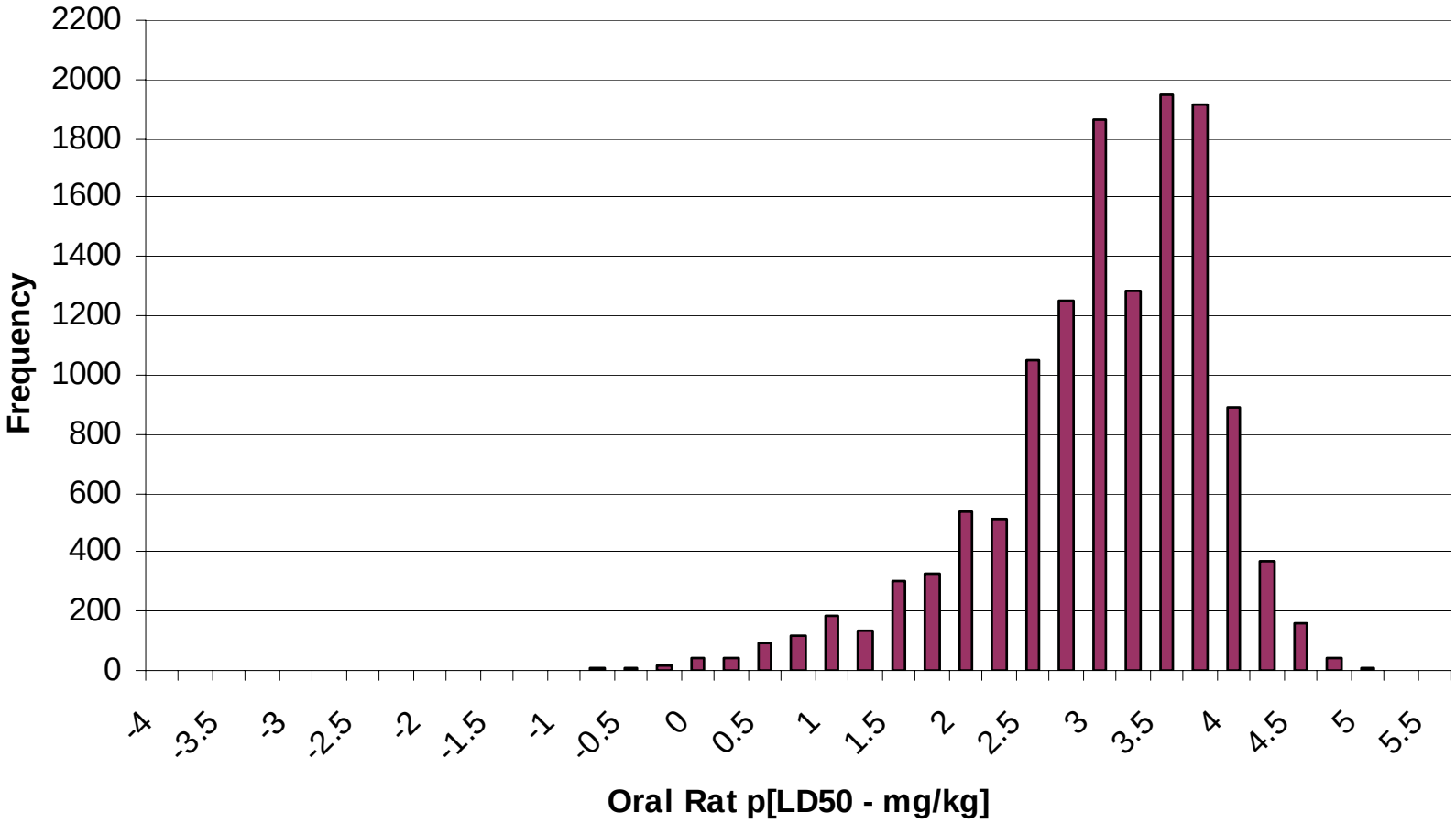


Notes: LD50 data points in DrugSet which have company source data in CMC 2002.1, legacy company-source names (e.g. Novartis/Sandoz/Ciba-Geigy, ICI/Zeneca, Wyeth/Ayerst, SmithKline/Glaxo/Wellcome, etc.) notwithstanding. Other_8, Other_7 represent companies with 8, 7, etc. data points. Over 300 source-company entries have 1 compound-LD50 data-point.

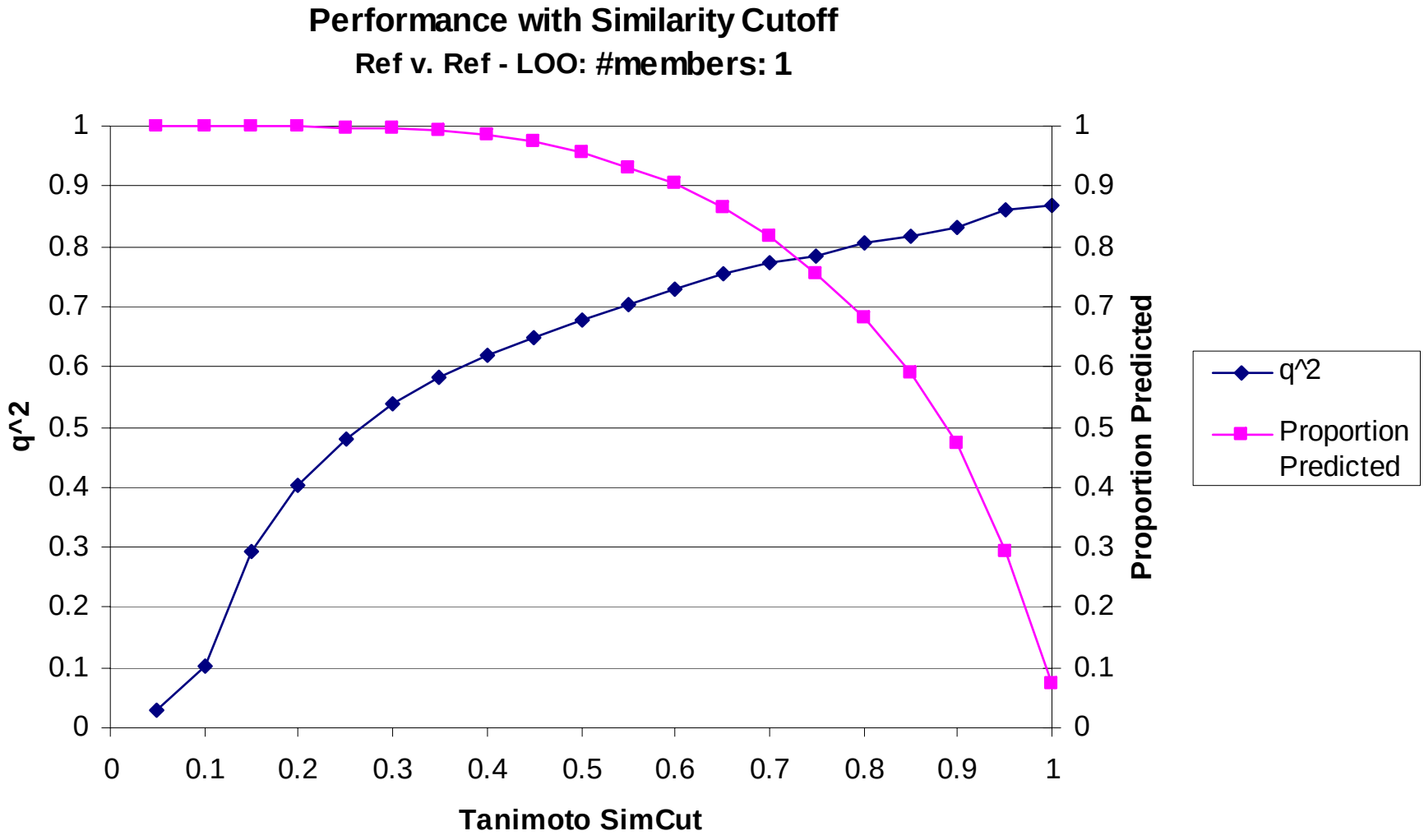
Reference Set Distribution

Oral Rat log(LD50) – pLD50

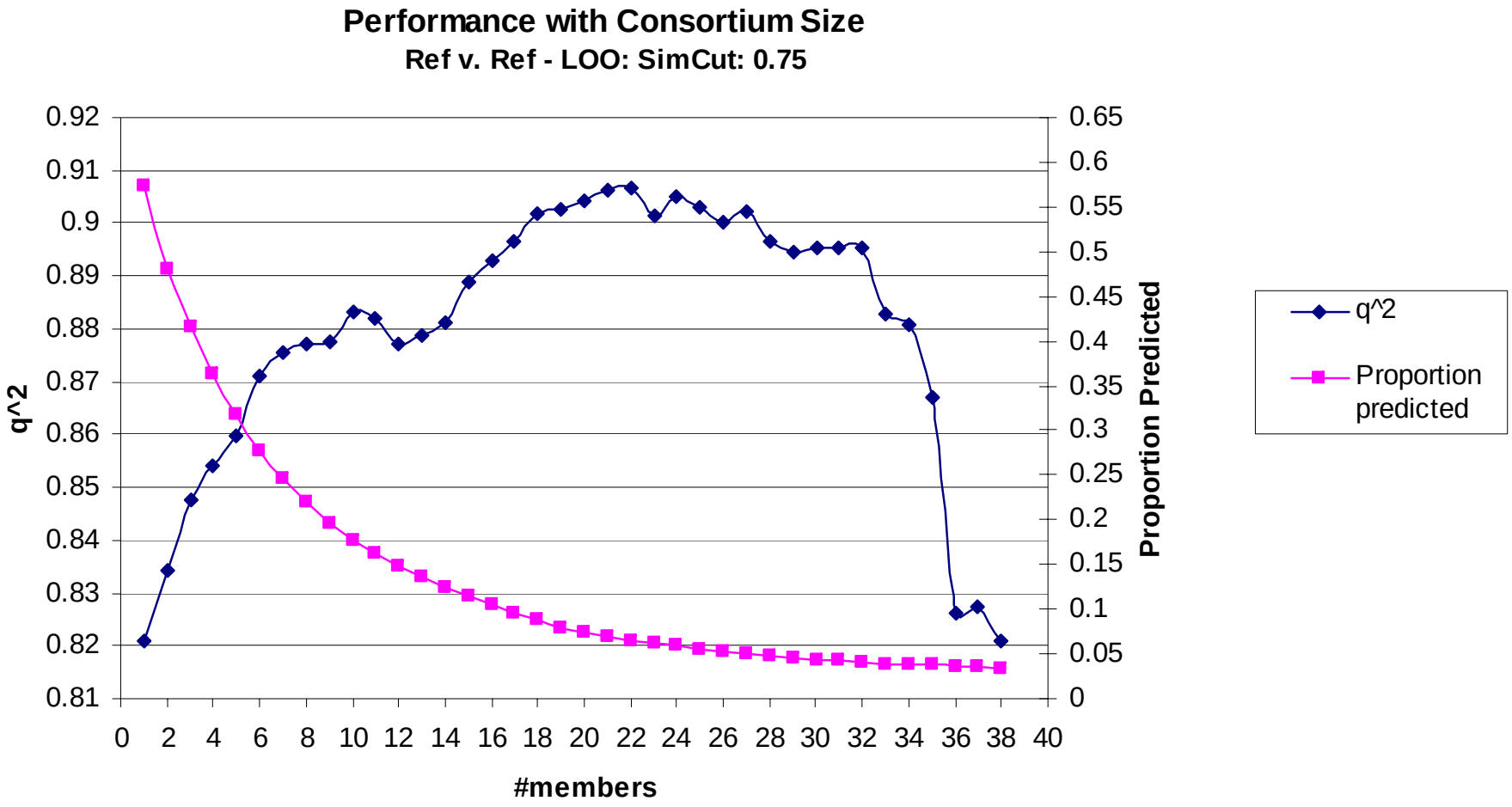
Range: [-3.85, 5.27] Avg: 2.92 SDev: 0.85



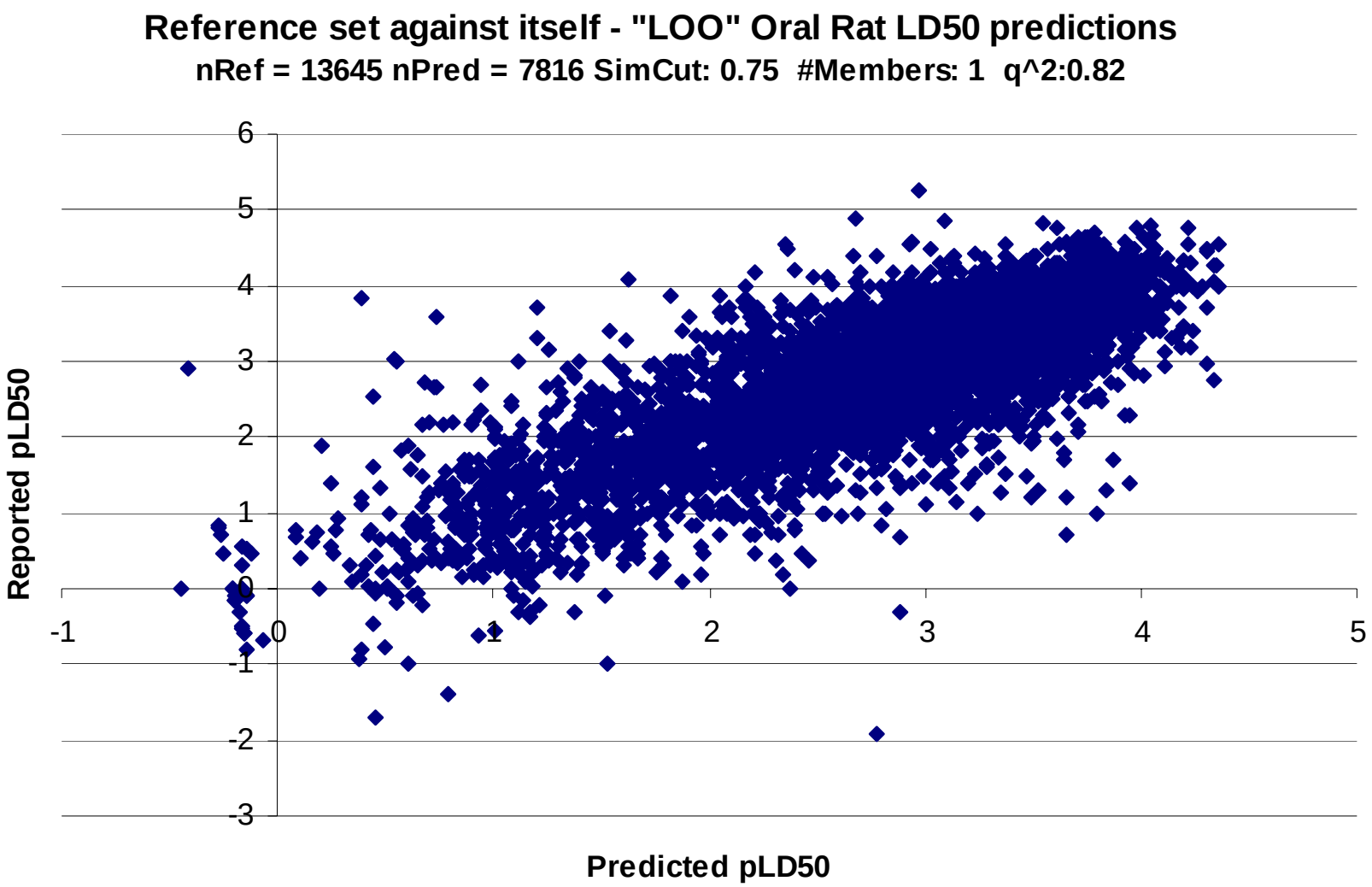
What is the Optimal Similarity Cutoff?



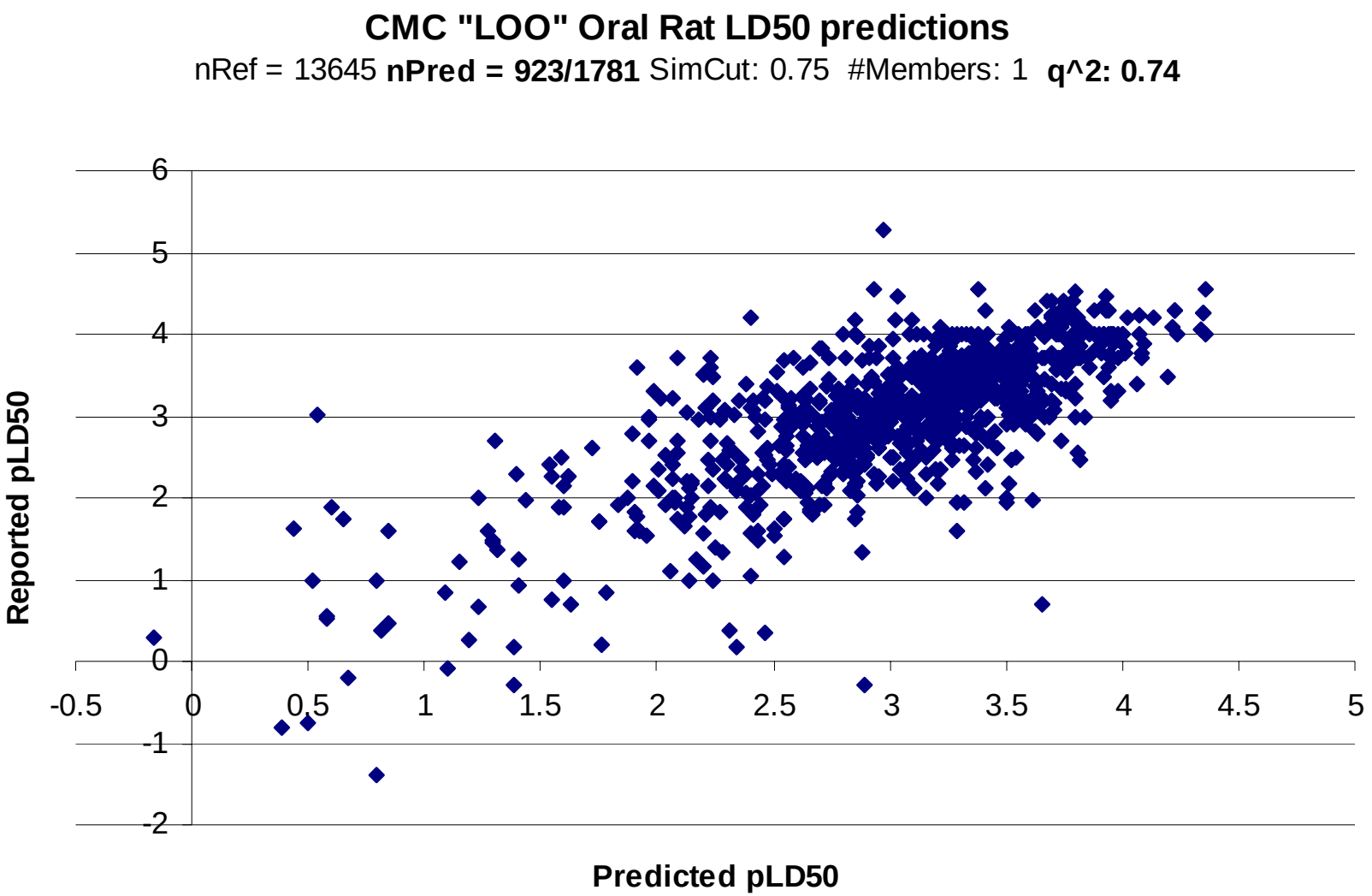
What is the Optimal Consortium Size?



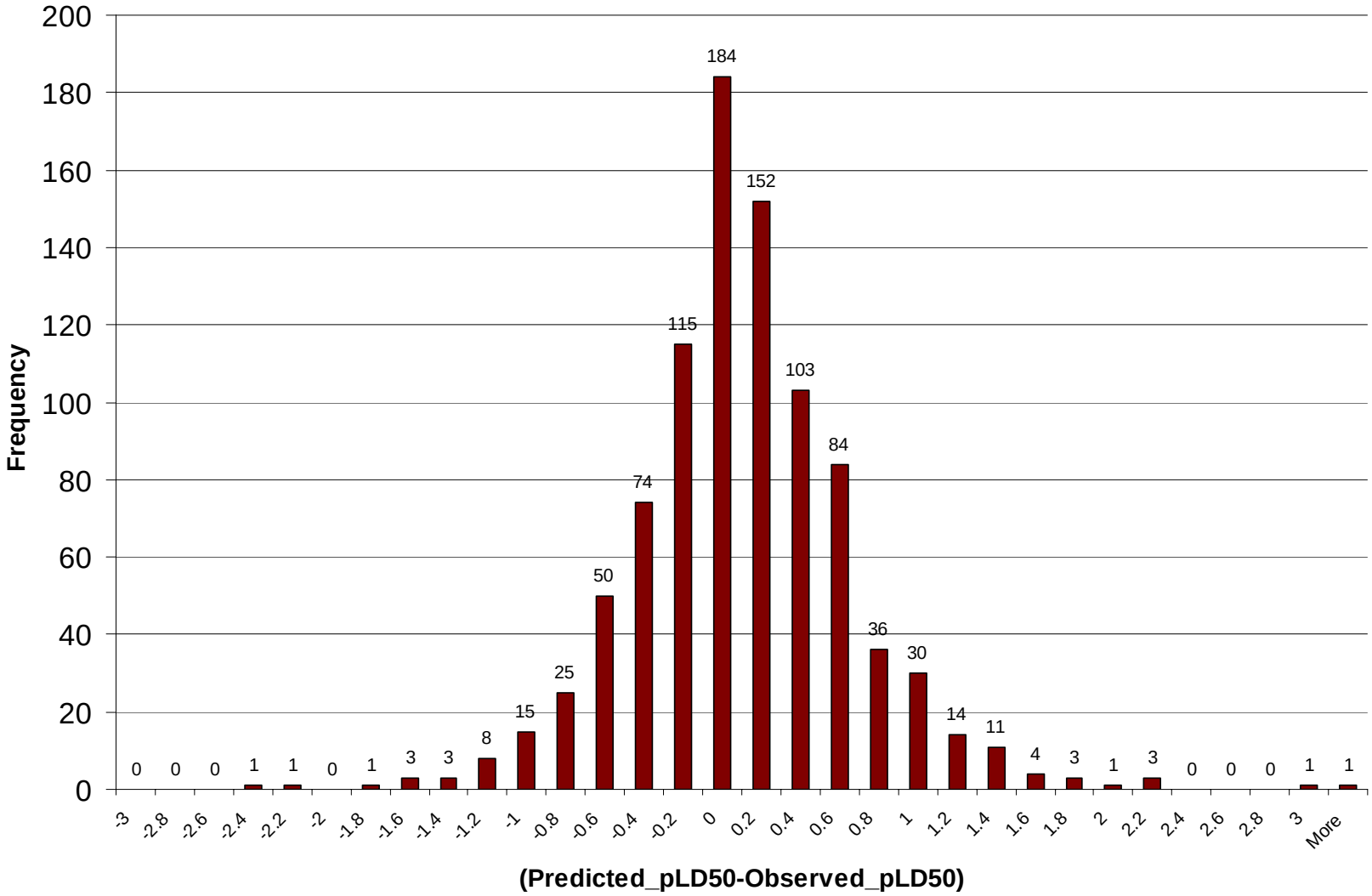
Evaluating RefSim – A Leave-One-Out Simulation



Predicting Drug Toxicity – “DrugSet”



DrugSet Prediction Errors



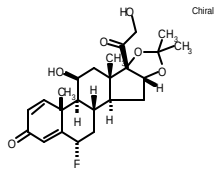
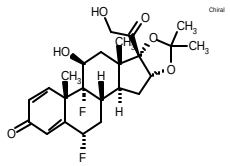
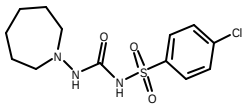
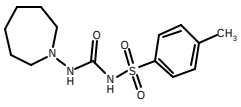
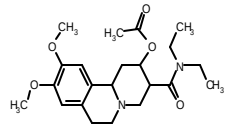
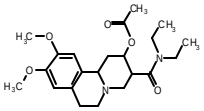
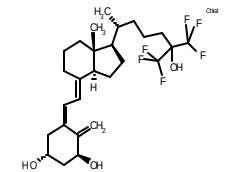
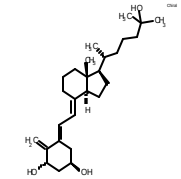
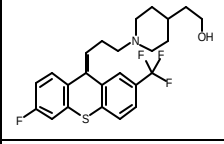
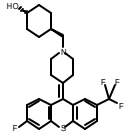
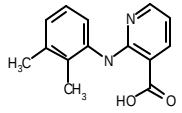
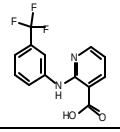
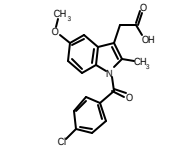
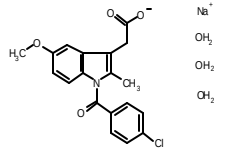
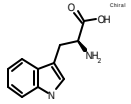
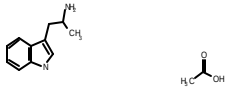
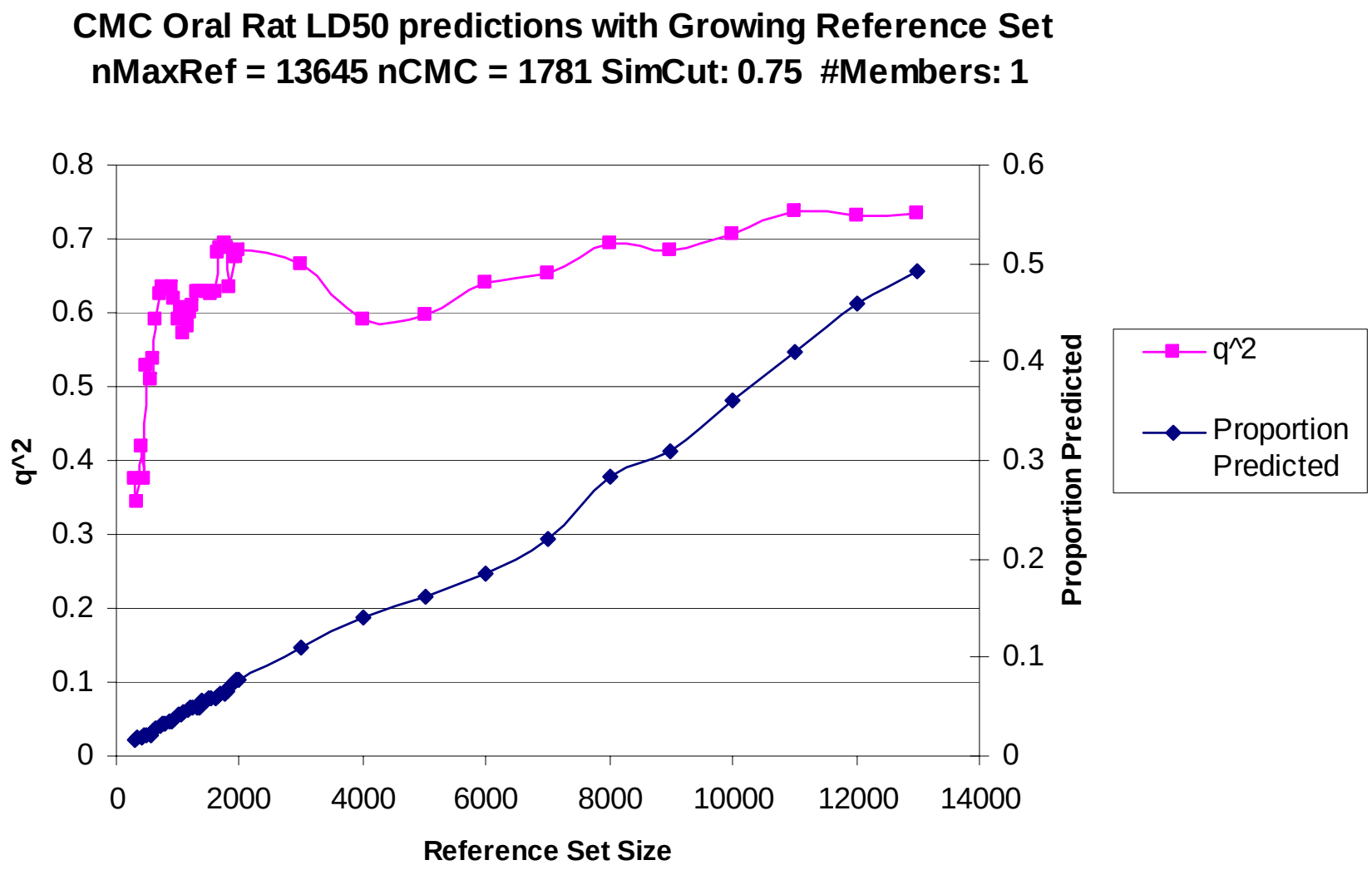
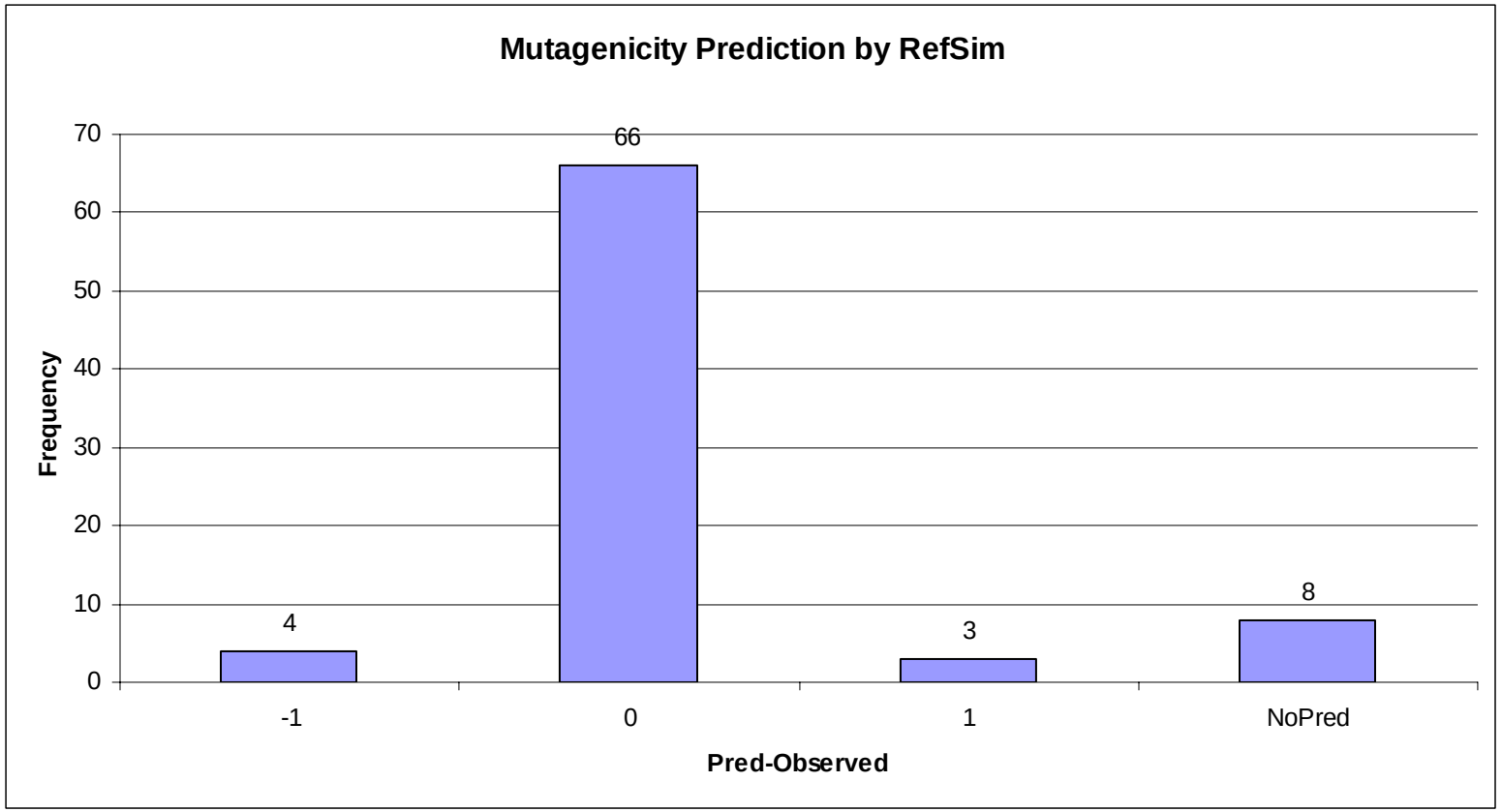
DrugSet Molecule Name	DrugSet Molecule (DSM)	DSM Observed LD50	DSM Reference	DSM Activity Class (CMC)	RefSet Most Similar Molecule (RSM)	RSM Observed LD50	RSM Reference	DSM Prediction Error: (Pred-Obs)	Num Ref Sim Mol	Tanimoto DSM&RSM
FLUNISOLIDE		>500 ug/kg	Gekkan Yakuji 26,501,1984	Glucocorticoid		>4 gm/kg	Drugs in Japan (Ethical Drugs) 6,694,1982	3.19	4	0.94
GLYPINAMIDE		>5 mg/kg	Patent, French Medicament Document #1087M	Antidiabetic		>5 gm/kg	Drugs in Japan (Ethical Drugs) 6,511,1982	2.96	2	0.87
BENZQUINAMIDE		1050 mg/kg	Psychotropic Drugs and Related Compounds - ,208,1972	Antiemetic		990 mg/kg	Toxicology and Applied Pharmacology 18,185,1971	-2.48	2	0.99
FLOCACITRIOL		41700 ng/kg	Kiso to Rinsho 30,2695,1996	Ca regulator		620 ug/kg	Patent, Japanese Kokai Tokyo Koho #94-247858	2.17	7	0.86
PIFLUTIXOLE		1500 ug/kg	Patent, United States Document #4309429	Neuroleptic		>60 mg/kg	Patent, United States Document #4309429	2.16	2	0.81
NIXYLIC ACID		2300 ug/kg	Therapie 22,157,1967	Antiinflammatory		250 mg/kg	Journal of Medicinal Chemistry 16,780,1973	2.10	2	0.94
INDOMETHACIN		2420 ug/kg	Arzneimittel-Forschung 25,1526,1975	Antiinflammatory		21 mg/kg	Gekkan Yakuji 37,952,1995	1.93	24	0.86
TRYPTOPHAN		>16 gm/kg	Iyakuhin Kenkyu 11,635,1980	Antidepressant		22 mg/kg	Toxicology and Applied Pharmacology 4,547,1962	-1.81	4	0.99

Table 2. DrugSet prediction outliers. The reported LD50's for the query DrugSet molecules (DSMs) and for their respective maximally similar molecules in the reference set (RSMs) show significant differences despite high Tanimoto similarity between DSM and RSM. This helps explain the fairly large absolute, signed DSM prediction error: pLD50(pred)-pLD50(obs).

How to Enhance Toxicity Predictions?



eMutagenicity by RefSim



Data From:
Handbook of Carcinogenic Potency and Genotoxicity Databases (Lois Swirsky, Gold and Errol Ziegler, Eds.), CRC Press, 1997, pp. 694-724
Data Files contain the experimental mutagenicity data in 1 or 0's. 742 Reference set examples; 81 Test set examples; Refsim cutoff: 0.5
One indicates a mutagen and zero indicates that the compound has no mutagenic effect on Salmonella typhimurium bacteria (AMES test).

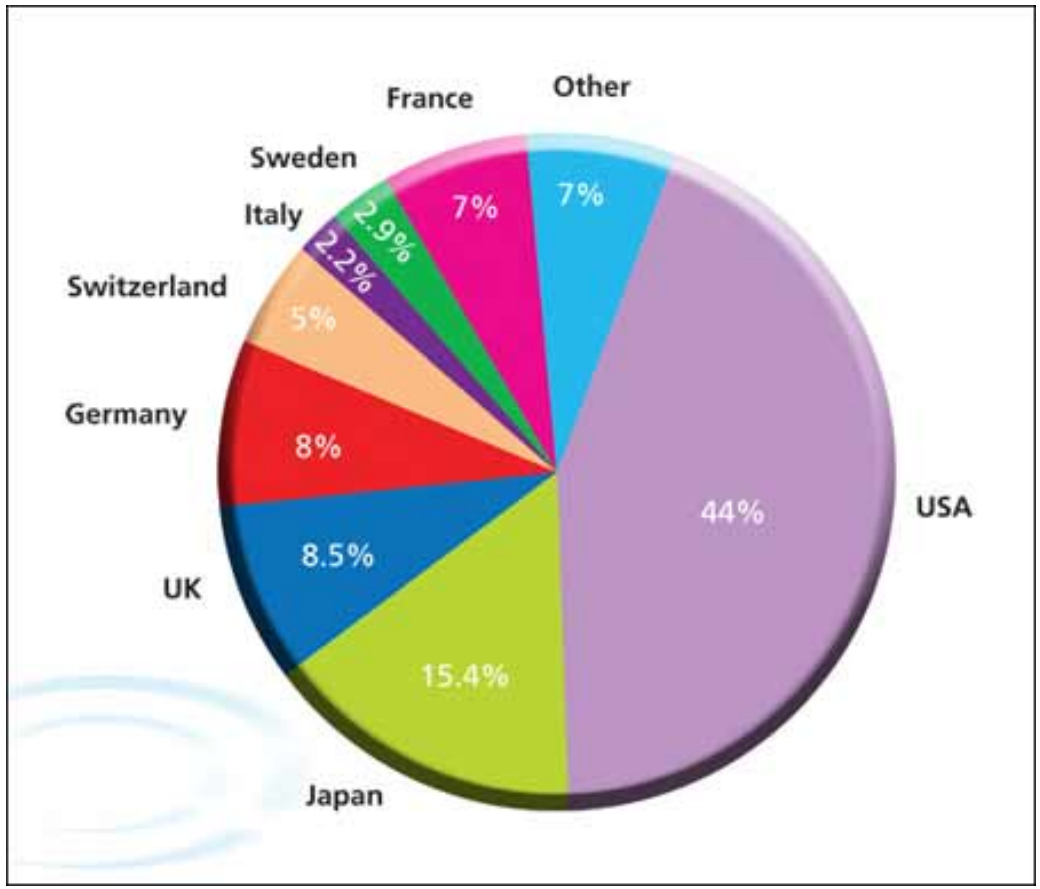
Summary & Conclusions

- Reasonably accurate and robust toxicity predictions can be achieved with a reference similarity approach
- Small to moderate consortia per compound-class may suffice to build a well rounded reference set
- An increase in reference set size is likely to improve both the quality and quantity of toxicity predictions
- Significant opportunity exists to enhance the “Starter” reference set, which only scratches the surface…
- Prediction technique can be readily incorporated into a high-throughput in silico eScreening strategy (e.g. compound prioritization, filtration, etc.)

There Can Be Safety In Numbers…

Supplementary Slides

Global Pharmaceutical R&D Expenditure by Country



Source:CMR International - http://www.cmr.org/pdfs/springnews2002.pdf

ADRs are a major cause of Death

Adverse Drug Reactions may be the fourth to sixth leading cause of death

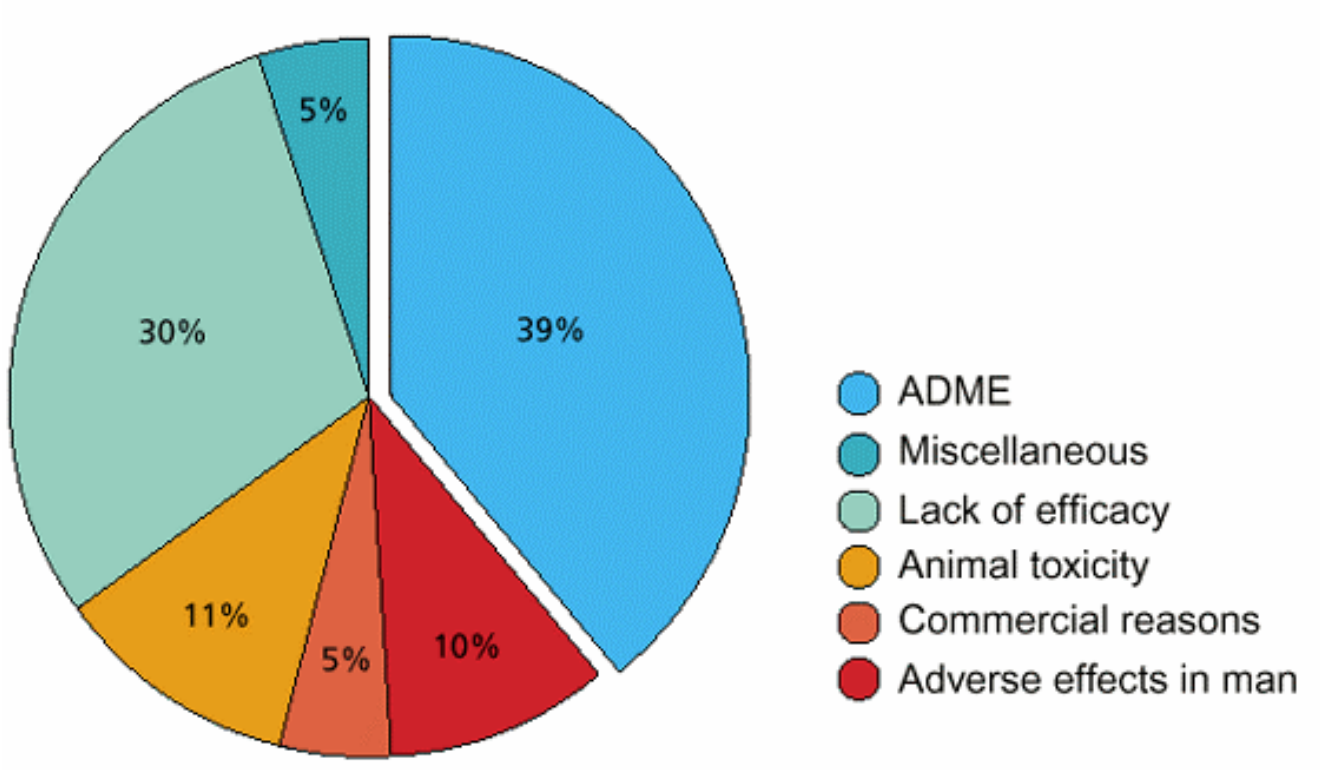
Deaths Per Year	Cause
106,000	**Non-error, negative effects of drugs[1]**
80,000	**Infections in hospitals[4]**
45,000	**Other errors in hospitals[4]**
12,000	**Unnecessary surgery[2]**
7,000	**Medication errors in hospitals[3]**
250,000	**Total deaths per year from iatrogenic* causes**

* The term *iatrogenic* is defined as "induced in a patient by a physician's activity, manner, or therapy. Used especially to pertain to a complication of treatment."

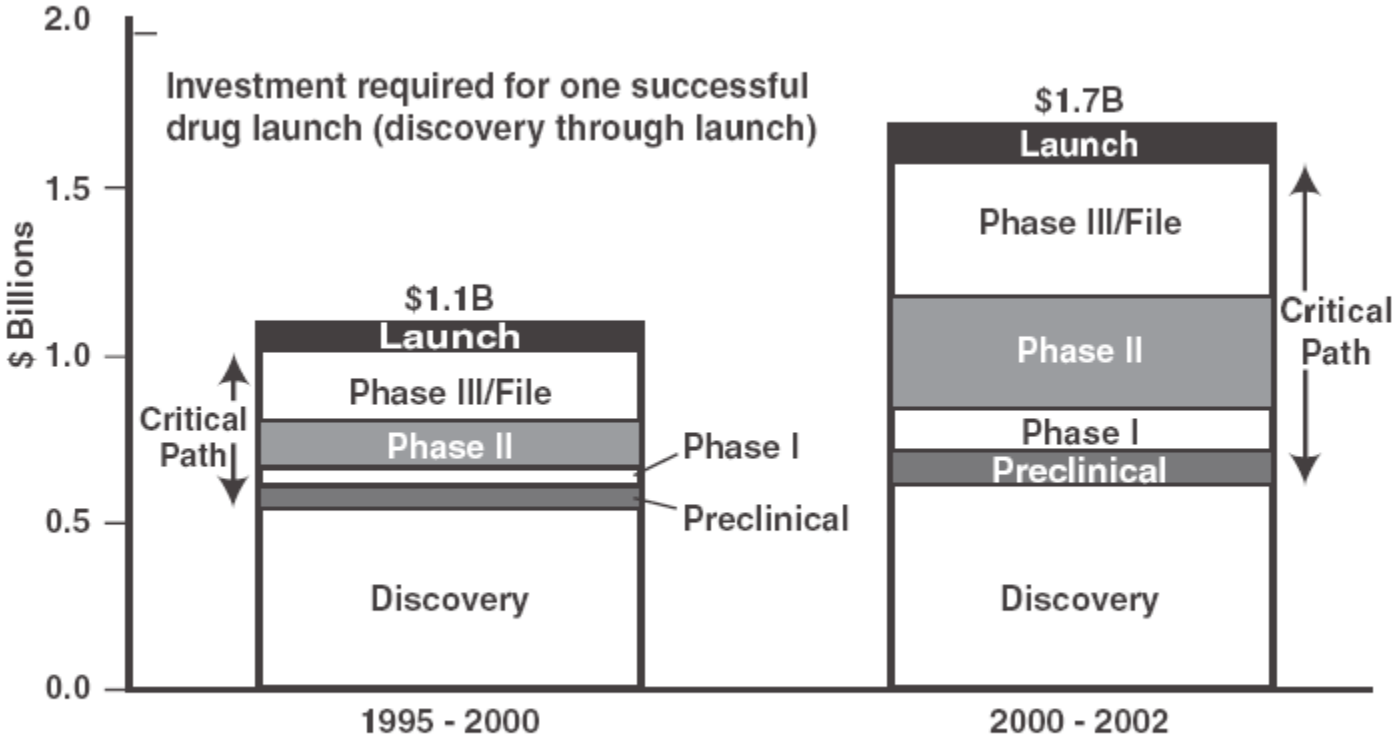
1. Kohn L, ed., Corrigan J, ed., Donaldson M, ed. To Err Is Human: Building a Safer Health System. Washington, DC: National Academy Press, 1999.
2. Leape L. Unnecessary surgery. Annual Rev. Public Health. 1992; 13:363-383.
3. Phillips D, Christenfeld N, Glynn L. Increase in U.S. medication-error deaths between 1983 and 1993. Lancet, 1998; 351:643-644.
4.Lazarou J, Pomeranz B, Corey P. Incidence of adverse drug reactions in hospitalized patients. JAMA. 1998; 279:1200-1205.

See: http://www.chelationtherapyonline.com/articles/p62.htm

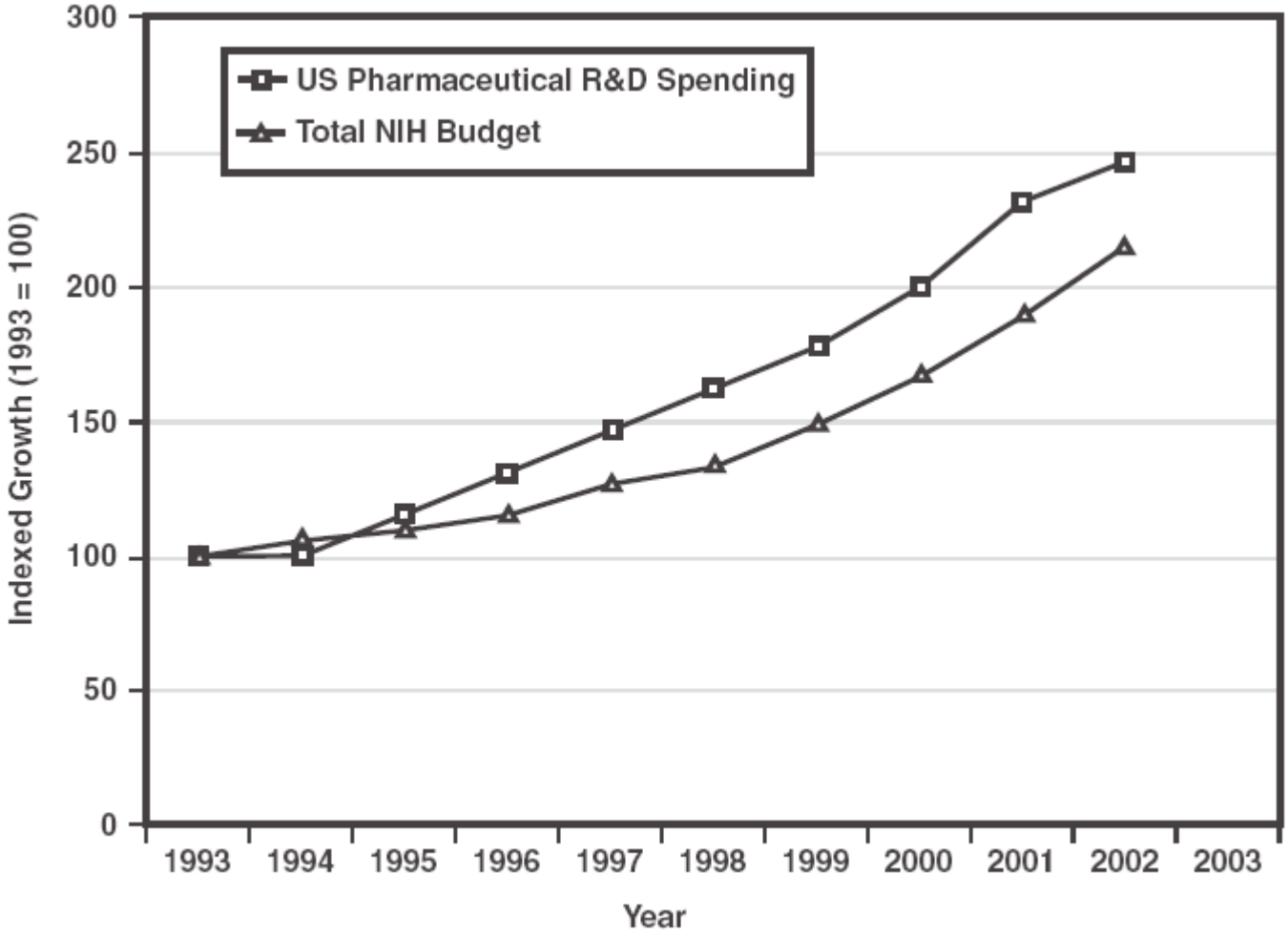
Causes of attrition during drug development

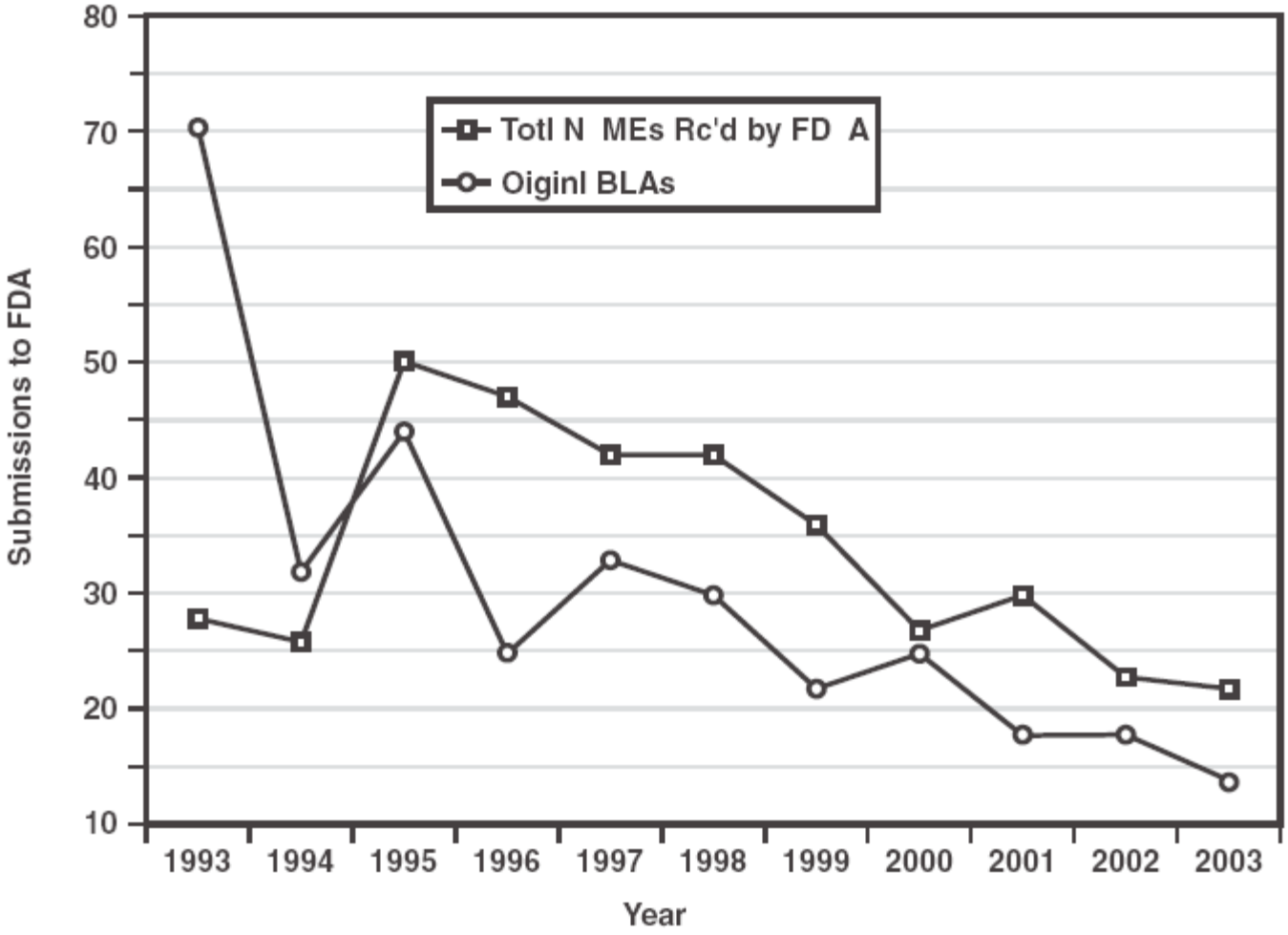


Source: Kennedy, T. *Drug Discovery Today* **2, 436–444, 1997.**



SOURCE: Windhover's In Vivo: The Business & Medicine Report, Bain drug economics model, 2003


US Pharmaceutical R&D Spending
Total NIH Budget
Indexed Growth (1993 = 100)
300
250
200
150
100
50
0
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
Year


Totl N MEs Rc'd by FD A
Oiginl BLAs
Submissions to FDA
80
70
60
50
40
30
20
10
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
Year