Some Statistical Issues in Modeling Frailty

Dane Van Domelen

January 28, 2019

Outline

- Goals of modeling frailty
- Modeling considerations for frailty…
  - …as a potential novel risk factor
  - …to help classify risk
  - …to control for confounding
- Claims-based frailty indexes

Path to COAH

- 2005-2009: BS in biomedical engineering, MSOE
- 2009-2012: Postbaccalaureate research fellow, NIA
- 2012-2018: PhD in biostatistics, Emory University

Modeling frailty: 4 broad goals

- Study frailty as a clinical syndrome
- Assess frailty as a novel risk factor
- Use frailty to improve risk prediction
- Use frailty to control for confounding

Bandeen-Roche et al. (*J Gerontol* 2015)

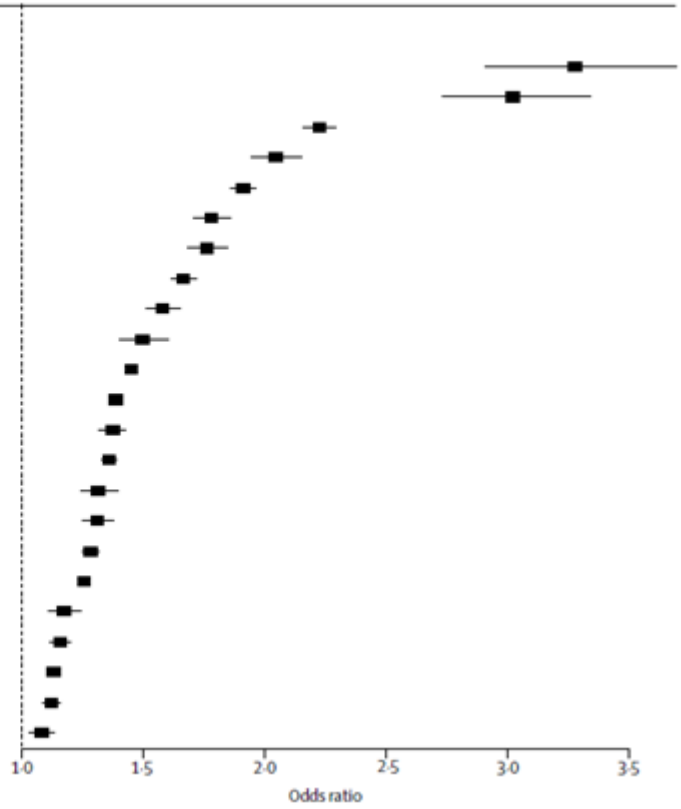
Table 1. Prevalence of Frail Status by Demographic Subgroups: National Health and Aging Trends Study, 2011; $n = 7{,}439$

Characteristic	Subgroup Prevalence (%)	Frail Status Prevalence Within Subgroup (%)		
		Robust	Prefrail	Frail
Age***	28.1	51.6	39.5	8.9
65–69	25.0	44.2	45.1	10.7
70–74	19.1	37.2	49.3	13.5
75–79	14.6	29.3	50.6	20.1
80–84	9.0	19.3	47.4	33.3
85–89	4.2	13.4	48.7	37.9
90+				
Sex***				
Male	43.6	43.8	43.3	12.9
Female	56.4	35.6	47.2	17.2
Race/ethnicity***				
White non-Hispanic	81.7	40.8	45.3	13.8
Black non-Hispanic	8.3	31.1	46.0	22.9
Hispanic	6.7	28.8	46.5	24.6
Other	3.4	40.5	45.8	13.7
Residence***				
Community	94.6	40.3	45.2	14.5
Residential care	5.4	19.8	50.7	29.5
Income***				
Lowest quartile	23.6	23.8	50.4	25.8
2nd quartile	24.1	32.4	47.9	19.7
3rd quartile	26.9	43.0	46.1	11.0
Highest quartiles	25.4	56.0	38.1	5.9

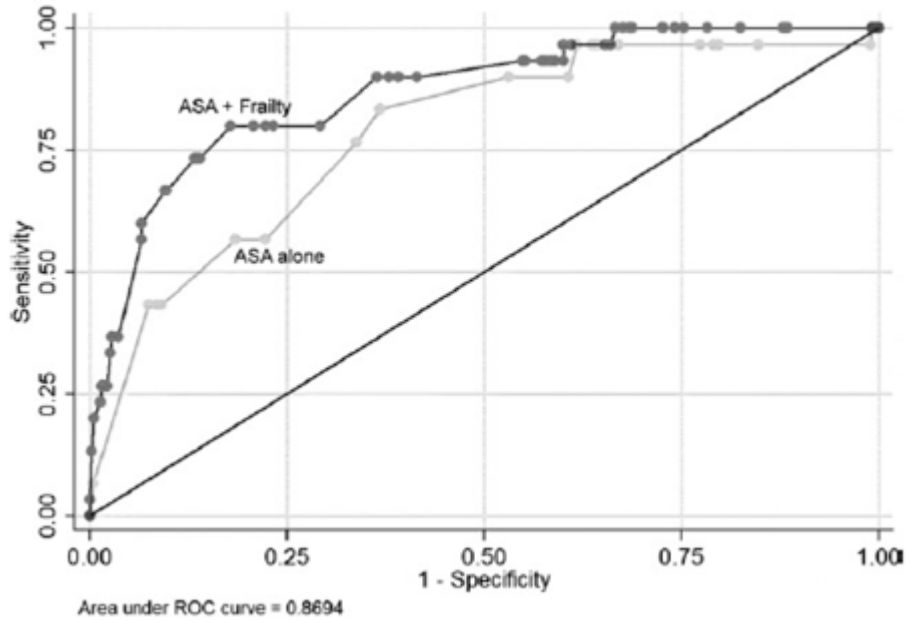
Notes: Overall prevalence estimates were 15.3% frail (95% CI: 14.2%, 16.4%), 45.5% prefrail (95% CI: = 44.0%, 46.9%), and 39.2% robust (95% CI: 37.7%, 40.8%). Per characteristic, comparisons were statistically significant at ***$p < .001$.

Hanlon et al. (*Lancet* 2017)

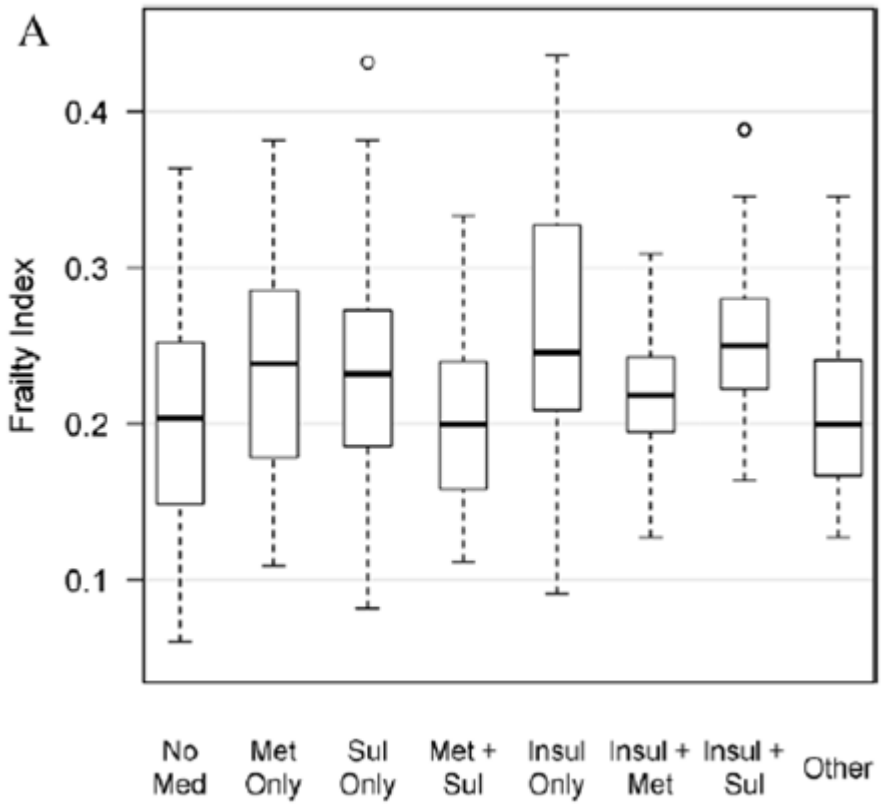
Condition	N	n	Pre-frailty (%)	Odds ratio (99.75% CI)
Multiple sclerosis	1540	845	54·9%	3·278 (2·91–3·70)
Chronic fatigue syndrome	2115	1139	53·9%	3·022 (2·74–3·34)
Diabetes	24696	13538	54·8%	2·226 (2·16–2·29)
Chronic obstructive pulmonary disease	8074	4119	51·0%	2·046 (1·95–2·15)
Depression	28022	14225	50·8%	1·911 (1·86–1·96)
Connective tissue disease	10778	5099	47·3%	1·782 (1·71–1·86)
Stroke or transient ischaemic attack	8513	4123	48·4%	1·764 (1·68–1·85)
Coronary heart disease	22074	10531	47·7%	1·666 (1·62–1·72)
Anxiety	8890	4184	47·1%	1·58 (1·51–1·65)
Epilepsy	3936	1788	45·4%	1·499 (1·40–1·61)
Painful condition	82620	37181	45·0%	1·453 (1·43–1·48)
Hypertension	130699	58406	44·7%	1·389 (1·37–1·41)
Irritable bowel syndrome	11354	4970	43·8%	1·375 (1·32–1·43)
Treated dyspepsia	38531	17292	44·9%	1·361 (1·33–1·39)
Diverticular disease	5334	2383	44·7%	1·316 (1·24–1·40)
Osteoporosis	7865	3294	41·9%	1·312 (1·25–1·38)
Thyroid disease	28775	12874	44·7%	1·283 (1·25–1·32)
Asthma	57169	24073	42·1%	1·258 (1·24–1·28)
Glaucoma	5245	2159	41·2%	1·176 (1·11–1·25)
Migraine	14267	5739	40·2%	1·16 (1·12–1·20)
Cancer*	37981	15325	40·3%	1·133 (1·11–1·16)
Psoriasis or eczema	17665	7000	39·6%	1·124 (1·09–1·16)
Prostate disease	8184	3016	36·9%	1·084 (1·03–1·14)



Makary et al. (*J Am Coll Surg* 2010)



Presley et al. (*J Gerontol* 2018)



Assessing frailty as a novel risk factor

Frailty phenotype as a novel risk factor

- Individual items:
  - X_1 = Weight loss in past year (lb)
  - X_2 = Grip strength (kg)
  - X_3 = Gait speed (m/s)
  - X_4 = Energy expenditure (kcal/d)
  - X_5 = 1 if exhaustion, else 0
- Frailty phenotype:
  - $X^* = I(X_1 > k_1) + I(X_2 < k_2) + I(X_3 < k_3) + I(X_4 < k_4) + X_5$
  - $X = I(X^* \geq 3)$
- If we fit:
  - $\text{logit}[P(Y = 1)] = \beta_0 + \beta_x X + \boldsymbol{\beta_c^T C}$
- We will almost certainly reject $H_0 : \beta_x = 0$

Frailty phenotype as a novel risk factor

- Why is this a problem?
  - In general, functions of risk factors are risk factors.
  - Expect FP correlate with any Y linked to X_1-X_5.
  - Known dose-response for gait speed, PA, etc.
- "Frailty is a novel risk factor"
  - How do we prove this?

Notions about frailty

"…a critical mass of phenotypic components…"

"…an adverse, potentially downward spiral…"

"…collectively often represent an insurmountable burden of disease…"

"…a marker of people who are very vulnerable for short-term mortality, for disability and falling…"

$\Rightarrow$ Sounds like an interaction!

Our burden of proof

- Show that frailty increases risk of outcomes, beyond what we would expect from the underlying items, which are well-established risk factors.
- E.g., HR > additive effects of individual items

The model we have to beat

$$\text{logit}[P(Y=1)] = \beta_0 + \boldsymbol{\beta_c^T C} + \beta_1 X_1 + \ldots + \beta_5 X_5$$

- (This would be too easy!)

$$\text{logit}[P(Y=1)] = \beta_0 + \boldsymbol{\beta_c^T C}$$

Approach #1:

$$\text{logit}[P(Y=1)] = \beta_0 + \boldsymbol{\beta_c^T C} + \beta_1 X_1 + \ldots + \beta_5 X_5 + \beta_{x^*} X^*$$

- Test for "frailty is a novel risk factor":
  - $H_0 : \beta_{x^*} = 0$
- Model implies:
  - Each item has (unique) effect on Y
  - In addition, crossing a cutpoint is associated with e^{β_x} times odds
  - Each deficit has same effect
- Some problems:
  - Incompatible with notion of extra risk at $X^* \geq 3$
  - Interpretation of β_{x^*}
  - A contradiction? Effect of each item unique, each deficit constant.

Approach #2:

$$\text{logit}[P(Y=1)] = \beta_0 + \boldsymbol{\beta_c^T C} + \beta_1 X_1 + \ldots + \beta_5 X_5 + \beta_x X$$

- Test for "frailty is a novel risk factor":
  - $H_0 : \beta_x = 0$
- Model implies:
  - Each item has (unique) linear effect on log-odds of Y
  - In addition, accumulating 3 deficits is associated with e^{β_x} times odds
- Some problems:
  - Interpretation of β_x, contradiction

Approach #3:

$$\text{logit}[P(Y=1)] = \beta_0 + \boldsymbol{\beta_c^T C} + \beta_1 X_1 + \ldots + \beta_5 X_5 + \beta_1^* X_1 X + \ldots + \beta_5^* X_5 X + \beta_x X$$

- Test for "frailty is a novel risk factor":
  - $H_0 : \beta_1^* = \beta_2^* = \beta_3^* = \beta_4^* = \beta_5^* = \beta_x = 0$
  - $H_0^* : \beta_x = 0$
- Model implies:
  - β_i = effect of item i in non-frail
  - $\beta_i + \beta_i^*$ = effect of item i in frail
- Plausible?
  - May or may not reflect our thinking about frailty
  - "Different model kicks in" once frail - appealing!

Why is this important?

- May dictate whether HR's are ~5 or ~1-1.5
- Could lose associations with certain outcomes
- Will this weaken the evidence base?
    - Yes: associations will be smaller
    - No: rejecting a stronger null is more convincing

Using frailty to improve risk prediction

Using frailty to improve risk prediction

- FP is a highly summarized statistic of $(X_1, \ldots, X_5)$
- If FP improves risk prediction, $(X_1, \ldots, X_5)$ should improve it even more
- If skip right to FP, likely to underfit
- What does "frailty improves risk prediction" mean?
  - In my opinion, a **significant and informative** interaction

Using frailty to control for confounding

Using frailty to control for confounding

- Three stages of underadjustment:
  - Dichotomize variables with dose-response
  - Convert 5 binary confounders to scalar sum
  - Dichotomize the sum
- Residual confounding $\Rightarrow$ false positives!

Watch out for Frank Harrell!

"Categorization is a data crime."

"Dichotomization kills."

"We see dichotomization doing untold damage in statistical analysis in all fields."

Claims-based frailty indexes

Basic idea

- Want to model frailty, but don't have frailty items
- Have Medicare claims data
- Many of those variables *inform* frailty
- Two-step procedure:
  - Build classifier for frailty given claims data
  - Impute frailty variable, and proceed with modeling
- Presumably, goal is to model CFI to similar effect as FP/FI

Two proposed estimation procedures

Med Care. 2017 Jul;55(7):716-722. doi: 10.1097/MLR.0000000000000729.

Development of a Claims-based Frailty Indicator Anchored to a Well-established Frailty Phenotype.

Segal JB[1], Chang H-Y, Du Y, Walston JD, Carlson MC, Varadhan R.

J Gerontol A Biol Sci Med Sci. 2018 Jun 14;73(7):980-987. doi: 10.1093/gerona/glx229.

Measuring Frailty in Medicare Data: Development and Validation of a Claims-Based Frailty Index.

Kim DH[1,2], Schneeweiss S[1], Glynn RJ[1,3], Lipsitz LA[2,4], Rockwood K[5], Avorn J[1].

Segal procedure

- Reference standard frailty measure:
  - Frailty phenotype
- Fitting imputation model:
  - 312 candidate variables, 21 selected
  - Logistic regression w/ lasso
- Imputation:
  - $\hat{p}$ = Predicted probability from logistic regression
  - $\hat{X} = I(\hat{p} > 0.2)$ (this is the CFI)
- Modeling: Just use the $\hat{X}$'s

TABLE 2. Best Model: Variables in Claims-based Frailty Indicator

B-Coefficient	Variables
1.24	Impaired mobility
0.54	Depression
0.50	Congestive heart failure
0.50	Parkinson disease
−0.49	White race
0.43	Arthritis (any type)
0.33	Cognitive impairment
0.31	Charlson comorbidity index (>0, 0)
0.28	Stroke
0.24	Paranoia
0.23	Chronic skin ulcer
0.21	Pneumonia
−0.19	Male sex
0.18	Skin and soft tissue infection
0.14	Mycoses
0.09	Age (in 5 y categories)
0.09	Admission in past 6 mo
0.08	Gout or other crystal-induced arthropathy
0.08	Falls
0.05	Musculoskeletal problems
0.05	Urinary tract infection

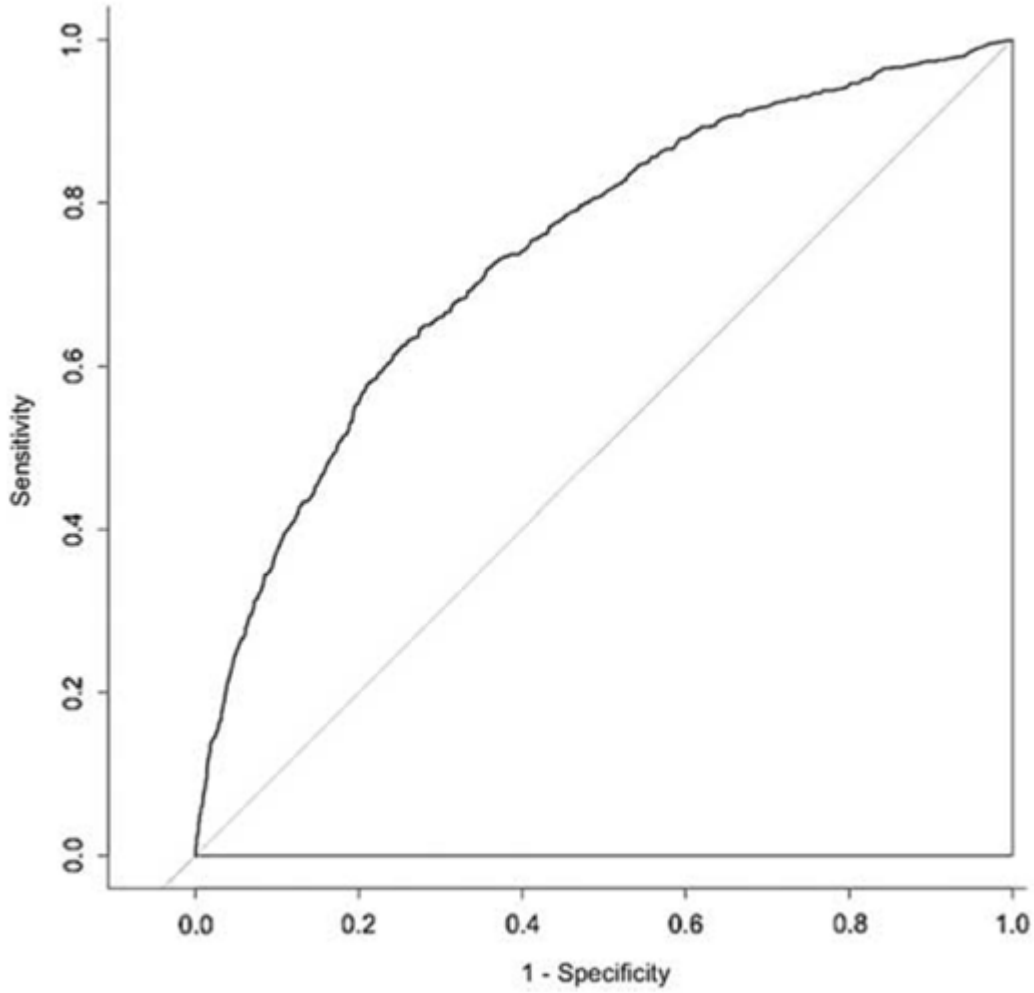
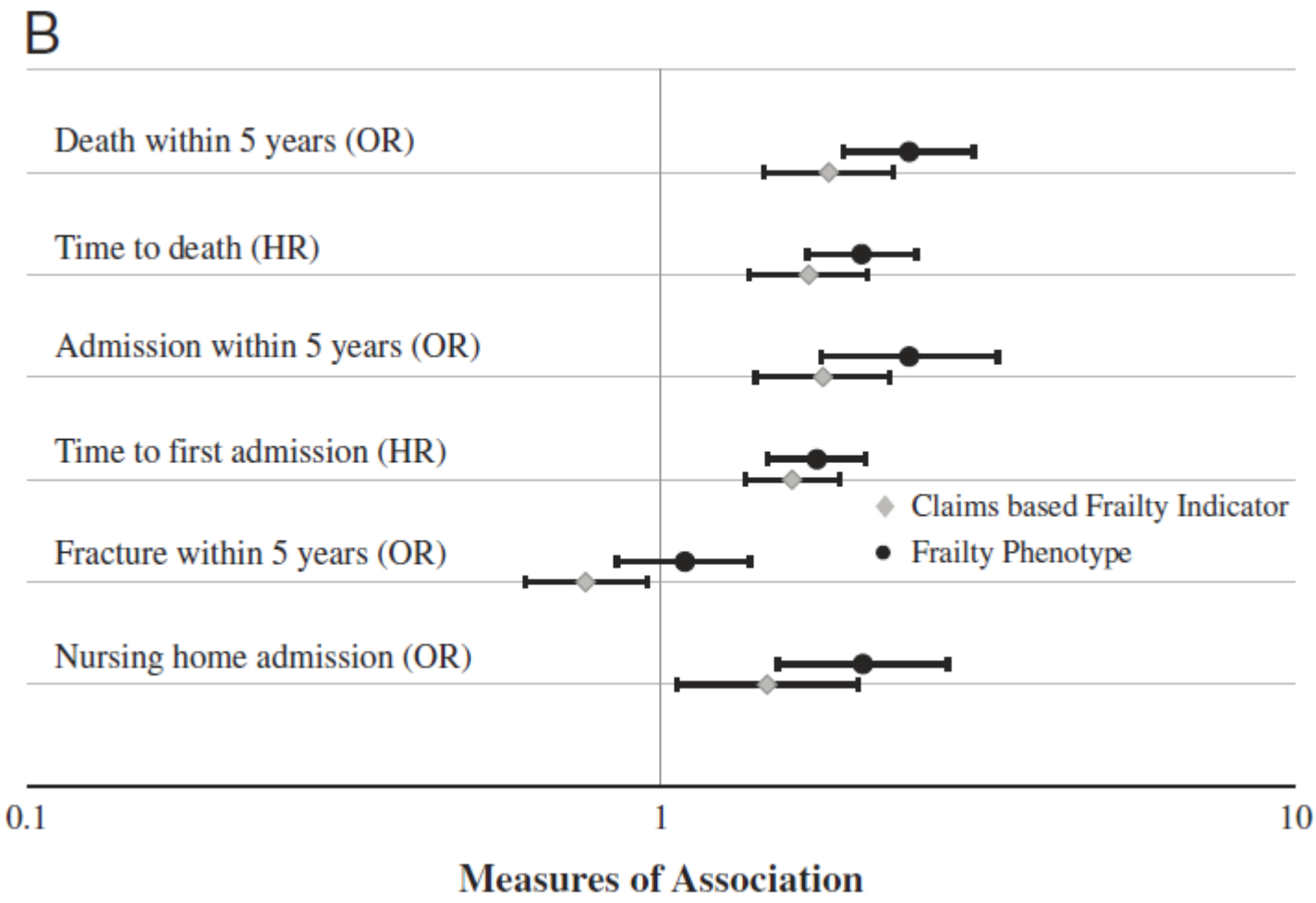


FIGURE 1. Receiver operator curve for Claims-based Frailty Indicator.


B
Death within 5 years (OR)
Time to death (HR)
Admission within 5 years (OR)
Time to first admission (HR)
Fracture within 5 years (OR)
Nursing home admission (OR)
Claims based Frailty Indicator
Frailty Phenotype
0.1
1
10
Measures of Association

Kim procedure

- Reference standard frailty measure:
  - 56-item frailty index
- Fitting calibration model:
  - 580 candidate variables, 93 selected
  - Linear regression w/ lasso
- Imputation:
  - $\hat{X}$ = Predicted values from regression (this is the CFI)
- Modeling: Just use the $\hat{X}$'s.

Table 2. Selected Codes Associated with Frailty in Medicare Current Beneficiary Survey

Type	Codes	Description of Claims-Based Variables	Prevalence	Coefficient
HCPCS	E0250-E0373	Hospital beds and associated supplies	0.018	0.086
HCPCS	K0001-K0462 K0669	Wheelchairs, components, and accessories	0.035	0.078
ICD9 Dx	290–294	Organic psychotic conditions	0.052	0.047
ICD9 Dx	330–338	Hereditary and degenerative diseases of the central nervous system	0.086	0.040
HCPCS	E0100-E0159	Walking aids and attachments	0.048	0.028
HCPCS	E1353-E1406	Accessories for oxygen delivery devices	0.051	0.027
HCPCS	A4244-A4290	Other supplies including diabetes supplies and contraceptives	0.125	0.024
HCPCS	A5500-A5513	Diabetic footwear	0.029	0.024
ICD9 Dx	295–299	Other psychoses	0.036	0.021
ICD9 Dx	420–429	Other forms of heart disease	0.375	0.020
ICD9 Dx	890–897	Open wound of lower limb	0.017	0.020
ICD9 Dx	410–414	Ischemic heart disease	0.310	0.019
ICD9 Dx	401–405	Hypertensive disease	0.752	0.017
ICD9 Dx	430–438	Cerebrovascular disease	0.172	0.016
ICD9 Dx	300–316	Neurotic disorders, personality disorders, and other nonpsychotic mental disorders	0.154	0.014
ICD9 Dx	710–719	Arthropathies and related disorders	0.482	0.014
CPT-4	99308	Nursing facility care—subsequent	0.016	0.014
ICD9 Dx	490–496	Chronic obstructive pulmonary disease and allied conditions	0.235	0.013
ICD9 Dx	030-041	Other bacterial diseases	0.031	0.012
ICD9 Dx	451–459	Diseases of veins and lymphatics, and other diseases of circulatory system	0.154	0.012
ICD9 Dx	480–487	Pneumonia and influenza	0.066	0.012
ICD9 Dx	250–259	Diseases of other endocrine glands	0.312	0.011
ICD9 Dx	590–599	Other diseases of urinary system	0.289	0.011
ICD9 Dx	797–799	Ill-defined and unknown causes of morbidity and mortality	0.046	0.011
ICD9 Dx	920–924	Contusion with intact skin surface	0.058	0.011
ICD9 Dx	580–589	Nephritis, nephrotic syndrome, and nephrosis	0.084	0.010
HCPCS	A0021-A0999	Transportation services including ambulance	0.110	0.010

Predictors	Mortality	ADL Disability	IADL Disability
A. Survey-based FI			
Age (per 1 year)	1.06 (1.04, 1.09)	1.05 (1.03, 1.06)	1.07 (1.05, 1.08)
Female (vs. male)	0.68 (0.48, 0.97)	1.20 (1.00, 1.45)	1.18 (1.00, 1.38)
CCI (per 1 point)	1.35 (1.27, 1.43)	1.01 (0.97, 1.06)	1.03 (0.99, 1.08)
FI (per 0.1 point)	1.41 (1.24, 1.59)	4.23 (3.79, 4.74)	4.30 (3.82, 4.86)
C statistic	0.83	0.76	0.70
B. Claims-based FI			
Age (per 1 year)	1.07 (1.04, 1.09)	1.05 (1.04, 1.07)	1.07 (1.05, 1.09)
Female (vs. male)	0.68 (0.48, 0.97)	1.42 (1.11, 1.82)	1.22 (0.98, 1.52)
CCI (per 1 point)	1.30 (1.22, 1.39)	1.01 (0.94, 1.08)	0.98 (0.91, 1.05)
FI (per 0.1 point)	1.82 (1.45, 2.29)	2.53 (1.96, 3.28)	2.30 (1.75, 3.05)
C statistic	0.82	0.69	0.67

Comparison of CFI procedures

	Segal et al.	Kim et al.
Reference standard	Frailty phenotype	56-item frailty index
Number of classifiers	312 (21 selected)	580 (93 selected)
Method to fit classifier	Logistic regression w/ lasso	Linear regression w/ lasso
Imputation Method	$\hat{X} = I(\hat{p} > 0.2)$	$\hat{X}$

Potential problem with CFI procedures

- Kim:
  - Regression calibration with 93 surrogates
- Segal:
  - Somewhere between regression calibration and MI
  - Akin to MI but with Y left out of imputation model

Rationale for regression calibration

$$Y = \beta_0 + \beta_x X + \boldsymbol{\beta_c^T C} + \epsilon,\ \epsilon \sim (0, \sigma_\epsilon^2)$$

- Problem: Measure X with error; observe "noisy version" W
- Solution: Fit model with $\hat{X} = E(X|W, \boldsymbol{C})$ in place of X
- Typically assume:

$$X = \alpha_0 + \alpha_w W + \boldsymbol{\alpha_c^T C} + \delta,\ \delta \sim (0, \sigma_\delta^2)$$

$$\begin{aligned} E(Y|\hat{X}, \boldsymbol{C}) &= E(\beta_0 + \beta_x X + \boldsymbol{\beta_c^T C} + \epsilon|\hat{X}, \boldsymbol{C}) \\ &= \beta_0 + \beta_x \hat{X} + \boldsymbol{\beta_c^T C} \end{aligned}$$

What if W informs Y?

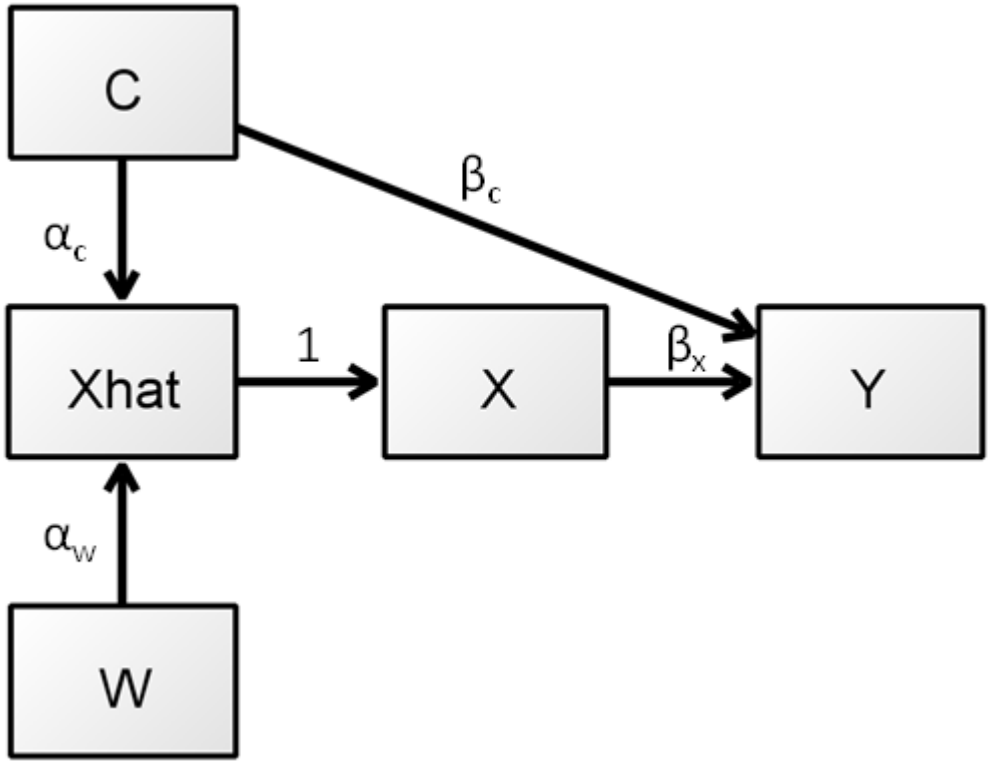
$$Y = \beta_0 + \beta_x X + \boldsymbol{\beta_c^T C} + \beta_w W + \epsilon,\ \epsilon \sim (0, \sigma_\epsilon^2)$$

- Typically assume:

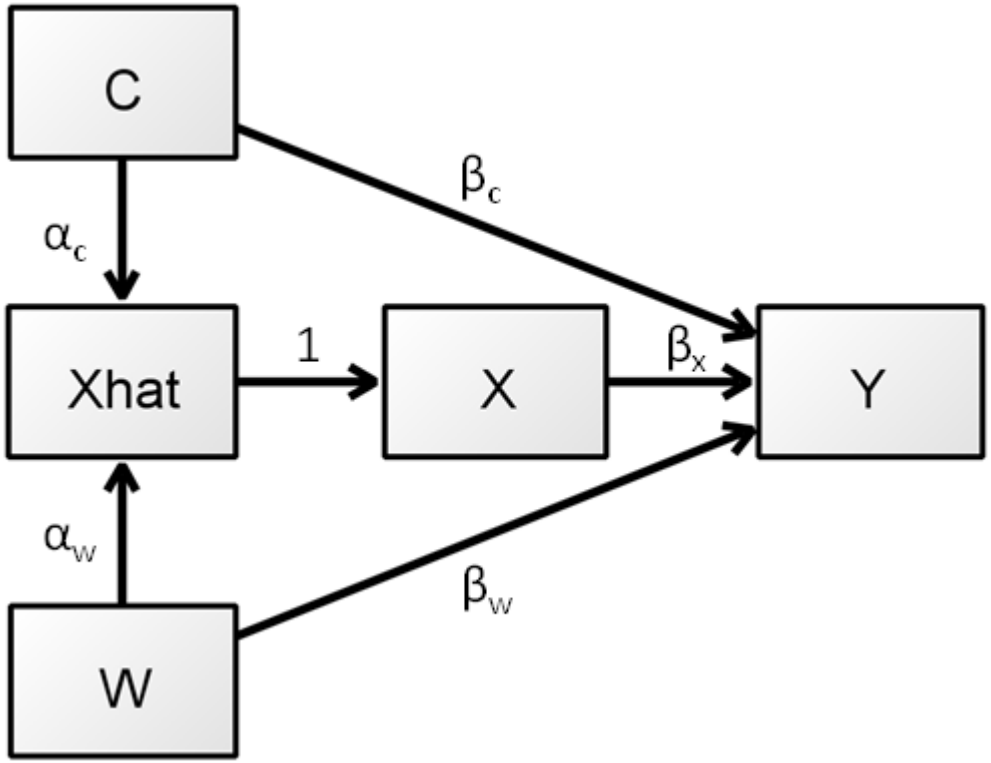
$$X = \alpha_0 + \alpha_w W + \boldsymbol{\alpha_c^T C} + \delta,\ \delta \sim (0, \sigma_\delta^2)$$

$$\begin{aligned} E(Y|\hat{X}, \boldsymbol{C}) &= E(\beta_0 + \beta_x X + \boldsymbol{\beta_c^T C} + \beta_w W + \epsilon|\hat{X}, \boldsymbol{C}) \\ &= \beta_0 + \beta_x \hat{X} + \boldsymbol{\beta_c^T C} + \beta_w E(W|\hat{X}, \boldsymbol{C}) \\ &= \beta_0 + \beta_x \hat{X} + \boldsymbol{\beta_c^T C} + \beta_w (-\frac{\alpha_0}{\alpha_w} + \frac{1}{\alpha_w}\hat{X} - \frac{1}{\alpha_w}\boldsymbol{\alpha_c^T C}) \\ &= (\beta_0 - \frac{\beta_w \alpha_0}{\alpha_w}) + (\beta_x + \frac{\beta_w}{\alpha_w})\hat{X} + (\boldsymbol{\beta_c^T} - \frac{\beta_w}{\alpha_w}\boldsymbol{\alpha_c^T})\boldsymbol{C} \end{aligned}$$

$\Rightarrow$ Actually need to compare to $E(Y|X, \boldsymbol{C})$...



$$E(Y|X,C) = k + \beta_x X + \beta_c C$$
$$E(Y|\hat{X},C) = k + (1)\beta_x \hat{X} + \beta_c C$$



$$E(Y|\hat{X}, C) = k + (1)\beta_x \hat{X} + \beta_c C + r(\alpha_w)\beta_w \hat{X} + \beta_c r(\alpha_w)\beta_w C$$

Regression calibration fails if any of the variables used to inform X, but omitted from the model for Y, truly inform Y

TABLE 2. Best Model: Variables in Claims-based Frailty Indicator

B-Coefficient	Variables
1.24	Impaired mobility
0.54	Depression
0.50	Congestive heart failure
0.50	Parkinson disease
−0.49	White race
0.43	Arthritis (any type)
0.33	Cognitive impairment
0.31	Charlson comorbidity index (>0, 0)
0.28	Stroke
0.24	Paranoia
0.23	Chronic skin ulcer
0.21	Pneumonia
−0.19	Male sex
0.18	Skin and soft tissue infection
0.14	Mycoses
0.09	Age (in 5 y categories)
0.09	Admission in past 6 mo
0.08	Gout or other crystal-induced arthropathy
0.08	Falls
0.05	Musculoskeletal problems
0.05	Urinary tract infection

Table 2. Selected Codes Associated with Frailty in Medicare Current Beneficiary Survey

Type	Codes	Description of Claims-Based Variables	Prevalence	Coefficient
HCPCS	E0250-E0373	Hospital beds and associated supplies	0.018	0.086
HCPCS	K0001-K0462 K0669	Wheelchairs, components, and accessories	0.035	0.078
ICD9 Dx	290–294	Organic psychotic conditions	0.052	0.047
ICD9 Dx	330–338	Hereditary and degenerative diseases of the central nervous system	0.086	0.040
HCPCS	E0100-E0159	Walking aids and attachments	0.048	0.028
HCPCS	E1353-E1406	Accessories for oxygen delivery devices	0.051	0.027
HCPCS	A4244-A4290	Other supplies including diabetes supplies and contraceptives	0.125	0.024
HCPCS	A5500-A5513	Diabetic footwear	0.029	0.024
ICD9 Dx	295–299	Other psychoses	0.036	0.021
ICD9 Dx	420–429	Other forms of heart disease	0.375	0.020
ICD9 Dx	890–897	Open wound of lower limb	0.017	0.020
ICD9 Dx	410–414	Ischemic heart disease	0.310	0.019
ICD9 Dx	401–405	Hypertensive disease	0.752	0.017
ICD9 Dx	430–438	Cerebrovascular disease	0.172	0.016
ICD9 Dx	300–316	Neurotic disorders, personality disorders, and other nonpsychotic mental disorders	0.154	0.014
ICD9 Dx	710–719	Arthropathies and related disorders	0.482	0.014
CPT-4	99308	Nursing facility care—subsequent	0.016	0.014
ICD9 Dx	490–496	Chronic obstructive pulmonary disease and allied conditions	0.235	0.013
ICD9 Dx	030-041	Other bacterial diseases	0.031	0.012
ICD9 Dx	451–459	Diseases of veins and lymphatics, and other diseases of circulatory system	0.154	0.012
ICD9 Dx	480–487	Pneumonia and influenza	0.066	0.012
ICD9 Dx	250–259	Diseases of other endocrine glands	0.312	0.011
ICD9 Dx	590–599	Other diseases of urinary system	0.289	0.011
ICD9 Dx	797–799	Ill-defined and unknown causes of morbidity and mortality	0.046	0.011
ICD9 Dx	920–924	Contusion with intact skin surface	0.058	0.011
ICD9 Dx	580–589	Nephritis, nephrotic syndrome, and nephrosis	0.084	0.010
HCPCS	A0021-A0999	Transportation services including ambulance	0.110	0.010

Some intuition

If we fit:

$$Y = \beta_0 + \beta_x \hat{X} + \epsilon$$

and our imputed variable is:

$$\hat{X} = \alpha_0 + \alpha_w W$$

effects of W on Y will contribute to $\hat{\beta}_x$.

More broadly…

- RC proposed by Rosner et al. (*Stat Med* 1989)
  - Just 1 surrogate: the "imprecise version"
- Extension by Weller et al. (*J Stat Plan Inference* 2007)
  - Multiple surrogates
  - Surrogates still have to be surrogates!

Option 1: Bias correction

- Recall:

$$E(Y|\hat{X}, \boldsymbol{C}) = (\beta_0 - \frac{\beta_w \alpha_0}{\alpha_w}) + (\beta_x + \frac{\beta_w}{\alpha_w})\hat{X} + (\boldsymbol{\beta_c^T} - \frac{\beta_w}{\alpha_w}\boldsymbol{\alpha_c^T})\boldsymbol{C}$$

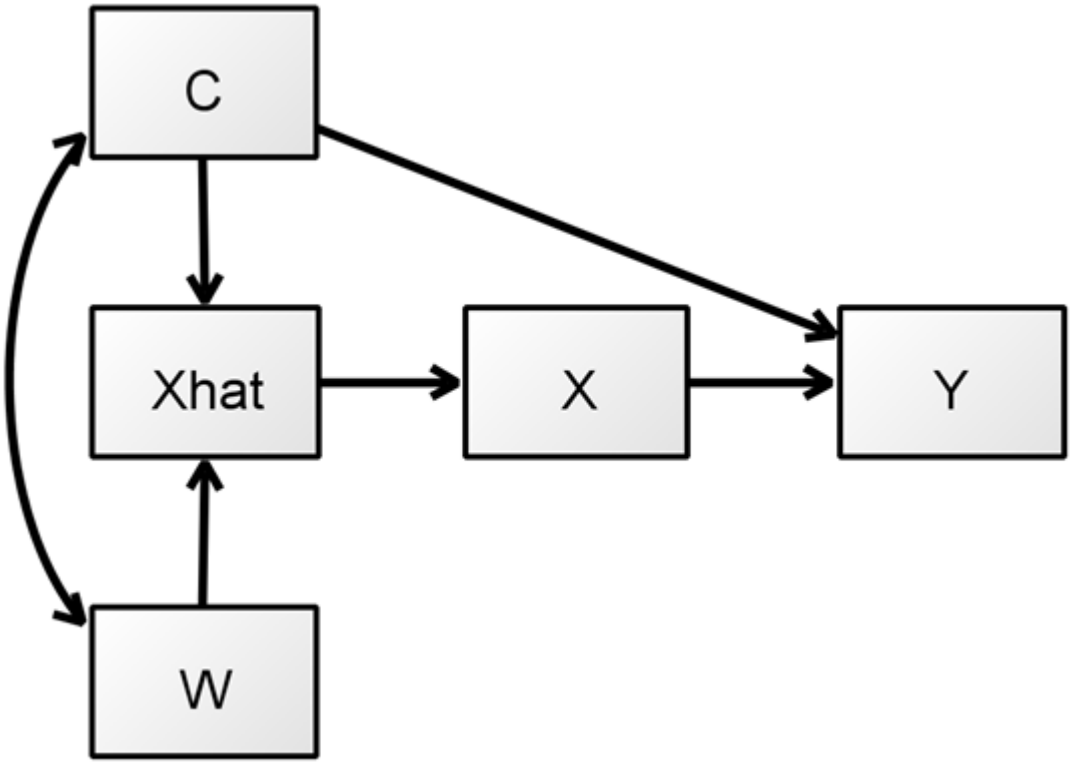
- Bias in $\hat{\beta}_x$ is $\frac{\beta_w}{\alpha_w}$
- Expect $\alpha_w > 0$ and probably $\beta_w > 0$
- Coefficient on $\hat{X}$ is an **upper bound**

$\Rightarrow$ This is wrong; bias should be based on *confounded* associations with Y

Option 2: Identify surrogate(s)

- For unbiasedness, W's must not inform Y.
- For precision, W's must substantially inform X.
- What might inform FP but not outcomes given FP?

Even with surrogate, need to be careful



$\Rightarrow W$ is a surrogate given (X, C), but not given X

Option 3: Standard MI

- Include Y in imputation model
  - Need **internal validation data**
  - In other words, assess FP on subsample
- Classifier variables
  - Optional as "auxiliary variables" to improve efficiency
- Abandons notion of CFI; no longer helpful for other studies.

$\Rightarrow$ Measuring FP on subset leads to easier estimation problem

Option 4: Don't anchor to FP/FI

- Classify frailty from claims data and theory
- Claims-based deficit accumulation index feasible (Kim)

Conclusions on CFI

- In my view, 2 options:
  - Measure frailty on subset, use MI (or RC, or ML)
  - Classify frailty directly from claims

Conclusions on regression calibration

- Developed for ME, where surrogacy is no problem
- Applied to missing data, several issues:
  - Need informative surrogate(s)
  - Surrogacy violated if any omitted variable links W to Y

Thanks!