7th Biennial CDC and ATSDR Symposium on Statistical Methods
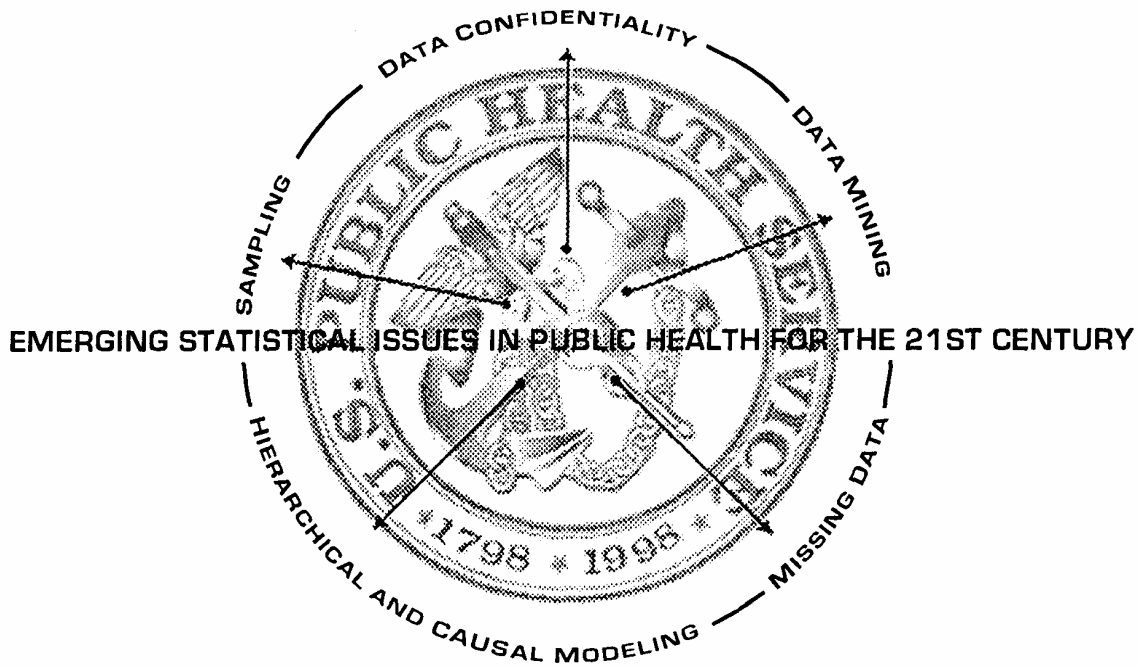
Emerging Statistical Issues in Public Health for the 21st Century

DATA CONFIDENTIALITY

DATA MINING

SAMPLING

EMERGING STATISTICAL ISSUES IN PUBLIC HEALTH FOR THE 21ST CENTURY

MISSING DATA

HIERARCHICAL AND CAUSAL MODELING

PROGRAM AND ABSTRACTS

January 27-29, 1999

Sic Transit Gloria Numeri: Methodology and Practical Tools for Mining Diagnostic Information From Useless and Seemingly Normal Data

V.I. Kvitash
Department of General Internal Medicine, School of Medicine
University of California
San Francisco, California

Objective: To develop a scientifically sound and clinically verifiable concept, axiomatic theory, methodology, and applied domain-free tools for mining diagnostic/prognostic information from useless data, defined as seemingly normal data which do not support or contradict clinical presentation and create significant uncertainties regarding diagnosis, treatment decision-making, and outcomes.

Methods: For simultaneous quantitative, qualitative, relational, and directional evaluation of interactions among multiple data points, their values are translated into Natural System Equivalent Units scaled from 0 to 100 to allow direct measurements of their interactions in abstract multi-dimensional space. Information from twelve-dimensional metabolic space is translated into quantitative networks of six distinct types of homeostatic dysfunctions designated as *(1)* homeostatic disintegration, *(2)* homeostatic integration, (3) homeostatic inversion, (4) combined homeostatic disintegration-inversion, (5) combined homeostatic integration- inversion, and (6) combined homeostatic inversion. All types of homeostatic dysfunctions are represented as a cascade of six windows of metabolic networks.

Results: Clinical usefulness of developed tools were successfully verified in ischemic heart diseases, cardiac surgery, pulmonary function testing, primary immune-deficiency, allergy, HIV-related diseases, and geriatrics. These tools identified previously unknown patterns of metabolic networks in sixty-six diseases, and discovered a novel class of knowledge --detectable immuno-metabolic pathology.

Conclusion: The presented methodology demonstrated its usefulness in (1) mining clinical evidence; (2) facilitating accurate and reliable diagnosis, effective treatment decisions, and disease management; (3) early recognition of diseases; (4) discovering new clinical entities, identifying new subtypes of known diseases, and diagnostic pattern cognition/recognition; (5) having explanatory and heuristic value; and (6) representing a novel class of useful knowledge which is not available from any other existing modalities.