



MAX-PLANCK-INSTITUT
FÜR DEMOGRAFISCHE
FORSCHUNG

MAX PLANCK INSTITUTE
FOR DEMOGRAPHIC
RESEARCH

Discrete-time multistate regression models in Stata: The `dtms` module

Daniel C. Schneider

German Stata Conference, Berlin, June 16, 2023



Outline

- Multistate methodology
- The **dtms** package
 - Example
 - Package highlights / contributions
 - Results organization and storage: Tree management
 - Analytical contributions
 - Computational aspects
 - Package shortcomings / To be implemented



Discrete-Time Multistate Estimation

- Discrete set of states, e.g., employment status, health status
- Discrete-time
 - Evenly spaced (intermittently observed) data
- Under Markov assumption, use
 - **(1)** regression models to
 - **(2)** predict transition probabilities, then
 - **(3)** apply matrix formulas to get stats like:
 - state and life expectancies (LEXP)
 - lifetime risk (LRSK), mean age at first entry (MAFN)
 - other newly introduced stats, like state at absorption (STAB) and mean duration of episodes (MDUR)



Continuous v. Discrete

- Data dependent choice
- Choice also depends on research question
- Computational advantage depends on application
- Ease of application
- Software exists mostly for continuous time



Package Installation

- `net install dtms, from(https://user.demogr.mpg.de/schneider/stata)`
- At some point, installation will be via SSC.
- Requires Stata version 16.1 or higher.
- Announcement / thread on Stata Forum:
<https://www.statalist.org/forums/forum/general-stata-discussion/general/1690703-dtms-new-stata-command-for-discrete-time-multistate-model-estimation>



Example

- All example statements to follow are listed in the addendum slides.
- dtms estimation proceeds in sequential steps:

(2) regression (`mlogit`)

(3) predict transition probabilities

(4) calculate various results from them

Very first step: (1) Model setup, so there are a total of four steps



Example

```
. dtms exampledata 12
(dtms example data set loaded)
(fictitious single-year data (II) on cognitive impairment (age centered at 50))
. describe
```

Contains data

```
Observations:      28,654      fictitious single-year data (II) on cognitive impairment
                               (age centered at 50)
Variables:          8         10 Jun 2023 11:01
                               (_dta has notes)
```

Variable name	Storage type	Display format	Value label	Variable label
id	int	%12.0g		Subject ID
n	float	%9.0g		Subject obs number
year	int	%9.0g		Survey year
age	float	%9.0g		Exact age at interview (years, centered at 50)
cog3	byte	%9.0g	COG3	Cognitive impairment, 3-cat
sex	byte	%8.0g	SEX	Sex
educ	byte	%9.0g	EDUC	Education level, 3-cat
numdrinks	byte	%9.0g		# of alc drinks / week

Sorted by: **id year**



Example

```
. table cog3
```

	Frequency
Cognitive impairment, 3-cat	
1.none	21,534
2.mild	4,729
3.severe	1,181
4.dead	1,210
Total	28,654

```
. xttrans cog3
```

Cognitive impairment , 3-cat	Cognitive impairment, 3-cat				Total
	1	2	3	4	
1	90.66	6.84	0.04	2.46	100.00
2	20.82	60.30	9.22	9.66	100.00
3	0.61	20.52	56.77	22.10	100.00
Total	74.77	16.63	4.06	4.54	100.00



Example

```
. dtms clear
. dtms setup (impex) , states(1 none 2 mild 3 sevr 4 dead) abs(1) ages(50/110)
. dtms dir
(S) impex : (no label) | tra IDs: 1 2 3 | abs IDs: 4 | 61 ages: 50-110
.
. mlogit cog3 iL.cog3 c.age c.age#c.age sex##educ c.numdrinks , cluster(id) rrr
  (output omitted)
. dtms estimate frome (impex mlog) , agevar(age, center(50)) orgnvar(L.cog3)
. dtms dir
(S) impex : (no label) | tra IDs: 1 2 3 | abs IDs: 4 | 61 ages: 50-110
(E) mlog : (no label) | cmdline: mlogit cog3 iL.cog3 c.age c.age#c.age se..

. dtms transprob atmeans (impex mlog allmeans)
. dtms dir
(S) impex : (no label) | tra IDs: 1 2 3 | abs IDs: 4 | 61 ages: 50-110
(E) mlog : (no label) | cmdline: mlogit cog3 iL.cog3 c.age c.age#c.age se..
(T) allmeans : (no label) | dtms trans atmeans: L.cog3=(1 2 3) age=(1 2..
```



Example

- . dtms proportion fixed (impex pfixed) , values(0.88 0.10 0.02)
- . dtms result lexp (impex mlog allmeans lexp) , initprop(pfixed) format(%7.1f)

Life expectancy

Remaining life expectancy

		init			
		none	mild	sevr	total
state	none	24.5	21.3	14.3	24.0
	mild	3.8	4.9	3.7	3.9
	sevr	0.8	0.9	2.1	0.8
	total	29.1	27.1	20.2	28.7

Simplification: a single fixed initial proportion is assumed for all groups in the example

- . dtms dir
- (S) impex : (no label) | tra IDs: 1 2 3 | abs IDs: 4 | 61 ages: 50-110
- (P) pfixed : (no label) | fixed: 0.880 0.100 0.020
- (E) mlog : (no label) | cmdline: mlogit cog3 iL.cog3 c.age c.age#c.age se..
- (T) allmeans : (no label) | dtms trans atmeans: L.cog3=(1 2 3) age=(1 2..
- (R) lexp : (no label) | prop: pfixed | timing: mid | calc: analytic |..



Example

- . dtms transprob atmeans (impex mlog edlow) , at(educ=1)
- . dtms transprob atmeans (impex mlog edhigh) , at(educ=3)
- . dtms result lexp (impex mlog edlow lexp) , initprop(pfixed) format(%7.1f)
- . dtms result lexp (impex mlog edhigh lexp) , initprop(pfixed) format(%7.1f)

Remaining life expectancy

	init			
	none	mild	sevr	total
state				
none	18.5	14.7	9.8	17.9
mild	5.2	6.4	4.8	5.3
sevr	1.2	1.4	2.6	1.2
total	24.9	22.5	17.2	24.5

Remaining life expectancy

	init			
	none	mild	sevr	total
state				
none	28.2	25.3	15.7	27.7
mild	2.3	3.2	2.3	2.4
sevr	0.3	0.4	1.6	0.4
total	30.8	28.9	19.5	30.4



Example

- . dtms result stab (impex mlog edlow stab) , initpr(pfixed) form(%7.2f)
- . dtms result stab (impex mlog edhigh stab) , initpr(pfixed) form(%7.2f) post nopv

State distribution before absorption

		init			
		none	mild	sevr	total
state	none	0.38	0.32	0.23	0.37
	mild	0.40	0.44	0.32	0.40
	sevr	0.22	0.23	0.46	0.23
	total	1.00	1.00	1.00	1.00

State distribution before absorption

		init			
		none	mild	sevr	total
state	none	0.60	0.55	0.36	0.59
	mild	0.29	0.33	0.21	0.30
	sevr	0.11	0.12	0.43	0.11
	total	1.00	1.00	1.00	1.00

State distribution before absorption:

		Coefficient	Std. err.	[95% conf. interval]	
none	none	0.598	0.027	0.545	0.650
	mild	0.295	0.023	0.250	0.340
	sevr	0.107	0.019	0.070	0.144
mild	none	0.552	0.026	0.500	0.604
	mild	0.332	0.023	0.286	0.378
	sevr	0.116	0.020	0.077	0.156
sevr	none	0.358	0.033	0.293	0.422
	mild	0.212	0.024	0.165	0.258
	sevr	0.431	0.048	0.336	0.525
cond	none	1.000	0.000	1.000	1.000
	mild	1.000	0.000	1.000	1.000
	sevr	1.000	0.000	1.000	1.000
state	none	0.588	0.026	0.537	0.640
	mild	0.297	0.023	0.253	0.341
	sevr	0.115	0.019	0.078	0.151
total	total	1.000	0.000	1.000	1.000



Example

```
. dtms dir
(S) impex : (no label) | tra IDs: 1 2 3 | abs IDs: 4 | 61 ages: 50-110
(P) pfixed : (no label) | fixed: 0.880 0.100 0.020
(E) mlog : (no label) | cmdline: mlogit cog3 iL.cog3 c.age c.age#c.age se..
(T) allmeans : (no label) | dtms trans atmeans: L.cog3=(1 2 3) age=(1 2..
(R) lexp : (no label) | prop: pfixed | timing: mid | calc: analytic |..
(T) edlow : (no label) | dtms trans atmeans: L.cog3=(1 2 3) age=(1 2 3 ..
(R) lexp : (no label) | prop: pfixed | timing: mid | calc: analytic |..
(R) stab : (no label) | prop: pfixed | timing: | calc: analytic | ST..
(T) edhigh : (no label) | dtms trans atmeans: L.cog3=(1 2 3) age=(1 2 3..
(R) lexp : (no label) | prop: pfixed | timing: mid | calc: analytic |..
(R) stab : (no label) | prop: pfixed | timing: | calc: analytic | ST..

. dtms file save using temptree.dtms , replace
```



Example

```
. dtms combine (impex mlog edhigh lexp) , with(edlow lexp) post(state total)
```

2 combined dtms results

result type(s) : **lexp**

model name : **impex mlog**

result names : **1: edhigh lexp**

2: edlow lexp

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
1_lexp_state						
none	27.654	0.755	36.644	0.000	26.175	29.133
mild	2.388	0.271	8.810	0.000	1.857	2.919
sevr	0.370	0.100	3.693	0.000	0.174	0.566
1_lexp_total						
total	30.412	0.737	41.285	0.000	28.968	31.855
2_lexp_state						
none	17.940	0.643	27.901	0.000	16.680	19.201
mild	5.292	0.367	14.433	0.000	4.574	6.011
sevr	1.243	0.154	8.048	0.000	0.940	1.545
2_lexp_total						
total	24.476	0.676	36.193	0.000	23.150	25.801



Example

```
. dtms combine (impex mlog edhigh lexp) , with(edlow lexp) post(state total) diff
```

2 combined dtms results

```
result type(s) : lexp difference  
model name     : impex mlog  
result names   : 1: edhigh lexp  
                2: edlow lexp
```

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
lexp_state						
none	9.714	0.923	10.529	0.000	7.906	11.522
mild	-2.905	0.436	-6.662	0.000	-3.759	-2.050
sevr	-0.873	0.182	-4.805	0.000	-1.229	-0.517
lexp_total						
total	5.936	0.881	6.738	0.000	4.209	7.663



Package Highlights / Contributions

- A **framework for organizing** and saving results
- Up to 9 states
- Automated estimation of transition probabilities
- **14+ different outcome statistics**
- **Asymptotic covariance matrices**
- **Partial age ranges**
- **Group comparisons**
linear and nonlinear hypothesis testing on any number of and on any type of results
- **Speed** (interactive time range)
- **Markov chains with rewards**
general implementation; includes the possibility of user-defined rewards
- Works with `svy` estimation
- Generation of data sets with simulated trajectories
- Simulation-based results as alternative to analytical ones

All help files and/or subcommands of the **dtms** package:

package and conceptual overview ([help dtms](#))

managing the dtms tree or its elements ([help dtms tree](#))

dtms dir	list a dtms tree
dtms describe	describe a dtms tree element
dtms label	label a dtms tree element
dtms rename	rename a dtms tree element
dtms drop	drop a dtms tree element
dtms usedby	list dtms tree elements that use a particular side tree element
dtms file	save and load setups and all of their downstream elements
dtms settings	query and modify global dtms settings
dtms clear	delete the entire dtms tree and all global dtms settings

adding elements to the dtms tree ([help dtms add](#))

dtms setup	add basic model setup
dtms proportion	add initial proportion
dtms rewards	add non-standard transition timing specification
dtms estimate	add model regression estimate
dtms transprob	add transition probabilities
dtms simdata	add simulated trajectories

calculate results ([help dtms result](#))

dtms result	calculate and add one of the 14+ different outcomes to the dtms tree
-----------------------------	--

extract information from the dtms tree ([help dtms extract](#))

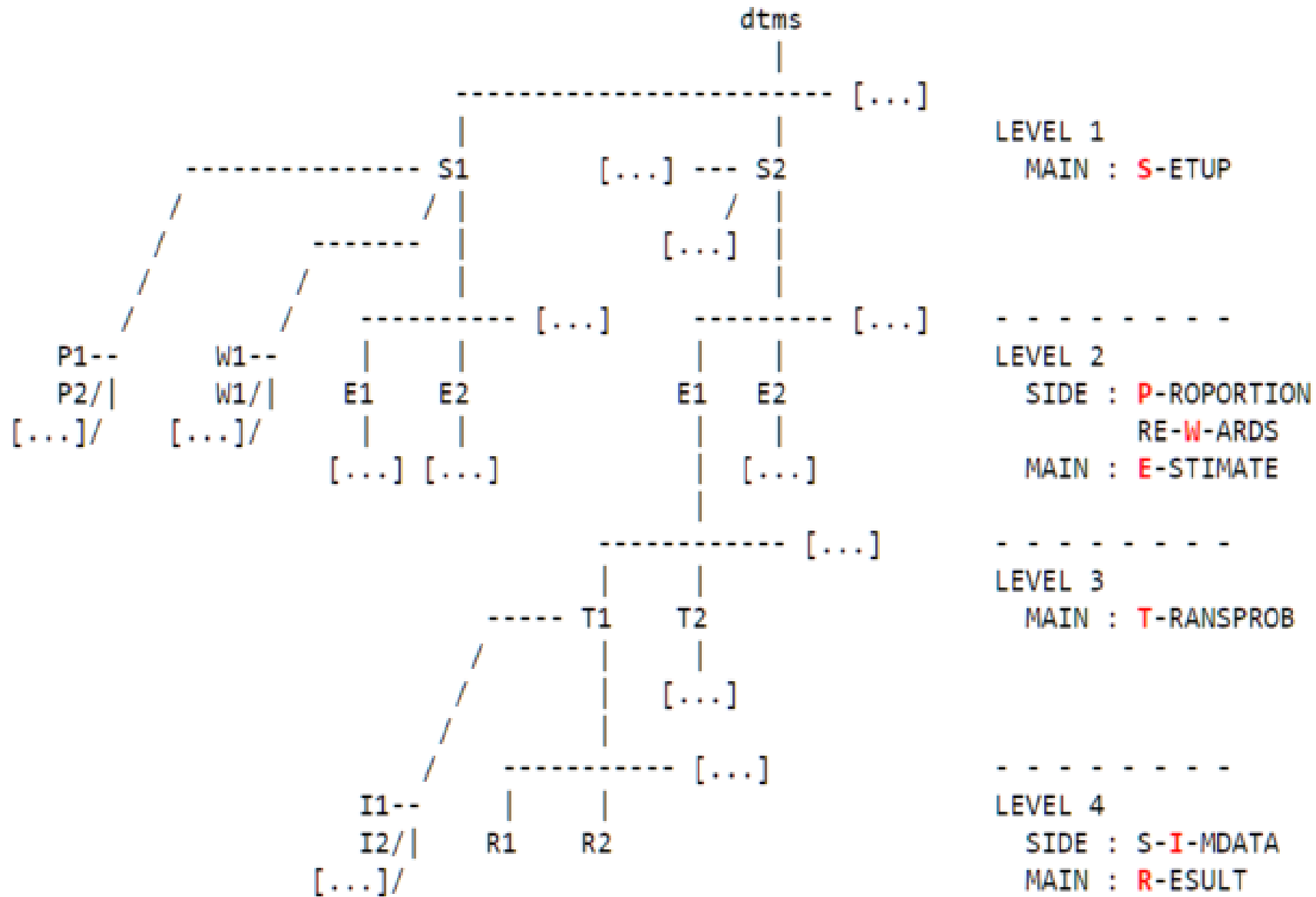
dtms erestore	restore e()-results that are held in some dtms tree elements
dtms combine	post combined estimates of two result tree elements to e()
dtms getmatrix	query various kinds of matrices related to dtms calculations
dtms matbrowse	browse Stata matrix in Stata's data browser

extensive examples and helper commands for executing them ([help dtms examples](#))

dtms exempladata	load example data sets
dtms exampletree	load example tree elements into the dtms tree



Tree Management





Analytical Contributions

- The material that follows will soon be available in two working papers: Schneider (2023) and Schneider and Myrskylä (2023).
- Large sections of the mathematical derivations are already in the methods doc included in the `dtms` package.



Analytical Contributions: Preliminaries

Markov transition matrix:

$$P = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ U_2 & \mathbf{0} & & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & U_3 & & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \vdots & & \ddots & & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & U_{\bar{a}-1} & \mathbf{0} & \mathbf{0} \\ d_2 & d_3 & \cdots & d_{\bar{a}-1} & d_{\bar{a}} & 1 \end{bmatrix}$$

Submatrices have trans. prob. at age a :

$$U_a = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1\bar{s}} \\ u_{21} & u_{22} & & u_{2\bar{s}} \\ \vdots & & \ddots & \vdots \\ u_{\bar{s}1} & u_{\bar{s}2} & \cdots & u_{\bar{s}\bar{s}} \end{bmatrix}$$

Sometimes only the transient states are needed:

$$U = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ U_2 & \mathbf{0} & & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & U_3 & & \mathbf{0} & \mathbf{0} \\ \vdots & & \ddots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & U_{\bar{a}-1} & \mathbf{0} \end{bmatrix}$$

The **fundamental matrix** has elements for LEXP: $F = (I - U)^{-1}$



Analytical Contributions: Asymptotics, LEXP

The first column of the fundamental matrix contains all the elements of state expectancies:

$$F = (I - U)^{-1} = \sum_{a=0}^{\bar{a}-2} U^a = \left[\begin{array}{c|ccc} I & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ U_2 & I & & \mathbf{0} & \mathbf{0} \\ U_3 \cdot U_2 & U_3 & & \mathbf{0} & \mathbf{0} \\ \vdots & & \ddots & & \vdots \\ U_{\bar{a}-1} \cdot \dots \cdot U_2 & U_{\bar{a}-1} \cdot \dots \cdot U_3 & \dots & U_{\bar{a}-1} & I \end{array} \right]$$

$$F_1 = \left[U_2 \quad U_3 U_2 \quad \dots \quad \prod_{a=\bar{a}-1}^2 U_a \right] = [f_2 \quad f_3 \quad \dots \quad f_{\bar{a}-1}]$$

Let $\tilde{U} = [U_2 \quad \dots \quad U_{\bar{a}-1}]$, then $V^{tr} = \text{cov}[\text{vec}[\tilde{U}]]$.

Use the delta method to compute the covariance matrix:

$$V^{F_1} = G^{F_1} V^{tr} G^{F_1}' \quad G^{F_1} = \frac{\partial \text{vec } V^{F_1}}{\partial \text{vec } \tilde{U}'}$$



Analytical Contributions: Asymptotics, LEXP

To summarize: Life expectancy covariance matrix:

$$\mathbf{V}^{tr} = \mathbf{G}^{tr} \mathbf{V}^{ml} \mathbf{G}^{tr'}$$

$$\mathbf{V}^{F_1} = \mathbf{G}^{F_1} \mathbf{V}^{tr} \mathbf{G}^{F_1'}$$

$$\mathbf{V}^E = (\mathbf{I}_{\bar{s}^2} \otimes \mathbf{r}_{+1}) \mathbf{V}^{F_1} (\mathbf{I}_{\bar{s}^2} \otimes \mathbf{r}_{+1})'$$

Calculations repeatedly use the **delta method**.



Analytical Contributions: Asymptotics, LEXP

$$V^{F_1} = G^{F_1} V^{tr} G^{F_1'}$$

$$\frac{\partial \text{vec}(f_i)}{\partial \text{vec}(U_j)'} = \frac{\partial \text{vec}(\prod_{a=i}^{j+1} U_a \times U_j \times \prod_{a=j-1}^2 U_a)}{\partial \text{vec}(U_j)'}$$

$$G^{F_1} = \begin{array}{c|ccc|c|c} & U_2 & U_3 & U_4 & \dots & U_{a-1} \\ \hline U_2 & I \otimes I & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ U_3 U_2 & I \otimes U_3 & U_2' \otimes I & \mathbf{0} & & \mathbf{0} \\ U_4 U_3 U_2 & I \otimes U_4 U_3 & U_2' \otimes U_4 & (U_3 U_2)' \otimes I & & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ U_{a-1} \dots U_2 & I \otimes U_{a-1} \dots U_3 & U_2' \otimes U_{a-1} \dots U_4 & (U_3 U_2)' \otimes U_{a-1} \dots U_5 & \dots & (U_{a-2} \dots U_2) \otimes I \end{array}$$

$$V^E = (I_{s^2} \otimes r_{+1}) V^{F_1} (I_{s^2} \otimes r_{+1})'$$



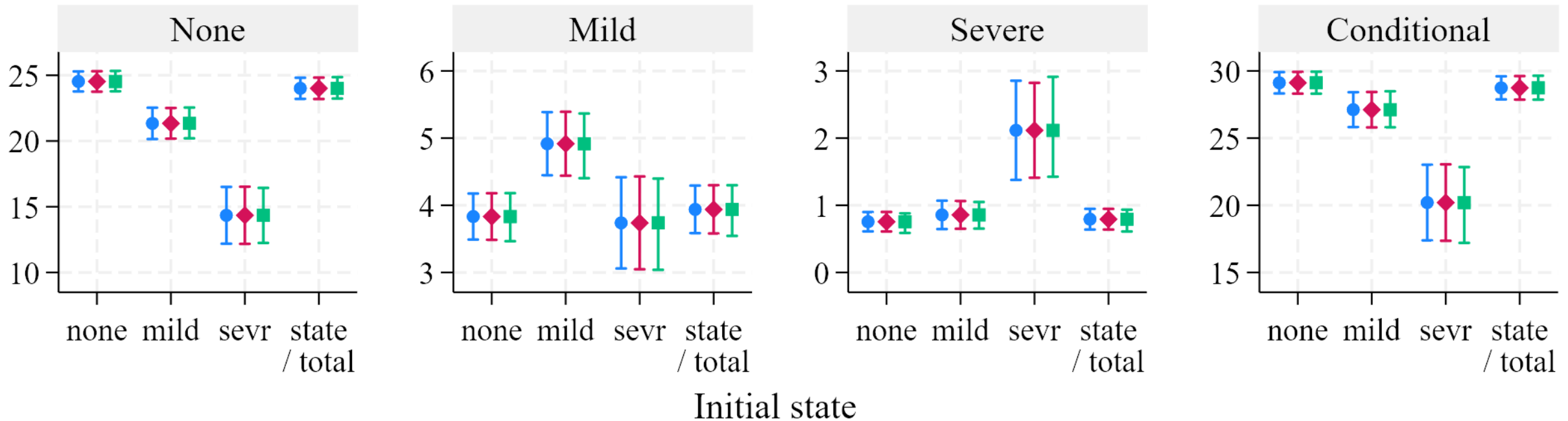
LEXP: Comparison Analytic v. Bootstrap CIs

. dtms erestore (impex mlog allmeans lexp) , replay

CI: Remaining life expectancy, for each state and total

All subjects

● SE-based, asymptotic ◆ SE-based, bootstrap ■ Percentile, bootstrap





Analytical Contributions: Asymptotics, LRSK/MAFN

Output for lifetime risk / mean age at first entry:

```
. dtms result mafn (impex mlog allmeans mafn) , initst(1) intermed(2) target(3) post(lrsk mafn)
```

Mean age (at) first entry / lifetime risk:

		Coefficient	Std. err.	z	P> z	[95% conf. interval]	
lrsk	lrsk	0.281	0.016	17.372	0.000	0.249	0.312
mafn	mafn	76.385	0.671	113.776	0.000	75.069	77.701

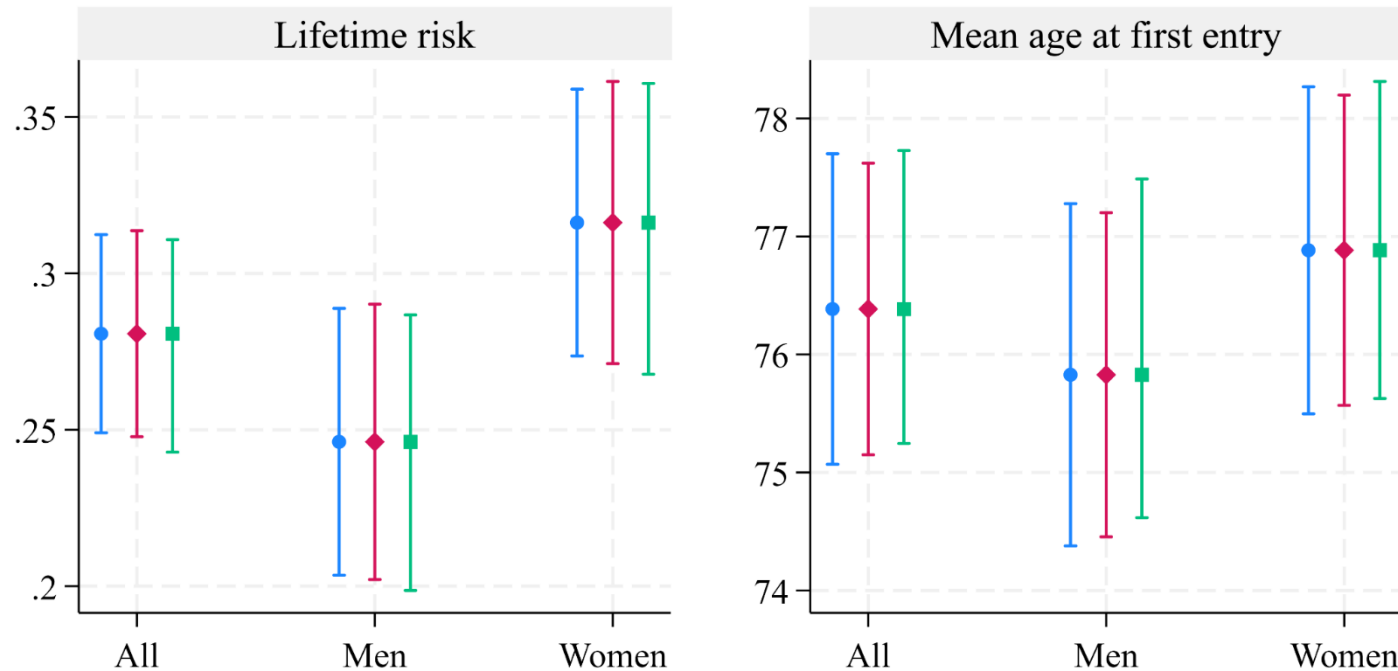


MAFN: Comparison Analytic v. Bootstrap CIs

- . dtms trans atmeans (impex mlog men) , at(sex=1)
- . dtms trans atmeans (impex mlog wmn) , at(sex=2)
- . qui dtms result mafn (impex mlog men mafn) , initst(1) intermed(2) target(3)
- . qui dtms result mafn (impex mlog wmn mafn) , initst(1) intermed(2) target(3)
- . qui dtms combine (impex mlog allmeans mafn) , with(men mafn) post(lrsk mafn)
- . dtms combine e() , with(wmn mafn)

CIs: Lifetime risk / mean age at first entry

• SE-based, asymptotic ♦ SE-based, bootstrap ■ Percentile, bootstrap





Analytical Contributions: Rewards

- The behind Markov chains with rewards: link probabilities of reaching certain states with transition probabilities out of that state, where each of these out-transitions is assigned a reward.
- LEXP is a special case.
- 12+ results in the package are based on the rewards method.
- Explaining rewards in more detail would need its own presentation.

```
epis - number of episodes
stab - state at absorption

mdur - mean duration of episodes
maan - mean age, all entries
maax - mean age, all exits
maax# - mean age, all entries, detail for
state encoded #
maan# - mean age, all exits, detail for state
encoded #
maab - mean age at absorption
mais - mean age in state

ncnt - entry count
xcnt - exit count
ncnt# - entry count, detail for state #:
breakdown w.r.t. exits to other
states
xcnt# - exit count, detail for state #:
breakdown w.r.t. entries from other
states
```



Analytical Contributions: Asymptotics, Rewards

Derivations similar to LEXP, but more complicated.

$$\mathbf{G}^{FR^S} = \begin{array}{c|cccccc}
 & \mathbf{U}_2 & \mathbf{U}_3 & \mathbf{U}_4 & \dots & \mathbf{U}_{\bar{a}-1} \\
 \hline
 \overline{\mathbf{PR}}_2^S & [I \otimes 1] \frac{\partial \text{vec}(\overline{\mathbf{PR}}_2^S)}{\partial \text{vec}(\mathbf{U}_2)'} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\
 \overline{\mathbf{PR}}_3^S \mathbf{U}_2 & I \otimes \overline{\mathbf{PR}}_3 & [\mathbf{U}_2' \otimes 1] \frac{\partial \text{vec}(\overline{\mathbf{PR}}_3^S)}{\partial \text{vec}(\mathbf{U}_3)'} & \mathbf{0} & & \mathbf{0} \\
 \overline{\mathbf{PR}}_4^S \mathbf{U}_3 \mathbf{U}_2 & I \otimes \overline{\mathbf{PR}}_4 \mathbf{U}_3 & \mathbf{U}_2' \otimes \overline{\mathbf{PR}}_4 & [(\mathbf{U}_3 \mathbf{U}_2)' \otimes 1] \frac{\partial \text{vec}(\overline{\mathbf{PR}}_4^S)}{\partial \text{vec}(\mathbf{U}_4)'} & & \mathbf{0} \\
 \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
 \overline{\mathbf{PR}}_{\bar{a}-1}^S \mathbf{U}_{\bar{a}-2} \dots \mathbf{U}_2 & I \otimes \overline{\mathbf{PR}}_{\bar{a}-1} \dots \mathbf{U}_3 & \mathbf{U}_2' \otimes \overline{\mathbf{PR}}_{\bar{a}-1} \dots \mathbf{U}_4 & (\mathbf{U}_3 \mathbf{U}_2)' \otimes \overline{\mathbf{PR}}_{\bar{a}-1} \dots \mathbf{U}_5 & & [(\mathbf{U}_{\bar{a}-2} \dots \mathbf{U}_2)' \otimes 1] \frac{\partial \text{vec}(\overline{\mathbf{PR}}_{\bar{a}-1}^S)}{\partial \text{vec}(\mathbf{U}_{\bar{a}-1})'} \\
 \overline{\mathbf{PR}}_{\bar{a}}^S \mathbf{U}_{\bar{a}-1} \dots \mathbf{U}_2 & I \otimes \overline{\mathbf{PR}}_{\bar{a}} \dots \mathbf{U}_3 & \mathbf{U}_2' \otimes \overline{\mathbf{PR}}_{\bar{a}} \dots \mathbf{U}_4 & (\mathbf{U}_3 \mathbf{U}_2)' \otimes \overline{\mathbf{PR}}_{\bar{a}} \dots \mathbf{U}_5 & \dots & (\mathbf{U}_{\bar{a}-2} \dots \mathbf{U}_2)' \otimes \overline{\mathbf{PR}}_{\bar{a}}
 \end{array}$$

with

$$\frac{\partial \text{vec}(\overline{\mathbf{PR}}_a^S)}{\partial \text{vec}(\mathbf{U}_a)'} = \frac{\partial \text{vec}(\mathbf{1}_{\bar{s}} \cdot (\tilde{\mathbf{P}}_a \odot \mathbf{R}_a^S))}{\partial \text{vec}(\mathbf{U}_a)'} = (\mathbf{I}_{\bar{s}} \otimes \mathbf{1}_{\bar{s}+1}) \cdot \mathbf{K}_{\bar{s}\bar{s}+1} \cdot \begin{bmatrix} \text{diag}(\text{vec}(\tilde{\mathbf{R}}_a^{S'})) \\ -(\mathbf{1}_{\bar{s}} \otimes \text{diag}(\tilde{\mathbf{r}}_a^S)) \end{bmatrix} \cdot \mathbf{K}_{\bar{s}\bar{s}}$$



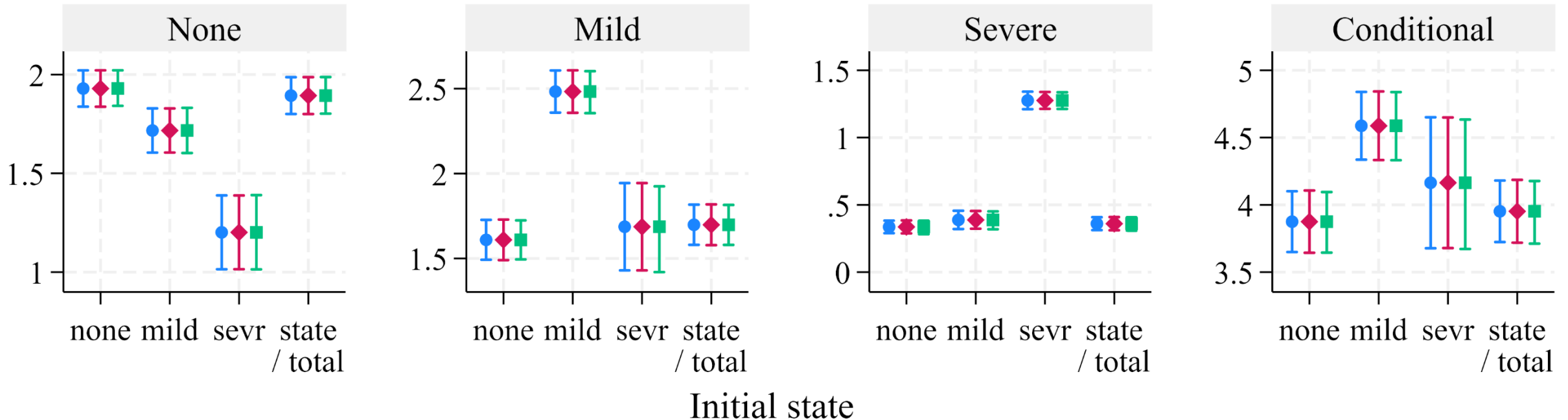
Rewards (EPIS): Comparison Analytic v. Bootstrap CIs

. dtms result epis (impex mlog allmeans epis) , initpr(pfixed) post

CIs: Number of episodes, for each state and total

All subjects

● SE-based, asymptotic ◆ SE-based, bootstrap ■ Percentile, bootstrap





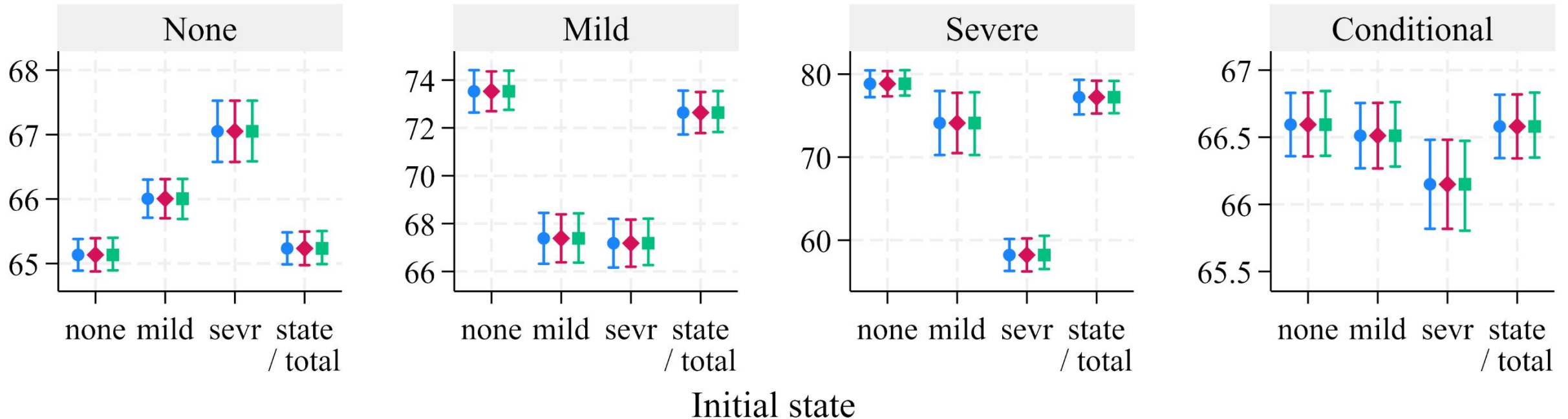
Rewards (MAIS): Comparison Analytic v. Bootstrap CIs

```
. dtms result mais (impex mlog allmeans mais) , initpr(pfixed) post
```

CIs: Mean age in state, for each state and total

All subjects

● SE-based, asymptotic ◆ SE-based, bootstrap ■ Percentile, bootstrap





Analytical Contributions: Group Comparisons

Life expectancy covariance matrix, single group

$$\begin{aligned}
 \mathbf{V}^{tr} &= \mathbf{G}^{tr} \mathbf{V}^{ml} \mathbf{G}^{tr'} \\
 \mathbf{V}^{F_1} &= \mathbf{G}^{F_1} \mathbf{V}^{tr} \mathbf{G}^{F_1'} \\
 \mathbf{V}^E &= (\mathbf{I}_{\bar{s}^2} \otimes \mathbf{r}_{+1}) \mathbf{V}^{F_1} (\mathbf{I}_{\bar{s}^2} \otimes \mathbf{r}_{+1})'
 \end{aligned}$$

Life expectancy covariance matrix, two groups

$$\begin{aligned}
 \mathbf{G}^{tr} &= \begin{bmatrix} \mathbf{G}_1^{tr} \\ \mathbf{G}_2^{tr} \end{bmatrix} \\
 \mathbf{G}^{F_1} &= \begin{bmatrix} \mathbf{G}_1^{F_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_2^{F_1} \end{bmatrix} \\
 \mathbf{V}^E &= \begin{bmatrix} \mathbf{I}_{\bar{s}^2} \otimes \mathbf{r}_{+1,1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\bar{s}^2} \otimes \mathbf{r}_{+1,2} \end{bmatrix} \mathbf{V}^{F_1} \begin{bmatrix} \mathbf{I}_{\bar{s}^2} \otimes \mathbf{r}_{+1,1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\bar{s}^2} \otimes \mathbf{r}_{+1,2} \end{bmatrix}'
 \end{aligned}$$

The above generalizes to N groups in the obvious way.

The methods doc extends this to any type of result, and any mixture of results.

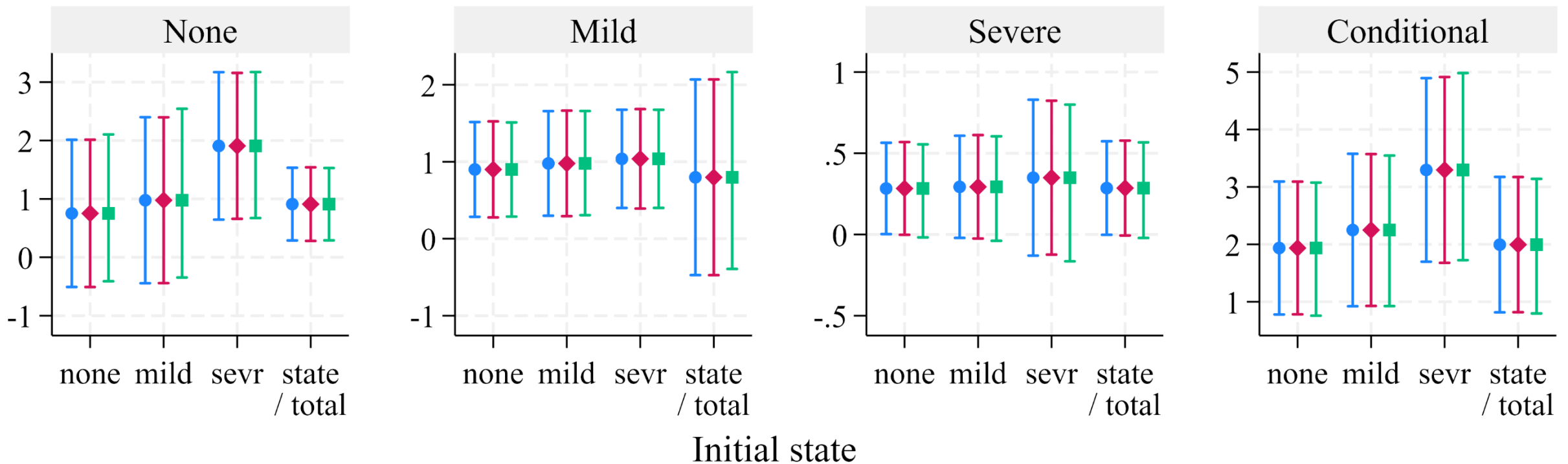


Analytical Contributions: Group Comparisons

- . dtms result lexp (impex mlog wmn lexp) , initpr(pfixed)
- . dtms result lexp (impex mlog men lexp) , initpr(pfixed)
- . dtms combine (impex mlog wmn lexp) , with(men lexp) difference

CIs: Life exp., women v. men (dif), for each state and total

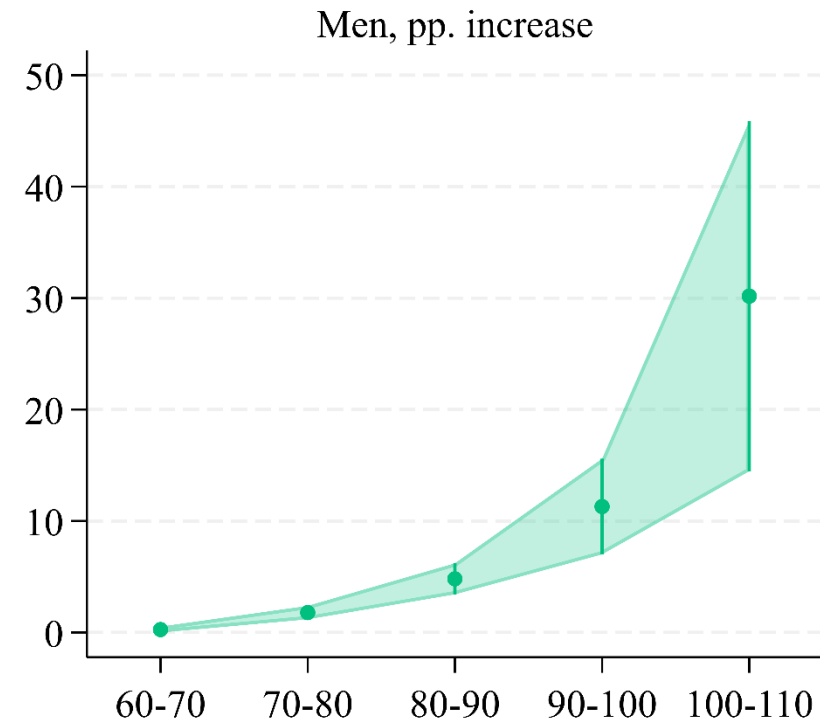
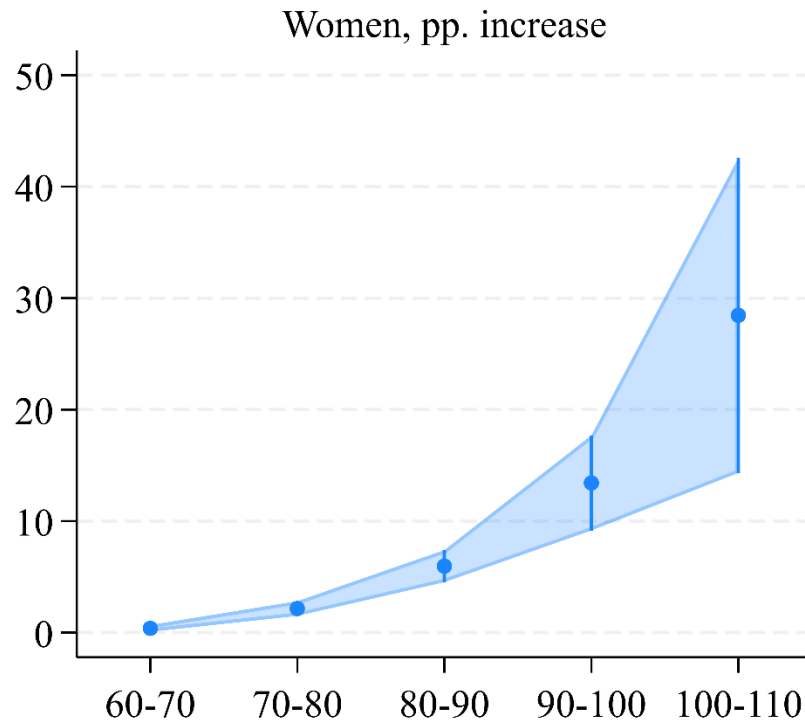
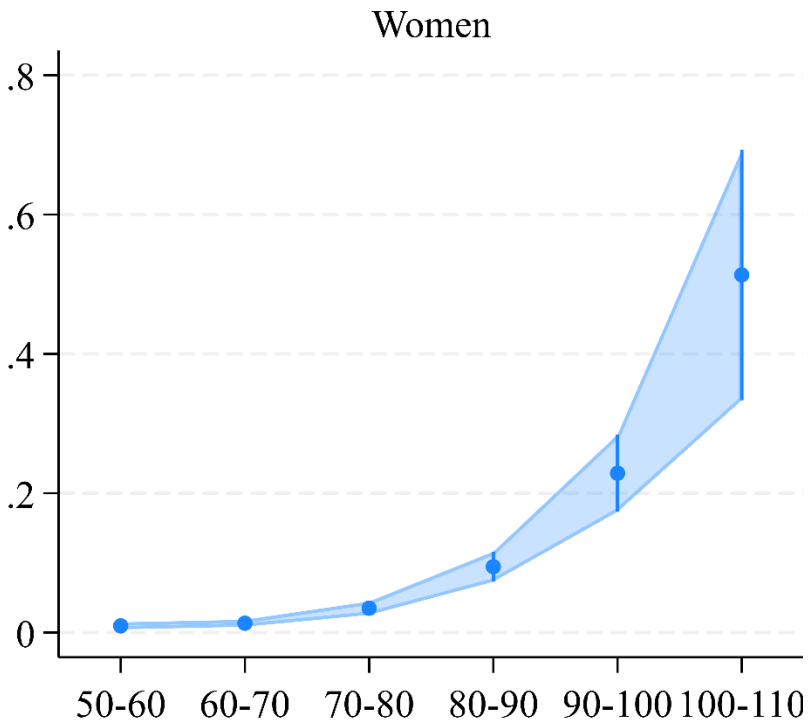
● SE-based, asymptotic ◆ SE-based, bootstrap ■ Percentile, bootstrap





Analytical Contributions: Partial Age Ranges

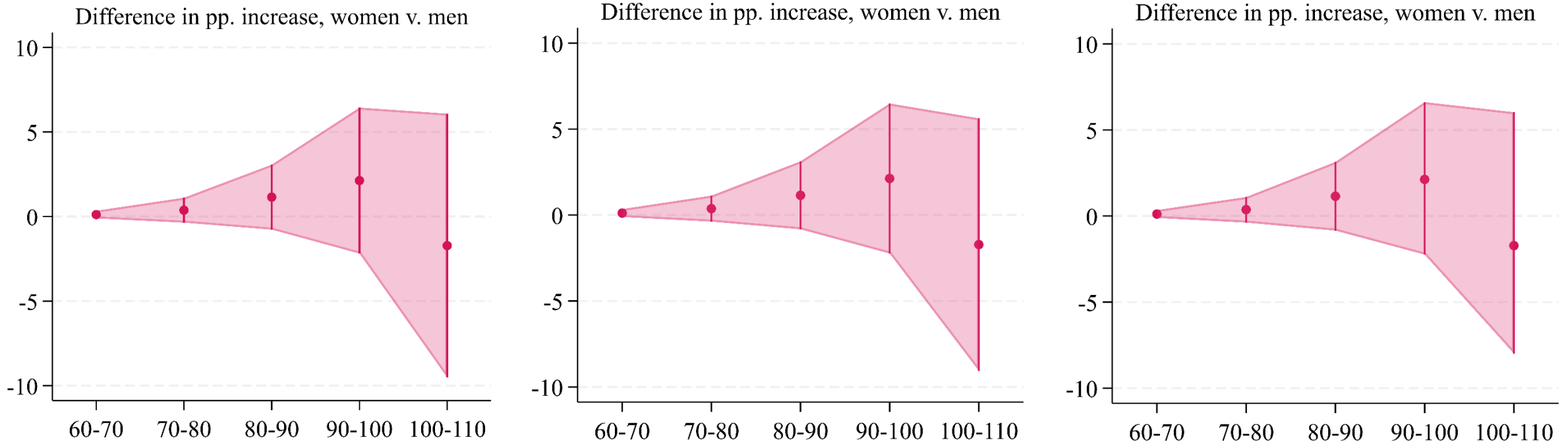
Illustrative example: Fraction of lifetime lived in severe impairment, by decade





Analytical Contributions: Partial Age Ranges

Fraction of lifetime lived in severe impairment, by decade
SE-based asymptotic v. SE-based bootstrap v. percentile bootstrap





Computational Aspects

- Computational speed is one of the key contributions
 - What once took a day on 30 cores, now takes 5 minutes on a single core
 - Replacement of bootstrap by asymptotics
 - Calculation of covariance matrix of transition probabilities:
code tailored to the purpose much faster than `margins`
- Mata object hierarchy underneath the dtms tree
- Self-threading code (Gould 2018)
- Memory consumption
 - Constant concern: Blows up with number of states and age classes
 - Intermediate results are not stored when a dtms tree is saved



Remaining Shortcomings / Future Improvements

- Asymptotic standard errors ignore the stochasticity of initial proportions
- Saved dtms trees incompatible across command versions
- Restrictions on state transitions not yet possible
 - disease progression models without recovery
 - relaxation of Markov assumption
- Functional form of age relatively restricted
- Import and processing of externally calculated covariance matrices for externally supplied transition probabilities not yet implemented
- Transition probabilities only calculated as conditional margins (this will not change)



Thank you
schneider@demogr.mpg.de



References

- Gould, William W. (2018). *The Mata Book: A Book for Serious Programmers and Those Who Want to Be*. Stata Press.
- Lièvre, Agnès, Nicolas Brouard, and Christopher Heathcote. 2003. "The Estimation of Health Expectancies from Cross-Longitudinal Surveys." *Mathematical Population Studies* 10 (4): 211–48. Available at: <https://doi.org/10.1080/713644739>.
- Lynch, Scott M. and J. Scott Brown (2005). "A New Approach to Estimating Life Tables with Covariates and Constructing Interval Estimates of Life Table Quantities." *Sociological Methodology*, 35(1): 189–237. Available at: <https://doi.org/10.1111/j.0081-1750.2006.00168.x>.
- Schneider, Daniel C. (2023). "Inference for Discrete-Time Multistate Models: Asymptotic Covariance Matrices, Partial Age Ranges, and Group Comparisons." [working title] *MPDIR Working Paper*, forthcoming.
- Schneider, Daniel C. and Mikko Myrskylä (2023). Extending Discrete-Time Multistate Models Using Markov Chains with Rewards: New Outcome Measures and Inference Results. [working title] *MPDIR Working Paper*, forthcoming.



Code from Slides

```
// introductory example
. dtms exempladata 12
. describe
. table cog3
. xttrans cog3

. dtms clear
. dtms setup (impex) , states(1 none 2 mild 3 sevr 4 dead) abs(1) ages(50/110)
. dtms dir

. mlogit cog3 iL.cog3 c.age c.age#c.age sex##educ c.numdrinks , cluster(id) rrr
. dtms estimate frome (impex mlog) , agevar(age, center(50)) orgnvar(L.cog3)
. dtms dir

. dtms transprob atmeans (impex mlog allmeans)
. dtms dir

. dtms proportion fixed (impex pfixed) , values(0.88 0.10 0.02)
. dtms result lexp (impex mlog allmeans lexp) , initprop(pfixed) format(%7.1f)
. dtms dir
```



Code from Slides

```
// introductory example (continued)
. dtms transprob atmeans (impex mlog edlow) , at(educ=1)
. dtms transprob atmeans (impex mlog edhigh) , at(educ=3)
. dtms result lexp (impex mlog edlow lexp) , initprop(pfixed) format(%7.1f)
. dtms result lexp (impex mlog edhigh lexp) , initprop(pfixed) format(%7.1f)

. dtms result stab (impex mlog edlow stab) , initprop(pfixed) format(%7.2f)
. dtms result stab (impex mlog edhigh stab) , initprop(pfixed) format(%7.2f) post nopvalues

. dtms dir

* dtms file save using temptree.dtms , replace

. dtms combine (impex mlog edhigh lexp) , with(edlow lexp) post(state total)
. dtms combine (impex mlog edhigh lexp) , with(edlow lexp) post(state total) diff
```



Code from Slides

```
// Comparison Analytic v. Bootstrap CIs (bootstrap code not included)
//      LEXP (life expectancy)
. dtms erestore (impex mlog allmeans lexp) , replay

//      MAFN (mean age at first entry)
. dtms result mafn (impex mlog allmeans mafn) , initst(1) intermed(2) target(3) post(lrsk mafn)

. dtms trans atmeans (impex mlog men) , at(sex=1)
. dtms trans atmeans (impex mlog wmn) , at(sex=2)
. qui dtms result mafn (impex mlog men mafn) , initst(1) intermed(2) target(3)
. qui dtms result mafn (impex mlog wmn mafn) , initst(1) intermed(2) target(3)
. qui dtms combine (impex mlog allmeans mafn) , with(men mafn) post(lrsk mafn)
. dtms combine e() , with(wmn mafn)
```



Code from Slides

```
// Comparison Analytic v. Bootstrap CIs (continued)
//      Rewards: EPIS (# of episodes)
. dtms result epis (impex mlog allmeans epis) , initpr(pfixed) post

//      Rewards: MAIS (mean age in state)
. dtms result mais (impex mlog allmeans mais) , initpr(pfixed) post

// Group Comparisons
dtms result lexp (impex mlog wmn lexp) , initpr(pfixed)
dtms result lexp (impex mlog men lexp) , initpr(pfixed)
dtms combine (impex mlog wmn lexp) , with(men lexp) difference

// Partial Age Ranges [code not shown]
```