Assessing Health Disparities in Intensive Longitudinal Data: Gender Differences in Granger Causality Between Primary Care Provider and Emergency Room Usage, Assessed with Medicaid Insurance Claims

May 24, 2017

Emil Coman[1]
Yinghui Duan[2]
Daren Anderson[3]

[1] UConn Health Disparities Institute, [2] UConn Health, [3] Weitzman Institute

Granger causality

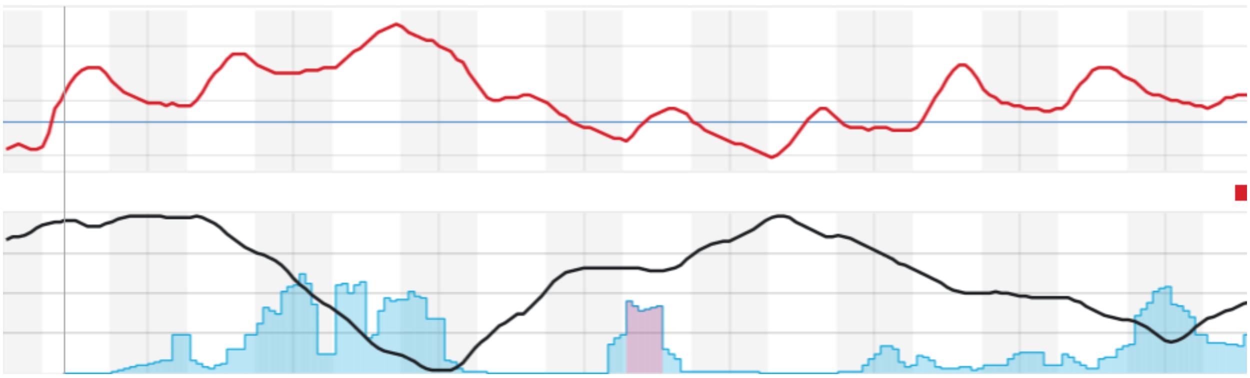
Goals

1. Introduce an analytical method to 'extract direction of causal effects' from time series/panel data/intensive longitudinal data

2. Show software code and output

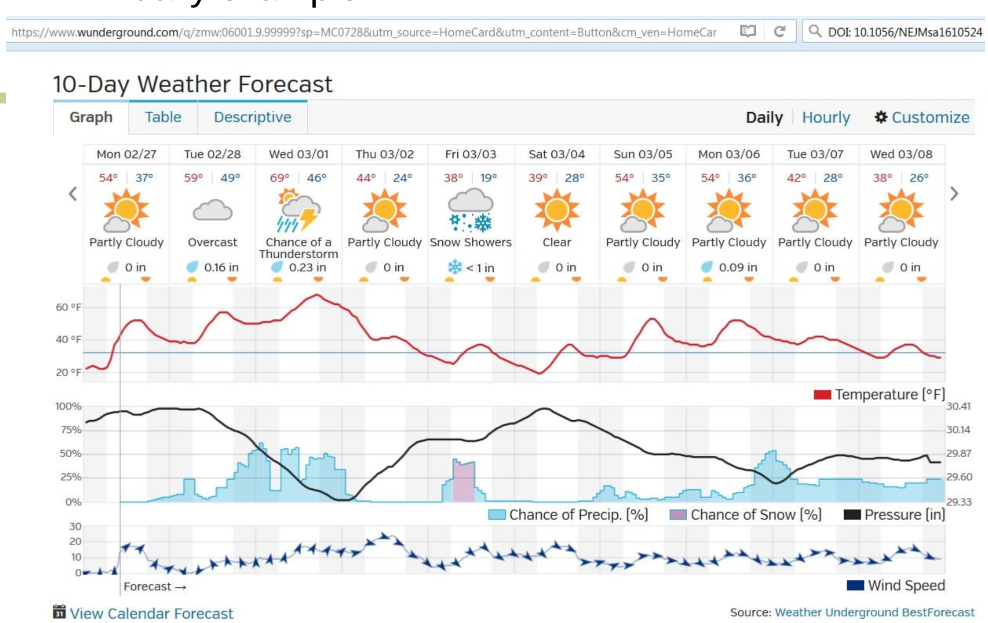
3. Interpret results

A daily example



Guess?

A daily example



1. Wunderground

Chicken-egg

"To conclude that one of the two "came first," we must find unidirectional causality from one to the other.

In other words, we must reject the *noncausality* of the one to the other and at the same time fail to reject the *noncausality* of the other to the one."

1. Thurman, W. N., & Fisher, M. E. (1988). Chickens, eggs, and causality, or which came first. *American journal of agricultural economics, 70(2), 237-238.*

Granger *noncausality*

"Granger (1969, 1980, 1988) proposed a notion of causality that is based on the following important points: the cause occurs before the effect and the cause must contain unique information that is otherwise not available and helps to predict the effect."

Eichler, M., & Didelez, V. (2010). On Granger causality and the effect of interventions in time series. *Lifetime Data Analysis, 16(1), 3-32. doi:10.1007/s10985-009-9143-3*

Granger *noncausality*

"We say that X Granger-causes Y if the current value of Y can be better predicted from the past values of all three series X, Y, and Z than from the past values of the two processes Y and Z alone. Here, "better predicted" means a smaller mean square prediction error.
We note that the definition depends on the set of variables Z included in the analysis."

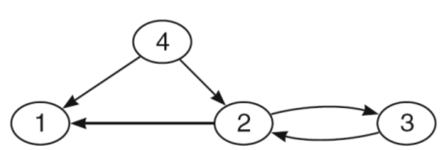
Tschacher W, Ramseyer F: Modeling psychotherapy processes by time-series panel analysis (TSPA). Psychother Res 2009, 19:469-481

Granger *noncausality*

Example 2.2 Consider the following multivariate Gaussian process X with

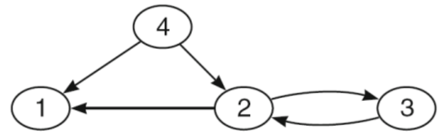
$$\begin{aligned} X_1(t) &= \alpha_1 X_4(t-2) + \beta_{12} X_2(t-1) + \varepsilon_1(t), \\ X_2(t) &= \alpha_2 X_4(t-1) + \beta_{23} X_3(t-1) + \varepsilon_2(t), \\ X_3(t) &= \beta_{32} X_2(t-1) + \varepsilon_3(t), \\ X_4(t) &= \varepsilon_4(t) \end{aligned}$$

where εi , i = 1, 2, 3, 4, are independent Gaussian white noise processes with mean 0 and variance σ^2.



Eichler, M., & Didelez, V. (2010). On Granger causality and the effect of interventions in time series. *Lifetime Data Analysis, 16(1), 3-32. doi:10.1007/s10985-009-9143-3*

Granger *noncausality*



"It is immediately obvious that with respect to the set {*X*1*, X*2*, X*3*, X*4}, the component *X*1 is Granger noncausal for all other variables, and for instance *X*3 is Granger noncausal for *X*1.

However, with respect to the reduced set {*X*1*, X*3*, X*4} it cannot be assumed anymore that *X*3 is Granger noncausal for *X*1.

It is less intuitive, but also clear from the full model that *X*3*(t)* is also not Granger noncausal anymore for *X*1 with respect to {*X*1*, X*2*, X*3} due to the selection effect when conditioning on the past of *X*2 which induces an association between *X*3 and *X*4."

Eichler, M., & Didelez, V. (2010). On Granger causality and the effect of interventions in time series. *Lifetime Data Analysis, 16(1), 3-32. doi:10.1007/s10985-009-9143-3*

Chicken-egg real example

The following equation was estimated by OLS:

$$Eggs_t = \mu = \sum_{i=1}^{L} \alpha_i \; Eggs_{t-i} = \sum_{i=1}^{L} \beta_i \; Chickens_{t-i} + \epsilon_t;$$

$H_0 : \beta_1 = \ . \ . \ . \ \beta_L = 0$ (chickens do not Granger cause eggs).

L = no. of lags	F-statistic	P-value	R^2 of the regression
1	.04	.85	.96
2	1.71	.19	.97
3	1.10	.36	.97
4	.79	.54	.97

"The validity of our test statistic requires lack of serial correlation, homoskedasticity, and normality of the disturbances in the distributed lag equations, which we of course assume."

1. Thurman, W. N., & Fisher, M. E. (1988). Chickens, eggs, and causality, or which came first. *American journal of agricultural economics, 70(2), 237-238.*

Chicken-egg

"The validity of our test statistic requires lack of serial correlation, homoskedasticity, and normality of the disturbances in the distributed lag equations, which we of course assume."

Part 2: Did the Egg Come First?

The following equation was estimated by OLS:

$$Chickens_t = \mu + \sum_{i=1}^{L} \alpha_i\, Chickens_{t-i} + \sum_{i=1}^{L} \beta_i\, Eggs_{t-i};$$

$H_0 : \beta_1 = \ldots = \beta_L = 0$ (eggs do not Granger cause chichens).

L = no. of lags	F-statistic	P-value	R^2 of the regression
1	1.23	.27	.73
2	10.36	.0002	.81
3	5.85	.0019	.81
4	4.71	.0032	.82

Data source: U.S. Department of Agriculture, 1983 and others.
Note: The data are annual, 1930–83.

1. Thurman, W. N., & Fisher, M. E. (1988). Chickens, eggs, and causality, or which came first. *American journal of agricultural economics, 70(2), 237-238.*

Chicken-egg

"While our test is agnostic regarding this instantaneous causality, we suspect that **eggs are endogenous** in the sense that chickens cause eggs within the sampling period.

A Wu-Hausman test of the predeterminedness of eggs could address the issue and would require a valid instrumental variable (correlated with eggs and uncorrelated with the chicken forecast error),perhaps bacon."

"*Suggestions for Future Research*
The **structural** implications of our results are not yet clear. To draw them out fully will require collaboration between economists and poultry scientists. The potential here is great."

1. Thurman, W. N., & Fisher, M. E. (1988). Chickens, eggs, and causality, or which came first. *American journal of agricultural economics, 70(2), 237-238.*

Illustration

<0?

Primary care
Costs $US

Emergency care
Costs $US

?

Medicaid costs:
Does spending for primary health care *reduce* later on emergency care costs?

Granger noncausality – applied in Stata

"Time-series vector autoregression (VAR) models originated in the macroeconometrics literature as an alternative to multivariate simultaneous equation models (Sims 1980).“

“Panel VAR model selection, estimation, and inference in a generalized method of moments (GMM) framework.“

“"Model selection, estimation, and inference about the homogeneous panel VAR model above can be implemented with the new commands *pvar, pvarsoc, pvargranger, pvarstable, pvarirf,* and *pvarfevd*. The syntax and outputs are closely patterned after Stata’s built-in *var* commands to easily switch between panel and time-series VAR."

1. Abrigo, M. R., & Love, I. (2016). Estimation of panel vector autoregression in Stata. *Stata Journal, 16(3), 778-804.*

Granger noncausality – applied in Stata

"*pvar* fits homogeneous panel VAR models by fitting a multivariate panel regression of each dependent variable on lags of itself, lags of all other dependent variables, and lags of exogenous variables, if any. The estimation is by GMM . The command is implemented using the interactive version of Stata's gmm command with analytic derivatives."
"*pvarsoc* provides various summary measures to aid the process of panel VAR model selection."
"The postestimation command *pvargranger* performs Granger causality Wald tests for each equation of the underlying panel VAR model. It provides a convenient alternative to Stata's built-in test command."
"The postestimation command *pvarstable* checks the stability condition of panel VAR estimates by calculating the modulus of each eigenvalue of the fitted model."
"The postestimation command *pvarirf* calculates and plots IRF s(impulse–response functions)."
"The postestimation command ichecks the stability condition of panel VAR estimates by calculating the modulus of each eigenvalue of the fitted model."
"The postestimation command *pvarfevd* computes FEVD based on a Cholesky decomposition of the residual covariance matrix of the underlying panel VAR model. Standarderrors and confidence intervals based on Monte Carlo simulation may be optionally computed."

1. Abrigo, M. R., & Love, I. (2016). Estimation of panel vector autoregression in Stata. *Stata Journal, 16(3), 778-804.*

Medicaid claims data

14,358 patients in CT (238404=63.79% female)

386,613 claims, up to 45 months.

Average age 44 years old

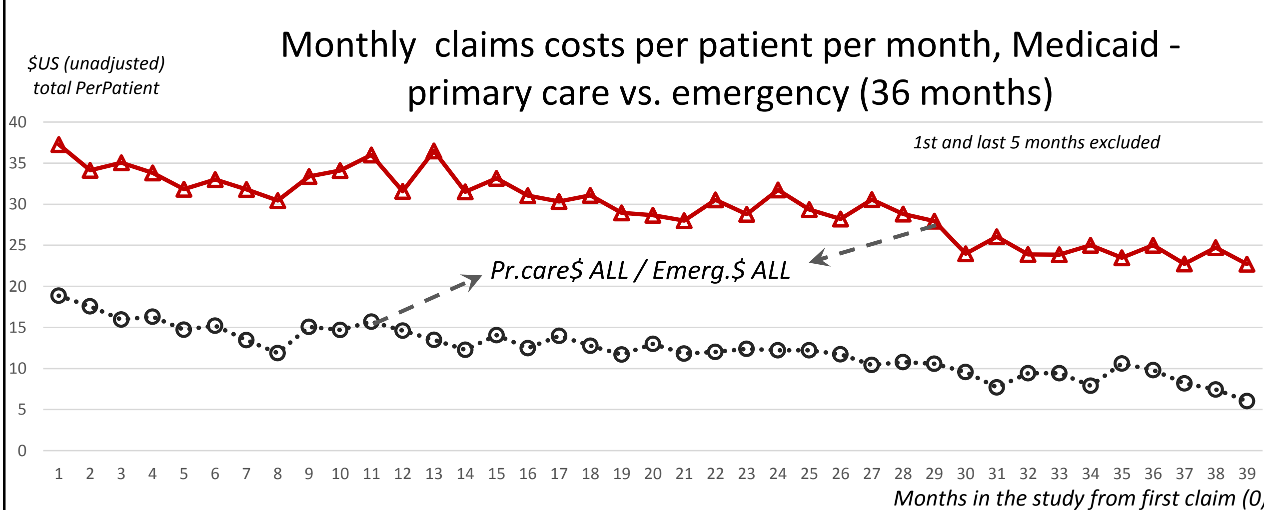
```
. xtsum age

Variable         |      Mean   Std. Dev.       Min        Max |    Observations
-----------------+--------------------------------------------+----------------
age      overall |  44.12122   17.60198          0        100 |     N =  373725
         between |             18.80543          0        100 |     n =   14358
         within  |                    0   44.12122   44.12122 | T-bar =  26.029
```

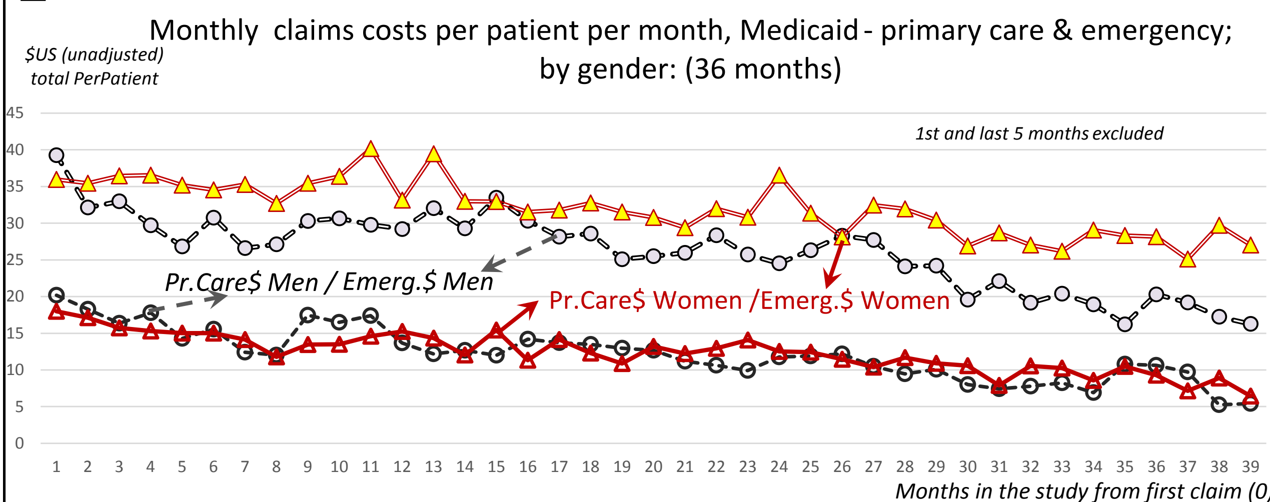
Medicaid claims Primary care (1) vs. Emergency (2)

92	Family Nurse Practitioner	1	
99	Community Health Nurse Practitioner	1	
124	Primary Care Nurse Practitioner	1	
306	Preventative Medicine	1	
318	General Practice Medicine	1	
345	General Pediatrics	1	
521	Medical FQHC & Tribal Svs Medical FQHC	1	
997	Primary Care Physician Assistant	1	
100	Critical Care Nurse Practitioner		2
102	Neonatal Critical Care Nurse Practitioner		2
104	Pediatric Critical Care Nurse Practitioner		2
260	Ambulance		2
261	Air Ambulance		2
262	Critical Care Helicopter		2
315	Emergency Medicine		2
611	Pediatric Emergency Department Medicine		2
612	Pediatric Emergency Medicine		2
621	Pediatric Critical Care Medicine		2

Medicaid claims Primary care vs. Emergency

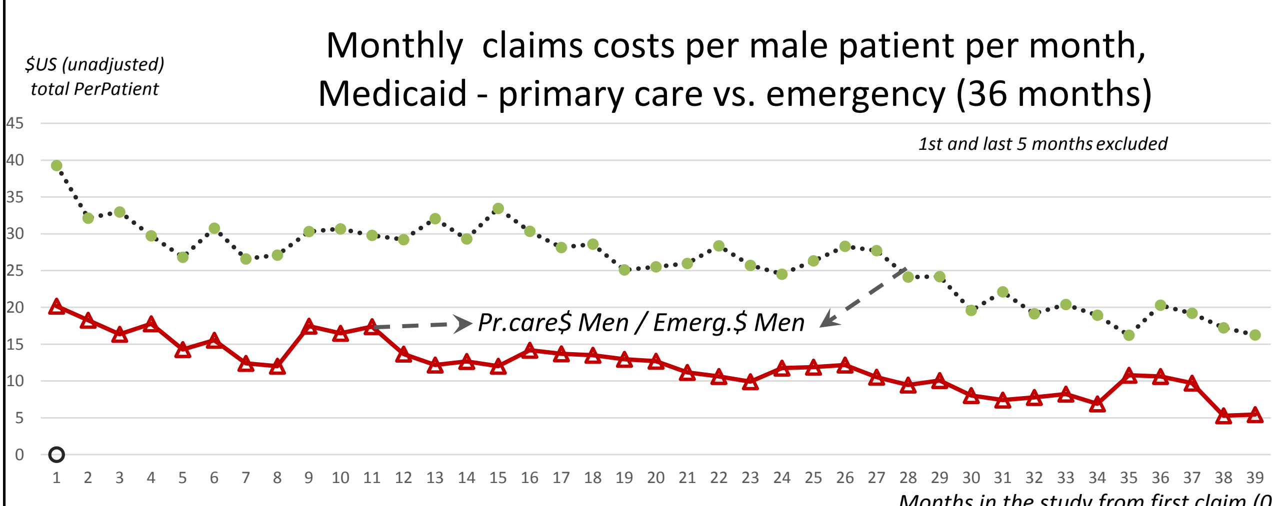


Medicaid claims by gender



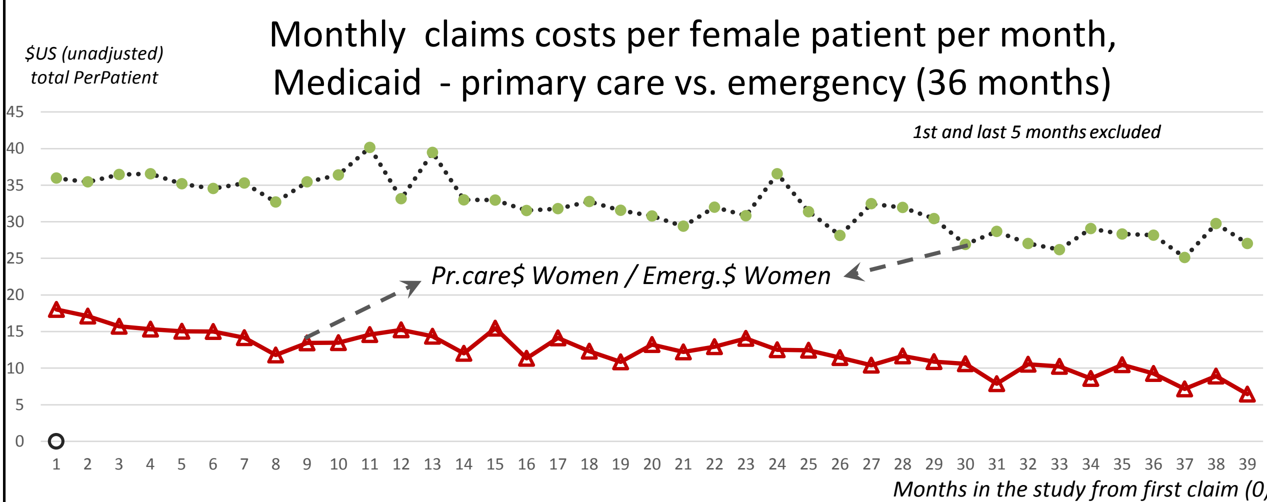
Modern Modeling conference, May 22-24, 2017 19

Medicaid claims - males



Modern Modeling conference, May 22-24, 2017 20

Medicaid claims - females



Panel Regression
Emergency $s-[-.018/.001*]->Primary care $s

```
xtreg dolprimc [Primary care $s] dolemerg [Emergency $s]

Random-effects GLS regression                   Number of obs      =    373,725
Group variable: id                              Number of groups   =     14,358
R-sq:                                           Obs per group:
     within  = 0.0004                                         min =          1
     between = 0.0007                                         avg =       26.0
     overall = 0.0004                                         max =         45
                                                Wald chi2(1)       =     146.57
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =     0.0000
------------------------------------------------------------------------------
    dolprimc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    dolemerg |  -.0181991   .0015032   -12.11   0.000    -.0211454   -.0152528
       _cons |   24.33821   .5940643    40.97   0.000     23.17387    25.50256
-------------+----------------------------------------------------------------
```

Panel Regression
Primary care $s -[-.021/.002*]->Emergency $s

```
xtreg dolemerg [Emergency $s] dolprimc [Primary care $s]
Random-effects GLS regression                 Number of obs      =    373,725
Group variable: id                            Number of groups   =     14,358
R-sq:                                         Obs per group:
     within  = 0.0004                                        min =          1
     between = 0.0007                                        avg =       26.0
     overall = 0.0004                                        max =         45
                                              Wald chi2(1)       =     146.31
corr(u_i, X)   = 0 (assumed)                  Prob > chi2        =     0.0000
------------------------------------------------------------------------------
    dolemerg |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    dolprimc |  -.0214857   .0017763   -12.10   0.000    -.0249672   -.0180043
       _cons |   47.62394   .6888659    69.13   0.000     46.27378    48.97409
-------------+----------------------------------------------------------------
```


pvar Emergency $s ->Primary care $s

```
. pvar  dolprimc  dolemerg  , lags(6) No. of obs        =    168681
No. of panels    =      9644            Ave. no. of T  =    17.491
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
dolprimc     |
    dolprimc |
         L1. |   .2746698   .1290072     2.13   0.033     .0218204    .5275192
         L2. |  -.0485268   .0453748    -1.07   0.285    -.1374597    .0404062
         L3. |   .0190768   .0153223     1.25   0.213    -.0109544     .049108
         L4. |  -.0008771   .0096073    -0.09   0.927    -.0197071    .0179529
         L5. |  -.0072563   .0102543    -0.71   0.479    -.0273544    .0128418
         L6. |   .0043254   .0049237     0.88   0.380     -.005325    .0139757
    dolemerg |
         L1. |   .0052949   .0023697     2.23   0.025     .0006504    .0099395
         L2. |    .001567   .0016548     0.95   0.344    -.0016764    .0048104
         L3. |   .0003862   .0009709     0.40   0.691    -.0015167    .0022891
         L4. |   .0017168   .0009404     1.83   0.068    -.0001262    .0035599
         L5. |    .001087   .0010514     1.03   0.301    -.0009737    .0031477
         L6. |   .0012177   .0009675     1.26   0.208    -.0006785    .0031139
```

Primary care $s ->Emergency $s

```
. pvar  dolprimc  dolemerg  , lags(6) No. of obs       =    168681
No. of panels    =       9644              Ave. no. of T   =    17.491
                 |       Coef.   Std. Err.       z    P>|z|     [95% Conf. Interval]
dolemerg         |
    dolprimc     |
         L1.     |    .0047209   .0031455     1.50   0.133    -.0014441    .0108859
         L2.     |    .006232     .003401     1.83   0.067    -.0004337    .0128978
         L3.     |    .0032212   .0028914     1.11   0.265    -.0024459    .0088882
         L4.     |    .0098327    .004332     2.27   0.023     .0013422    .0183233
         L5.     |   -.0036448   .0039193    -0.93   0.352    -.0113265    .0040368
         L6.     |    .0245599   .0061225     4.01   0.000     .0125601    .0365598
    dolemerg     |
         L1.     |    .1230638   .0233041     5.28   0.000     .0773886    .1687391
         L2.     |    .0796443   .0134472     5.92   0.000     .0532882    .1060004
         L3.     |    .0649163    .010089     6.43   0.000     .0451422    .0846904
         L4.     |    .0522287   .0124232     4.20   0.000     .0278798    .0765777
         L5.     |    .0411237   .0113239     3.63   0.000     .0189292    .0633181
         L6.     |    .0303653   .0088422     3.43   0.001     .0130349    .0476957
```


pvargranger Primary care $s -!*->Emergency $s

```
. pvargranger

  panel VAR-Granger causality Wald test
    Ho: Excluded variable does not Granger-cause Equation variable
    Ha: Excluded variable Granger-causes Equation variable

  +------------------------------------------------------+
  |  Equation \ Excluded |     chi2     df    Prob > chi2 |
  |----------------------+-------------------------------|
  |dolprimc              |                               |
  |             dolemerg |      9.425     6          0.151 |
  |                  ALL |      9.425     6          0.151 |
  |----------------------+-------------------------------|
  |dolemerg              |                               |
  |             dolprimc |     19.265     6          0.004 |
  |                  ALL |     19.265     6          0.004 |
  +------------------------------------------------------+
```

Males *neither* vs. Females PrC-!*->Em.

```
+-----------------------------------------------------+ MALES
|  Equation \ Excluded |    chi2     df    Prob > chi2  |
 dolprimc             |                                 |
|             dolemerg |    6.326    6         0.388   |
|                  ALL |    6.326    6         0.388   |
|dolemerg             |                                 |
|             dolprimc |    4.375    6         0.626   |
|                  ALL |    4.375    6         0.626   |
+-----------------------------------------------------+ FEMALES
|  Equation \ Excluded |    chi2     df    Prob > chi2  |
|dolprimc             |                                 |
|             dolemerg |    7.462    6         0.280   |
|                  ALL |    7.462    6         0.280   |
 dolemerg             |                                 |
|             dolprimc |  113.310    6         0.000   |
|                  ALL |  113.310    6         0.000   |
```


Primary care $s <->Emergency $s by gender

```
. pvar  dolprimc  dolemerg  , lags(6)
[ONLY approaching sig. shown)
MALES
|      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
PrimC$s = DV
      Emerg$s  |
           L1. |     .01222    .007332     1.67   0.096    -.0021505    .0265905
FEMALES
PrimC$s = DV
 Emerg$s   L3. |   .0018045   .0009995     1.81   0.071    -.0001545    .0037635
           L4. |   .0017593   .0010084     1.74   0.081    -.0002171    .0037357
Emerg$s =Dv
  PrimC$s  L2. |   .0169106   .0061186     2.76   0.006     .0049184    .0289028
           L6. |   .0295973   .0028833    10.27   0.000     .0239461    .0352485
```

Conclusions – email for >: comanus@gmail.com

1. Causal dynamic processes captured by costs of care may differ by gender

2. The pattern of differences merits further investigations

3. Causality with such intensive data offers some opportunities

"As a limitation, we have to emphasize that Granger causality is not true causality."

Tschacher W, Ramseyer F: Modeling psychotherapy processes by time-series panel analysis (TSPA). Psychother Res 2009, 19:469-481