Quantifying Harvesting:
Estimating and forecasting the effect of mortality shocks on population dynamics

Trifon I. Missov

Interdisciplinary Center on Population Dynamics CPop
University of Southern Denmark SDU

SCOR FOUNDATION FOR SCIENCE

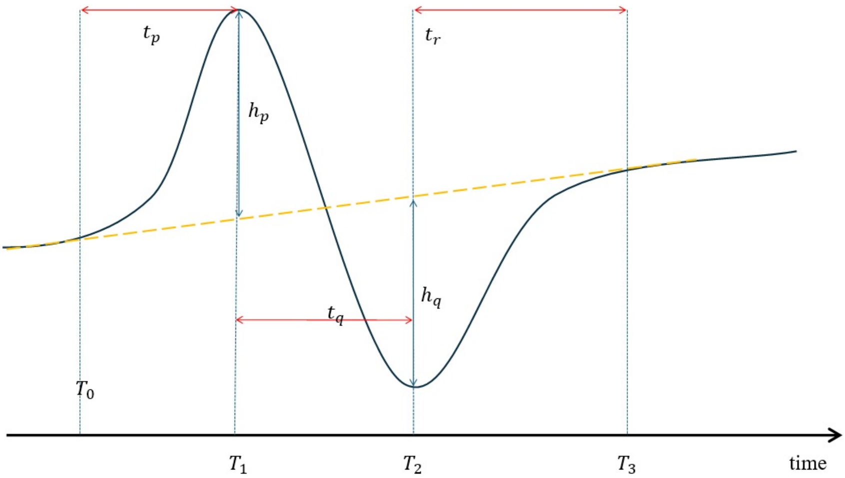
2025 Annual Meeting of the Population Association of America
Washington, DC, April 12, 2025

Harvesting

Mortality displacement when a sudden shock eliminates imminently a higher than anticipated number of individuals, especially those that are expected to die soon.

The outcome, in terms of death counts or death rates, is a curve with a sharp peak, corresponding to the excess deaths caused by the shock, followed by a trough, reflecting the lack of deaths as the expected individuals at risk have already been eliminated.

Harvesting: a simple visualization



COVID-19 Pandemic: A Natural Setup

- Mortality shock: pandemic (multiple waves)
- Data availability: daily/weekly/monthly

  STMF (www.mortality.org)
  MoCy (github.com/jschoeley/mocy)
  COVerAGE-DB (www.coverage-db.org)

- Main task: calculate expected deaths

Expected deaths

Weekly death counts $D_t \sim \text{Poisson}(\lambda_t, \varphi)$ with $\lambda_t = \exp(\xi_t)$.

1. 5-year average model (simple, but with low accuracy)
   $\xi_t = \alpha_t + \ln E_t$

2. Serfling method (accounts for time trends, seasonality)
   $\xi_t = \nu_t + \gamma_{1,t} sin\frac{2\pi t}{52} + \gamma_{2,t} cos\frac{2\pi t}{52} + \delta_{t,k[t]} + \ln E_t$

   Extensions (used by Léger, Rizzi & Basellini 2025):
   smooth trend & seasonality (SP-STSS, Eilers et al. 2008)
   smooth trend & fixed seasonality (SP-STFS)

3. Latent Gaussian model
   $\xi_i = \alpha + \nu_t + \gamma_t + \delta_{t,k[t]} + \zeta_{t,t-1} + \varepsilon_t + \ln E_t$

CPop Expected deaths: Accuracy & Model Choice

Schöley (2021), using weekly data, and Léger, Rizzi & Basellini (2025), using monthly data, compared the accuracy of forecasts on multiple countries. They arrived at a similar conclusion in favor of a Serfling model with relaxed specification (assuming smoothness as in SP-STFS or something else). We make a small extension:

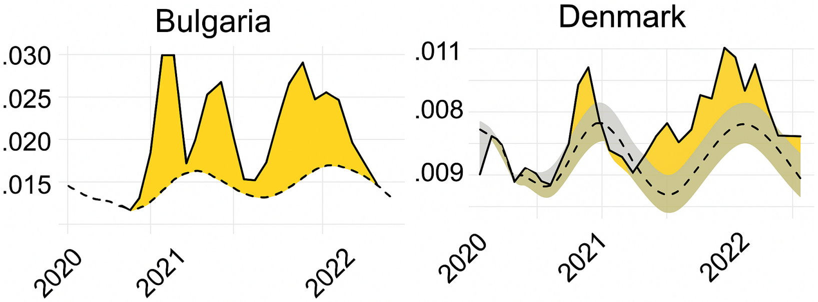
SQP-STFS: quasi-Poisson distributed deaths in week t, D_t:
$D_t \sim Poisson(\lambda_t, \phi)$, $\phi > 0$, $\lambda_t = \exp \xi_t$, and

$$\xi_t = \nu_t + \beta_1 \sin \frac{2\pi t}{52} + \beta_2 \cos \frac{2\pi t}{52} + \ln E_t$$

$$\nu_t = \sum_i \alpha_i B_i(t) = \mathbf{B}\boldsymbol{\alpha} \, ,$$

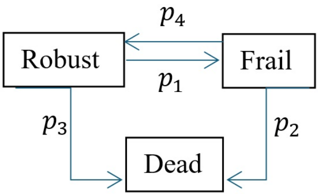
where $\mathbf{B} = [b_{ti}] = [B_i(t)]$ is a B-spline basis (Eilers and Marx 1996, Léger, Rizzi & Basellini 2025). Exposures E_t are an offset.

Harvesting?


Bulgaria
.030
.025
.020
.015
2020
2021
2022
Denmark
.011
.008
.009
2020
2021
2022

Delayed Effects (Work in Progress)

Is there something more we can do? Are there any delayed effects?

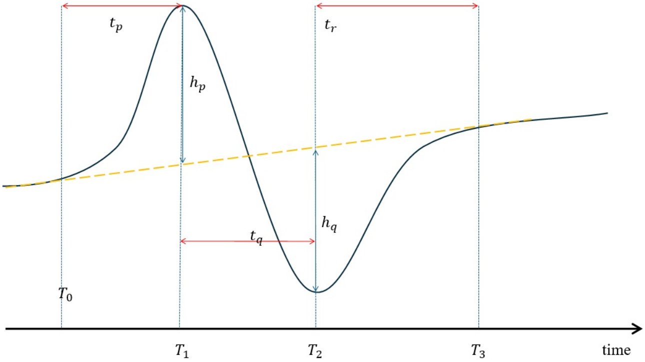


What if we consider densities?

Suppose

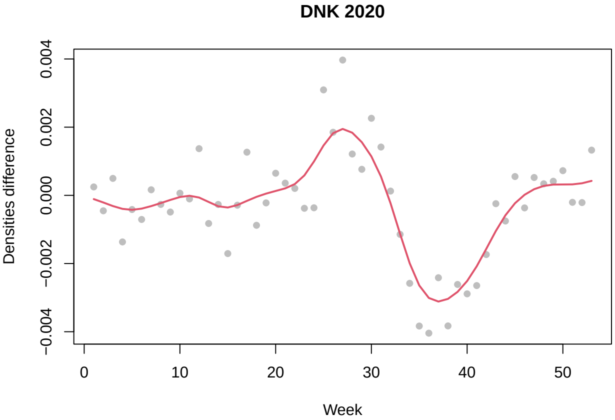
- $D_{t,y}$ = death counts in week t in year y
- $d_{t,y} = D_{t,y} / \sum_y D_{t,y}$ is the density of deaths for that week, normalized by the total deaths across years
- $\bar{d}_{t,y}$ represents the five-year average of densities for week t from year from year y to $y + 4$
- $diff_t = d_{t,y} - \bar{d}_{t,y}$
  For the COVID year, y^*, we calculate $diff_t$ as the difference between d_{t,y^*} and the five-year average density, $\hat{d}_{t,y^*-5}$, for the same week.
- To smooth the variations in $diff_t$, we apply the Nadaraya–Watson kernel regression.

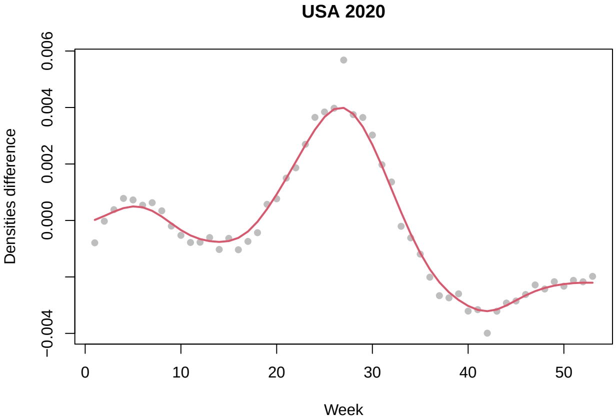
CPop

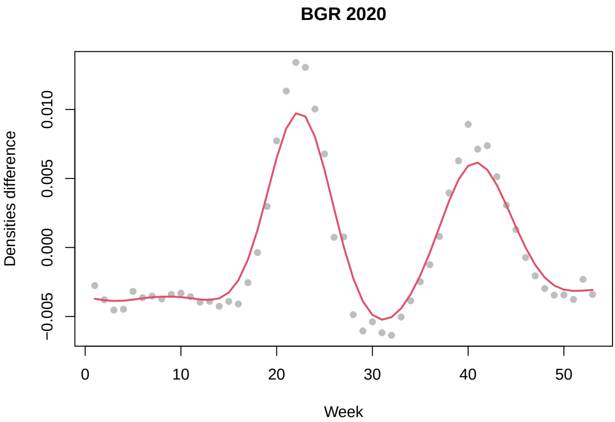


To determine T_1 and T_2, we carry out global optimization by the differential evolution technique (R library DEoptim).

T_3, the inflection point occurring after T_2, is calculated by using a combination of numerically estimated Hessians and applying the Richardson extrapolation method, implemented through the numDeriv package in R.

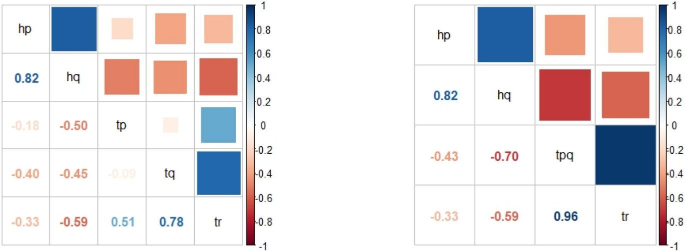

DNK 2020
Densities difference
0.004
0.002
0.000
−0.002
−0.004
0
10
20
30
40
50
Week


USA 2020
Densities difference
0.006
0.004
0.002
0.000
−0.004
0
10
20
30
40
50
Week



BGR 2020
Densities difference
0.010
0.005
0.000
−0.005
0
10
20
30
40
50
Week

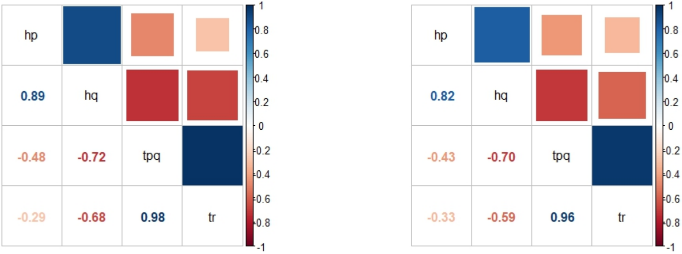
CPop



Correlation between pairs of h_p, h_q, t_p, t_q, and t_r (left panel), as well as between pairs of h_p, h_q, $t_{pq} = t_p + t_q$, and t_r (right panel) based on the 14 studied populations (COVID-19 epi-years).

←

Immunity?



Correlation between pairs of h_p, h_q, t_p, t_q, and t_r (left panel), as well as between pairs of h_p, h_q, $t_{pq} = t_p + t_q$, and t_r (right panel) based on 14 studied populations (COVID-19 epi-years, left panel) and 10 studied populations (2003 heat wave, right panel).

Granger Causality?

- Consider series of expected deaths, D_t^1, and series of excess deaths, D_t^2, separately.
- Predict total death counts after peak D_t in a lagged autoregressive model setting w.r.t. D_t^1 and D_t^2
- If magnitude and timing of the peak increase prediction accuracy, there is harvesting

Granger Causality?

Country	Wave 1	Wave 2	Wave 3
Bulgaria	Yes	Yes	No
Denmark	Yes	No	—
France	Yes	No	No
Netherlands	Yes	No	No
Norway	Yes	No	—
Sweden	Yes	No	No
USA	Yes	Yes	No

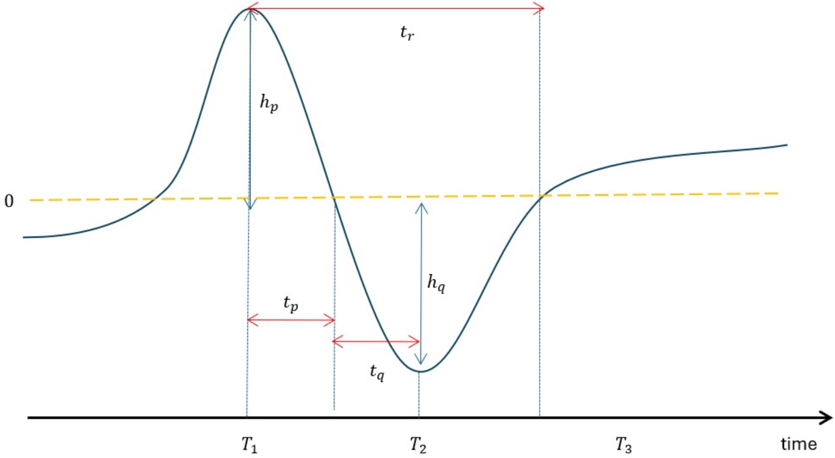
Further Steps: Use Individual Data

- Estimate shares of people in different states (Robust, Frail, Dead) prior to and after the shock
- Study what causes of death have been affected by COVID-19
- Study impact of COVID-19 on other subgroups (e.g., by residence)

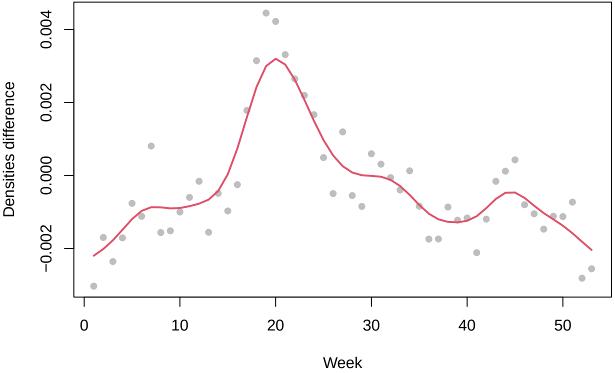
trim@sam.sdu.dk

FOUNDATION FOR SCIENCE

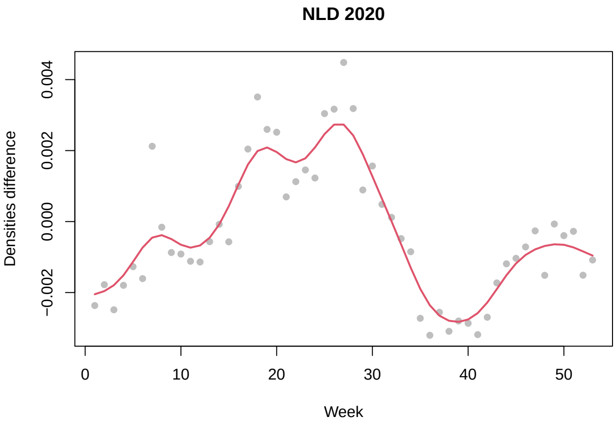
Harvesting: a simple visualization

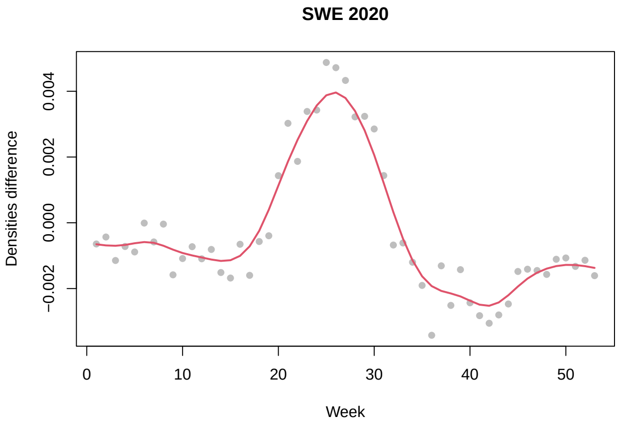


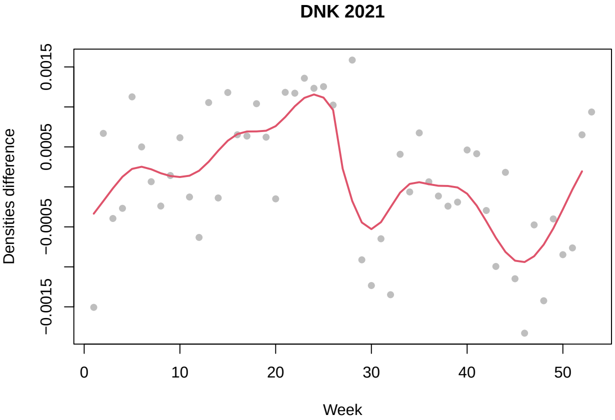
FRATNP 2020

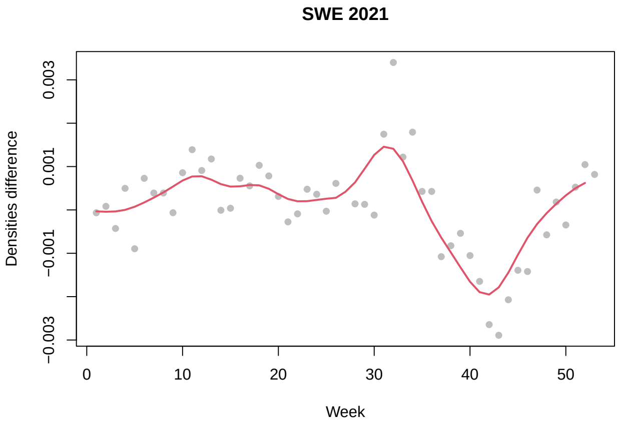


Densities difference
0.004
0.002
0.000
−0.002
0
10
20
30
40
50
Week


NLD 2020
Densities difference
0.004
0.002
0.000
−0.002
0
10
20
30
40
50
Week


SWE 2020
Densities difference
0.004
0.002
0.000
−0.002
0
10
20
30
40
50
Week


DNK 2021
Densities difference
0.0015
0.0005
−0.0005
−0.0015
0
10
20
30
40
50
Week


SWE 2021
Densities difference
0.003
0.001
-0.001
-0.003
0
10
20
30
40
50
Week