

# What Factors Determine Health? Understanding Health Histories Using Medicare Claims Data

---

James Sorace MD MS  
Senior Medical Officer

Office of Assistant Secretary for Planning and  
Evaluation



**U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES**  
Office of the Assistant Secretary for Planning and Evaluation

# Disclaimer

---

THE FINDINGS AND CONCLUSIONS OF THIS REPORT ARE THOSE OF THE AUTHORS AND DO NOT NECESSARILY REPRESENT THE VIEWS OF THE ASSISTANT SECRETARY FOR PLANNING AND EVALUATION, THE CENTERS FOR MEDICARE & MEDICAID SERVICES, OR THE DEPARTMENT OF HEALTH AND HUMAN SERVICES.



**U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES**  
Office of the Assistant Secretary for Planning and Evaluation

# What Factors Determine Health?

---

- At the individual level, this fundamental question underlies the issues involved in projecting and managing an individual's future health resource requirements.
- At the population level, risk adjustment of a population is easier to accomplish to the extent individuals' health needs can be predicted.
- When can these factors be quantitated?



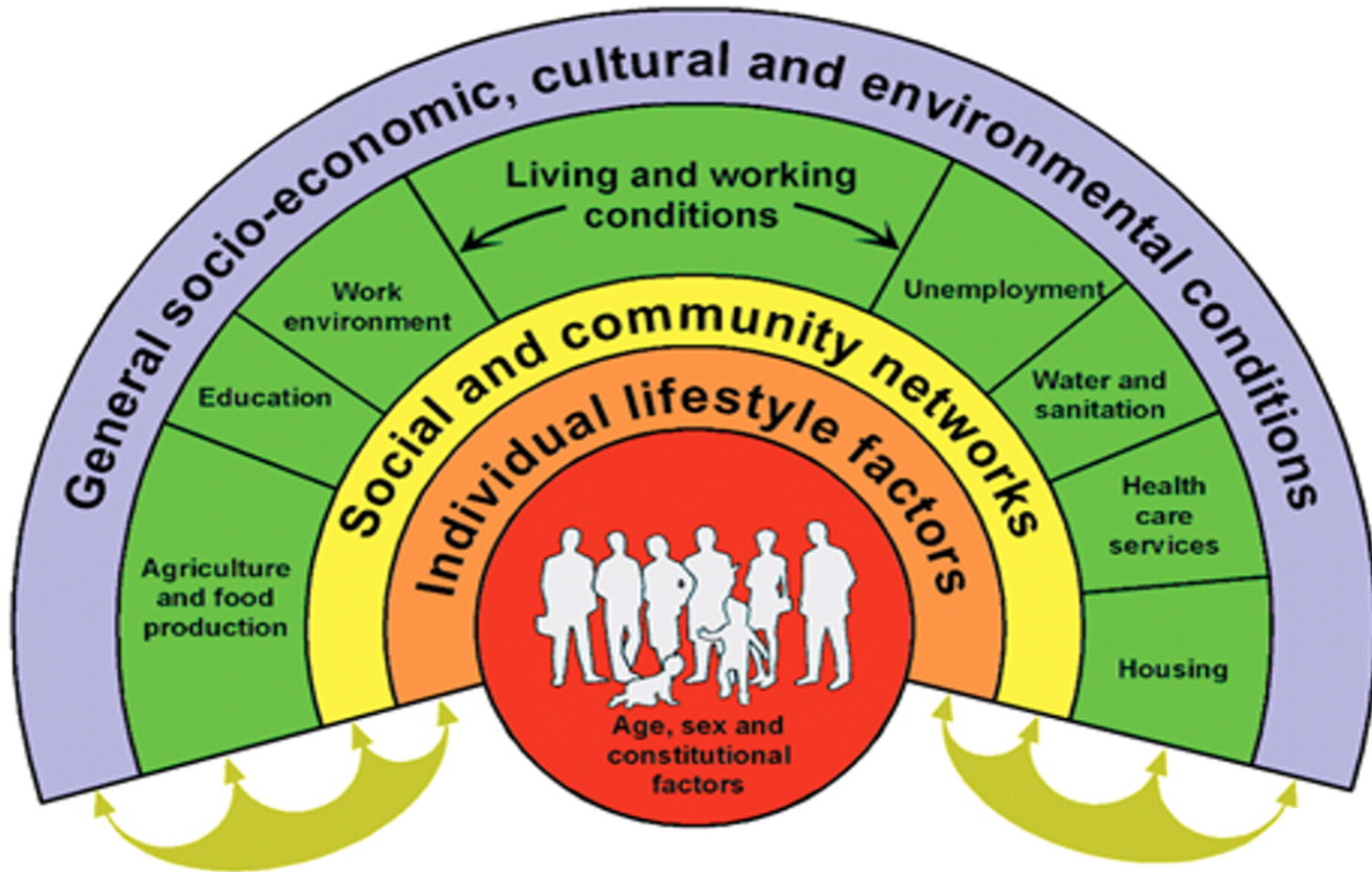
# Social Determinants of Health

---

- These results frame the importance of considering socioeconomic and behavioral factors in health.
- Socioeconomic Factors: “Social determinants of health reflect the social factors and physical conditions of the environment in which people are born, live, learn, play, work, and age.”
- Behavioral Factors: “individual behaviors such as substance abuse, diet, and physical activity. ”
- For more information see Health People 2020 at:  
[https://www.healthypeople.gov/2020/about/foundation-health-measures/Determinants-of-Health#individual behavior](https://www.healthypeople.gov/2020/about/foundation-health-measures/Determinants-of-Health#individual%20behavior)



# SDOH framework



DEPARTMENT OF HEALTH



U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

Office of the Assistant Secretary for Planning and Evaluation

Source: Dahlgren, G. and Whitehead, M. (1991) Policies and Strategies to Promote Social Equity in Health. Stockholm Institute for Futures Studies

# Presentation Overview

---

- We will review 3 projects:
  - **Hypothesis 1: Prior cost can predict future health.**
    - Project : Cost Swim Lanes
  - **Hypothesis 2– Heredity can predict future health.**
    - Project : The ASPE/CMS Twin Study
  - **Hypothesis 3– Disease burden (i.e. the set of diseases a patient currently has) can predict future health.**
    - Project : ASPE/CMS studies regarding the combinatorial complexity of disease combinations in Medicare.



# ASPE/CMS Data Infrastructure

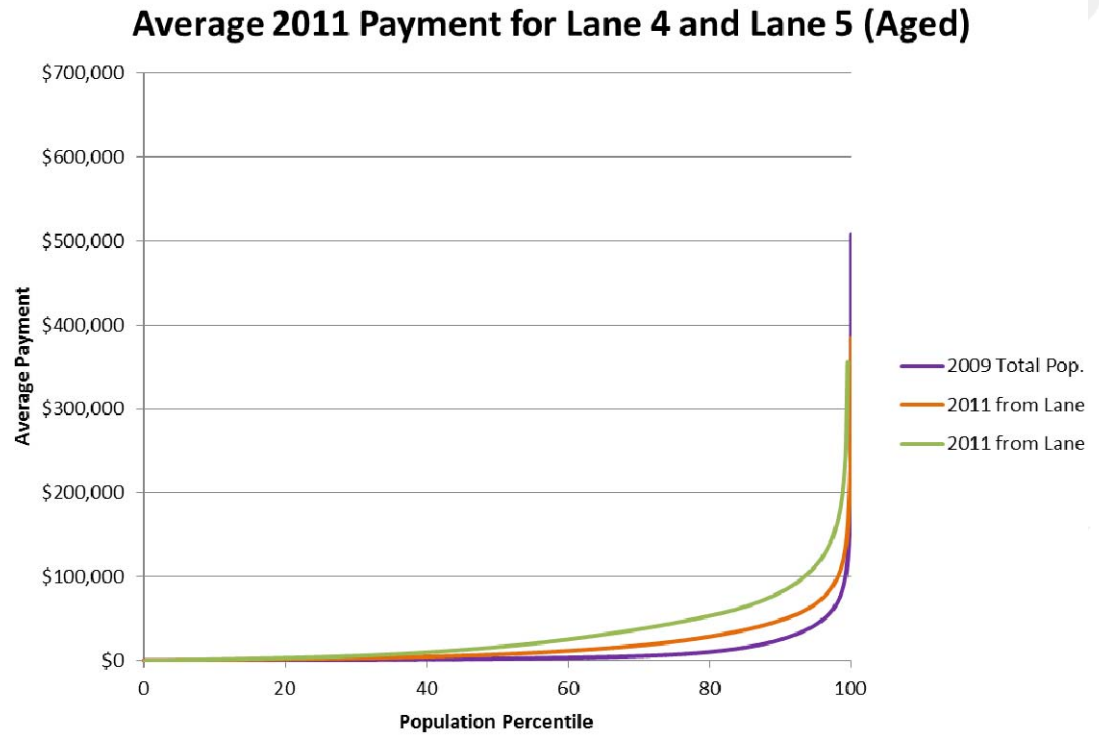
---

- Done as part of the “Medicare DataLink” project with Acumen LLC.
- Database contains:
  - All Medicare Fee For Service Part A and B claims since 1991.
  - All Medicare Part-D drug claims since start of the benefit in 2006.
  - Nursing Home MDS and Home Health OASIS.
  - Medicare Part-C.
- Projects 2 and 3 used CMS’s Hierarchical Conditions Category (HCC) system to group diseases.



# Hypothesis 1: Prior Cost Do Not Predict Future Cost

- Purple curve: 2009 expenditure curve represents the distribution of 2009 Expenditures of all Beneficiaries who entered Medicare at age 65 or above.
- Green curve : 2011 expenditures of Beneficiaries in lane 5, 2009 (top 96.5 % to 100%). Note 65.8% survived to and 56.2% survived through 2011.
- Red curve: 2011 expenditure of Beneficiaries in lane 4, 2009 (top 80% to 96.5%) . Note 84.6% survived to and 78.7% survived through 2011.





## Hypothesis 2: Heredity Predicts Future Health

---

- Twin studies allow the relative contributions of genetics and shared family environment (nature vs. nurture) to be quantitated.
- These effects may be estimated and then used for resource allocation decisions.



# Project 2: ASPE/CMS Twin Study

---

- In collaboration with VCU's Mid-Atlantic Twin Registry, we matched 396 pairs of Monozygotic (MZ) twins and 378 pairs of Dizygotic (DZ) twins to their Medicare claims data from 1991 through 2011.
- Predominantly white, male and Mid-Atlantic. Studied pairs in which both members survived to age 65.
- We used the Medicare claims database to construct unrelated matched control pairs (MCPs) for both the MZ and DZ twins. Specifically we matched the MCPs on sex, age, race and current county of residence.



# Project 2: ASPE/CMS Twin Study

---

- We now have 4 groups to compare:
  - Monozygotic twins: MZ
  - Dizygotic twins: DZ
  - Demographically Matched Control Pairs for the MZ Population: MZ-MCP
  - Demographically Matched Control Pairs for the DZ Population: DZ-MCP



# MCP Methodology Advantages

---

- Controls for differences in traits that may be associated with demographic differences between the MZ and DZ populations and therefore not due to heredity.
- Provides for 3 independent comparisons: MZ vs. MZ-MCP, DZ vs. DZ-MCP, and the traditional MZ vs. DZ.
- In comparison to the traditional MZ vs. DZ comparison, the twin vs. MCP comparisons:
  - Do not share family environment.
  - Exhibit a greater genetic difference. For example the MZ vs. MZ-MCP group differ in “100%” of their genes as opposed to the “50%” difference found in the traditional comparison.



# Twin Study Results: shared HCCs

---

- MZ shared 26.3% of the HCCs as compared to 19.8% shared by the MZ-MCP ( $P < 0.001$ ). Representing a 6.5% absolute and 33% relative increase.
- DZ shared 25.6% of the HCCs as compared to 21.8% shared by the DZ-MCP ( $P < 0.001$ ). Representing a 3.8% absolute and 17% relative increase.
- MZ/DZ ( $p = 0.52$ ) and MZ-MCP/DZ-MCP ( $p = 0.029$ ) comparisons were not significant.
- The only HCC that showed significance for both the MZ vs. MZ-MCP and DZ vs DZ-MCP comparisons was HC-92, Specific Heart Arrhythmias, and in particular ICD-9 code 427.3: Atrial Fibrillation and Flutter.
- The MZ twins (but not the DZ twins) also had a significantly higher concordance for HCC\_96, Ischemic or Unspecified Stroke, which disappeared after MZ twins concordant for ICD-9 code 427.3 were dropped.



U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES  
Office of the Assistant Secretary for Planning and Evaluation

---

**MZ-MCP:** Monozygotic Twins Matched Control Pairs  
**DZ-MCP:** Dizygotic Twins Matched Control Pairs

**MZ:** Monozygotic Twins  
**DZ:** Dizygotic Twins

# Twin vs MCP: KS-Test Curves of Monthly Expenditure Difference

Abbreviations:

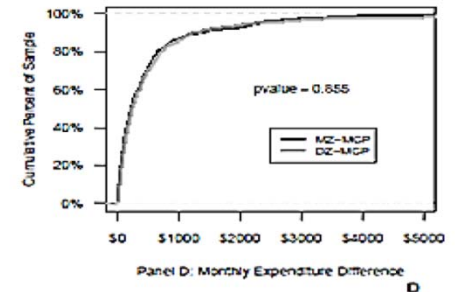
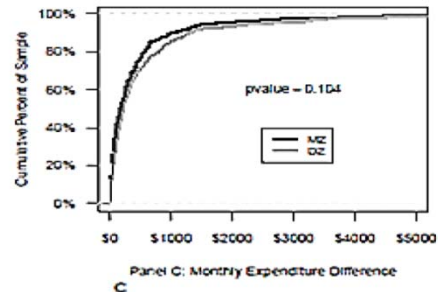
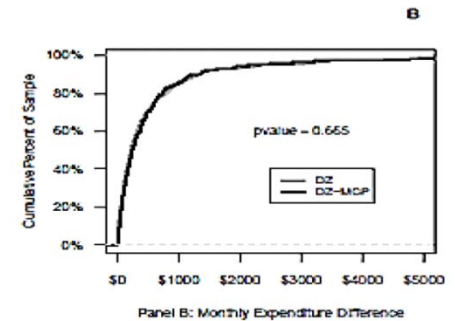
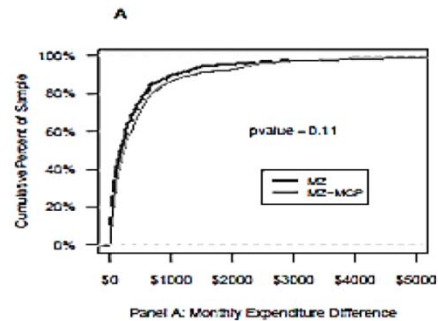
MZ – Monozygotic twin group

DZ – Dizygotic twin group

MZ-MCP - Monozygotic  
matched control pair group

DZ-MCP - Dizygotic matched  
control pair group

P-values were calculated  
using the Kolmogorov–  
Smirnov (KS) test



**U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES**  
Office of the Assistant Secretary for Planning and Evaluation

# Twin Study Conclusions

---

- Within the limitations of the study, the role of heredity is limited in predicting health.
- Additional factors need to be considered when making resource allocations and risk adjustment. These might include:
  - Disease Burden
  - Socioeconomic Factors
  - Behavioral Factors



## Hypothesis 3: Disease Combinations in Medicare

---

- There is a large body of evidence that patients with comorbidities account for a large share of Medicare expenditures.
- Can clusters of disease combinations (DCs) be defined that predict future health?
- Is there a national set of DCs in the Medicare population that is relatively simple and stable overtime that can be used to estimate expenditures and resource requirements?





# Disease Combination Analysis

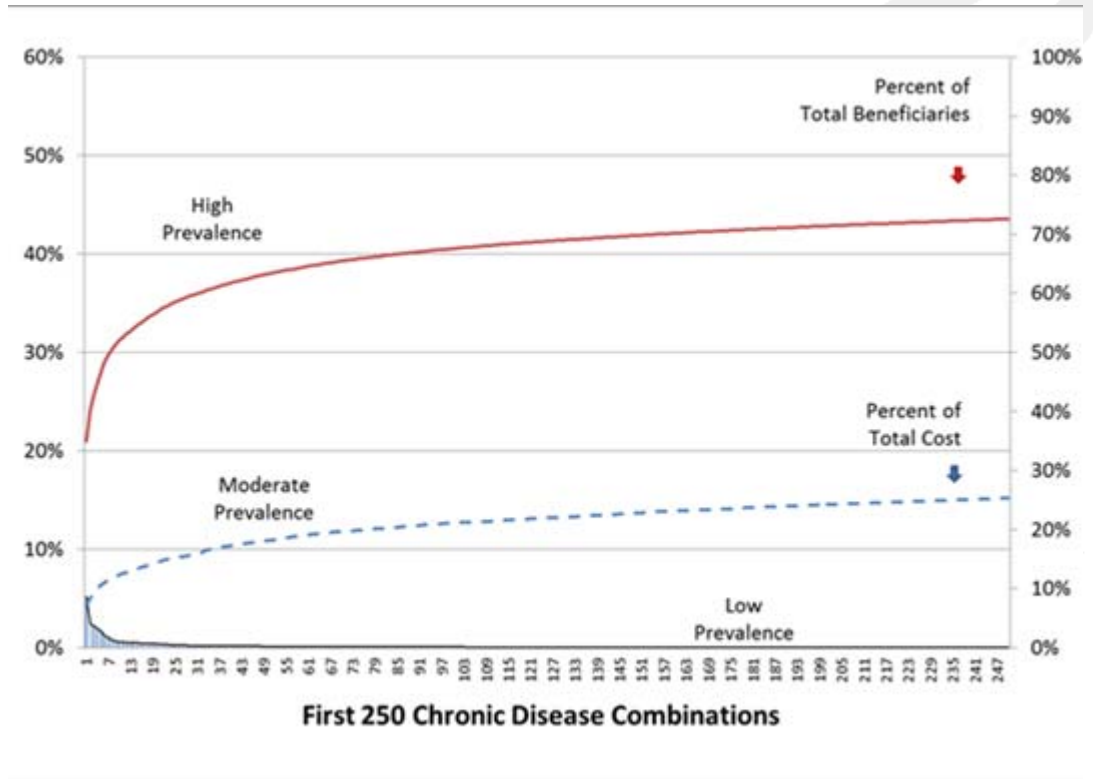
---

- All 2008 Beneficiaries with continuous fee for service claims history.
  - 32,220,634 Beneficiaries
  - \$283,088,306,347
- 2,027,394 Disease Combinations (DCs) were identified at the 70 HCC level with 1,658,233 (81.8%) containing a single beneficiary.
- Three distinct populations:
  - No HCC: representing 35% of Beneficiaries and 6% of expenditures.
  - 100 most prevalent DCs: representing 33% of Beneficiaries and 15% of expenditures.
  - Remaining 2,027,294 DCs: representing 32% of Beneficiaries and 79% of expenditures.



# Long Tailed Distribution of Medicare Disease Combinations

The graph displays the first 250 Diseases Combinations, ranked by prevalence, from the baseline HCC analysis. Note that the left Y-axis represents the proportion of the population that is included in each unique disease combination. The right Y-axis represents the cumulative percent of the total population (red format) and the total expenditure (blue format) and is adjusted for the 32% of beneficiaries and 6% of expenditures that are associated with the no-HCC population. As there are over 2 million disease combinations calculated by this methodology, the figure's X-axis would need to be extended over 8,000 fold to the reader's right before both cumulative lines reached 100%.



# Disease Combination Analysis

---

- In additional studies we determined that the national set of combinations changes over time.
- This illustrates the challenges in traditional risk adjustment.



---

**U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES**  
Office of the Assistant Secretary for Planning and Evaluation

# Role for Behavioral and Socioeconomic Factors

---

- It has been recognized for many years that behavioral and socioeconomic factors are significant.
- Blum's model dating back to the 1970s stressed the importance of interplays between:
  - Environment
  - Behavior
  - The Healthcare System
  - Heredity
- Recent advances in data science enable us to design experiments to test the contribution of these factors.



# References

---

- [A Comparison of Disease Burden Between Twins and Control Pairs in Medicare: Quantification of Heredity's Role in Human Health.](#)  
Sorace J, Rogers M, Millman M, Rogers D, Price K, Queen S, Worrall C, Kelman J. Popul Health Manag. 2015 Feb 6. [Epub ahead of print]  
PMID: 25658666
- [The complexity of disease combinations in the Medicare population.](#) Sorace J, Wong HH, Worrall C, Kelman J, Saneinejad S, MaCurdy T. Population Health Management. 2011 Aug;14(4):161-6.
- [Temporal variation in patterns of comorbidities in the Medicare population.](#) Sorace J, Millman M, Bounds M, Collier M, Wong HH, Worrall C, Kelman J, MaCurdy T. Popul Health Manag. 2013 Apr;16(2):120-4. doi: 10.1089/pop.2012.0045. Epub 2012 Oct 31. PMID: 23113637



# Thank You!

---

**JAMES SORACE MD, MS**  
**ASSISTANT SECRETARY FOR PLANNING AND**  
**EVALUATION: OFFICE OF SCIENCE AND DATA POLICY**

**EMAIL: [JAMES.SORACE@HHS.GOV](mailto:JAMES.SORACE@HHS.GOV)**

**OFFICE: 202-205-8678**



**U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES**  
Office of the Assistant Secretary for Planning and Evaluation

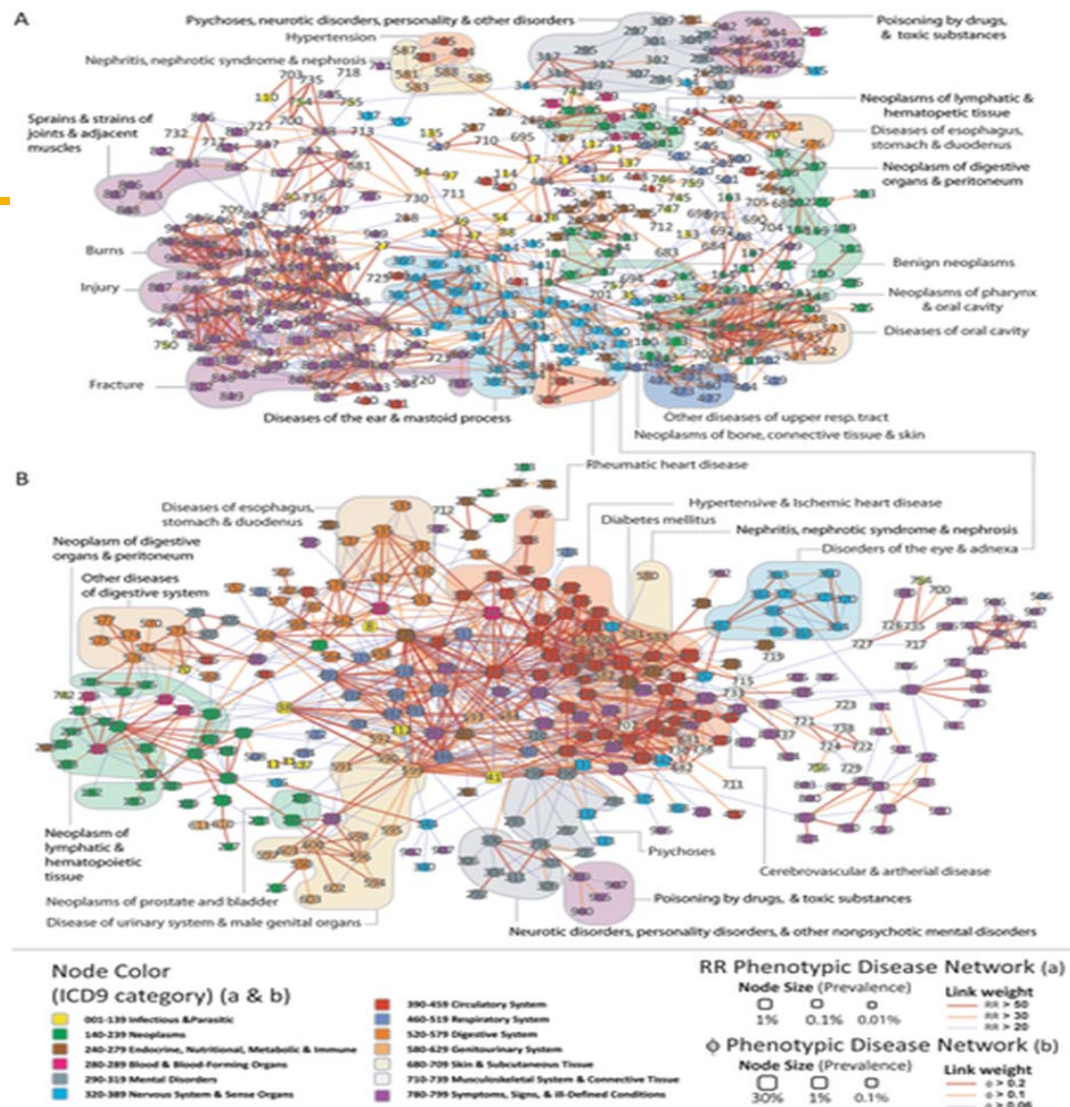
## Figure 2. Phenotypic Disease Networks (PDNs).

Nodes are diseases; links are correlations. Node color identifies the ICD9 category; node size is proportional to disease prevalence. Link color indicates correlation strength. A. PDN constructed using  $RR$ . Only statistically significant links with  $RR_{ij} > 20$  are shown. B. PDN built using  $\varphi$ -correlation. Here all statistically significant links where  $\varphi > 0.06$  are shown.

Hidalgo CA, Blumm N, Barabási A-L, Christakis NA (2009) A Dynamic Network Approach for the Study of Human Phenotypes. *PLoS Comput Biol* 5(4): e1000353.

doi:10.1371/journal.pcbi.1000353

<http://www.ploscompbiol.org/article/info:doi/10.1371/journal.pcbi.1000353>



**U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES**  
Office of the Assistant Secretary for Planning and Evaluation