Modeling age-specific mortality by detailed age between 0 and 5 years: Results from a log-quadratic model applied to high-quality vital registration data

Michel Guillot,[1,2] Julio Romero Prieto,[3]
Andrea Verhulst,[2] Patrick Gerland[4]

[1]French Institute for Demographic Studies (INED), [2]University of Pennsylvania,
[3]Banco de la República de Colombia, [4]United Nations

Funding: Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), Grant R01HD090082

Objectives

- Model the shape of the mortality curve between age 0 and 5 years by detailed age group (weeks, months, trimesters, years)
- For predicting a full mortality schedule between 0 and 5 with only 1 or 2 parameters

Significance

- Importance for health policy of examining how the risk of death varies within the 0-5 age range (NNMR, IMR, but also detailed information by days, weeks, and months of age)
- In spite of their importance, age patterns of U5M are difficult to establish in LMICs, due to lack of reliable data
- Use of model life tables (MLT) to address these deficiencies
  - Coale & Demeny, United Nations models
  - Used for estimating IMR on the basis of U5MR
- Drawbacks
  - Only 0 vs. 1-4 as age details; based on limited number of country-years
- Proposed model builds on MLT approach but with new data and finer age granularity

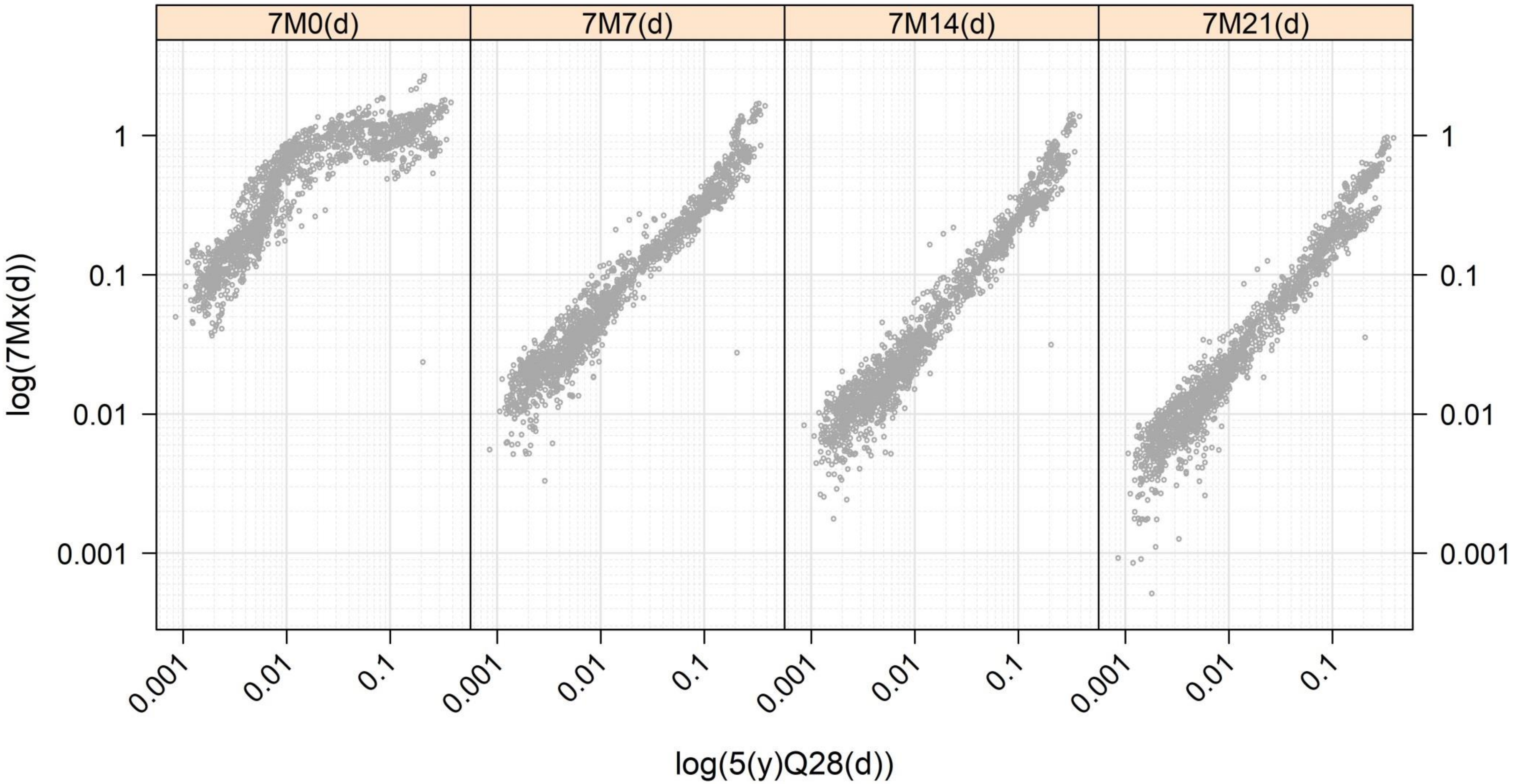
Scope of current model

- Based on high-quality VR countries represented in the Human Mortality Database (HMD)
  - European countries + Australia, Canada, Chile, Israel, Japan, New Zealand and the US
  - Small populations (Iceland, Luxembourg) and Former Soviet bloc countries are excluded

A new mortality database for under-five mortality by detailed age

- Annual distributions of under-five deaths by sex and detailed age (days, weeks, months, trimesters, years)
- Two components:
  - UN database since 1970
  - Archival work for the pre-1970 period
- 22 harmonized age groups: 0d, 7d, 14d, 21d, 28d, 2m, 3m, 4m, 5m, 6m, 7m, 8m, 9m, 10m, 11m, 12m, 15m, 18m, 21m, 2y, 3y, 4y, 5y
- Use of HMD exposure terms for denominator of rates
- Exclusion of a number of country-years for the late 19th – early 20th century due to data quality concerns at early neonatal ages
- Final database for modeling: 1235 country-years, by sex

${}_nM_x$ vs q(28d,5y) for each of the first four weeks



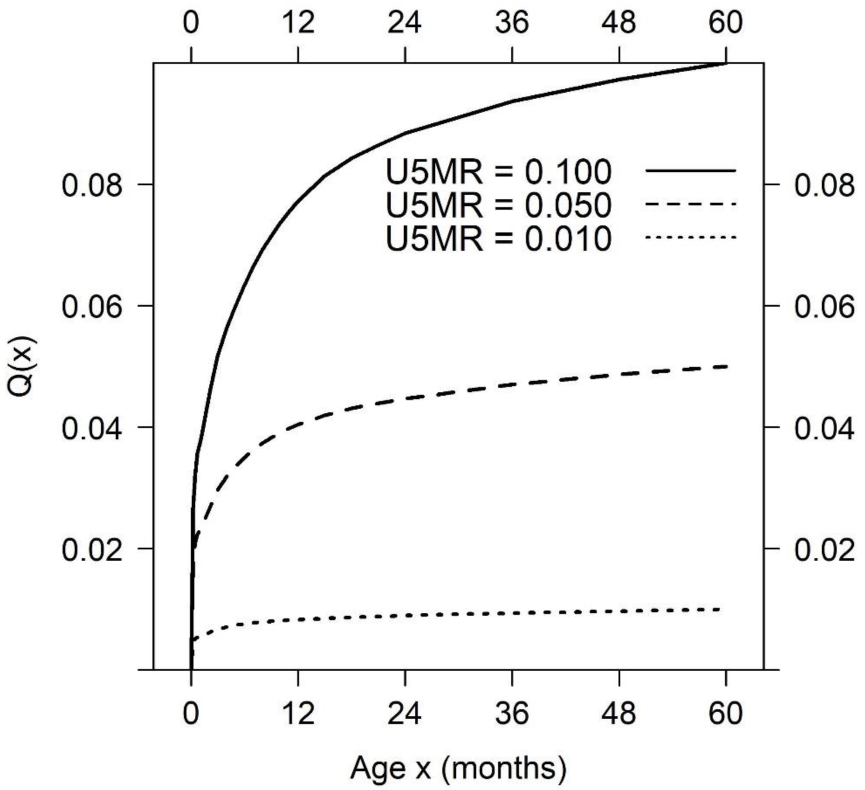
Modeling approach: log-quadratic model

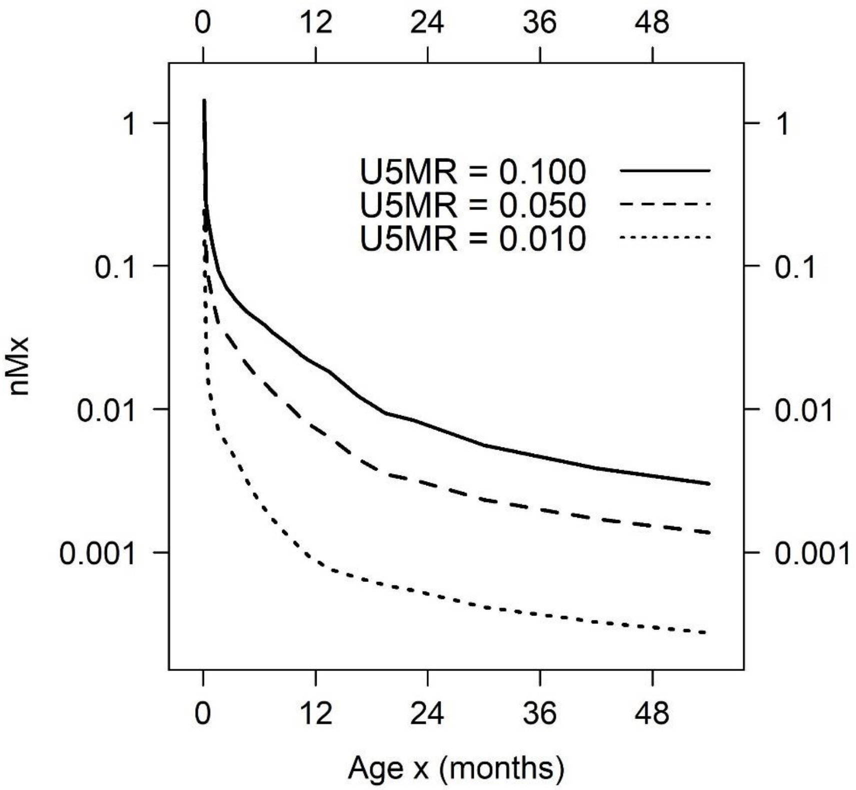
- Adapted from Wilmoth et al. (2012)
- Using U5MR (=q(5)) as main explanatory variable
- Using q(x), the cumulative probability of death from birth to age x (x=7d, 14d, …, 3y, 4y) as response variable

$$\ln[q(x)] = a_x + b_x \ln[q(5)] + c_x \ln[q(5)]^2 + v_x k$$

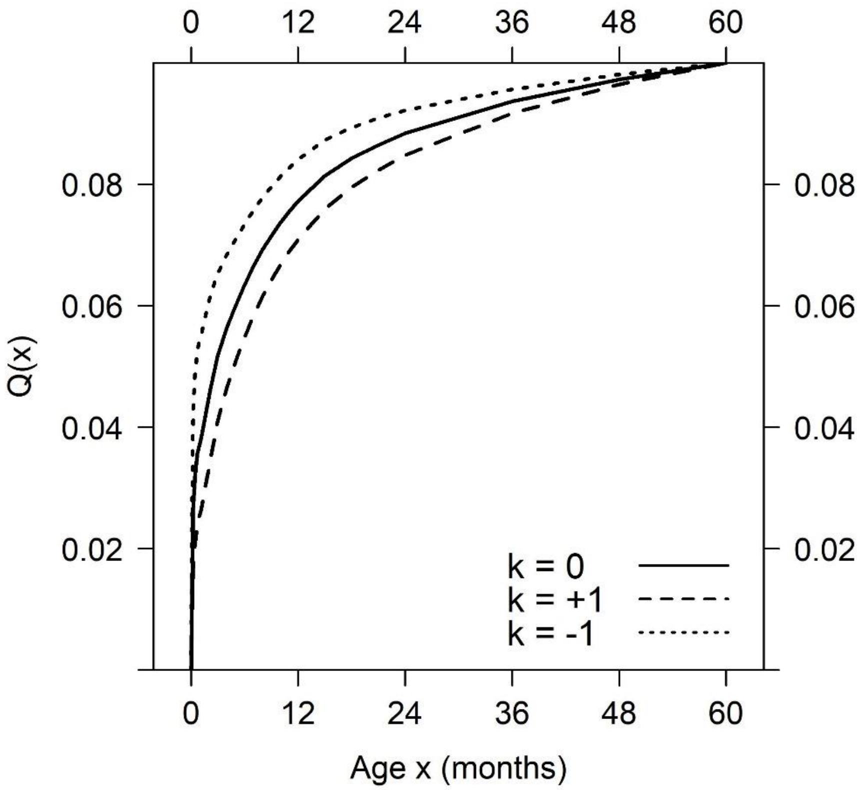
- q(5) determines overall level of mortality
- k affects the shape of the age pattern (k=0 -> average age pattern)
- a_x, b_x, c_x estimated via OLS; v_x via Singular Value Decomposition (SVD)

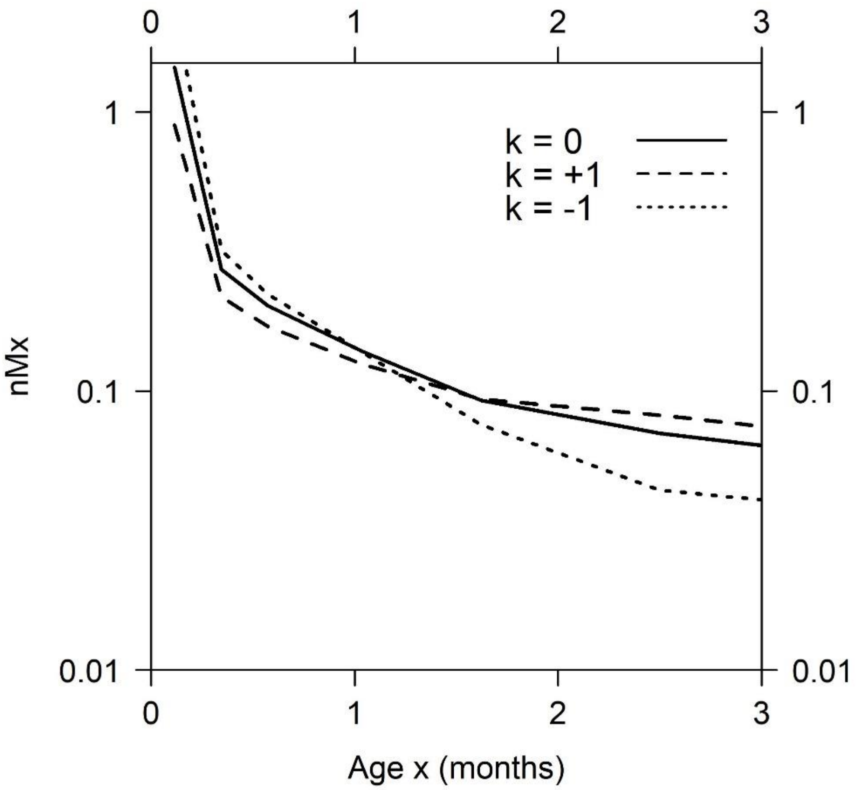
Effect of varying U5MR on predicted q(x) and $_nM_x$ (with k=0)



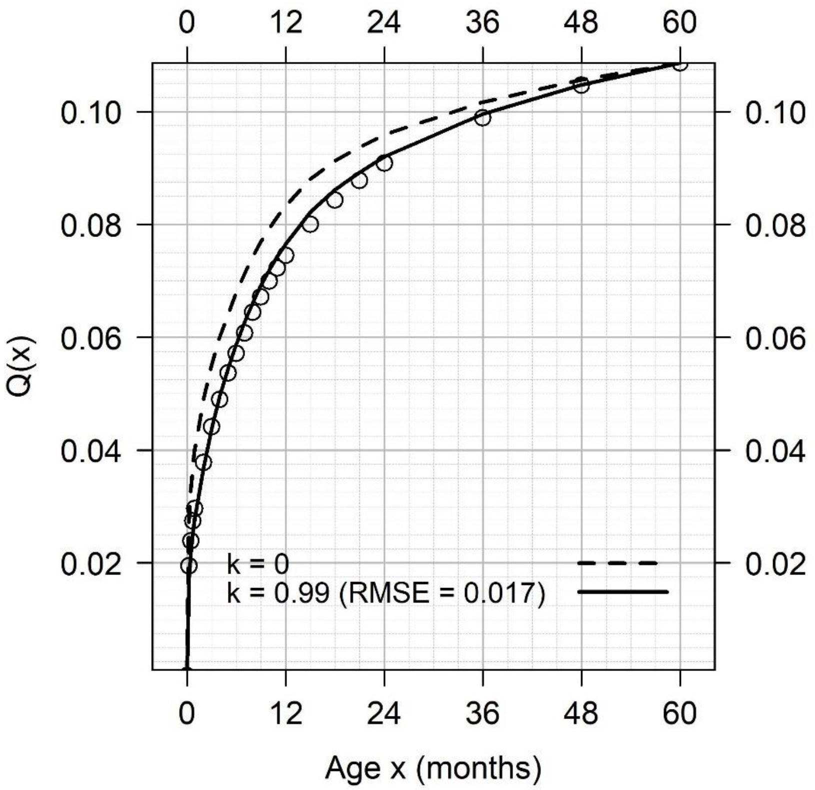


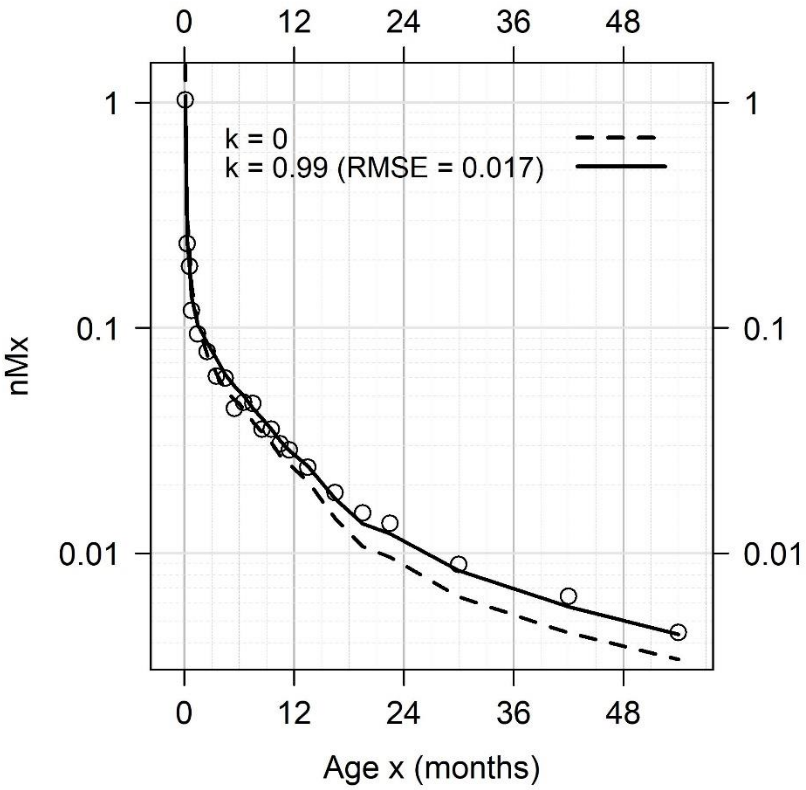
Effect of varying k on predicted q(x) and ${}_nM_x$ (with U5MR=.100)





Fitting the model to data for a given population (Example: Finland 1933, both sexes)





U5MR is given by the data; k is the solution that minimizes the RMSE of predicted q(x)’s.

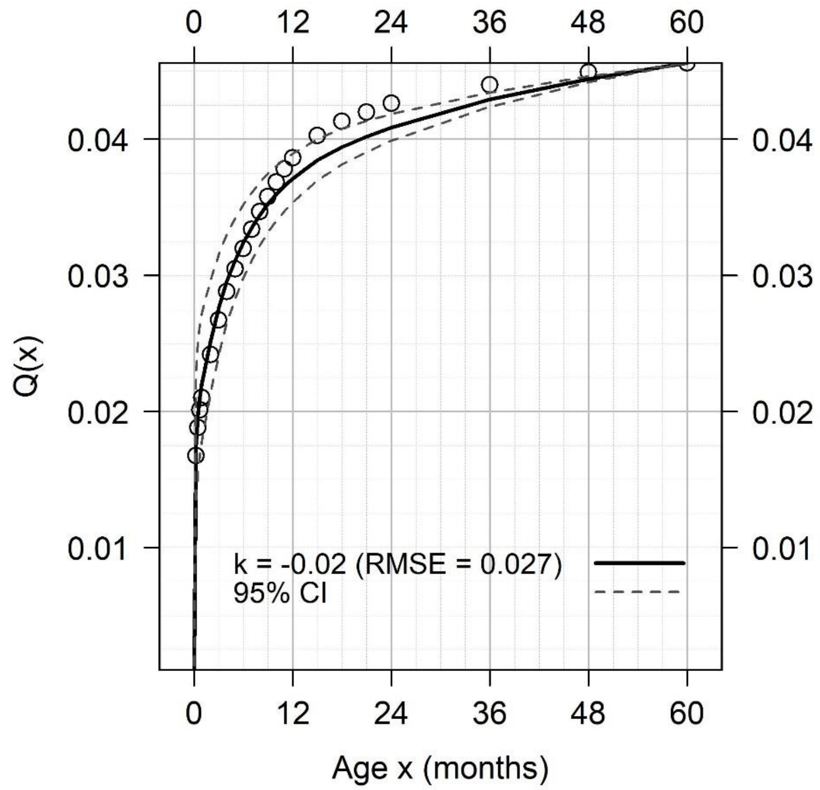
How does the model fit the entire database? Overall RMSE of predicted q(x)'s:

Entry points		Female	Male	Both
$q(5y)$ only		0.04131	0.04238	0.04049
$q(5y)$ +	$q(7d)$	0.02698	0.02637	0.02491
	$q(28d)$	0.02408	0.02442	0.02271
	$q(3m)$	0.02078	0.02086	0.01922
	$q(6m)$	0.02396	0.02440	0.02269
	$q(12m)$	0.03399	0.03468	0.03265
All q(x) values		0.01901	0.01911	0.01771

60-40 Monte Carlo cross validation: 741 life tables used for estimation and 494 for evaluation. Reported values are the mean of the RMSE of predictions based on 10,000 random samples of 494 country-years.

Estimating confidence intervals around predicted q(x) values (France, 1955, both sexes)

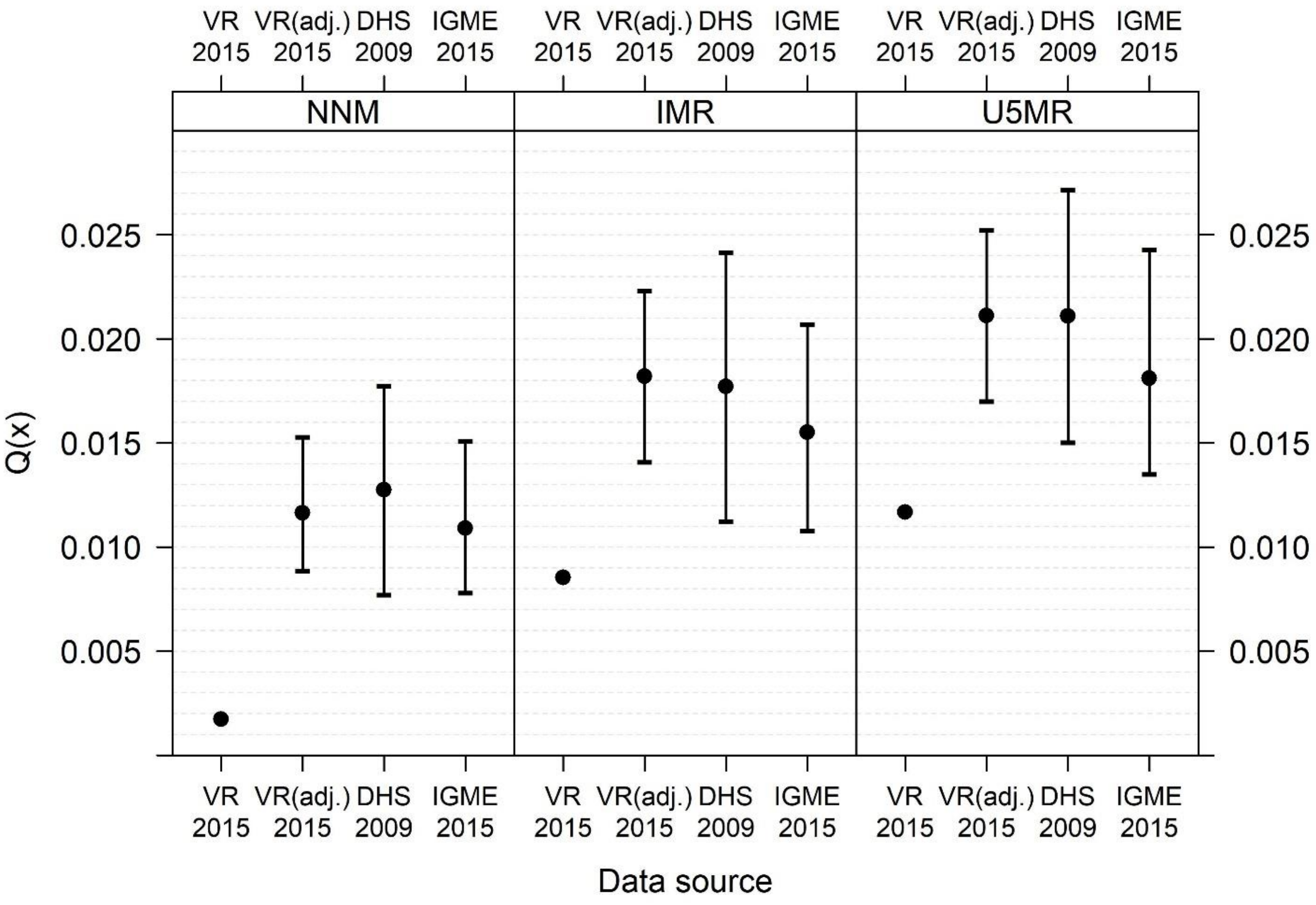
- Approach based on variance around the optimal value of k for given population
- Takes into account prediction errors for each age group and the width of each age interval
- $$var(k_i) = \left[\frac{\sum_{x \in X} w(x) \cdot e_i(x) \cdot e_i(x)}{\sum_{x \in X} w(x) \cdot \mathrm{v}_x \cdot \mathrm{v}_x} - k_i^2\right]$$



Using the model for indirect estimation

- Model can be used for:
  - Smoothing noisy age schedules
  - Correcting mortality estimates in the presence of age heaping (Romero Prieto, Verhulst and Guillot, 2019)
  - Adjusting mortality data for under-reporting in specific age ranges
- Application: VR data from Jordan (2015)
  - Concerns about quality of the VR neonatal information in that country
  - Model parameters estimated using reported mortality from 28 days to 5 years, i.e., excluding reported mortality at neonatal ages
  - Adjusted values of neonatal, infant and under-five mortality, with 95% CI

Reported vs adjusted neonatal, infant, and under-five mortality in Jordan, 2015, both sexes



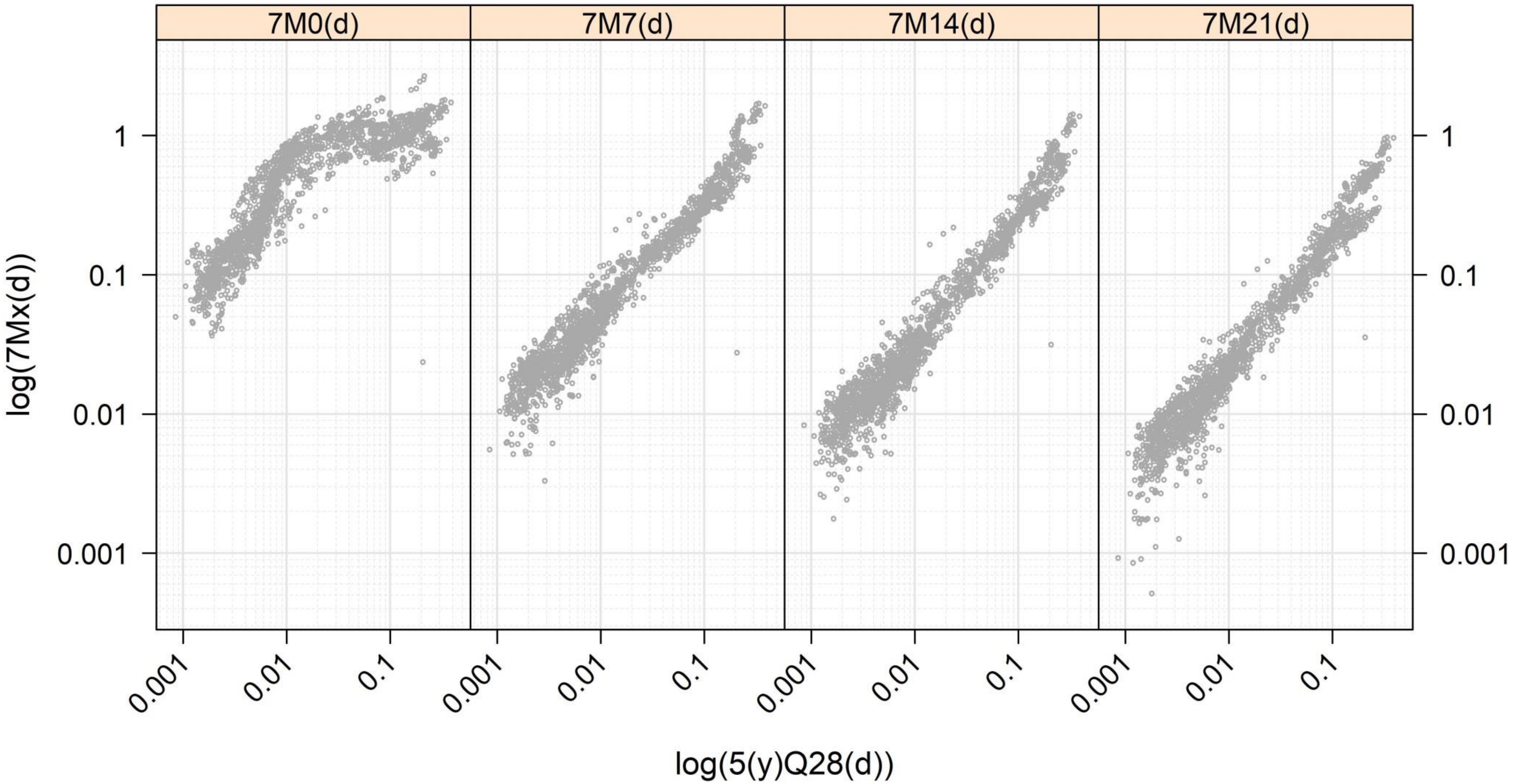
Conclusion: strengths

- Log-quadratic model provides more detailed age groups than existing model life tables
- Flexible choice of predictors
- Model fits the historical VR data well
- Jordan example: promising results for correcting incomplete VR data

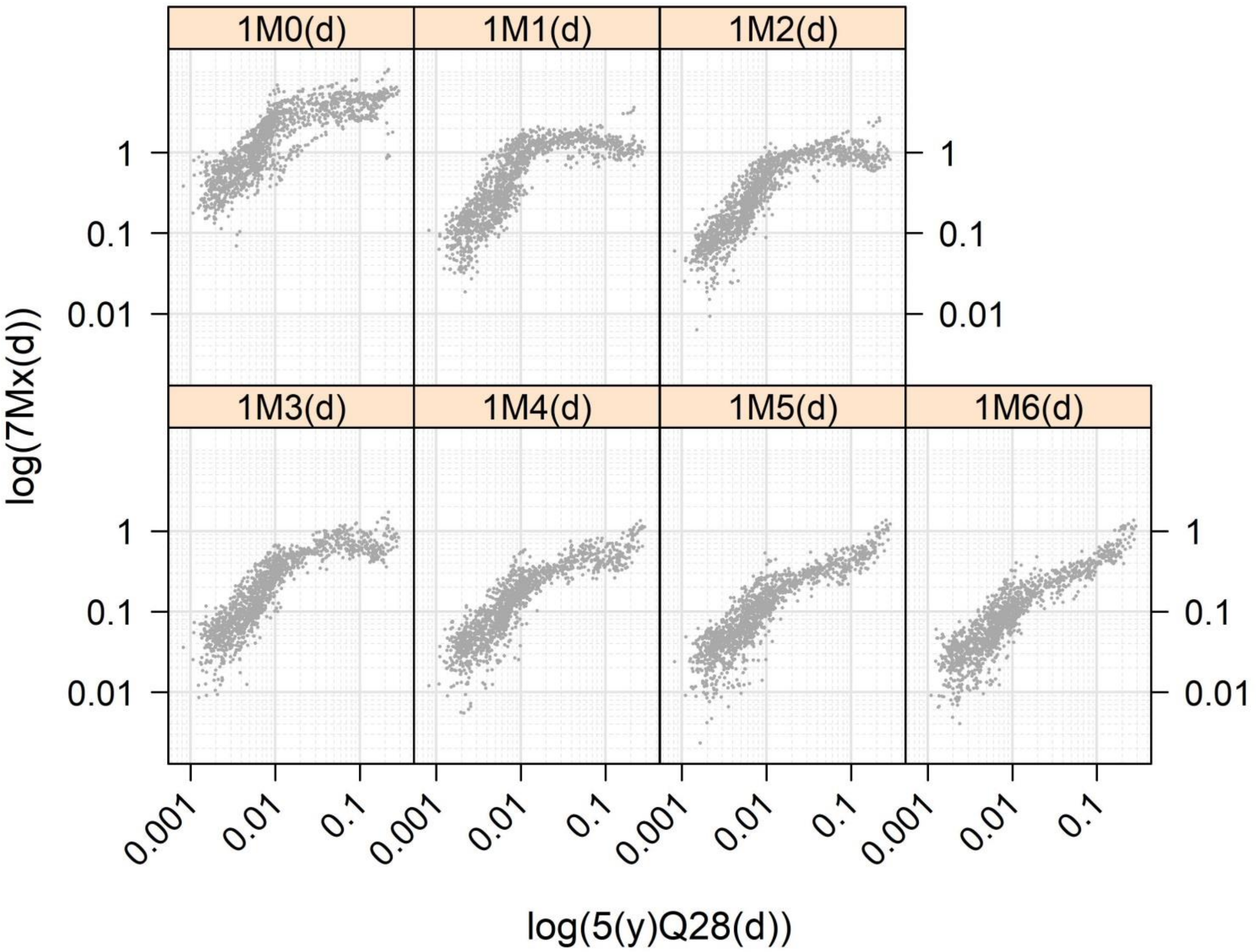
Conclusion: limitations

- The model's empirical basis does not include mortality data from low and middle-income countries
- Model's applicability goes beyond HMD countries (e.g. Jordan), but model will need to be updated with information from non-HMD countries
- Compilation of additional sources from developing countries (SRS, HDSS, Cohort studies) under way as part of ongoing R01 project

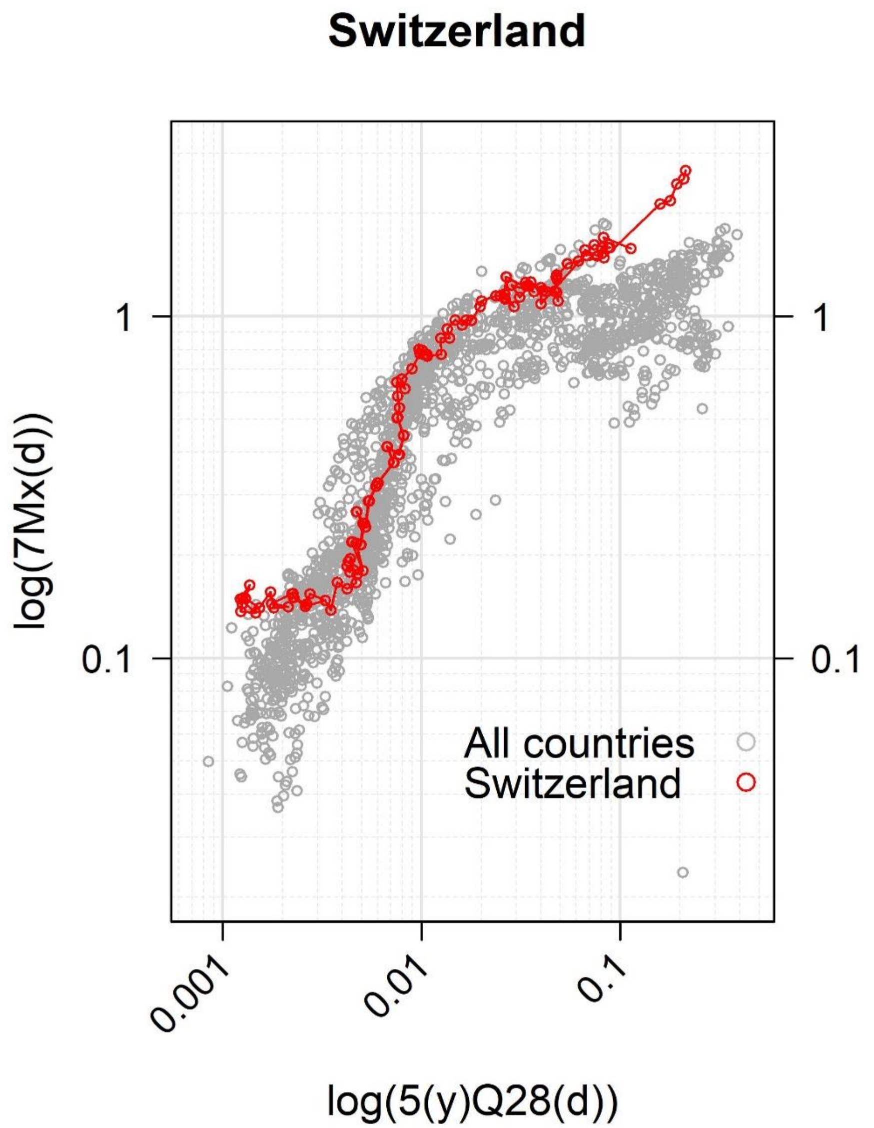
${}_{n}M_{x}$ vs q(28d->5y) for each of the first four weeks



$_nM_x$ vs q(28d->5y) for each of the first seven days



$_7M_0$(d) vs q(28d->5y) in Switzerland



${}_{7}M_{0}$(d) vs q(28d->5y) in the DHS