Regarding any statistical modelling the only thing I can think of would be to formally investigate the correlation between the residual proportions and the mortality rate. But on the lexis surface, both variables exhibit the equivalence of spatial correlation (is that right?) Therefore a straightforward r^2 violates the independence assumption. So whether or not the relationship between mortality rates and residuals is significant has to probably be assessed using some sort of simulation to get the distribution?

Another thought: *the importance of the data source itself.* So if the argument is the denominators i.e. the population counts are inflated, the question is to what extent are the individual statistical offices that collate this data, and the HMD which harmonizes it, aware of this and adjust for it. E.g. In countries where the data is census based it is probably reasonable to assume the intercensal estimates underestimate both immigration and emigration, the net effect on the error depending on which is larger. So e.g. in the UK the mortality rates might look artificially high in the groups where the in-migration has been most underestimated, but this is presumably fixed when the retrospective adjustments are made when the new census comes in.

- So one idea would be to see if this adjustment effect can be seen anywhere, is old HMD data available i.e. of the sort before it was adjusted for new census numbers?

Then there is also the question of EU harmonization and to what extent this has already happened. There is a 2008 regulation (reference in .docx) that has redefined at least in some countries what the usual population is. To what extent has this affected the data, was it fixed retrospectively in some countries?

OK, here's me trying to conceptualise the probelm. So the issue is that we don't only have *recorded* and *expected* population counts, but implicitly we also have *actual*. Which is the denominator for the true mortality rate. The possible relationships are:

positive residual proportion

	expected	<	recorded	<	actual	1.& 4. (& 3./5?)
	expected	< actual <	recorded			2.& 6. & 3.
actual <	expected	<	recorded			2.& 6. (& 3./5?)

negative residual proportion

	recorded	<	expected	<	actual	1.& 4 (& 3./5?)
	recorded	< actual <	expected			1 & 4.& 5.
actual <	recorded	<	expected			2. & 6.(& 3./5?)

And the possible values of, and explanations for, the actual counts are (see last column of the table):

1. larger than recorded if people are not registring/being picked up my the system. (Always with 4.)
2. lower than recorded if people are being double counted. (Always with 6.)

3. larger than expected if they are in-migrating from outside the observed countries. (Works with 4., counters 6., never with 5).

4. larger than expected if the previous year's age specific mortality rate is too high because it was based on the recorded count being too low (Always with 1., can be countered by 5.)

5. lower than expected if they are emigrating outside the observed countries(Works with 6., counters 4., never with 3).

6. lower than expected if if the previous year's age specific mortality rate is too low because it was based on the recorded count being too high (Always with 2., can be countered by 3.)

So if I've got those right I can't see how the relationship between recorded and expected, which is the only one we can actually see, tells us anything about the the actual population count. I'm nor sure any more I understand this, have to think about it a bit more.

Then there are the overlayed contourplots. I did one for 50-80 year olds and you can see the ww1 cohort effect on the mortality rates coincide precisely with a positive residual proportion. So why are more people being recorded in that cohort? Is this an artefact of the data collection (smoothing?)? A circularity problem (don't think so).