Estimation of Melting Points of Brominated and Chlorinated Organic Pollutants using QSAR Techniques

By: Marquita Watkins

Persistent Organic Pollutants

- Do not undergo photolytic, biological, and chemical degradation.
  - low water solubility
  - high lipid solubility
  - semi-volatility
  - high molecular masses
- Found in pesticides
  - Transport by air and water
- Found at extremely low levels in 20% of US Food Supply
  - Continual exposure toxicity and effects are unknown

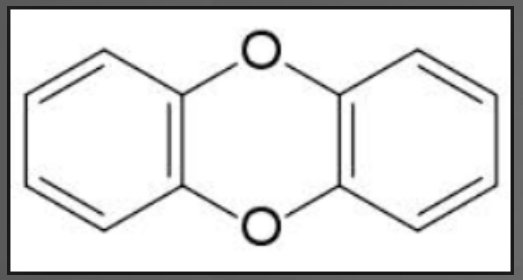
Previous Research

- Reserved to small sets (10 to 100 compounds) of hydrocarbons, substituted aromatics, aldehydes, amines, and ketones utilizing other descriptor types, regressions, and analysis techniques.
- Larger data sets were only of approximately 400 compounds.

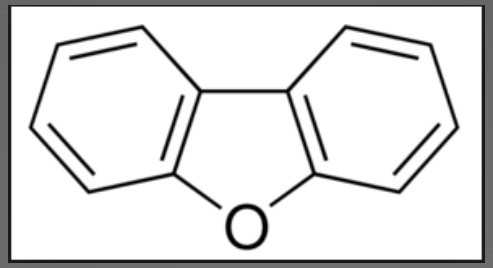
The Endeavor

- To estimate the melting points of 1436 chloro- and bromo-analogues of dibenzo-p-dioxins, dibenzofurans, biphenyls, naphthalenes, diphenyl ethers, and benzenes by utilizing quantitative structure—property relationship (QSAR) techniques.

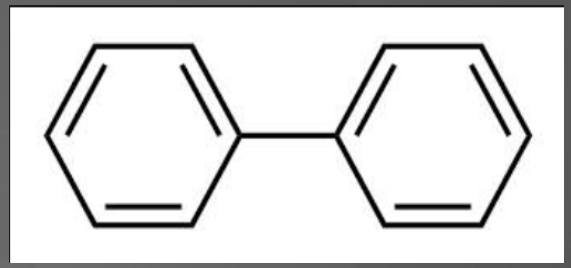
Chosen Structure Families



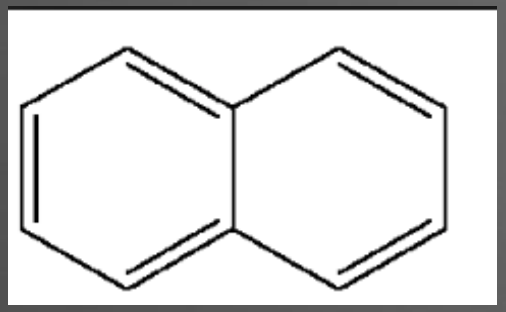
dibenzo-p-dioxins



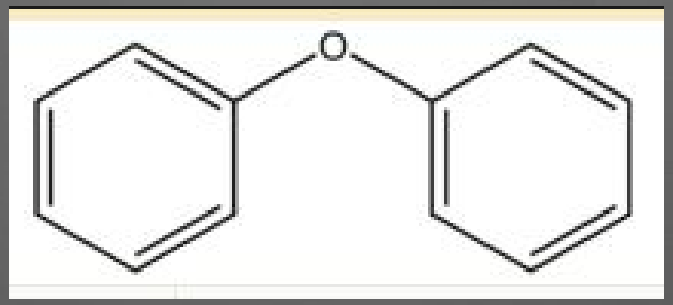
dibenzofurans



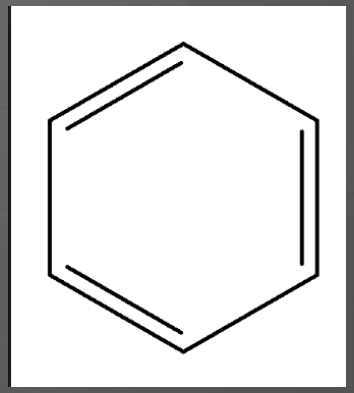
biphenyls



naphthalenes



diphenyl ethers



benzenes

Melting Point

- Specifies transition temperature
- Essential in biochemical and environmental studies
- Solubility estimation
- Difficult QSAR endpoint to predict

QSAR Significance

- Inexpensive
- Time-efficient
- Safe
- Enhances the efficiency of drug design

Basic concept of QSAR modeling

Endpoint
(experimentally measured)

y

- *Activity (EC_{50})*
- *Phys/Chem property (K_{OW}, $t_{1/2}$)*
- *Retention parameters (t_R)*
- *Toxicity (LD_{50}, LC_{50})*
- *…*

QSAR model

y = f(X)
(*eg.* $y = b_0+b_1x_1+b_2x_2$)

- *Linear Regression (LR)*
- *Multiple Linear Regression (MLR)*
- *Partial Least Squares (PLS)*
- *Artificial Neural Networks (ANN)*
- *…*

Structural descriptors

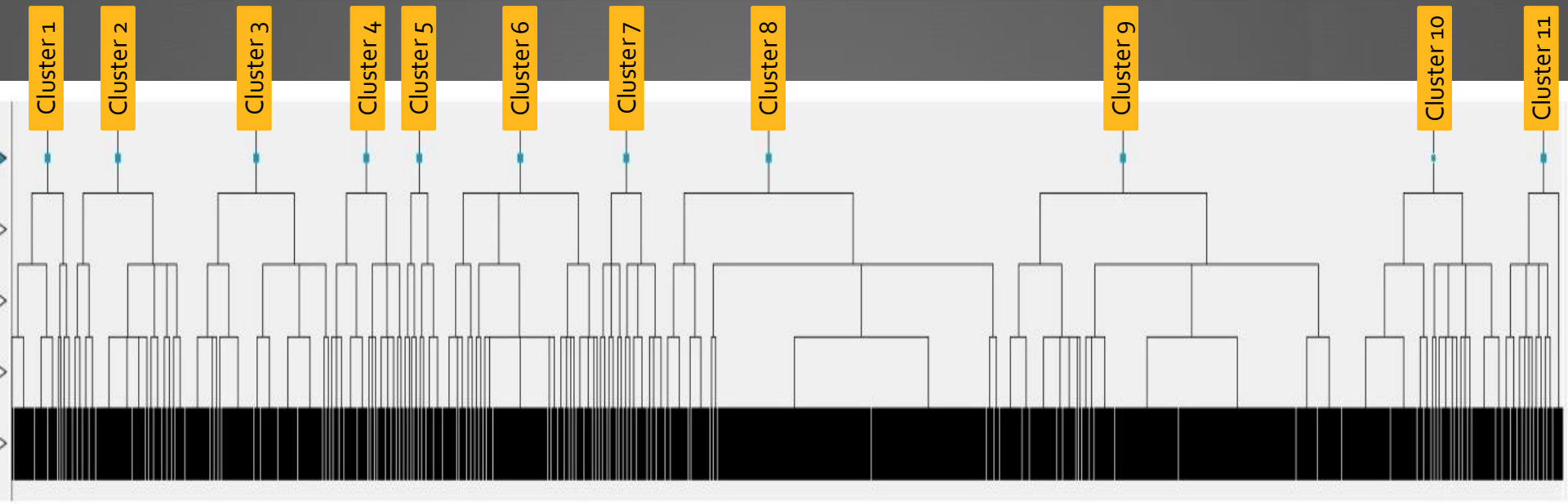
X

- *Dipole moment*
- *Polarizability*
- *HOMO, LUMO*
- *Topological indexes*
- *Number of specific atoms/groups*
- *…*

Software Employed

- Structures were built and optimized using HyperChem 5.0
- Hierarchal clustering in ChemAxon's MCS Library Software.
- Descriptor Generation Software
  - DRAGON5
- PLS Toolbox
  - Automated Descriptor Selection
    - Genetic Algorithm
  - Model Building

Hierarchal Cluster Results



Minimal MCS Size = 25
MCS Mode = Very Fast
Matching Parameters = Atom Type, Bond Type, Charge
Clustering Options = Keep Rings
Required Cluster Count =15
Maximum Level Count = 15

Cluster Analysis

Cluster #	Calibration Set	Validation Set	Prediction Set
1	**R^2 Cal:** 0.89	**R^2 CV:** 0.88	**R^2 Pred:** 0.80
	Bias:0	**CV Bias:**0.004	**Pred Bias:** -1.07
	RMSEC: 8.04	**RMSECV:** 8.43	**RMSEP:** 13.24
3	**R^2 Cal:** 0.76	**R^2 CV:** 0.72	**R^2 Pred:** 0.66
	Bias: -4.26e-14	**CV Bias:** -0.20	**Pred Bias:** 5.46
	RMSEC: 21.66	**RMSECV:** 23.64	**RMSEP:** 23.72
4	**R^2 Cal:** 0.89	**R^2 CV:** 0.86	**R^2 Pred:** 0.91
	Bias: -2.84e-14	**CV Bias:** -0.13	**Pred Bias:** -5.29
	RMSEC: 17.32	**RMSECV:** 19.30	**RMSEP:** 15.71
7	**R^2 Cal:** 0.94	**R^2 CV:** 0.80	**R^2 Pred:** 0.63
	Bias: -2.84e-14	**CV Bias:** -5.24	**Pred Bias:** -3.82
	RMSEC: 18.56	**RMSECV:** 34.79	**RMSEP:** 26.26

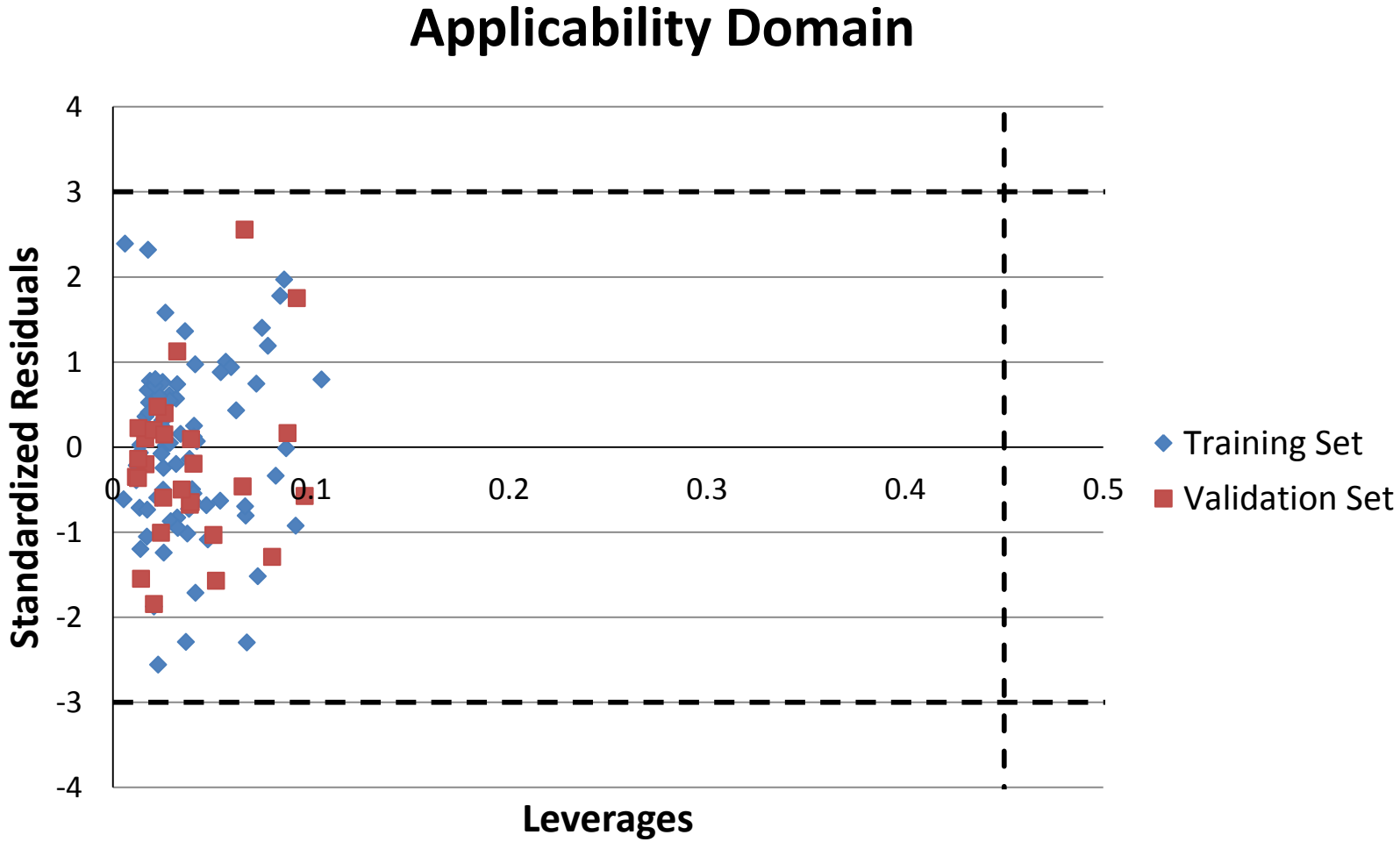
Cluster Analysis

Cluster #	Calibration Set	Validation Set	Prediction Set
8	**R^2 Cal:** 0.75	**R^2 CV:** 0.56	**R^2 Pred:** 0.81
	Bias: 4.26e-14	**CV Bias:** -1.19035	**Pred Bias:** 13.38
	RMSEC: 21.12	**RMSECV:** 28.5243	**RMSEP:** 34.0649
9	**R^2 Cal:** 0.84	**R^2 CV:** 0.54	**R^2 Pred:** 0.83
	Bias: 0	**CV Bias:** 2.41	**Pred Bias:** -13.43
	RMSEC: 35.51	**RMSECV:** 67.81	**RMSEP:** 44.05
10	**R^2 Cal:** 0.85	**R^2 CV:** 0.791392	**R^2 Pred:** 0.709519
	Bias: 5.68e-14	**CV Bias:** 1.66	**Pred Bias:** 6.91
	RMSEC: 22.39	**RMSECV:** 27.13	**RMSEP:** 56.34
11	**R^2 Cal:** 0.77	**R^2 CV:** 0.73	**R^2 Pred:** 0.80
	Bias: 0	**CV Bias:** -0.14	**Pred Bias:** -0.059
	RMSEC: 26.61	**RMSECV:** 29.19	**RMSEP:** 25.52

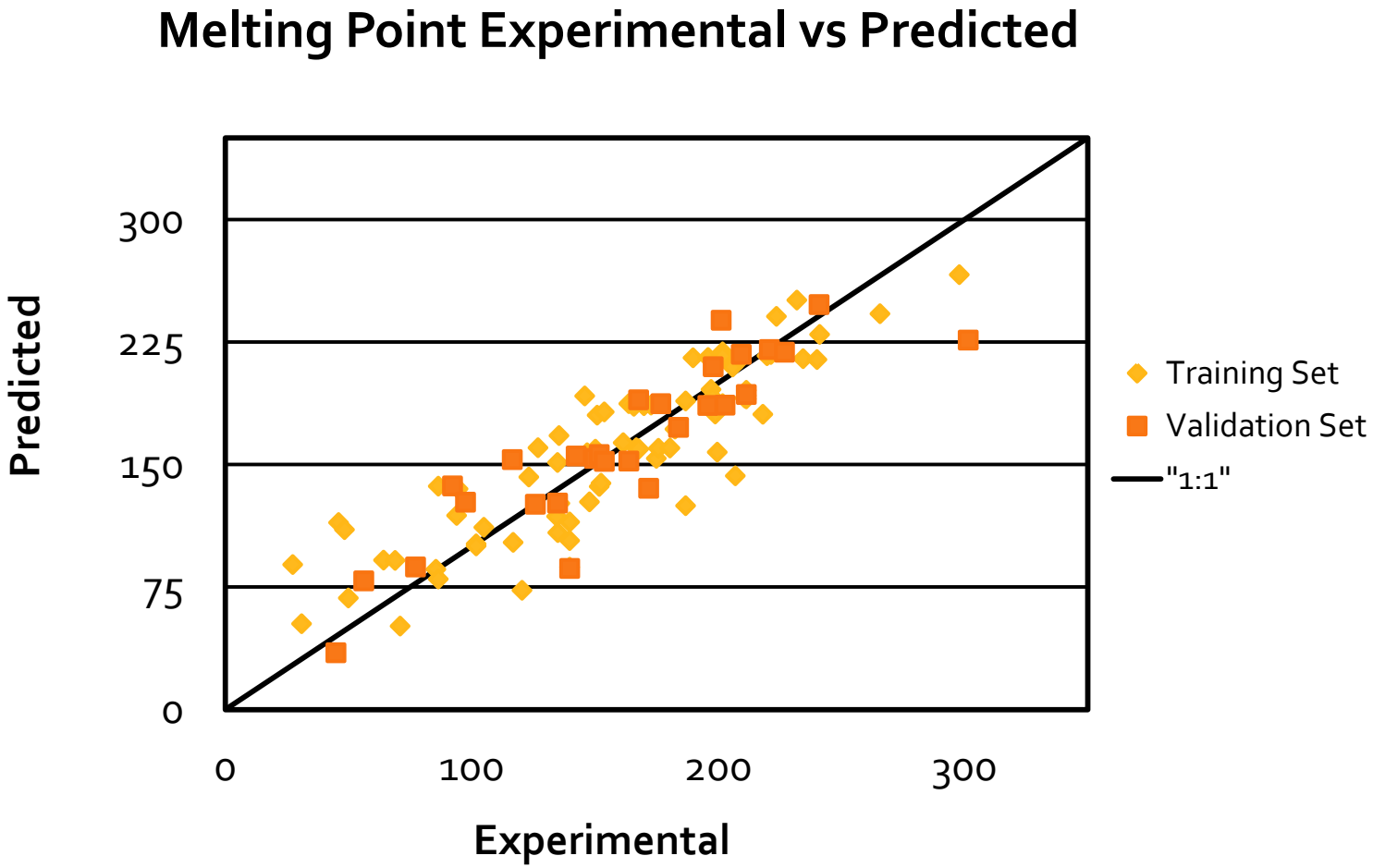
Cluster 11 Results/Discussion

- Algorithm
  - MP=0.213208(D/D) + 0.011536(Wap) + 101.442(J) + 2.9733(S1K) + 0.093732(TPC) + 0.000737(pilD) + -0.238566(PCR) + 0.063346(PCD) + 1.99058(MATS8m) + 161.158(MATS1v) + 46.4475(MATS2v) + 67.106(MATS3v) + -194.013
- Data Characteristics
  - N_{trn}=80 N_{val}=28 N_{pred}=386

Cluster 11 Validation



Cluster 11 Validation



Missing Clusters

- Cluster 2: still being validated and analyzed
- Cluster 5: has been validated and results are currently being compiled
- Cluster 6: shown to have a poor correlation in multiple aspects
  - Possibly due to data set
  - Endpoint more complex than originally believed

Conclusions

- The melting points of the series of POPs have been modeled using a set of calculated descriptors
  - Suggests that existing cheminformatics descriptors fail to adequately describe interactions in the crystalline solid phase
    - May be a significant cause of error in melting point prediction
- Original data set contains too many similar compounds to produce a valid model
  - Hierarchical clustering method employed via MCS Library
    - Improved predictions significantly

Future Endeavors

- Explore other possible methodology and descriptors
  - Nonlinear methods
  - Other descriptors - for example, quantum-chemical descriptors
- Model other endpoints (biological activity, solubility, etc.) with these data sets

Acknowledgements

- Dr. Jerzy Leszczynski
- Dr. Bakhtiyor Rasulev
- Dr. Tomasz Puzyn and group
- Dr. Hill and group
- NSF-CREST

References

- Czub, G., & Mclachlan, M. S. (2004). Bioaccumulation Potential of Persistent Organic Chemicals in Humans. *Environmental Science and Technology , 38* (8), 2406-2412.
- Katritzky, A. R., Maran, U., Karelson, M., & Lobanov, V. (1997, April 25). Prediction of Melting Points for the Substituted Benzenes: A QSPR Approach. *Advance ACS Abstracts* .
- Ritter, L., Solomon, K. R., Forget, J., Stemeroff, M., & O' Leary, C. (1995, December 1). *A Review Of Selected Persitent Organic Pollutants.* Retrieved Aug 29, 2013, from World Health Organization: http://www.who.int/ipcs/assessment/en/pcs_95_39_2004_05_13.pdf
- Tomasz Puzyn, Aleksandra Mostrag-Szlichtyng, Agnieszka Gajewicz, Michal Skrzynski, and Andrew P. Worth. "Investigating the influence of data splitting on the predictive abilkity of QSAR/QSPR moldels." *Structural Chemistry* 22 (February 2011): 795-804.
- Leszczynski, J. (2013). Presentation of Toxicity of Nanomaterials – Development Of New Theoretical Approaches at the JSU Interdisciplinary Nanotoxicity Center.

References

- Katritzky, A.R., Ritu, J., Lomaka, Petrukhin R., Maran U., and Karelson, M (2001). Perspective on the Relationship between Melting Points and Chemical Structure. *Crystal Growth, and Design.* 1(4), 261-265.
- Katritzky, A.R., Maran U., Victor S., and Karelson, M (2000). Property Relationship Correlations of Technologically Relevant Physical Properties. *J. Chem. Inf. Comput. Sci.* 1(4), 261-265.
- Nilcolopowlou-Stamati and Pitsos, M.A. The Impact of Endocrine Disrupters on the Female Reproductive System. *Human Reproduction Update.* 2001. Vol 7(3) 323-330.
- Damstra T, Page S.W., Herrman J.L., Merideth T. *J Epidemiol Community Health.* 2002; 56: 824-825

Questions???