Day Four: 66Daysof(Health)Data

1. **Define terms: face validity, content validity, construct validity, criterion validity**

   Validity: refers to the accuracy of the data

   When assessing, look for the ***extent to which data elements measure the quantity they are intended to measure***; often defined in the context of surveys or diagnostic tests.

   Face validity: expert assessment to determine if the metric appears to measure the quantity it was intended to measure; assessed by a subject matter expert (SME).

   > Example, pg. 20
   >
   > *Can you use timestamps in a clinic EHR to measure the amount of time a patient waits in the exam room before seeing the physician?* After consulting the clinic manager, you discover that the staff only regularly record check in and check out time, and not rooming time. Because of this, you discover that you cannot use time stamps to measure wait time.

   Construct validity: measured for survey and assessment tools; patient satisfaction surveys (HCAHPS) for example; *does the instrument measure the constructs or concepts of satisfaction that they are designed to measure*? Tested by asking questions that measure a similar concept.

   > Example, p. 20-21
   >
   > HCAHPS includes 21 questions measuring 9 constructs: communication with docs, communication with nurses, responsiveness of staff, pain management, communication about medications, discharge info, cleanliness, quietness, and transition of care. *If responses to questions within a construct (eg. communication with docs) are strongly associated with each other, the instrument displays construct validity in measuring that area*. Statistical tools: factor and component analysis, can be used to test for or confirm this

   Criterion validity: measurement of the agreement of a test or data element with a known gold standard. eg) a new diagnostic test results in the same conclusion as an accepted standard test, means the new test has strong

criterion validity. Another example: using the # of chronic conditions as a proxy to measure resource-intensiveness of patients -> this is a well-established relationship based on prior research.

> Example, p. 21
>
> O'Driscoll et al., 2020 studied validity of activity monitors as est. of energy expenditure comparing Fitbits with lab devices. *Used the lab devices as gold standard, and checked the agreement of commercially available devices such as FItbits as an assessment of criterion validity*, and found there was quite a bit of variation in agreement between commercial and gold-standard lab devices depending on the activity performed.

Reliability: a measure of the *repeatability or reproducibility of the results of a measurement*. **Three types**: inter-rater & intra-rater both measure the consistency of the data over time (measured using Kappa statistic or Cohen's kappa - measuring the rate of agreement between raters or scales), and internal consistency (Chronbach's alpha may be used to measure the strength of internal consistency among survey questions)

Kappa/Cohen's Kappa, indicative of rater reliability:

> 0.41-0.60 - moderate
>
> 0.61-0.80 - substantial
>
> 0.81-1.00 - almost perfect

Calculating Cohen's Kappa

1. Calculate the proportion of agreement between raters
    a. po = (Both said Yes + Both said No) / (Total Ratings)
2. Calculate the hypothetical probability of chance agreement (pe) between raters
    a. This is calculated as the total number of times that Rater 1 said "Yes" divided by the total number of responses, multiplied by the total number of times that Rater 2 said "Yes" divided by the total number of responses, added to the total number of times that Rater 1 said "No" multiplied by the total number of times that Rater 2 said "No."
3. use po and pe to calculate Cohen's Kappa: k = (po – pe) / (1 – pe)

Chronbach's alpha, indicative of internal consistency:

Values close to 1 - high reliability

Calculating Chronbach's alpha

1. Python:

```
import pingouin as pg

pg.cronbach_alpha(data=df)

(0.7734375, array([0.336, 0.939]))
```

2. Interpretation and presentation: "A satisfaction survey was sent to 200 customers. The survey consisted of 12 items and the value for Cronbach's Alpha for the survey was α = .84."

   The minimum recommended N for calculation of this statistic is 20.

2. Explore types of healthcare data and data coding systems

Electronic Health Record Data

Diagnostic Data -

1. ICD-10 codes
2. CMS Inpatient prospective payment system (IPPS)
   a. ICD-10-CM/PCS used to compute the MS-DRG (Medicare severity diagnosis-related group)
   b. MS-DRGs fall into larger MDC (Major Diagnostic Category) groups
   c. Both MS-DRGs and MDCs are *assigned using a grouper* based on codes and demographics of inpatient cases

Procedure Data -

1. ICD-10-PCS, code inpatient procedures; adopted 2015
2. Seven-character codes
   a. 1st character: section
   b. 2nd character: body system
   c. 3rd character: operation
   d. 4th character: actual body part
   e. 5th character: approach
   f. 6th character: device (Z indicates 'none')
   g. 7th character: qualifier (Z indicates 'none')

Example: Destruction of the right kidney, open approach: 0T500ZZ

3. CPT “Current Procedural Terminology” - used for outpatient and physician services, 5-digit codes, maintained by the AMA. Visit levels correspond to the amount of work, complexity of case with level 1 being a very basic visit, to level 5 a long, complex visit or higher acuity
4. HCPCS (Hic Picks) “Healthcare Common Procedure Coding System”, five-digits, for Medicare beneficiaries drugs and supplies, drug codes begin with J.
5. OPPS “Outpatient Prospective Payment System” using ambulatory payment classification groups (APCs) to group hospital outpatient services on the CPT or HCPCS codes assigned to that outpatient case. A claim or encounter may have multiple APCs but only 1 MS-DRG.

Pharmacy Data -

1. NDC “National Drug Codes”, describe drugs, indicate:
    a. Size of package
    b. Dosage formulation
    c. Drug name (generic and brand)
    d. Manufacturer
    e. Standardized by NCPDP, “National Council on Prescription Drugs Program”
    f. 11 digits
    g. 1-5 describe the manufacturer, 6-9 drug name, 10 & 11 package size and type
    h. Available via FDA website
2. Grouped into more-easily analyzed groups by the American Hospital Formulary Service (AHFS)’s Therapeutic Classification grouping system
    a. Categorizes drug products using six-digit #s
    b. Eg. 68:00 hormones, 68:20 anti-diabetic agents, 68:20:08 insulin (easier than all the specific NDCs for every type of insulin)
3. RxNorm - a naming system for prescription and OTC drugs produced by the National Library of Medicine (NLM) to allow providers to standardize the naming conventions and identifiers for drugs (less common)

Administrative Data -

1. Revenue code - a four-digit code that categorizes charges on the Uniform Bill-04 (UB-04) claim or 8371 electronic file associated with the claim.
2. Claims data now submitted using 5010 format.
3. Revenue codes stand for standardized departments in facilities, needed for each charge line when billing Medicare
4. Tracks Medicare costs by revenue centers for the Medicare cost report and categorizes revenue by similar types of services such as radiology or lab
5. Place of Service (POS) codes; indicate the location of health care services. Eg) 01 Pharmacy, 02 Telehealth, 09 Prison, etc.

3. Come up with at least 3 ways these data could be used in an analysis

I've used diagnostic and procedure data in analyses before, but I think it would be fascinating to try to use HCPCS data in an analysis. Does this data allow one to track claims related to a particular drug or class of drugs? A model created with this data might allow exploration of the relationship between adherence and various outcomes for Medicare beneficiaries with certain chronic conditions where condition maintenance relies heavily on treatment adherence, such as diabetes or depression.

I also think efforts to track utilization are too fragmented. Perhaps we are looking at re-admissions, or we are looking at high risk for inpatient stays, but are we looking across the spectrum at all utilization (both outpatient and inpatient) to see what are the differences in characteristics and/or behavior between those who are able to manage their conditions outpatient vs. those who can't? Is this a severity issue, or is it an issue of too many conditions? Do those with intensive case management do better, or is it really more about family support?

Place of service data seems like it could be helpful to CMS to see where care is being provided and what are rates of hospitalization, re-admission, or overall utilization as place of service varies. Are there places of service that are doing better or doing worse with these metrics? We have been having the argument in healthcare analysis for a long time over whether a given hospital or system has a higher burden of care, and thus should be "off the hook" for worse outcomes, but I think this is asking the wrong question. There must be safety net hospitals who have at least slightly better outcomes even with similar care burdens. How are these improvements being achieved, and how can we ensure that they can make their way to other contexts? How can we come together to get the best care possible for patients, in light of the circumstances, and what is standing in the way?