

## Håkon K. Gjessing

Professor/Principal Investigator

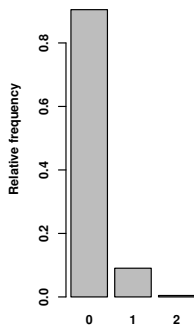
- Centre for Fertility and Health, Norwegian Institute of Public Health, Oslo
  - a Norwegian Research Council Centre of Excellence
- Department of Global Public Health and Primary Care, University of Bergen

---

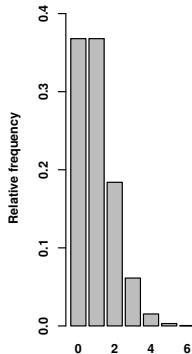
Kathmandu  
Tuesday, 6 Oct 2026

# POISSON DISTRIBUTION

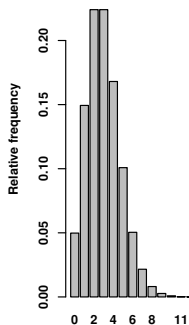
$\lambda = 0.1$



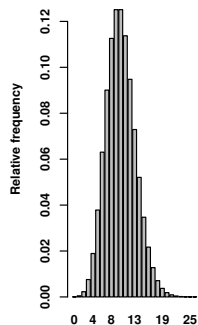
$\lambda = 1$



$\lambda = 3$



$\lambda = 10$

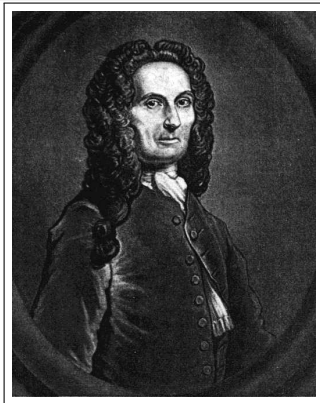


- Count data: 0, 1, 2, 3, .... no upper limit
- $\lambda$ : Both **mean** and **variance** of the Poisson distribution

# SIMÉON DENIS POISSON



- 1781-1840, France
- Contributions to
  - Statistics
  - Complex analysis
  - Partial differential equations
  - The calculus of variations
  - Analytical mechanics
  - Electricity and magnetism
  - Thermodynamics
  - Elasticity
  - Fluid mechanics



- 1667-1754, France, England
- Derived results on the Poisson distribution before Poisson



- 1941-, USA
- Professor at the University of Chicago
- Historian of Statistics
- Formulated  
Stigler's law of eponymy (1980)

- **Eponymy** = The practice of naming a discovery after its discoverer
- **Stigler's law of eponymy**  
A discovery is *never* named after its actual discoverer

## STIGLER'S LAW OF EPONYMY\*

Stephen M. Stigler  
Department of Statistics  
University of Chicago  
Chicago, Illinois 60637

No reader of Robert K. Merton's work on the reward system of science could fail to be struck by his insightful and engaging discussions of the role of eponymy in the social structure of science. The scientist should read (and reread) his 1957 address, "Priorities in Scientific Discovery,"<sup>1</sup> but for present purposes I must at least repeat his definition of eponymy, as "the practice of affixing the name of the scientist to all or part of what he has found, as with the Copernican system, Hooke's law, Plutarch's constant, or Halley's comet."<sup>2</sup> Merton went on to discuss three levels of a hierarchic order of eponymous practice: at the top there are a few men for whom an entire epoch is named, then comes a larger number of scientists designated as "father" of a particular science, and, finally, "thousands of eponymous laws, theories, theorems, hypotheses, instruments, constants, and distributions."<sup>3</sup> The present paper is an attempt by an Outsider to the sociology of science to shed some light on the workings



- 1941-, USA
- Professor at the University of Chicago
- Historian of Statistics
- Formulated  
**Stigler's law of eponymy** (1980)

- **Eponymy** = The practice of naming a discovery after its discoverer
- **Stigler's law of eponymy**  
A discovery is *never* named after its actual discoverer
- **Stigler's law of eponymy**  
was discovered by the sociologist Robert K. Merton (and others)
- ... and is thus self-confirming

## STIGLER'S LAW OF EPONYMY\*

Stephen M. Stigler  
Department of Statistics  
University of Chicago  
Chicago, Illinois 60637

No reader of Robert K. Merton's work on the reward system of science could fail to be struck by his insightful and engaging discussions of the role of eponymy in the social structure of science. The sociologist should read (and reread) his 1957 address, "Priorities in Scientific Discovery,"<sup>1</sup> but for present purposes I must at least repeat his definition of eponymy, as "the practice of affixing the name of the scientist to all or part of what he has found, as with the Copernican system, Hooke's law, Plutarch's constant, or Halley's comet."<sup>2</sup> Merton went on to discuss three levels of a hierarchic order of eponymous practice: at the top there are a few men for whom an entire epoch is named, then comes a larger number of scientists designated as "father" of a particular science, and, finally, "thousands of eponymous laws, theories, theorems, hypotheses, instruments, constants, and distributions."<sup>3</sup> The present paper is an attempt by an Outsider to the sociology of science to shed some light on the workings

# POISSON DISTRIBUTION

- Sometimes referred to as the “law of rare events” (but that’s not entirely precise)
- Typically appears when counting events that are rare, but where there are a vast number of “opportunities”
- ... and events happen more or less independently

Example:

- Number of patients admitted to an Intensive Care Unit during a week
- Traffic accidents in a city a given week

BUT:

- Infectious diseases (Covid 19)
- Traffic accidents involving buses, trains etc.

may lead to *overdispersion* if counting the number of victims.

... because events are not independent!

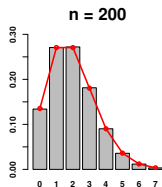
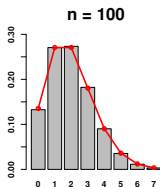
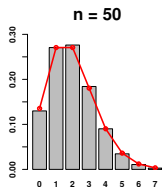
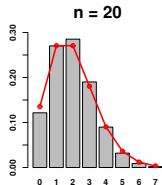
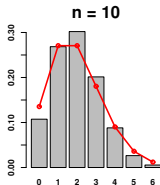
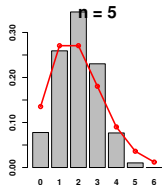
## Binomial:

- N possible events
- Select randomly, each with probability  $p$
- Count the number of selected,  $X \sim \text{Bin}(N, p)$

## Poisson:

- N is very large
- $p$  is very small
- ... but  $N \cdot p = \lambda$  is fixed as the mean value
- Count,  $X \sim \text{Pois}(\lambda)$

# BINOMIAL $\rightarrow$ POISSON, $\lambda = 2$ , N INCREASES, P DECREASES



Red line is Poisson

**Common requirement:**

$$n \geq 50$$

$$p \leq 0.05$$

# UNCERTAINTY, AND THE “RULE OF THREE”

## Uncertainty

If we observe, for instance,  $X = 64$ , then the uncertainty (SE) is  $\sqrt{64} = 8$

---

# UNCERTAINTY, AND THE “RULE OF THREE”

## Uncertainty

If we observe, for instance,  $X = 64$ , then the uncertainty (SE) is  $\sqrt{64} = 8$

---

## Rule of three

Binomial:

- $X \sim \text{Bin}(n, p)$  where  $n$  is fairly large
- Observe  $X = 0$
- Then  $[0, \frac{3}{n}]$  is a 95% Confidence Interval for  $p$

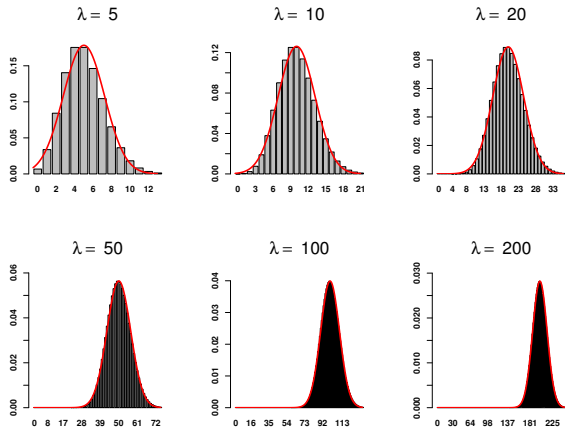
Poisson:

- $X \sim \text{Pois}(\lambda)$
- Observe  $X = 0$
- Then  $[0, 3]$  is a 95% Confidence Interval for  $\lambda$

# SOME “FUN” (AND IMPORTANT) PROPERTIES OF THE POISSON DISTRIBUTION

- **If:**  $X_1$  and  $X_2$  are independent Poisson variables with means  $\lambda_1$  and  $\lambda_2$ ,  
**Then:**  $X_1 + X_2$  has a Poisson distribution with mean  $\lambda_1 + \lambda_2$
- **If** you condition that  $X_1 + X_2 = N$ ,  
**Then**  $X_1 \sim \text{Bin}(N, p)$ , with  $p = \frac{\lambda_1}{\lambda_1 + \lambda_2}$
- **Thinning:**
  - $X \sim \text{Pois}(\lambda)$
  - You throw away randomly any of the counts in  $X$ , keep each with probability  $p$
  - The final result is then  $\text{Pois}(\lambda \cdot p)$

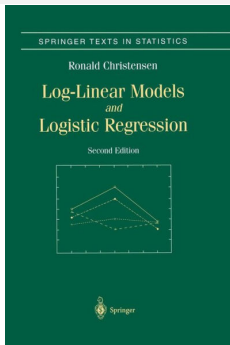
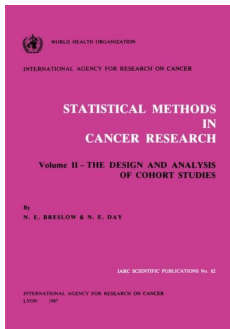
# POISSON $\rightarrow$ NORMAL, $\lambda$ INCREASES



**Common requirement:**

$$\lambda \geq 5 \text{ (or 10?)}$$

# (A FEW SELECTED) SAMPLES FROM HISTORY



## Software

- BMDP
- GLIM
- Epicure
- StatXact, LogXact

# POISSON REGRESSION

Count traffic accidents in an area in two consecutive weeks

- `count`: number of accidents counted
- `week`: week number 1 and 2
- `week01`: week number, coded 0 and 1
- `at.risk`: amount of traffic, for instance in units 1000 kilometers driven

| Data  |      |        |         |
|-------|------|--------|---------|
| count | week | week01 | at.risk |
| 50    | 1    | 0      | 2000    |
| 36    | 2    | 1      | 1000    |

- `week01` is coded as a standard “dummy” variable, slightly easier in interpretation than 1,2
- `at.risk` = 2000 means, for instance, 2 000 000 kilometers driven

# RATES

## Accident rates

- Week 1:  $\frac{50}{2000} = 0.025$  per 1000 kilometers
- Week 2:  $\frac{36}{1000} = 0.036$  per 1000 kilometers

## Rate Ratio

$$RR = \frac{0.036}{0.025} = 1.44$$

## Rate Difference

$$RD = 0.036 - 0.025 = 0.011 \text{ per 1000 kilometers}$$

# POISSON REGRESSION, MODEL

To begin with: ignore `at.risk`

## GLM (Generalized Linear Model)

- Family **Poisson**
- Link function **log**, i.e.  $\log(\lambda) = \beta_0 + \beta_1 \cdot x_1 + \dots$

**Model:**

$$\text{week 1:} \quad \lambda_1 = \exp(\beta_0)$$

$$\text{week 2:} \quad \lambda_2 = \exp(\beta_0 + \beta_1) = \exp(\beta_0) \cdot \exp(\beta_1)$$

$$\text{Rate Ratio RR:} \quad \frac{\lambda_2}{\lambda_1} = \exp(\beta_1)$$

**R model:**

```
res <- glm(count ~ week01, family = poisson, data = .data)
```

# POISSON REGRESSION, RESULTS

## Estimates and confidence intervals on log scale:

NOTE! In Stata there is the **eform** option that gives you the exponentiated values

```
> summary(res)
```

Coefficients:

|             | Estimate | Std. Error | z value | Pr(> z )   |
|-------------|----------|------------|---------|------------|
| (Intercept) | 3.9120   | 0.1414     | 27.662  | <2e-16 *** |
| week01      | -0.3285  | 0.2186     | -1.503  | 0.133      |

```
> confint(res)
```

|             | 2.5 %      | 97.5 %     |
|-------------|------------|------------|
| (Intercept) | 3.6214057  | 4.17693452 |
| week01      | -0.7635849 | 0.09655906 |

# POISSON REGRESSION, RESULTS

Estimates and confidence intervals on a multiplicative scale:

```
> exp(res$coefficients)
```

|             |        |
|-------------|--------|
| (Intercept) | week01 |
| 50.00       | 0.72   |

```
> exp(confint(res))
```

|             |            |           |
|-------------|------------|-----------|
|             | 2.5 %      | 97.5 %    |
| (Intercept) | 37.3900896 | 65.165782 |
| week01      | 0.4659929  | 1.101375  |

Interpretation:

- Estimate for (Intercept) is simply the count in week 1 (when we use week01)
- Estimate for week01 is the ratio week 2/week 1 =  $36/50 = 0.72$ , i.e. expected count in week 2 is 72% of week 1
- Count in week 2 can be found as  $50 \times 0.72 = 36$
- **Saturated model** number of parameters same as number of observations
- In a saturated model, model expected numbers  $\lambda$  are the same as the observed counts

# POISSON REGRESSION, MODEL

Now, **do it correctly** and take `at.risk` into account

**Model:**

$$\text{week 1: } \lambda_1/2000 = \exp(\beta_0)$$

$$\text{week 2: } \lambda_2/1000 = \exp(\beta_0 + \beta_1)$$

**Model:**

$$\text{week 1: } \lambda_1 = 2000 \times \exp(\beta_0)$$

$$\text{week 2: } \lambda_2 = 1000 \times \exp(\beta_0 + \beta_1)$$

**Model, on log scale:**

$$\text{week 1: } \log(\lambda_1) = \log(2000) + \beta_0$$

$$\text{week 2: } \log(\lambda_2) = \log(1000) + \beta_0 + \beta_1$$

- The values  $\log(2000)$  and  $\log(1000)$  have no parameter to be estimated... known as an “offset”
- $\beta_0$  is still an intercept to be estimated, but same for all (and “offset” by the value of  $\log(\text{at.risk})$ )

## POISSON REGRESSION, MODEL

The offset has two effects:

- Balance the two groups correctly against one another, according to `at.risk`
- Give the intercept  $\beta_0$  a more relevant interpretation in terms of expected rates, not expected numbers

**Again, a GLM with Poisson family, log link, but with offset!**

**R model:**

```
> res <- glm(count ~ week01 + offset(log(at.risk)),  
  family = poisson, data = .data)
```

NOTE that Stata can use either an [offset](#) or an [exposure](#). The latter is just 2000, 1000.... thus it needs no log, and the output may be easier to read.

# POISSON REGRESSION, RESULTS

## Estimates and confidence intervals on log scale:

```
> summary(res)
```

Coefficients:

|             | Estimate | Std. Error | z value | Pr(> z )   |
|-------------|----------|------------|---------|------------|
| (Intercept) | -3.6889  | 0.1414     | -26.084 | <2e-16 *** |
| week01      | 0.3646   | 0.2186     | 1.668   | 0.0953 .   |

```
> confint(res)
```

|             | 2.5 %       | 97.5 %     |
|-------------|-------------|------------|
| (Intercept) | -3.97949677 | -3.4239679 |
| week01      | -0.07043773 | 0.7897062  |

# POISSON REGRESSION, RESULTS

Estimates and confidence intervals on multiplicative scale:

```
> exp(res$coefficients)
```

|             |        |
|-------------|--------|
| (Intercept) | week01 |
| 0.025       | 1.440  |

```
> exp(confint(res))
```

|             |            |            |
|-------------|------------|------------|
|             | 2.5 %      | 97.5 %     |
| (Intercept) | 0.01869504 | 0.03258289 |
| week01      | 0.93198577 | 2.20274925 |

**Interpretation:**

- Estimate for (Intercept) is the rate in week 1, i.e. count in week 1 per number at .risk = 50/2000
- Estimate for week01 is the rate ratio

$$\frac{\# \text{ week 2/1000}}{\# \text{ week 1/2000}} = \frac{36/1000}{50/2000} = 1.44,$$

i.e. expected count in week 2 (per number at .risk)  
is 44% higher than in week 1

# WEIGHTS...

How about the following “trick”?

|       |      |        |         | Data  |
|-------|------|--------|---------|-------|
| count | week | week01 | at.risk | rate  |
| 50    | 1    | 0      | 2000    | 0.025 |
| 36    | 2    | 1      | 1000    | 0.036 |

```
> res <- glm(rate ~ week01,  
  family = poisson, weights = at.risk, data = .data)
```

... produces a warning:

Warning messages:

```
1: In dpois(y, mu, log = TRUE) : non-integer x = 0.025000  
2: In dpois(y, mu, log = TRUE) : non-integer x = 0.036000
```

... but the result is identical to the offset result:

|             | Estimate | Std. Error | z value | Pr(> z )   |
|-------------|----------|------------|---------|------------|
| (Intercept) | -3.6889  | 0.1414     | -26.084 | <2e-16 *** |
| week01      | 0.3646   | 0.2186     | 1.668   | 0.0953 .   |

# WEIGHTS...

... and then additive?

| count | week | week01 | at.risk | rate  |
|-------|------|--------|---------|-------|
| 50    | 1    | 0      | 2000    | 0.025 |
| 36    | 2    | 1      | 1000    | 0.036 |

```
> res <- glm(rate ~ week01,  
  family = poisson(link = identity), weights = at.risk, data = .data)
```

... produces a warning:

Warning messages:

```
1: In dpois(y, mu, log = TRUE) : non-integer x = 0.025000  
2: In dpois(y, mu, log = TRUE) : non-integer x = 0.036000
```

... but provides correct results for an additive scale!:

Coefficients:

|             | Estimate | Std. Error | z value | Pr(> z )     |
|-------------|----------|------------|---------|--------------|
| (Intercept) | 0.025000 | 0.003536   | 7.071   | 1.54e-12 *** |
| week01      | 0.011000 | 0.006964   | 1.580   | 0.114        |

## ... JUST BE CAREFUL WITH WEIGHTS, THOUGH

- Different software has different implementations of weights (four types in Stata)
- A manual check is ALWAYS a good idea!  
Check BOTH of
  - point estimate
  - standard error

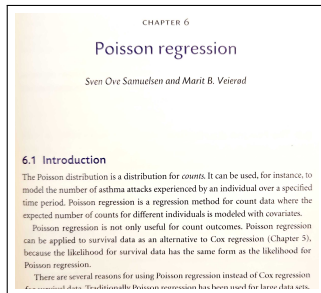
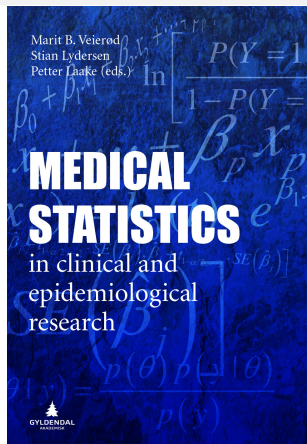
### Rate Difference

$$RD = 0.036 - 0.025 = 0.011 \text{ per 1000 kilometers}$$

Usual formula for Standard Error

$$SE(RD) = \sqrt{\frac{50}{2000^2} + \frac{36}{1000^2}} = 0.006964$$

OK!



# “REGULAR” LINEAR (GAUSSIAN) REGRESSION FOR POISSON OUTCOMES?

Does it make sense?

## Poisson:

- For a “true” Poisson distribution, mean = variance, so there is only one parameter
- A “true” Poisson regression model knows that mean = variance, and can use this to extract more information
- In the previous example, the Poisson model can do estimation from only two groups
- BUT if there is overdispersion, care must be taken to include it!

## Linear:

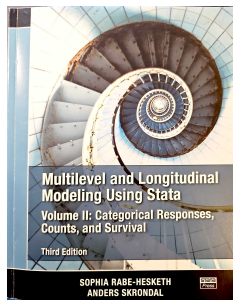
- Sometimes faster
- A linear regression model can be told that mean = variance through an assumption of heteroscedasticity
- It is still just an approximation
- However, it usually includes a residual variance that partly corrects for overdispersion

# OVERDISPERSION, POSSIBLE SOLUTIONS

- 1 Include a single overdispersion parameter
- 2 Assume that the distribution is a mixture of many Poisson distributions with different  $\lambda$ 's

Often ends up with a *Negative Binomial* distribution

- 3 Mixed effects models for Poisson outcomes



# POISSON REGRESSION FOR SURVIVAL ANALYSIS

- Create time intervals, for instance weeks
- Aggregate (count) events in each interval
- Also count – i.e. keep track of – the number “at risk” in each interval
- Update time-dependent covariates for each interval
- Then: Run Poisson regression on the event counts, with appropriate offset etc.

## Some benefits:

- Aggregation can provide large speed gains, particularly when
  - The number of covariate categories are moderate
  - The “raw” file contains lots of individuals
  - Particularly with time-dependent covariates
- (Perhaps) more natural when including multiple time scales in the same analysis, e.g. age and calendar time
- Easy to estimate rates from interval to interval, i.e. more in line with “traditional” epidemiological thinking

# TASK: ASTRAZENECA VACCINE, WEEK 14, 2021

- **In Norway:** 4 deaths had been confirmed as side effects among 134,000 vaccinated with the AstraZeneca vaccine

Risk:  $4/134,000 \approx 3$  out of 100,000

- If the vaccine is removed from the program, this will cause delays in vaccination
- Delays will cause further deaths from infections
- Continuing with the vaccine may cause more deaths as side effects

---

## Question:

To an individual, what is most/least dangerous:

- 1 Take the vaccine
  - 2 Wait until a safer vaccine is available, but then risk dying from Covid-19 in the meantime
- 
- Should be considered in light of the current situation of infections in Norway at that time
  - Should take into consideration differences in mortality rates between women and men, and across different age groups

## Data source: BeredtC19

- Several Norwegian registries linked together
- Exclusively to answer questions related to the pandemic
- Infections (positive Covid-19 PCR-test)
- What vaccines received, and when
  - Pfizer (mRNA vaccine)
  - Moderna (mRNA vaccine)
  - AstraZeneca (adeno-vector)

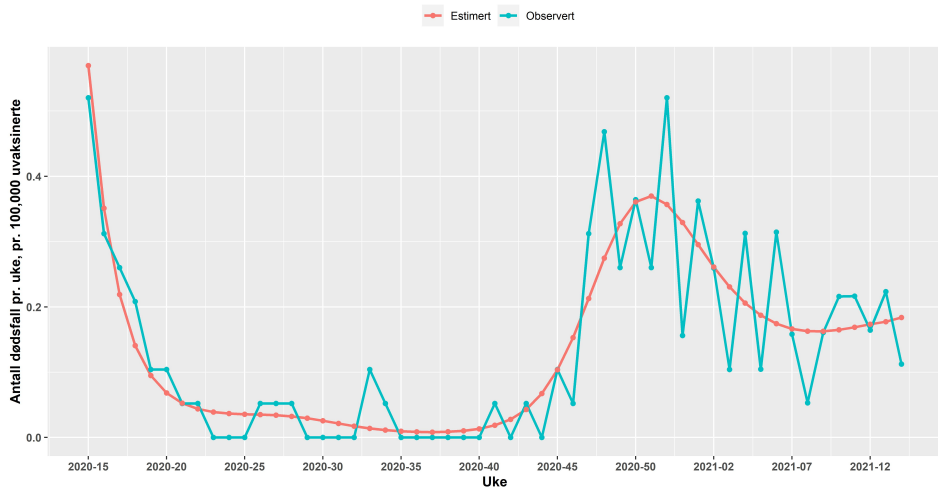
## Outcomes

- 1 Hospital admissions related to Covid-19
- 2 ICU admissions related to Covid-19
- 3 Deaths related to Covid-19

## Survival analysis

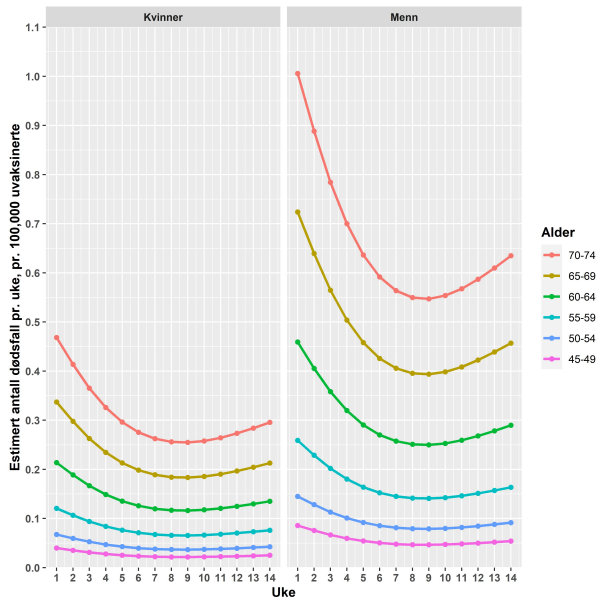
- Outcomes: Hospitalization, ICU admission, Death
- Data in aggregated form are easier to obtain and work on
- Poisson regression with splines, GAM models
- Covariates: age, sex and others
- “Censored”, i.e. protected, after vaccination
- Infections are common, deaths are rare

# MORTALITY RATES ACROSS PANDEMIC, WEEK BY WEEK



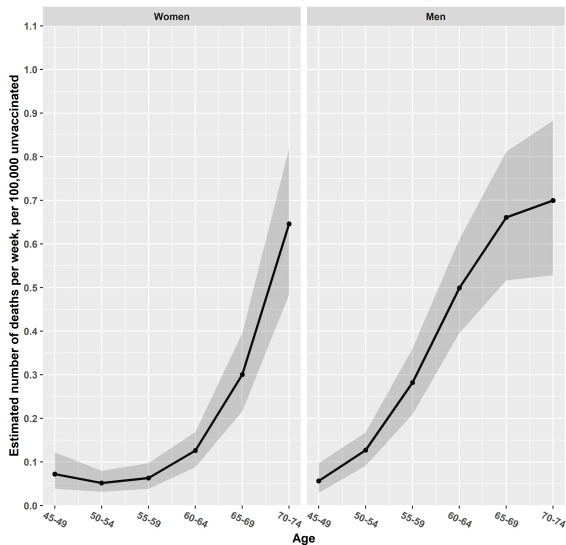
(Aggregated over selected age groups)

# MORTALITY RATES IN AGE CATEGORIES, WEEK BY WEEK, 2021



# MORTALITY RATES IN AGE CATEGORIES, WITH INTERACTION

Estimated rate of death related to COVID-19, across weeks 11-13, 2021



(Aggregated over weeks 11-13, 2021)

# Thank you!