# Value of environmental information and how to gather it

**David Hunter**

**Nuffield Department of Population Health**
**University of Oxford**

**Harvard TH Chan School of Public Health**

**Channing Laboratory, Brigham and Women's Hospital**

**Broad Institute of MIT and Harvard**

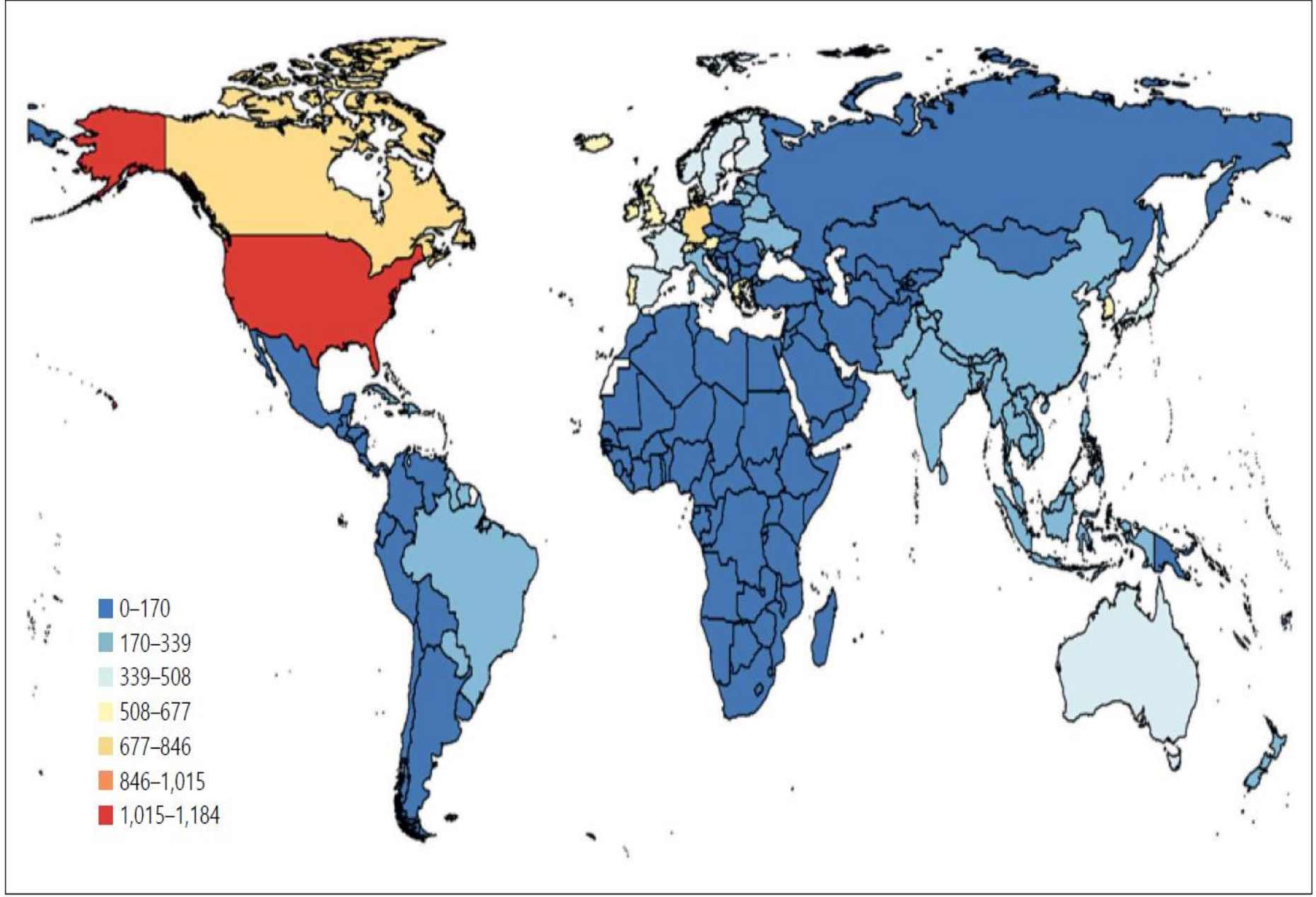
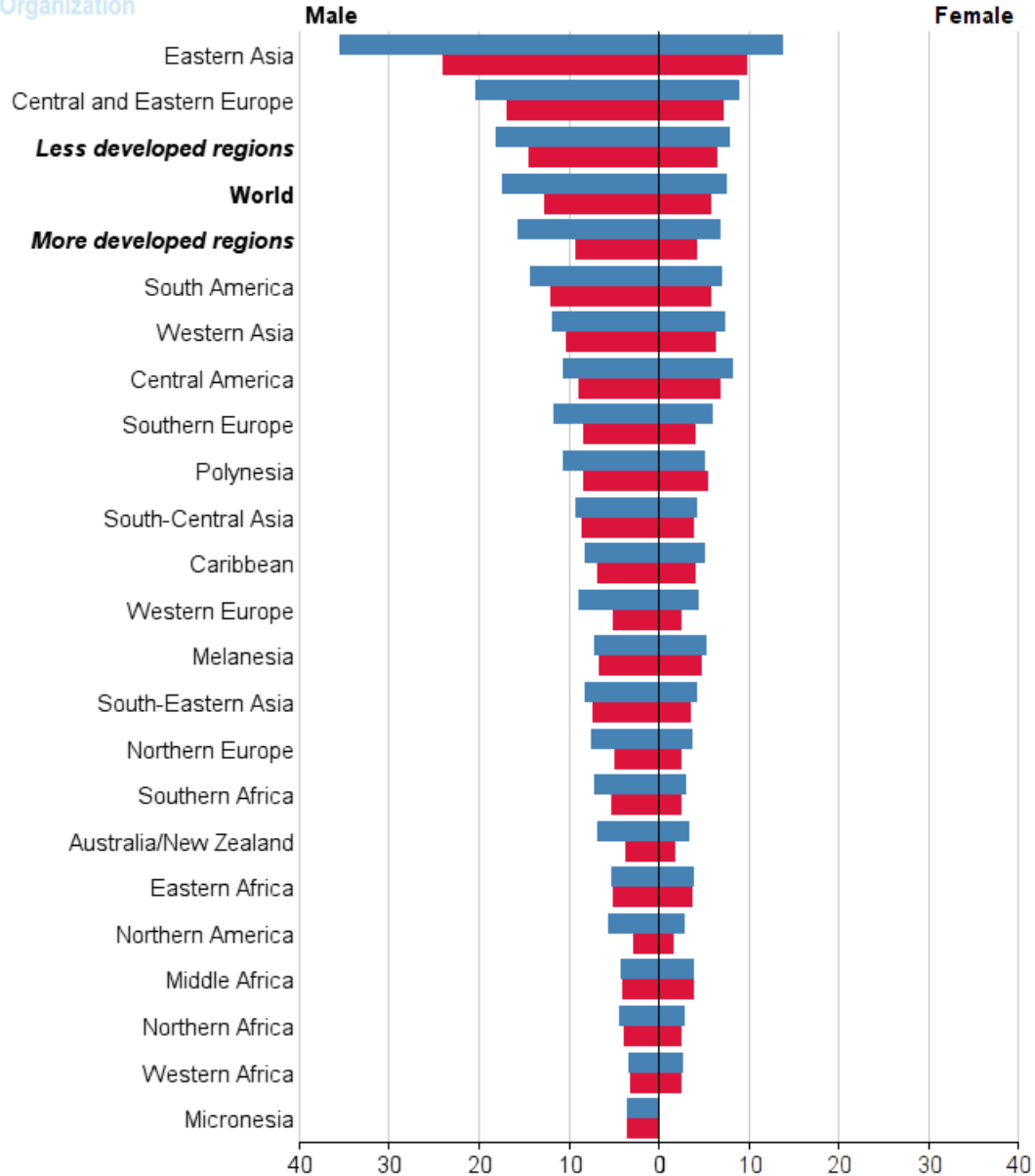



**Fig. 1.** Age-standardized annual prevalence (per 100,000) of ischemic stroke in 2013.

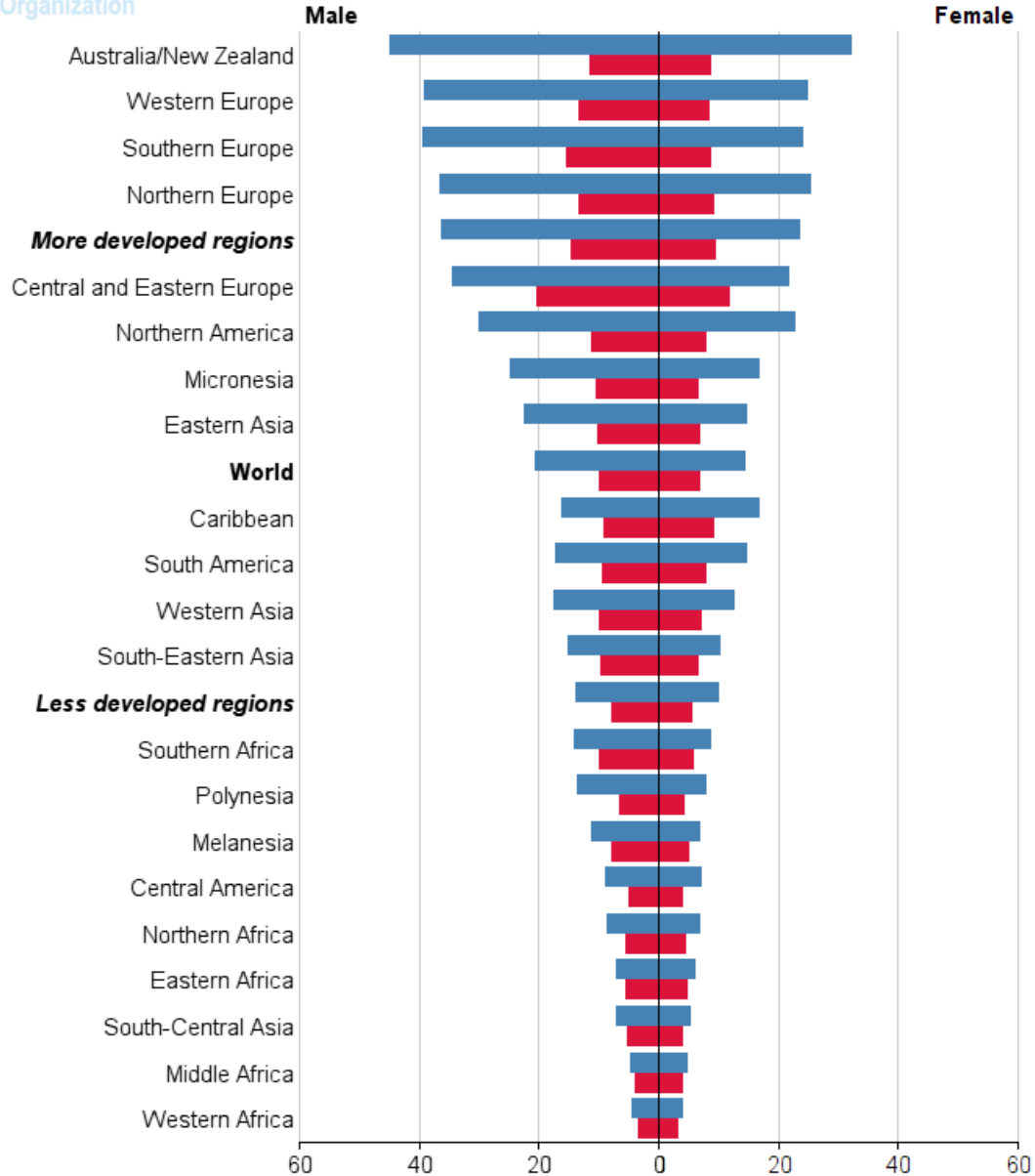
GBD Feigin et al. Neuroepidemiology 2015;45:230–236

# Stomach
## ASR (W) per 100,000, all ages

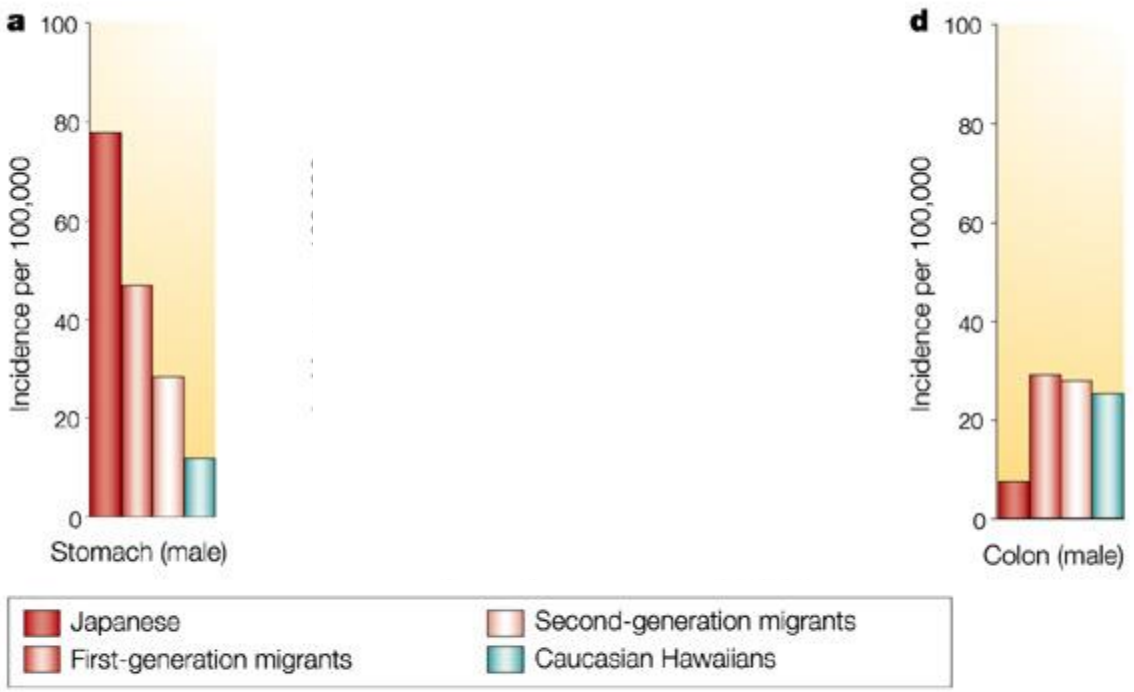



GLOBOCAN 2012 (IARC) (15.3.2018)

# Colorectum
## ASR (W) per 100,000, all ages



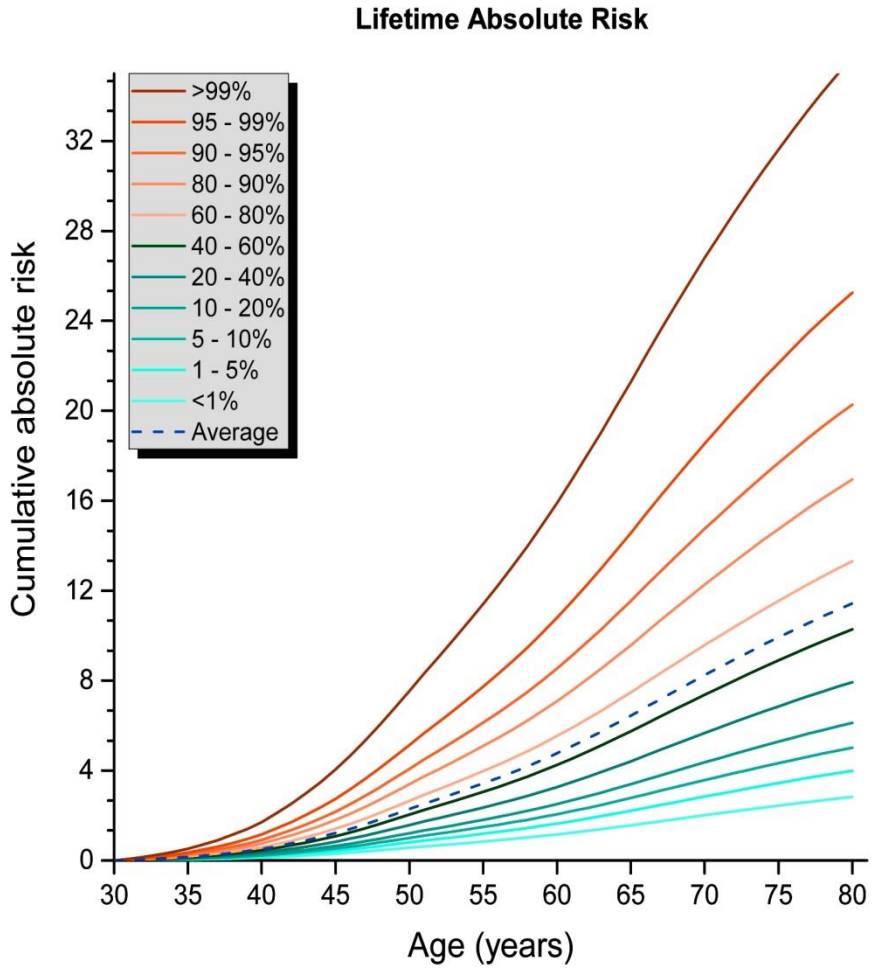

GLOBOCAN 2012 (IARC) (15.3.2018)




**Cancer Rates in Hawaii 1973-1977**

The multiethnic cohort study: exploring genes, lifestyle and cancer risk
Laurence N. Kolonel, David Altshuler & Brian E. Henderson
Nature Reviews Cancer 4, 519-527 (July 2004)

Differences in rates of most diseases between countries (and over time within countries) are due to differences in environmental and “lifestyle” risk factors – not genetic differences



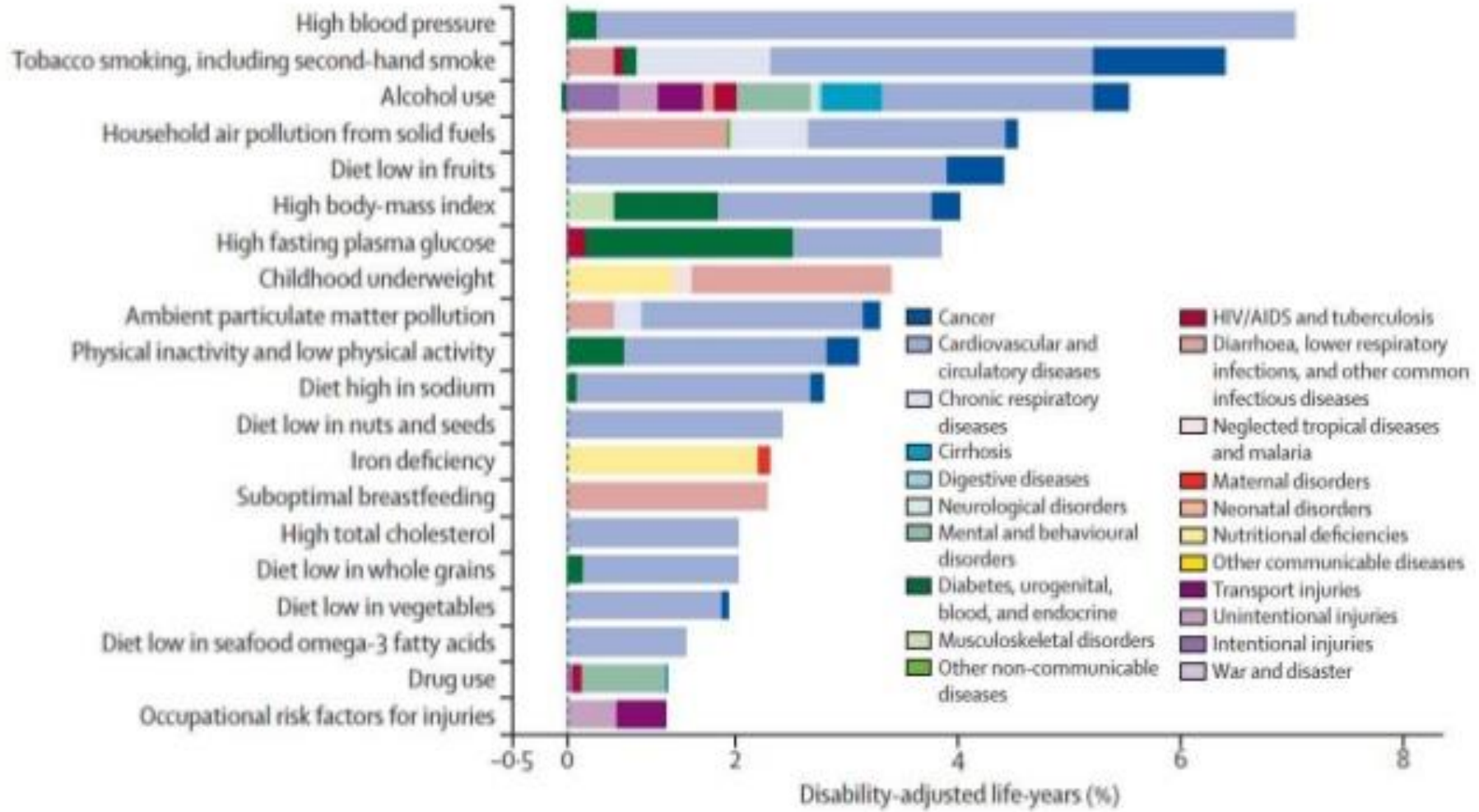

# Post-GWAS Polygenic Risk Scores are predictive – Breast Cancer

**Maas, Chatterjee et al. JAMA Oncol 2016**

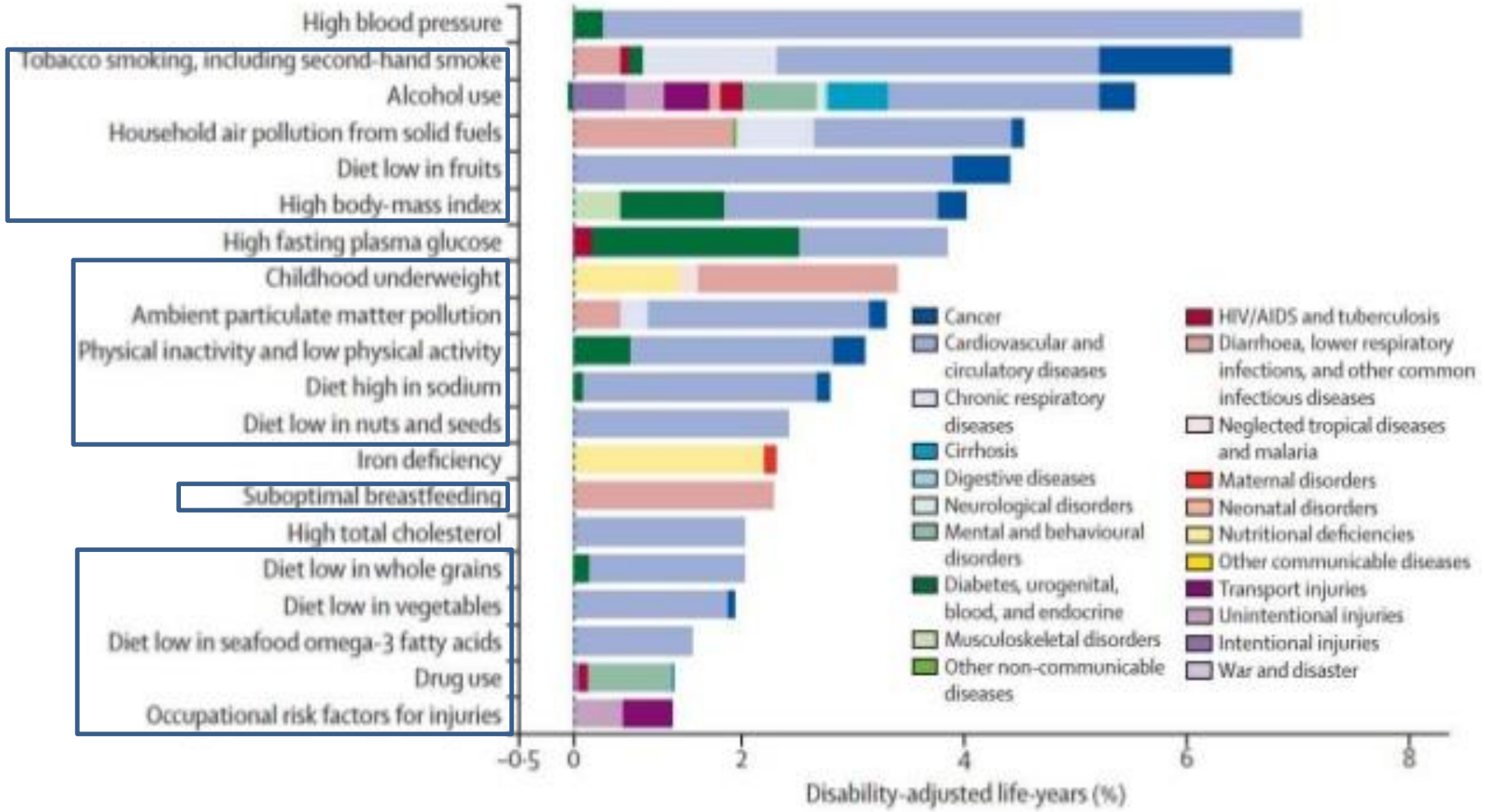
Differences in individual risk of most diseases within countries are due to differences in both genetic and environmental and "lifestyle" risk factors.

We need to measure both.

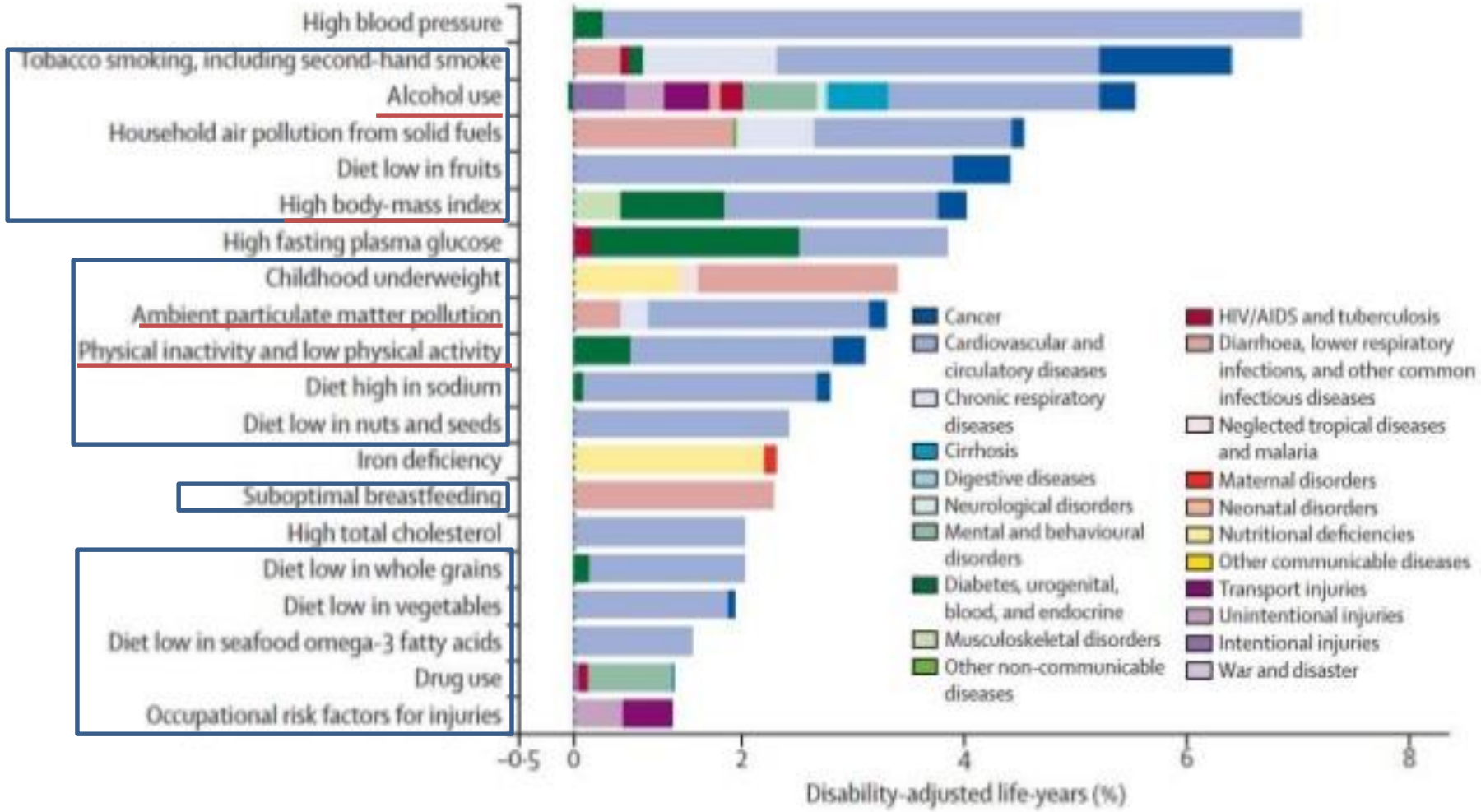
# Burden of disease attributable to 20 leading risk factors in 2010, expressed as a percentage of global disability-adjusted life years, both sexes



# Burden of disease attributable to 20 leading risk factors in 2010, expressed as a percentage of global disability-adjusted life years, both sexes



# Burden of disease attributable to 20 leading risk factors in 2010, expressed as a percentage of global disability-adjusted life years, both sexes



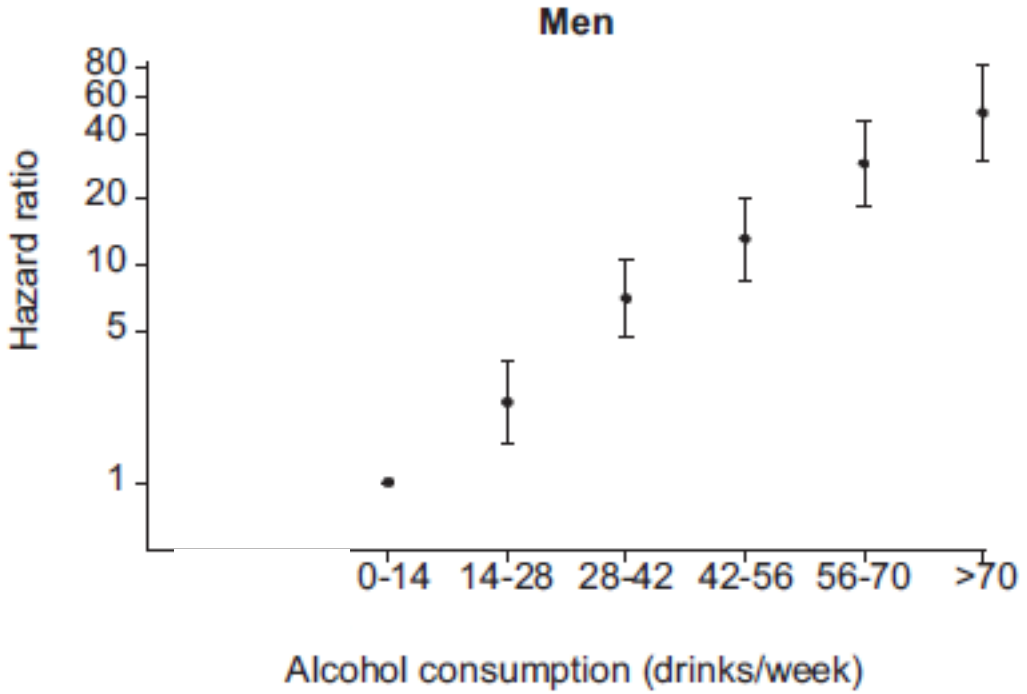
Reasons given for skepticism about exposure measurement in observational studies:

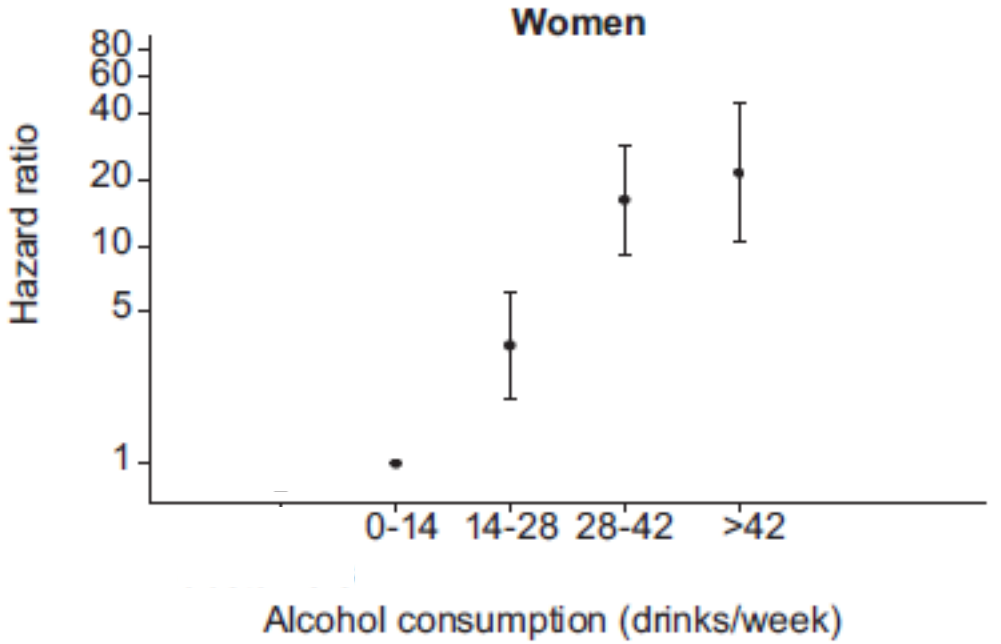
- People lie
- People don’t know
- People can’t estimate complex exposures – we need gizmos
- All of the above – Mendelian Randomization is the solution

Risk of alcoholic liver cirrhosis
Danish Cancer, Diet and Health cohort

Adj. smoking, education, waist circumference

Askgaard et al. J Hepatol. 2015

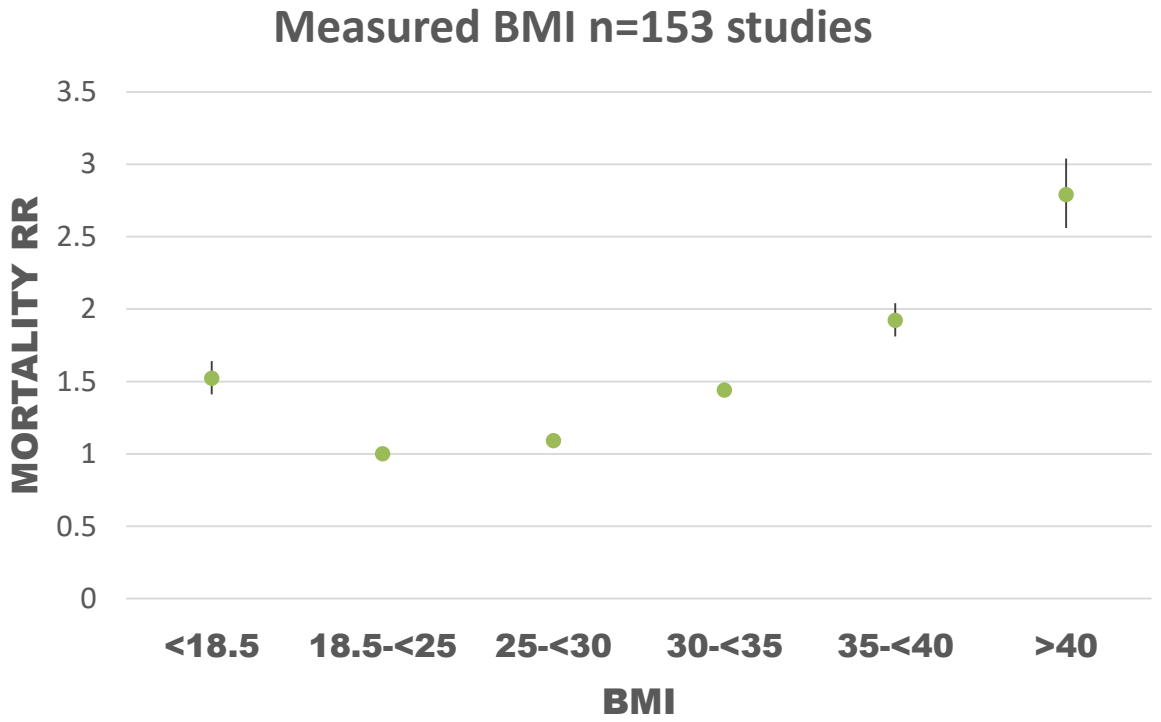


BMI
Vs Mortality

Global BMI
Mortality
Collaboration

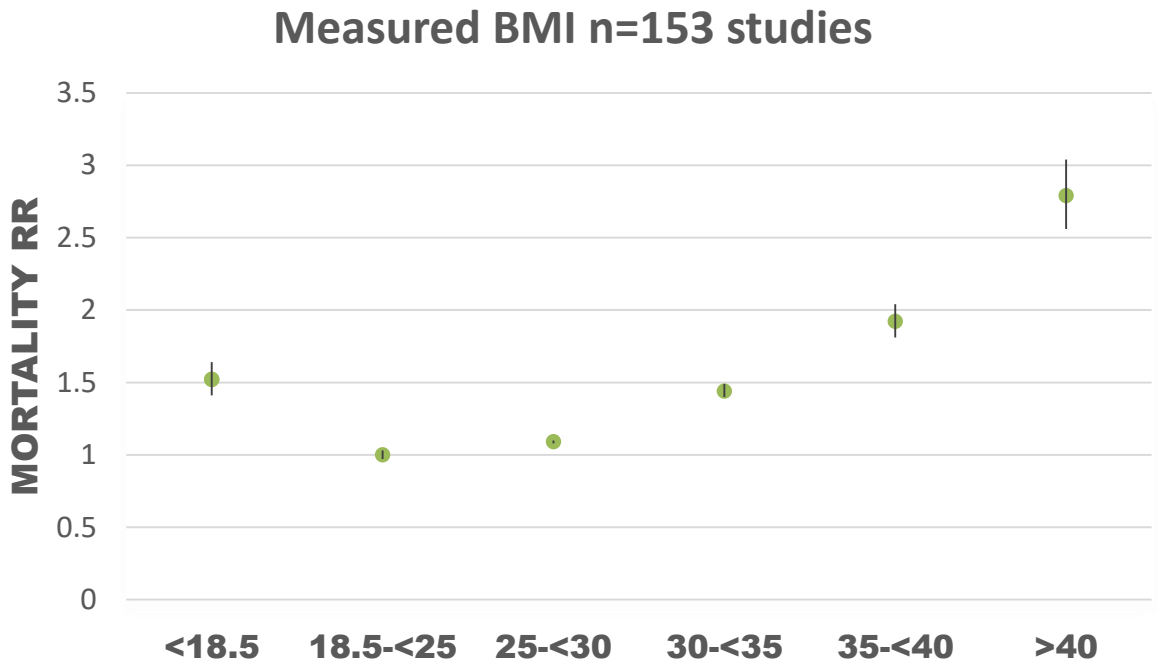
Lancet 2016

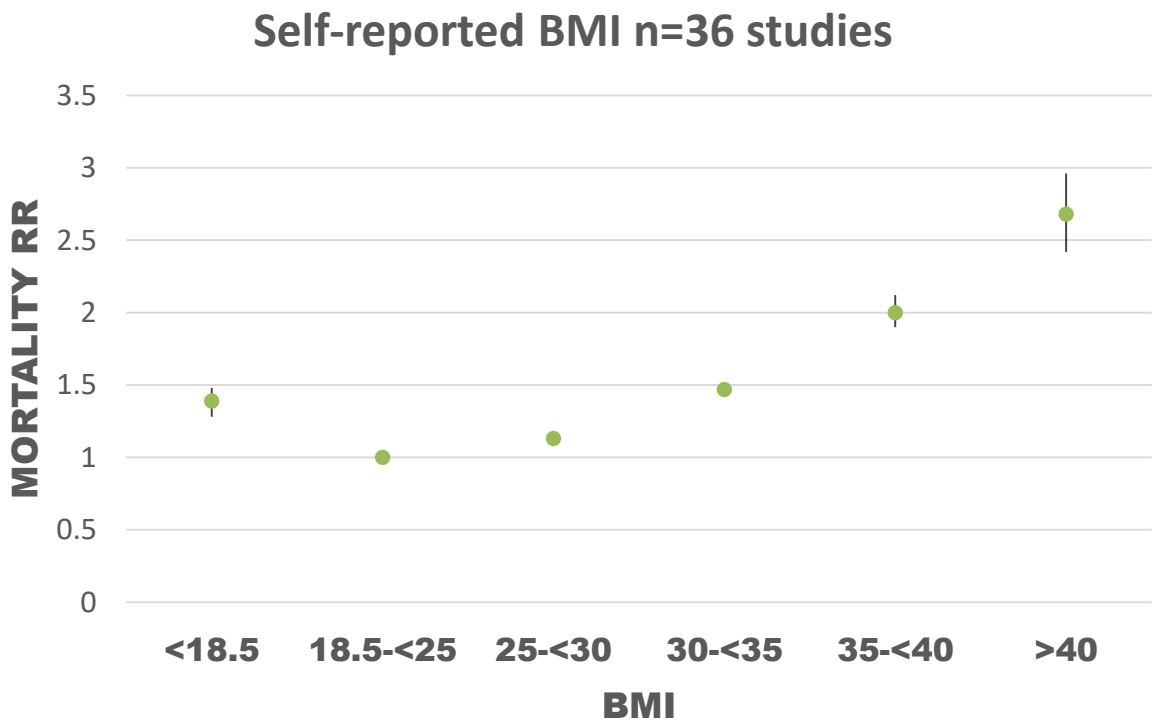


BMI
Vs Mortality

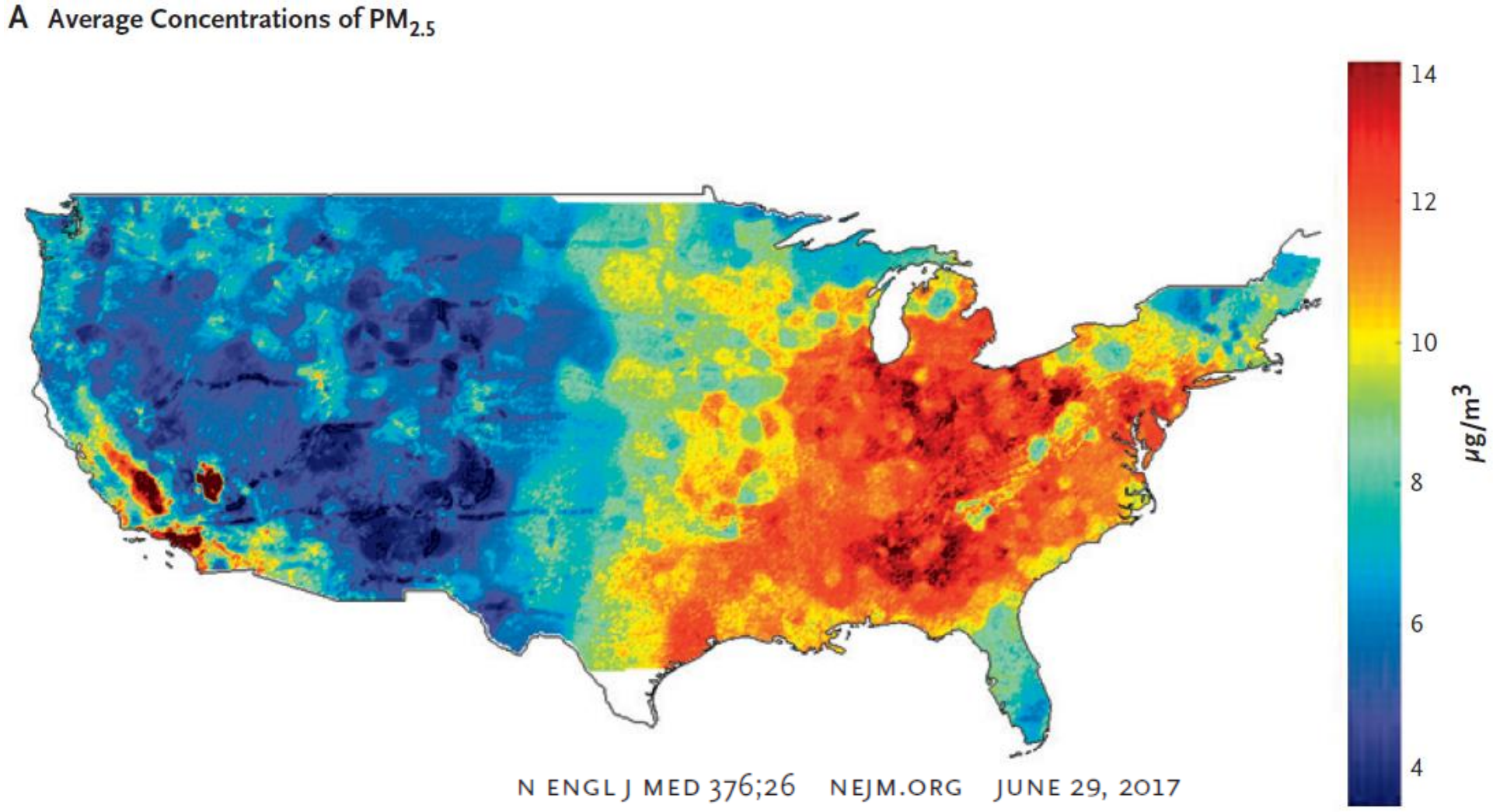
Global BMI
Mortality
Collaboration

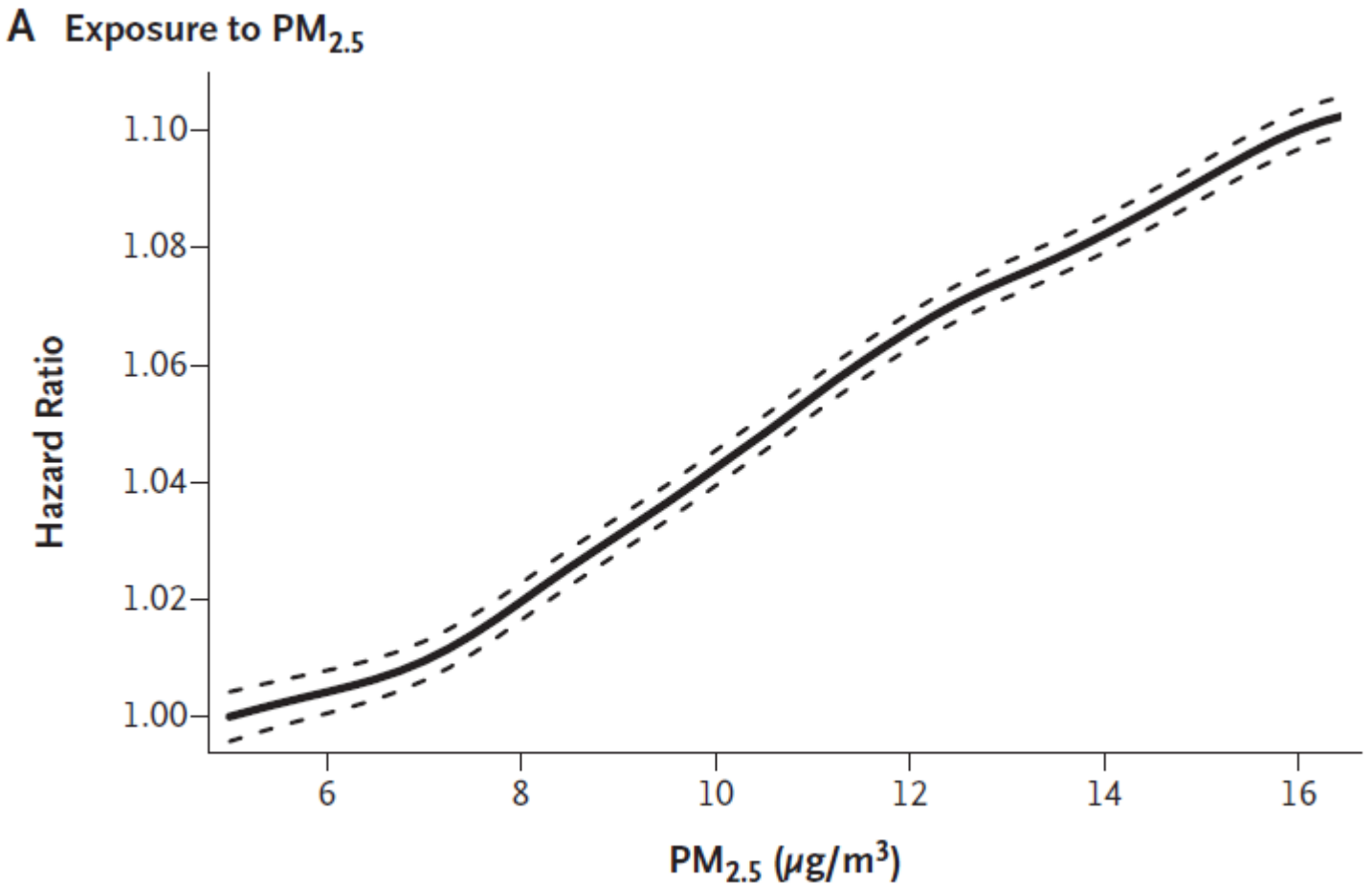
Lancet 2016



# Air Pollution and Mortality - Geolocation

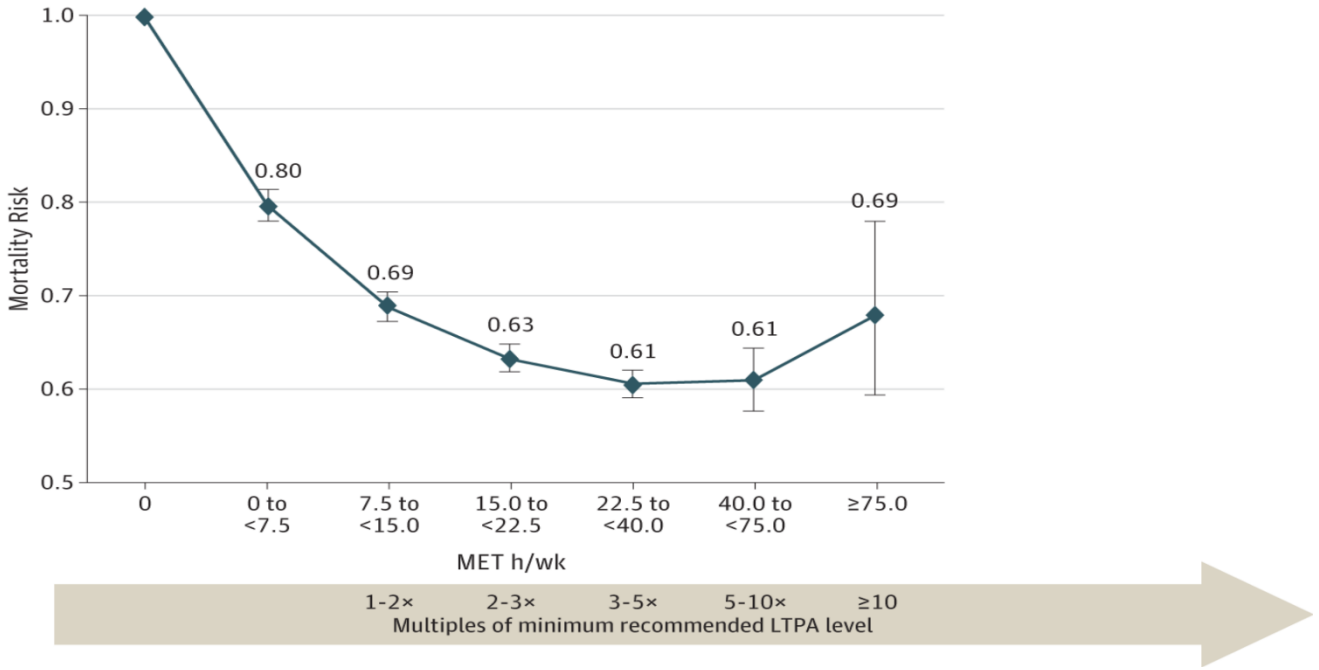





Concentration of PM 2.5 vs Mortality - >60M Medicare recipients 2000-2012

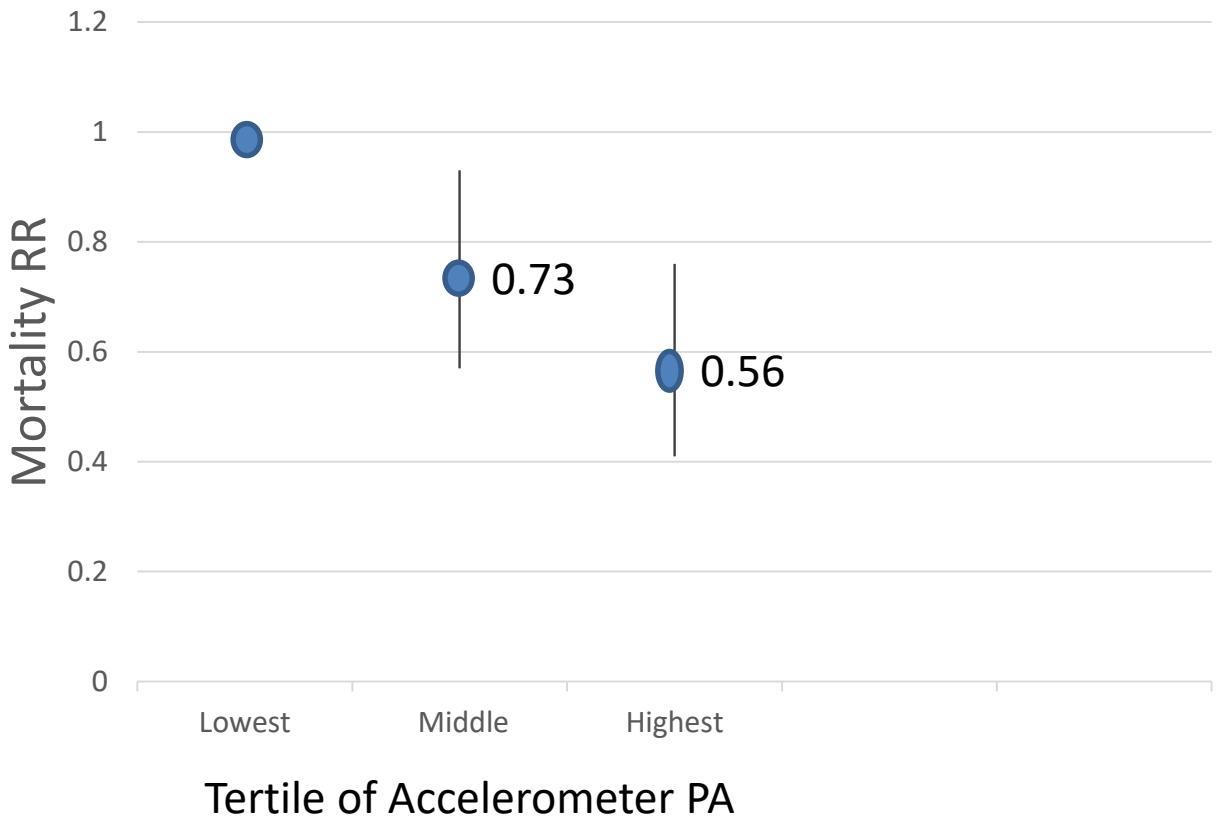
From: **Leisure Time Physical Activity and MortalityA Detailed Pooled Analysis of the Dose-Response Relationship**

JAMA Intern Med. 2015;175(6):959-967. doi:10.1001/jamainternmed.2015.0533



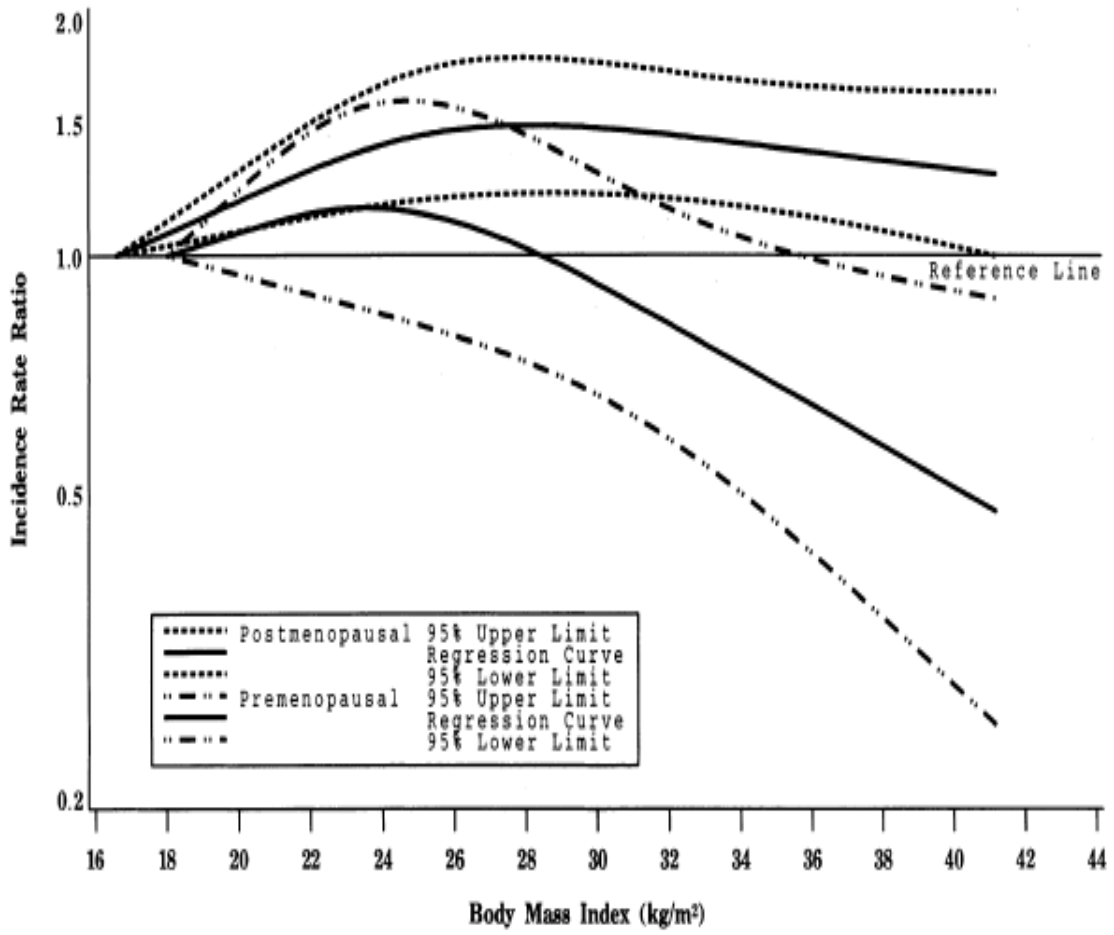

Hazard Ratios (HRs) and 95% CIs for Self-reported Leisure Time Moderate- to Vigorous-Intensity Physical Activity and Mortality

# Triaxial accelerometer-measured PA vs Mortality in the WHI




Womens' Health Initiative, n=6,382, 450 deaths.  LaMonte et el. J Am Geriatr Soc. 2017

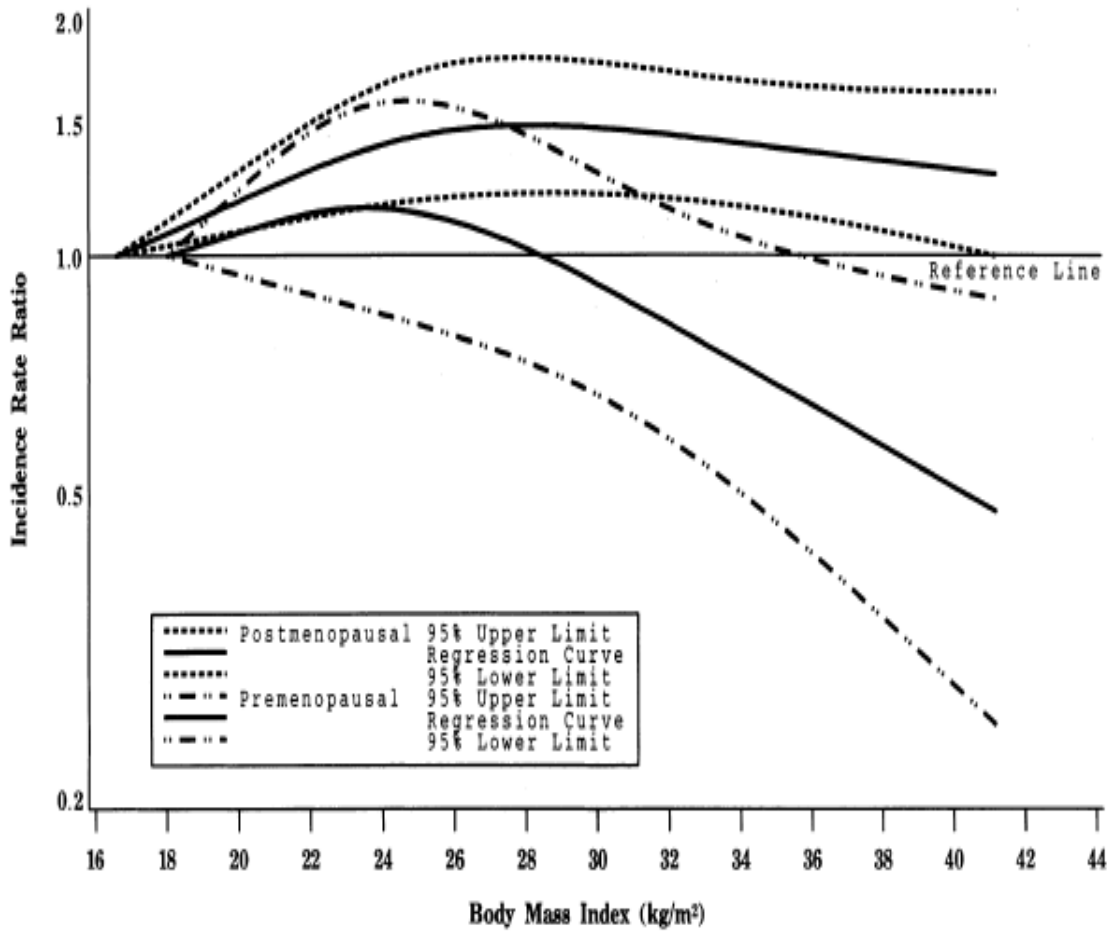
# Is Mendelian Randomization the solution?




BMI (kg/m$^2$) AND PRE- and POST-MENOPAUSAL BREAST CANCER RISK

**Pooling Project of Prospective Studies, 1991-, Van den Brandt et al., AJE 2000**

# Is Mendelian Randomization the solution? Not always




**MR ANALYSIS**
**OR per 5kg/m$^2$**

**PRE- 0.44 (0.31–0.62)**
**POST- 0.57 (0.46–0.71)**

**Guo et al. PLOS MED 2016**

BMI (kg/m$^2$) AND PRE- and POST-MENOPAUSAL BREAST CANCER RISK

**Pooling Project of Prospective Studies, 1991-, Van den Brandt et al., AJE 2000**

# OC's and Pre-Menopausal Breast Cancer – integrating Registries


1.8 M women, followed for avg 10.9 years
19.6M person-years, 11,517 cases of breast cancer
Exposure from prescription registries
Confounders from other databases

RR current use = 1.20 (1.14-1.26)

(estimates for 18 different formulations and other routes of delivery)

Cf. self-reported OC use

RR current use = 1.33 (1.03-1.73) (NHS2, CEBP 2010

RR current use = 1.24 (95% CI, 1.15 to 1.33) (Collaborative Group, Lancet 1996)

- Risk persists despite changes in formulation, longitudinal monitoring needed

# MODELLING GENE-ENVIRONMENT INTERACTIONS

**DO CLASSIC BREAST CANCER RISK FACTORS SYNERGIZE WITH GWAS SNPS?**

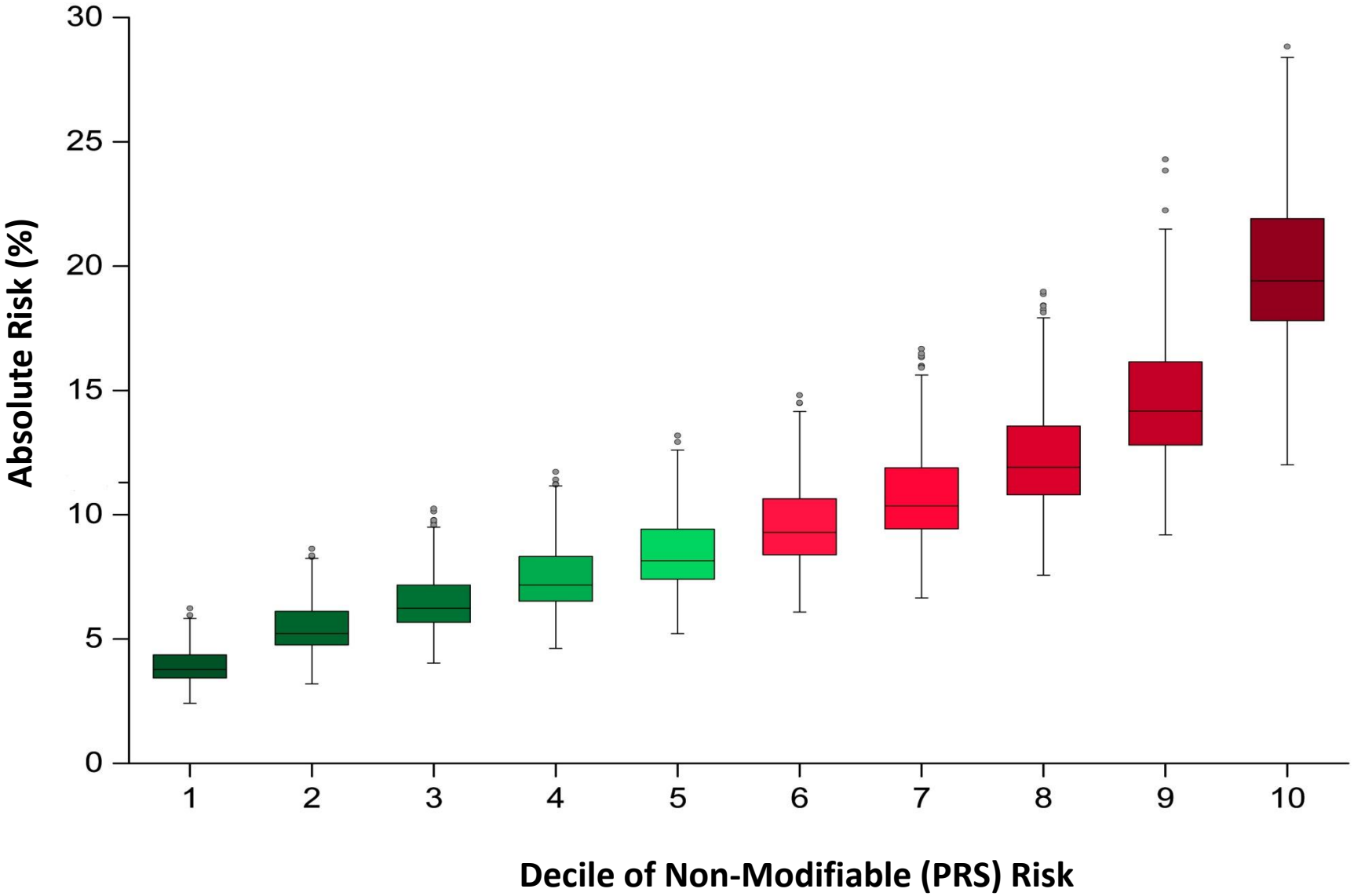
**16,285 BC cases and 19,376 controls**

**39 GWAS-assoc SNPS x 8 "Env" Risk Factors**

**AAM**
**Parity**
**AAMeno**
**Height**
**BMI**
**FH**
**Smoking**
**Alcohol**

**"After correction for multiple testing, no significant [multiplicative] interaction between SNPs and established risk factors...was found."**

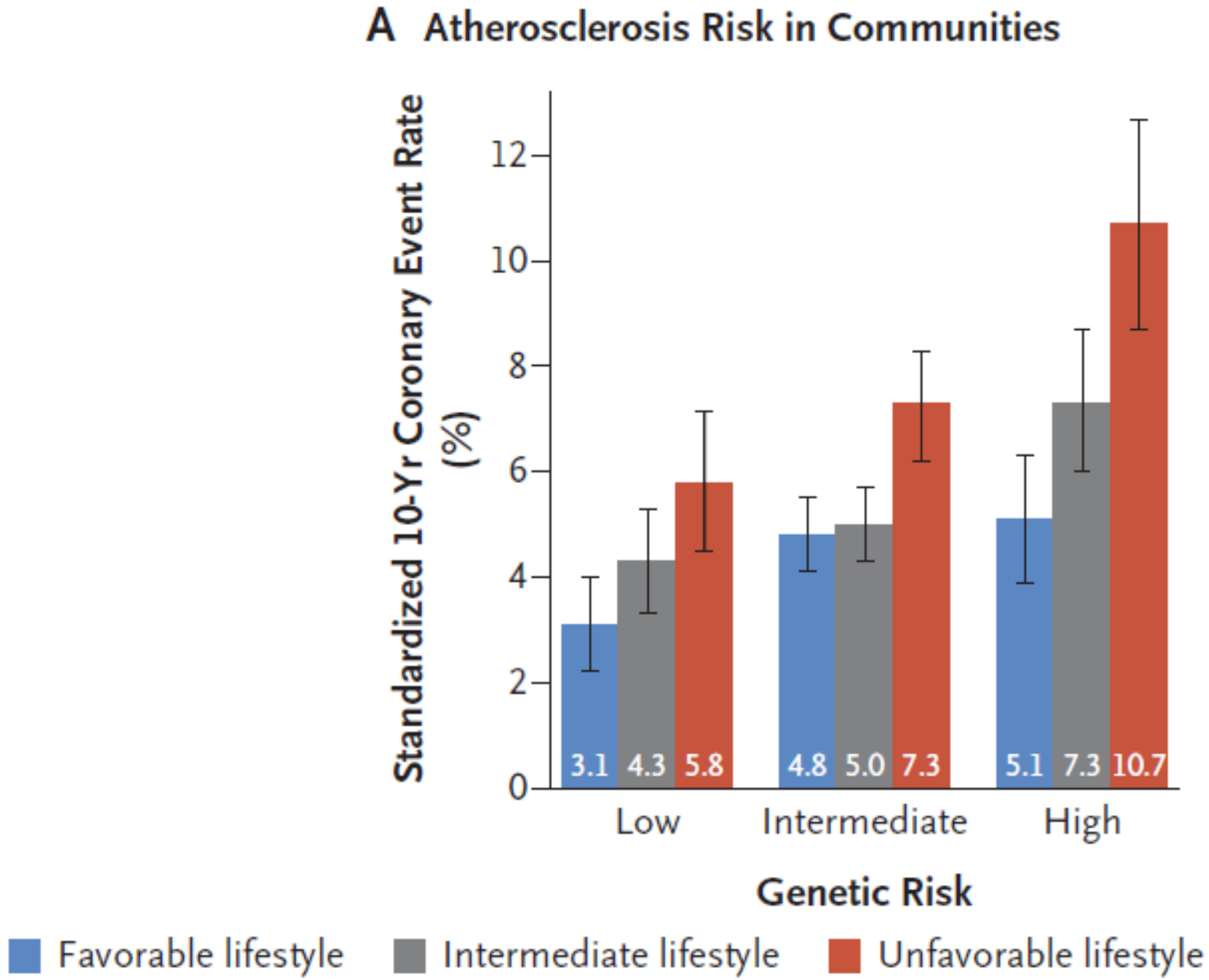
**Campa et al, JNCI 2011, Barrdahl et al, BPC3, HMG 2014**




Maas, Chatterjee et al. JAMA Oncol, 2016

**PRS, Lifestyle and CHD**

**Khera et al. NEJM 2016**



With some exceptions (e.g. drug idiosyncracies) genetic and environmental and “lifestyle” risk factors are independent and the risks multiply.

# Cohort studies – pitfalls

- Measurement error
- Confounding
- Reverse causation
- P hacking
- False positives
- Outcome data access

# Cohort studies – practice

- Measurement error
- Confounding
- Reverse causation
- P hacking
- False positives
- Outcome data access
- Regression-dilution
  Calibration methods
- Causal inference methods
  Mendelian randomization
- Latency analysis
- Prespecified analyses
- Fewer, larger studies
- EMR, Single payer systems

# Cohort studies – practice

- Measurement error
- Confounding
- Reverse causation
- P hacking
- False positives
- Outcome data access
- Regression-dilution
  Calibration methods
- Causal inference methods
  Mendelian randomization
- Latency analysis
- Prespecified analyses
- Fewer, larger studies
- EMR, Single payer systems

# Cohort studies – practice

## Self-reported data via Questionnaires

- **“low tech”**

**BUT: The NHGRI-funded PhenX project, after exhaustive analysis of exposure assessment options presented mainly questionnaire-based instruments**

PhenX
consensus measures for Phenotypes and eXposures

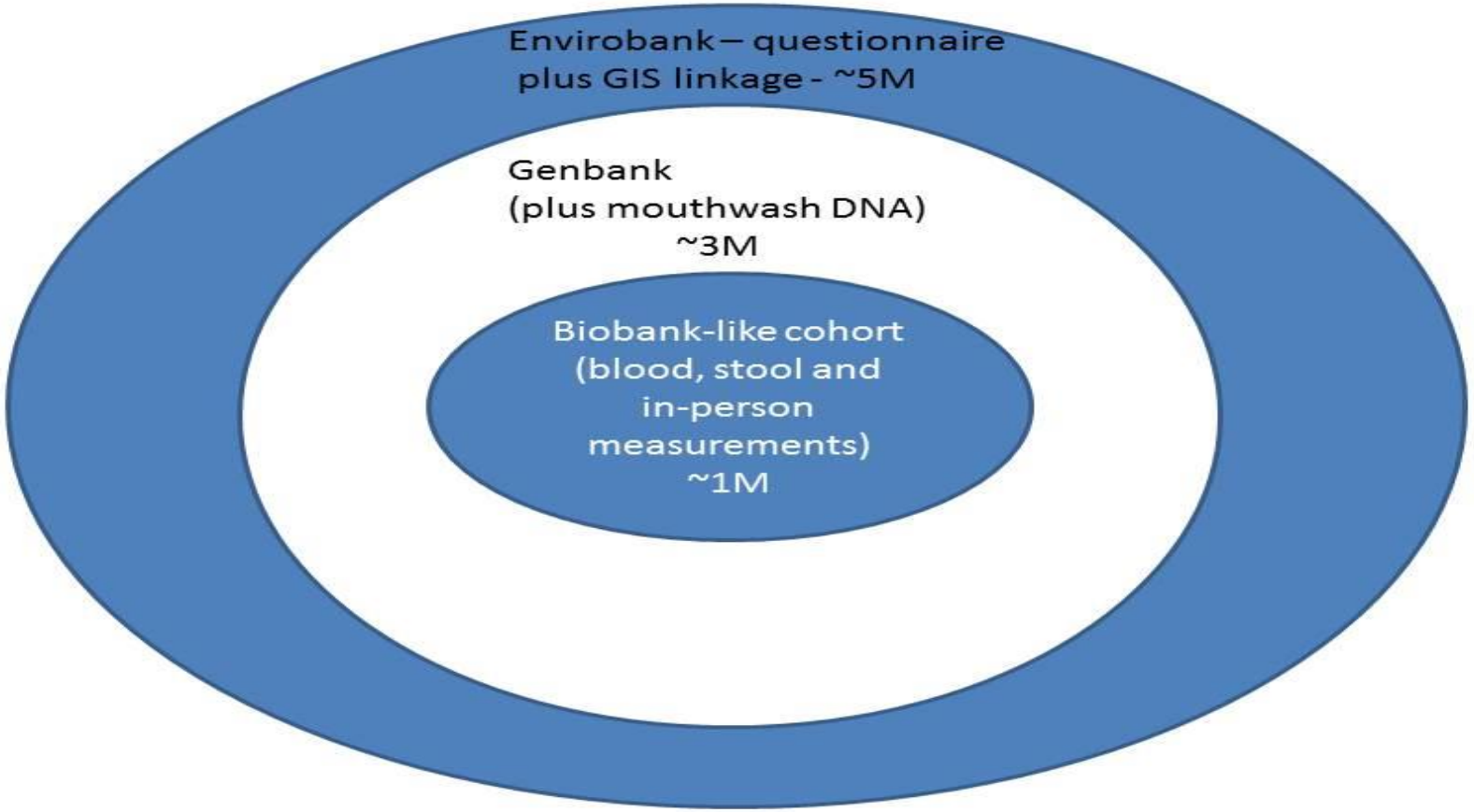
## Opportunity for harmonization

*https://www.**phenx**toolkit.org*

# Cohort studies – practice

| Era | Collection mechanism | Maximum sample sizes |
|---|---|---|
| Mid 20$^{th}$C | Interview | Thousands |
| Late 20$^{th}$C | Paper questionnaire | Hundreds of thousands |
| 21stC | Digital questionnaire (touchscreen, online, mobile, Apps) | Millions – tens of millions |

**SUCCEED cohorts** (**S**tudies in the **U**K of disease **C**auses and **C**ontrol through pre-**E**mption and **E**arly **D**etection)




UK SUCCEED Cohorts for disease prevention and early detection

# SUMMARY

- The major causes of most diseases are environmental – so we must measure them
- Rumors of the futility of measurement by self-reported data greatly exaggerated
- Gizmos will play a role – mainly in calibrating other sources of data
- Biomarkers will play a role – providing unique measurements as well as calibrating other sources of data
- Mendelian randomization useful – but no panacea
- We need larger sample sizes to analyze less common diseases prospectively
- We need larger sample sizes to analyze common diseases in the era of stratified medicine
- These can be best obtained by combining self-reported data, geolocation data, clinical data and outcome data from interoperable EMRs and system-wide outcome coding e.g. UK HES data

# Acknowledgements - DRIVErs

## DRIVE Discovery, Biology, and Risk of Inherited Variants in Breast Cancer


**Cancer Research UK**
Jack Cuzick

**Dana Farber Cancer Institute**
John Quackenbush
Judy Garber

**Fred Hutchinson CRI**
Paul Auer
Ross Prentice

**German Cancer Research Center (DKFZ)**
Federico Canzian
Rudolf Kaaks
Daniele Campa

**BWH**
Rulla Tamimi
Sue Hankinson
Meir Stampfer

**Imperial College**
Elio Riboli

**HSPH**
Peter Kraft
Connie Chen
Sara Lindstroem
Amit Joshi
Walter Willett
Donna Spiegelman
Frank Hu
Eric Rimm

**Mayo Clinic**
Fergus Couch

**National Cancer Institute**
Stephen Chanock
Joe Fraumeni

**Stanford University**
Alice Whittemore

**University of Cambridge**
Douglas Easton
Kyria Mikailidou

**University of Chicago**
Brandon Pierce
Habibul Ahsan

**University of Southern California**
Chris Haiman
Fred Schumacher

**University of Utah**
David Goldgar

**Vanderbilt University**
Wei Zheng
Jirong Long

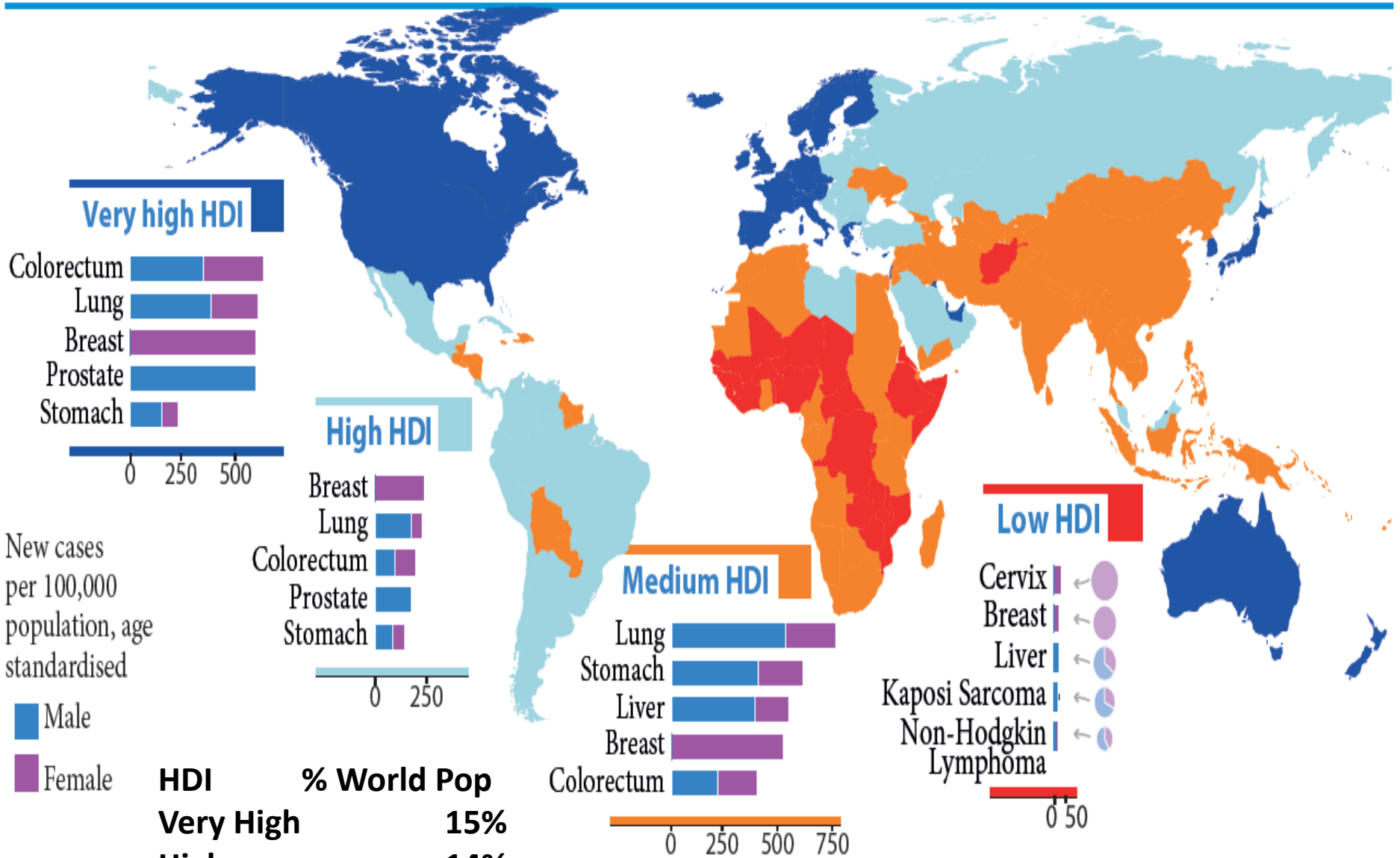
**JHU**
Nilanjan Chatterjee

**University Hawaii**
Loic Le Marchand

**NCI Program Office**
Daniela Seminara

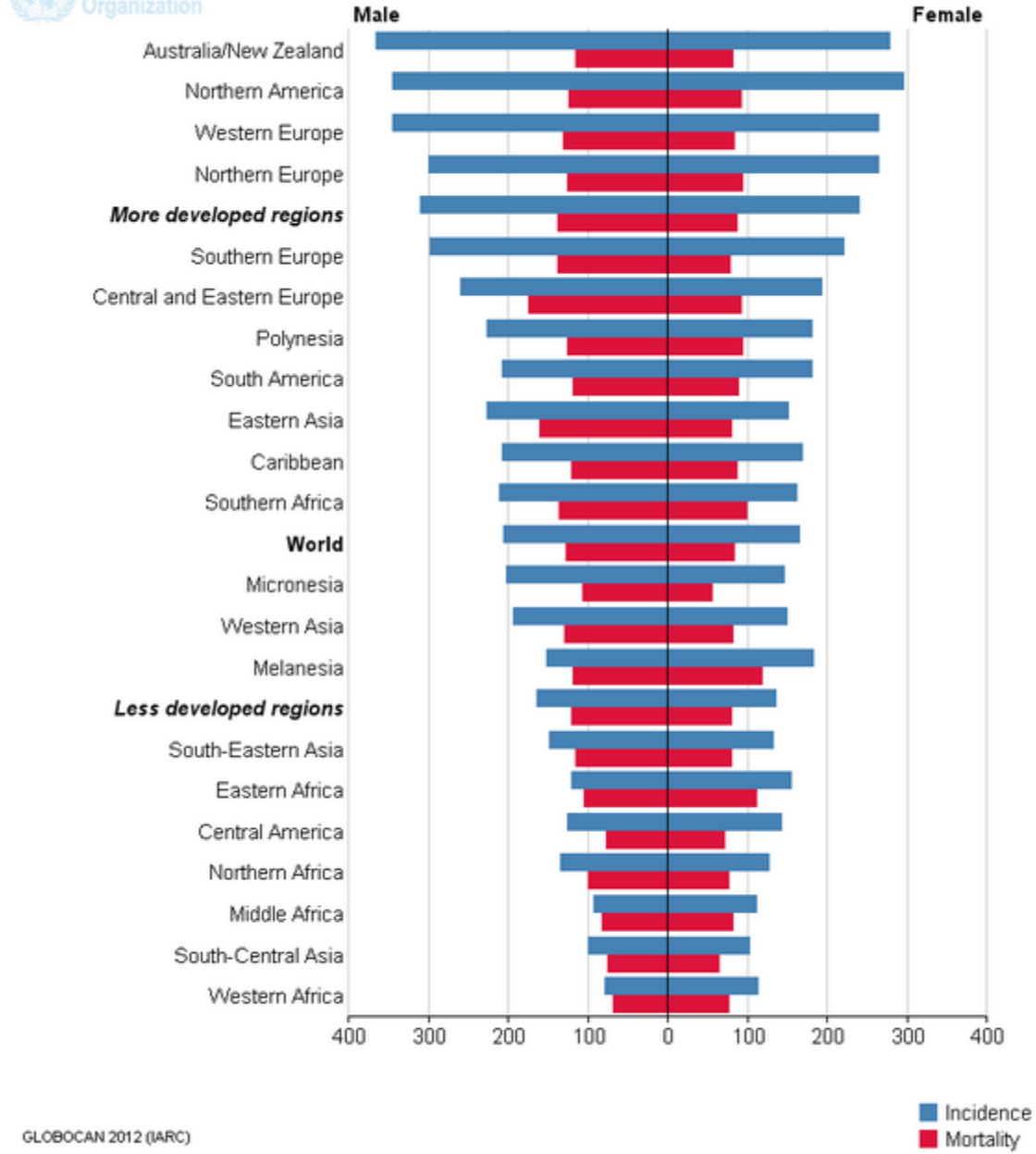
**University of Oxford**
Naomi Allen
Valerie Beral
Rory Collins
Richard Peto

# Most commonly diagnosed cancers by Human Development Index



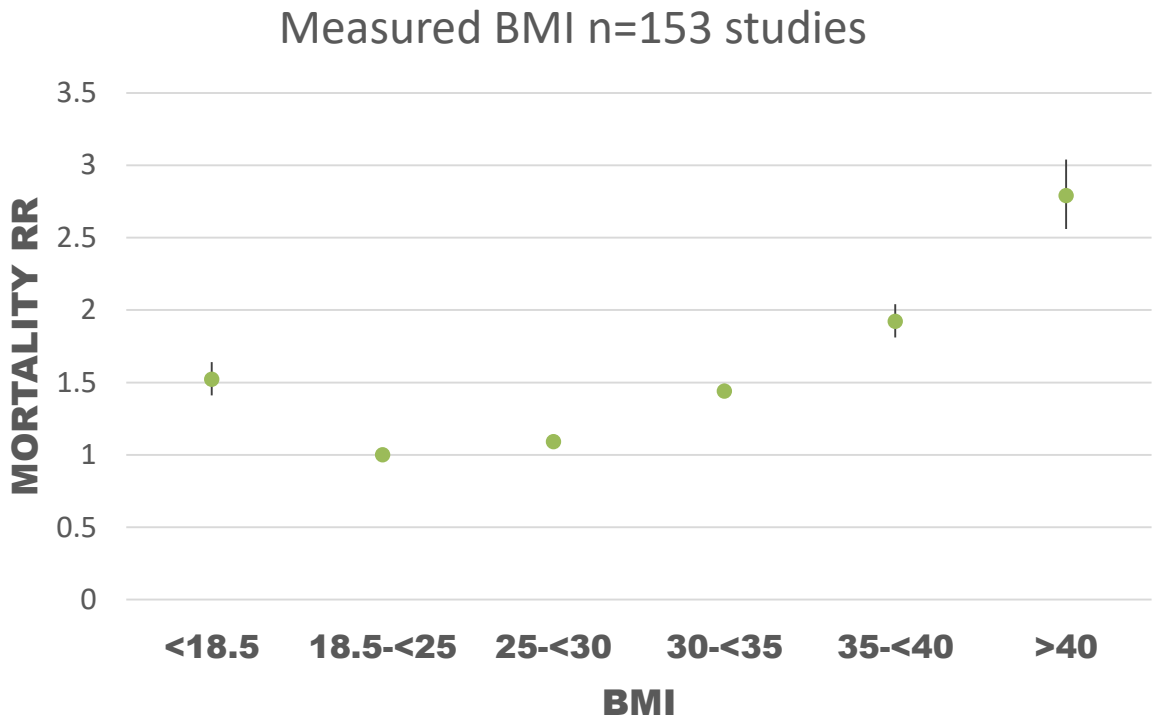

| HDI | % World Pop |
|---|---|
| Very High | 15% |
| High | 14% |
| Medium | 66% |
| Low | 6% |

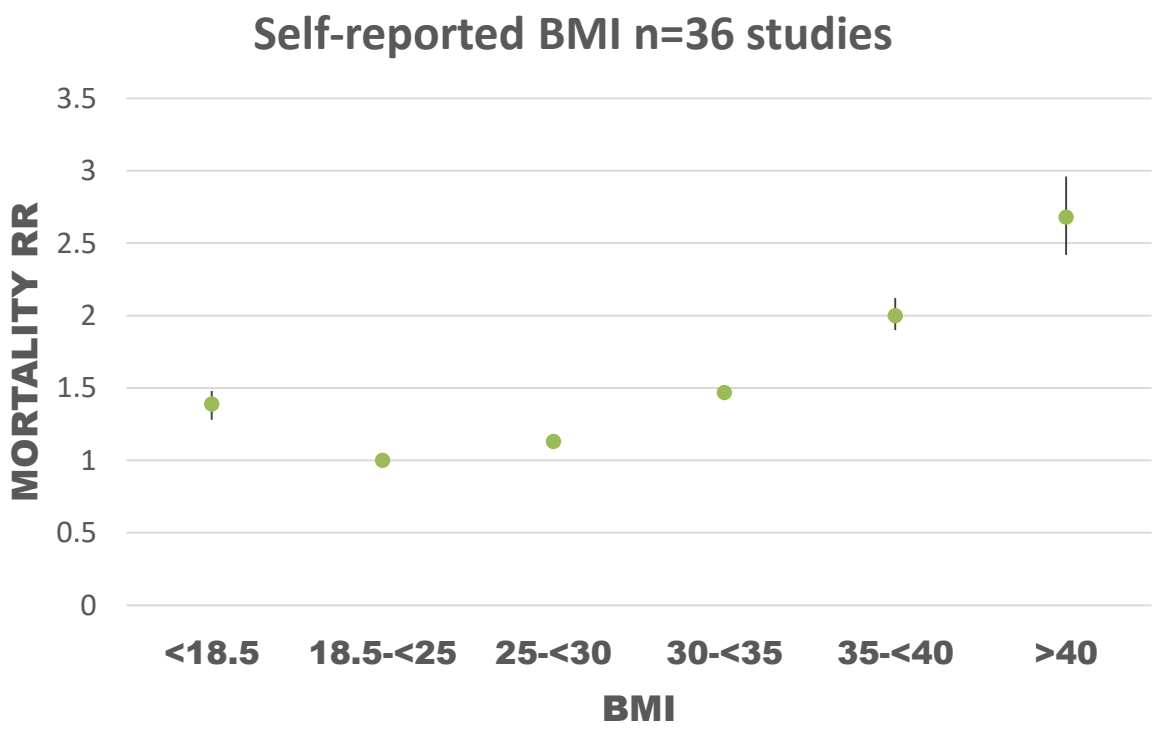
CRUK, 2012




GLOBOCAN 2012 (IARC)

Estimated age-standardised rates (World) per 100,000


Measured BMI n=153 studies
MORTALITY RR
3.5
3
2.5
2
1.5
1
0.5
0
<18.5
18.5-<25
25-<30
30-<35
35-<40
>40
BMI


Self-reported BMI n=36 studies
MORTALITY RR
3.5
3
2.5
2
1.5
1
0.5
0
<18.5
18.5-<25
25-<30
30-<35
35-<40
>40
BMI

# MORE BREAST CANCERS COULD BE PREVENTED IN HIGH RISK STRATA

Percentage preventable breast cancers by removal of modifiable risk-factors (overall and in categories of non-modifiable risk quintiles)

| | All Modifiable Factors Simultaneously | |
|---|---|---|
| | % Preventable | % Total |
| NonMod Risk Quintile 1 | 12.3 | 4.03 |
| NonMod Risk Quintile 2 | 16.0 | 5.23 |
| NonMod Risk Quintile 3 | 18.7 | 6.14 |
| NonMod Risk Quintile 4 | 22.4 | 7.34 |
| NonMod Risk Quintile 5 | 30.6 | 10.01 |
| Overall | 100.0 | 32.75 |

**Maas, Chatterjee et al. JAMA Oncol, in press**