Introduction to Quantile Regression

Matteo Bottai, Sc.D.

Karolinska Institutet

BIOSTATISTICS
IMM Karolinska Institutet

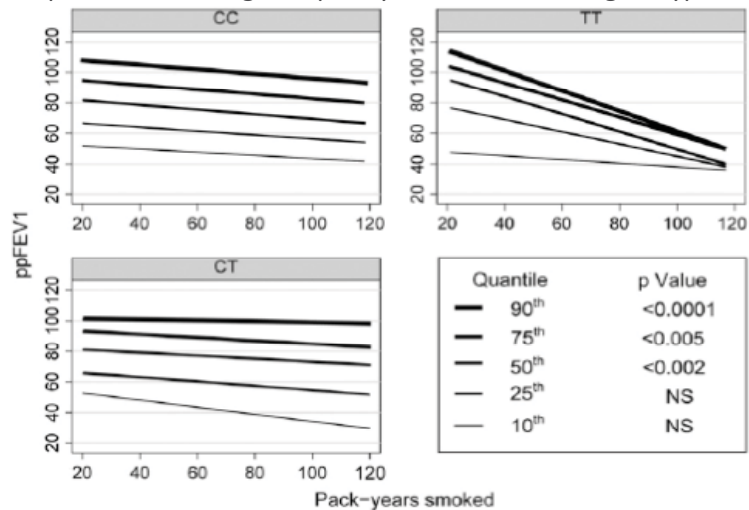## Motivating Example 1: Interest in Percentiles

Median and 5th percentile of post-bronchodilator forced vital capacity against age

Johannessen et al, *Am J Respir Critical Care Med*, 2006

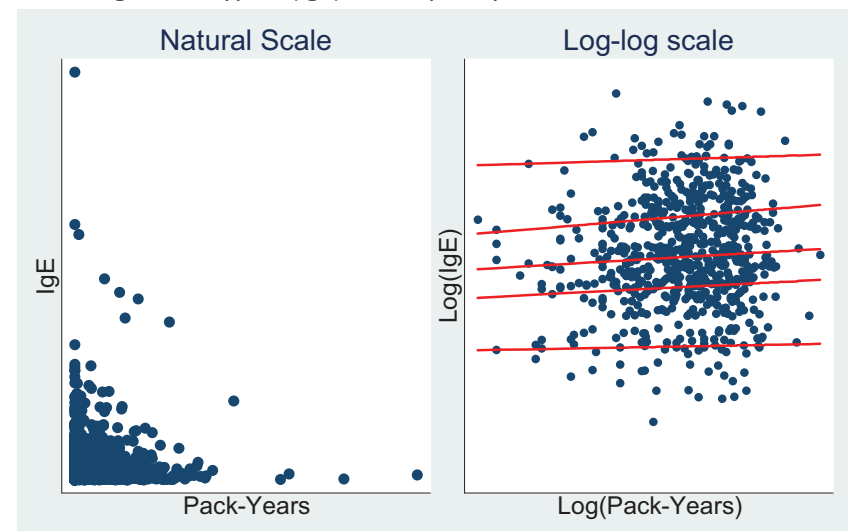## Motivating Example 2: Percentiles Provide Insight

Percentiles of predicted FEV1 against pack-years for –1112C/T genotypes of IL13 gene

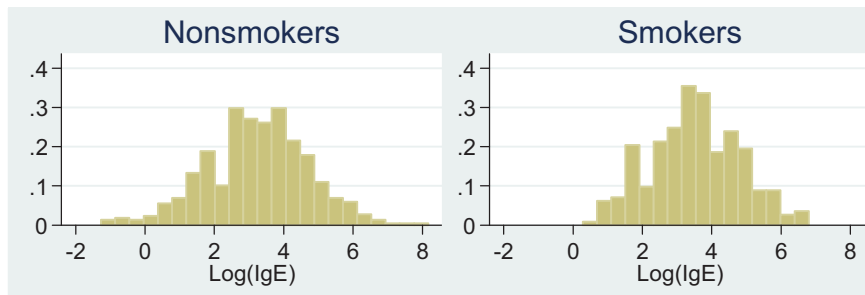| Quantile	p Value
90th	<0.0001
75th	<0.005
50th	<0.002
25th	NS
10th	NS

Sadeghnejad et al, *Am J Respir Critical Care Med*, 2007

Motivating Example 3: Percentiles Allow Outcome Transformation

Serum immunoglobulin type E (IgE) versus pack-years

Natural Scale — Log-log scale

Motivating Example 4: Transforming the Outcome

Differences between groups in mean IgE and mean log(IgE).

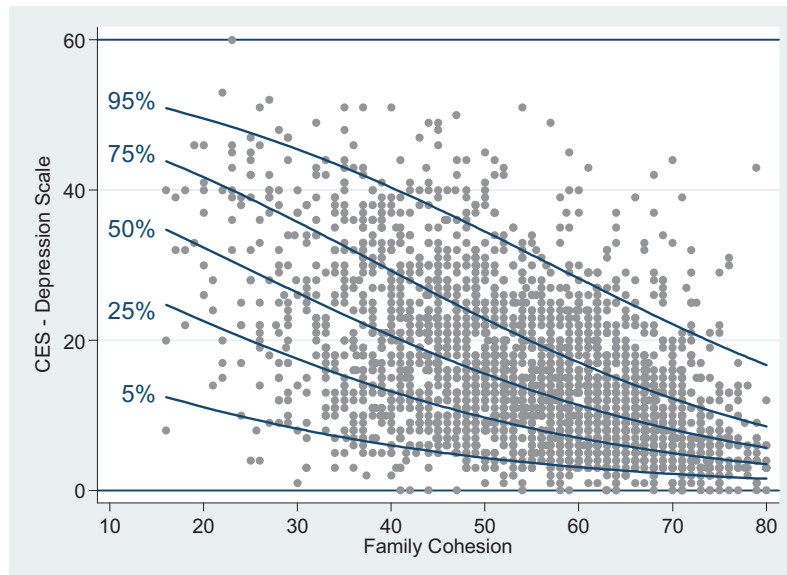|	IgE	Log(IgE)
Smokers vs. Nonsmokers	1.06	0.26
P-value	0.94	0.01

Percentiles are invariant to monotone transformations.

Motivating Example 5: Logistic Quantile Regression for Bounded Outcomes

The Center for Epidemiologic Studies depression scale ranges from 0 to 60. Its distribution changes with family cohesion in shape and location.
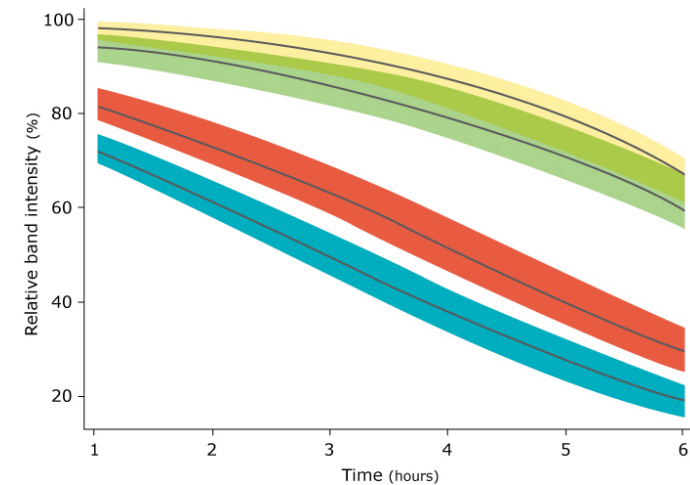
Since the outcome is bounded, we apply a logit transform.
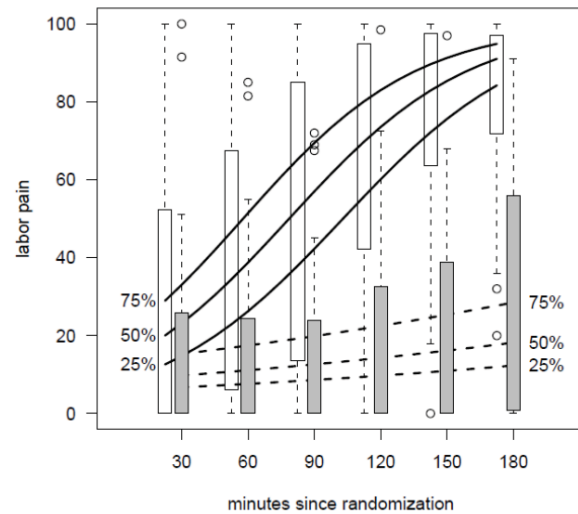

Bottai et al, *Stat Med,* 2010

Motivating Example 6: Logistic Quantile Regression for Bounded Outcomes

Quartiles of fibrin in pulmonary hypertension other than CTEPH, chronic thromboembolic pulmonary hypertension, pulmonary embolism, and control group

Miniati et al, *Am J Respir Criti Care Med*, 2010

Motivating Example 7: Logistic Quantile Regression with Longitudinal Data

Labor pain in a medication and a placebo group

Liu and Bottai, *Int J Biostat*, 2009
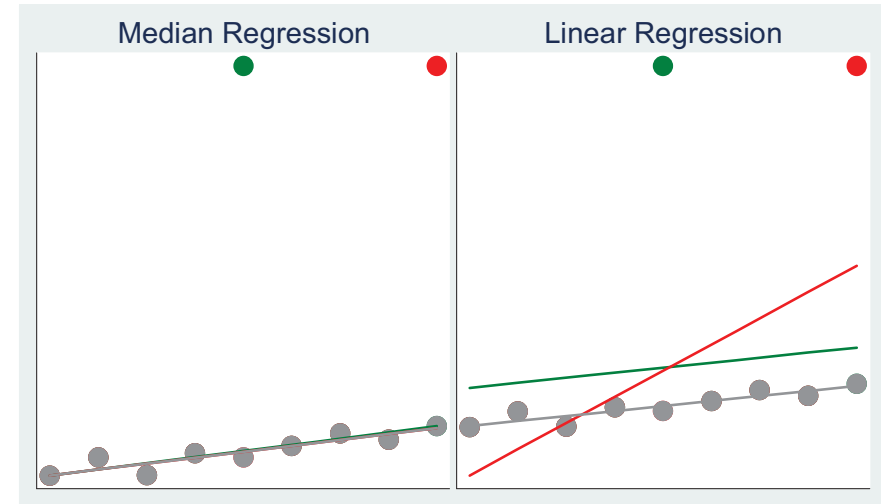
Motivating Example 8: Percentiles Are Robust to Outliers

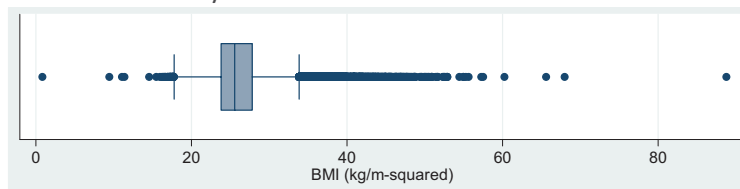Three scenarios: no outliers (●), one in the middle (●), one at the extreme (●)

Bottai et al, *Eur Respir J*, 2002

Motivating Example 9: Percentiles Are Robust to Outliers

Body mass index in 25 to 75 year old men

	Median Regression		Linear Regression	
	All sample	14<BMI<60	All sample	14<BMI<60
	n = 90,244	n = 90,232	n = 90,244	n = 90,232
Age	0.802	0.801	0.316	0.257
P-value	0.000	0.000	0.367	0.035

Even with a large sample removing outliers changes our inference on the mean. Inference on the median is nearly unaffected.

Motivating Example 10: Percentiles Can Be Efficient

Inference on quantiles can be very efficient.

Suppose we test differences in IgE between smokers and non-smokers.

These are the bootstrap estimated standard errors.

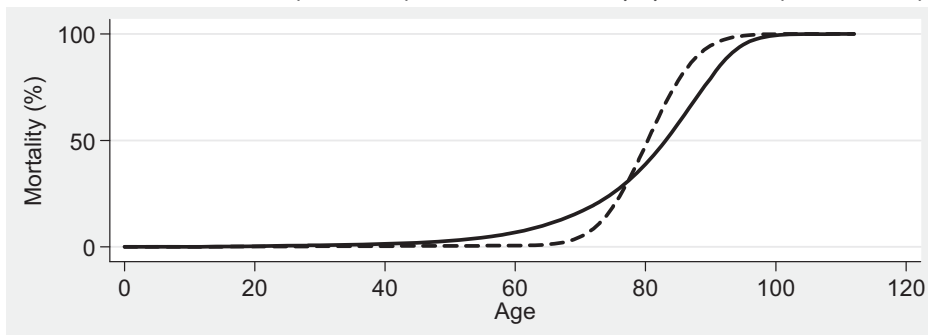| Difference	Standard Error	Ratio
Mean	13.00	–
0.01 quantile	0.53	24.5
0.10 quantile	0.64	20.3
0.50 quantile	3.04	4.3
0.75 quantile	9.57	1.4
0.90 quantile	34.00	0.4

To get the same power as 0.10-quantile regression with 100 observations, linear regression (or t-test) needs about 40,000 observations.

Motivating Example 11: Percentiles of Survival

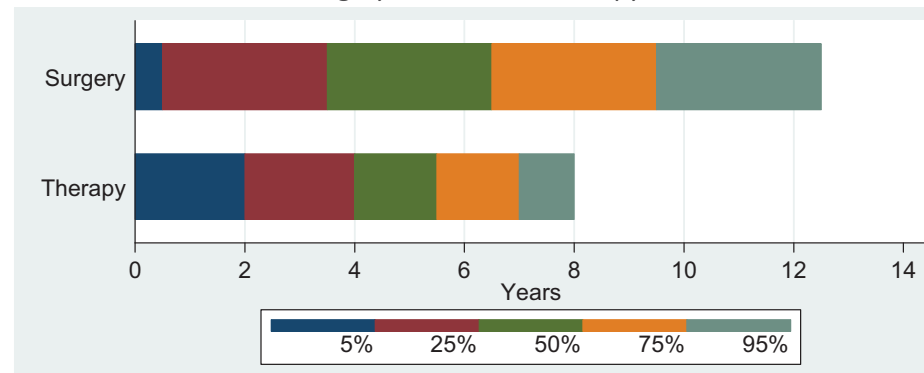Sweden from 2006 to 2010 (solid line) and in a fictitious population A (dashed line)

	Life Expectancy	Lifespan	Mortality Percentiles		
			5%	50%	95%
Sweden	80.6	112.0	56.0	84.0	96.0
Population A	80.6	112.0	70.7	80.8	90.8

Sweden has more early deaths and longer survivors than Population A.

Motivating Example 12: Laplace Regression with Censored Data

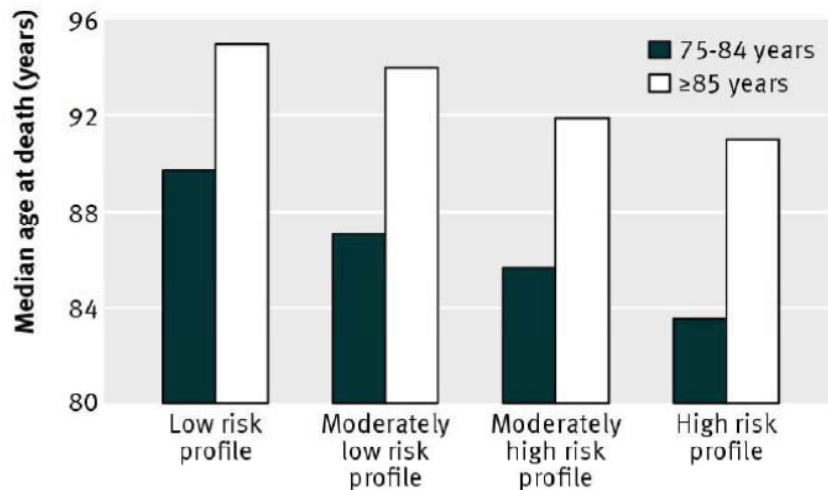Percentiles of survival after surgery and standard therapy

Patients respond to surgery differently.
The frailest 5% die within 6 months but the strongest 25% live at least 9.5 years.

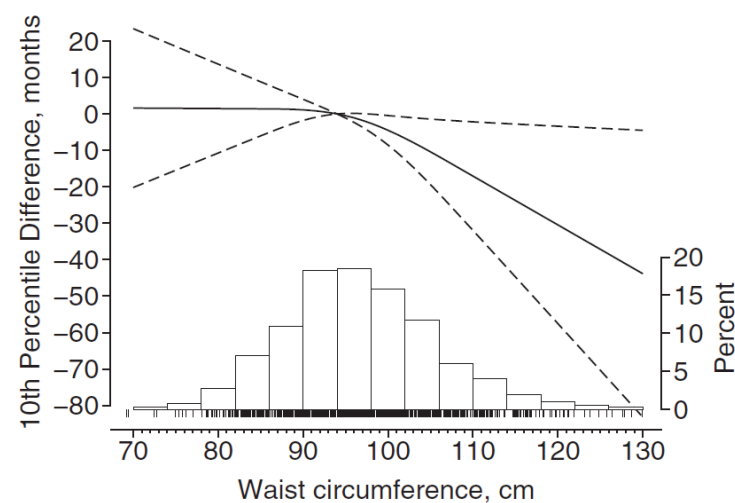## Motivating Example 13: Laplace Regression with Censored Data

A population-based study on lifestyle and survival
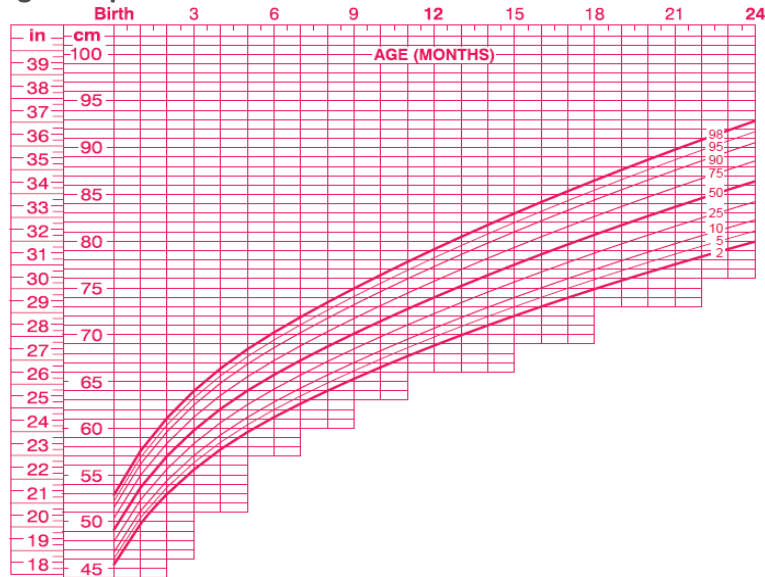

Rizzuto et al, *BMJ*, 2012

Motivating Example 14: Laplace Regression with Censored Data

A population-based study on waist circumference and survival

Orsini et al, *Epidemiology*, 2012

Motivating Example 15: Growth Chart for Girls

World Health Organization Child Growth Standards

Do We Think Percentiles?

Why are percentiles of children's weight more interesting than kilograms?

How do we define tall/short, fast/slow, and rich/poor?

The actual inches, miles per hour, or dollars a year, are not as relevant as their corresponding percentiles in a referent population.

Summary

Quantiles
 – can be of research interest
 – provide a thorough assessment of continuous variables of interest
 – are invariant to monotone transformations
 – may be efficient and robust to outliers
 – can be estimated with popular statistical programs (e.g. Sas, Stata, R)

Few relevant references
 – Count data (Machado and Santos-Silva, *J Am Stat Ass,* 2005)
 – Longitudinal data (Geraci and Bottai, *Biostatistics,* 2007)
 – Laplace regression with censored data (Bottai and Zhang, *Biometical J,* 2008)
 – Bounded outcomes (Bottai et al, *Statistics in Medicine,* 2009)
 – Bayesian approaches (Yu and Moyeed, *Stat Prob Letters*, 2001)

More information can be found at www.imm.ki.se/biostatistics.