Artificial Intelligence: Introduction, Applications, and Overview of Colloquium

Jim E. Riviere, DVM, PhD, ATS, DACVCP

Distinguished Professor Emeritus,
North Carolina State University and Kansas State University

Jim.Riviere@gmail.com

Conflict of Interest Statement

- I have no real or perceived conflicts of interests with the research described in this presentation.
- There is no animal experimentation discussed in this presentation.
- All opinions presented are solely those of the author.

Outline/Objectives

- We now have a great idea of the power and breadth of what Artificial Intelligence (AI) is and fascinating applications.
- Start by reviewing concepts of modeling and how AI applications have been entering my modeling research in food and cosmetic safety
  - FARAD, 1DATA Consortium, dermal absorption of complex mixtures, Nanoparticle modeling
- What is Artificial Intelligence (AI) and Machine Learning (ML)?
- Brief introduction to background, approaches and potential applications of these tools to cosmetic and food safety
- Strengths and weaknesses of AI and ML

Personal Philosophy of Modeling

- Model building involves testing assumptions
- *Lack of Fit* often indicates a specific relationship is not valid. This is not a bad thing !!
  - Species has different metabolic pathway
  - Distribution is different
  - *In vitro* data is not predictive of *in vivo*
- Models often identify data gaps which can then be used to design experiments to fill these gaps
- Modern computational tools allow all models to be easily updated with new data, stressing that all models should be viewed as *works in progress*.

Philosophical

- Experimental Design
  - Trained to block subjects before treatment
  - This assumes we know what to block on!
  - What happens when blocking variable is not known?
  - Do we reduce inference by seeking low variability?
- Interfacing between different levels of models are inherently difficult.
- Disease is nonlinear and multifactorial: Do we simplify too much? Do we know how to simplify?
- We model for different reasons: discovery vs. validation vs. regulatory testing
- We use animals to model humans.

Statistical Testing vs. Model Validation for Drugs

- Traditional experimental design divides testing treatment effects into blocks that are then tested against one another:

A vs. B vs. C vs. D

- Regression may analyze this as a time-series and assess whether a slope is present, implying differences amongst points:

A > B > C > D

- Alternate approach is to validate a specific drug's pharmacometric model in a population:

Is A, B, C, or D within the model-predicted Confidence Interval

(***Predominant tool use to validate AI predictions***)

Historically What Were Good Traditional Models?

- Mathematically and scientifically elegant
- Contains few arbitrary or adjustable elements
- Agrees with and explains existing observations
- Makes detailed predictions about future observations that can disprove or falsify the model
  - These are classic mechanistic hypothesis-based “scientific” models commonly used in many fields, where the model explicitly defines relationship between input and output: x = *f* [a,b,c,---]
  - May use large datasets and are done on computers

NOT what most Big Data AI and ML approaches do

In AI, data defines the relationship, not the programmer or model

Proliferation of New Modeling Approaches

- There are many ways to approach modeling and statistical analysis
- Historically, most have been model based and relatively structured
- As computing power and data storage/availability increased, new approaches to less “rigid” analytics became prevalent
- Use my work to illustrate evolution of this philosophy to contrast with true AI approaches now being developed
- We have been using many of these tools for a long time. Major difference now is ability to automate, speed, and increased data
- As these progress, the variety of statistical perspectives have grown and large datasets have changed the rules

Food Animal Residue Avoidance and Depletion Program (FARAD)

FARAD was founded in 1981 as a university consortium and funded by USDA.

FARAD is a consortium for identifying, gathering, extracting, analyzing, generating, and extending residue avoidance information to ensure that animal derived foods will be free of illegal chemical & drug residues and safe for consumers

FARAD's Mission

Keeping the Food Supply Safe
by Avoiding Drug, Pesticide and Biotoxin Residues

FARAD utilizes a computerized databank of information on chemical residues in food animals. Data are categorized into five major areas:

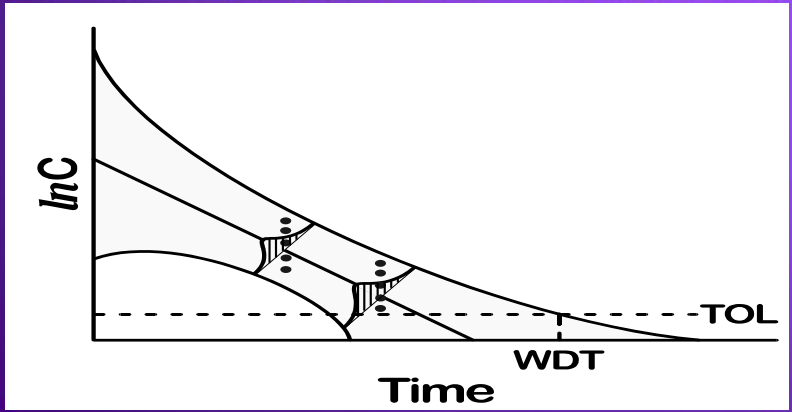
1. Approved Drugs for Food Animals
2. Tolerances for Residues
3. Pharmacokinetic data
4. Bibliographic Citations
5. Rapid Residue Screening Tests

Primary Task of FARAD Is to Estimate Withdrawal Times for Meat, Milk, and Eggs

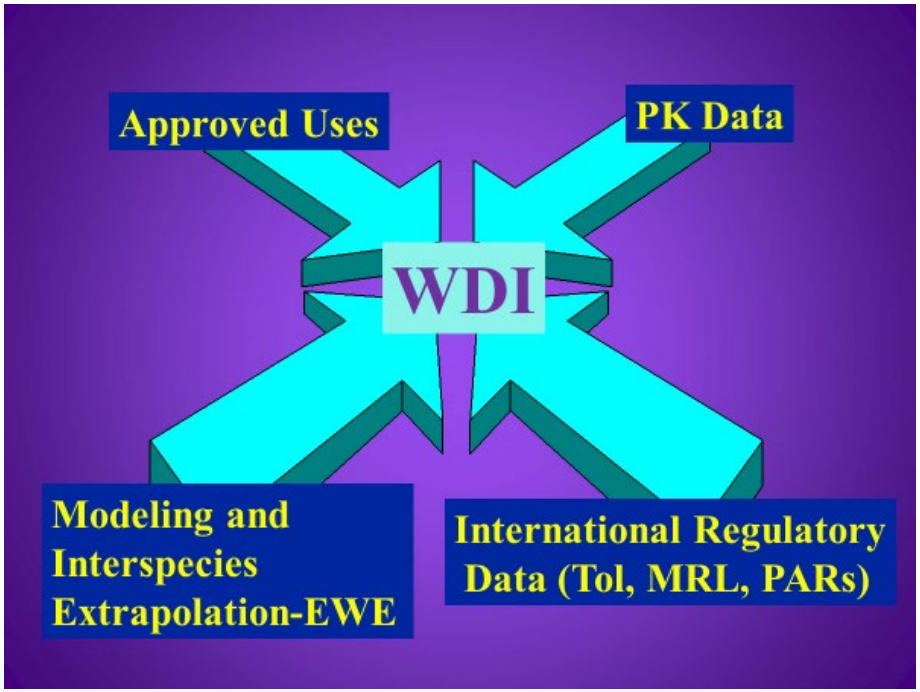
Establish Withdrawal Intervals (WDI) For:

1. Environmental Contaminants
2. Extra-label Drug Use for Major Animal Species
3. Extra-label Drug Use for Minor Animal Species
4. Bioterrorism
5. Accidental Exposure (Hurricanes, Meltdowns, Mistakes)

FDA has rigid guidelines for approved label drug withdrawal times (WDT)

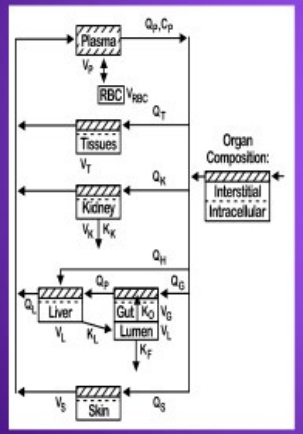


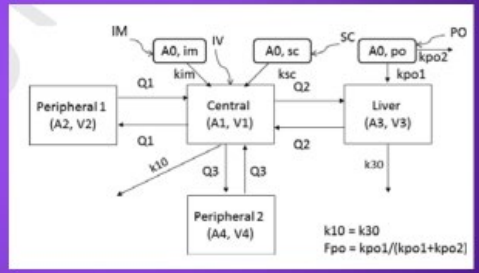
FARAD Uses Multiple Data Sources



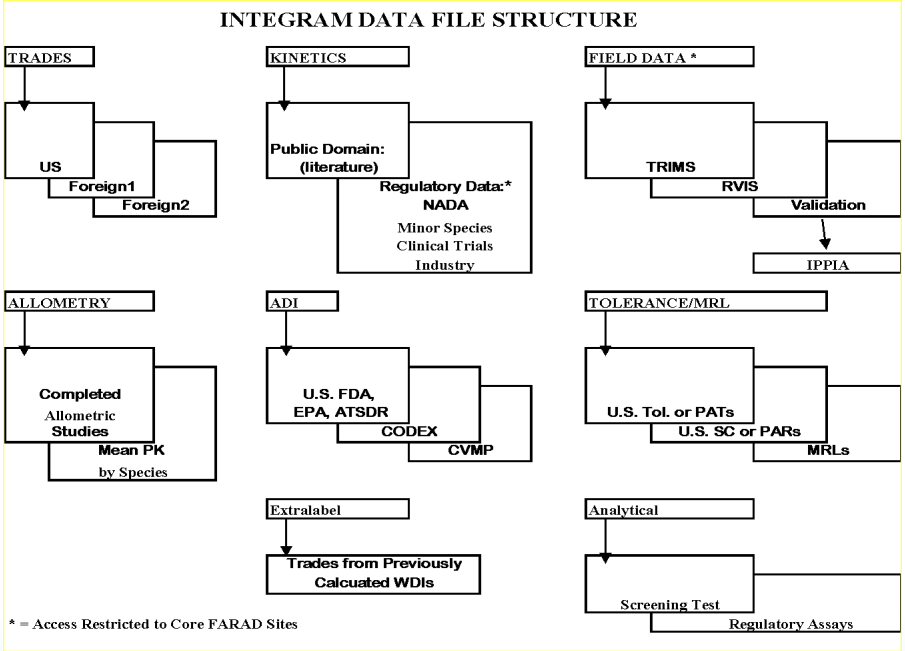
Two pharmacokinetic modeling approaches, a mechanistic and a stochastic are employed.

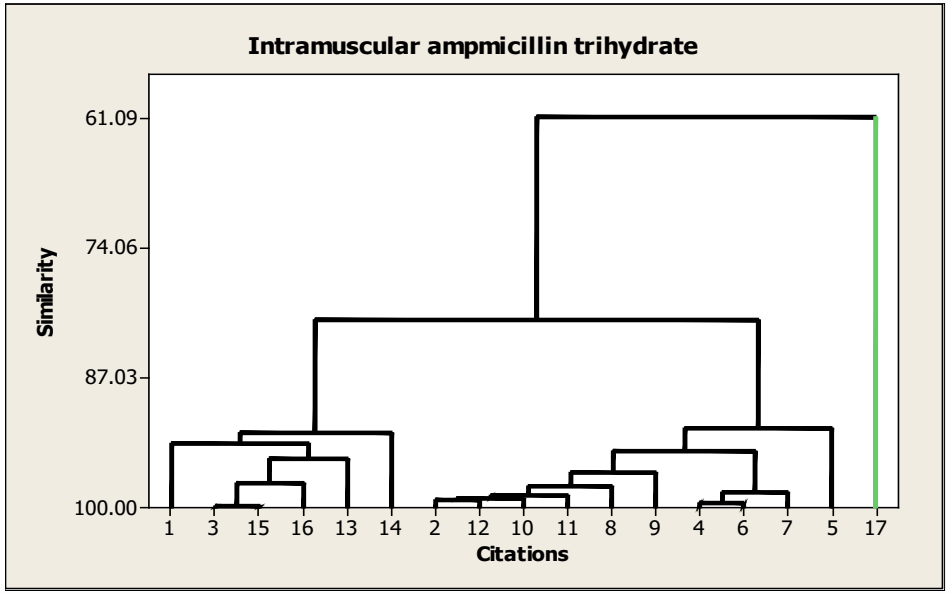
- Population Pharmacokinetic models (Pop PK; MEM)
- Physiological Based Pharmacokinetic Models (PBPK)





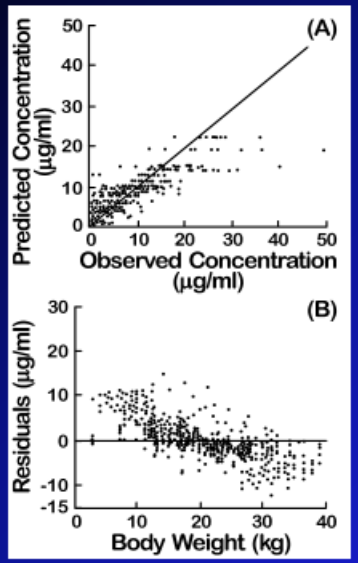
Relational Databases and Algorithms for Quality Control



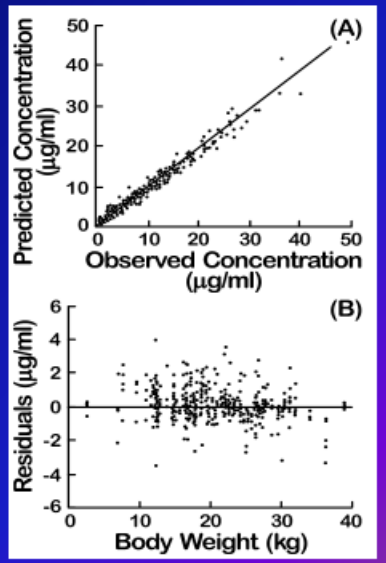


Mixed Effect Modeling (MEM)
Population Pharmacokinetics (PopPK)

Data Set Modeled with No Covariates



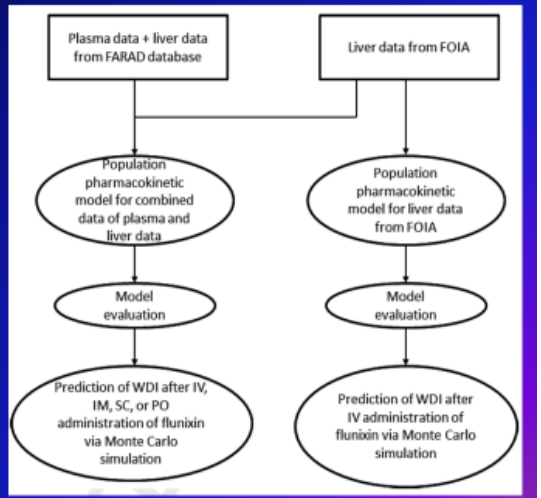
Data Set Modeled with Body Weight as a Covariate



Flunixin Studies in Cattle

Used data from six published studies of flunixin plasma deposition using various dosage regimens to a diverse groups of cattle

Liver data obtained from one publication and US FDA FOIA from sponsor submission.

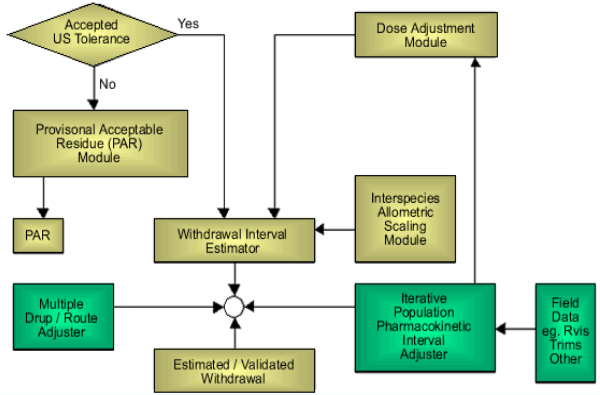


Wu, Baynes, Leavens, Tell, Riviere. *J. Vet. Pharmacol. Therap.* 36: 248-257, 2013.

FARAD aka 2020

When contaminant exposure or extra-label drug use occurs, a detailed request is filed on-line from a licensed veterinarian (www.FARAD.org). A number of independent datasets and tools are then queried.

- Is this legal? What is health status of animals?
- Does request match previous request?
- Search databases to find approved drug data, pharmacokinetic data, and target food safety tolerances in US and abroad (AI tools being increasingly employed here)
- Use available PK models to estimate WDI

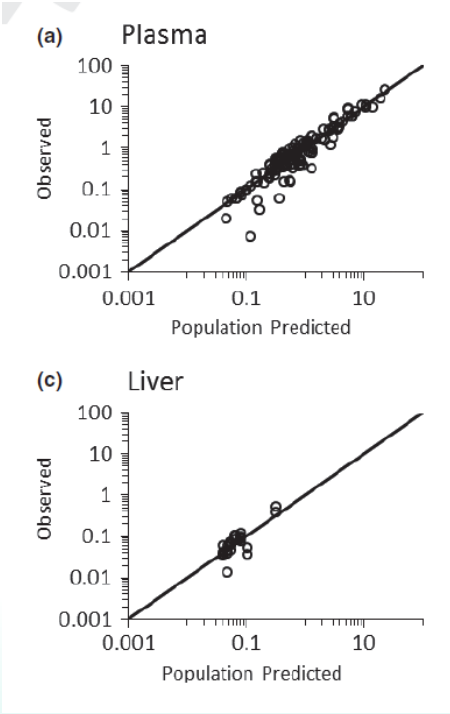


US Patent 6,066,091

Use Monte Carlo Techniques

- In certain situations like minor species or disease, we have knowledge of PK profile and some estimate of variance. We use these parameters to generate “synthetic populations” upon which we run the regulatory FDA tolerance limit algorithm (99% of population with 95% confidence of no residue)* to estimate WDI
- This is done on defined stochastic PBPK models or PopPK mixed-effect pharmacostatistical models
- These models can be updated with new data

* https://www.fda.gov/media/70028/download

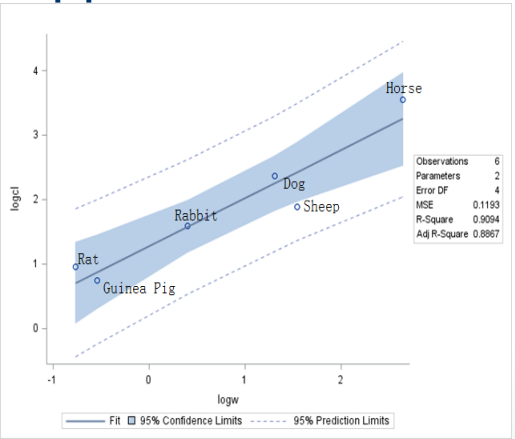


Decision by a “Committee”

Results of these analyses is a list of calculated withdrawal times using all data sources (including proprietary data released just for this use)

- Historical response (*has anything significantly changed–new data?*)
- Heuristic Half-Life Dose Multiplier algorithm based on approved withdrawal time adjusting for dose or disease
- Foreign approvals using MRL to TOL algorithm
- Interspecies extrapolations using allometry
- Population Pharmacokinetic estimates
- Real-time stochastic PBPK model estimate (RShiny)

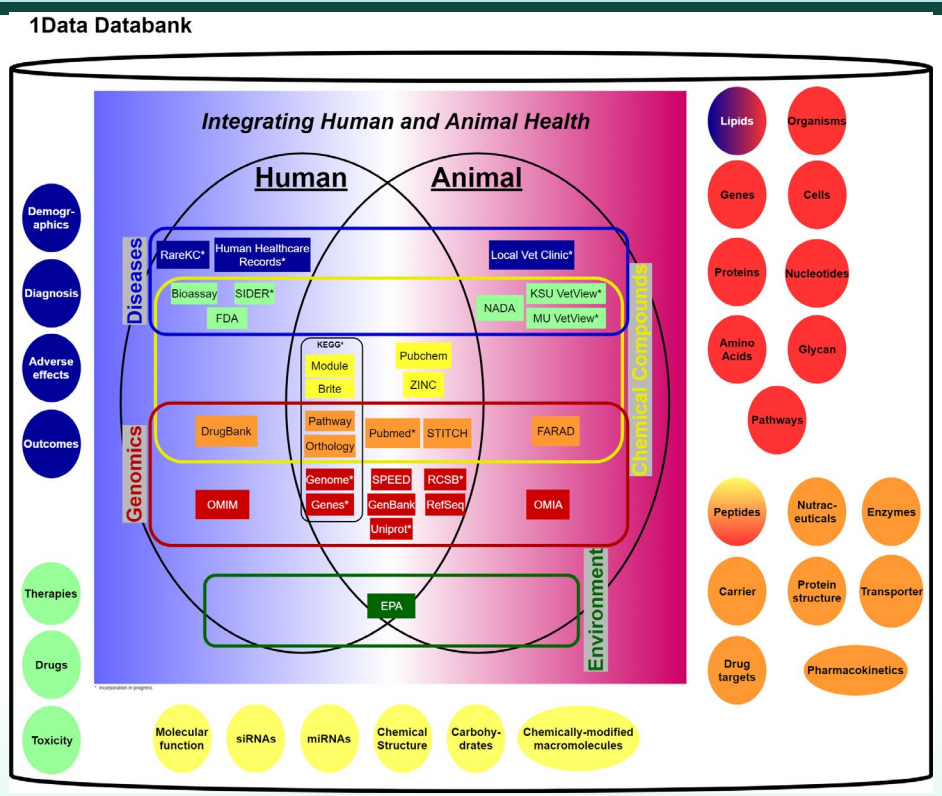
With enough experience, we can train AI to do this!

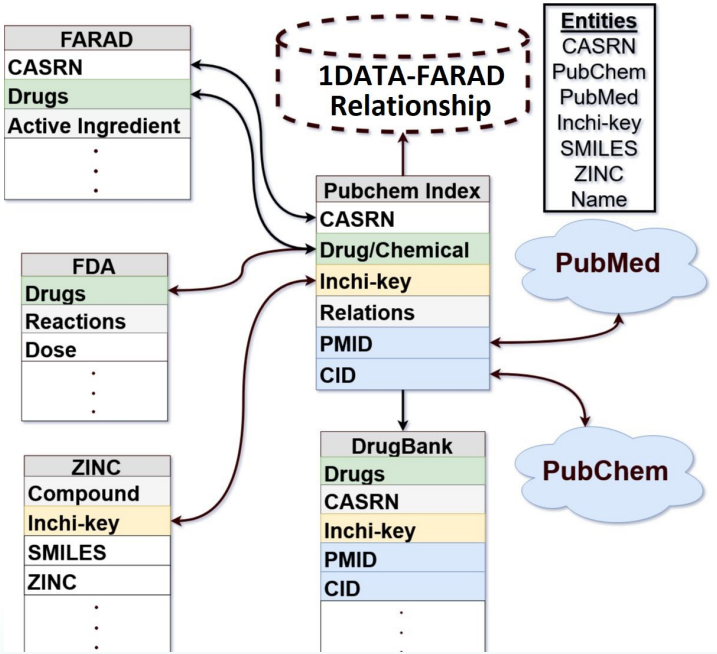


1DATA Program

- Collaboration between KSU and UMKC focused on integrating human and animal health data
- Concept is to create a curated database from disparate sources containing significantly different but overlapping data elements
  - Adverse Drug Reactions (ADR)
  - New Animal Drug Approval (NADA), Freedom of Information (FOI) requests and preclinical data from New Drug Applications (NDA)
  - FARAD
  - Health records (clinics, universities)
  - Genomic databases
  - Animal disease diagnostic laboratories

1DATA Comparative Integrated Database

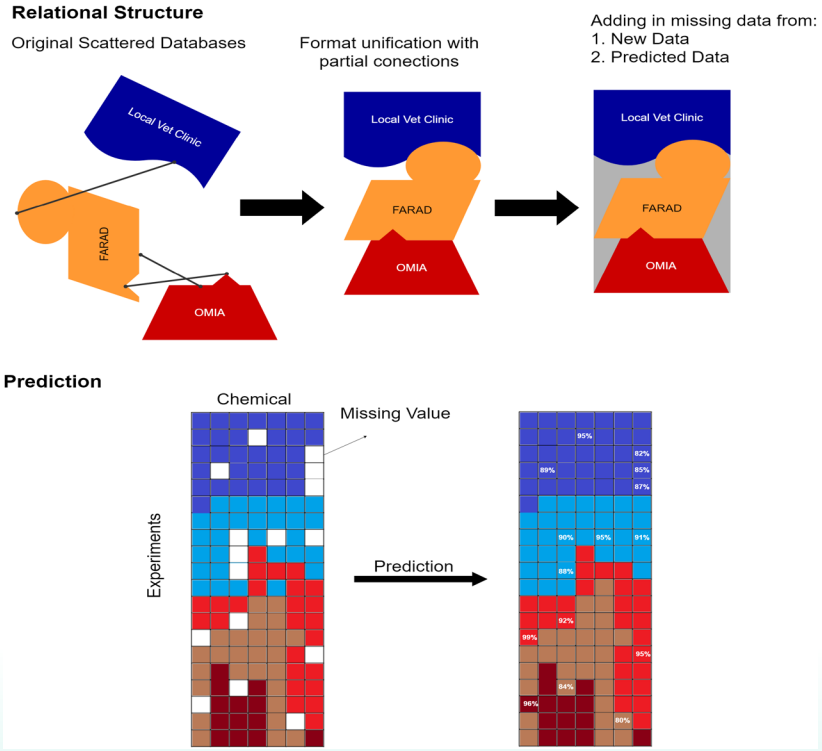




Data Curation and Correlating 1DATA Elements

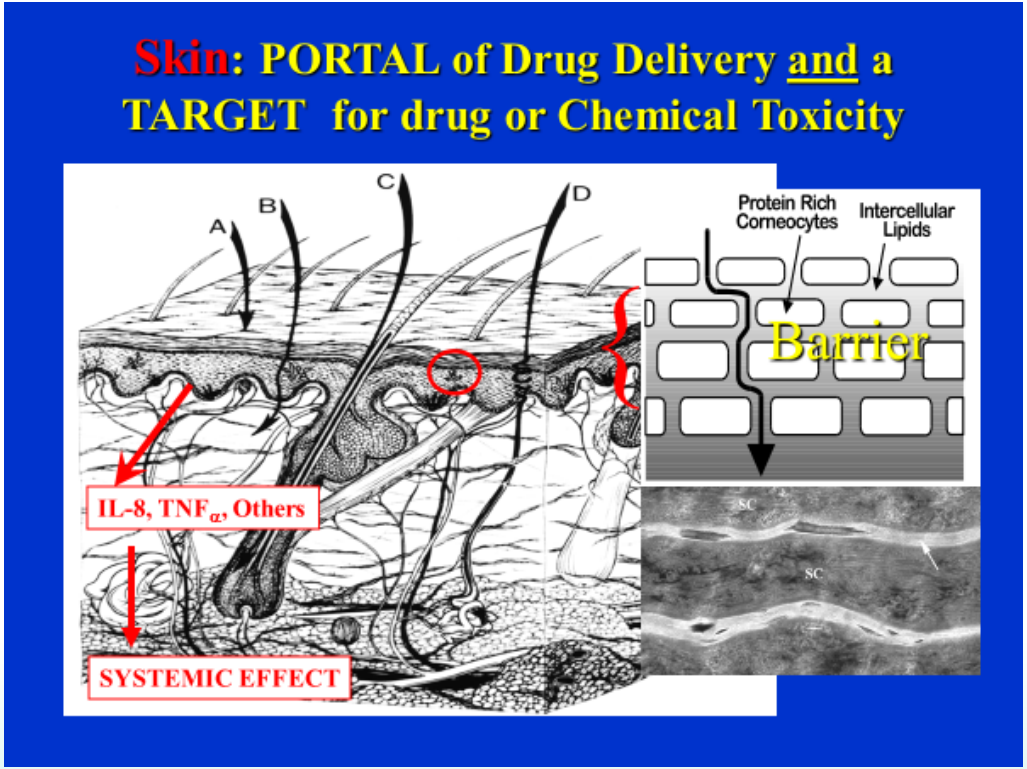
Data Curation

This step has proven essential in correlating data elements from different sources with similar names but very different meanings. Includes translating across foreign databases. Output can be a richer synthetic database to conduct more robust analyses. Recently used an AI tool called 1DrugAssist in this system to help identify what specific hypertension drugs might worsen COVID-19 prognosis based on an analysis of their ADEs.



Dermal Absorption–Cosmetic Safety

- Shift gears from food safety to cosmetics–dermal absorption
- Well-developed models often used to define structure activity modeling
- AI and ML techniques are starting to be used to analyze these types of data based on permeability (Kp)



Quantitative Structure Activity Analysis (QSAR)

- QSAR is a large area where AI is being applied. Fully discussed by later speakers
- QSAR applied to skin penetration and absorption is of concern to cosmetic safety [QSPR–Permeation vs. Activity]
- Attempt to correlate permeability through skin (Kp) to log $K_{oct/water}$ or other physical chemical properties. Extension of linear free energy relationships (LFER). First widespread use in skin was by Potts and Guy

$$\text{Log } K_p = 0.71 \text{ log } K_{o/w} - 0.0061 \text{ MW} - 6.3$$

$$\text{Log } K_p = 0.03MV - 1.7\Sigma\alpha_2^H - 3.9\Sigma\beta_2^H - 4.8$$

$$\text{Log } K_p = c + rR + s\pi + a\alpha + b\beta + vV$$

Account for Vehicle Effects in LFER

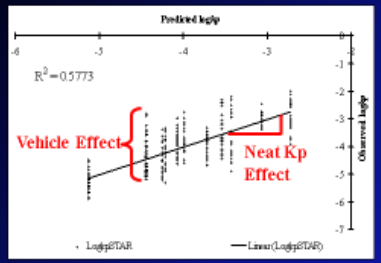
Account for Vehicle Effects in LFER

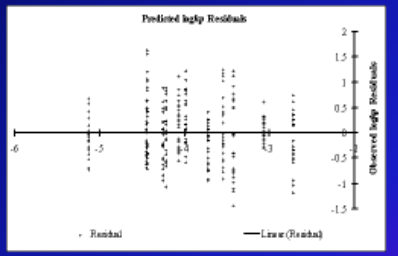
- Problem is that all this modelling work has been based on single penetrants in simple aqueous vehicles which is uncommon in many natural and experimental exposure scenarios. How do you account for component interactions?
- Assumption is that solvatochromatic interactions between a penetrant and other mixture-vehicle components would alter partitioning into stratum corneum and vehicle solubility and thus change the LFER relationship.
- Our laboratory has developed a method to quantitate this modulation using a *Mixture Factor (MF)* calculated from physical chemical properties of mixture components.

$$\text{Log Kp} = c + mMF + rR + s\pi + a\alpha + b\beta + vV$$

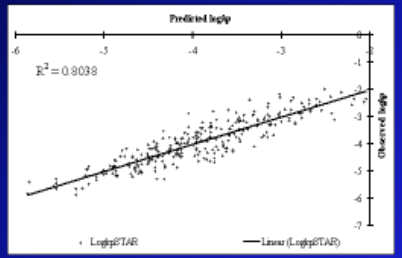
- Works with any type of LFER model (Potts and Guy, Abraham, etc.).

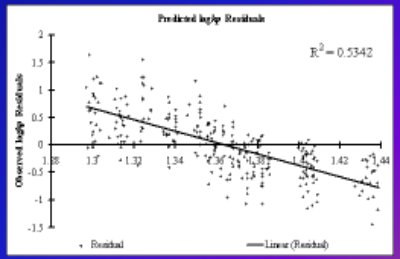
No Mixture Factor





Mixture Factor





Natural Compounds

Natural Compounds

- Having reviewed approaches used to assess mixture interactions for many compounds (e.g. drugs, pesticides, contaminants), we asked whether so-called "natural" ingredients can modulate dermal absorption of simultaneously dosed compounds.
- We are defining natural compounds as primarily herbal or plant products commonly applied to skin because of several purported beneficial and health promoting effects (e.g. exfoliants, antiaging creams, cleansers, soothing lotions, etc.).
- Many are components of globally marketed cosmeceuticals.
- Compounds are the source of many active pharmaceutical dermal penetration enhancers (e.g. terpenes) ∴ expect some effects.
- Obtained from a wide range of fruits, vegetables and plants.
- Not the purpose of this work to assess efficacy of compounds.
- Concern is whether use of such products modulate absorption of simultaneously applied topical drugs.

Mixture Interaction Models

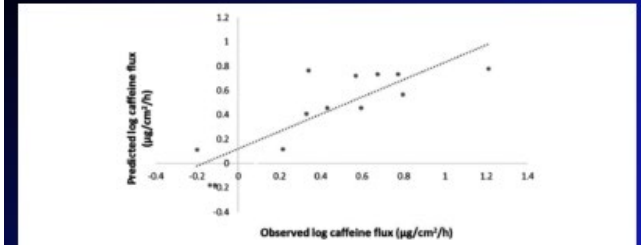


Figure 4. Observed versus Predicted Log Caffeine Flux. A plot illustrating the relationship between observed versus predicted log caffeine flux with QSPR model using two molecular descriptors PSA and log P ($r^2 = 0.71$).

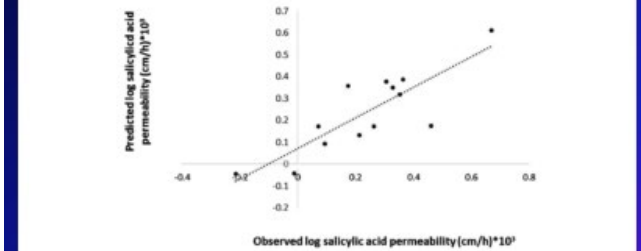


Figure 5. Observed versus Predicted Log Salicylic Acid Permeability. A plot illustrating the relationship between observed versus predicted log Salicylic acid permeability with QSPR model using three molecular descriptors HK, FRB, and WS ($r^2 = 0.70$).

Mixture Factors
PSA: Polar Surface Area
Log P: Octanol/Water PC

Mixture Factors
HK: Henry constant
FRB: # Freely Rotating Bonds
WS: Water Solubility

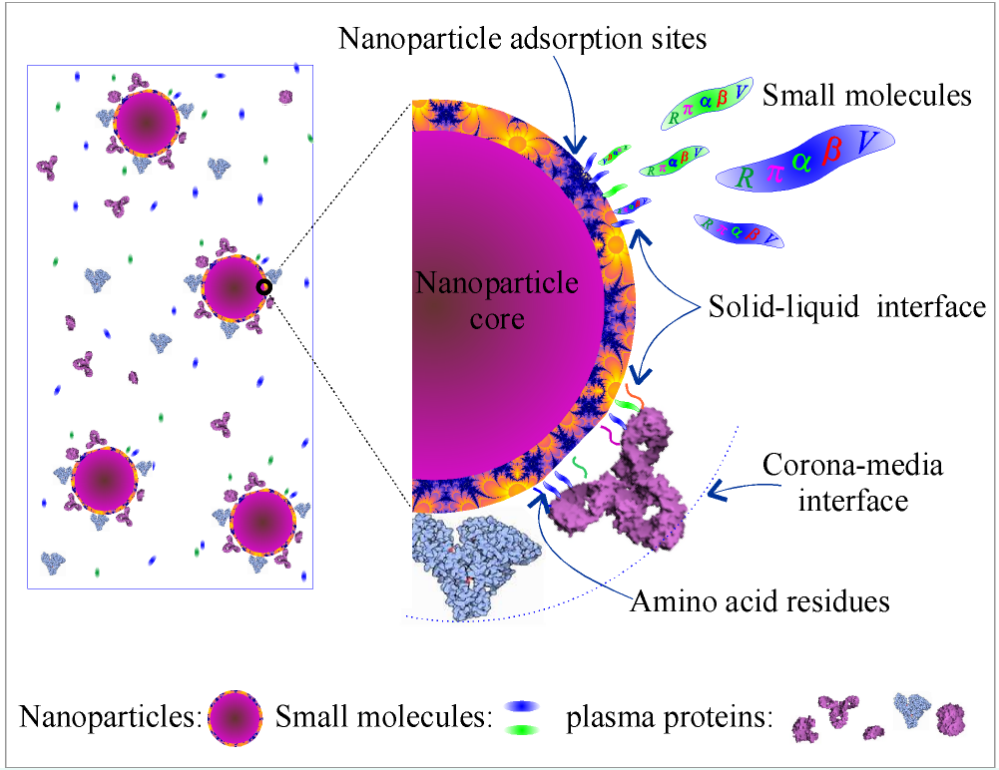
Biological Surface Adsorption Index (BSAI)
A Molecular Signature for NP Interactions

We have worked on developing metrics that could be used to characterize nanoparticles.

Our approach has been to probe the surface of NP using LFER approaches.

Such parameters are useful as input into AI classification algorithms to model disparate types of nanoparticles.

Xia et al. *Nature Nanotech* 5: 671-675, 2010



BSAI Statistics

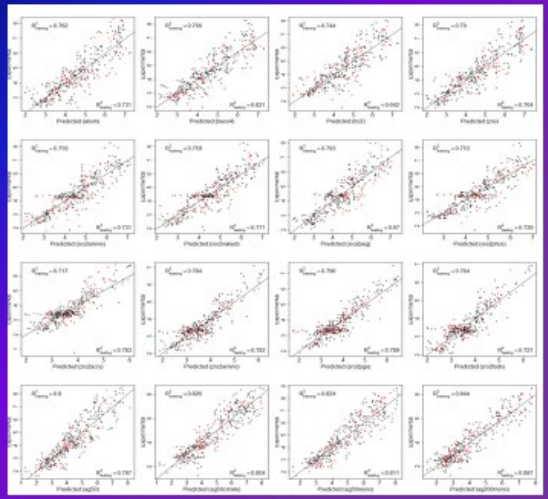
Multiple Regression Analysis

- Experimentally generate adsorption coefficients (K_p) for 28 molecules with different physical chemical properties
- Simultaneously solve 28 linear free energy relationships (LFER) for optimal fit of the 5 molecular descriptors
- Do this for each NP with appropriate replicates

$$\text{Log } K_p\,(1) = c + rR + s\pi + a\alpha + b\beta + vV$$
$$\text{Log } K_p\,(2) = c + rR + s\pi + a\alpha + b\beta + vV$$
$$\text{Log } K_p\,(3) = c + rR + s\pi + a\alpha + b\beta + vV$$

$$\text{Log } K_p\,(27) = c + rR + s\pi + a\alpha + b\beta + vV$$
$$\text{Log } K_p\,(28) = c + rR + s\pi + a\alpha + b\beta + vV$$

BSAI Model Fits and Validation

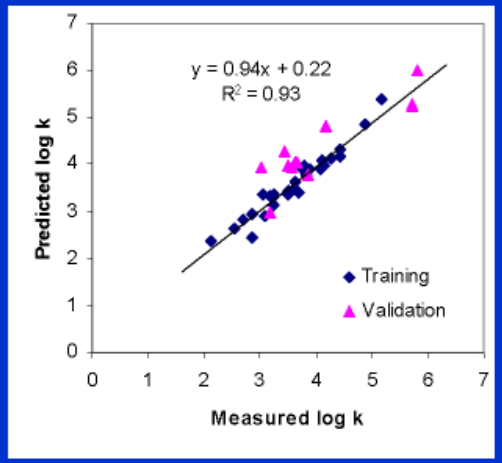


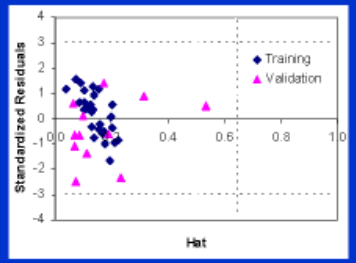
Demonstrates ability to predict pesticide absorption

Chen, Zhang, Monteiro-Riviere, Riviere. *Nanotoxicology*. 10: 1118-1128, 2016

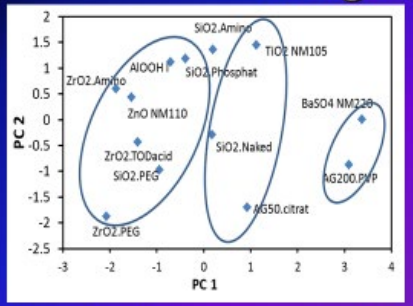
Data Analysis Approach

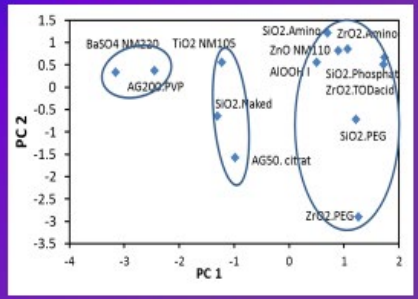
Nano-Descriptors Provide Better Prediction





Langmuir Model





- Principal components of the five nanodescriptors obtained from multivariate linear regression:
  - (left), original BSAI model,
  - (right), low conc approximation by Langmuir model
- Identical categorization, but tighter differentiation
- Eliminates error inherent to assaying adsorption at low concentrations

Take Home Lessons From These Projects

- Use models with different assumptions to ensure full outcome space is captured
- Different models have different built-in biases and sensitivity to errors
- Use structured/mechanistic models to define input parameters to focus analysis
  - Preprocess raw input data using mechanistic models (Langmuir example) or equations and eliminate errors
- More data, and more diverse the data, better the inference
  - Food Safety: Melanine–Never included in any databases before incident
- Need to continuously update datasets, find new sources of data as more become available in the cloud, and retrain and modify models if necessary
- Developed strategy to use and credit individual data elements from private sources that are incorporated into large analysis
  - Use a blockchain to tag individual data elements

What Is AI? Plethora of Definitions

- *Science and engineering of making intelligent machines, especially intelligent computer programs, some capable of learning and problem solving*
- *A branch of computer science that studies the properties of intelligence by synthesizing intelligence*
  - Subcomponents of AI are ML and within this, deep learning
  - Supervised vs. unsupervised learning
  - Neural nets, back propagation, deep reinforcement, random forest, Bayesian hyperparameter optimization ----- *All these are specific techniques*
  - Narrow AI versus General AI ----- *Stuff of superintelligence and the Singularity*
    - *Narrow is for a specific domain while General is domain and task independent*

AI Implies More Than ML

Confluence of multiple “computer-digital” developments facilitated by exponential increases in power and decreases in size and cost of computer hardware, resulting in:

- ***Elastic cloud computing***
- ***Big Data–High volume, velocity and variety***
- ***AI and Deep Learning***
- ***Internet of Things (IoT)*** *{think RFID, sensors, etc.}*

  (Thomas Siebel Digital Transformation 2019)
- ***Digitization of everything***
- ***Blockchain***

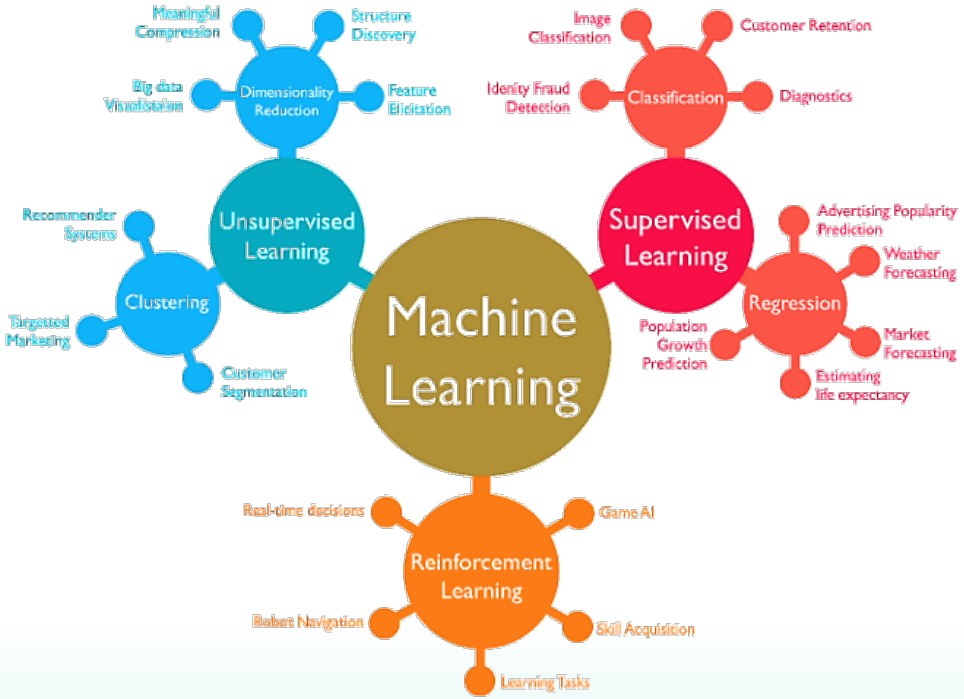
They can then be integrated into complex systems for specific purposes

AI and ML Cover a Universe of Modeling Approaches

Machine-Intelligence Hierarchy

0. Basic automation
1. Rule-based systems
2. Supervised learning
3. Unsupervised learning
4. Multiagent interactions
5. Creative AI
6. Aspirational (general) AI

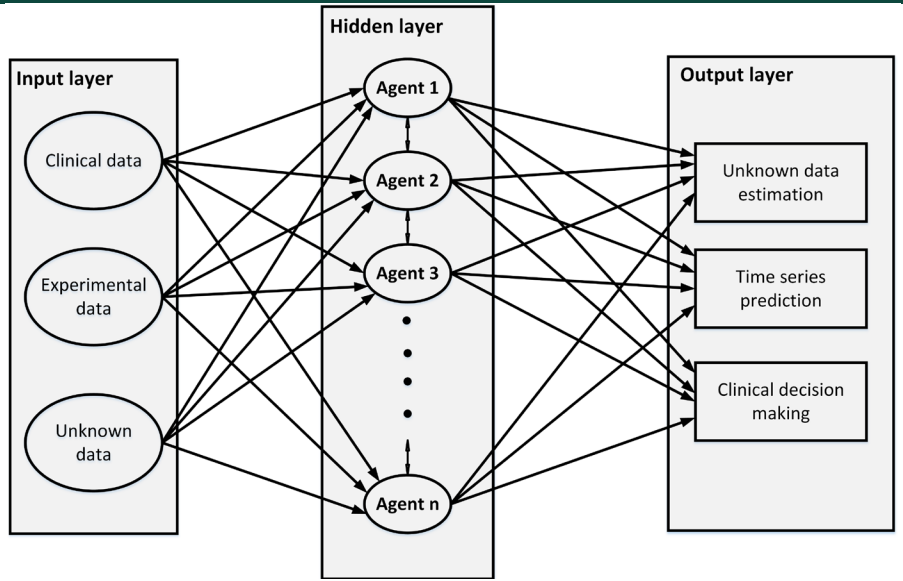
(Denning and Lewis. *American Scientist* 107: 346-349, 2019)

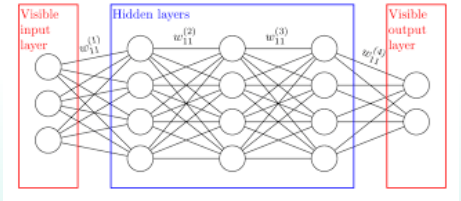


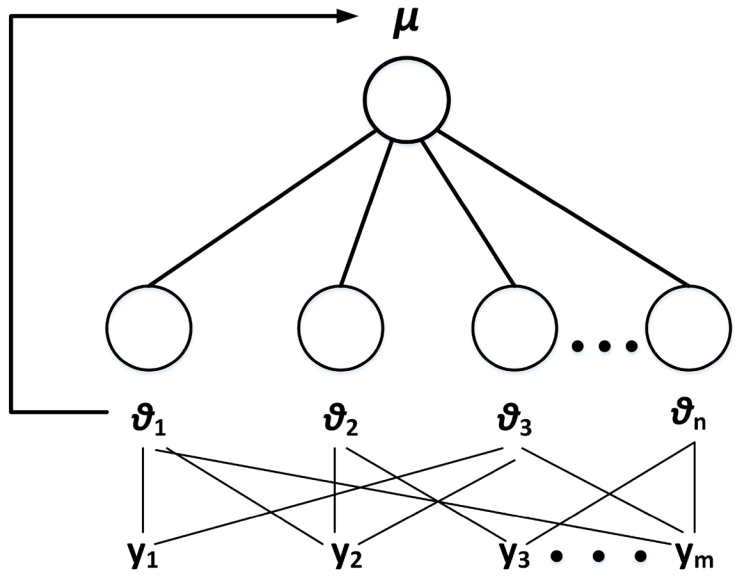
Typical Study Design

1. Assemble and organize/prepare data into proper format for analysis
2. Identify features which will be identified as input
3. Label outcomes being modeled as output
4. Identify training sets
5. Select and train algorithm
6. Improve labels, refine and tune features, rerun analysis
7. Validate with separate data set
8. Deploy

Neural Nets and Bayesian Optimization







Expert, Rule-Based, and Logic-Based Systems

- System is a set of rules, defined by “knowledge/domain” experts
  - Grammar syntax rules for translation
  - Signs and symptoms for possible diagnosis
- Early examples were in diagnosis and language translation
  - Translation now uses Natural Language Processing (NLP) systems
- Basis of AI’s original success in chess. All are rule based games
  - However, deep learning AI neural net systems (Alpha Zero) beat these rule-based systems yet are still doing so in a rule-defined environment
- Logic-Based algorithms require defined input and domain specific algorithms
- Programmer defines the relationships

Image Analysis Systems

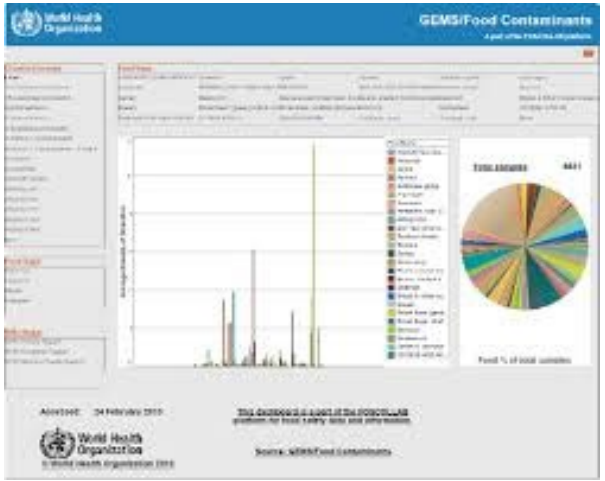
- Example of pattern recognition which is major focus of AI research
- One of major applications in consumer products, self-driving cars, facial recognition, and identifying cats and dogs!
- Use “human identified” images (use “Mechanical Turks”) as input to train networks. Algorithms actually focus on small set of data elements assuming input images are from a certain class (e.g., faces)
- Needs very large and diverse datasets
- Employed in surveillance systems
- “Tricked” by manipulating identifying elements–adversarial examples
  - Classic examples of complex dot patterns fool deep neural networks, glasses

Classification Systems

- Define associations between users and interests
- Designed to identify clusters of similar objects and detect anomalies
- Functionality is based on strengths of Big Data
  - Increase granularity–better the predictions
- Basis of Facebook, Amazon, and Google business models
  - You search sports sites, maybe you want to buy a ball or travel to a game
- Good strategy to associate risks and causes of biological or chemical contamination in a food system
- Basis of some QSAR systems which cluster chemicals based on physical chemical properties and toxicological outcomes

Big Data in Food Safety

- Primarily related to on-line databases containing food safety relevant data (Marvin et al.)
- WHO FOSCOLLAB food safety platform. Integration of multiple food safety related information sources (accessed by Dashboards based on subject domain–contaminants, pesticides, diets)
- Potential to link scientific and legal endpoints, pathogen genomic databases, monitoring data, social media alerts, etc.



Marvin et al. Big Data in Food Safety. *Crit Rev Food Sci Nutrition* 57: 2286-2295, 2017.

Big Data in Food Safety

- Major AI efforts are in visualization and geospatial linkages
- Used to map out and track food supply chain and spread of pathogens responsible for food safety outbreaks, detect food fraud
- Supply chain systems are in place that integrate GPS, sensors (T^o), and RFID systems
  - Cheesecake Factory with IBM Big Data Analytics
  - WalMart's SPARK
- Food companies are using AI to develop new recipes and flavors

AI at the FDA

- CDRH published regulatory framework for AI in medical devices
- Considers software as a medical device
- Pilot project using blockchains for food traceability
- AI used in skin lesion identification/classification by dermatologists and diagnosis of mammary cancer
- Emerging Chemical Hazard Intelligence Platform (ECHIP)
  - Scans news, social media, and scientific literature for signals of interest to CFSAN ultimately allowing issues to be identified in hours rather than days
- Surveillance, monitoring, and epidemiological investigations linked to genomic and pathogen data
  - NCBI-WGS databases, GenomeGraphR, Genome Trakr, PulseNet,

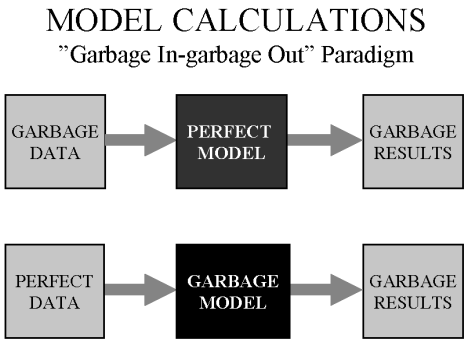
AI at the FDA

- Text mining tool to scan inspector notes
- CDER post-market surveillance system Sentinel
- FDA Adverse Event Reporting System (FAERS) uses NLP and ML tools
- NCTR, CFSAN, and ORA developing image analysis systems to identify insect pests
- Program to study drug-induced liver injury

Potential Weaknesses of AI and ML

- Modern AI does not conduct true statistical inference from small datasets. Instead it relies on assumed completeness of data and an absence of sampling error since entire inference space is studied
- Data thus needs to be representative of the inference population
- Linking output of one complex system as input to another complex system can be fraught with issues of complexity and chaos
  - Strange attractors in a non-linear network
- Many systems remain challenged by complex stochastic environments that define the real-world applications
- Large networks are “brittle,” resulting in instability in certain conditions.
  - Unanticipated input or changed environment from that originally modeled

Specific Problems



- Data entry and sensor errors
- Misspecification or errors in input data
- Incomplete or unbalanced data
  - Unknown co-factors (important in food and cosmetics)
  - Mismatch of training and validation datasets with the real inference space
  - New toxin–e.g., melamine (although AI could actually identify an outlier faster)
- Malicious intervention–Hacking
- Computer/network/sensor malfunctions

Potential Weaknesses of AI and ML

- One problem is sociological: scientists and engineers are trained by classic definition of modeling as presented earlier and are simply bothered by lack of explanatory details
- “Hidden Layers” in a deep neural net are hidden
  - Some visualization approaches have been developed to assess state of a neural net
- May tolerate errors in home-based systems (Alexa, Siri voice commands; lost robotic vacuums), but how does one guarantee such errors do not propagate in a food system with real consequences?

If we don’t understand an AI system, how can we trust it?

Colloquium Overview

- Dr. Steve Bennett has given you a feel for AI and its applications
- I have provided a brief introduction on how AI has evolved in modeling and data managements projects and how it has been applied to some domains within CFSAN's purview
- Next three speakers will:
  1. Illustrate its applications in a food cleaning system (Nik Watson)
  2. Use AI to extend QSAR models (Joe Zhang)
  3. Use ML tools for safety determination of cosmetics and cosmetic ingredients (Tim Allen)

Suggested Readings

Riviere JE, Brooks JD: Predicting skin permeability from complex chemical mixtures. *Toxicol. Appl. Pharmacol*. 208: 99-110, 2005.

Xia XR, Monteiro-Riviere NA, Riviere JE: An index for characterization of nanomaterials in biological systems. *Nature Nanotech*. 5: 671-675, 2010.

Chen R, Zhang Y, Sahneh F, Scoglio CM, Wohlleben W, Haase A, Monteiro-Riviere NA, Riviere JE. Nanoparticle surface characterization and clustering through concentration-dependent surface adsorption modeling. *ACS Nano* 8: 9446-9456, 2014.

Baynes RE, Riviere JE (eds): *Strategies for Reducing Drug and Chemical Residues in Food Animals: International Approaches to Residue Avoidance, Management and Testing*. New York: John Wiley, 2014

Huang Q, Gehring R, Tell LA, Li M, Lin Z, Riviere JE. Interspecies allometric meta-analysis of the comparative pharmacokinetics of 85 drugs across veterinary and laboratory animal species. *J. Vet. Pharmacol. Therap*. 38: 214-226, 2015.

Riviere JE, Gabrielsson J, Fink M, Mochel J. Mathematical modeling and simulation in animal health. I. Moving beyond pharmacokinetics. *J. Vet. Pharmacol. Ther*. 39: 213-223, 2016.

Lin Z, Monteiro-Riviere NA, Kanan R, Riviere JE. A computational framework for interspecies pharmacokinetics, exposure and risk assessment of gold nanoparticles. *Nanomedicine (London)* 11: 107-119, 2016.

Suggested Readings

Marvin HJP, Janssen EM, Bouzembrak Y, Hendriksen PJM, and Staats M: Big data in food safety: An overview. *Crit Rev Food Sci Nutrition* 57: 2286-2295, 2017.

Riviere JE, Tell LA, Vickroy T, Baynes RE, Gehring R. FARAD Digest: A user's guide to FARAD resources: Historical and future perspectives. *J. Am. Vet. Med. Assoc*. 250: 1131-1139, 2017.

Li M, Gehring R, Riviere JE, Lin Z. Probabilistic physiologically based pharmacokinetic model for penicillin G in milk from dairy cows following intramammary or intramuscular administrations*. Toxicol Sci.* 164:85-100. 2018.

Li M, Cheng YH, Chittenden JT, Baynes RE, Tell LA, Davis JL, Vickroy TW, Riviere JE, Lin Z..Integration of Food Animal Residue Avoidance Databank (FARAD) empirical methods for drug withdrawal interval determination with a mechanistic population-based interactive physiologically based pharmacokinetic (iPBPK) modeling platform: example for flunixin meglumine administration. Arch. Toxicol. 93:1865-1880. 2019. Siebel T*: Digital Transformation*. Rosetta Books, New York, 2019.

Xu X, Mazloom R, Goligerdian A, Stahley J, Amini M, Wyckoff GJ, Riviere JE, Jaberi-Douraki M. Making sense of pharmacovigilance and drug adverse event reporting: Comparative similarity association analysis using AI machine learning algorithms in dogs and cats. Topics Companion Anim Med. 37: 100366, 2019.

Acknowledgements

- 1DATA (Professors Majid Jaberi-Douraki and Gerry Wycoff) and FARAD colleagues (Professors Art Craigmill, Steve Sundlof, Alistair Web, Lisa Tell, Tom Vikroy, Jennifer Davis, Zhoumeg Lin, and Ron Baynes) who have done all the work
- Continuous support from USDA NIFA and NIH, AFOSR, and EPA in support of research presented