Estimated Resident Population in the IDI: What happens if you want a denominator pre-2008?

THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

MEDICAL AND HEALTH SCIENCES
SCHOOL OF POPULATION HEALTH

ANNA HOWE

Disclaimer

The results of this presentation are not official statistics. They have been created for research purposes from the Integrated Data Infrastructure (IDI), managed by Stats NZ.

The opinions, findings, recommendations, and conclusions expressed in this presentation are those of the author(s), not Stats NZ or individual data suppliers.

Access to the anonymised data used in this study was provided by Stats NZ under the security and confidentiality provisions of the Statistics Act 1975. Only people authorised by the Statistics Act 1975 are allowed to see data about a particular person, household, business, or organisation, and the results in this presentation have been confidentialised to protect these groups from identification and to keep their data safe.

Careful consideration has been given to the privacy, security, and confidentiality issues associated with using administrative and survey data in the IDI. Further detail can be found in the Privacy impact assessment for the Integrated Data Infrastructure available from www.stats.govt.nz.

Outline

- How is the Estimated Resident Population determined?
- Where are the dataset problems if we want a population from 2006?
- What is the impact if we remove datasets?

NB: Not a discussion on the method for determining ERP, per se

Why do I care?

To examine vaccination data from 2006-2016 to evaluate:

1. the uptake of the national pneumococcal immunisation schedule for young people, considered to be at high-risk of invasive pneumococcal disease;
2. how effective the enhanced immunisation schedule is against IPD and pneumonia in this group.

Therefore, I require a denominator and information for censoring from 2006

Estimate Resident Population

Activity in education datasets in last 5 years

- Tertiary enrolments
- Industry training
- School enrolments

Activity in tax datasets in last year

- EMS (tax at source) dataset
- Self-employment income

Activity in health datasets in last year

- GMS claims
- Laboratory tests
- Non-admissions events
- Prescriptions dispensed
- Consultation with PHO-registered GP
- Discharged from publicly funded hospitals

Activity in ACC in the last year

Births in the last 5 years

Visa approvals for ages under 5 years

Remove individuals from the population if living overseas

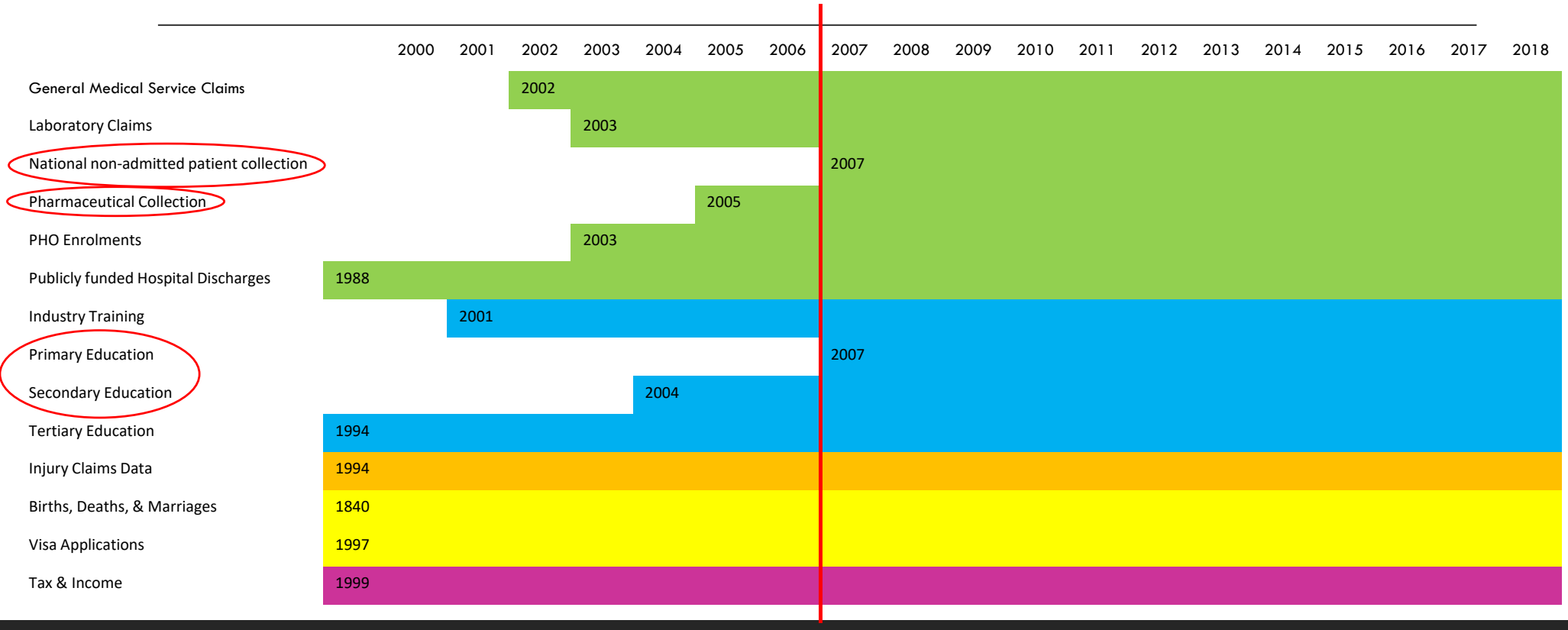
Estimate Resident Population

Combine all activity sources

Combine all individuals who have activity and are in the spine

- Remove those with no DOB or sex, who died prior to reference date, are not NZ residents (non-resident or > 182 days overseas)

ERP dataset coverage



Modified Estimate Resident Population

Removed:

- Pharmacy dispensings
- Non-admission events
- School enrolments

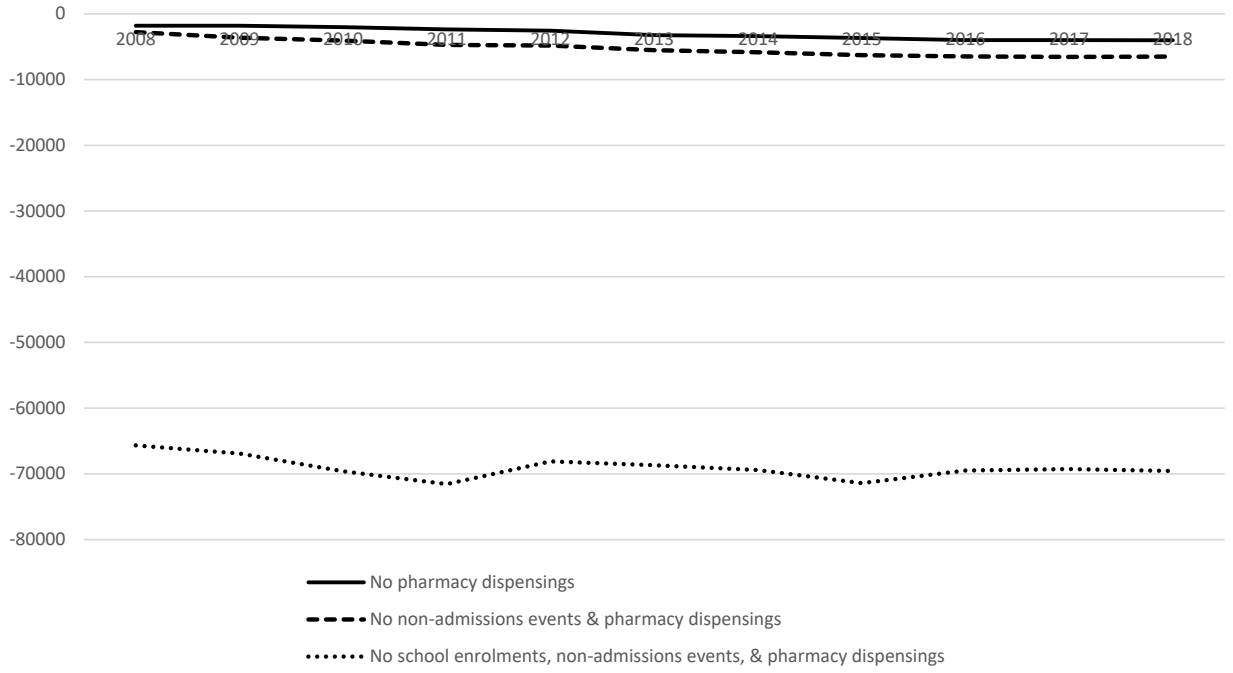
Calculated absolute and relative differences between modified and original counts

- on rounded values (RR3)
- by sub-groups of interest

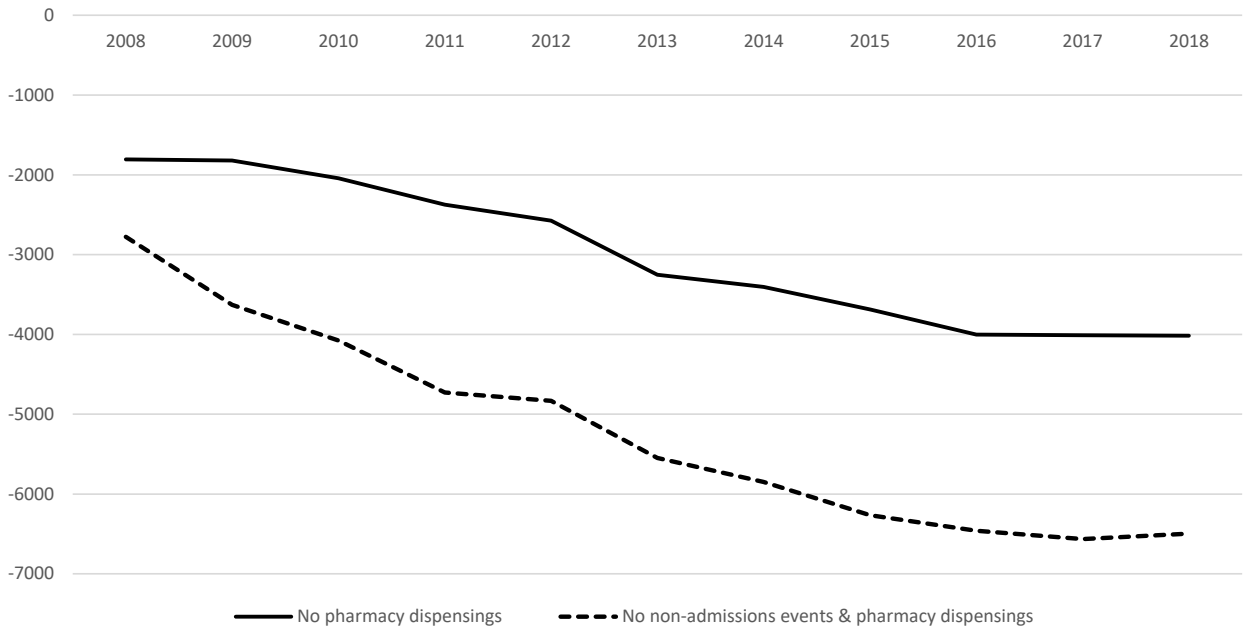
Year	Original Count	Without pharmacy dispensings	Without non-admissions events & pharmacy dispensings	Without school enrolments, non-admissions events, & pharmacy dispensings
2008	4,297,935	4,296,129	4,295,157	4,232,274
2009	4,348,617	4,346,796	4,344,987	4,281,684
2010	4,393,161	4,391,118	4,389,087	4,323,561
2011	4,416,912	4,414,539	4,412,184	4,345,347
2012	4,433,382	4,430,808	4,428,546	4,365,282
2013	4,464,543	4,461,291	4,458,996	4,395,858
2014	4,523,424	4,520,019	4,517,577	4,454,016
2015	4,602,141	4,598,454	4,595,880	4,530,747
2016	4,691,079	4,687,077	4,684,620	4,621,611
2017	4,778,067	4,774,056	4,771,503	4,708,815
2018	4,847,496	4,843,479	4,841,001	4,777,941

Estimated Resident Population counts

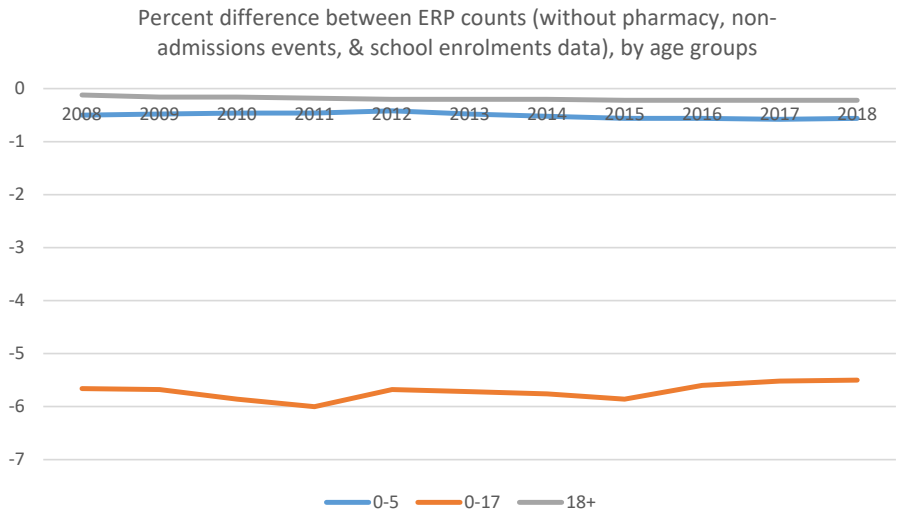
Absolute difference between modified and original ERP counts

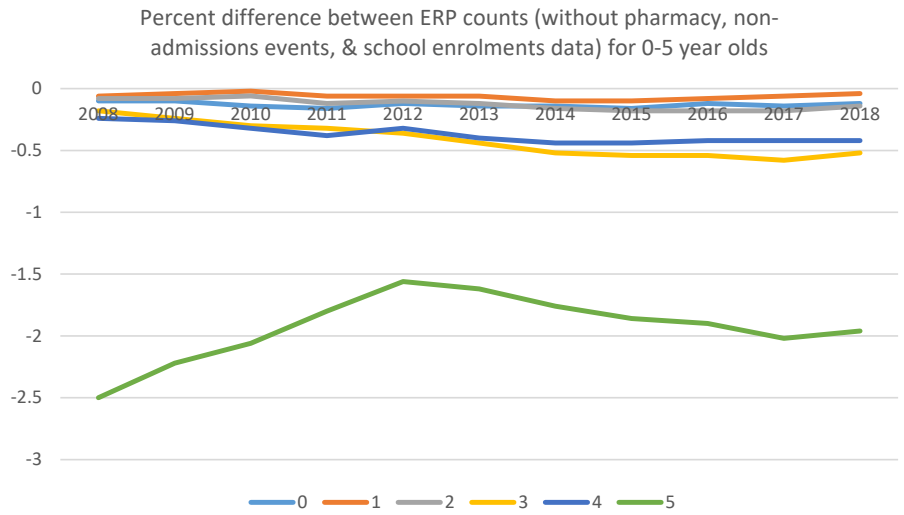


Absolute difference between modified and original ERP counts

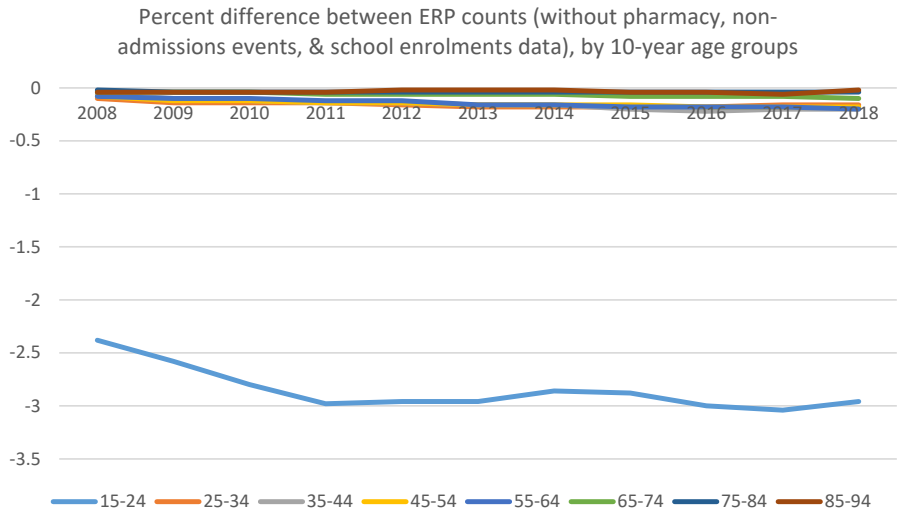


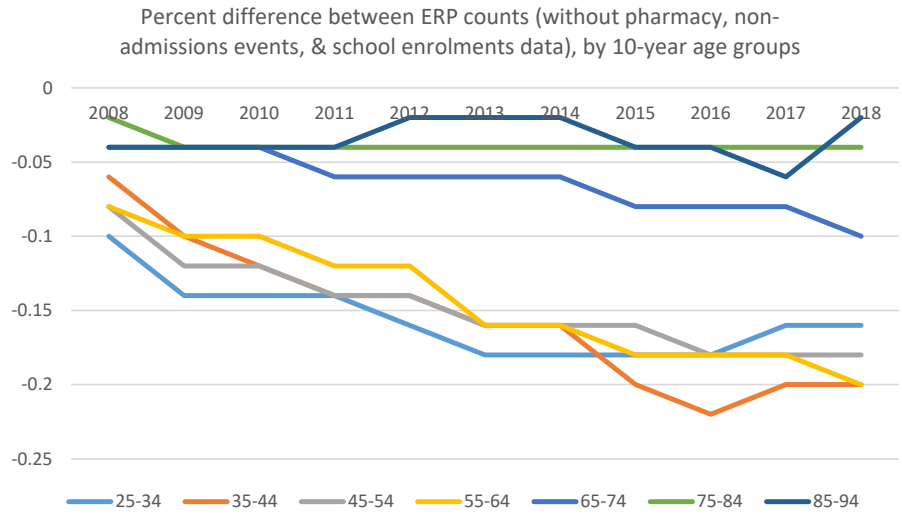
Percent difference between modified and original ERP counts, by age



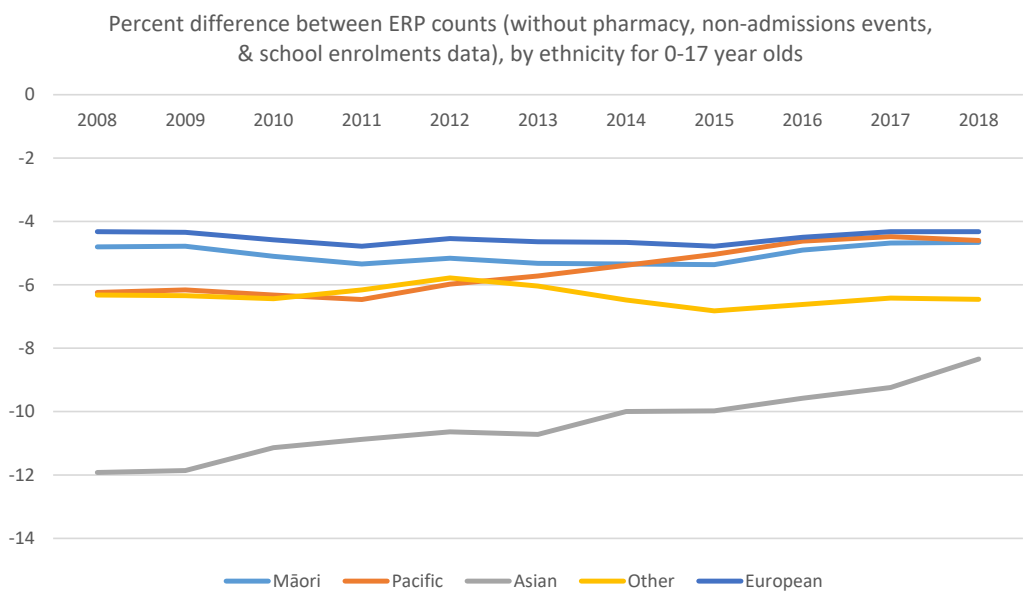


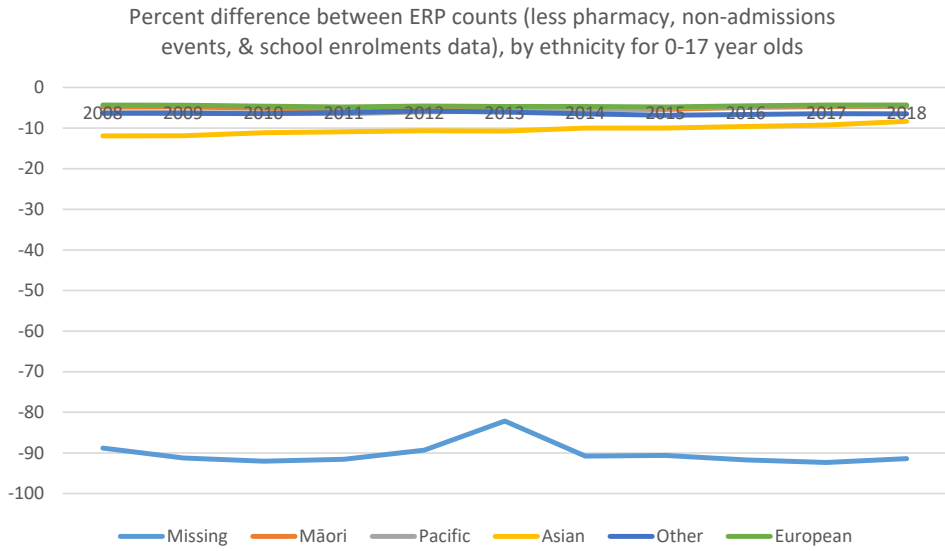
Percent difference between modified and original ERP counts, by age



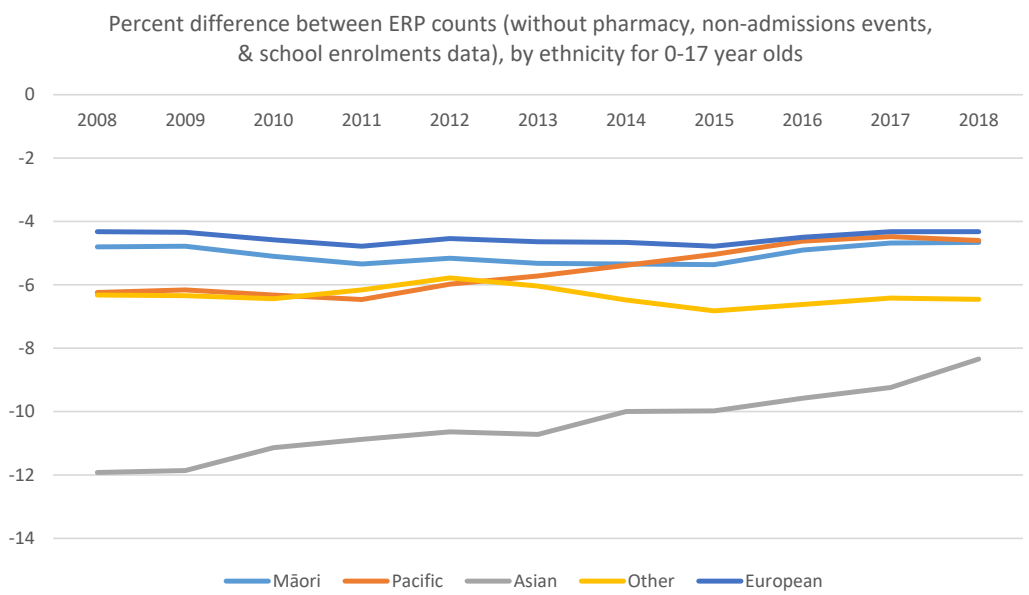


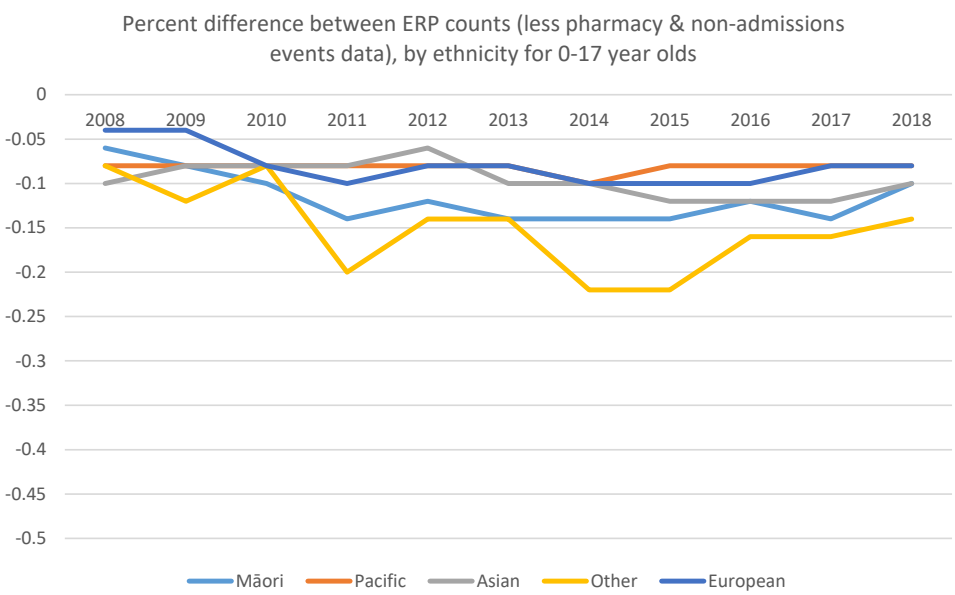
Percent difference between modified and original ERP counts, by ethnicity



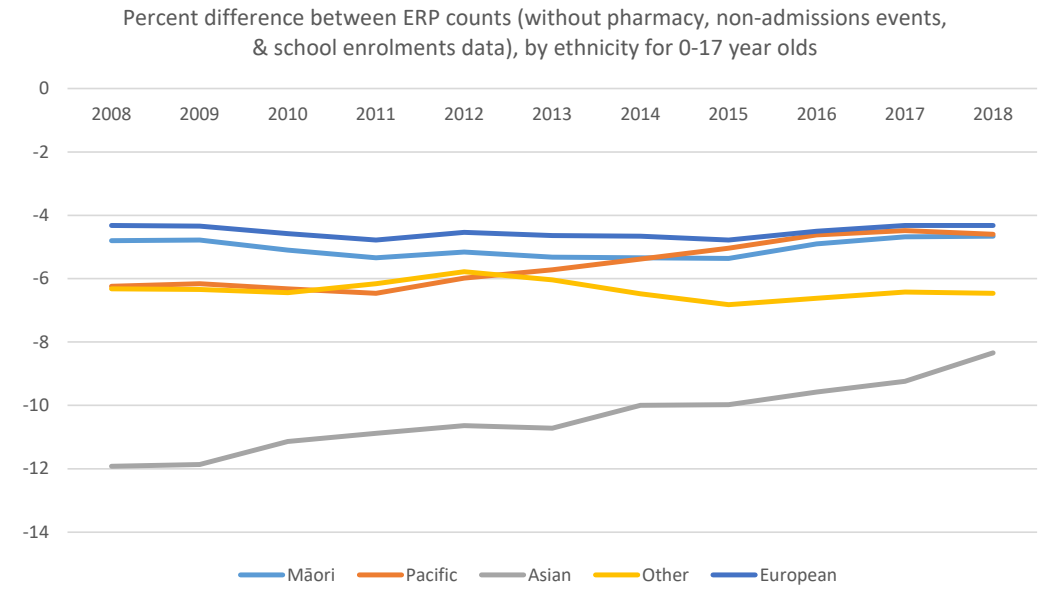


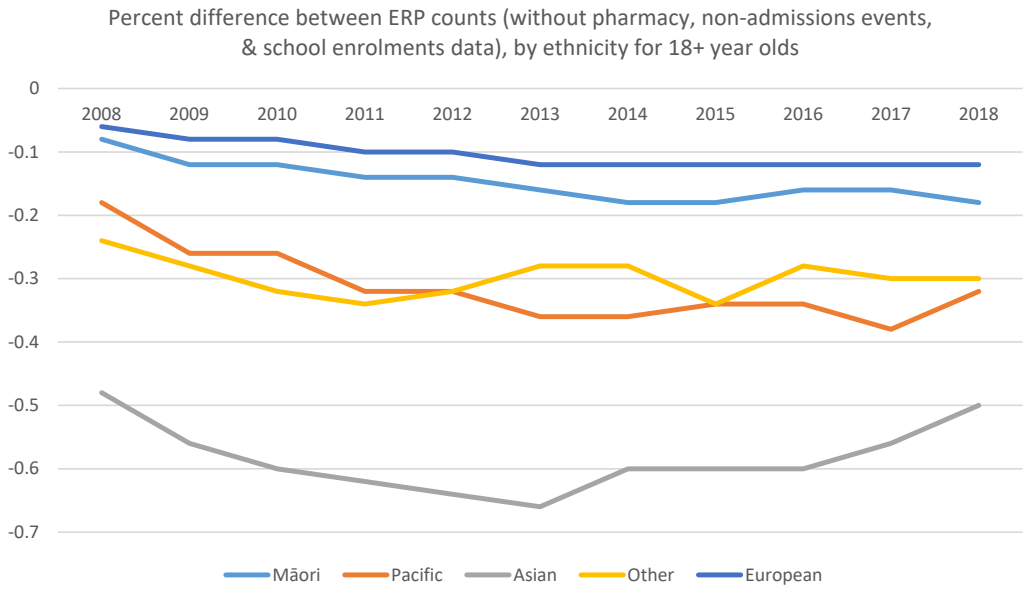
Percent difference between modified and original ERP counts, by ethnicity



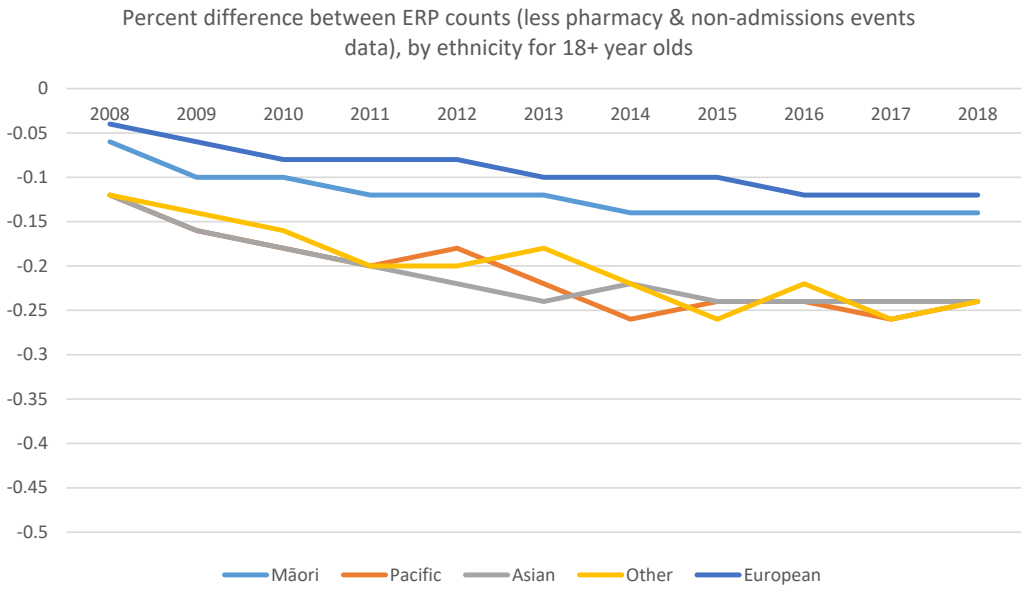


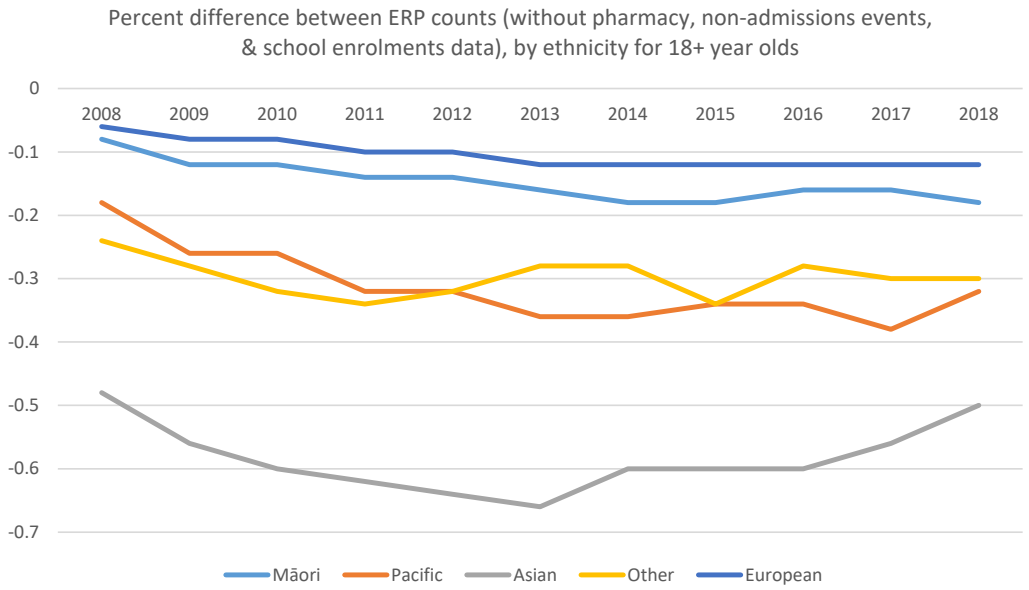
Percent difference between modified and original ERP counts, by ethnicity



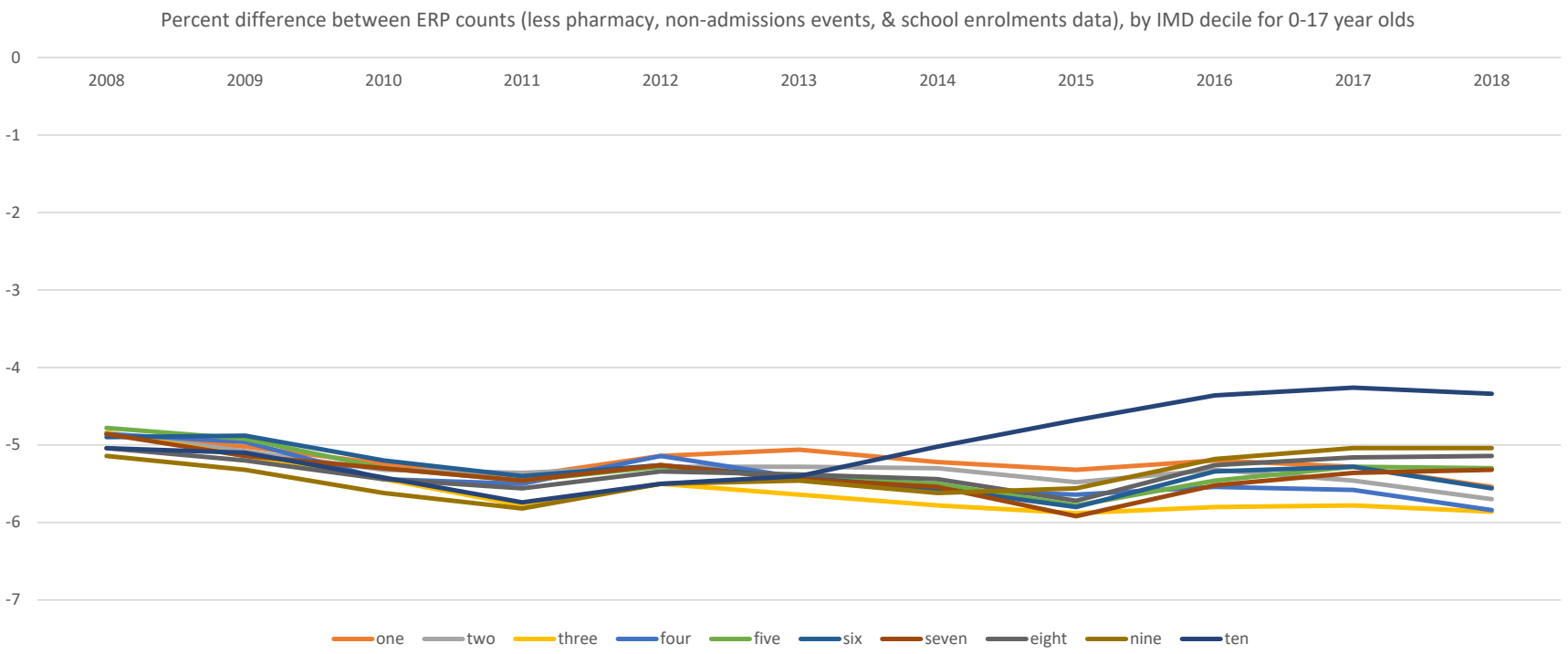


Percent difference between modified and original ERP counts, by ethnicity

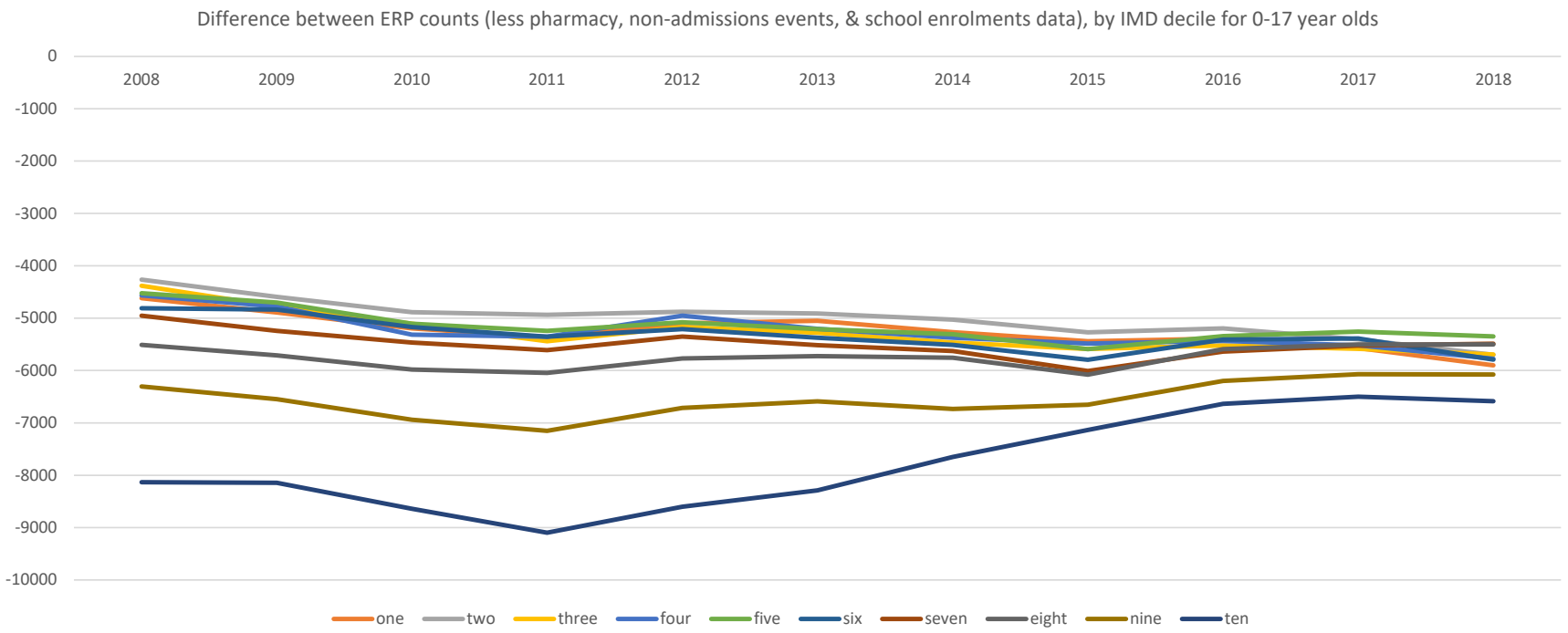




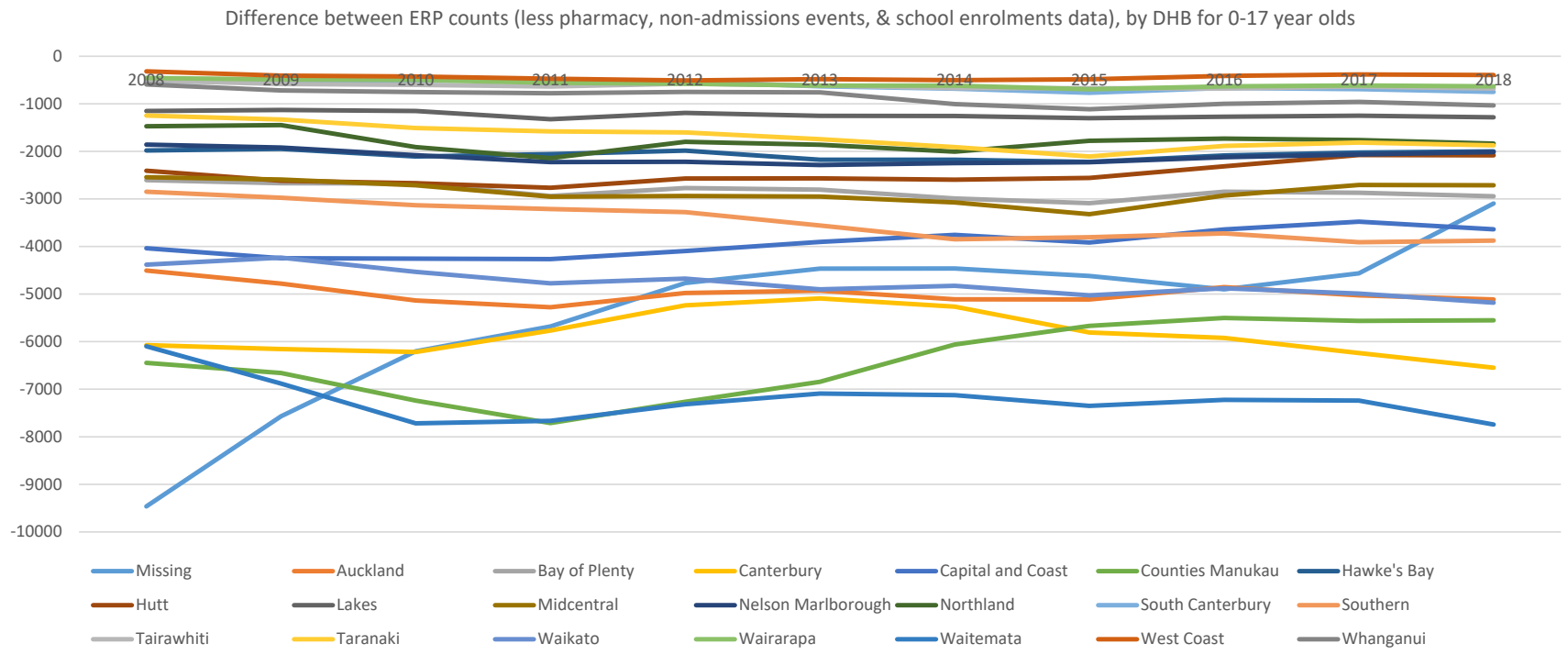
Percent difference between modified and original ERP counts, by IMD decile



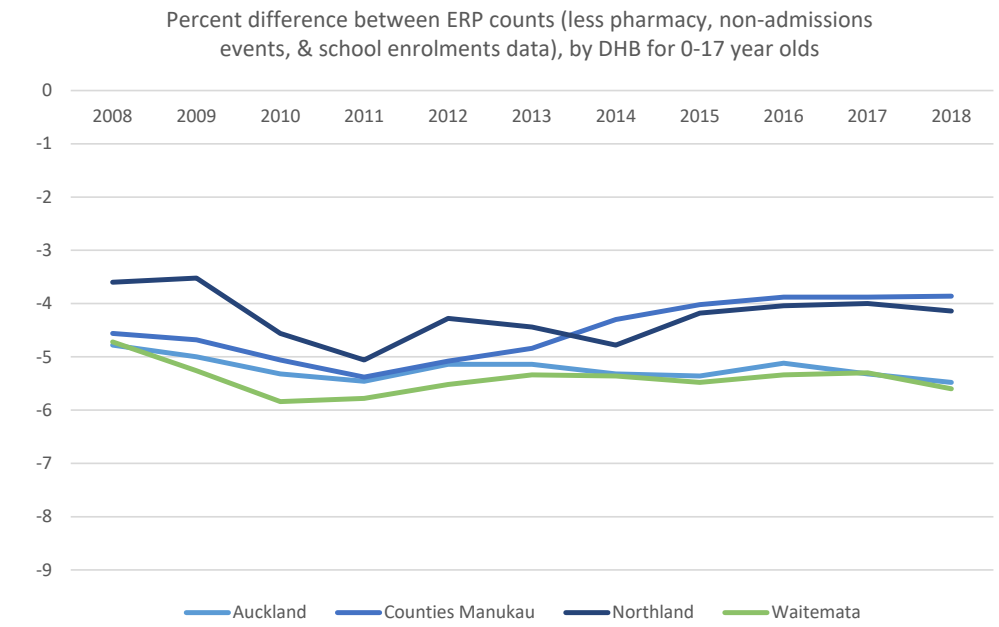
Absolute difference between modified and original ERP counts, by IMD decile

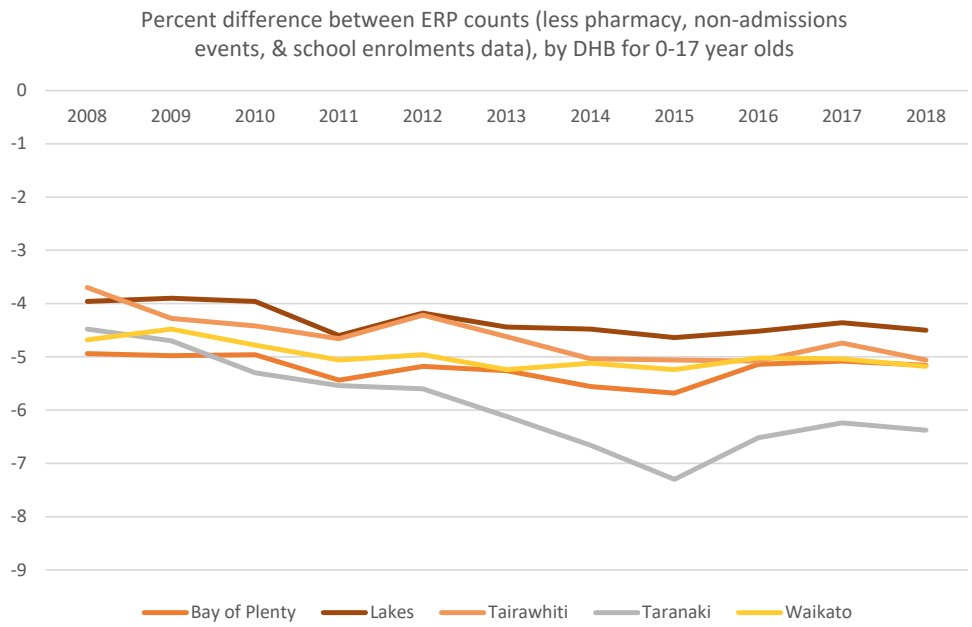


Absolute difference between modified and original ERP counts, by DHB

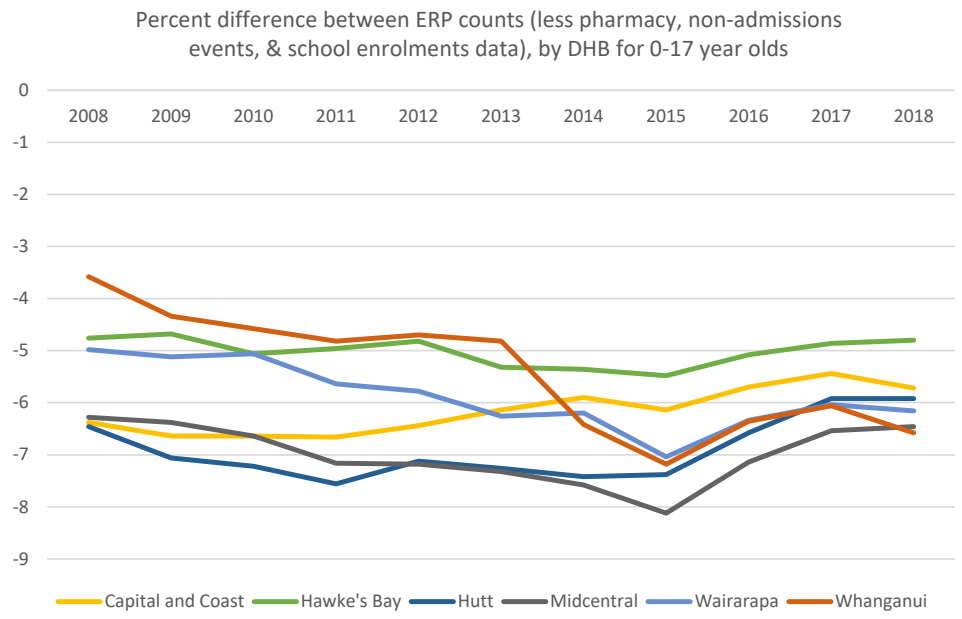


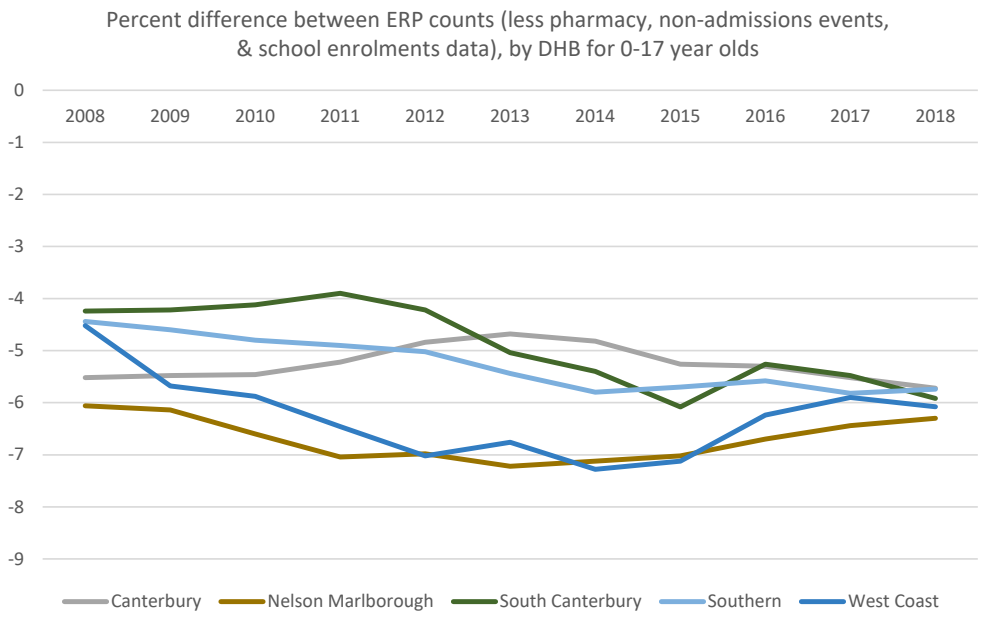
Percent difference between modified and original ERP counts, by DHB



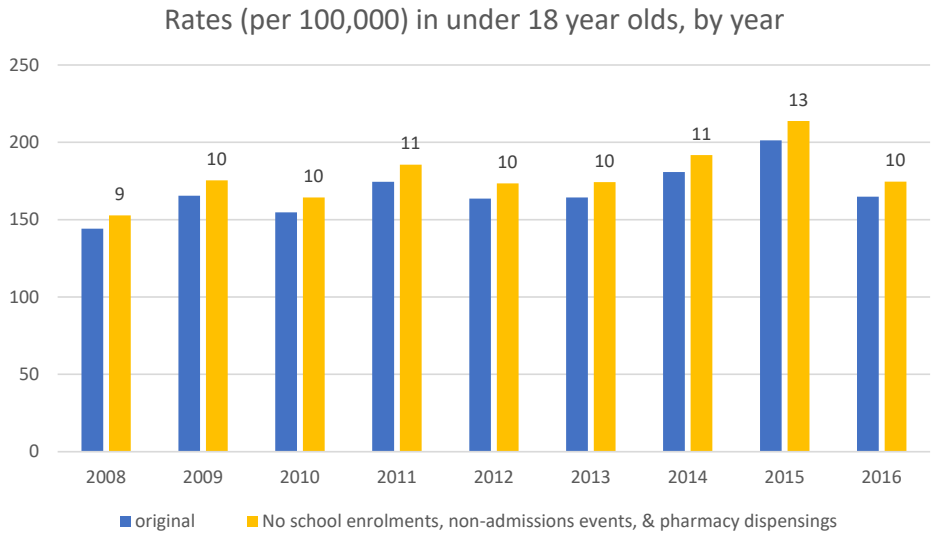


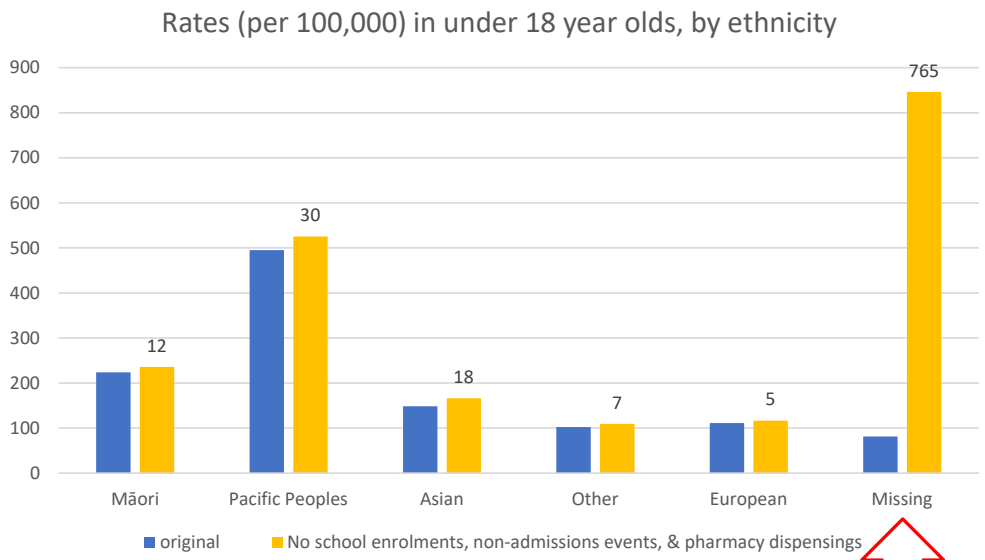
Percent difference between modified and original ERP counts, by DHB





Undercount results in over-estimation of rates





ERP of
25,842 *cf* 2,481

Conclusions

Possible to determine ERP back to 2006 by removing some datasets

BUT need to

- understanding where your bias is
  - For example age, ethnicity, region, deprivation
- and the likely impact on your results

Thank You