

The Power of Multiple Data Streams and Big Data in Health and Health Care

Betsy Shenkman, PhD

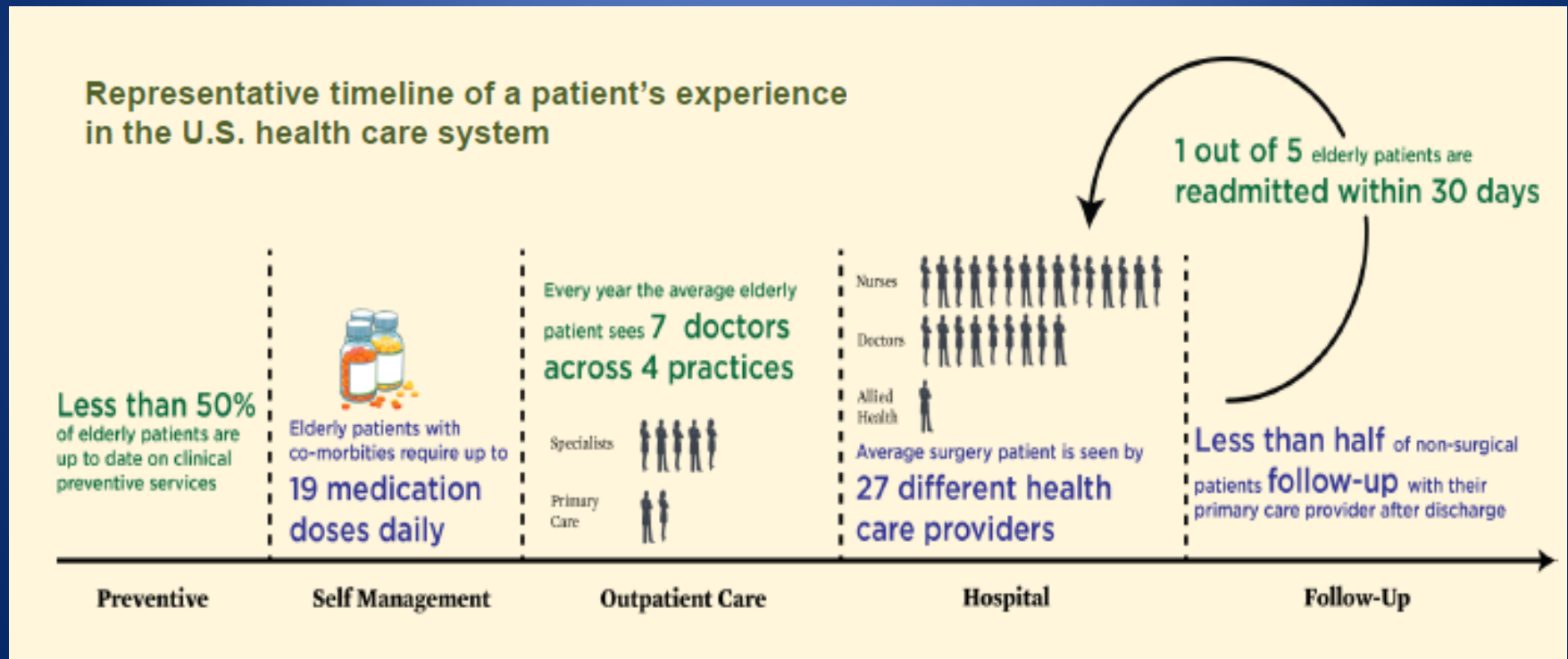
Department of Health Outcomes and
Policy, College of Medicine
Institute for Child Health Policy



"This work supported in part by the NIH/NCATS Clinical and Translational Science Award to the University of Florida UL1 RR029890."

What We Do - Example: Use of Large Datasets to:

- Identify patients at risk for poor quality of care
- Link data across providers and settings to reduce error and improve care
- Conduct comparative effectiveness research
- Engage patients in their health care through sharing information on web portals and other strategies



Health Outcomes and Policy Data Repository

- 4 TB data received annually
- 300-350 Million claims/encounter records/Yr
 - Florida and Texas - Medicaid, CHIP
 - Health care claims and encounter data
 - Dxs, services, paid amount, rendering provider, location of service
 - Pharmacy
 - Lab data
 - Electronic health record
- 25 TB of healthcare data spanning 10 years
- Linked to patient-reported outcomes of care
- All geocoded to census tract
- Receiving Medicare data
- Health care data from the commercially insured has same characteristics and is being received from selected private practices in Florida

How Do We Use the Data?

Clinical Effectiveness Research

- Conduct of community trials, observational studies and natural experiments
 - **Ecological validity**
 - Is the care clinically effective under conditions where the care will be delivered?
 - **Assessment of Heterogeneity of Treatment Effects**
 - What works best for whom, under what conditions?
 - **Longitudinal design**
 - How long are effects observed or is exposure long enough to observe effects?
 - **Large Data Streams and the Power of Numbers**
 - Integrating data from multiple sources

How Do We Use the Data?

Randomized Community Trial Example

- The Wellness Incentives and Navigation Project
 - Funded: Centers for Medicare and Medicaid Innovation (ACA) - \$10M over 4 years
 - Goal - To examine the effects of a personal navigator and a flexible wellness account on improving wellness behavior and health among adults 21-55 years old in Medicaid Managed Care who have comorbid physical and mental health conditions
 - Our “Big Data” used to identify the sample for randomization, track their health care use and expenditures , and biomarkers longitudinally

How Do We Use the Data?

Information Dissemination

- Texas healthCare Learning Collaborative uses the “big data”
- Interactive portal
- 300,000 pre-computed tables to respond to queries about health care use and expenditures
- Purpose
 - Identify areas for future research
 - Provide applied tools for health care providers, policy makers

Big Challenges and Growth

- Increasing recognition of limits of traditional randomized trials in identifying most efficacious health care interventions
 - Need to randomized community-based trials and observational studies with large populations in real world settings – need to link large datasets from multiple sources
- Increasing need to test novel methods to transmit information to patients and providers to engage patients in their care and reduce costs
- Expect continued growth and demand due to high cost of health care, aging society, and increasing numbers of people with chronic conditions

Infrastructure Needs

- Increase team science (examples)
 - Linkages with College of Engineering (Human-Machine Interaction)
 - Linkages with College of Journalism (Structuring how health care information is presented to diverse audiences)
 - Linkages to clinical scientists, population scientists
- Increased professional staff
 - Programmers, programmers, programmers