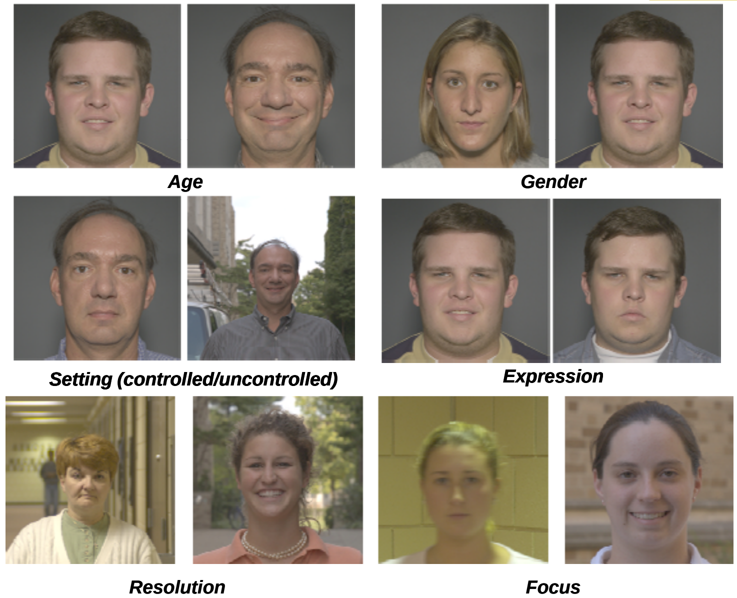
Covariates & Quality Measures

Bruce A. Draper, J. Ross Beveridge,
Yui-Man Lui, David Bolme, Geoff Givens
Colorado State University
P. Jonathan Phillips
National Institute of Standards and Technology

Outline

- Covariate Analysis (a quick review)
  - Methodology: GLMMs
  - Some typical results
- Covariate Meta-analysis
  - Are the conventional wisdoms true?
  - Where is more work needed?
- Quality measures
  - Properties
  - Illumination
  - Focus
- This talk covers results from 3 papers:
  - FRVT 2006: Quo Vidas Face Quality, *To appear in Image and Vision Computing.*
  - A Meta-analysis of Face Recognition Covariates, *IEEE International Conference on Biometric Theory, Applications and Systems (BTAS), 2009*
  - Quantifying How Lighting and Focus Affect Face Recognition Performance, *submitted to IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*

Covariates-Examples



Covariate Analysis

- Ongoing collaboration between CSU & NIST
  - Since 2003
- Six papers
  - Two journal
  - Four conference
  - One workshop
- Three data-sets/challenge problems/evaluations
  - FERET
  - FRGC
  - FRVT 2006

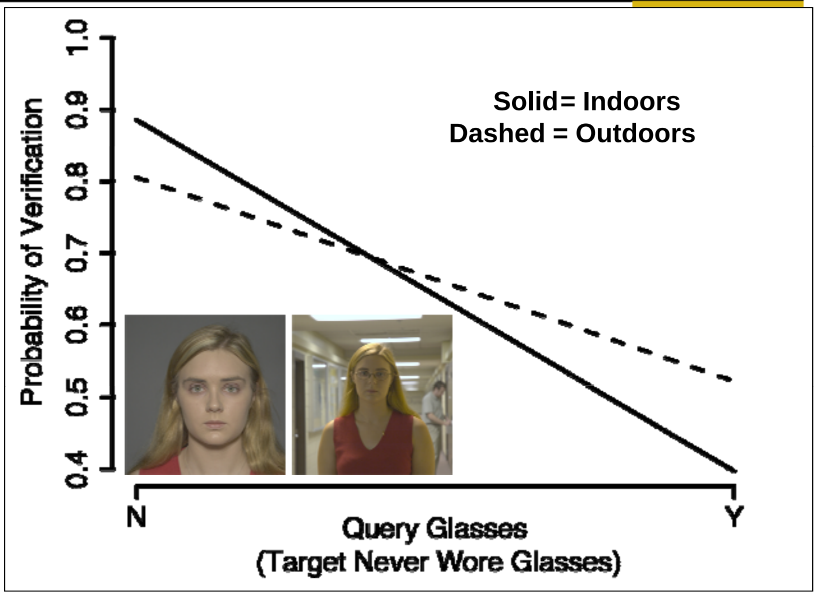
BEST PAPER AWARD

FACE RECOGNITION VENDOR TEST 2006

- Methodology
  - Generalized Linear Mixed Effect Model

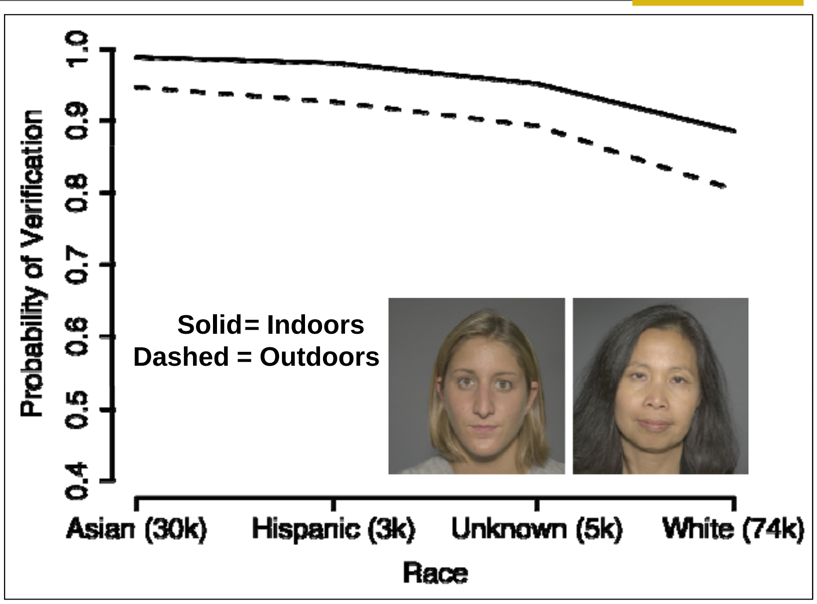
$$\log\left(\frac{p_{padj}}{1-p_{padj}}\right) = \mu + \gamma_a + \gamma_b B + \gamma_j + \gamma_{aj} + \pi_p$$

Covariate Effect - Glasses



5

Covariate Effect - Race



6

Meta-analysis

- The quantitative synthesis or analysis of results from **multiple** experiments or studies
  - Examples
    - Education – Bilingual Education (BRJ, 1997)
    - Medicine – Coronary Heart Disease (BMJ, 2000)
    - Face Recognition, Philips & Newton (AFGR, 2002)
      - Concluded that the majority of FR research papers were working on "easy" problems and that testing of novel algorithms should be accompanied by a control algorithm.
    - Iris Recognition, Newton & Phillips (BTAS, 2007)
      - Concluded the results from ITIRT, Iris '06, and ICE 2006 are comparable.

Meta-Analysis

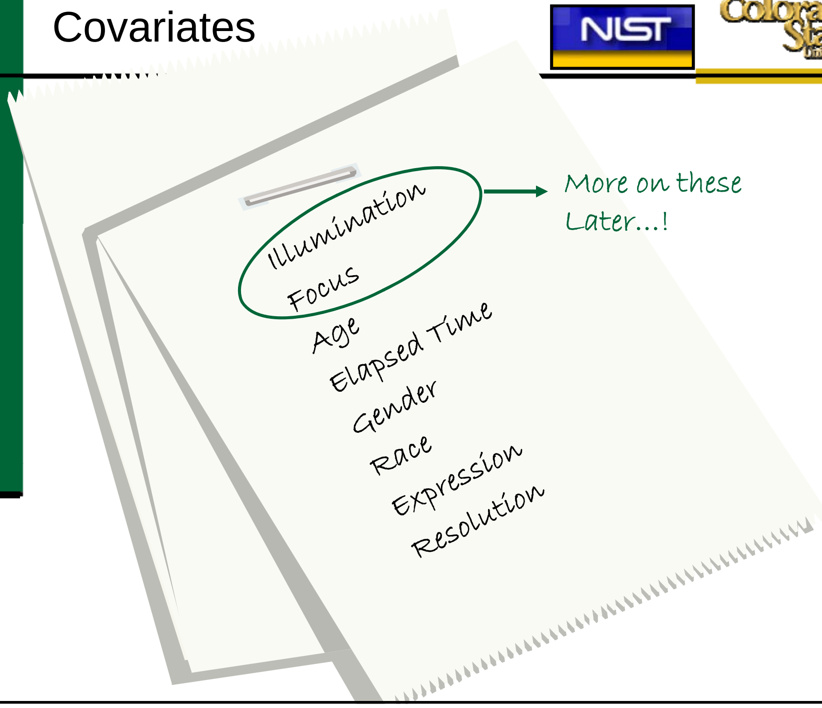
Methodology:

Step 1: Assemble over 100 candidate papers from 1993 through 2008

Step 2: 1st Filter — Paper must relate a factor X to a measured change in recognition performance.

Step 3: 2nd Filter — We must be able to map the effect to a coarse but common quantitative scale.

Covariates



More on these Later…!

Testing Conventional Wisdom

Motivation:

Make sense of …

Women are harder to recognize

Caucasians are hard to recognize

Old people are easier to recognize

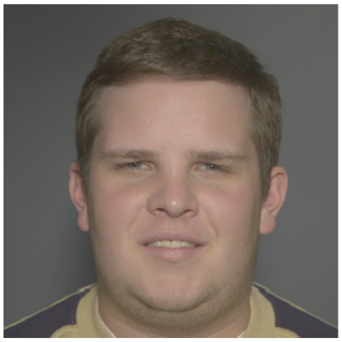
Don't let people smile

Face images bigger than 64x64 are a waste

Same year is as good as same week

"A Meta-Analysis of Face Recognition Covariates," Y. Man Lui, D. Bolme, B. A. Draper, J. R. Beveridge, G. Givens, P. J. Phillips, In Proceedings, *Third IEEE International Conference on Biometrics: Theory, Applications, and Systems*, 2009.

Who is easier to recognize?



Old people are easier to recognize

Young ***…***

Conventional wisdom: older subjects are easier to recognize than young ones

Analysis of Age as a Covariate

Here is a glimpse of detail presented in Meta-Analysis …

Older Easier							Younger Easier								
+17.5%	15%	10%	5%	0%	5%	10%	15%	+17.5%	Interactions with other covariates	Controls for other covariates (Table IX)	Algorithm (Table VII)	Dataset (Table VIII)	Author	Year	Source
		*	*						Outdoor/Indoor	Ctrl	Anon1	FRVT2006	Beveridge	2009	[9]
	*	*							Outdoor/Indoor	Ctrl	Anon2	FRVT2006	Beveridge	2009	[9]
			*							Ctrl	Anon3	FRVT2006	Beveridge	2009	[9]
		*								Ctr2	NJIT	FRGC	Beveridge	2009	[7]
			*							Ctr2	PCA1	FRGC	Beveridge	2009	[7]
*										Ctr2	CMU	FRGC	Beveridge	2009	[7]
*											Bayes2	FG-NET	Park	2008	[34]
*											Gabor	FG-NET + WVU	Singh	2007	[44]
				*						Ctr4	PCA1	FERET	Ho	2007	[23]
	*									Ctr5	PCA1	FERET	Givens	2005	[20]
			*							Ctr6	EBGM	FERET	Givens	2004	[19]
		*								Ctr6	PCA1	FERET	Givens	2004	[19]
		*								Ctr6	Bayes2	FERET	Givens	2004	[19]
		*								Ctr6	PCA1	FERET	Givens	2003	[21]
	*								Gender		Cog	HCINT	Phillips	2002	[35]
*									Gender		Idtx	HCINT	Phillips	2002	[35]
		*							Gender		Eyem	HCINT	Phillips	2002	[35]
			*						Gender		Imagis	HCINT	Phillips	2002	[35]
		*							Gender		Viisage	HCINT	Phillips	2002	[35]
		*							Gender		VisSph	HCINT	Phillips	2002	[35]
		*							Gender		C-Vis	HCINT	Phillips	2002	[35]
				*					Gender		DrmMIRH	HCINT	Phillips	2002	[35]

… and this is more detail then appropriate here.

Summary of Findings

AGE:

Older people are easier to recognize.
(9 Studies)

Elapsed Time

The older the image, the poorer the match

14 days

140 days

238 days

target

Conventional wisdom: smaller time delays (query to target) are easier

Summary of Findings

AGE:
Older people are easier to recognize. (9 Studies)

TIME BETWEEN IMAGES:
Recognition degrades with time between images.

Gender

Conventional wisdom: men are easier to recognize than women

Summary of Findings

AGE:
Older people are easier to recognize. (9 Studies)

TIME BETWEEN IMAGES:
Recognition degrades with time between images. Months and years matter. (8 Studies)

GENDER:
Murky outcome, modest and depends upon study, algorithm, setting, etc. (8 Studies)

Resolution




Face images bigger than 64x64 are a waste

Conventional wisdom: low resolution imagery is sufficient for face recognition

Summary of Findings

AGE:
Older people are easier to recognize. (9 Studies)

TIME BETWEEN IMAGES:
Recognition degrades with time between images. Months and years matter. (8 Studies)

GENDER:
Murky outcome, modest and depends upon study, algorithm, setting, etc. (8 Studies)

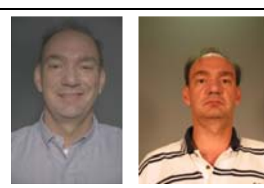
RESOLUTION:
Older algorithms don't care. Newer algorithms like more pixels. (10

Expression

Don't let people smile

Neutral to Neutral

Smiling to Smiling



Smiling to Neutral

Neutral to Smiling

Conventional wisdom:
always match neutral expressions

Summary of Findings

AGE:
Older people are easier to recognize. (9 Studies)

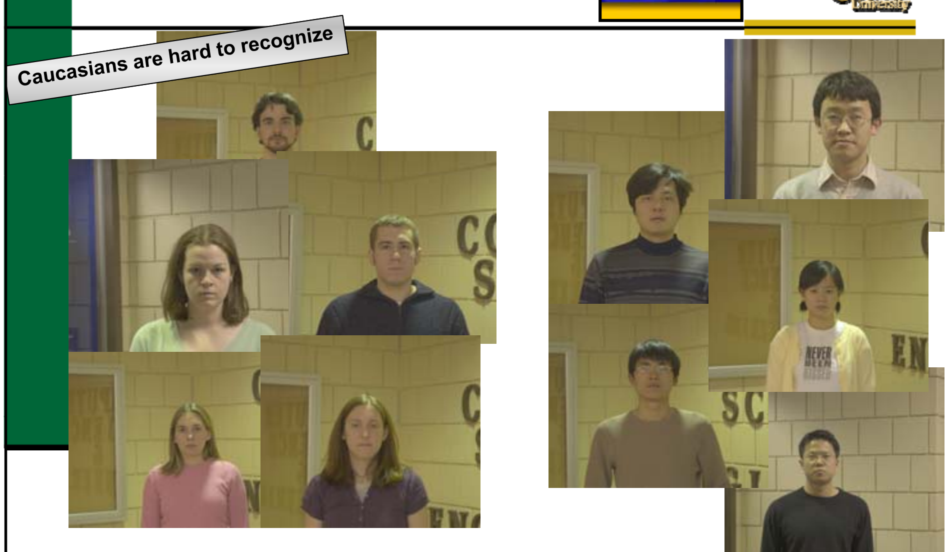
TIME BETWEEN IMAGES:
Recognition degrades with time between images. Months and years matter. (8 Studies)

GENDER:
Murky outcome, modest and depends upon study, algorithm, setting, etc. (8 Studies)

RESOLUTION:
Older algorithms don't care. Newer algorithms like more pixels. (10 Studies)

EXPRESSION:
Same expression better, Otherwise smile/neutral same. (4 Studies)

Race



Conventional wisdom: Caucasians are more difficult to recognize than East Asians

Meta-analysis of Race

All of these studies confound Race with sampling effects

All systems trained on majority-Caucasian data sets

All systems tested on majority-caucasian data sets (fewer possible East Asian confusions)

Therefore, no conclusion is supported

Summary of Findings

AGE:
Older people are easier to recognize. (9 Studies)

TIME BETWEEN IMAGES:
Recognition degrades with time between images. Months and years matter. (8 Studies)

GENDER:
Murky outcome, modest and depends upon study, algorithm, setting, etc. (8 Studies)

RESOLUTION:
Older algorithms don't care. Newer algorithms like more pixels. (10 Studies)

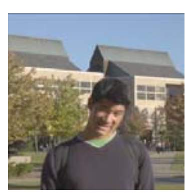
EXPRESSION:
Same expression better, Otherwise smile/neutral same. (4 Studies)

RACE:
East Asians easier, BUT, this may be because fewer East Asians in data sets. (6 Studies)

What is image quality?

Depends on the query/target pair

Smiling



Size

Focus

Quality Measures

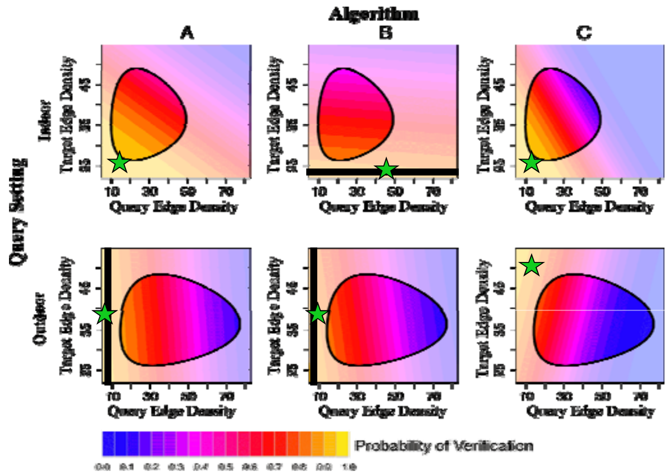
- Quality measures should be:
  - Statistically predictive of success
  - Directly computable from an image pair
  - Explainable
  - Operationally Controllable

Quality Measures

Edge Density

Edge Density Effects

Algorithms and Location Matter



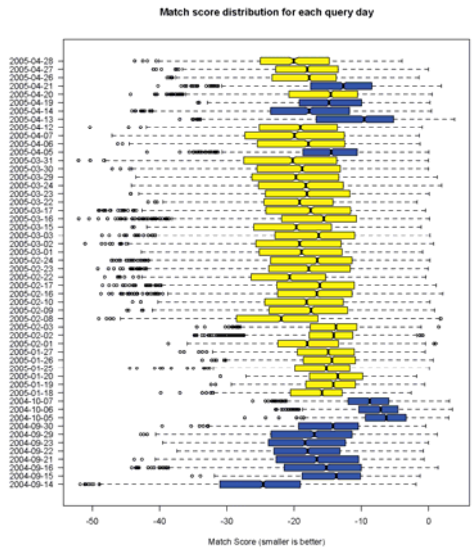
FRVT 2006: Quo Vidas Face Quality" J. Ross Beveridge, Geof H. Givens, P. Jonathon Phillips, Bruce A. Draper, David S. Bolme, Yui Man Lui. Image and Vision Computing, Under Review.

Edge Density: Why?

- Why is edge density predictive of recognition performance across algorithms?
- Why is *lower* edge density better than high?
  - One post-hoc explanation: edge density as a focus measure
    - Implies that out-of-focus is good
  - Another post-hoc explanation: edge density as an indirect lighting measure
- Focus & illumination as possible quality measures

Post-hoc Quality

- In FRVT 2006, the best predictor of recognition rate is *date of acquisition*.
- Date of acquisition corresponds to setting.



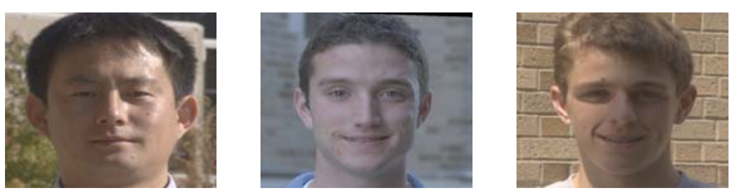
Date of Acquisition

- Why? Date of acquisition subsumes:
  - Image location (camera set up once per day)
    - Backgrounds
    - Illumination
  - Approximate time of day (short sessions)
    - Relates to lighting in outdoor settings
  - Possibly focus
    - Some sessions in better focus than others?
- A good quality measure should be as good as date of acquisition on FRVT 2006, but generalizable across data sets

Illumination

- Extensively studied
  - PIE (shown) & Yale B
- Question: have modern algorithms "solved" lighting?

Illumination Models



-0.23879	0.067728	-0.237399
Left	Frontal	Right

$$BS = X \longrightarrow S = B^{-1}\hat{X}$$

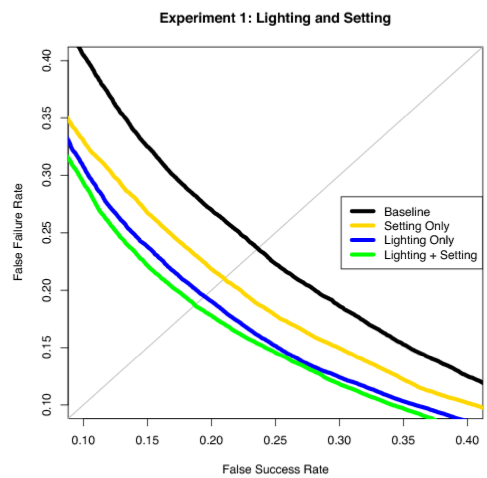
where $\hat{X} = \frac{\alpha_k \mu_k}{\sum_k \alpha_k}$ and $\alpha_k = \exp\left(-\frac{\| X - \mu_k \|^2}{\sigma^2}\right)$

Y.M Lui, JR Beveridge, and M. Kirby, Canonical Stiefel Quotient and its Application to Generic Face Recognition in Illumination Spaces, BTAS, 2009.

T. Sim and T. Kanade, Combining Models and Exemplars for Face Recognition: An Illumination Example, CVPR Workshop, 2001.

Lighting Direction as Quality Measure

- Directly estimated from query image
  - Target images had frontal illumination
- Highly predictive of success
- Explains most of the information in setting (date)

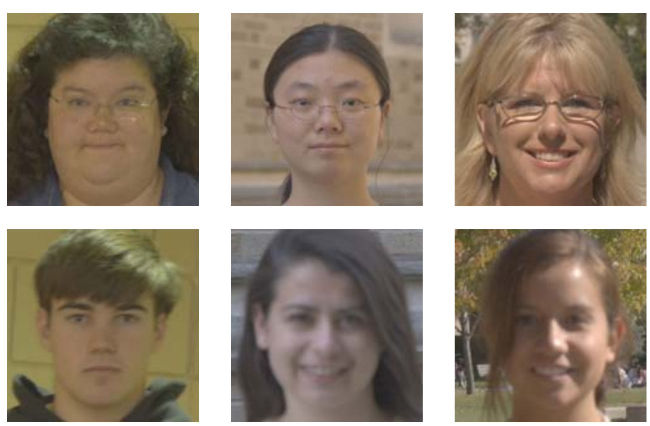


SEMC Focus Measure

- We developed a “true” focus measure

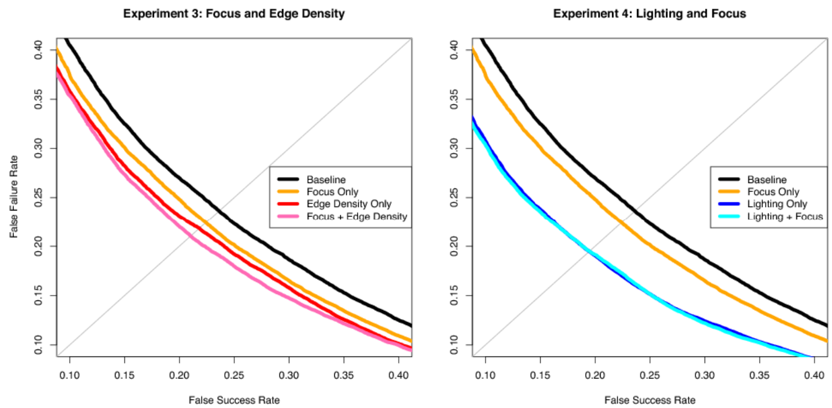
In Focus

Out of Focus



Focus ≠ Edge Density

- Focus less predictive than as edge density
- Focus is subsumed by lighting



Recent Conclusions

- Lighting direction is an important quality measure
  - Implies lighting has not been “solved”
- Lighting direction explains previous edge density result
  - Edge density loses significance when lighting direction feature is added
- Focus not significant in FRVT 2006.

Summary

- Some Covariates Matter
  - Age, Time Delay are important
  - Gender less so
  - Resolution depends on algorithm
  - Race, Expression : more work needed
- Quality Measures
  - Its the lighting, stupid
  - Focus is insignificant in FRVT