Calibrated Spline Graduation of Age-Group Fertility Rates

UN Expert Group Meeting
Methods for the WPP 2021+

April 2020

Carl Schmertmann
Florida State University
Tallahassee, Florida USA

Objective

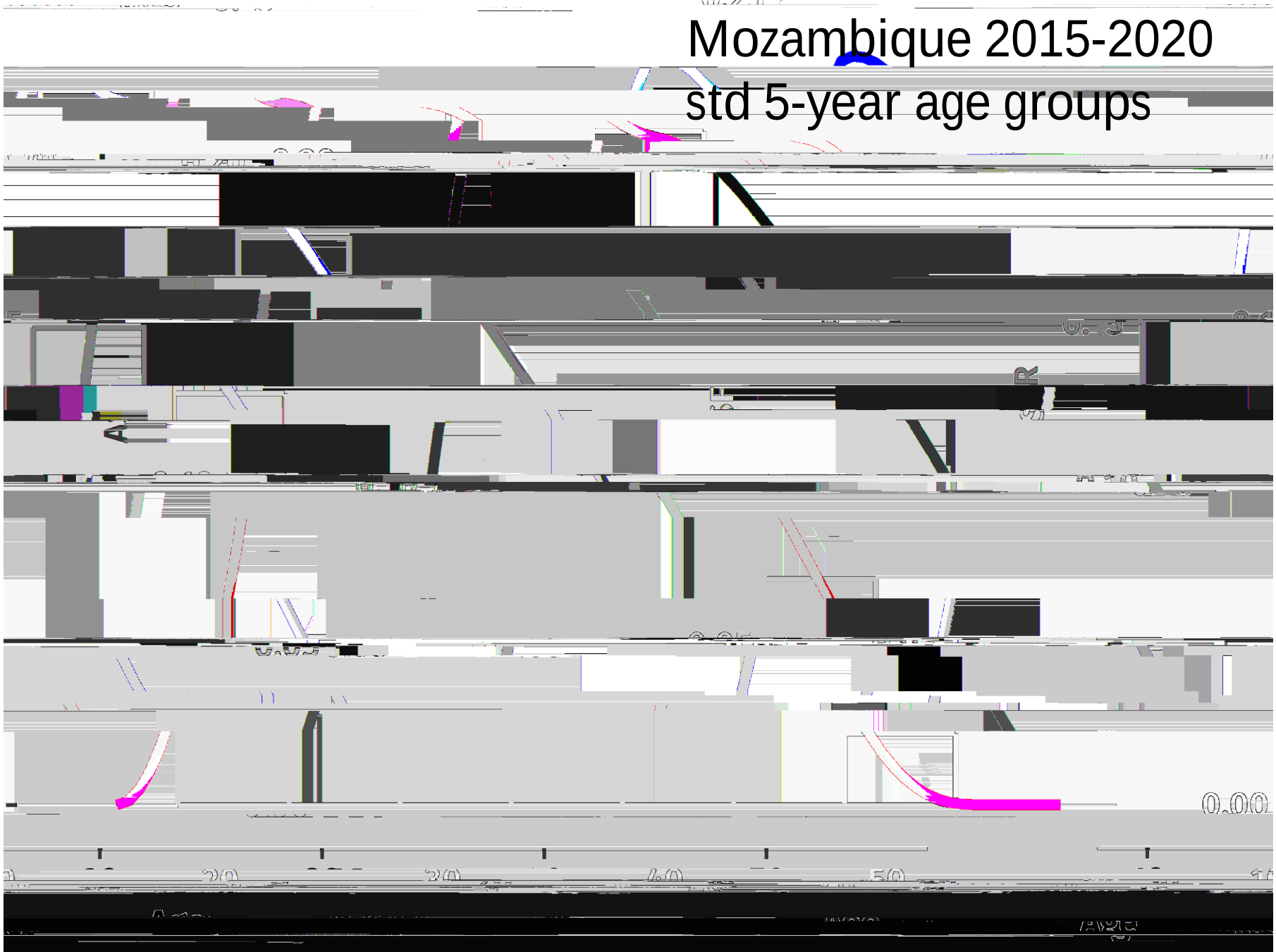
For age-group fertility data { }
find a continuous fertility schedule

Principles

Construct ( ) such that it can be estimated

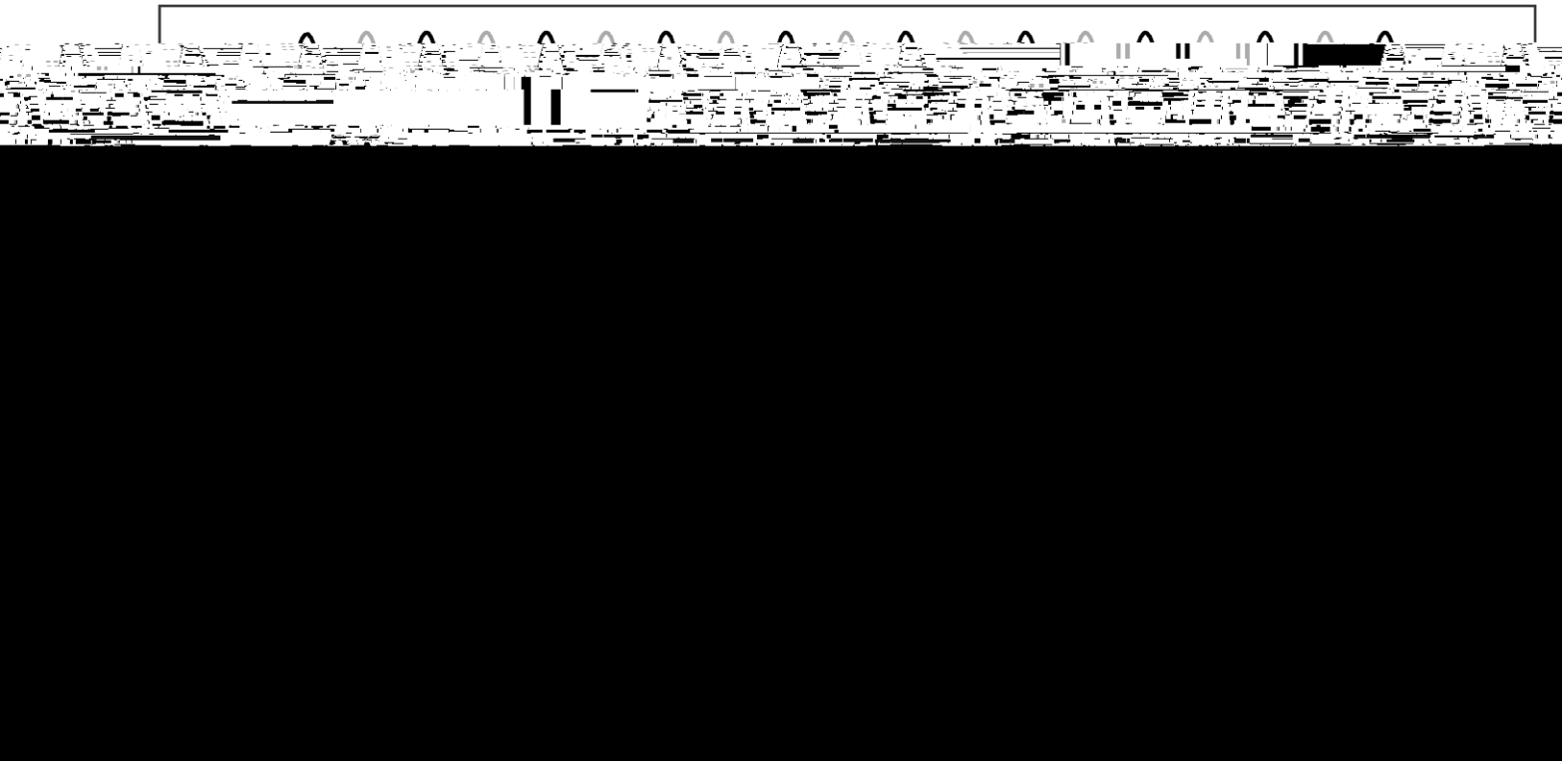
1. from any age grouping { }
2. for ages outside available { }
   e.g. <15 or >50
3. using simple spreadsheet arithmetic

WPP 2019: Mozambique 2015-2020 std 5-year age groups



Main Ideas

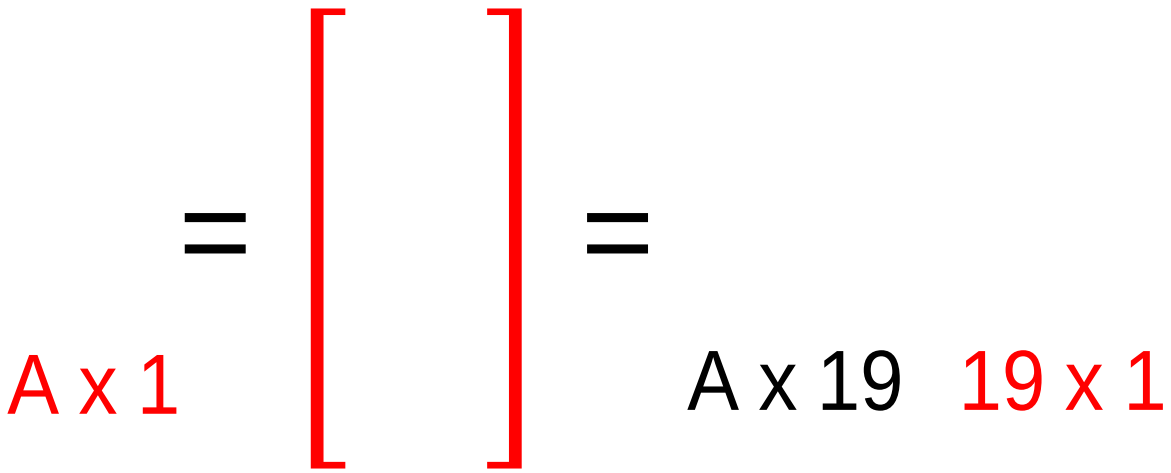
(



Main Ideas

Discretize over a fine grid of A ages such as

$(x_1 \ldots x_A) = (12.25, 12.75, \ldots, 54.75)$



Main Ideas

Observed rates for groups are averages of the detailed rates

{

Any particular θ implies spline fitting errors

$$= \{ \quad \} -$$

$$7 \times 1 \qquad 7 \times 1 \qquad 7 \times A \quad A \times 19 \quad 19 \times 1$$

Lower SSE → better θ

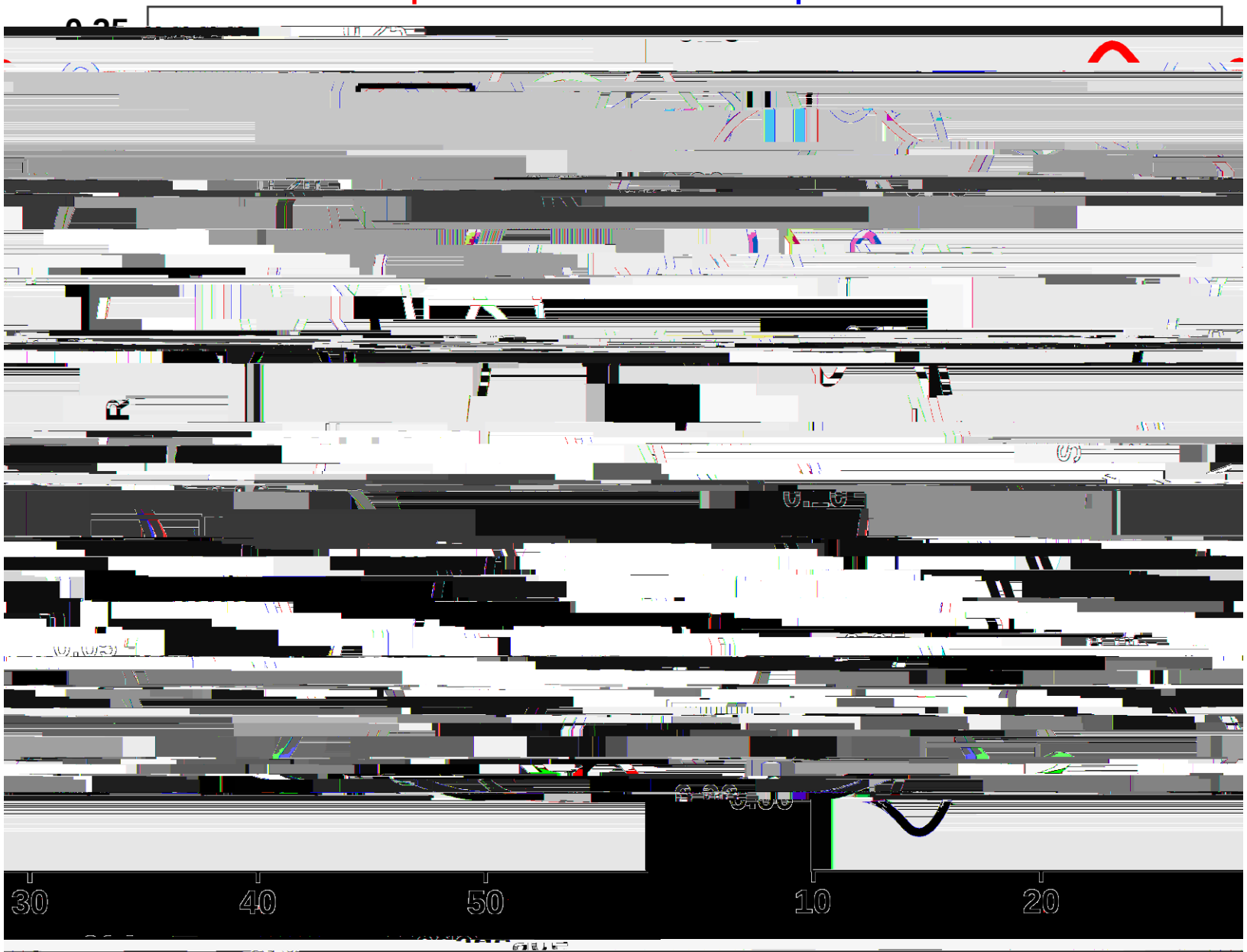
$$(\ ) = \quad '$$

Singular Value Decomposition

find principal components of single-year age schedules over *CALIBRATION DATA*

$f(x)$

Same fit. SSE_{shape}=91, SSE_{shape} = 1050



Objective Function:

select θ to minimize

$$(\quad) = \underbrace{\qquad\qquad (\quad)}_{\text{quadratic in } \theta, \text{ includes } \{{}_nF_x\}} + \underbrace{\qquad\qquad (\quad)}_{\text{quadratic in } \theta}$$

Unique, closed-form solution

$$= \quad C \{ \quad \}$$

$$\equiv \qquad = \quad K \{ \quad \}$$

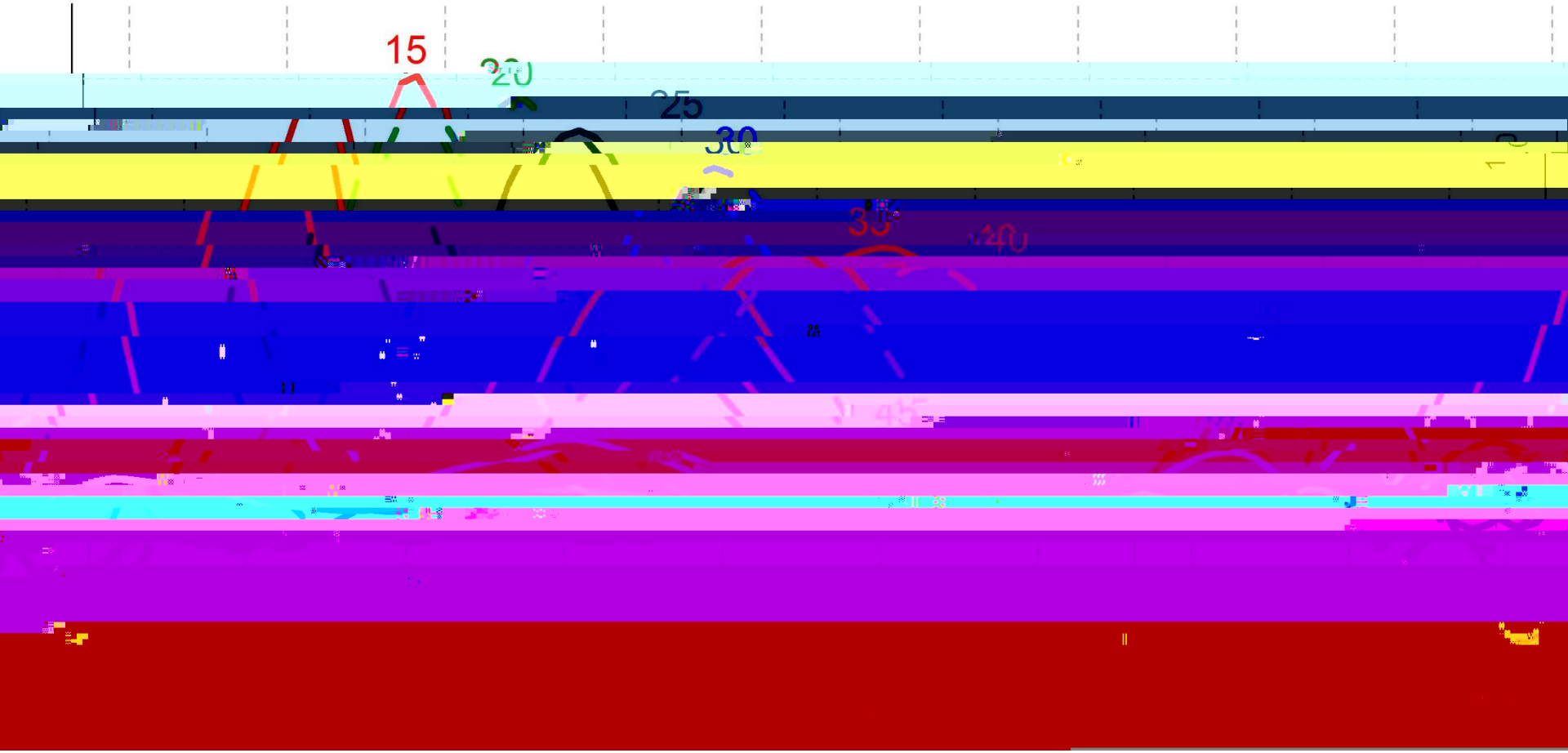
$A \times 1 \qquad\qquad A \times 7 \quad 7 \times 1$

A simple final product

Fitted schedule is a wtd sum of K's columns
Weights = observed ${}_nF_x$ values

$$f = k_1 \cdot \quad \ldots + k_7 \cdot$$

Columns of K matrix (Calibrated splines)



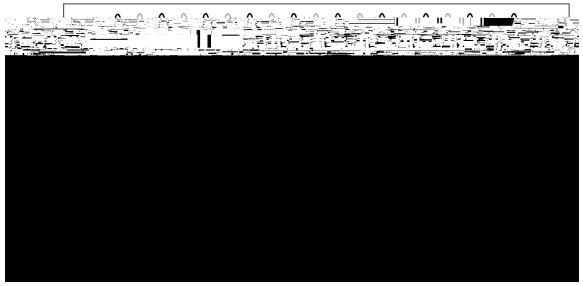
Examples (change windows, Carl!)

Issues/Problems

Graduated schedule does not *exactly* match age-group rates (unless $\lambda \rightarrow \infty$)

Graduated rates occasionally have (very!) small negative values at highest and lowest ages

WPP Modeling Choices



Number of spline knots for B

Order of splines for B (quadratic, cubic, ...)

Age grid ($x_1 ... x_A$) for discretizing f(x)

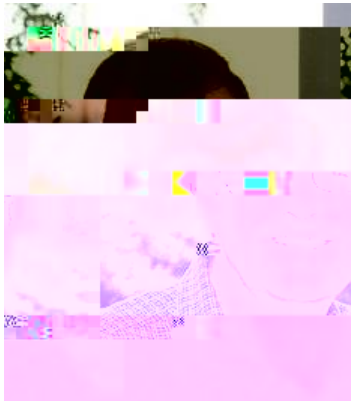
CALIBRATION DATA (HFD, HFC, smoothed WPP,...)

Number of principal shape components (3,4,...)

Relative weight on fitting errors λ (should vary with sample size, data quality)

Add a smoothing penalty? (e.g. squared 2nd diffs in rates)

THANKS!

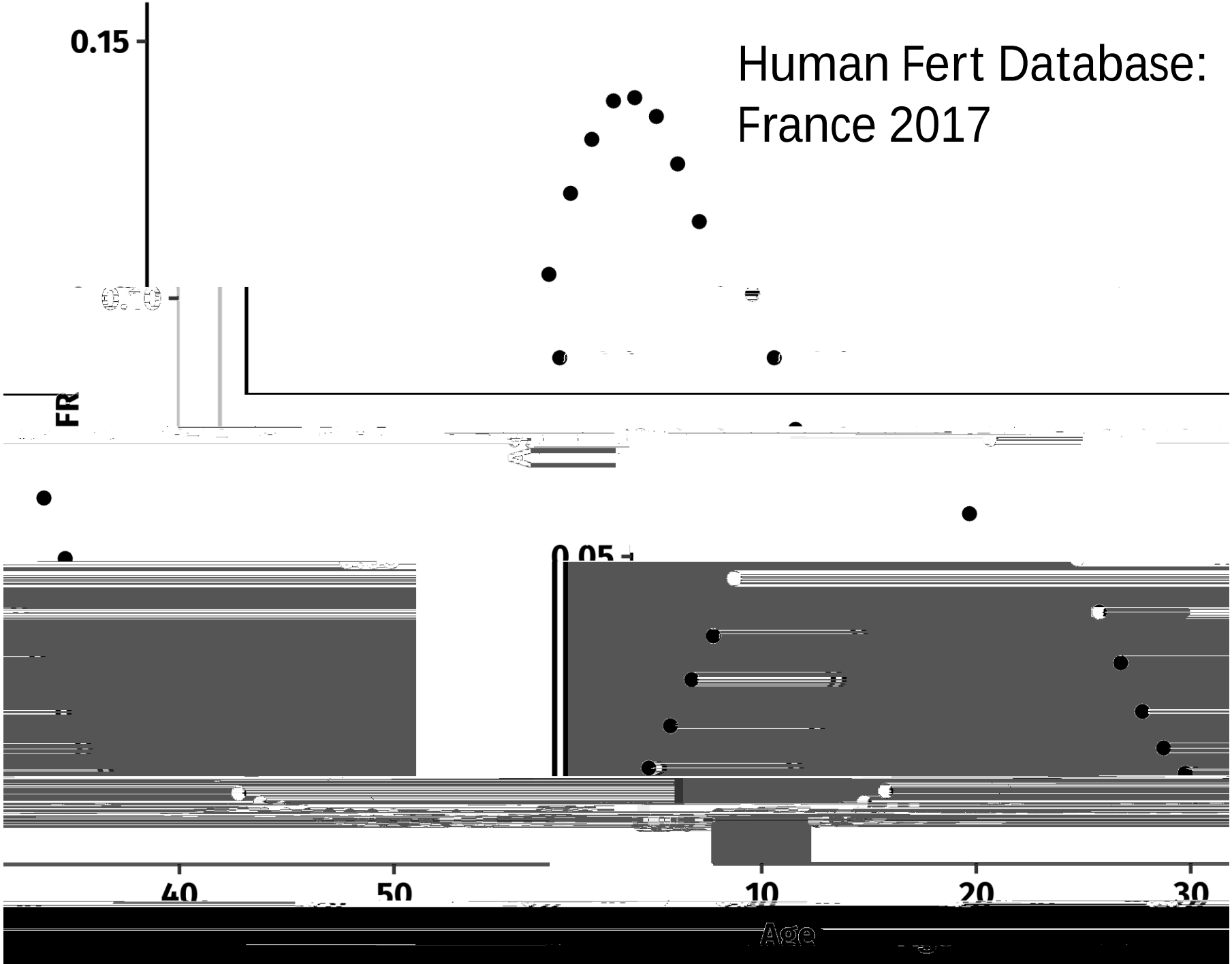


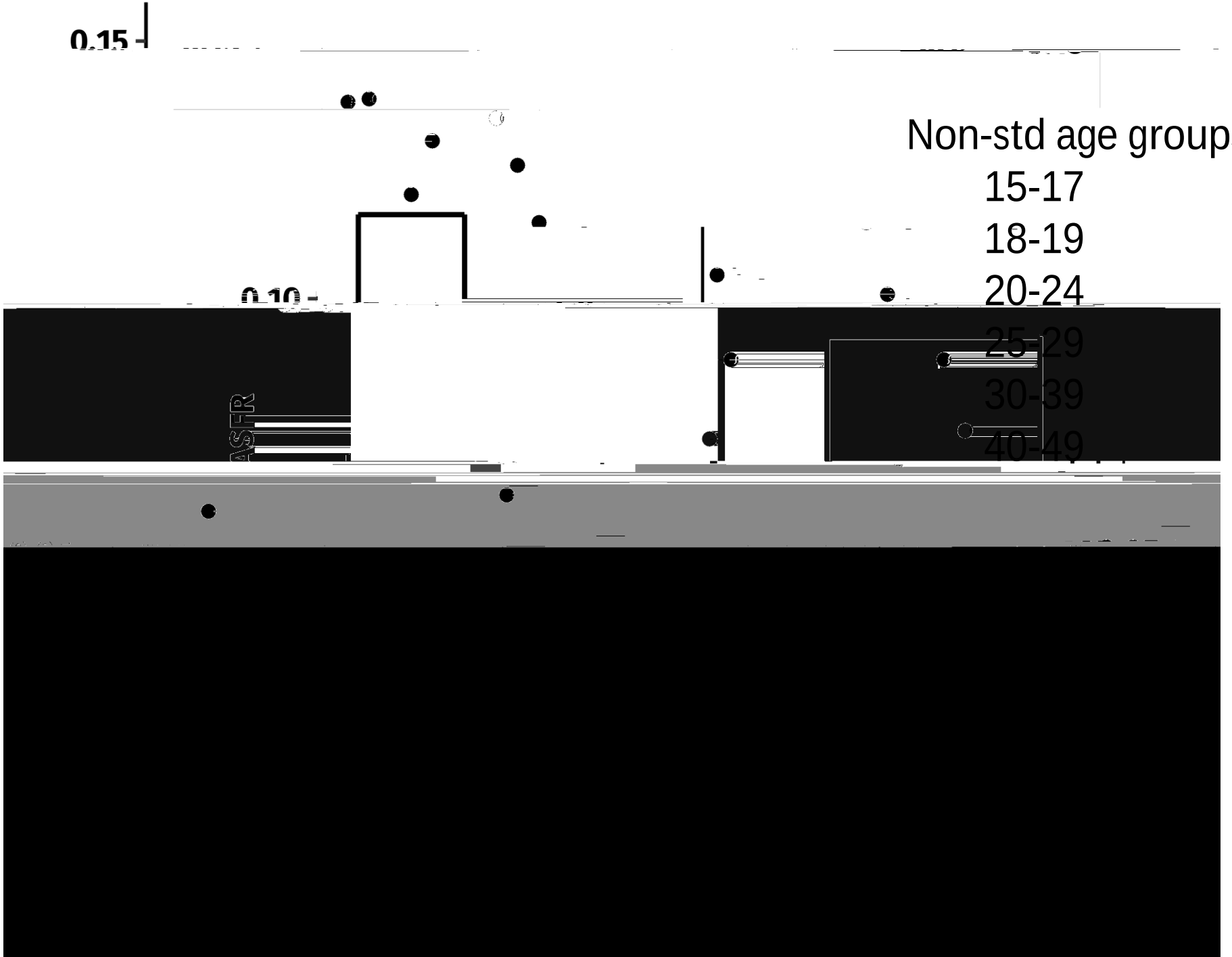
Carl Schmertmann

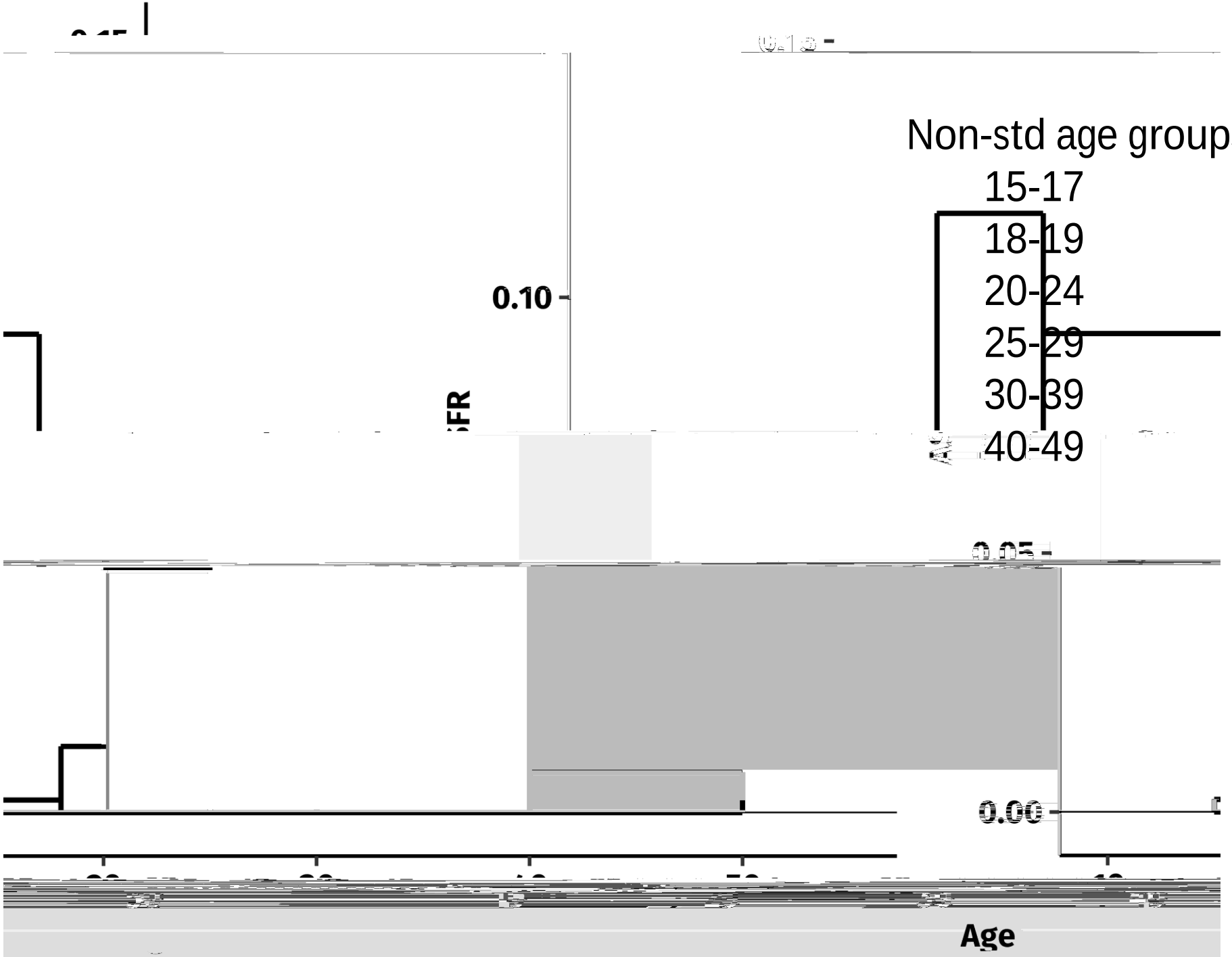
Article at tinyurl.com/fertility-splines

Data/code at tinyurl.com/fertility-splines-replication

Non-standard age groups
(same procedure, different G matrix)


Human Fert Database:
France 2017
0.15
FR
0.05
40
50
10
20
30
Age


0.15
0.10
ASFR
Non-std age groups
15-17
18-19
20-24
25-29
30-39
40-49


Non-std age groups
15-17
18-19
20-24
25-29
30-39
40-49
0.10
0.05
0.00
Age


Non