Risk Adjustment

Can we balance pressures for payment, quality, and equity?

Todd Wagner, PhD

Todd Wagner, PhD

Acknowledgements

- No conflicts of interest

- Colleagues: Alex Sox Harris (VA Palo Alto & Stanford), Amy Rosen (Boston VA & BU), Michael Shwartz (BU), Megan Vanneman (SLC VA & Univ Utah), John Lin (Penn), Christine Chee (Stanford)

- Creation of the Nosos risk score was made possible by support from the Office of Reporting, Analysis, Performance, Improvement and Deployment and the Office of Productivity, Efficiency and Staffing

Outline

- Background on recurring challenges in the US health care system and the importance of risk adjustment

- Define risk adjustment

- Conceptual and measurement challenges

- The tension between payment, quality, and equity

Recurrent Themes over the past 30 years

- ## First is the financing of health care

- Physicians and hospitals want to get paid fairly

- Struggle to find payment methods that reward socially desirable outcomes
  - Capitation rewards providing too little care
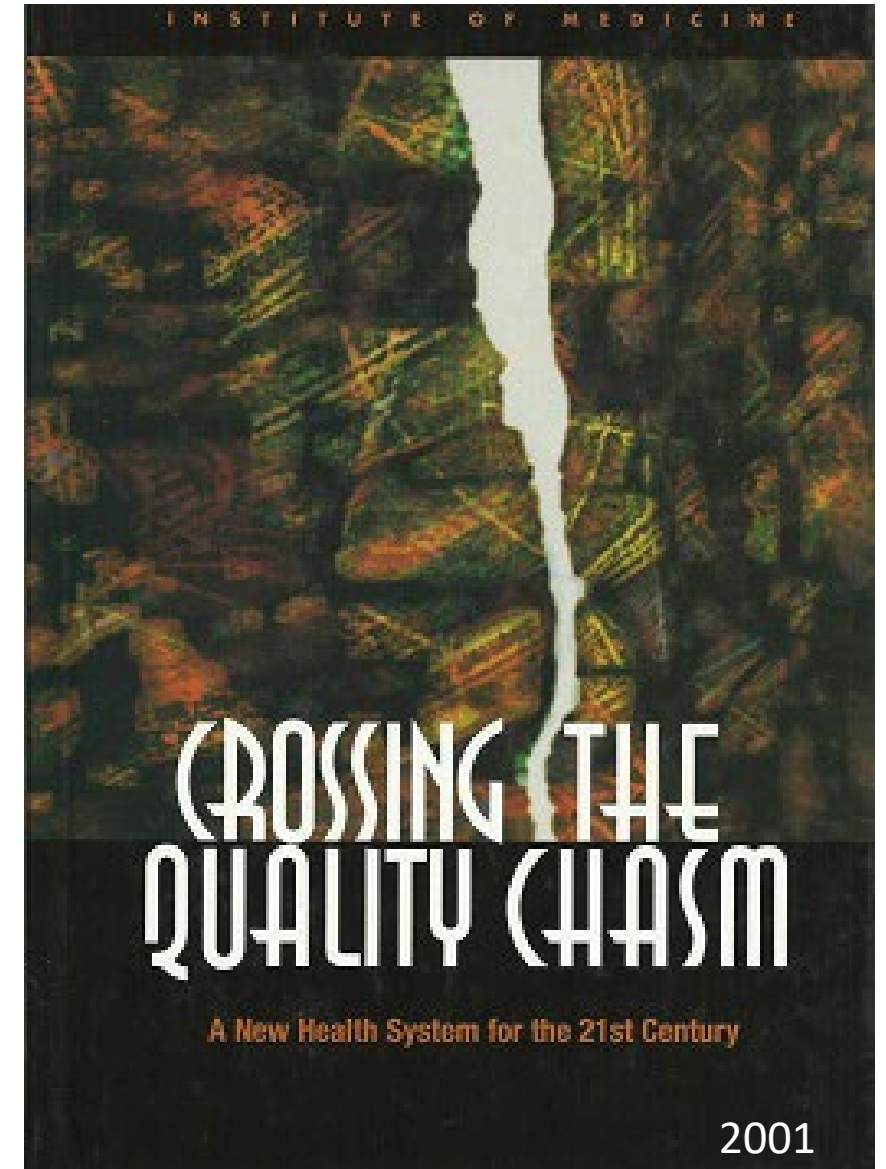 - Fee for service rewards providing too much care

FIGURE A.
U.S. Health-Care Expenditures as a Share of GDP, 1960–2018

Source: CMS 1960–2018.
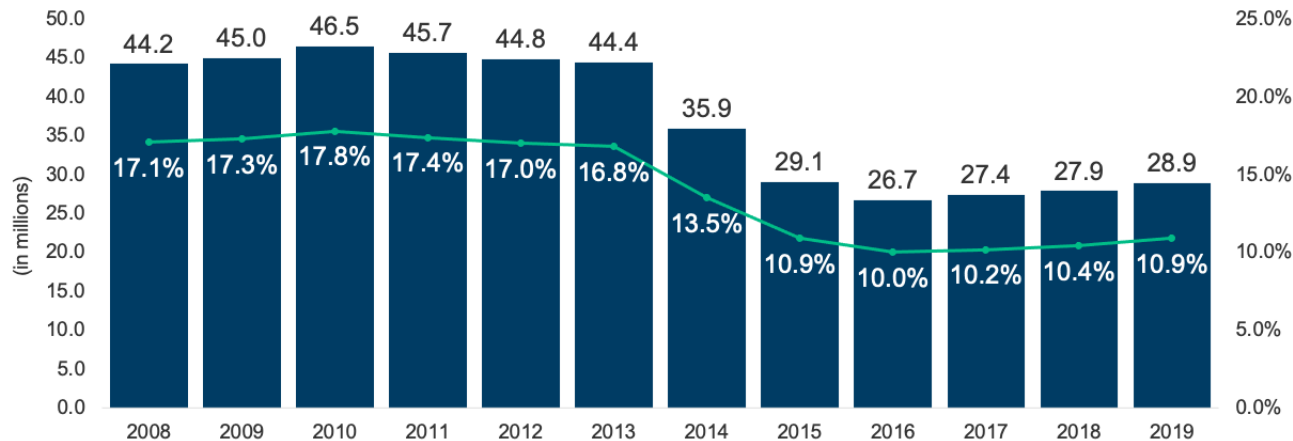
THE HAMILTON PROJECT
BROOKINGS
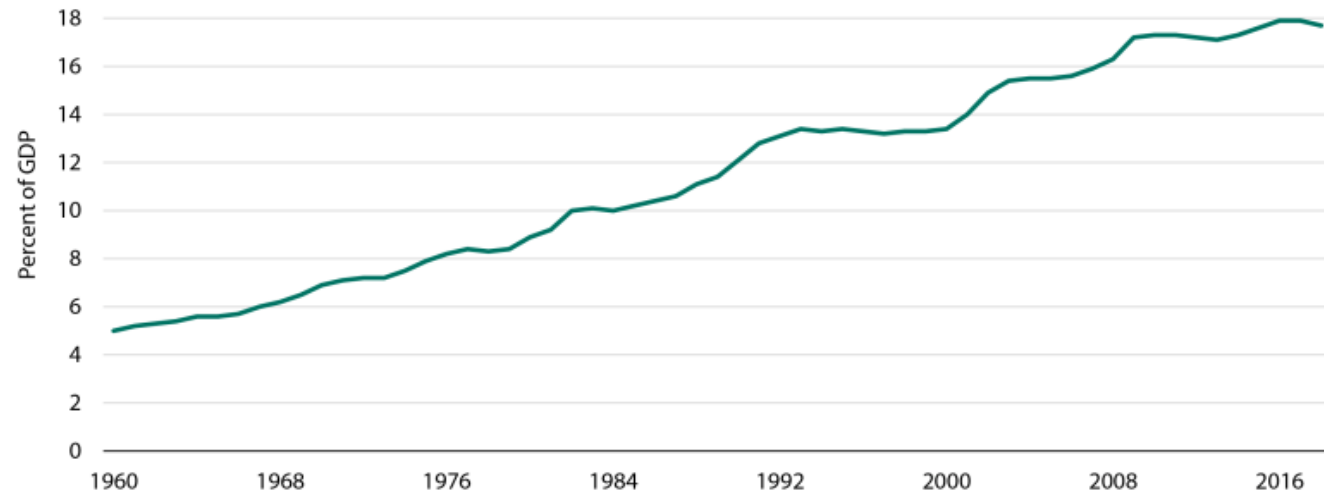
Second Theme: Quality Gaps

- Large and persistent quality gaps

- An entire field of study, implementation science, has developed to reduce these gaps.

INSTITUTE OF MEDICINE

CROSSING THE QUALITY CHASM

A New Health System for the 21st Century

2001

Equity

Figure 1

Number of Uninsured and Uninsured Rate among the Nonelderly Population, 2008-2019

NOTE: Includes nonelderly individuals ages 0 to 64.
SOURCE: KFF analysis of 2008-2019 American Community Survey, 1-Year Estimates.

KFF

- Insurance coverage is often used as a measure of equity

- Affordable Care Act decreased the number of those without insurance

- Growth of high deductible health plans
  - HDHP were associated with higher rates of harmfully delayed care: hernias → strangulated hernia.

Aligning Incentives

- Improving efficiency, quality and equity have proven to be persistently challenging.

- Well-known efforts
  - Prospective payment
  - Pay for performance
  - Accountable care organizations
  - Hospital readmission reduction program
  - Bundled payments

Why is alignment persistently difficult?

- Example: ICU Care

Aligning Incentives: ICU care

- Up to 50% of ICU admissions in the US are not for life sustaining care (don't belong in the ICU).

- Paradoxically, prior to COVID, many hospitals were expanding their number of ICU beds

Chen LM et al. Intensive care unit admitting patterns in the Veterans Affairs health care system. Arch Intern Med. 2012;172(16):1220-1226.
Chang DW et al. Priority Levels in Medical Intensive Care at an Academic Public Hospital. JAMA internal medicine. 2017;177(2):280-281.

Why are Hospitals Expanding ICUs?

Four reasons

1. **Heterogenous patients**: ICUs treat different types of patients
   - Those who need life sustaining care
   - Dying patients
   - Lower acuity patients where the clinician wants extra monitoring of vitals

2. **Imperfect information**: Hard to differentiate between heterogenous patients in administrative data.

3. **Marginal benefits >0**: There can be benefits for treating low acuity patients in an ICU.[1]

4. **Higher reimbursement**: Fee for service payments are higher for an ICU than a medical surgical bed.

1. Vranas KC, Jopling JK, Scott JY, Badawi O, Harhay MO, Slatore CG, Ramsey MC, Breslow MJ, Milstein AS, Kerlin MP. The Association of ICU with Outcomes of Patients at Low Risk of Dying. Critical Care Medicine. 2018 Mar;46(3):347.

Resource Constraints

- This wouldn't be a problem with unlimited resources.

- Average annual health insurance contribution in 2020
  - Avg Individual plans: $7,000
  - Avg Family plans: $13,700.

- How do we create health care systems that maximize quality, efficiency, and equity?

"It helps cut post-op recovery times, as well as health-care costs."

Data Driven Policy

- Federal agencies, including CMS, AHRQ, and VA are increasingly using administrative data to shape policies and programs

- Even the FDA is facing these questions given increases in low-risk devices (wearables)

- Correlational data are subject to all sorts of confounding

Risk Adjustment

- Because we aren't randomizing patients to health plans and providers, if we want to compare patients, we need risk adjustment to control for variation in patient sickness / acuity

- Risk adjustment creates incentives
  - Selecting low risk cases ("cherry picking")
  - Providing more more care over time
  - Upcoding[1]
  - Selective contracting

- Increasing awareness of equity beyond health insurance (i.e., social determinants of health)

1. Geruso M, Layton T. Upcoding: Evidence from Medicare on Squishy Risk Adjustment. Journal of Political Economy. 2019 Jun 5;128(3):984–1026.

Programs that Use Risk Adjustment

- Medicare Advantage (MA)

- The Affordable Care Act (ACA) exchanges

- CMS alternative payment models
  - Bundled Care Payment Initiative
  - Merit Based Incentive Payment System

- Hospital Compare

http://www.medicare.gov/hospitalcompare

	Stanford Health Care	Kaiser Foundation Hospital -...	Sequoia Hospital
	(650) 723-5708	(650) 299-2000	(650) 369-5811
Central line-associated bloodstream infections (CLABSI) in ICUs and select wards ↓ *Lower numbers are better* National benchmark: 1.000	**0.615** Better than the national benchmark	**1.836** No different than national benchmark	**2.203** No different than national benchmark
Catheter-associated urinary tract infections (CAUTI) in ICUs and select wards ↓ *Lower numbers are better* National benchmark: 1.000	**1.285** No different than national benchmark	**0.913** No different than national benchmark	**0.515** No different than national benchmark
Surgical site infections (SSI) from colon surgery ↓ *Lower numbers are better* National benchmark: 1.000	**1.241** No different than national benchmark	**0.618** No different than national benchmark	**2.455** No different than national benchmark
Surgical site infections (SSI) from abdominal hysterectomy ↓ *Lower numbers are better* National benchmark: 1.000	**1.356** No different than national benchmark	**Not available** [13]	**Not available** [13]

Almost impossible to find the methods behind these numbers, including risk adjustment

What is Risk Adjustment?

Risk Adjustment Goal

- Produce an estimate of central tendency that
  - Accurately reflects the average "risk" across different groups (healthy, sick).
  - Not easily gameable.
  - Is transparent.

- For costs or payments, averages are preferred over medians
  - This estimate should be influenced by the extremes, otherwise we create incentives for how we treat very sick (i.e., expensive) patients

- This risk score can then be used in statistical models to control for differential risk profiles across groups.

Risk Score Vs Co-morbidities

- Many studies control for specific illness using diagnostic groups with dummy variables

- Advantage of using actual risk score
  - The score means something (probability of death; expected cost)
  - Prior experience using the risk score (strengths and weaknesses)
  - Ease of use. Nosos is updated and available for patients on the VA CDW
  - Degrees of freedom (an issue for small studies)

Risk Scores Rely on Algorithms

- Most of my research focuses on risk adjustment for costs / payments.

- Many risk adjustment methods use regression models, to estimate the total costs.

- These risk adjustment models are often used for quality or outcome assessment.
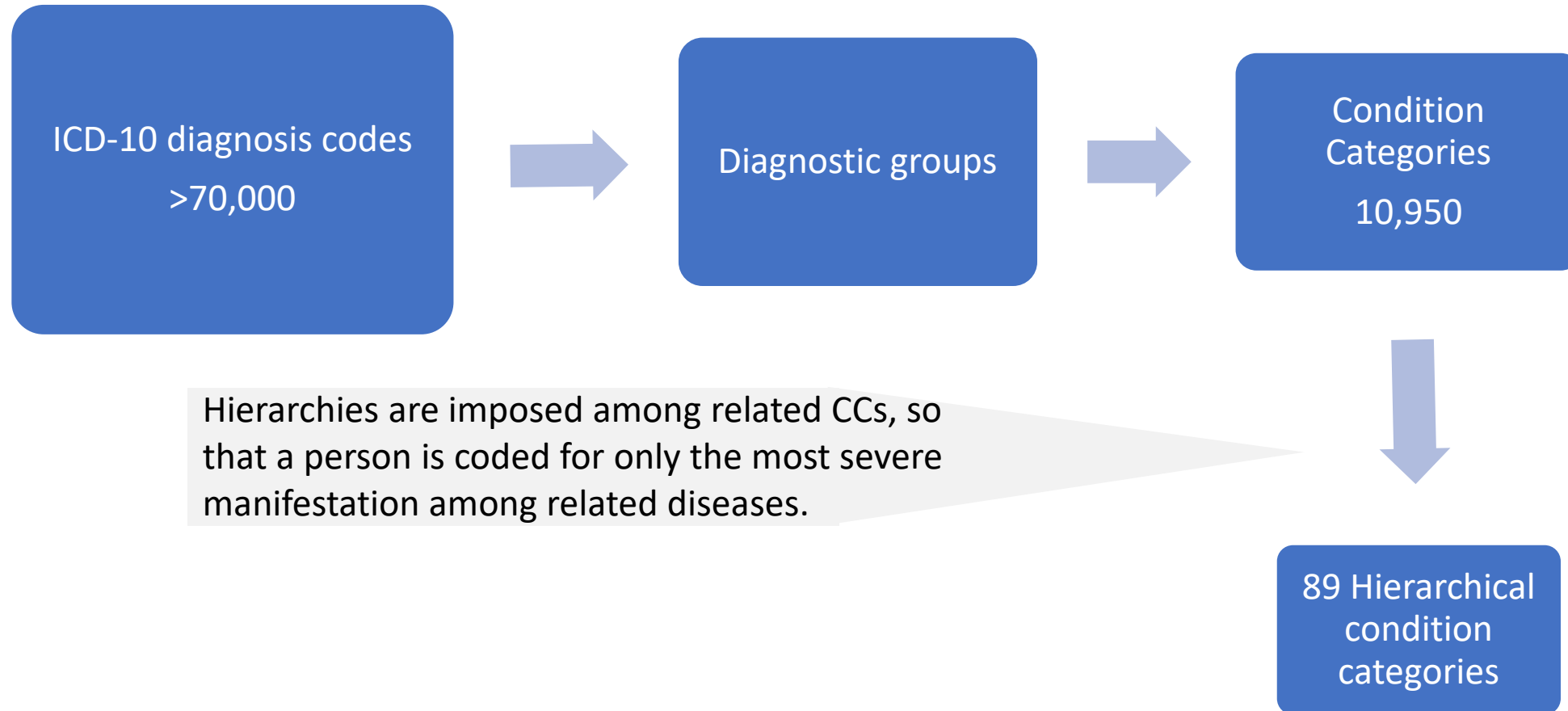
Risk Focus

- <u>Disease specific</u>: computed for a specific group
  - Examples: STS Short-Term Risk Calculator calculates a patient's risk of mortality and morbidities for cardiac surgery.

- <u>Generic</u>: models that estimate risk for a population
  - Hierarchical condition categories (HCCs)
  - Charlson co-morbidity Index
  - Elixhauser co-morbidity Index

- Generic measures are most appropriate when addressing questions of value or payment.

Medicare and Medicaid

- CMS relies uses hierarchical condition categories (HCCs) for risk adjustment
  - HCCs were created for Medicare Advantage
  - Exchanges and alternative payment models now also use the HCCs

- HCC model (V21-V24)
  - V21. 2009-2015: ICD-9 codes.
  - V22/23. 2016-2019: Updated V21 to work with ICD-10 codes
  - V24. 2020 onwards. Many more HCCs especially in mental health

- The HCCs are based on diagnostic information

Hierarchical Categorization

ICD-10 diagnosis codes >70,000 → Diagnostic groups → Condition Categories 10,950

Hierarchies are imposed among related CCs, so that a person is coded for only the most severe manifestation among related diseases.

89 Hierarchical condition categories

Medicare Risk Score

- CMS uses a simple linear regression model to estimate the relative contribution of each HCC

$$\text{Cost} = \alpha + \sum HCCs + age + sex + age * sex + \varepsilon$$

- Calibration: Risk score= predicted cost from regression / average spending for Medicare ~$9300
  - 1= average risk, >1 is higher risk, <1 is lower risk

- CMS produces software that allows plans to enter diagnostic information and obtain a risk score

Timing

- <u>Prospective risk</u>. Using data from this year to estimate risk in the next year (Medicare Advantage uses this for payments).

- <u>Concurrent risk</u>. Using data this year to estimate fair payments or penalties for this year (MIPS, Bundled payments)

Prospective Risk

- Many people are familiar with prospective risk models.

Risk Score based on prior data

Surgical Intervention

Develops CKD

Develops Depression
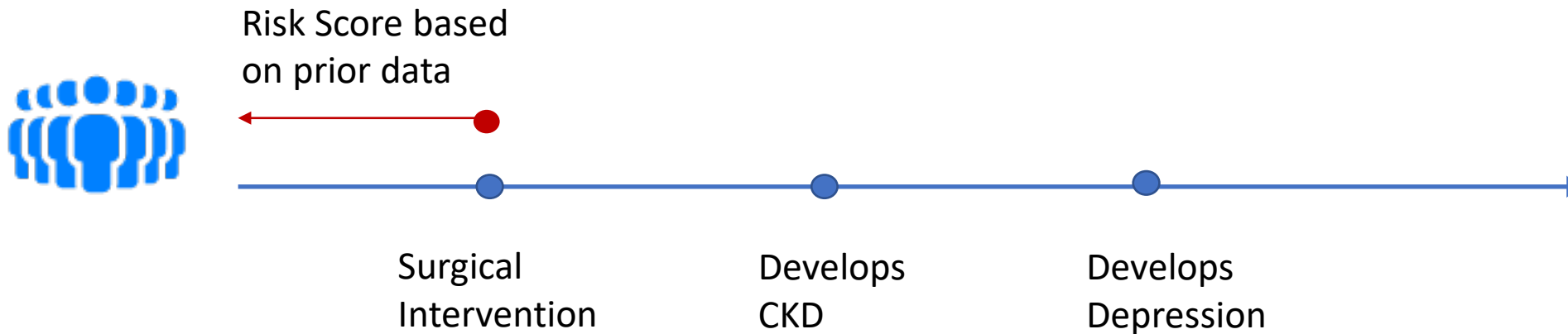
CKD and Depression would not be included in the prospective risk score estimated at the time of surgery
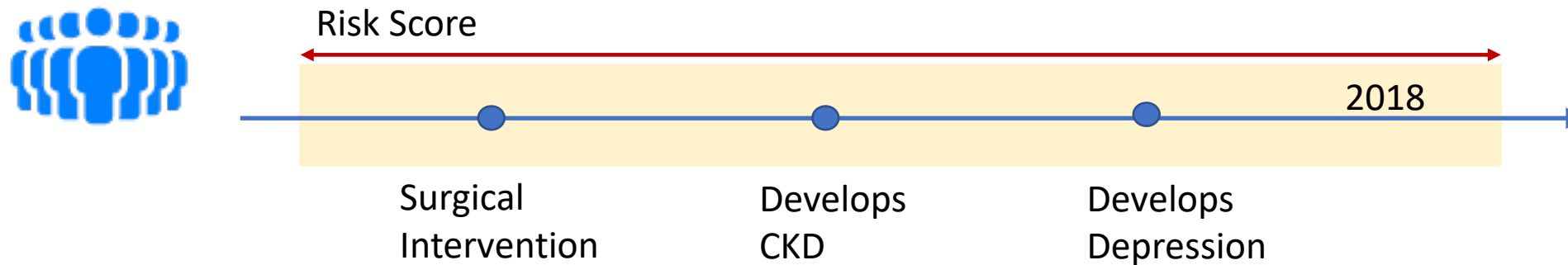
Prospective Risk Models

- Prospective Risk is used in establishing payments for Medicare Advantage→ pay health plans this year based on the patient's risk profile from the prior year

- But <u>generic</u> prospective risk scores are less commonly used in broad studies of payment, quality and equity
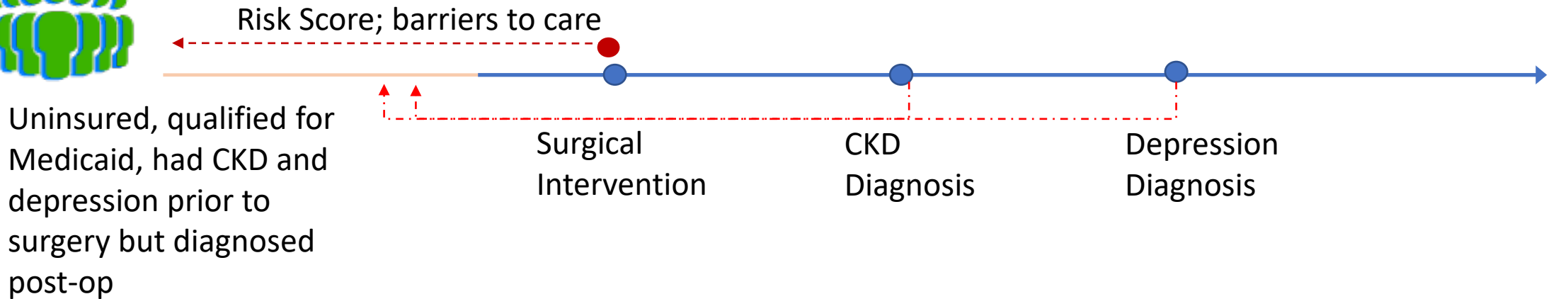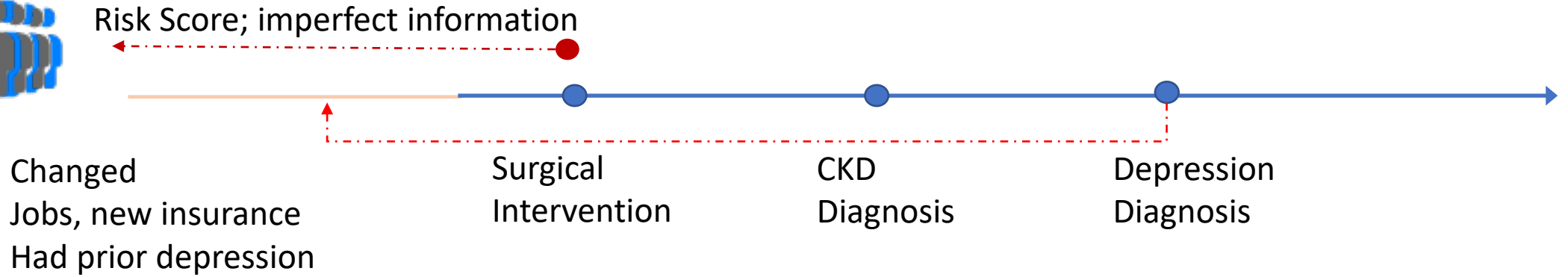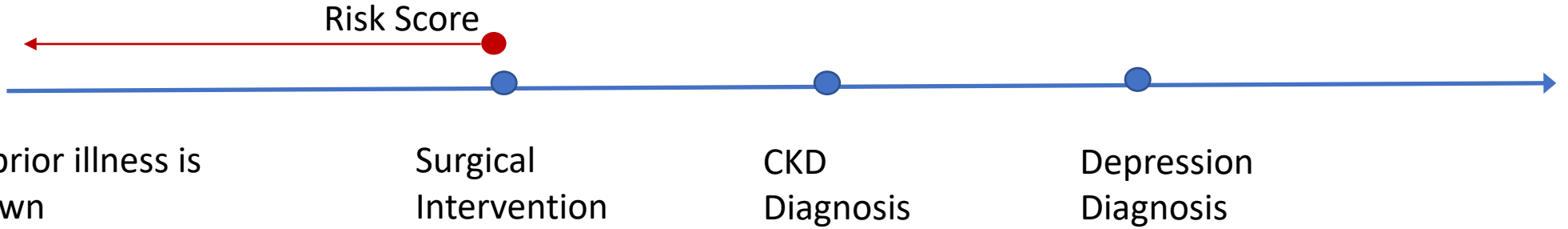 - Partly a result of the "timing problem."
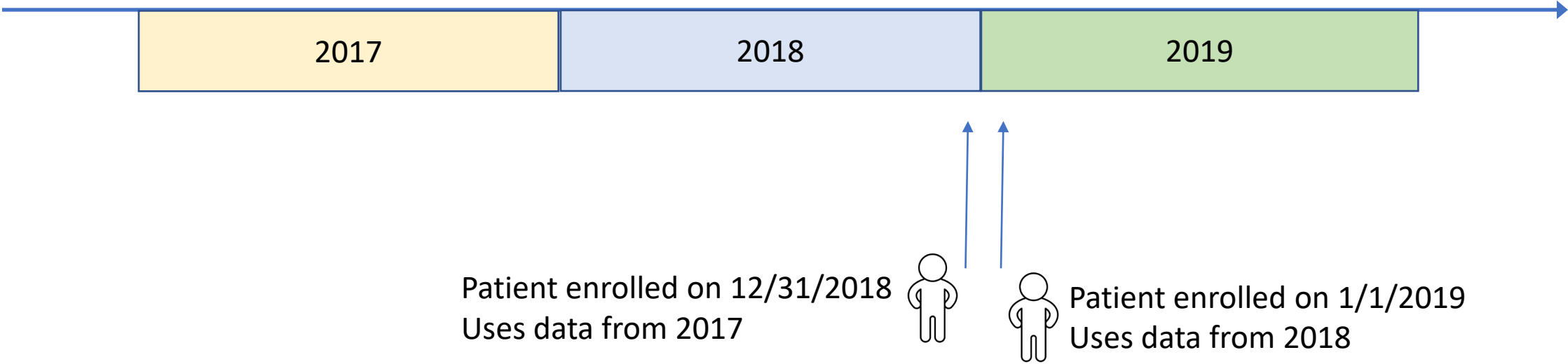
Concurrent Risk

- Risk score is based on information from the entire year
- This method ignores the temporal sequencing. You could risk adjust the surgical intervention for care that happens after the surgery

Risk Score

2018

Surgical
Intervention

Develops
CKD

Develops
Depression
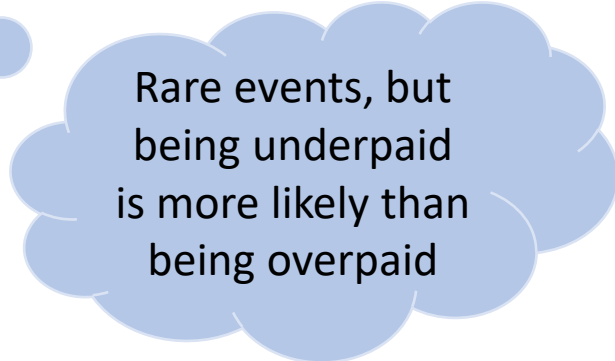
- This has led some to conclude that concurrent risk is biased

Risk Score

All prior illness is known

Surgical Intervention

CKD Diagnosis

Depression Diagnosis

Risk Score; imperfect information

Changed Jobs, new insurance Had prior depression

Surgical Intervention

CKD Diagnosis

Depression Diagnosis

Risk Score; barriers to care

Uninsured, qualified for Medicaid, had CKD and depression prior to surgery but diagnosed post-op

Surgical Intervention

CKD Diagnosis

Depression Diagnosis

These Models are Imperfect
But they Matter

- McGuire et al found imbalance in the extremes in the US, Germany and Netherlands

- In the United States exchanges, in 2017
  - one in 1,000 enrollees were underpaid by > $190,000
  - one in 1,000 enrollees were overpaid > $95,000

Rare events, but being underpaid is more likely than being overpaid

McGuire, T.G., Schillo, S. & van Kleef, R.C. Very high and low residual spenders in private health insurance markets: Germany, The Netherlands and the U.S. Marketplaces. *Eur J Health Econ* **22,** 35–50

Are You Risk Adverse?

- Physicians / insurers seek to manage risk
  - Risk pooling
  - Selective contracting
  - Accepting patients / transfers

- Physicians should care about risk adjustment
  - Poor risk adjustment increases each physician's gamble
  - Over reliance on diagnostic codes that do not discriminate at the extremes
  - Over reliance on simple statistical models that are transparent but less robust
  - Concerns about whether the risk score creates incentives for gaming

Model & Model Fit

- New statistical models and variables (e.g., pharmacy or functioning) make this is interesting area of study

- Multiple gauges of statistical success
  - Area under the curve (AUC)
  - R^2
  - MSAE
  - Calibration curves, Brier scores / Hosmer Lemeshow deciles (predictive accuracy)

- Often R^2 is reported, but these statistics provide different information under different circumstances.

The Allure of Measurement

Measurement issues often take center stage

Conceptual issues are often downplayed.
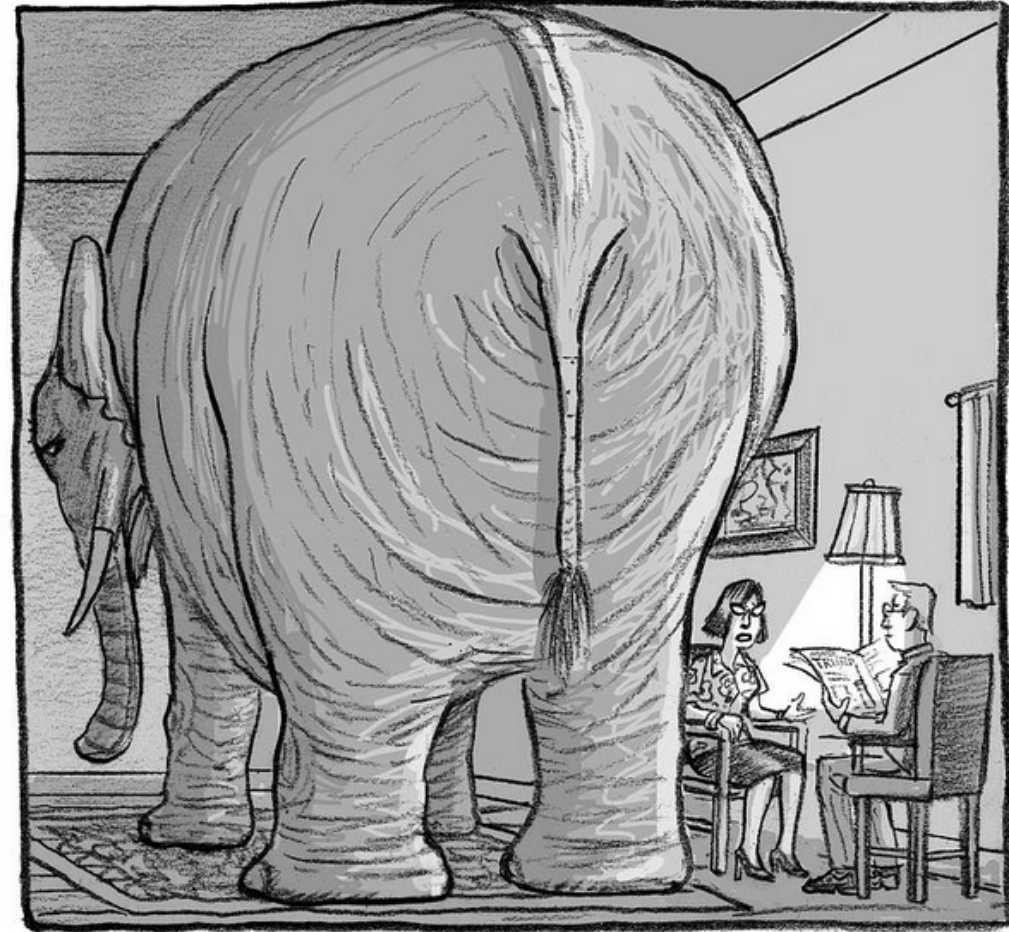
"*Actually, it's all we ever talk about.*"

Risk Adjustment for VA

- In 2015, VA was spending > $200,000 a year for commercial risk adjustment software.

- We embarked on a study to determine whether we could create open-source software to save the VA money.

- We started the CMS HCC V21 model and ran analyses with 6 samples

Wagner TH, Upadhyay A, Cowgill E, Stefos T, Moran E, Asch SM, Almenoff P. Risk Adjustment Tools for Learning Health Systems: A Comparison of Dx CG and CMS-HCC V21. Health services research. 2016 Oct;51(5):2002-19.

Initially We Struggled to Match DxCG

	V21 Risk	DxCG
R^2		
General	0.5793	0.6274
Older	0.5728	0.6233
MH-SUD	0.5820	0.6268
High cost	0.3559	0.4244
Multimorbid	0.5350	0.5957
Low risk	0.2922	0.3508

We Succeeded with Three Changes

Pharmacy data

- CMS V21 did not include pharmacy information

- We created 25 groups based on drug class

Mental health

- CMS V21 had 4 mental health & substance use groups

- We created 62 groups (PsyCMS)

More sophisticated statistical models

- CMS V21 used linear regression

- We used square root transformed model to address extreme high costs

This risk model, known as Nosos, is now produced quarterly for all VA users.

Nosos

- The resulting risk score is Nosos-- Greek for chronic condition

- Computed annually and quarterly
  - Use the annual risk scores, if you have a choice
  - Quarterly scores based on spending projections

- We've updated it over time.
  - FY20, we switched to HCC V24
  - Also include Fee and PIT data

- Partnership between HERC and OPES. Updated technical documentation are on the HERC website.

Follow-up Studies

- Does the risk model fit improve when you add more clinical history?

  - Answer: Surprisingly no

  - Adding 2 or 3 years of clinical history provided almost no additional diagnostic leverage

Lin JK, Hong J, Phibbs C, Almenoff P, Wagner T. Evaluating the Role of Past Clinical Information on Risk Adjustment. Medical Care. 2020;58(2):169.

High Risk Patients

New Yorker, 2011

- Given that 5% of patients are responsible for 50% of costs, can you predict high risk patients?

  - Answer: Not easily

  - Among patients with the highest 10 percent of costs at baseline:
    - 68 percent did not remain high cost in subsequent years
    - High mortality largely explained low persistence
    - High-costs were most persistent among people with a spinal cord injury (16 percent)

Yoon J, Chee CP, Su P, Almenoff P, Zulman DM, Wagner TH. Persistence of high health care costs among VA patients. Health Services Research. 2018 Oct;53(5):3898-916.

Is risk score fair for safety net providers?

- We examined V21 in the VA
  - CMS V21 identified 4 mental health and substance use HCCs'
  - CMS V21 model identified 694,706 as having a mental health or substance use condition.

- Of the 5,472,629 VA patients
  - Using another psychiatric comorbidities score, we identified another 1,266,938 patients with a mental health condition.
    - Depression NOS: 396,062 (31.3%)
    - Posttraumatic stress disorder: 345,338 (27.3%)
    - Anxiety: 129,808 (10.2%).

- Overall, the V21 model underestimated the cost of care by $2,314 for every person with a mental health diagnosis.

Wagner TH, Almenoff P, Francis J, Jacobs JC, Chee CP. Assessment of the Medicare Advantage Risk Adjustment Model for Measuring Veterans Affairs Hospital Performance. JAMA Network Open. 2018;1(8):e185993-.

Conceptual Issues are Re-emerging

Conceptual issues

Measurement issues

Photo Credit: Wildflower Schools

Conceptual Challenges

- What variables are missing from our risk adjustment algorithms?

- Should we control for those factors?

- All risk algorithm include age and sex. But what about homelessness, access to education or healthy food?

VA Purchased Care

- VA is purchasing more care, and the risk matters.

	VA		Community Care		
	N	Mean	N	Mean	P value
TKA					
Nosos score	6,293	3.04	7,357	1.50	<.001
Cataracts					
Nosos score	65,799	1.90	25,342	1.44	<.001

Is the playing field truly level?

- Selective contracting with providers who see vulnerable populations[1]

- Equity, structural racism and social determinants of health[2]

1. Shepard M. Hospital Network Competition and Adverse Selection. National Bureau of Economic Research; 2016 Sep No.: w22600.
2. Nerenz DR, et al. Adjusting Quality Measures For Social Risk Factors Can Promote Equity In Health Care. Health Affairs. 2021 Apr 1;40(4):637-44

Should We Include Social Determinants?

Social Determinants of Health

Education Access and Quality

Health Care Access and Quality

Economic Stability

Neighborhood and Built Environment

Social and Community Context

Social Determinants of Health
Copyright-free

Healthy People 2030
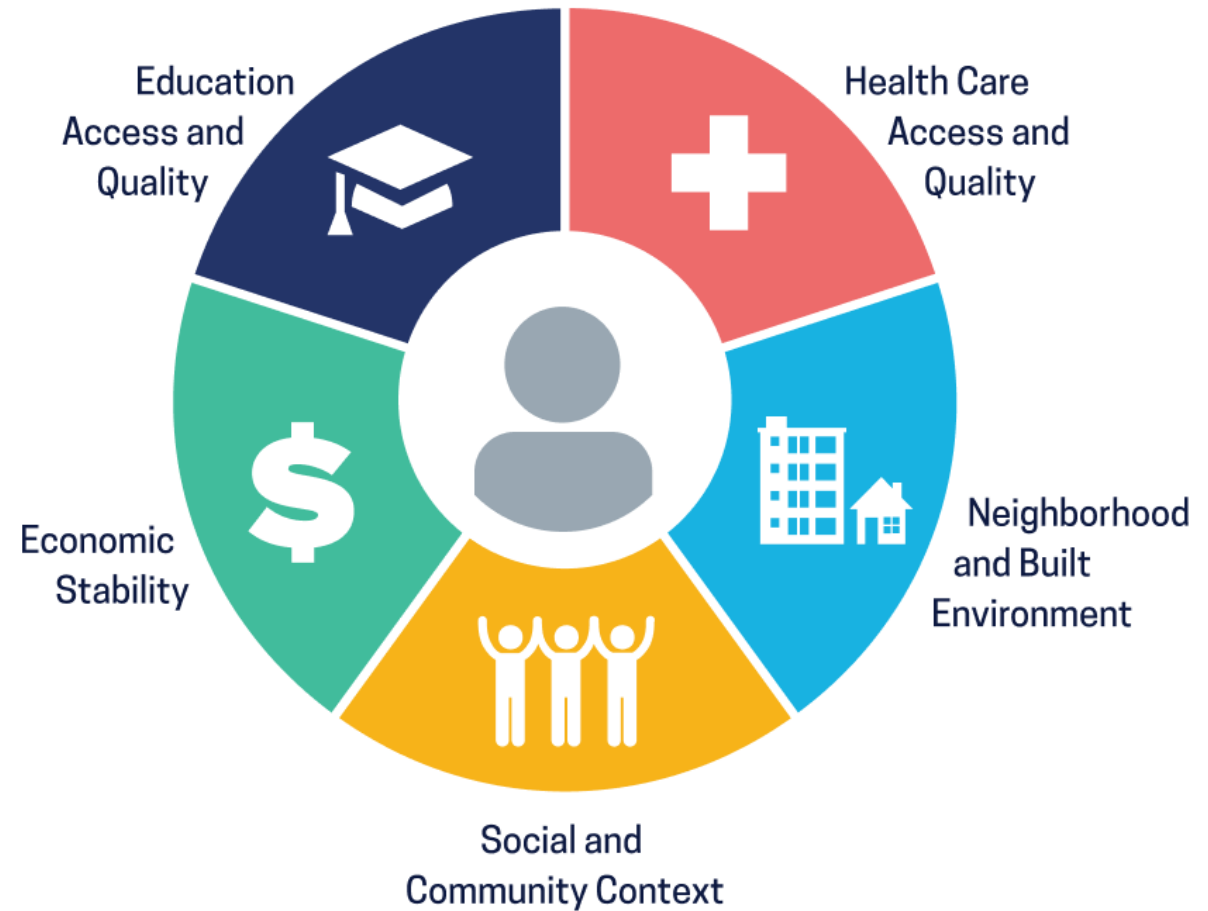
Housing

- Considerable literature identifying homelessness and housing instability as a risk factor for poorer health

  - Mental health: suicide, depression, substance use

  - Health behaviors: cigarette smoking a chronic stress, depression,

  - Diagnosis: logistical challenges to may lead to delays

  - Treatment: Barriers to medication adherence, communication, and timely follow-up.

Competing Demands

- Measurable: can you measure the risk factor accurately and precisely?
  - Homelessness is hard to measure.

- Coding and "gaming": does including the risk factor create perverse incentives that could result in gaming?
  - Including homelessness probably won't lead to gaming.

- Financing:
  - We don't want to finance housing through health care (too expensive).
  - A growing concern that value-based payments are creating "wrong pockets" problem

Fixing the Wrong Pockets Problem in Healthcare

Recommendations

- More research– a number of recently funded grants and more under review

- Greater transparency– make it easy for readers to understand what we are controlling in data displays.

- Consider separate risk systems for performance measurement and payment.

Thanks

- Questions?

- twagner@stanford.edu or todd.wagner@va.gov