Sun exposure and longevity: a blunder involving immortal time

Jane E Ferrie* and Shah Ebrahim

*Corresponding author. E-mail: jane.ferrie@bristol.ac.uk

Unfortunately we have to start this Editor’s Choice with an acknowledgment that we have fallen prey to a common, perennial problem; immortal time bias.

To illustrate the concept we borrow an example from William Farr, as used by James Hanley and Bethany Foster in a full and entertaining exposition of the problem in this issue of the journal.1 Generals and bishops live longer than corporals and curates—but this is not necessarily because an elevated occupational status makes you live longer—it may simply be because you have to reach a certain age before it is possible to hold such positions. People become generals and bishops in middle age so their deaths arise after this point in time, whereas corporals and curates can die at any age above 20 or so.2 This difference in time during which an event can occur to one group but not the other produces a bias favouring longer life expectancy—immortal time bias. In the figure on the next page, the problem is evident at a glance (Figure 1).3

In the October issue of the International Journal of Epidemiology (IJE) last year, we published a paper by Peter Brøndum-Jacobsen and colleagues in which they examined the effects of sunlight exposure on mortality among the whole population of Denmark aged above 40 years, using linked data from national registries.4 They used non-melanoma skin cancer as a proxy for sun exposure, which is a clever idea but it should have been obvious that the findings were ‘too good to be true’—an apparent halving of all-cause mortality and reductions in myocardial infarction and hip fracture. The authors concluded: ‘Causal conclusions cannot be made from our data. A beneficial effect of sun exposure per se needs to be examined in other studies’.

The Danish media picked up the story and it became front page news—‘Sunbathers live longer’.5 Although the authors never made this claim in their published paper, their interviews with the press did not appear to emphasize their non-causal conclusion. The Danish Cancer Association claims that this paper has undone all their good work in persuading Danes to keep out of the sun to avoid skin cancers.

Commentators on the story identified a likely problem of immortal time bias. People in the ‘sun exposure’ group had to live long enough to be diagnosed with skin cancer but the comparison group only had to be over 40 years old—the design of the study had built in a potential bias in favour of longevity among those presumed to be more highly exposed to sunlight. Theis Lange and Neils Keiding, in a letter commenting on the paper, pose questions about how such highly improbable findings got through the editorial process at IJE.6

In response to this criticism, Brøndum-Jacobsen and colleagues argue that their paper used both cohort and case-control analyses, and that the latter should be free from immortal time bias as cases and controls were matched on age.7 They acknowledge that the case-control analyses—which showed much smaller survival advantage [odds ratio (OR): 0.97, 95% confidence interval (CI) 0.96 to 0.99; vs hazard ratio (HR): 0.52, 95% CI 0.52 to 0.53]—should have been included in their abstract. In addition, they conducted a revised Cox proportional hazards analysis stratified by 10-year, 5-year and 2-year age strata in an attempt to control for immortal time bias, and interpret these findings as similar to those in their original paper. However, they fail to stress that the effect sizes become increasingly attenuated as the age matching becomes more exact, suggesting that the apparent effect of sun exposure may indeed be produced by immortal time bias.

Ironically, in parallel with the review and publication of this paper we had commissioned an ‘Education Corner’
paper by Hanley and Foster on ‘Avoiding blunders involving immortal time’. At the editors’ request they added a postscript commenting on the Danish analyses. Using a Danish population of over 4 million people drawn from the Human Mortality Database, they modelled the effect on all-cause mortality of an annual prize allocated at random. This mimics the incidence of non-melanoma skin cancer, but clearly the prize could have no biological effect on longevity. However, the analysis almost exactly mirrored both the original published findings and the revised age-strata analyses produced in response to Lange and Keiding’s criticism. The effects reported by Brøndum-Jacobsen and colleagues could clearly be spurious.

Should the IJE have identified these flawed findings during the editorial process? The short answer is ‘Yes’ and our reviewers did indeed spot the problem: ‘For the non-melanoma skin cancer group, you have to survive long enough to get non-melanoma skin cancer before you can die—that is, you cannot die before the age of acquiring non-melanoma skin cancer’. In response to this and several other comments, the authors conducted a revised analysis excluding people under 40 years and applying different methods of analysis, and seemed to consider that by truncating the age range they had dealt with the reviewer’s comment above. Our reviewer considered the revised analysis to be an improvement and did not comment on the issue again. The paper was considered ‘clever’ and ‘innovative’ by our reviewers and was a large study apparently confirming earlier findings. The handling editor considered that the authors had done a sufficiently good job in dealing with the criticisms, and an editor-in-chief then accepted the paper for publication. The authors’ matched case control analyses provided more plausible findings but we failed to ensure that these findings were given prominence or substituted for the misleading Kaplan-Meier and Cox model analyses.

Should this paper be retracted now? There are many examples of flawed analyses and inappropriate conclusions in the biomedical literature. Neither the authors nor the editors and reviewers who let such papers slip through the net are guilty of intentional mischief or fraud. We all learn from mistakes, and removing authorial and editorial mistakes from the public record is not a good solution. On the editorial side, like all who have fallen into this trap, we need to be more vigilant in the future. We have added a brief description of the problem and links to the material
in this issue, which in effect amount to ‘post-publication peer review’, to the online version of the paper by Brøndum-Jacobsen and colleagues.\(^9\) We believe that this editorial comment, the accompanying letters and Hanley and Foster’s excellent overview of immortal time bias provide a better understanding of the problem, how to detect it and how to deal with it properly.

**Do Republican presidents kill babies?**

Further methodological debate is provoked by a paper in this issue of the journal by Javier Rodriguez and colleagues.\(^8\) The provocative title—above—of a US blog on the paper\(^9\) reminded us that US presidents, Democrat and Republican, have frequently been indicted with killing babies and children in other countries: ‘Hey, hey LBj, How many kids did you kill today?’ Rodriguez and colleagues look at the potential of presidents to contribute to infanticide at home.\(^3\) As the authors point out, infant mortality rates in one of the world’s wealthiest countries are shocking. Despite a dramatic downward trend between 1965 and 2010, for the period 2005-10 the rate exceeded that of most other developed countries as well as some less developed countries, like Cuba.\(^10\)

Judging by its widespread acceptance, if not translation into effective policy, few quibble with the assertion by Sir Michael Marmot that the causes of the causes of inequalities in health reside in the social and economic arrangements of society.\(^11\) Rodriguez and colleagues seem to have drawn particular flak for having located their study within the emerging sub-discipline of ‘political epidemiology’. In his many writings on inequalities, Marmot rarely strays into the overtly party-political. However, it is here that the causes of the causes can be at least partially addressed. Despite depressing similarities, in most countries there are real differences in health and social welfare policies between the main political parties. As Rodriguez and colleagues point out, it would be surprising were these not related to health outcomes, especially among vulnerable groups.\(^12\)

To test this, Rodriguez and colleagues examined associations between the party of the last nine US presidents—four Democratic and five Republican—and infant mortality rates from 1965 to 2010. Their regression estimates show that, relative to trend, infant mortality rates during Republican administrations have been, on average, 3% higher than during Democratic administrations. These findings remained after adjustment for factors like unemployment, smoking, abortion rates, education and income.\(^12\) The authors finish their paper on a cautionary note: ‘Further research is needed to determine whether the association we have uncovered is causal, and to identify the mechanisms involved’. Coverage of the paper in the Washington Post is similarly cautious: ‘There is a correlation here that persists after accounting for some obvious alternative explanations. However, the mere existence of this correlation does not permit any strong conclusions’.\(^13\)

In a commentary on the paper, Ralph Catalano takes Rodriguez and colleagues to task for providing so little in the way of explanatory mechanisms,\(^14\) although the authors do indicate possibilities, such as austerity vs increased social welfare in response to economic crisis.\(^8\) Danny Dorling in his commentary suggests psychosocial and behavioural as well as material mechanisms.\(^15\) However, Catalano’s main criticism is reserved for the methods. Using data from the Human Mortality Database he sets out to show that the findings of Rodriguez and colleagues are a product of their methods. In so doing he adds artefact to the potential explanations proffered and completes the quartet of potential explanations: material, psychosocial, behavioural and artefact, proposed by the Black Report on inequalities in health.\(^16\) Rodriguez and colleagues respond to Catalano with arguments about the relative merits of cubic spline and Box-Jenkins methods.\(^12\) The editors of the *IJE* are of the opinion that there is no definitively ‘right’ answer for interpreting time trends, and so welcome this informative debate which will no doubt continue.

**Methods of measurement and the mirror test**

‘If you really want to know whether you are obese, just undress and look at yourself in the mirror’. This, according to Henry Blackburn and David Jacobs,\(^17\) was the advice given by Ancel Keys to participants visiting his laboratory who wanted to know if they were too fat. Undoubtedly the best possible indicator at the individual level, at the population level the mirror test is of more limited use. However, despite his reported fatophobe attitudes,\(^17\) it was Keys and colleagues who gave the ratio weight/height squared its now familiar name ‘body mass index’ and, in a comprehensive comparison of various measures of relative weight, endorsed it as the optimum obesity index.\(^18\)

In 2010 a new *IJE* series was launched with an editorial by Debbie Lawlor and Nish Chaturvedi. The aim of the new series—‘Methods of measurement in epidemiology’—was to help ‘population health scientists to make informed decisions about the best measurement tools to use in different contexts and to understand the impact of using any one measurement tool’.\(^19\) Included in their editorial was an example of an area which the editors felt would benefit from inclusion in such a series—the measurement of body size and composition. Sadly, as yet, no one has risen to the challenge of addressing this issue. However, Keys’ original
paper and the accompanying three commentaries provide a comprehensive review of the merits and limitations of body mass index (BMI) as a measure of body fat and a risk factor for disease. Lawlor and Chaturvedi ponder the relative value of BMI compared with modern methods of measuring adiposity such as dual-energy X-ray absorptiometry scans, bioelectrical impedance and computed tomography or magnetic resonance imaging, particularly in children. The commentators give more prominence to alternative measures, cheaper and simpler to collect at population level, such as waist circumference and waist:hip ratio. However, time and again the discussion returns to the simplicity of BMI, its utility and its value as a marker of cardiometabolic risk.

Several of the empirical papers in this issue also focus on anthropometric measurements, mostly measures of adiposity, most often in relation to chronic conditions in later life. An elegant paper by Adam Hulman and colleagues presents the simultaneous effects of ageing and secular trends on the distribution of major cardiovascular risk factors in the UK over 25 years from 1985, using five phases of data from the Whitehall II study. In addition to blood pressure and lipids, the authors examined BMI and waist circumference in women and men aged 57-61. The bell curves for both measures flattened over successive phases. Smaller shifts to the left than to the right (higher values) indicate that most weight gain was seen among those already overweight and obese. However, although BMI changed little in the lean they still increased in girth. The authors suggest a simultaneous loss of muscle mass and accumulation of abdominal fat—a situation all too familiar to those of us in the requisite age group.

<table>
<thead>
<tr>
<th>Box 1. Daily intake of the overfed convicts</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Breakfast</strong></td>
</tr>
<tr>
<td>12 oz. bread</td>
</tr>
<tr>
<td>16 oz. meat</td>
</tr>
<tr>
<td>½ oz. salt</td>
</tr>
<tr>
<td>1 pint soup</td>
</tr>
<tr>
<td>1 pint tea with ¾ oz. sugar &amp; ½ oz. tea</td>
</tr>
<tr>
<td><strong>Total weight</strong></td>
</tr>
<tr>
<td>2 lb.</td>
</tr>
</tbody>
</table>

The total weight of food allowed per diem was thus seven pounds six ounces, including fifty-nine ounces of solid ailment; while to the blacksmiths and sawyers, an extra ration of 4 oz. of bread and 4 oz. meat was given, bringing their diet up to 7 lbs 14 oz., of which 67 oz. was solid ailment. (N.B. 67 oz. = 1.9 kilograms)

Emily Williams and colleagues used longitudinal data from a UK multi-ethnic population of older adults to examine associations between weight gain over two decades and disability. They found both weight gain and moving up a BMI category to be associated with higher risks of three measures of later life disability: objectively measured locomotor dysfunction, self-reported functional limitation and problems with activities of daily living. Risks associated with weight gain were mitigated if accompanied by an increase in physical activity. However, highest levels of risk were observed among those who remained obese throughout.

While the potential for an association between weight gain and disability is immediately obvious, associations between BMI and autoimmune diseases are less so. In an 11-year follow-up of 75,000 women in the Danish National Birth Cohort, Maria Harpsøe and colleagues examined associations between pre-pregnancy BMI and 43 of the most common autoimmune diseases identified via national hospital in- and outpatient registers. Risks of any autoimmune disease, dermatitis herpetiformis and type 1 diabetes increased with each unit increase in BMI, although the risks of celiac disease and Raynaud’s phenomenon decreased. There were also higher risks of psoriasis, rheumatoid arthritis and sarcoidosis in obese compared with normal-weight women. The authors discuss potential explanations, including a common aetiology linking adiposity to autoimmunity, for example via changes in adipokine and cytokine levels, or shared risk factors, but also suggest that their novel findings need confirmation.

In addition to the potential risk of subsequent autoimmune disease, pre-pregnancy BMI has been associated with...
a number of adverse offspring outcomes including, more recently, general cognitive ability or intelligence. However, the problem with any study of associations between maternal BMI before pregnancy and offspring outcomes is the propensity for confounding by genetic factors and the postnatal environment. One approach to this problem is to compare associations between maternal BMI and offspring outcomes with those for paternal BMI. A recent study which took this approach concluded that the similar association between maternal and paternal BMI and offspring intelligence suggests that it is not a specific pregnancy-related adiposity effect. Another approach to minimizing the effects of familial factors is to use a sibling design. This is one approach applied by Lisu Huang and colleagues in data from the Collaborative Perinatal Project. The association they observed between maternal pre-pregnancy obesity and intelligence at age 7 years counter the findings of Mette Bliddel and colleagues but confirm associations observed in earlier studies.

Confounding, criminality and overfed convicts

Intelligence is the main confounder of an observation by Amber Beckley and colleagues that short men are more likely to be convicted of violent crimes. The authors took advantage of Swedish register data for men who underwent military conscription tests between 1980 and 1992, to examine associations between height and first conviction for acts of violence such as homicide, assault and kidnapping. Over a mean time at risk of 27 years, just under 7% of the 713,877 conscripts were convicted. However, after adjustment for other anthropometric measures, sociodemographic factors and general cognitive ability, a weak but positive association between height and crime emerged. Muscle strength as well as height were measured during the conscript examination and although, intuitively, we might expect stronger men to be more likely to engage in violence, the negative association between strength and conviction survived adjustment.

Nothing is mentioned of the crimes which brought 1,534 convicts into the care of Dr Rennie, a medical officer in the penal colony at Freemantle, Western Australia, between 1 July and 31 December 1854. Almost all suffered from diseases of the skin, diseases of the digestive system or inflammatory eye disease: all, in Rennie’s opinion, ‘the results of overfeeding, assisted occasionally by a deficiency of vegetable matter’. Perusal of the daily diet of these men inclines to immediate endorsement of Rennie’s view. Observing an absence of such disease in the general population, the efficacy of purgatives, and the cure effected by solitary confinement on a reduced diet of bread and water, Rennie suggests a reduction in the diet. Despite opposition from his superior, Rennie’s suggestion was adopted and intake reduced from 59 to 46 ounces of solids per day, a course of action endorsed by reduced hospital admission rates 6 months later.

Throughout, the paper is littered with snide remarks and asides from the editor, e.g. ‘The report of Dr Rennie met with some opposition from his superior, Dr Galbraith, who evidently has no leanings to commonsense deductions’. Later the editor asks the rhetorical question ‘Why are convicts thus over replenished?’ His answer is: ‘The authorities find a body of men are more easily managed when well clad, well lodged, and supplied with more food than will satisfy their animal cravings’. David Cameron and other proponents of austerity worldwide should pause for thought although possibly, in their liaisons with Big Beverage, Big Food and Rupert Murdoch, they feel they have their bread and circuses. Lastly, the editor of the Journal of Public Health and Sanitary Review, in an early forerunner of the IJE’s Data Resource Profile series describes a data resource for the observational epidemiologist of his time—Box 2.