Diagnosing Change in a Sparse, Bursty Variable

Eh in Pākehā English

SIMON TODD

Background

Eh in New Zealand

Eh is a tag (like *you know*)

- Found at end of clauses
- Non-propositional effect on meaning

 "they just weren't cool at all in Europe eh"

 "I wouldn't let on eh but they were really accurate"

Typical functions:

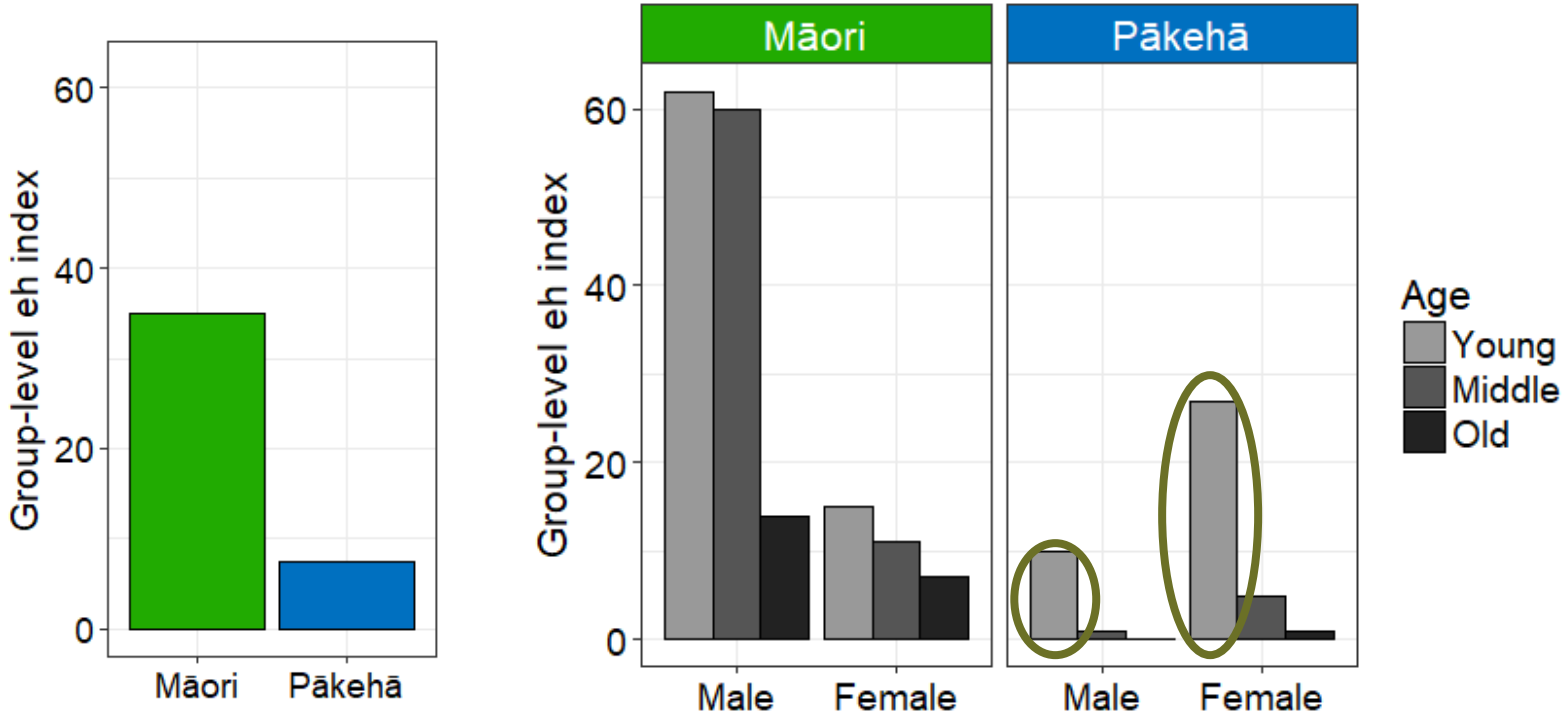
- Mark something as noteworthy
- Engender / reflect shared knowledge, beliefs, & values

(Meyerhoff, 1992)

Eh and ethnicity

Additional function: in-group marker of Māori

- 60 working-class speakers, Porirua City, 1989-90



Why do some young Pākehā use *eh*?

(Meyerhoff, 1992, 1994)

Eh among young Pākehā

Apparent time hypothesis → change

- All Pākehā are adopting *eh* (from Māori)
- *Eh* is a feature of NZ English

Alternative → age-grading

- Young speakers use *eh*, older speakers don't
- *Eh* is a feature of youth speech

Previous studies claim evidence for change

… but there are some sticking points

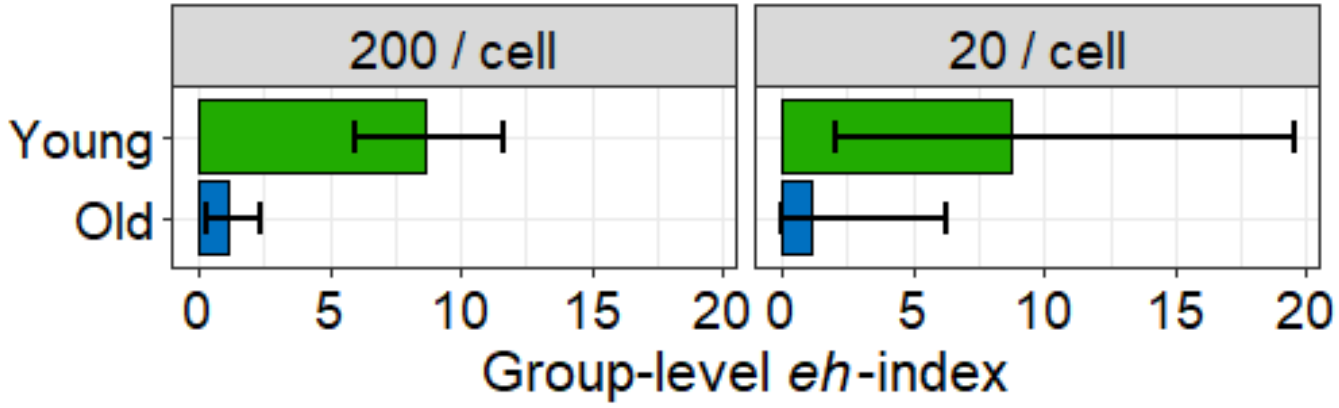
Te Kāhui Roro Reo • New Zealand Institute of Language Brain & Behaviour

Sticking point 1: Representativeness

Meyerhoff (1994):

- Young female pattern → driven by living arrangement
- Excluding extreme speaker → reduce group *eh* by 4x

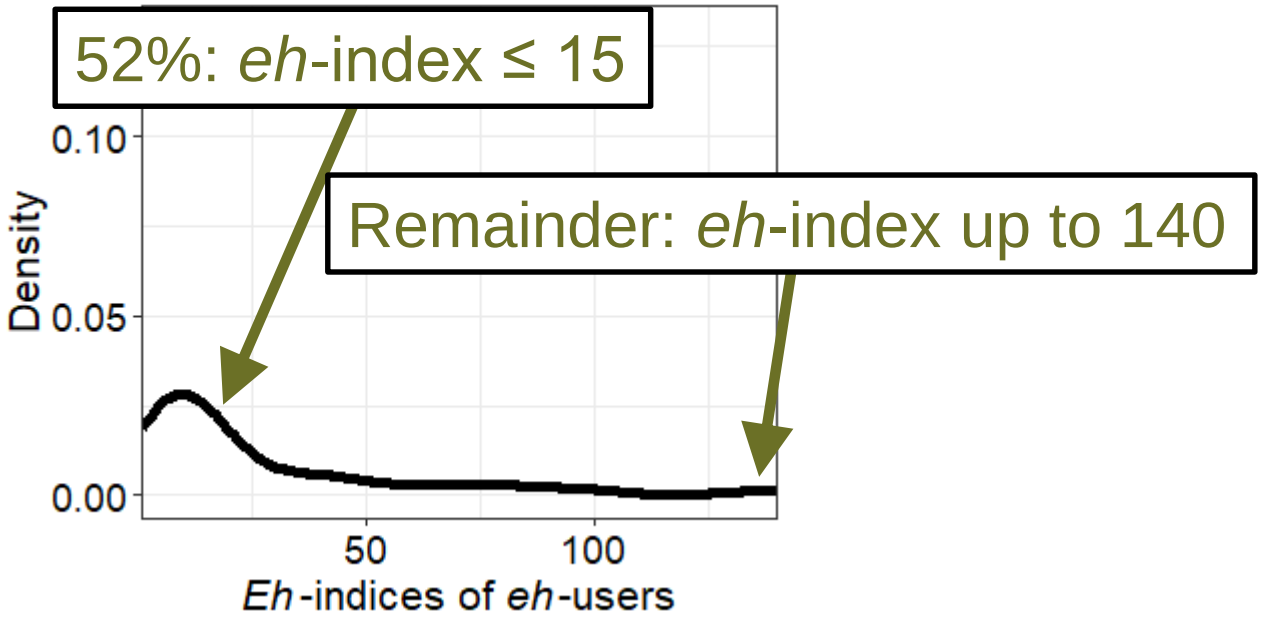
Small sample size → less representative



Solution: use a big corpus

Sticking point 2: Burstiness

A few people use *eh* to extreme degrees



Problem: can skew group-level averages

Need: a way to downweight extreme users

Te Kāhui Roro Reo • New Zealand Institute of Language Brain & Behaviour

Sticking point 3: Sparsity

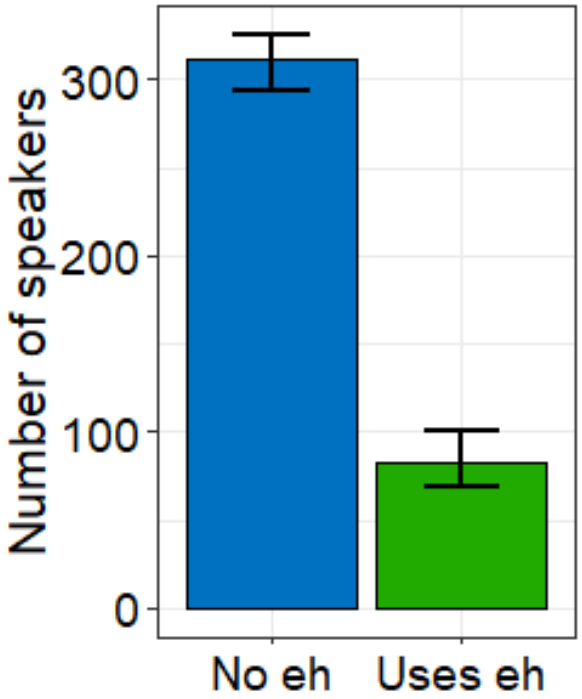
Many people don't use *eh* at all during a recording

ONZE (contemporary):

79% → no *eh* in interview

Do these speakers:

- not use *eh* at all?
- not use *eh* enough to appear?



Need: a way to tell (while controlling for burstiness)

Diagnosing population-level change

How can we measure *eh*-adoption?

"Adoption" ≈ adding to production repertoire

- Distinct from differences in *usage*

Measurement must:

- Disentangle repertoire from usage
- Address burstiness and sparsity

Solution: new addition to quantitative toolbox

Zero-inflated negative binomial regression

Te Kāhui Roro Reo • New Zealand Institute of Language Brain & Behaviour

Zero-inflated negative binomial regression

Addresses burstiness: extreme users

Extension of Poisson regression:

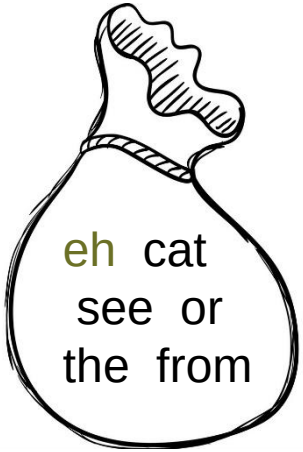
How many times do you use *eh*?

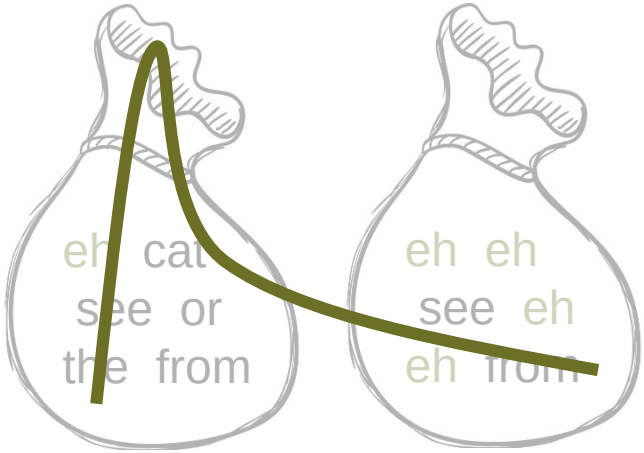
Poisson: each individual draws from same "bag" of words

NegBin: each individual may draw from a different "bag"

Distribution over "bags"

→ extreme users downweighted





Zero-inflated negative binomial regression

Addresses sparseness:

what does “zero” mean?

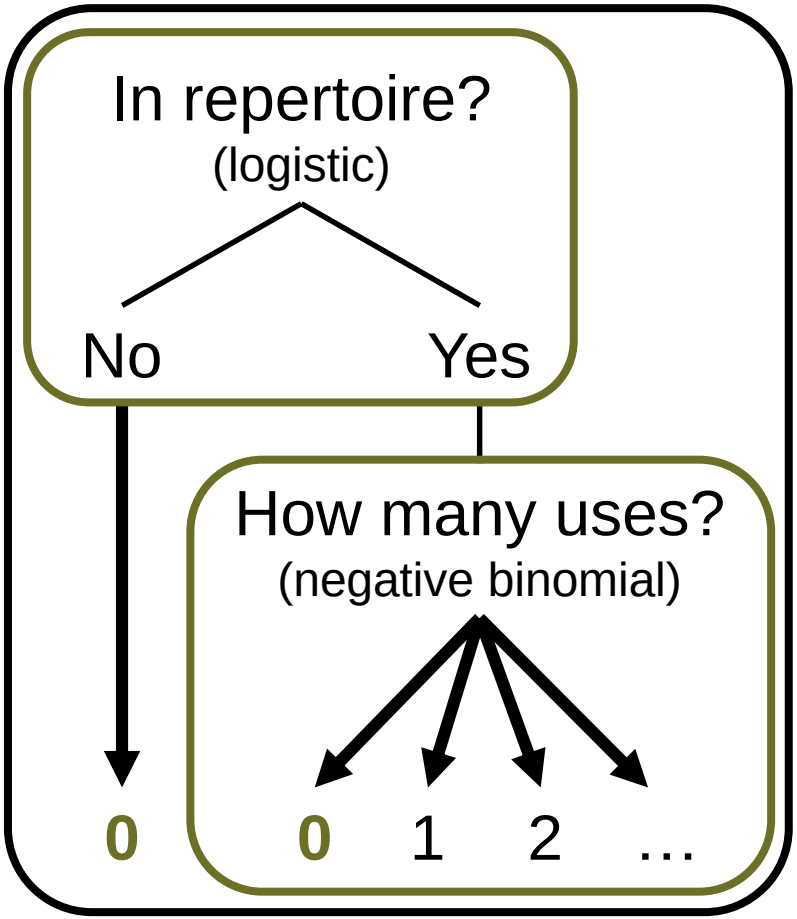
Two-component model

1. In production repertoire?
2. How many uses on this occasion?

Components balanced:

expected number of uses

~ P(0 ↔ not in repertoire)



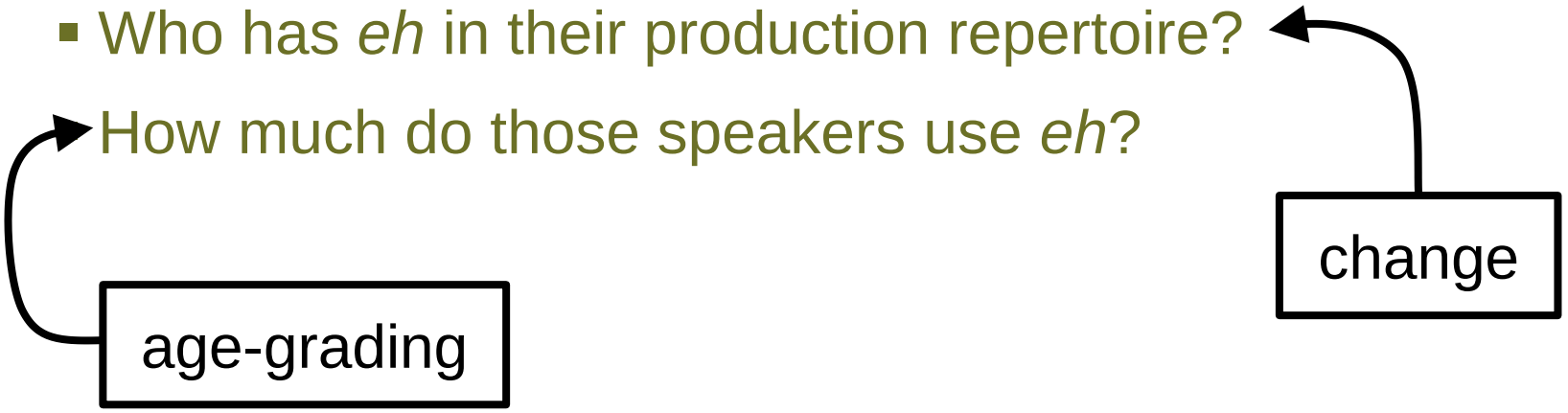
Predictions: Pākehā *eh* adoption

"Adoption" ≈ adding to production repertoire

- Distinct from differences in *usage*

Two separate questions:

- Who has *eh* in their production repertoire?
- How much do those speakers use *eh*?



Predictions: Change

Apparent-time → *eh* more in repertoire of young Pākehā

Basic diffusion model (Māori → Pākehā):

- Change spreads through interaction
- More interaction with *eh*-users → more likely to adopt
- Assume: demographic similarity → more interactions

Māori *eh*-users over-represented in working class
→ *eh* more in repertoire of working-class Pākehā

Male Māori highest *eh*-users
→ *eh* more in repertoire of male Pākehā

Predictions: Age-grading

Only consider speakers with *eh* in their repertoire

Fundamental idea → *eh* used more by young Pākehā

- Above and beyond other socio-demographic factors

Testing the predictions

Data

ONZE Corpus (contemporary component)

394 speakers, born 1926-87, interviewed 1994-2007

Demographic categories:

- Age: young (20-30) / old (45-60)
- Class: working / middle
- Sex: male / female

~50 speakers per cell

Testing predictions: Pākehā *eh* adoption

ZINB component 1

Who has *eh* in their production repertoire?

- Change: age: young > old
  class: working > middle
  sex: male > female

ZINB component 2

How much do those speakers use *eh*?

- Age-grading: young > old

Te Kāhui Roro Reo • New Zealand Institute of Language Brain & Behaviour

Who has *eh* in their production repertoire?

✓Age ***

young > old

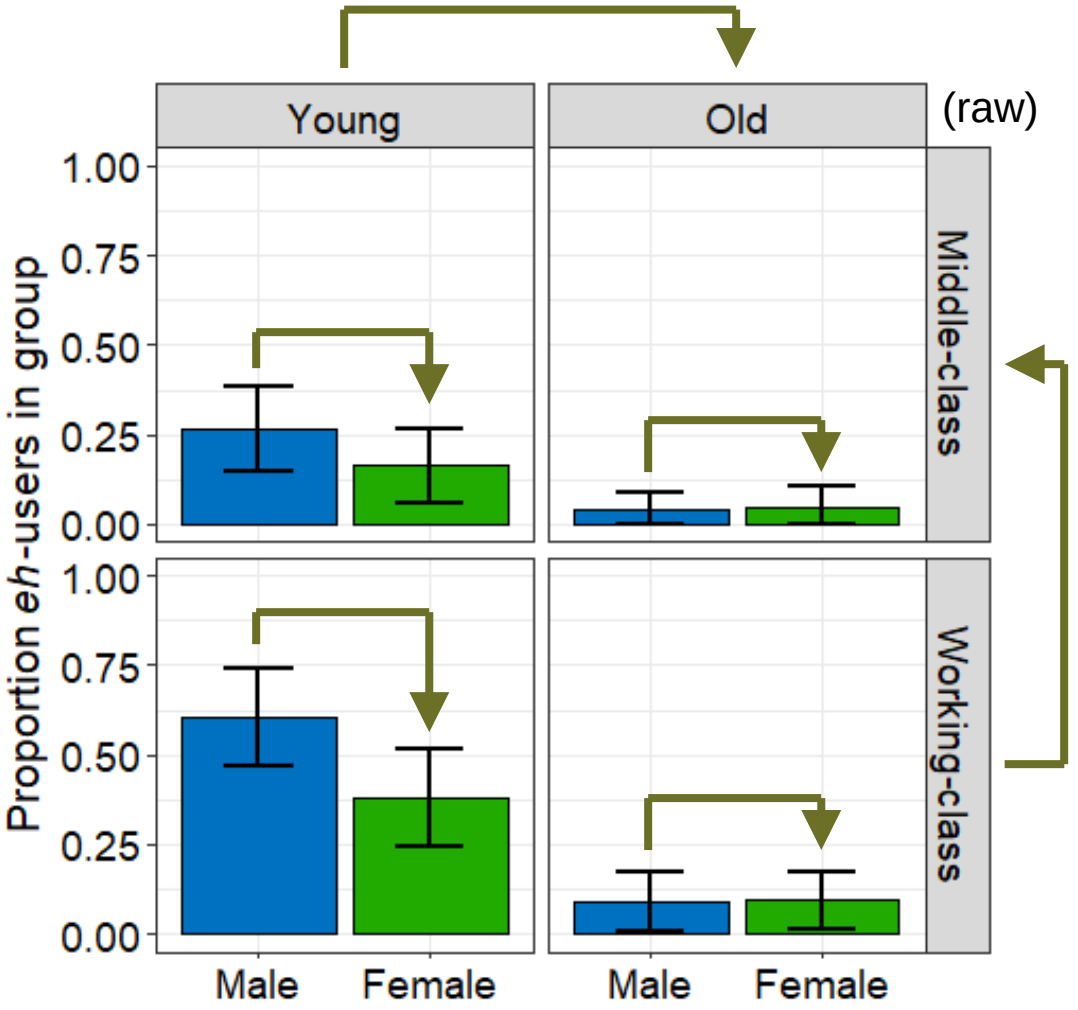
✓Class *

working > middle

✓Sex ·

male ≥ female

→ supports change

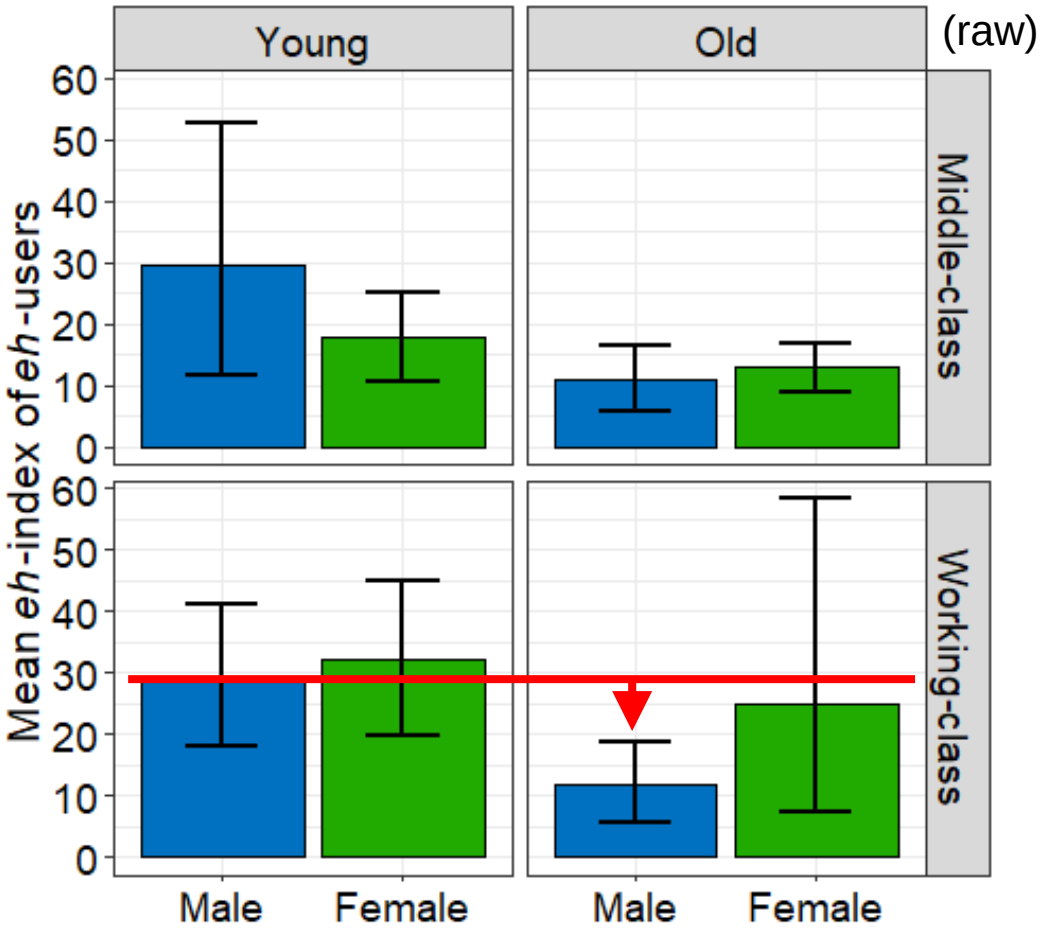


How much do those speakers use *eh*?

✘ Usages all similar

? Only difference:
old males
< everyone else

→ <u>little</u> support for age-grading



Weighing up the hypotheses

ZINB component 1: supports change

ZINB component 2: little support for age-grading

Model contributions: component 1 > component 2

change > age-grading

More support: prev. generation (b. 1890-1930; rec. 1990)

- Only 1/87 “uses” *eh* (metalinguistically)

Why adopt eh? Attitudinal change

Thank you for listening!

Questions / comments?

simon.todd@canterbury.ac.nz