Inter-Measure Agreement Among Metrics for Low-Value Imaging for Low Back Pain: The Devil is in the Exclusions

James Henderson, PhD[1,2], Katherine Wilkinson, MS[1], Timothy P. Hofer, MD MSc[2], Rob Holleman, MPH[2], Mandi L. Klamerus, MPH[2], Eve A. Kerr, MD, MPH[1,2]
[1] Michigan Program on Value Enhancement, University of Michigan, Ann Arbor, MI; [2] VA Center for Clinical Management Research, Ann Arbor, MI

Background

Low back pain is one of the most common presenting problems in primary care. Professional societies recommend conservative treatment prior to performing imaging tests for acute low back pain without red-flags.

Research Question

To what extant do overuse measures for low-value LBP imaging agree in identifying patients exposed to low-value imaging?

Population Studied

N = 2,001,135 MarketScan enrollees age 18-64 with an index diagnosis for acute low back pain in 2014 and 9 months continuous coverage.

Methods

We compared three overuse measures for LBP imaging using % mutual agreement and Fleiss's κ.

Results

LBP
100% ($\kappa = 1$)

LBP Imaging
99% ($\kappa = 0.98$)

Exclusions
76% ($\kappa = 0.62$)

Low-value LBP imaging
91% ($\kappa = .79$)

Conclusion

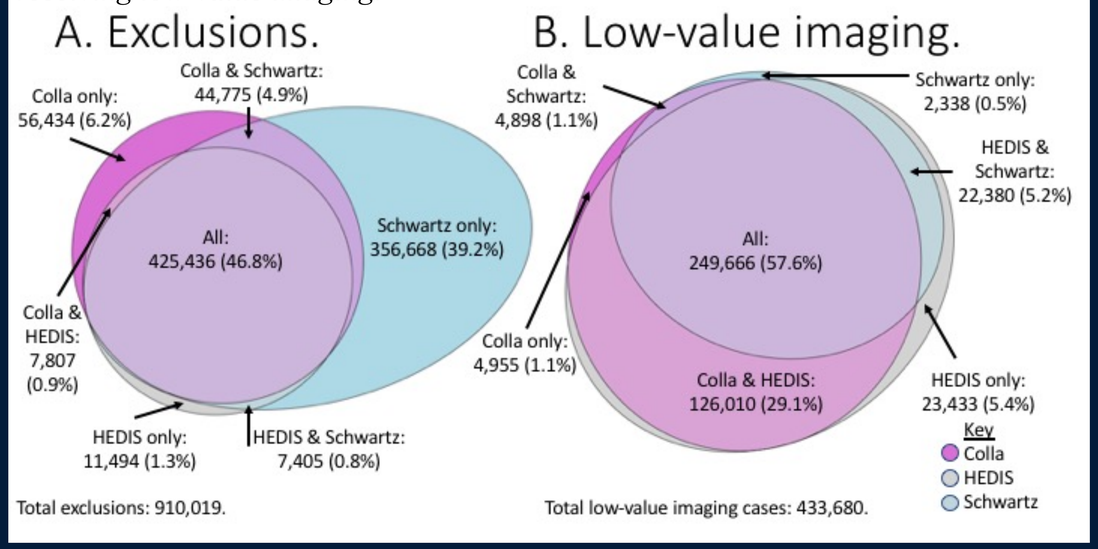
Among three overuse measures for imaging for acute LBP without red flags, there is high inter-measure agreement regarding diagnosis codes for LBP and procedure codes for LBP-related imaging, but *substantial disagreement in defining exclusions*. Consequently, there is *poor agreement as to which patients received low-value imaging*.

Implication

Additional consensus is need around which diagnoses should exclude a patient from the denominator of overuse measures for imaging for acute LBP without red-flags.

Overuse measures of "low-value imaging for acute lower back pain without red flags" agree in defining lower back pain and associated imaging, but *disagree when defining red flags*.

Venn diagrams comparing measures on excluded patients and patients flagged as receiving low-value imaging.



Consequently, there is *low inter-measure agreement* as to which patients were exposed to low-value imaging.

https://jbhender.github.io/research/hsr/lbp_agree.pdf

Design.

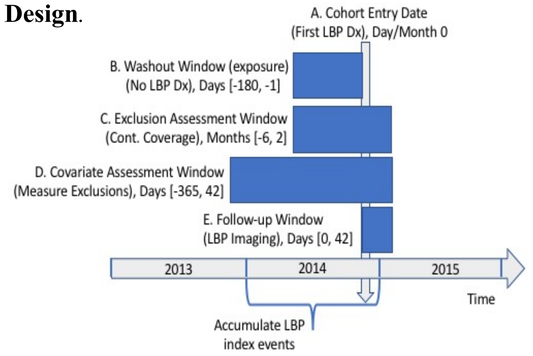


Table 1. *Standardized measurement periods.*

Measure	Clean Period (B)	Exclusion Period (D)	Imaging Period (E)
Colla	Study period	12 months[1,2]	42 days
HEDIS	6 months	12 months[3]	28 days
Schwartz	6 months	Study Period	42 days
Standard	180 days	365 days	42 days

1: Study period for cancer. 2: two professional or one IP claim.
3: Study period for cancer and HIV; 3 months for trauma.
4: Following Pham et al 2009.

Table 2. *Distribution of cases.*

Measure	LBP w/o red flags, n (%)	Exclusions, n (%)	Low-value imaging, n (%)
Total	1,575,699 (100)	910,019 (100)	433,680 (100)
All	1,091,116 (69.2)	425,436 (46.8)	249,666 (57.6)
C & H	356,668 (22.6)	7,807 (0.9)	126,010 (29.1)
C & S	11,494 (0.7)	44,775 (4.9)	4,898 (1.1)
H & S	56,434 (3.6)	7,405 (0.8)	22,380 (5.2)
C only	7,405 (0.5)	56,434 (6.2)	4,955 (1.1)
H only	44,775 (2.8)	11,494 (1.3)	23,433 (5.4)
S only	**7,807 (0.5)**	**356,668 (39.2)**	2,338 (0.5)

C = Colla (2015), H = HEDIS (2017), S = Schwartz (2014)

Top exclusion disagreements

S only: malaise & fatigue (28%), neuritis (20%), anemia (8%), fever (7%), weight loss (2%).
C + S: neuralagia (4%); **C only:** atypical pap (2%)
H only: history of malignant skin neoplasm (2%), unspecified lumbar disc disorder (2%)

Agreement statistics, % agreement (κ):

- LBP - 100% ($\kappa = 1$)
- LBP imaging - 99% ($\kappa = 0.98$)
- Red-flag exclusions - 76% ($\kappa = 0.62$)
- Low-value LBP imaging - 91% ($\kappa = 0.79$)

Fleiss's Kappa.

$$\kappa = \frac{P_o - P_E}{1 - P_E}, P_0 = \frac{1}{3N}\sum_{n=1}^{N}\sum_{k \neq j} Y_{nk} = Y_{nj},$$

$$P_E = p^2 + (1-p)^2,\ p = \frac{1}{3N}\sum_{n=1}^{N}\sum_{k} Y_{nk}\ .$$