Model assumptions & extending the twin model

Matthew Keller
Sarah Medland
Hermine Maes

Boulder 2010

Acknowledgments

- John Jinks
- David Fulker
- Robert Cloninger
- Lindon Eaves
- Andrew Heath

Files you will need are in F:/matt/Assumptions

- Assumptions_mck.pdf (the powerpoint)
- CTD.ACDE-param.indet.R (OpenMx script)

Structural Equation Modeling (SEM) in BG

- SEM is great because…
  - Directs focus to effect sizes, not “significance”
  - Forces consideration of causes and consequences
  - Explicit disclosure of assumptions
- Potential weakness…
  - Parameter reification: “Using the CTD we found that 50% of variation is due to A and 20% to C.”

True parameters vs. Estimated parameters

A C D E: true (unknowable) values of A, C, D, E in the population

$\hat{A}\ \hat{C}\ \hat{D}\ \hat{E}$: *estimated* values of A, C, D, E.

$\hat{A}\ \hat{C}\ \hat{D}\ \hat{E}$ will differ from A, C, D, E due to:

1) sampling variability
2) bias

Quiz Question 1

1) $\hat{\mathrm{C}}$ and $\hat{\mathrm{D}}$ cannot be estimated simultaneously in the classical twin design (i.e., the design that uses MZ and DZ twins only) model because: [choose all that apply]

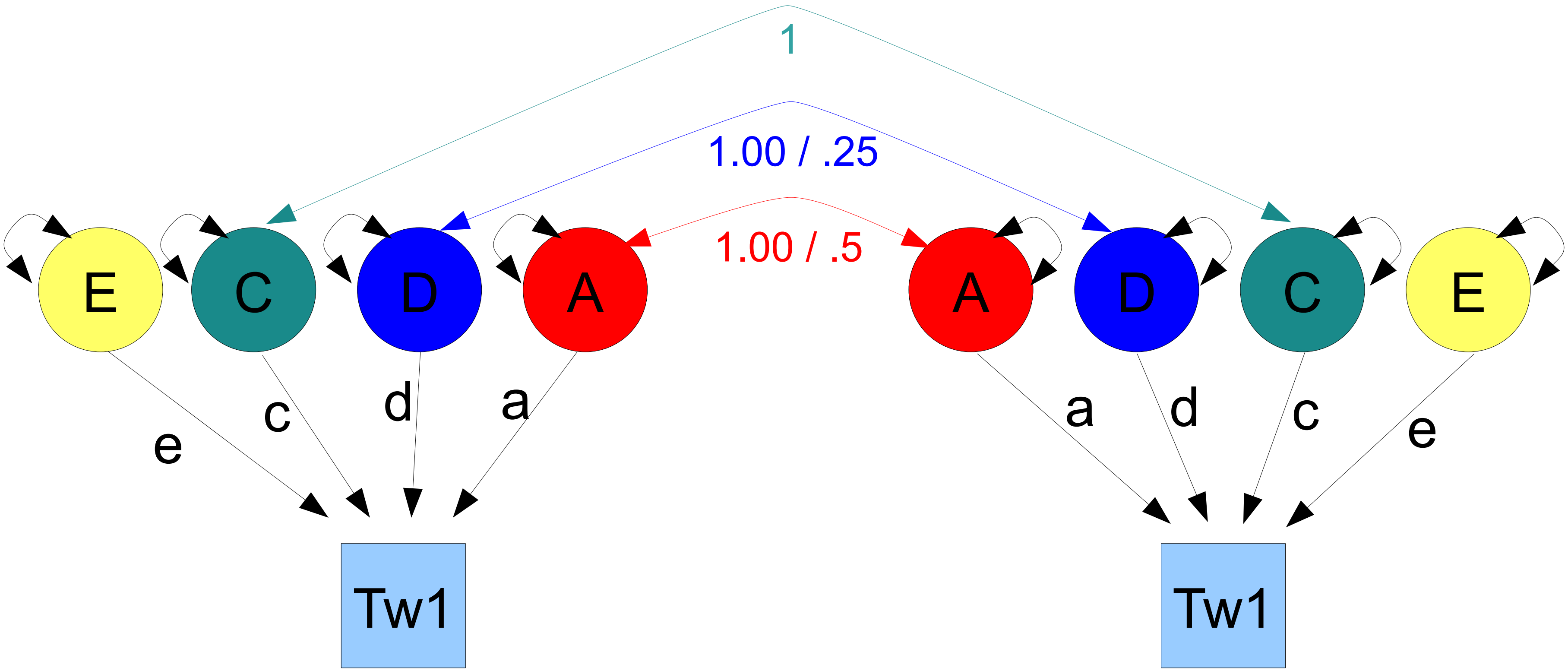
a) these two estimates are too highly correlated (multicolinearity problems)

b) they **can** be estimated simultaneously; you just have to fix $\hat{\mathrm{A}}$ to some specific value

c) there are more observed statistics than parameters to be estimated

d) because the workshop teachers said so

The Classical Twin Design



Why can't we estimate $\hat{C}$ & $\hat{D}$ at same time using twins only?

- Solve the following two equations for $\hat{A}$, $\hat{D}$, & $\hat{C}$:

$$CVmz = A + D + C$$

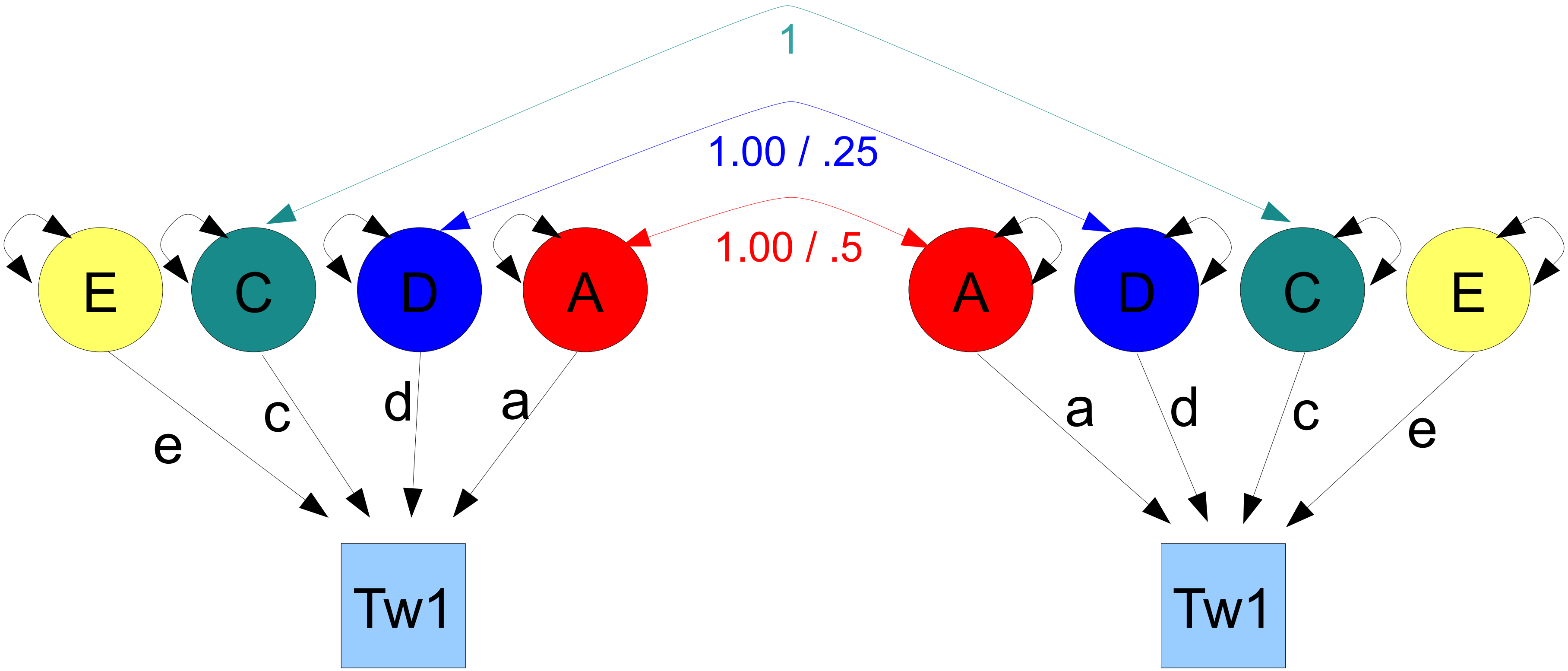
$$CVdz = 1/2A + 1/4D + C$$

- 3 unknowns, 2 informative equations. It can't be done. The model is "unidentified".
- In practice, you can detect non-identification by noting that (a) model estimates depend on starting values AND (b) all final models have identical likelihoods

Indeterminacy: Practical 1

- Open up CTD.ACDE-param.indet.R in R
- Run this script until you see "# END PRACTICAL 1." Don't close the script or R, as we'll use this same script again for Practicals 2 & 3.
- Write down your -2 log likelihood and your estimates of A, C, and D
- Compare these to your neighbor's results
- WHY is this occurring?

The CTD: Two statistics give info about within-family resemblance



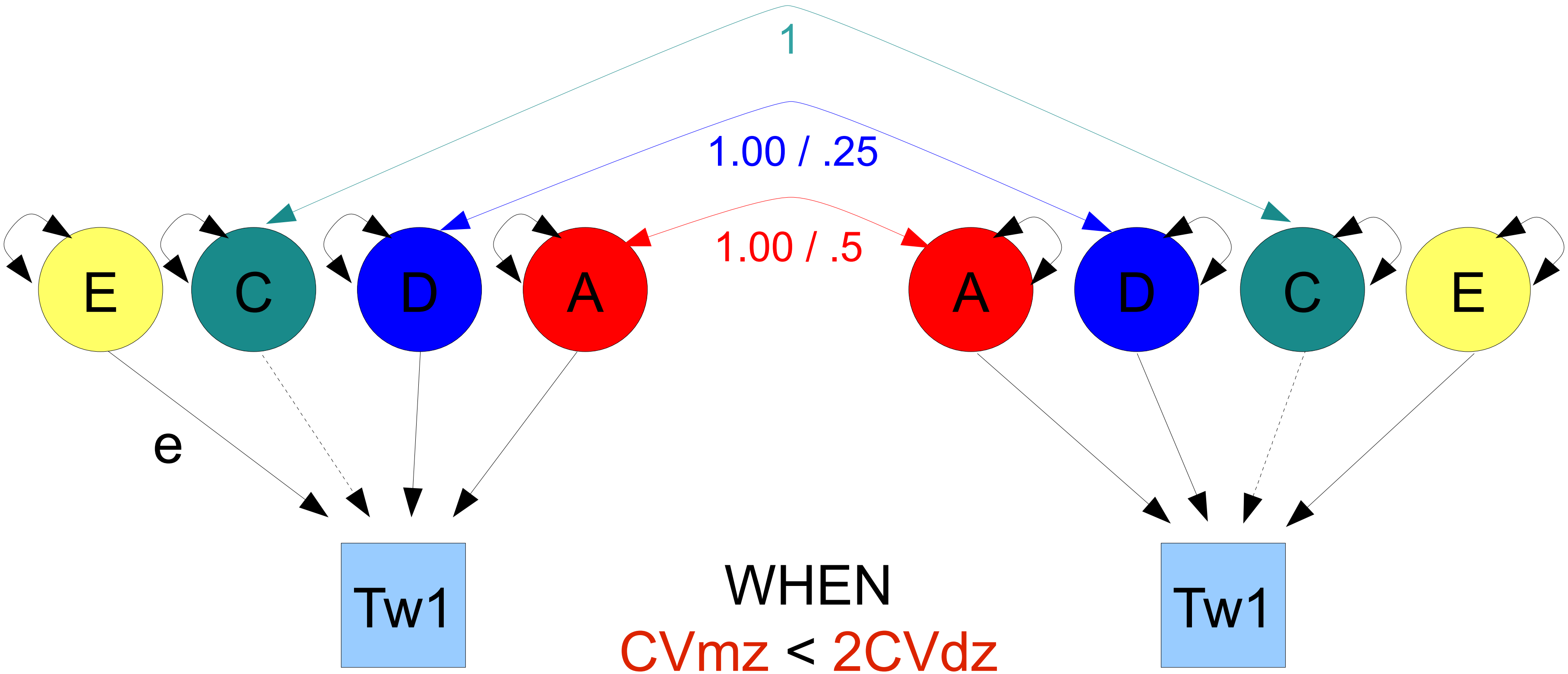
MZ covariance

Vp	
CVmz	Vp

DZ covariance

Vp	
CVdz	Vp

ACE Model



WHEN
CVmz < 2CVdz

Vp	
CVmz	Vp

Vp	
CVdz	Vp

ACE Algebra

- Assume D = 0. Solve for $\hat{A}$ & $\hat{C}$

CVmz = A + C

CVdz = 1/2 A + C

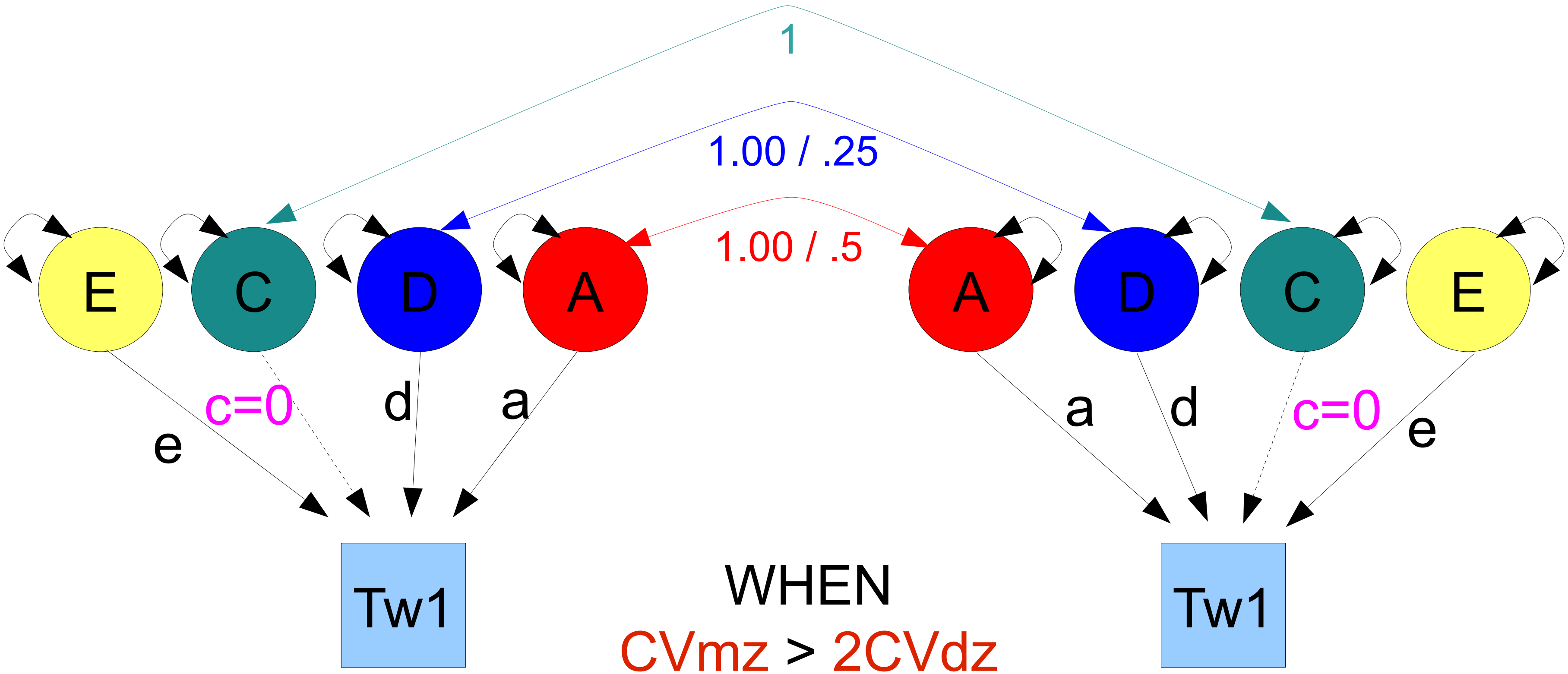
- 2 unknowns, 2 independently informative equations:

$\hat{A}$ = 2(CVmz-CVdz)

$\hat{C}$ = 2CVdz-CVmz

Note: if we tried to estimate $\hat{D}$, it would necessarily hit the 0 boundary anyway and the model wouldn't fit as well (because $\hat{D}$ 'wants' to go negative), so it makes sense to solve for $\hat{C}$

The CTD: ADE Model



WHEN

CVmz > 2CVdz

Vp	
CVmz	Vp

Vp	
CVdz	Vp

PRACTICAL 2: ADE Algebra & Indeterminacy

- Assume C = 0. Solve for $\hat{A}$ & $\hat{D}$

$CVmz = A + D$

$CVdz = \frac{1}{2}A + \frac{1}{4}D$

Solve for these yourselves algebraically now

- Then reopen CTD.ACDE-param.indet.R in R & run FROM "# START PRACTICAL 2"
  TO "# END PRACTICAL 2"

- Did you get roughly the same answer as your algebra suggested?

- What happened to estimates of C & D in the DCE model?

Review Question 1 again

1) $\hat{C}$ and $\hat{D}$ cannot be estimated simultaneously in the classical twin design (i.e., the design that uses MZ and DZ twins only) model because: [choose all that apply]

a) these two estimates are too highly correlated (multicolinearity problems)

b) they **can** be estimated simultaneously; you just have to fix $\hat{A}$ to some specific value

c) there are more observed statistics than parameters to be estimated

d) because the workshop teachers said so

Quiz Question 2

2) What are the *typical* assumptions of a classical twin model? [choose all that apply]

a) the MZ and DZ covariances are equal

b) either D or C is equal to zero

c) only genetic factors cause MZ twins to be more similar to each other than DZ twins

d) no assortative mating

e) no epistasis

f) no gene-environment interactions or correlations

What are the effects of violations of assumptions in the CTD?

Either D or C is equal to zero: A is overestimated and D and C are underestimated

Only genetic factors cause MZ twins to be more similar to each other than DZ twins: A and D are overestimated and C is underestimated

No assortative mating: A and D are underestimated and C is overestimated

No epistasis: D or A is overestimated and C is underestimated

No gene-environment interactions or correlations: AxC: A overestimated; AxE: E overestimated; Cov(A,C): depends

Bias in parameter estimates

- In ACE Models:

$\hat{A} = A + 3/2VD$

$\hat{C} = C - ½VD$

- In ADE Models:

$\hat{A} = A + 3C$

$\hat{D} = D - 2C$

Quiz Question 3

3) If the assumptions of a model are violated (i.e., not true in the real world)... [choose all that apply]

a) the interpretation of the estimated parameters should be altered; e.g., $\hat{A}$ should be considered an amalgam of A, D, and C and not just additive genetic effects

b) the standard errors of the estimated parameters may be wrong

c) the point estimates of the estimated parameters may be biased

Quiz Question 4

4) An ACE model finds that $\hat{A} = .30$ and $\hat{C} = .15$. This implies that D does not influence the trait in question, or has a minor (non-significant) effect.

a) TRUE

b) FALSE

Quiz Question 5

5) We run an ADE model and find that $\hat{A} = .68$ and that $\hat{D} = .04$. If in truth, C = .10, what will the effect on the estimated parameters be? [choose all that apply]

a) $\hat{D}$ will be biased (too low)

b) $\hat{D}$ will be biased (too high)

c) $\hat{A}$ will be biased (too low)

d) $\hat{A}$ will be biased (too high)

e) there is no affect on the estimated parameters; however by not estimating C (aka, fixing it to zero), we underestimated C

PRACTICAL 3: Sensitivity analysis

- Sensitivity analysis: studying what the effects are on estimated parameters when assumptions are wrong

- In CTD.ACDE-param.indet.R, run:

  FROM "# START PRACTICAL 3"
  TO "# END PRACTICAL 3"

- Change the value of C from 0 to other values (remember, C=c^2). What happens to estimates of A and D depending on different assumed values of C?

Quiz Question 6

6) In the CTD, we have two different relative covariance estimates (MZ covariance & DZ covariance). Let's say we add parents to the twin design. That gives us 2 additional relative covariance estimate to work with (parent-offspring and spousal) and allows us to ____________ [choose all that apply]

a) estimate A, C, & D simultaneously

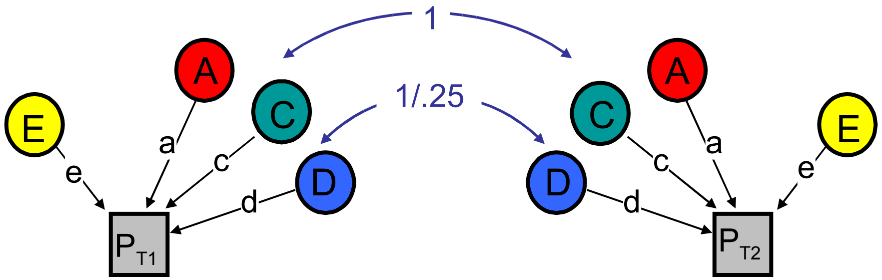
b) account for the effects of assortative mating

c) account for passive G-E covariance

d) reduce the bias in estimates of A, C, and D vis a vis the CTD

Classical Twin Design (CTD)

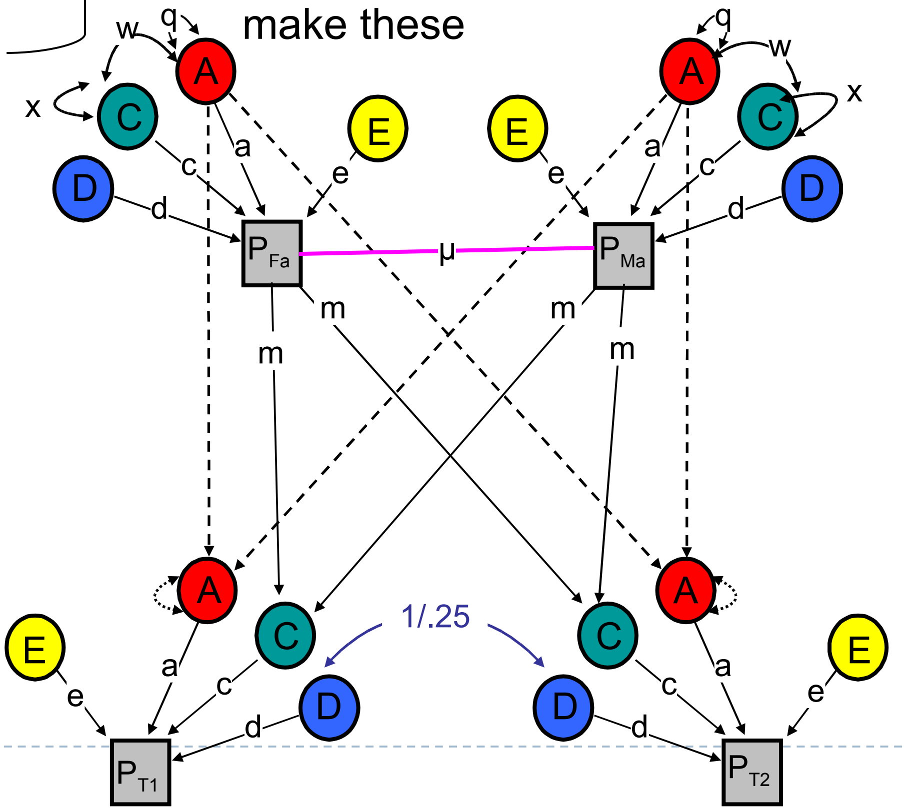
Assumption	biased up	biased down
Either D or C is zero	A	C & D
No assortative mating	C	D
No A-C covariance	C	D & A



Adding parents gets us around all these assumptions

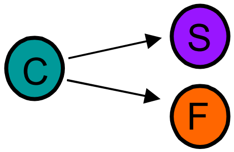
Assumption	biased up	biased down
Either D or C is zero		
No assortative mating		
No A-C covariance		

We don’t have to make these



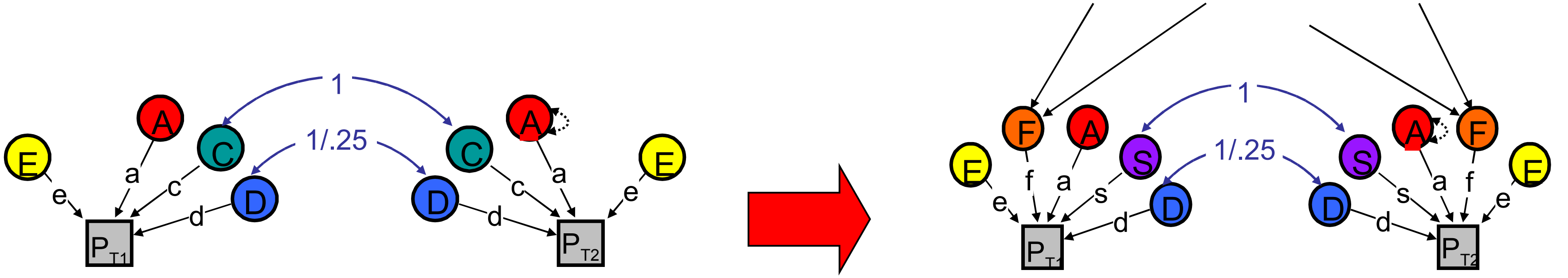
Parents also allow differentiation of S & F

With parents, we can break “C” up into:

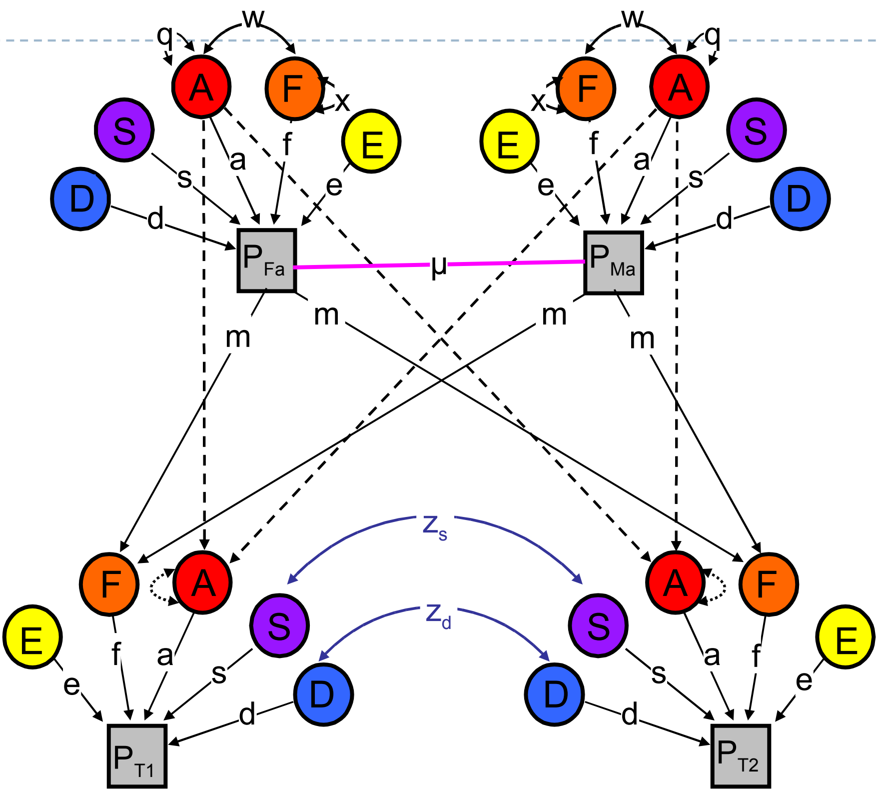


= env. factors shared only between **sibs**

F = **familial** env factors passed from parents to offspring



Nuclear Twin Family Design (NTFD)



Note: m estimated and

f fixed to 1

PRACTICAL 4: NTFD analysis

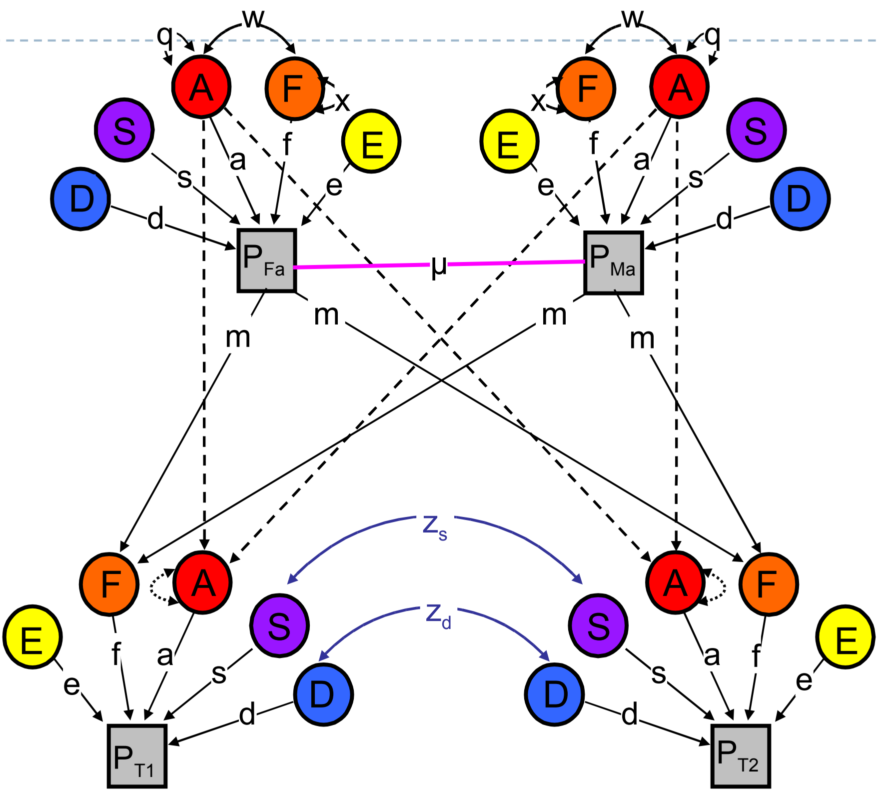
- In CTD.ACDE-param.indet.R, run:
  FROM "# START PRACTICAL 4"
  TO "# END PRACTICAL 4"

- What are the estimated values of A, D, & S [S = sib environment, equivalent to C in the CTD]?

CTD vs. NTFD vs. simulation results

TRUE values	CTD estimates	NTFD estimates
A = .30	A-hat = .68	A-hat = .32
D = .30	D-hat = .04	D-hat = .29
S = .10	S-hat = 0	S-hat = .13

Nuclear Twin Family Design (NTFD)



Note: m estimated and

f fixed to 1

- Assumptions:
  - Only can estimate 3 of 4: A, D, S, and F (bias is variable)
  - Assortative mating due to primary phenotypic assortment (bias is variable)

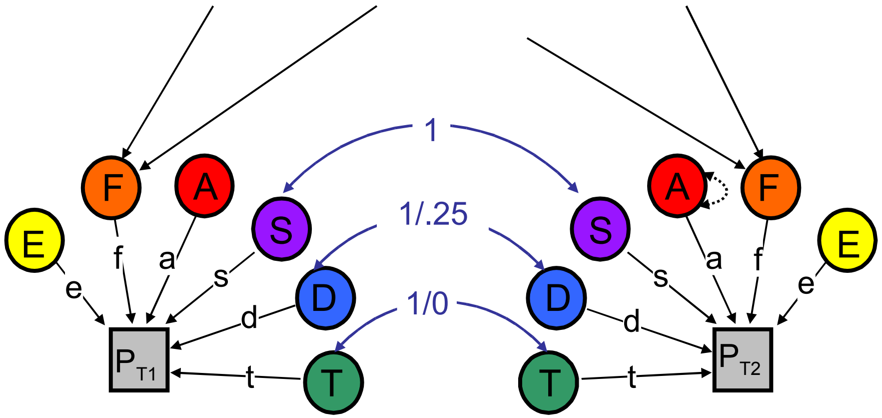
Stealth

- Include twins and their sibs, parents, spouses, and offspring…
  - Gives 17 unique covariances (MZ, DZ, Sib, P-O, Spousal, MZ avunc, DZ avunc, MZ cous, DZ cous, GP-GO, and 7 in-laws)
  - 88 covariances with sex effects

Additional obs. covs with *Stealth* allow estimation of A, S, D, F, T

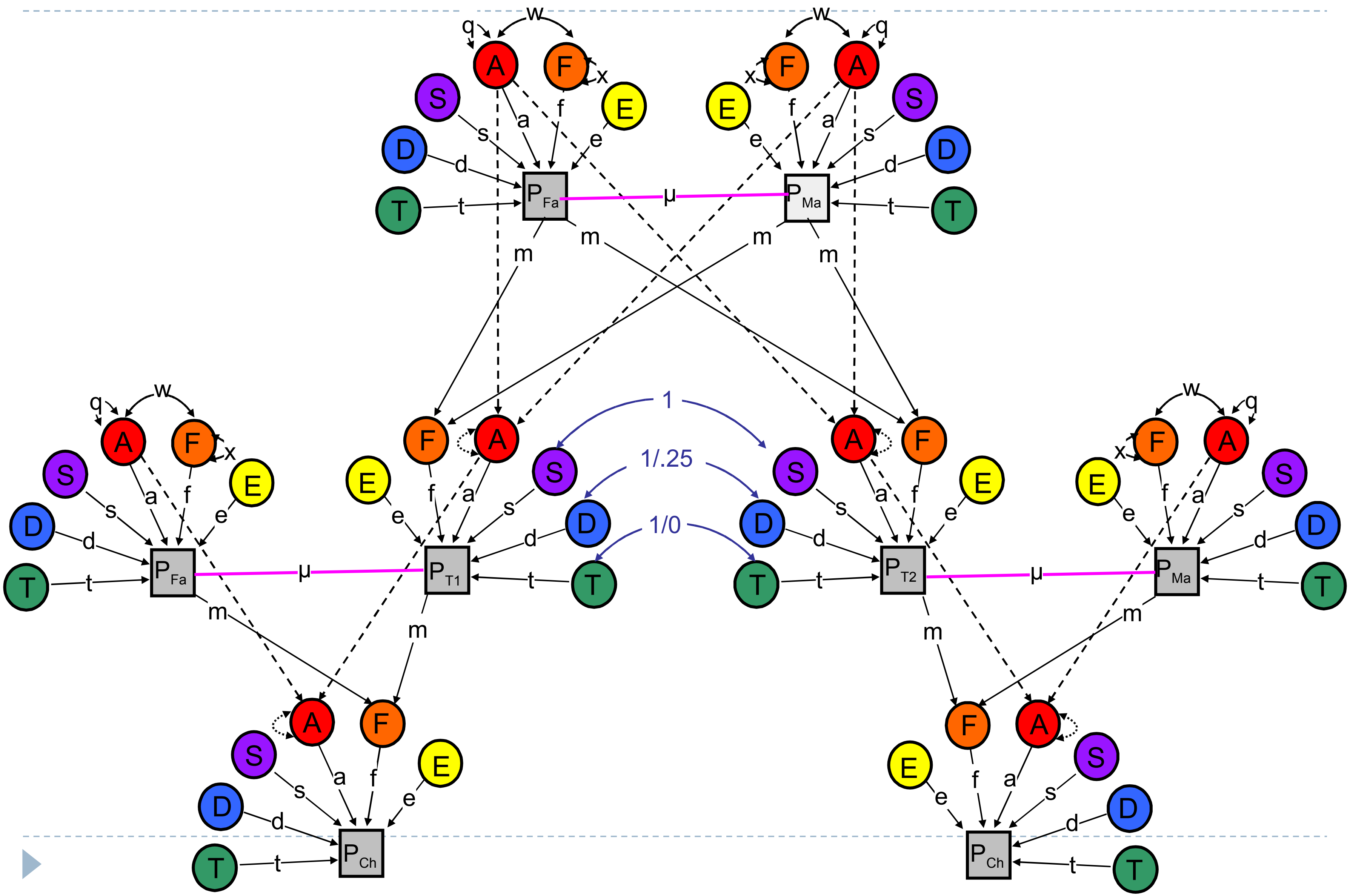
A S F D T can be estimated simultaneously

T = env. factors shared only between **twins**



(Remember: we're not just estimating more effects. More importantly, we're reducing the bias in estimated effects!)

Stealth



Stealth

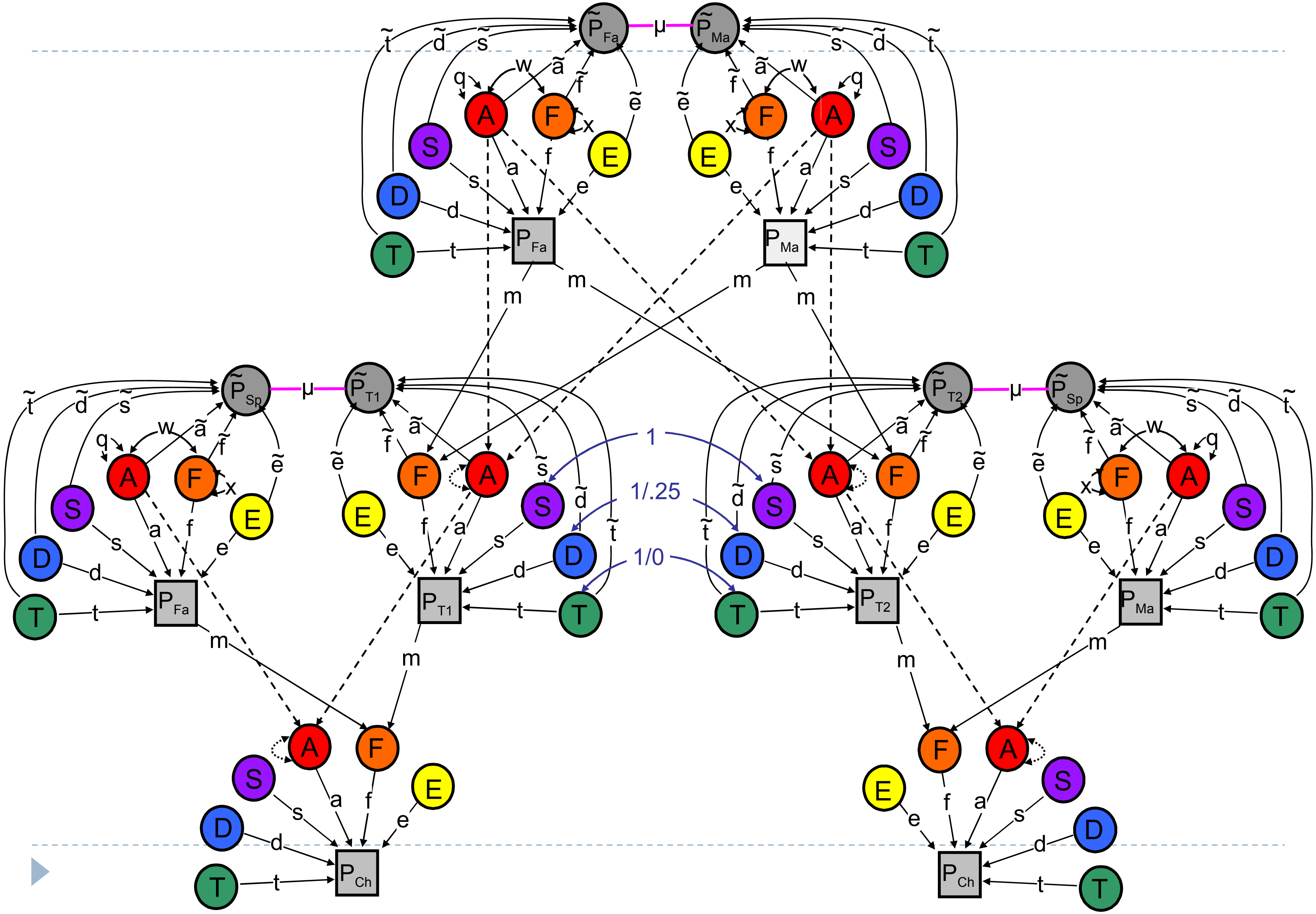
Assumption	biased up	biased down
Primary assortative mating	A, D, or F	A, D, or F
No epistasis	A, D	S
No AxAge	D, S	A

Stealth

Assumption	biased up	biased down
Primary assortative mating	A, D, or F	A, D, or F
No epistasis	A, D	S
No AxAge	D, S	A

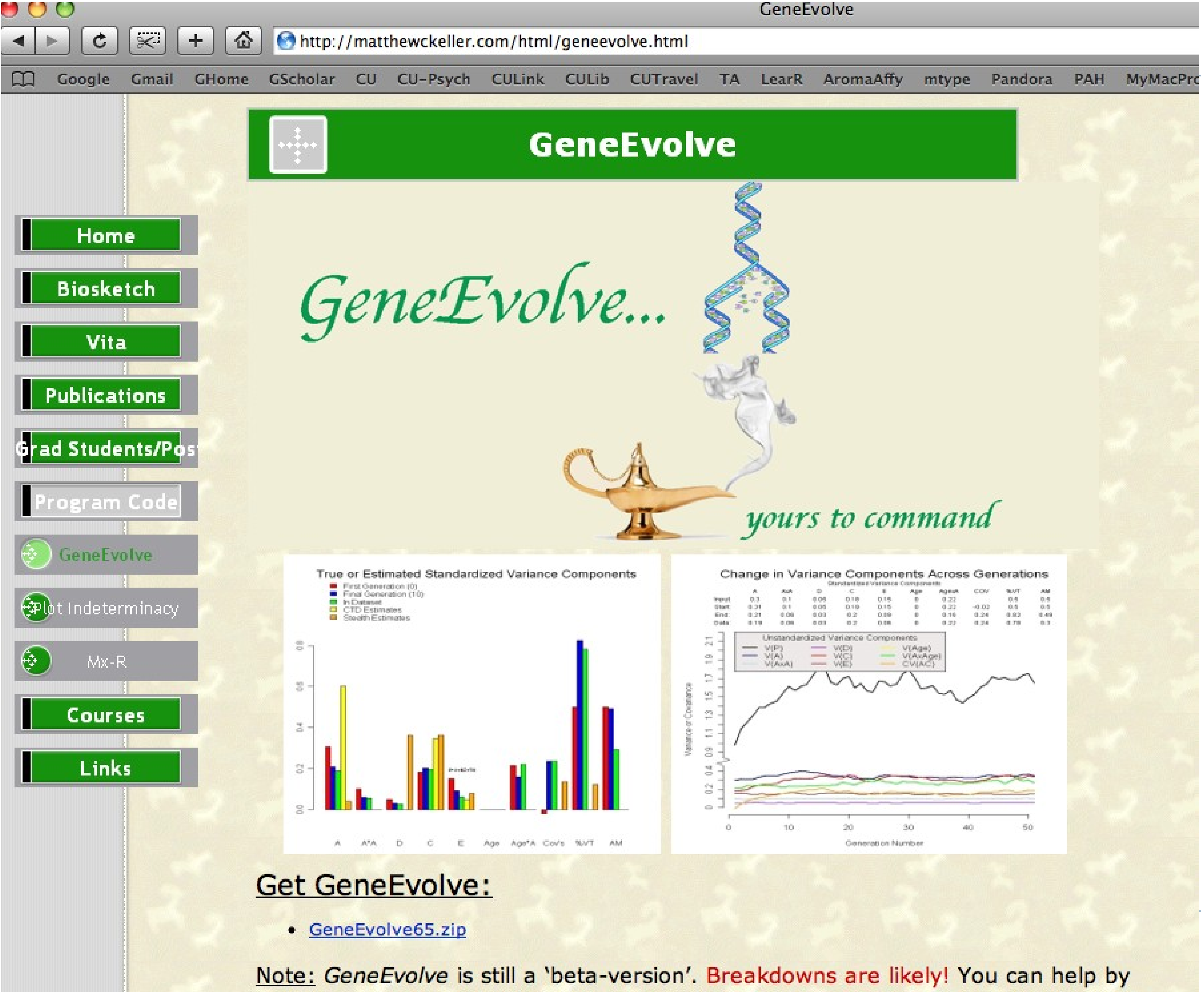
- Primary AM: mates choose each other based on phenotypic similarity
- Social homogamy: mates choose each other due to environmental similarity (e.g., religion)
- Convergence: mates become more similar to each other (e.g., becoming more conservative when dating a conservative)

Cascade


P̃ Fa
μ
P̃ Ma
t̃
d̃
s̃
q
w
ã
f̃
ẽ
A
F
x
E
S
D
T
a
f
s
e
d
t
P Fa
P Ma
m
P̃ Sp
P̃ T1
P̃ T2
P T1
P T2
1
1/.25
1/0
P Ch

Using complex models without independent verification (e.g., simulation) is like… well…

Simulation program: GeneEvolve

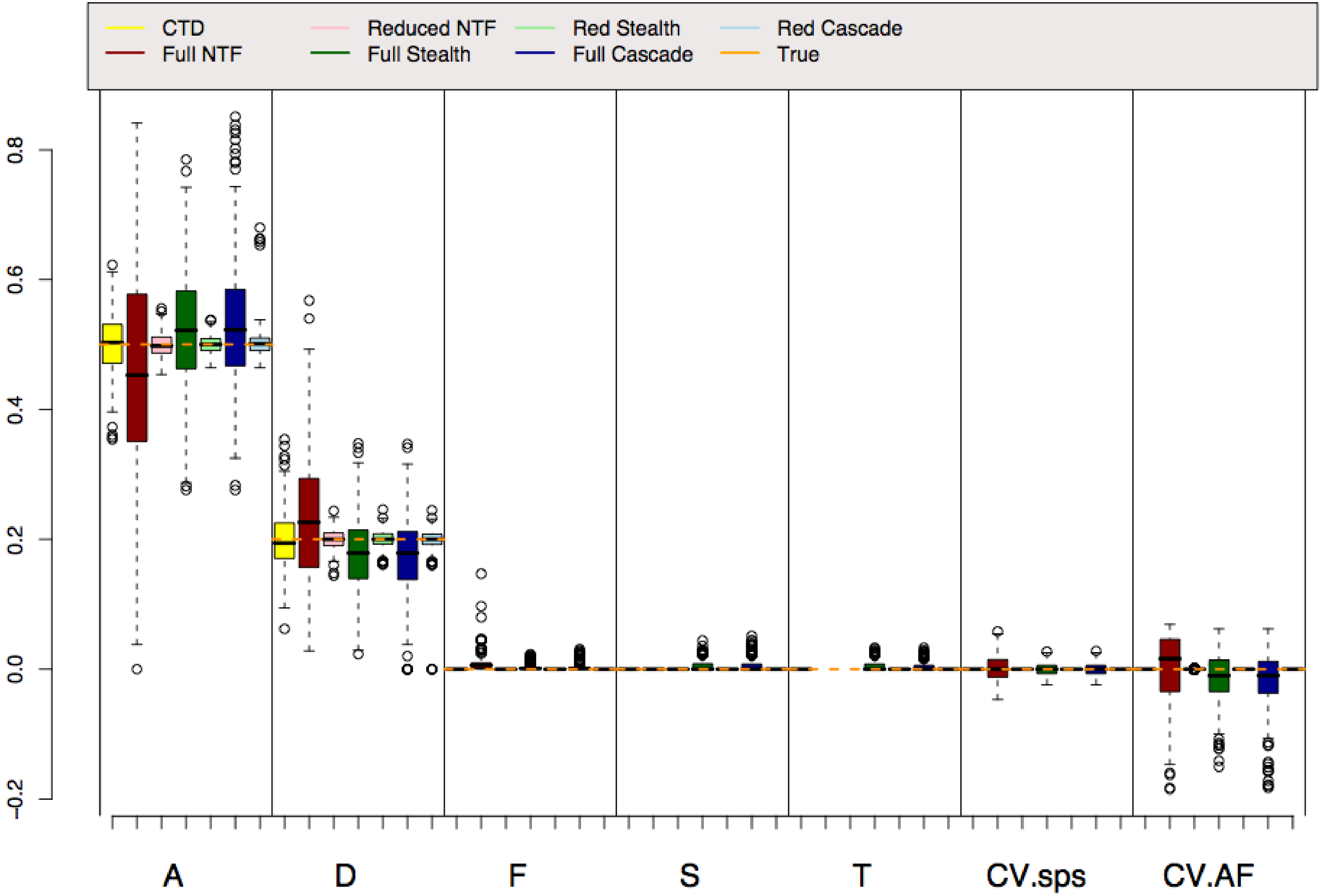


Get GeneEvolve:

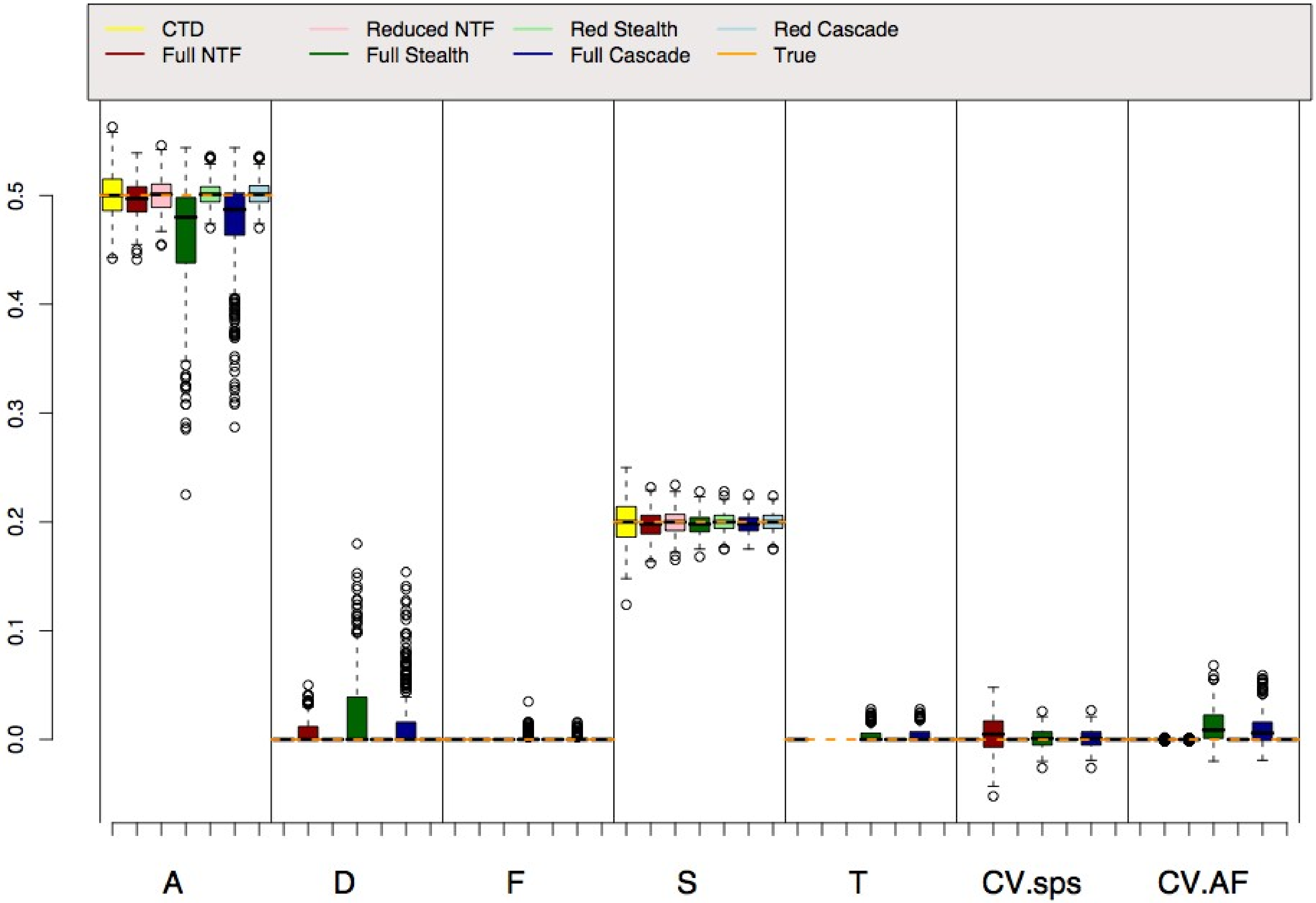
- GeneEvolve65.zip

Note: *GeneEvolve* is still a ‘beta-version’. Breakdowns are likely! You can help by

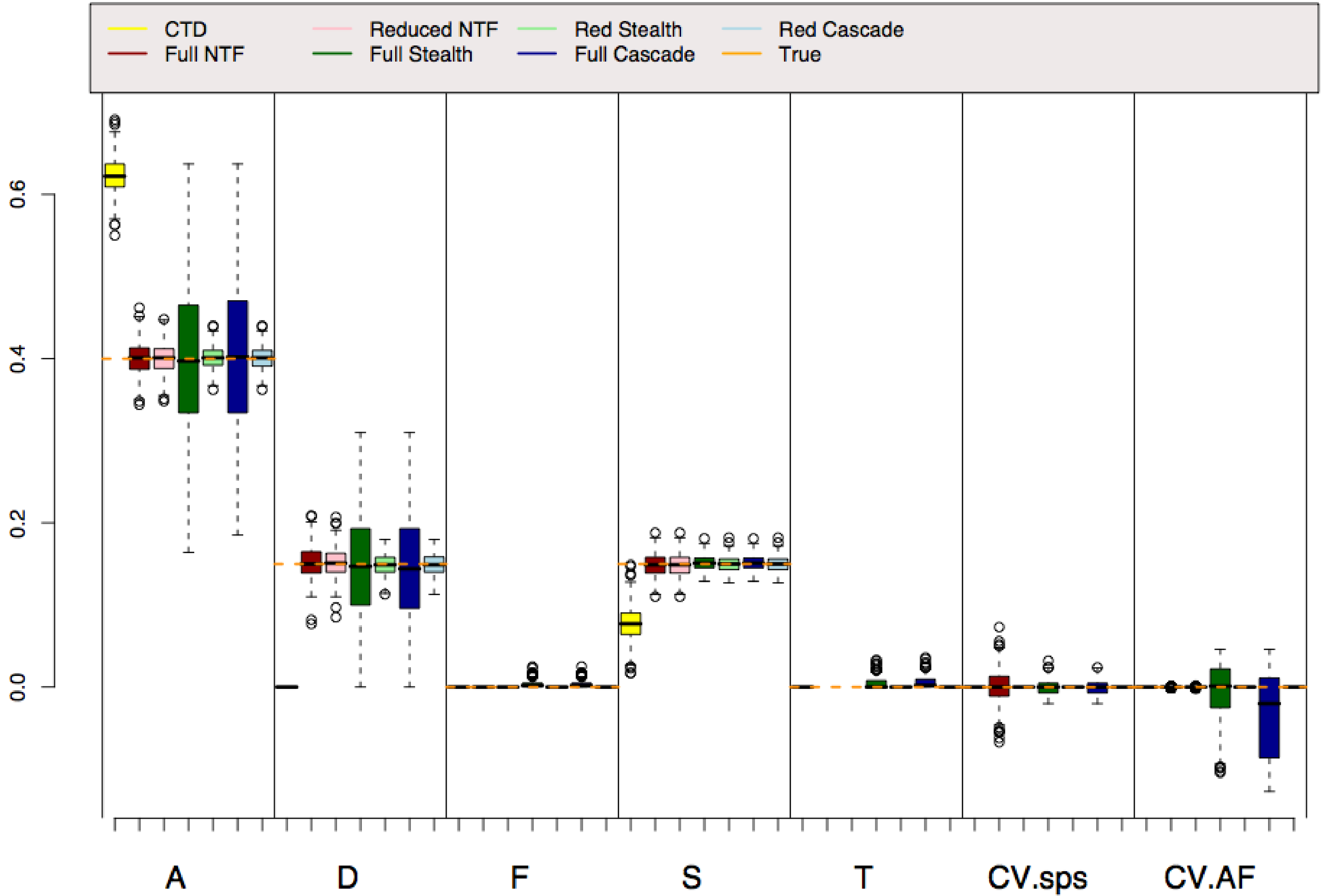
Reality: A=.5, D=.2



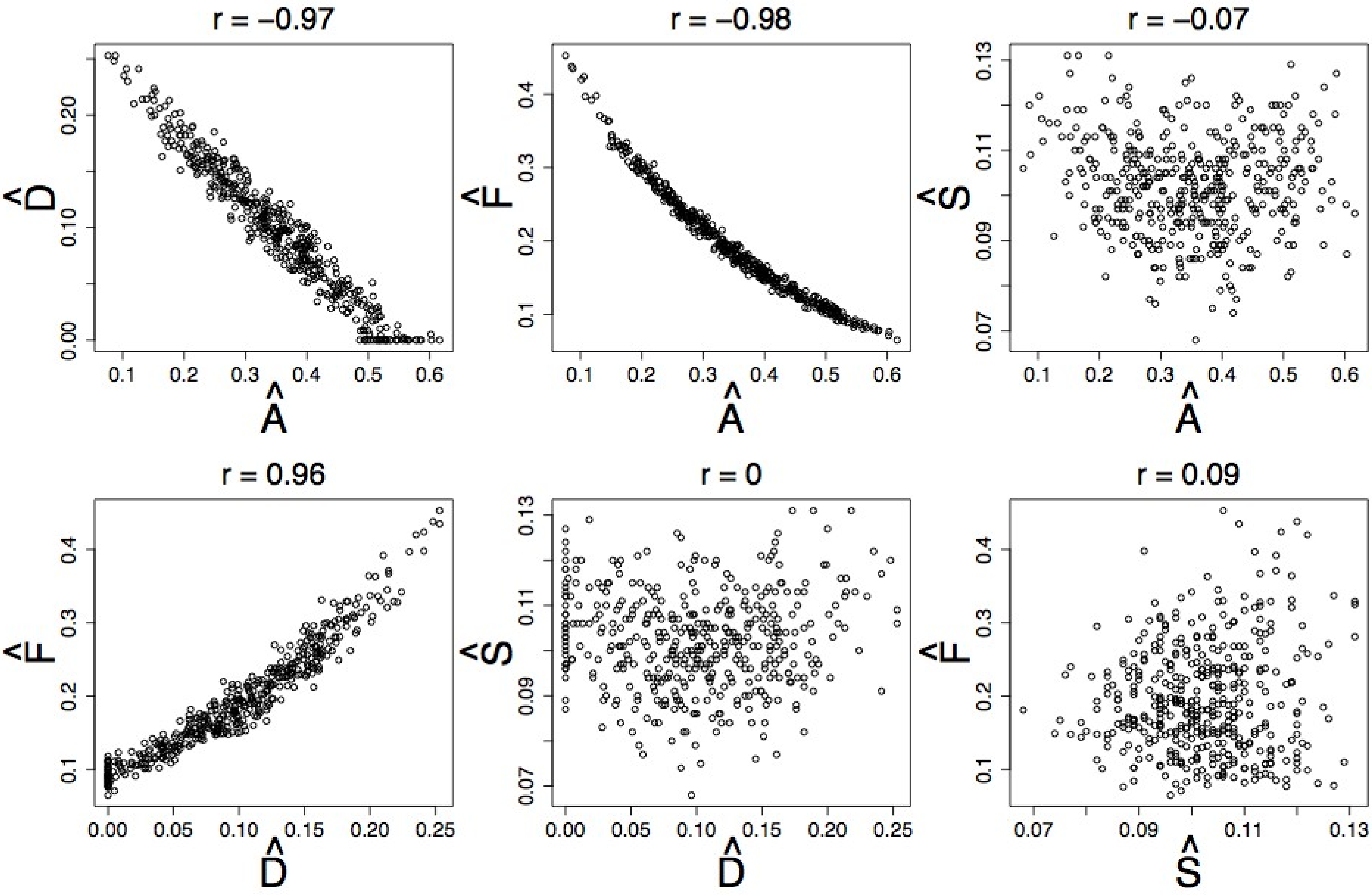
Reality: A=.5, S=.2



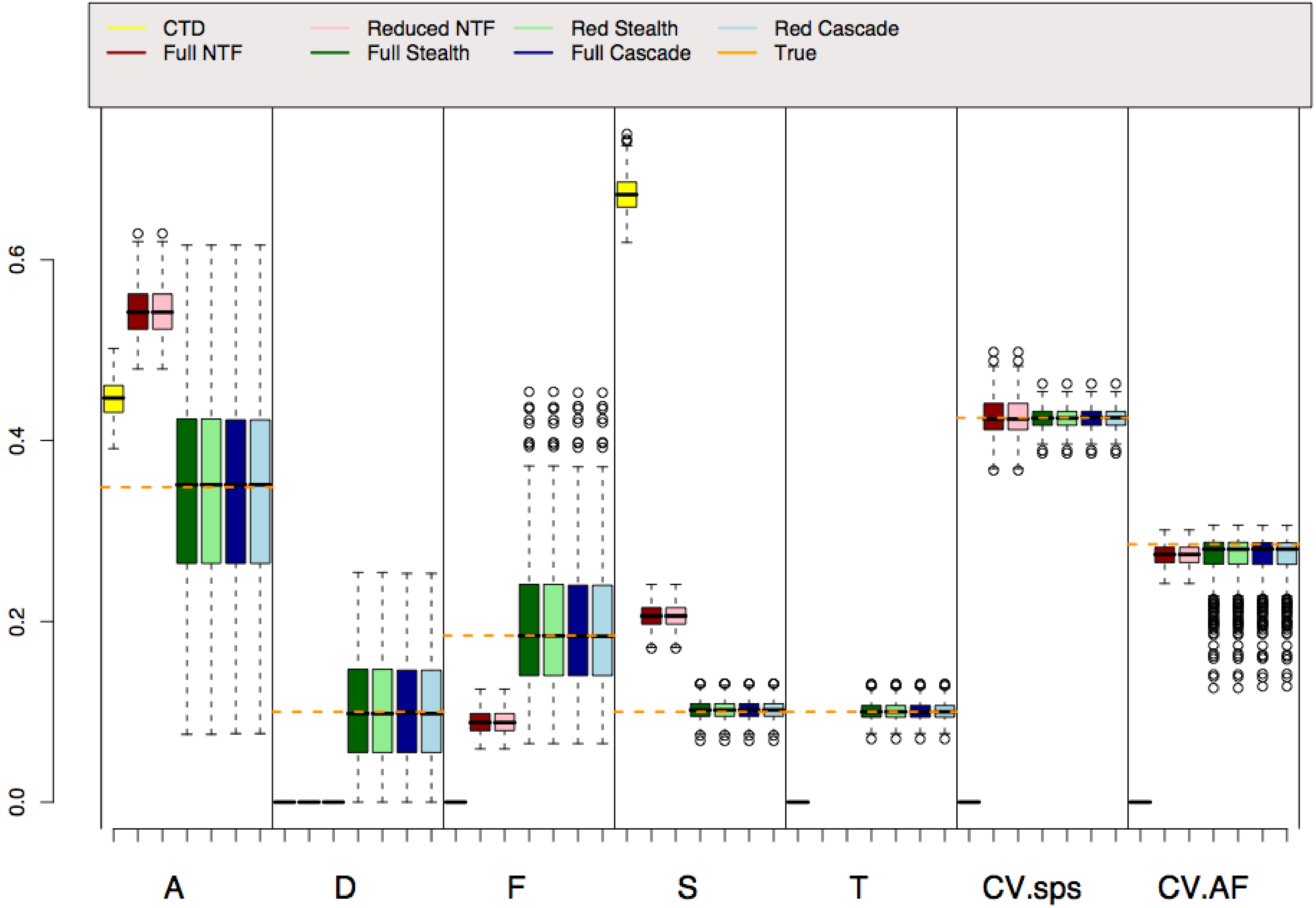
Reality: A=.4, D=.15, S=.15



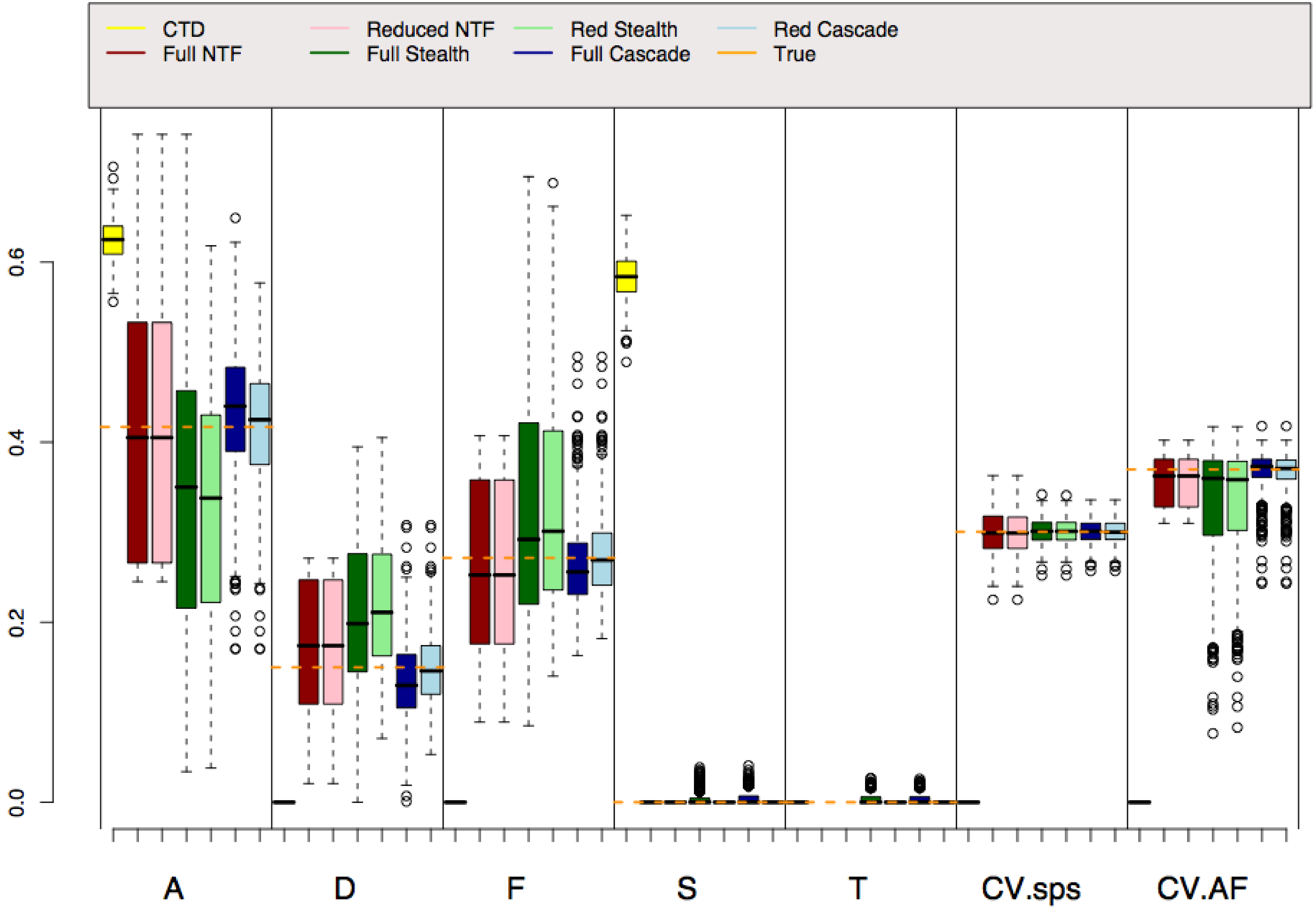
A,D, & F estimates are highly correlated in Stealth & Cascade



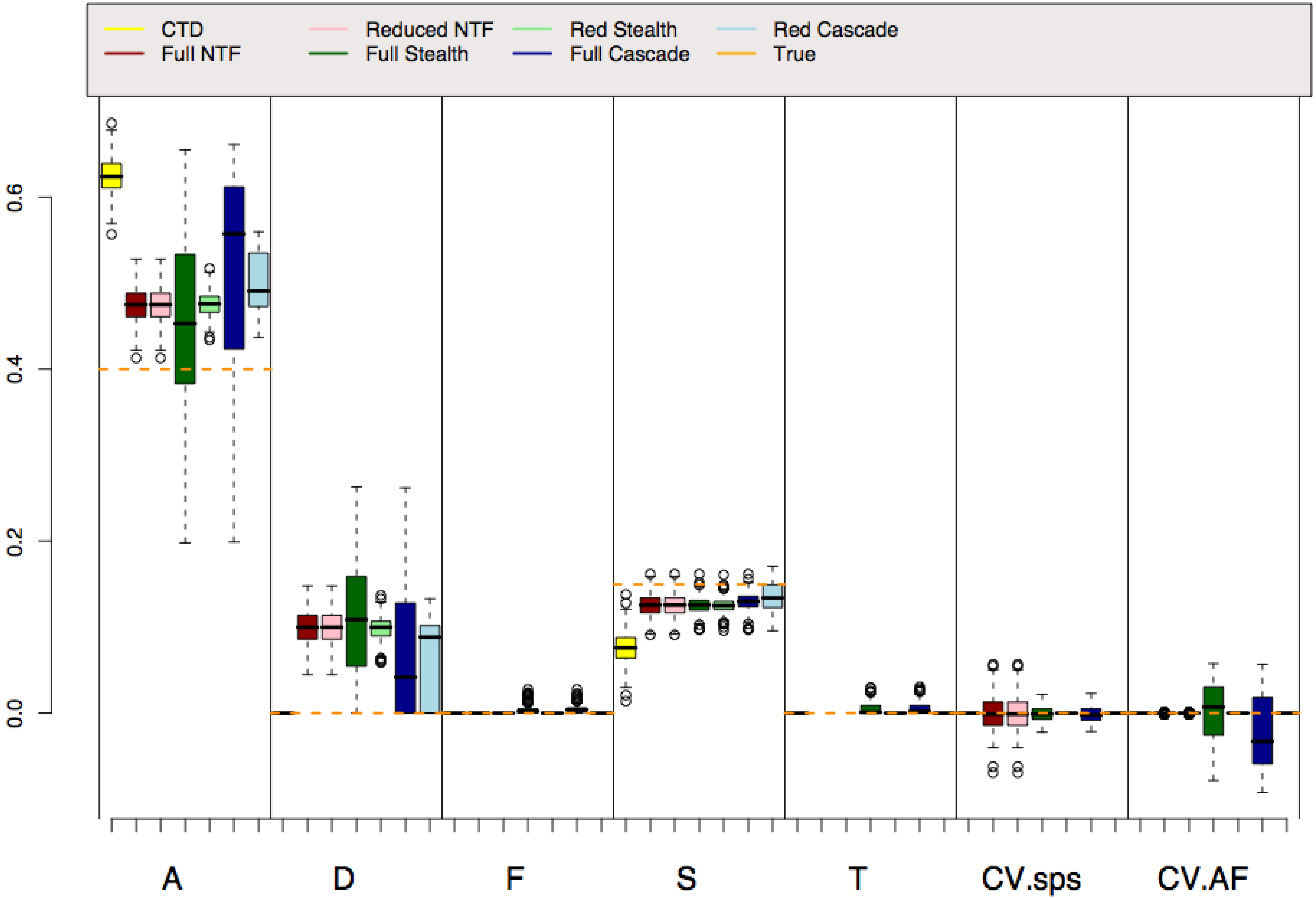
Reality: A=.35, D=.15, F=.2, S=.15, T=.15, AM=.3



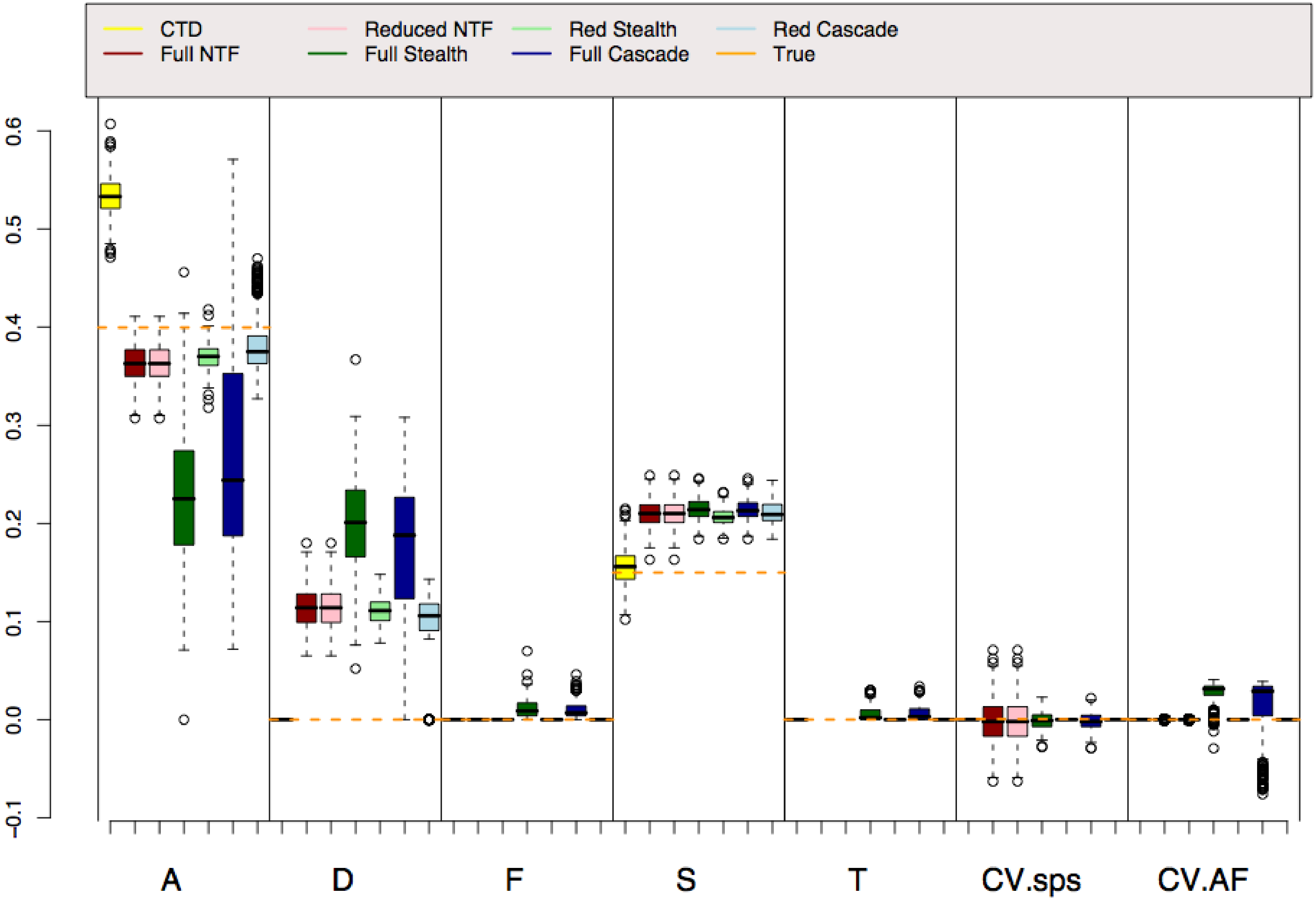
Reality: A=.45, D=.15, F=.25, AM=.3 (Soc Hom)



Reality: A=.4, A*A=.15, S=.15



Reality: A=.4, A*Age=.15, S=.15



Conclusions

- All models require assumptions. Generally, more assumptions = more biased estimates
- Simulations provide independent assessments of the NTFD, *Stealth*, and *Cascade* models
  - These complicated models work as designed!
  - In all models, but especially the CTD, be cautious of reifying parameter estimates (e.g., A, C, & D)!

Quiz Question 7

7) ANALOGY QUESTION

Dr. Frankenstein **is to** Dr. Eaves

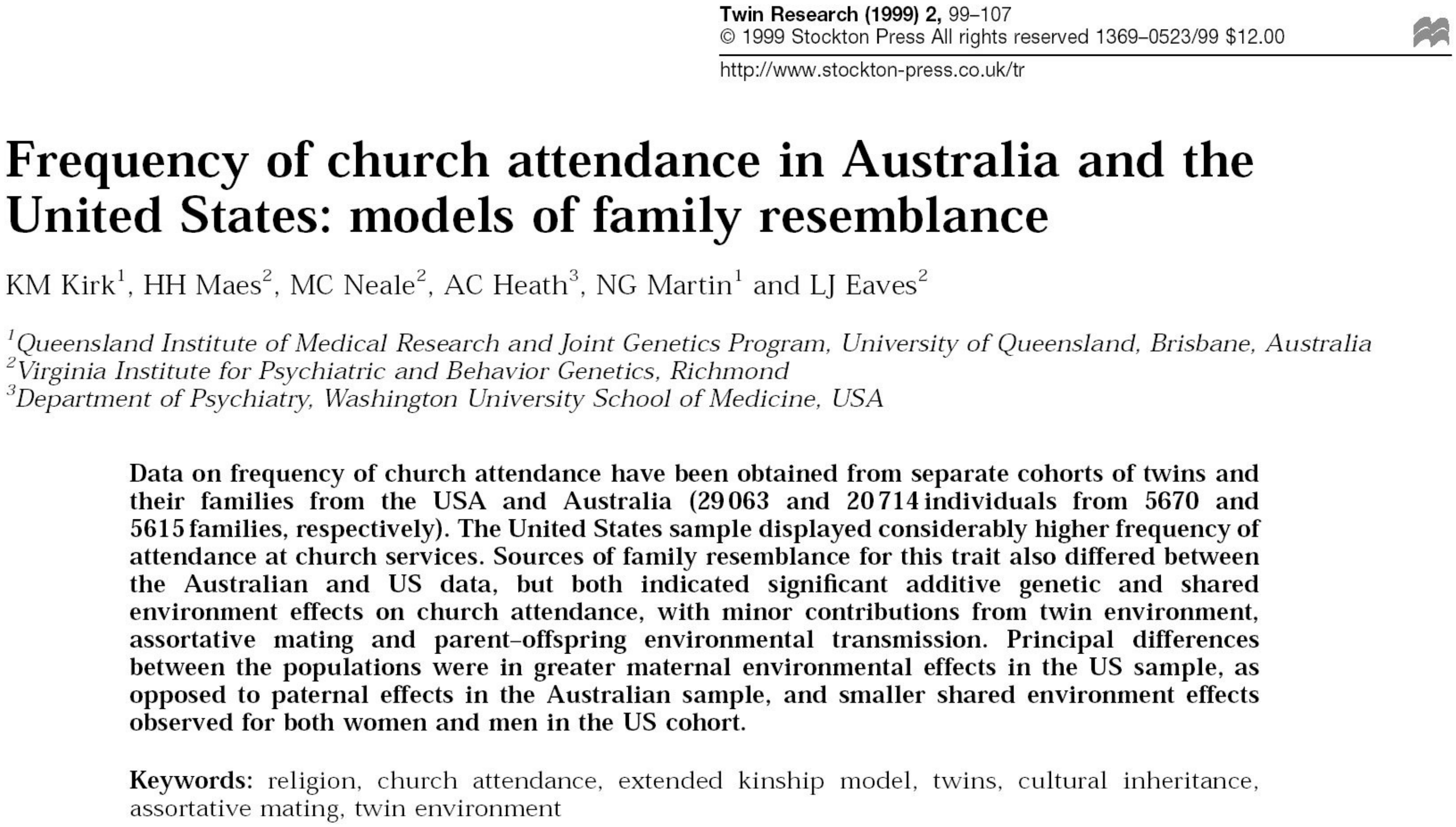
- AS -

the Monster **is to** __________

a) the obsession with twin-only (ACE & ADE) models

http://www.youtube.com/watch#v=W8GRQHsAVjI&feature=related

Stealth application



Twin Research (1999) 2, 99–107
© 1999 Stockton Press All rights reserved 1369–0523/99 $12.00
http://www.stockton-press.co.uk/tr

Frequency of church attendance in Australia and the United States: models of family resemblance

KM Kirk[1], HH Maes[2], MC Neale[2], AC Heath[3], NG Martin[1] and LJ Eaves[2]

[1] *Queensland Institute of Medical Research and Joint Genetics Program, University of Queensland, Brisbane, Australia*
[2] *Virginia Institute for Psychiatric and Behavior Genetics, Richmond*
[3] *Department of Psychiatry, Washington University School of Medicine, USA*

Data on frequency of church attendance have been obtained from separate cohorts of twins and their families from the USA and Australia (29 063 and 20 714 individuals from 5670 and 5615 families, respectively). The United States sample displayed considerably higher frequency of attendance at church services. Sources of family resemblance for this trait also differed between the Australian and US data, but both indicated significant additive genetic and shared environment effects on church attendance, with minor contributions from twin environment, assortative mating and parent–offspring environmental transmission. Principal differences between the populations were in greater maternal environmental effects in the US sample, as opposed to paternal effects in the Australian sample, and smaller shared environment effects observed for both women and men in the US cohort.

Keywords: religion, church attendance, extended kinship model, twins, cultural inheritance, assortative mating, twin environment

Further reading on this lecture

- Eaves LJ, Last KA, Young PA, Martin NG (1978) Model-fitting approaches to the analysis of human behaviour. *Heredity* 41:249-320
- Fulker DW (1982) Extensions of the classical twin method. Human Genetics. Part A: The Unfolding Genome (Progress in Clinical and Biological Research Vol 103A). p. 395-406
- Fulker DW (1988) Genetic and cultural transmission in human behavior. Proceedings of the Second International conference on Quantitative Genetics
- Eaves LJ, Heath AC, Martin NG, Neale MC, Meyer JM, Silberg JL, Corey LA, Truett K, Walter E (1999) Comparing the biological and cultural inheritance of stature and conservatism in the kinships of monozygotic and dizygotic twins. In: Cloninger CR (Ed) Proceedings of 1994 APPA Conference. p. 269-308
- Keller MC & Coventry WL (2005). Quantifying and addressing parameter indeterminacy in the classical twin design. *Twin Research and Human Genetics,* 8, 201-213
- Keller MC, Medland SE, Duncan LE, Hatemi PK, Neale MC, Maes HHM, Eaves LJ. Modeling extended twin family data I: Description of the Cascade Model. *Twin Research and Human Genetics*, 29, 8-18.
- Keller MC, Medland SE, & Duncan LE (2010). Are extended twin family designs worth the trouble? A comparison of the bias, precision, and accuracy of parameters estimated in four twin family models. *Behavior Genetics*.