Experian plc, 80 Victoria Street, London

The finer points of postcode mortality modelling

Stephen J. Richards, BSc, FFA, PhD
9th October 2013

Contents

1. About the speaker

1. About the speaker

- Consultant on longevity risk since 2005
- Founded longevity-related software businesses in 2006:

LONGEVITAS

mortalityrating.com

- Joint venture with Heriot-Watt in 2009:

PROJECTIONS TOOLKIT

2. Why should actuaries care about postcodes?

2. Why should actuaries care about postcodes?

Three reasons:

1. Improve your underwriting via geodemographic profiling.
2. Understand your data better.
3. Understand your risks better.

3. Improving your underwriting

3. Relative importance of risk factors

Financial impact of mortality rating factors:

Factor	Step change	Reserve	Change
Base case	-	13.39	
Gender	Female-male	12.14	-9.3%
Lifestyle	Top-bottom	10.94	-9.9%
Duration	Short-long	9.88	-9.7%
Pension size	Large-small	9.36	-5.2%
Region	South-North	8.90	-4.9%
Overall			-33.6%

Source: Richards and Jones (2004), page 39.

3. Anatomy of a UK postcode

$$\underbrace{\overbrace{\overbrace{\text{EH}}^{\text{region}}\ 11}^{\text{district}}\ 2}_{\text{sector}}\ \underbrace{\text{AS}}_{\text{walk}}$$

3. Improving your underwriting

- Compare the postcodes G1 2TD and G12 0PD.
- Both in Glasgow.
- Same mortality characteristics?
- Life expectancy "6.7 years less than the UK average"[1]?

Source: [1] Punter Southall, Postcode Life Expectancy Tool, accessed on 8th October 2013.

3. Anatomy of a UK postcode — G1 2TD

Source: Google Maps.

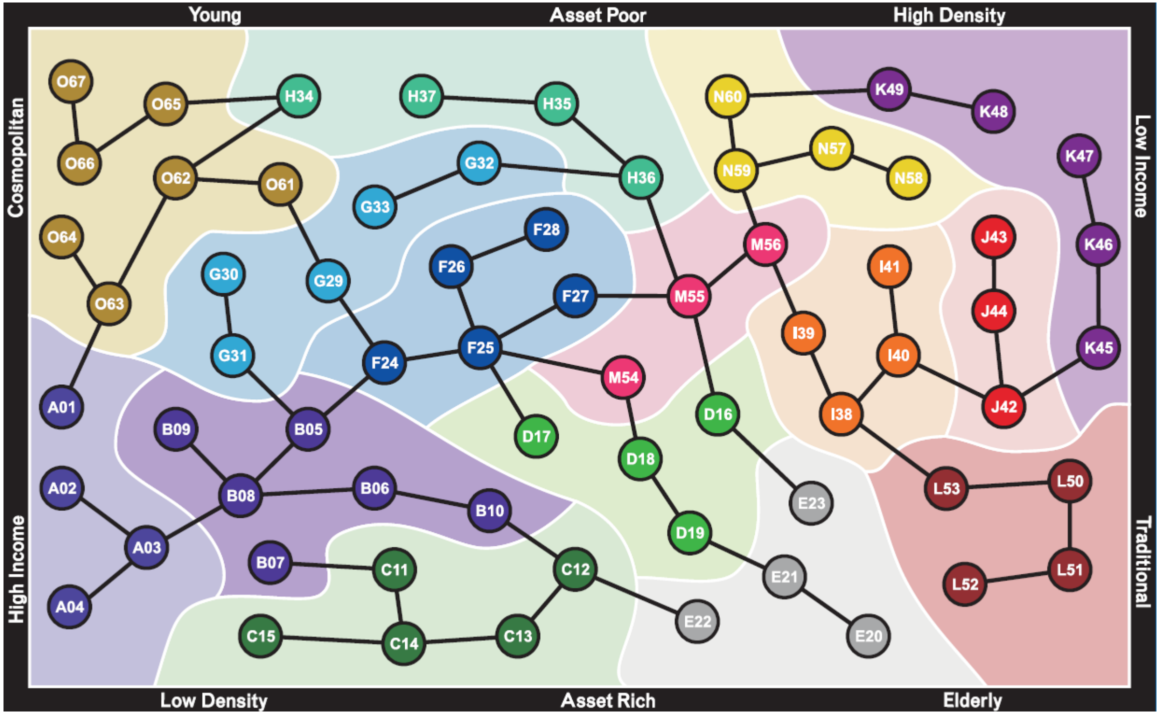
3. Anatomy of a UK postcode — G12 0PD

Source: Google Maps.

3. How to do postcode profiling

- 1.6 million current UK postcodes
- Each maps to a *geodemographic type*

3. Geodemographic example — Mosaic



Source: Experian Ltd.

3. Anatomy of a UK postcode — G1 2TD

Mosaic Type K47 — “Upper Floor Living, Deprived View”

Source: Google Maps, Experian Ltd.

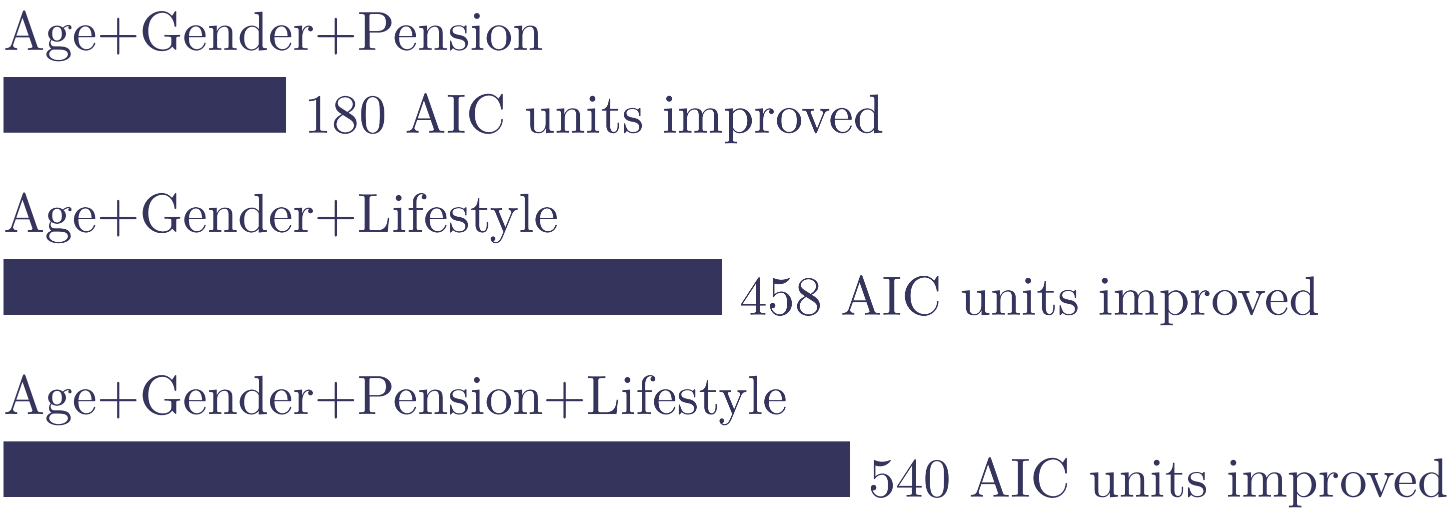
3. Anatomy of a UK postcode — G12 0PD

Mosaic Type A04 — "Alpha Territory, Serious Money"

Source: Google Maps, Experian Ltd.

3. Improving your underwriting

- Adding a geodemographic lifestyle factor improves most models.
- Example: compare alternatives to simple Age+Gender model:



→ Models with both pension size and lifestyle are usually better.

4. Understanding your data and risks

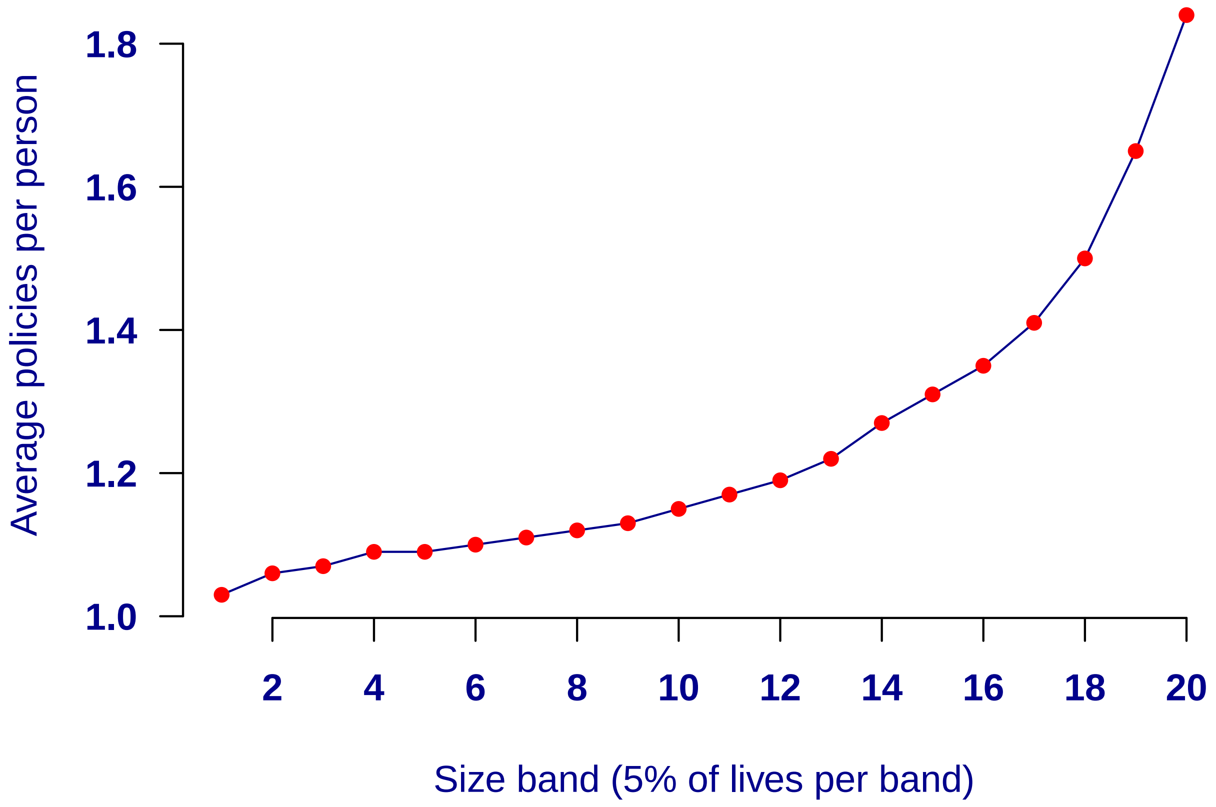
4. Understanding your data and risks

- Even without geodemographics, postcodes are very useful.
- A key application is *deduplication*.

4. Deduplication

- People often have multiple policies.
- Wealthier people tend to have more policies...

4. Wealth and duplicates



Source: Richards and Currie (2009).

4. Deduplication

Deduplication turns a dataset of policies:

PolicyNo	DateOfBirth	Gender	Surname	Forenames	Postcode	Pension
12345A	1968-03-13	M	Richards	Stephen	EH4 2AB	3,000.00
67890B	1968-03-13	M	Richards	Stephen	EH4 2AB	7,000.00

into a dataset of people:

DateOfBirth	Gender	Surname	Forenames	Postcode	Pension
1968-03-13	M	Richards	Stephen	EH4 2AB	10,000.00

4. Deduplication

We need deduplication to:

1. Build a more complete and accurate picture of each life.
2. Ensure the independence assumption for statistical modelling.

4. Deduplication challenges

- Problem: client identifier rarely reliable.
- Solution: use combination key made up from reliable fields, e.g.
  - — Date of birth
  - — Gender
  - — Surname
  - — First initial
  - — Postcode
- Names are often unavailable, so can just use:
  - — Date of birth
  - — Gender
  - — Postcode
- Deduplication must affect deaths and survivors equally to avoid bias.

5. Bias

5. Bias

All mortality investigations are based around the simple ratio:

$$\text{estimated mortality rate} = \frac{\text{deaths}}{\text{exposure}}$$

- Critical that whatever affects the top line also affects the bottom line.
- Failure to ensure this leads to biased estimates.

5. Bias

Sources of bias:

- Different rates of rejection during validation.
- Different rates of deduplication.
- Missing profiles.

5. Bias

Example of medium-sized pension scheme:

17,947	records
14	records rejected:
	— 1 missing date of birth
	— 9 missing dates of retirement
	— 4 inconsistent dates of retirement

→ All 14 rejected records were deaths, creating potential for bias.

→ However, there were 3,530 deaths in all so bias here is small.

5. Deduplication

We need deduplication to:

1. Build a more complete and accurate picture of each life.
2. Ensure the independence assumption for statistical modelling.

5. Bias

Pension scheme example:

17,832	records after deduplication
13,833	records with valid postcode
78%	records with valid postcode

- Bias possible if, for example, postcodes are only available for in-force.

$\rightarrow$ Only survivors would be deduplicated.

5. Bias

Pension scheme example — survivors:

14,311	survivors after deduplication
11,545	survivors with valid postcode
81%	survivors with valid postcode

Pension scheme example — deaths:

3,521	deaths after deduplication
2,288	deaths with valid postcode
65%	deaths with valid postcode

→ Bias has occurred during deduplication.

→ Potential over-statement of mortality rates.

5. Bias

- Bias also occurs when deaths less likely to have a profilable postcode.
- What is impact of not having a profilable postcode?
- First consider a simple Age+Gender model…

5. A simple Age+Gender model

Parameter	Estimate	Std. error	Z-value	p-value	Sig.	Lives	Deaths
Age	0.118179	0.0024	48.41	0	***	16,093	1,749
Gender.F	-0.434739	0.0562	-7.74	0	***	5,115	496
Intercept	-12.1213	0.1814	-66.82	0	***	16,093	1,749

- Mortality increases by about 12% for each year of age.
- Females have about 43% lower mortality than males.

Source: own calculations for Perks model fitted to data for medium-sized pension scheme.

5. Better than sex?

Parameter	Estimate	Std. error	Z-value	p-value	Sig.	Lives	Deaths
Age	0.121265	0.0024	49.6	0	***	16,093	1,749
Gender.F	-0.431421	0.0565	-7.63	0	***	5,115	496
Intercept	-12.4507	0.1838	-67.76	0	***	16,093	1,749
Unprofiled	0.696318	0.0673	10.35	0	***	3,138	325

- Can not having a Mosaic profile really be more important than gender?
- Problem stems from 35% of deaths not having a profile...

  ... while only 19% of survivors don't have a profile.

Source: own calculations for Perks model fitted to data for medium-sized pension scheme.

5. Where does profiling bias come from?

Two sources:

1. The profiler.
2. Your data processes.

5. Profiler bias

- Marketing profilers typically only hold recent postcodes (1.65 million).
- However, admin systems often hold historic postcodes for deaths.

    . . . and there are at least 0.8 million additional historic postcodes!

→ Actuaries need profiler with historic postcodes included.

    (Longevitas Mosaic has 2.45 million postcodes in total)

5. Data bias

- Incomplete postcode:

  EH21

  G15

- Careless data entry:

  1P18 6BH → IP18 6BH — E22 (Active Retirement, Beachcombers)

  BN14 OEX → BN14 0EX — D19 (Small Town Diversity, Innate Conservatives)

→ Need addresses and postcodes formatted to PAF standard.

→ Make sure address cleaning applied to deaths, not just survivors!

6. Actuaries v. marketeers

6. Actuaries v. marketeers

- Marketers only need current or recent postcodes (1.6 million).
- Actuaries need both current and historic postcodes (2.45 million).
- There are other differences…

6. Actuaries v. marketeers

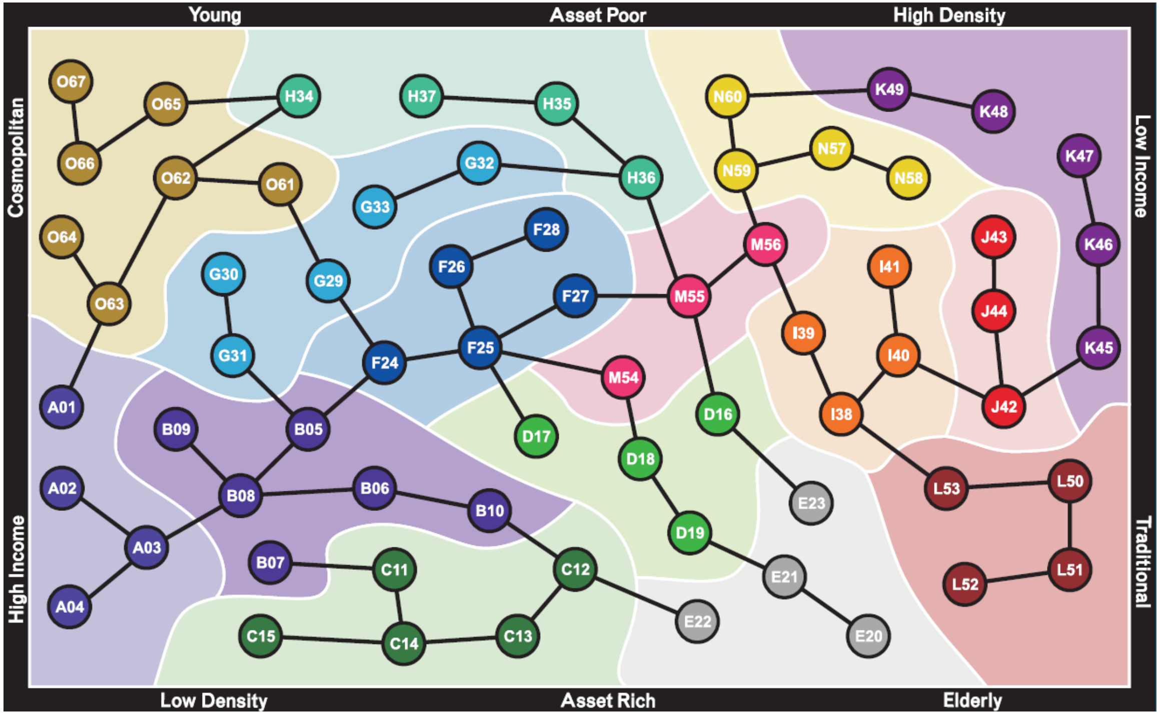
- Marketeers want lots of narrowly defined groups to target consumers.

$\rightarrow$ 67 Mosaic Types.

- Actuaries want a few homogeneous groups with similar mortality.

$\rightarrow$ Say 3–5 lifestyle categories.

6. Geodemographic example — Mosaic



Source: Experian Ltd.

6. Actuaries v. marketeers

- Geodemographic profilers like Mosaic are built for marketeers.
- 67 Mosaic Types can only be used directly with very large data sets.

- Even 15 Mosaic Groups are too many for most pension schemes.
- So how do actuaries get what they need?

6. Actuaries v. marketeers

Three approaches to mapping Type or Group to lifestyle:

1. Optimisation
2. Aggregation
3. Supergroups

6. Actuaries v. marketeers — optimisation

- Work out the best-fitting assignment of Type/Group to lifestyle.

  (The Longevitas optimiser does this)

- Drawback is over-fitting if number of events is small.

6. Actuaries v. marketeers — aggregation

- Aggregate the Types with small numbers of lives.
- Aggregate within Groups first.

- Avoids over-fitting risk…

  … but may still require Groups to be aggregated.

6. Actuaries v. marketeers — Supergroups

- Little used by marketeers, but useful for actuaries:

Supergroup	Groups
A	A, B
B	C, D
C	E, F
D	G, H
E	I, J, K
F	L, M
G	N, O

- May still require further aggregation.
- Ask Graham for suggestions!

7. Correlation

7. Correlation

- People in wealthier Mosaic Groups tend to have larger pensions:

Mosaic group	Average annuity (£ p.a.)	Average policies
Symbols of Success	4,348	1.33
Rural Isolation	3,405	1.30
Grey Perspectives	2,708	1.29
Suburban Comfort	2,203	1.24
Urban Intelligence	2,489	1.22
Happy Families	1,856	1.19
Ties of Community	1,592	1.19
Twilight Subsistence	1,394	1.17
Blue Collar Enterprise	1,444	1.16
Welfare Borderline	1,281	1.14
Municipal Dependency	1,093	1.12

Source: Richards and Currie (2009).

7. Correlation

Q. How do you know you aren't double-counting when both pension and lifestyle are included in a model?

A. With a survival model you can calculate the parameter correlations…

7. Correlations between parameters

Age+Gender+Pension+Lifestyle model:

	Age	Gender.F	Intercept	Lifestyle.2	Lifestyle.3	Lifestyle.4	Size.2	Size.3
Age	100%							
Gender.F	-19%	100%						
Intercept	-95%	9%	100%					
Lifestyle.2	0%	2%	-26%	100%				
Lifestyle.3	3%	1%	-30%	78%	100%			
Lifestyle.4	3%	2%	-28%	77%	79%	100%		
Size.2	3%	8%	-11%	-2%	3%	-1%	100%	
Size.3	-2%	14%	-5%	-7%	7%	0%	30%	100%

8. Conclusions

8. Conclusions

- Keep your addresses formatted to PAF standard.
- Look after postcodes for deaths as well as survivors!
- Ensure your profiler contains historic postcodes ($\approx$ 2.5 million or more).

References

Richards, S. J. and Jones, G. L. **2004** *Financial aspects of longevity risk*, Staple Inn Actuarial Society, London

Richards, S. J. **2008** *Applying survival models to pensioner mortality data*, British Actuarial Journal.

Richards, S. J. and Currie, I. D. **2009** *Assessing longevity risk and annuity pricing with the Lee-Carter model*, Faculty of Actuaries Sessional Meeting Paper, February 2009