
Big Data: Exploring Possibilities to Improve Predictive Exposure Capabilities



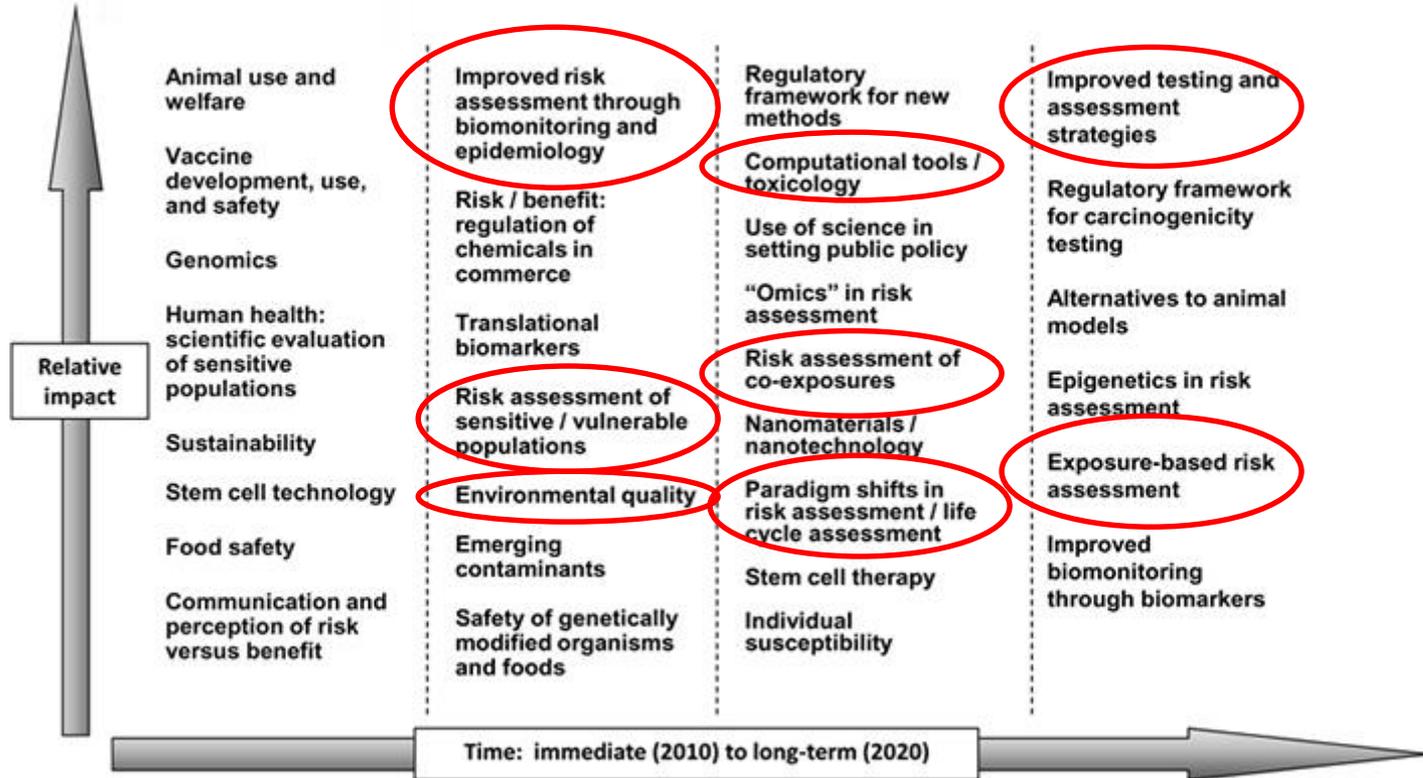
**R. Zaleski, T. Buckley,
H. Zenick**

**HESI Emerging Issues Session
June 2015, Washington, D.C.**

ILSI Health and
Environmental Sciences
Institute



2010-2020 HESI COMBINED CHALLENGES MAP



Each axis appearing on the 2010-2020 HESI Combined Challenges Map is a continuum. All issues on the map are of high importance/impact based on prioritization by the participants in the 2009 HESI mapping exercise. "Relative impact" is a qualitative measure of importance among high priority topics. The location of issues along the "time" continuum is an approximation of when the topic is likely to become a major issue in the timeframe from 2010 to 2020.

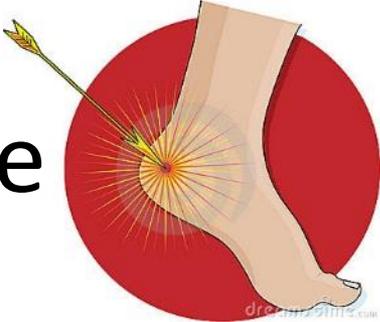
Why Exposure Science?



No exposure, no risk – period.

*International Society of Exposure Science Board
Letter to Science (2011) citation of NRC (1983).*

Risk = Fn (Hazard, Exposure)

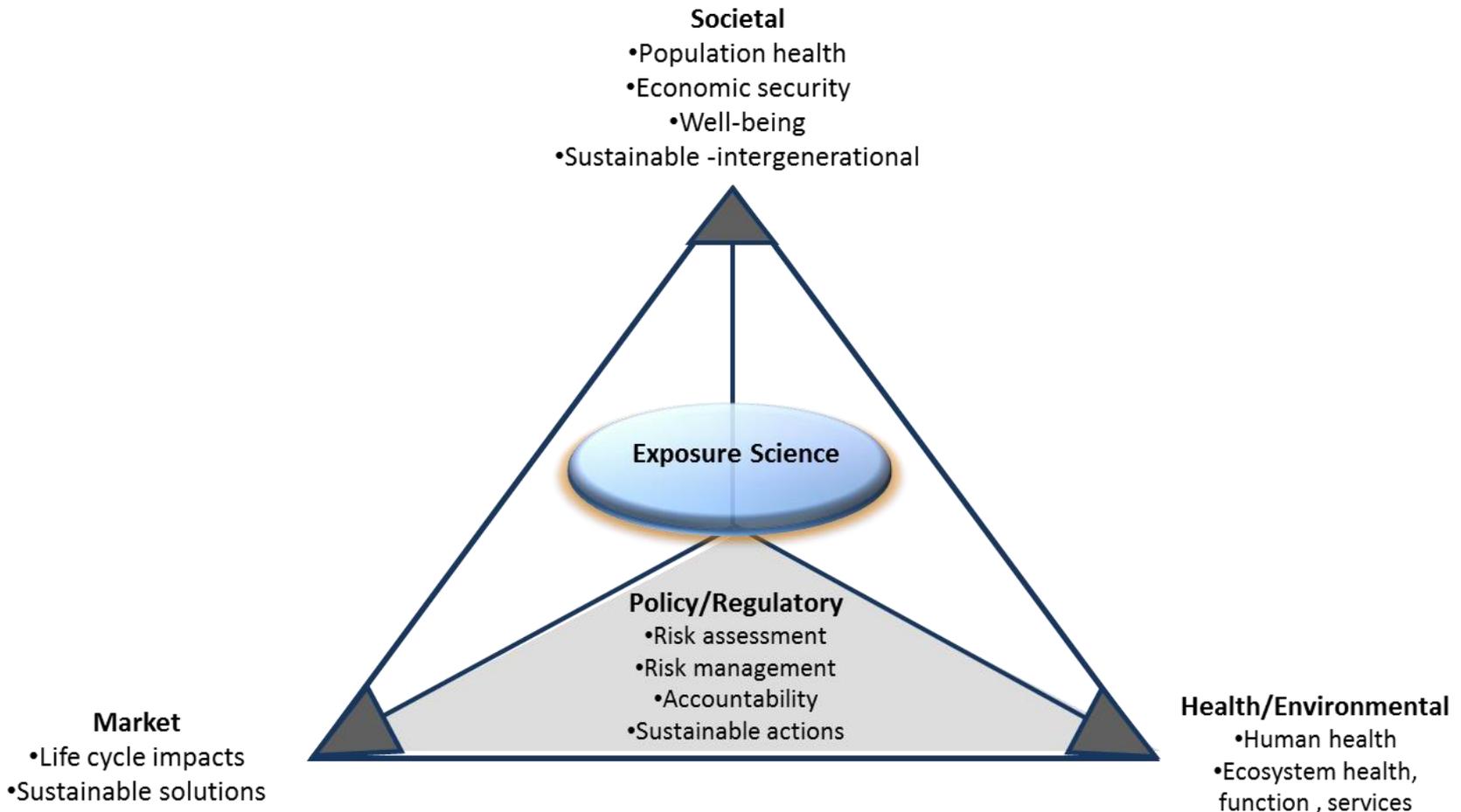
Exposure is the  of Risk Assessment

Expanded Needs for Exposure Information



NRC 2012: Demands for exposure science:

Plays a key role in public health protection, environmental regulation, urban and ecosystem planning, and disaster management



Many Exposures - Measurement Infeasible



Widespread dispersive (air, water, soil)



Point source emissions (facilities)



Direct exposures (food, products)



Expanded predictive capability is key!

Goal of proposed project:

- Increased ability to predict exposures
- Greater confidence in predicted estimates

Big Data, Informatics, and Exposure Science



- Increasing informatics capability to collect and generate large data sets from multiple linked sources
- Challenges include:
 - Spatial, temporal differences
 - Linking different data sets in a useful way
 - Confidentiality
 - Data quality
 - Data identification
 - Data access
- Exposure relevance: While generated for other reasons (marketing, social networking), identified sources can include information relevant to exposure, including product purchase, locations or purchase, activity patterns, product composition
 - Could it be used to improve exposure predictions?



This Proposal:



- Focus on consumer exposures:
 - Focus scope to be actionable/achievable
 - Direct product use indicated to be important exposure source (ExpoCast, ExpoDat)
- Design so findings relevant to approaches for obtaining, integrating and applying big data in multiple scientific fields, not just exposure
- Generate information that can be used to:
 - Provide a generic framework to help identify and integrate big data sets
 - Increase understanding of exposures
 - Test predictive exposure models
 - Potentially expand application of predictive models



Propose Two Phases:



- Phase I: Assemble interdisciplinary team
 - To identify potential data sets and considerations for use
- Phase II: Interdisciplinary workshop
 - To evaluate identified sets and generically address challenges related to access and integration

Proposal Phase 1: Establish Team



- Formation of new HESI committee
 - Engage RISK21 participants given expertise and broad experience
 - Consider additional scientific disciplines
- Committee charge to identify:
 - Types of data would be useful for improving consumer exposure prediction
 - Novel sources of these data, building upon existing work:
 - + ExpoCast
 - + HESI sponsored ISES symposium on Development and Application of Predictive Models from Big Data
 - Challenges to collecting and integrating big data



Proposal Phase 2: Workshop



- Evaluate data sources and address challenges for data use
- Data sources
 - What is available
 - Which are most relevant
 - Strengths/limitations for exposure science application
- What would be needed to use these data:
 - Confidentiality
 - Cost
 - Developing meaningful linkages between data sets collected under differing spatial and temporal conditions
 - Utility/quality of info collected for other purposes
 - Appropriate application domains
 - Building partnerships to enable data access



Anticipated Deliverables



- Develop white paper based upon workshop that:
 - 1) Identifies and assesses big data sources for exposure assessment
 - 2) Summarizes ways to address issues of how to use these data for EA (as listed above),
 - 3) Suggests approaches to address generic challenges of big data use, and
 - 4) Makes recommendations for how one specific data set could be tested and evaluated for contribution
- Follow up: Test the data set described in the workshop report for usefulness in testing/improving exposure estimates

Potential Advancements from Project



Findings can be used to:

- Contribute to success of efforts to utilize and integrate big data in general (any scientific area)
- Evaluate and improve exposure models
 - Approach used here will be directly applicable for consumer exposures, could form general basis of approach to assess other exposure sources
 - Potential to expand predictive capability
 - Reduce uncertainty in exposure estimates

Proposal Genesis



- Based upon 2014 proposal, modified based upon feedback:
 - Smaller, more focused “pilot” size
 - Big data aspects particularly relevant
- 2014 proposal key elements:
 - Gather information on emerging technologies and data sources including strengths and weaknesses
 - Examine how these multiple data sources and new technologies can be integrated using a systems approach
 - Provide a case example of such integration
 - Determine if systems approach improves the predictive value of exposure models

Thank you - Questions?

