Income inequality and population health: A review and explanation of the evidence

Richard G Wilkinson\textsuperscript{a,*}, Kate E. Pickett\textsuperscript{b}

\textsuperscript{a}Division of Epidemiology and Public Health, University of Nottingham Medical School, UK
\textsuperscript{b}Department of Health Sciences, University of York, UK

Available online 13 October 2005

Abstract

Whether or not the extent of income inequality is a determinant of population health is still regarded as a controversial issue. We decided to review the evidence and see if we could find a consistent interpretation of both the positive and negative findings.

We identified 168 analyses in 155 papers reporting research findings on the association between income distribution and population health, and classified them according to how far their findings supported the hypothesis that greater income differences are associated with lower standards of population health. Analyses in which all adjusted associations between greater income equality and higher standards of population health were statistically significant and positive were classified as “wholly supportive”; if none were significant and positive they were classified as “unsupportive”; and if some but not all were significant and supportive they were classified as “partially supportive”. Of those classified as either wholly supportive or unsupportive, a large majority (70 per cent) suggest that health is less good in societies where income differences are bigger.

There were substantial differences in the proportion of supportive findings according to whether inequality was measured in large or small areas. We suggest that the studies of income inequality are more supportive in large areas because in that context income inequality serves as a measure of the scale of social stratification, or how hierarchical a society is.

We suggest three explanations for the unsupportive findings reported by a minority of studies. First, many studies measured inequality in areas too small to reflect the scale of social class differences in a society; second, a number of studies controlled for factors which, rather than being genuine confounders, are likely either to mediate between class and health or to be other reflections of the scale of social stratification; and third, the international relationship was temporarily lost (in all but the youngest age groups) during the decade from the mid-1980s when income differences were widening particularly rapidly in a number of countries. We finish by discussing possible objections to our interpretation of the findings.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Income inequality; Population health; Review; Social class; Area

Introduction

Whether or not the extent of income inequality in a society is a determinant of population health remains a controversial issue despite a large body of research.
Although the findings of a substantial majority of studies suggest that more egalitarian societies do have better health and longevity (Lynch, Smith, & Harper, 2004a; Subramanian & Kawachi, 2004), a minority conclude otherwise and several authorities remain skeptical as to whether inequality has any implications for population health (Deaton, 2003; Lynch et al., 2004a). To gain a clearer understanding of the evidence and the nature of the disagreement, we decided to review all the research reports published in peer reviewed journals, and then to see if we could arrive at an interpretation of them which made sense of both the supportive and unsupportive findings.

The review

We compiled a list of 155 published peer reviewed reports of research on the relation between income distribution and measures of population health. This is much the most comprehensive list of studies yet compiled: as well as containing all the eligible studies listed in three previous reviews of parts of the literature (Hsieh & Pugh, 1993; Lynch et al., 2004a; Subramanian & Kawachi, 2004), we also found 37 additional papers either by using electronic searches or through informal contacts. Several papers contained analyses at more than one level of aggregation (for example analyses of international data and data from states or regions within one country). The 155 papers contained 168 separate analyses. To facilitate comparison, we decided to classify findings using the same criteria as Lynch et al. (2004a). Thus, we classified analyses into three categories according to their findings after adjustment for whatever control variables authors thought appropriate. We classified them as “wholly supportive” if they reported only statistically significant associations between greater income inequality and poorer population health; as “unsupportive” if they found no statistically significant positive associations; and as “partially supportive” or “mixed” if some, but not all, of the associations they reported showed significant positive associations. [These categories correspond to those labeled “positive”, “negative” and “mixed” in Lynch et al. (2004a)]

Findings

Table 1 provides a summary of the 168 analyses according to classification and the type of area over which inequality was measured. Table 2 lists all the analyses included according to their classification. A tally of numbers showed 87 wholly supportive analyses, 44 partially supportive, and 37 unsupportive. Almost three-quarters of all analyses were classified as either wholly or partially supportive. Of all analyses classified as wholly supportive or unsupportive, 70 per cent were wholly supportive. Given that almost every paper reported many different measures of association, for different health variables in different age, sex or ethnic subgroups, it is notable that only nine (classified as partially or unsupportive) contained any measures of an association suggesting a health variable was significantly better where inequality was greater. In no analysis were such associations the predominant finding.

However, our aim in this paper is to go beyond these crude categories, which take no account of methodological quality or statistical power, and to try to gain a theoretically coherent overview of the literature as a whole. We shall proceed by drawing attention to various patterns in the findings and the likely implications of each. In doing so we will outline a consistent interpretation of most of the evidence, both supportive and unsupportive. We shall end by discussing possible criticisms and alternative interpretations.

<table>
<thead>
<tr>
<th></th>
<th>wholly supportive</th>
<th>Partially supportive</th>
<th>Unsupportive</th>
<th>Total</th>
<th>wholly supportive as per cent of all analyses excluding partially supportive (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nations</td>
<td>30 (11)</td>
<td>9</td>
<td>6</td>
<td>45 (11)</td>
<td>83</td>
</tr>
<tr>
<td>States, regions, cities</td>
<td>45 (13)</td>
<td>21</td>
<td>17</td>
<td>83 (13)</td>
<td>73</td>
</tr>
<tr>
<td>Counties, tracts, parishes</td>
<td>12 (2)</td>
<td>14</td>
<td>14 (1)</td>
<td>40 (3)</td>
<td>45</td>
</tr>
<tr>
<td>Total</td>
<td>87 (26)</td>
<td>44</td>
<td>37 (1)</td>
<td>168 (27)</td>
<td>70</td>
</tr>
<tr>
<td>International studies</td>
<td>Supportive studies</td>
<td>Mixed studies</td>
<td>Unsupportive studies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------------------</td>
<td>--------------</td>
<td>---------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Groves et al. (1983)</td>
<td>Ross et al. (2005)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hansmann and Quigley (1982)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kick and LaFree (1985)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Krohn (1976)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lee and Bankston (1999)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Legrand (1987)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Macinko et al. (2004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marmot and Bobak (2000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McIsaac and Wilkinson (1997)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pampel and Zimmer (1989)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pickett, Kelly, Lobstein, Brunner, and Wilkinson (2005)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pickett, Mookherjee, and Wilkinson (2005)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rodgers (2002)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Steckel (1983)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waldmann (1992)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wennemo (1993)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wilkinson (1992)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>30 studies, including 11 of homicide only</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>States, regions, metropolitan areas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. US states</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baron and Straus (1988)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blakely et al. (2001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daly et al. (2001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Holtgrave and Crosby (2004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blakely et al. (2002)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daly, Duncan et al. (1998)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diez-Roux et al. (2000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Holtgrave and Crosby (2003)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kahn, Tatham et al. (1998)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laporte (2002)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deaton and Lubotsky (2003)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McLeod et al. (2004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Muller (2002)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shi et al. (2003)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Huff-Corzine et al. (1986)\textsuperscript{H}

Kahn, Wise, Kennedy, and Kawachi (2000\textsuperscript{M})

Kaplan, Pamuk, Lynch, Cohen, and Balfour (1996)

Kawachi and Kennedy (1997)

Kawachi, Kennedy, Lochner, and Prothrow-Stith (1997)

Kennedy, Kawachi, Glass, and Prothrow-Stith (1998\textsuperscript{H})

Kennedy et al. (1999)

Kennedy et al. (1998\textsuperscript{M})

Loftin and Hill (1974\textsuperscript{H})

Pickett, Mookherjee, and Wilkinson (2005\textsuperscript{N})

Ross et al. (2000\textsuperscript{RC})

Shi (1999)

Smith and Parker (1980\textsuperscript{H})

Subramanian, Blackely, and Kawachi (2003), Subramanian and Kawachi (2003\textsuperscript{M})

Subramanian and Kawachi (2003\textsuperscript{M}, 2004\textsuperscript{N})

Wilkinson et al. (1998)

Wolfson, Kaplan, Lynch, Ross, and Backlund (1999\textsuperscript{N})

\textbf{25 studies, including 6 of homicide only}

2. Canadian provinces—studies are of USA unless otherwise noted

Daly et al. (2001)\textsuperscript{H}

3. UK regions

Ross, Wolfson, Dunn et al. (2000)

Weich, Lewis, and Jenkins (2001\textsuperscript{M}, 2002\textsuperscript{M})


Regidor, Navarro, Dominguez, and Rodriguez (1997)

Messias (2003\textsuperscript{L})

4. Spanish regions

De Vogli et al. (2005\textsuperscript{N})

Subramanian et al. (2003\textsuperscript{M})

Shmueli (2004\textsuperscript{N})

Blomgren, Martikainen, Makela, and Valkonen (2004\textsuperscript{N, M})

5. Brazilian states

6. Italian regions

De Vogli et al. (2005\textsuperscript{N})

Walberg et al. (1998)

Subramanian et al. (2003\textsuperscript{M})

Blakely et al. (2002\textsuperscript{M, RC})

Deaton and Lubotsky (2003\textsuperscript{L})

7. Russian regions

8. Chilean regions

9. Israeli regions

10. Finnish regions

11. Ecuadorean regions

Larrea and Kawachi (2005\textsuperscript{N, M})

Larrea and Kawachi (2005\textsuperscript{N, M})—\textit{Ecuador}

McLeod, Lavis, Mustard, and Stoddart (2003\textsuperscript{M})—\textit{Canada}

Metropolitan areas and cities—studies are of USA unless otherwise noted

Bailey (1984\textsuperscript{H})

Blakely et al. (2002\textsuperscript{M, RC})

Deaton and Lubotsky (2003\textsuperscript{L})

Balkwell (1990\textsuperscript{H})

Lochner, Pamuk, Makuc, Kennedy, and Kawachi (2001\textsuperscript{M})

Lynch, Smith, Harper et al. (2004\textsuperscript{N})

Mayer and Sarin (2005\textsuperscript{N, M})

Mellor and Milyo (2001)

Reagan and Salsberry (2005\textsuperscript{N, M})

Shi et al. (2004\textsuperscript{N})

Subramanian et al. (2001\textsuperscript{M})
Table 2 (continued)

<table>
<thead>
<tr>
<th>Supportive studies</th>
<th>Mixed studies</th>
<th>Unsupportive studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chiang (1999) — Taiwan</td>
<td>Sanmartín et al. (2003&lt;sup&gt;N&lt;/sup&gt;) — Canada</td>
<td>Sturm and Gresenz (2002&lt;sup&gt;M,U&lt;/sup&gt;)</td>
</tr>
<tr>
<td>Cooper et al. (2001)</td>
<td>Shi, Starfield, Politzer, and Regan (2002&lt;sup&gt;M&lt;/sup&gt;)</td>
<td></td>
</tr>
<tr>
<td>Kahn et al. (1999&lt;sup&gt;N&lt;/sup&gt;)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kennedy et al. (1991&lt;sup&gt;H&lt;/sup&gt;) — Canada</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loftin and Parker (1985&lt;sup&gt;H&lt;/sup&gt;)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loper (2004&lt;sup&gt;N&lt;/sup&gt;)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lynch et al. (1998)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Messner (1983&lt;sup&gt;H&lt;/sup&gt;)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ronzio (2003&lt;sup&gt;N&lt;/sup&gt;)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ronzio et al. (2004&lt;sup&gt;N&lt;/sup&gt;)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sanmartín et al. (2003&lt;sup&gt;N&lt;/sup&gt;)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shi and Starfield (2000&lt;sup&gt;M&lt;/sup&gt;, 2001&lt;sup&gt;RC&lt;/sup&gt;)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 studies, including 6 of homicide only</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baldani et al. (2004&lt;sup&gt;N&lt;/sup&gt;)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galea et al. (2003&lt;sup&gt;M&lt;/sup&gt;)</td>
<td>Brodish et al. (2000)</td>
<td></td>
</tr>
<tr>
<td>Gold, Kawachi, Kennedy, Lynch, and Connell (2001&lt;sup&gt;N&lt;/sup&gt;)</td>
<td>Fiscella and Franks (2000&lt;sup&gt;M&lt;/sup&gt;)</td>
<td></td>
</tr>
<tr>
<td>Messner and Tardiff (1986&lt;sup&gt;H&lt;/sup&gt;)</td>
<td>Franzini et al. (2001)</td>
<td></td>
</tr>
<tr>
<td>Massing et al. (2004&lt;sup&gt;N&lt;/sup&gt;)</td>
<td>Gold et al. (2004&lt;sup&gt;N,M&lt;/sup&gt;)</td>
<td></td>
</tr>
<tr>
<td>Muramatsu (2003&lt;sup&gt;N&lt;/sup&gt;)</td>
<td></td>
<td>Blakely, Atkinson, and O’Dea (2003&lt;sup&gt;M,U&lt;/sup&gt;) — New Zealand</td>
</tr>
<tr>
<td>Pattussi, Marcenes, Croucher, and Sheiham (2001&lt;sup&gt;N&lt;/sup&gt;) — Brazil</td>
<td>Hou and Chen (2003)</td>
<td></td>
</tr>
<tr>
<td>Shi et al. (2005&lt;sup&gt;N&lt;/sup&gt;)</td>
<td>LeClere and Soobader (2000&lt;sup&gt;M&lt;/sup&gt;)</td>
<td></td>
</tr>
<tr>
<td>Soobader and LeClere (1999&lt;sup&gt;M&lt;/sup&gt;)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stanislaw et al. (1999)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subramanian, Delgado, Jadue et al. (2003) — Chile</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Szwarcwald, Bastos, Viacava et al. (1999&lt;sup&gt;H&lt;/sup&gt;) — Brazil</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small areas: Counties, Neighborhoods, Zip codes, Census Tracts, Parishes, etc.—studies are of USA unless otherwise noted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 studies</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This table contains the 98 studies reviewed by Lynch et al. (2004a, b) (including all peer-reviewed studies in Subramanian and Kawachi, 2004) and the 20 studies of income inequality and homicide reviewed by Hsieh and Pugh (1993), as well as 37 additional studies which were omitted from earlier reviews or have been published more recently. These 155 papers contain the 168 separate analyses listed here. To aid comparisons with other reviews the following notation is used in the table.

- \( N \): Denotes new and previously omitted studies \((n = 37)\).
- \( H \): Denotes studies where homicide was the only outcome \((n = 25)\).
- \( R \): Denotes studies that we re-classified from the Lynch et al. (2004) review, as results did not match the classification given previously \((n = 6)\).
- \( M \): Denotes multi-level, rather than ecological studies, although not all of these use appropriate multi-level statistical techniques.
- \( U \): Denotes “unsupportive” studies that reported an unadjusted association between income inequality and health that was removed when inappropriate control variables were added to models.

### 12 studies, including 2 of homicide only

- Robert and Reither (2004\( ^{N,M} \))
- Sohler et al. (2003)
- Szwarcwald, Andrade, and Bastos (2002\( ^{B} \))—Brazil
- Veenstra (2002a)—Canada

### 14 studies

- Lorant, Thomas, Deliege, and Tonglet (2001)—Belgium
- Muntaner et al. (2004\( ^{N,M,U} \))
- Osler et al. (2002\( ^{M,U} \))—Denmark
- Shibuya et al. (2002\( ^{M} \))—Japan
- Szwarcwald, Bastos, Viaesva et al., (1999)—Brazil
- Veenstra (2002b\( ^{U} \))—Canada
- Wen et al. (2003\( ^{M} \))
Explanations of findings

The size of area

Table 1 shows the per centage of analyses classified as either wholly supportive or unsupportive according to whether they were international analyses using data for whole countries, whether their data were for large subnational areas such as states, regions and metropolitan areas, or whether they were for smaller units such as counties, census tracts or parishes. The proportion of analyses classified as wholly supportive falls from 83 per cent (of all wholly supportive or unsupportive) in the international studies, to 73 per cent in the large subnational areas, to 45 per cent among the smallest units.

The tendency towards more positive findings in the largest areas compared to the smallest is important and has already received some attention in the literature (Subramanian & Kawachi, 2004). The same pattern was observed in a review of studies of the relation between homicide and inequality (Hsieh & Pugh, 1993). It was also shown in a study by Franzini, Ribble, and Spears (2001) comparing the strength of association among the counties of Texas according to population size. Wilkinson (1997) has argued that income inequality in small areas is affected by the degree of residential segregation of rich and poor and that the health of people in deprived neighbourhoods is poorer not because of the inequality within their neighbourhoods, but because they are deprived in relation to the wider society. If that is what matters, then it is to be expected that inequality will only be sensitive to this broader pattern of deprivation if inequality is measured across the wider framework in which the relevant social comparisons are made. The fact that measures of inequality made across larger areas are more closely related to health bears out this point.

This takes us to a familiar and difficult question: if inequality is important, what are the relevant social comparisons? Rather than suggesting any new causal processes or framework of comparisons which affect health, it is more parsimonious to suggest that inequality is related to health insofar as it serves as a measure of the extent of the same processes of class differentiation and social distances in a society which are responsible for class differences in health. The processes which lead to class differences in health are likely to be closely related to those which explain why greater inequality is related to worse health. If that is right, then the question becomes one of the scale of the social units in relation to which one’s class position is defined. The broad impression is that social class stratification establishes itself primarily as a national social structure, though there are perhaps also some more local civic hierarchies—for instance within cities and US states. But it should go without saying that classes are defined in relation to each other: one is higher because the other is lower, and vice versa. The lower class identity of people in a poor neighbourhood is inevitably defined in relation to a hierarchy which includes a knowledge of the existence of superior classes who may live in other areas some distance away.

Control variables

If, in the association between income inequality and health, we are seeing the effects of the scale of social class stratification, of bigger or smaller class differences, then it is hard to decide what are legitimate variables to use as controls when analysing that association. What is part of class and what is not? If we had classified analyses by their findings before the use of control variables, 21 of the 37 studies we have listed as unsupportive of an association between income distribution and health actually started off with supportive findings but then lost them as a result of the various control variables they used.

A wide variety of control variables have been used, including the per cent without a high school education (Muller, 2002), individual income (reviewed in Subramanian & Kawachi, 2004), perceived control (Bobak, Pikhart, & Rose, 2000), ethnicity (Blakely, Atkinson, & O’Dea, 2003; Deaton & Lubotsky, 2003), social capital (Veenstra, 2002a), and unemployment (Shi & Starfield, 2000). Subramanian and Kawachi (2004) have a useful discussion of possible confounding by education, individual income, race, and regional effects. To know which are genuine confounders and which are pathway, or mediating, variables means—for us—knowing what is part of social class and what is not. If ethnicity is related to health because it is a proxy for a classification by class, then perhaps we should not control for ethnicity.

Similarly, if Sahlins (1974) was right to say “Poverty is not a certain small amount of goods (but)...a relation between people...a social status...an invidious distinction between classes...” (p. 37), then it may be misconceived even to control for individual income. It has been suggested a number of times that the social gradient in health within countries is primarily a gradient in relative income, or social status, rather than a reflection of absolute material living standards. Marmot (2004) and others (Charlesworth, Gilfillan, & Wilkinson, 2004; Singh-Manoux, Adler, & Marmot, 2003; Wilkinson, 2005) have argued that the relation between health and social status may be primarily a reflection of the effects of social position itself. This view is strongly supported by the fact that the international relation between Gross National Income per capita and life expectancy not only grows progressively weaker as countries get richer, but disappears altogether among
the richest (Marmot & Wilkinson, 2001; Wilkinson, 1997; see also Discussion section below and Table 3). Although there are clearly aspects of rising material living standards which contribute directly to better health even in the richest countries, we believe their effects are not apparent in international comparisons either because they are relatively small, or because they are offset by other factors. It is surely unwise to ignore this evidence and conduct analyses as if any given level of individual income had the same effect on health regardless of the social status it buys. If a person’s income is a marker of their social position, then adjusting inequality effects for individual income may be like controlling measures of class stratification for individual social status differentiation.

However, even if this objection to controlling for individual income is ignored, it appears that despite often using small areas, analyses of inequality which use multilevel methods have usually been able to identify inequality effects even after controlling out the effects of individual income (Subramanian & Kawachi, 2004).

That there are so many correlates of income distribution is consistent with our view that income inequality is an indicator of the extent of social stratification and points to the need to think carefully about which factors are confounders and which are mediators in this relation.

Discussion

Taking account of the size of the area and the use of control variables reveals a high degree of consistency in the research findings. Thus, if we confine our attention to the 128 analyses which use data for areas the size of metropolitan areas or larger, only 23 fail to find some support for the hypothesis. If we were to reclassify analyses on the basis of results before the use of potentially problematic control variables (including individual income in multilevel models), then only eight (6 per cent) of the 128 analyses would remain classified as unsupportive.

The extent of social class divisions may vary substantially from country to country: we know that human beings have lived in every kind of society from the most egalitarian (Erdal & Whiten, 1996) to the most tyrannical. Given the importance of the social class gradient in health, the societal differences in the extent of social class inequality are not something we can ignore. Several variables may provide rough measures of the extent of social class differentiation. These might include educational differences, inequalities in the distribution of power or wealth, and perhaps scores on the Social Dominance Orientation Scale (Sidanius & Pratto, 1999). However, income inequality is likely to be one of the most widely applicable. Although it may not be the best measure of social hierarchy in all cultures, the fact that dominance hierarchies (in human societies as among animals) are fundamentally about privileged access to scarce resources, may mean that differences in income and/or wealth are particularly apposite indicators of rank difference across cultures. But even if there are better measures—perhaps ones which include measures of ownership of assets, income inequality has the substantial advantage that it is collected for numerous other purposes and so can be used in secondary data analyses. We hope that this interpretation may bring us closer to the thinking of others working in this field and narrow the area of controversy (Deaton, 2003; Lynch et al., 2004a, b).

The interpretation of the evidence which we have put forward has the advantage of simplicity. Instead of suggesting that inequality is a new risk factor for health, it may be telling us more about the already widely recognised health effects of socioeconomic status and class. It may simply be that larger class differences lead to a steeper social gradient in health, but it could also be that a more unequal society becomes more dominated by status competition and class differentiation and suffers a more widespread health disadvantage as a result. It is already clear from studies designed to illuminate this issue, that the health disadvantages of inequality are not confined to the poorest (Kennedy, Kawachi, & Glass, 1998; Lochner, Pamuk, & Makuc, 2001).

Counter arguments

What are the objections and counter arguments to these interpretations? The most important is undoubtedly that income is related to health because it is a determinant, not of class differences or social position,
but of material living standards which it is claimed continue to exert a major direct influence on health. However, although raising absolute material living standards continues to be important in developing countries, among the 25 or 30 richest countries there is no relation between Gross National Income per capita and health (Marmot & Wilkinson, 2001; Wilkinson, 1997)—even though curves are sometimes still fitted to the data to suggest otherwise (Lynch et al., 2004a). If absolute living standards were overwhelmingly important, it would be difficult to understand why, despite having a median income four times as high, life expectancy among black men in the USA was 9 years shorter than for men in Costa Rica (Marmot & Wilkinson, 2001). Similarly Greece, with half the average real income of the US has, like many other developed countries, better life expectancy. Indeed, looking at the relations between life expectancy, income distribution, and Gross National Product per capita among 21 rich countries, we found only income distribution had significant independent effects on life expectancy. Gross National Product per capita showed no sign at all—regardless of statistical significance—of an independent association. The unweighted partial correlations are shown in Table 3.

We emphasise that rather than meaning material factors can be ignored, psychosocial pathways provide a major new route through which they affect health. But the psychosocial link changes the nature of the relationship, particularly making us more sensitive to relativities and to the social connotations of material differences. And insofar as psychosocial risk factors have drawn attention to the importance of the social environment to health, we believe that the social structure is built substantially on material foundations—hence the importance of inequality.

Alongside reasons such as these for discounting the continued primacy of material influences on health in the rich countries, the relation between income inequality and homicide (Hsiev & Pugh, 1993) shows that inequality has powerful psychosocial and behavioural effects (Wilkinson, 2004). Indeed, the relation with homicide appears to be part of a more general effect which inequality has on social capital and the quality of social relations, both of which might be expected to influence health. As well as having higher levels of violence, people in more unequal societies also seem less likely to trust others and less likely to be involved in community life (Wilkinson, 2005). Other examples of behavioural effects of inequality include higher teenage pregnancy rates (Gold, Connell, & Heagerty, 2004; Pickett, Mookherjee, & Wilkinson, 2005a) and more obesity (Pickett, Kelly, & Lobstein, 2005b). While it is easy to understand why inequality and increased status differentiation should affect both health and behaviour through psychosocial stress (Wilkinson, 2005), it would be harder to argue that material factors could affect behaviour directly.

A potential second criticism of our interpretation of the studies of income inequality and health is the view, derived from relative deprivation theory, that people compare themselves with near equals (Runciman, 1966). No doubt this explains why some have chosen to measure inequality in small areas, intending to capture the effects of these social comparisons rather than focusing on the wider structure of inequality. However, people's judgment of who their near equals are is dependent on a prior recognition of their class identity and where they fit into the wider class structure. The logic of what is happening in a dominance hierarchy when comparisons appear to be between near equals was spelt out by Sapolsky when describing conflict over rank among baboons in the Serengeti. They also seem to compare themselves with near equals:

A pattern emerged that has grown familiar to me over the years. When you look at the frequencies of dominance interactions, the typical pattern you see is that, for example, number 4 is having his most interactions with 3 and 5, losing to the former, defeating the latter. Number 17 mostly interacts with 16 and 18 (Sapolsky, 2001, p. 95).

As Sapolsky points out, there is no point in animals fighting those which are clearly much higher or lower in the dominance hierarchy: because the outcome is predictable the subordinate recognises its inferiority and avoids making a challenge. The point concerns the logic of ranking systems, not whether humans are like baboons. So, for instance, when the results of a race are announced, a competitor placed second might claim he was really first, or the eighth might claim to have come seventh, but the eighth is unlikely to claim he really came first. However, that rank is only contestable among near equals does not mean that the rest of the social hierarchy is irrelevant: ignoring it may result in ridicule or injury. Who counts as a near equal is merely the converse of recognising who is not a near equal. Similarly, our recognition of our class status is constituted primarily by our recognition of uncontestable status differences.

The logic of ranking systems which leads baboons to conflict mainly with near neighbours, is the same logic which leads us to exclaim of other people “Who do they think they are?”—not so much at those we accept as higher status, but at our equals who pretend to superiority. To maintain rank we have to pay attention to the fine grain of social status: that means keeping up with the Joneses. And even if we live in a neighbourhood in which everyone is poor, that does not mean that we are unaware of those in richer neighbourhoods whose existence defines our lower status and relative poverty.
Sociologists agree that class identity is defined by position in the wider society (Bourdieu, 1984; Canadine, 1998). Because classes are mutually defining, researchers cannot identify the effects of social status differentiation, and of our class identity within it, in a statistical context (such as a small, residentially segregated, neighbourhood) which excludes much richer or poorer neighbourhoods.

Sometimes discussion of these issues is further confused by arbitrarily labelling an income variable as absolute rather than relative income. This is important because it is often assumed that a relation between health and absolute income reflects the direct effects of material living standard on health—regardless of the rest of society, whereas any effects of relative income are assumed to reflect psychosocial processes contingent on social status or social comparisons. But in different analyses income differences which are called absolute in one context may be called relative in another (Drukker, Feron, & van Os, 2004; Hsieh & Pugh, 1993). Income differences which make up income inequality within large areas can of course be broken down into inequalities within and between smaller constituent areas (Franzini et al., 2001; Lobmayer & Wilkinson, 2000; Soobader & LeClere, 1999). The smaller and more numerous the constituent areas used, the more of the income inequality in the larger areas gets converted into income differences between the small areas and the less that remains as inequality within them. That conversion can be done almost *ad infinitum* until, at the limit, all inequality is reduced to differences in income between the smallest (single household) areas. How much of the income differences in a society are analysed as differences in relative income within, and how as differences in absolute income between, small areas is inevitably a by-product of the choice of units of analysis.

In this situation what tends to happen is that people look for an effect of income inequality in small areas, find it weak or non-existent, and report an association between health and the average income of the small areas. Instead of interpreting that as an effect of low income *relative* to the wider society, it is interpreted as evidence of a direct effect of material living standards.

If health among the developed countries is unrelated to the big differences in material living standards between countries, then why should the same differences in living standards have an effect on health when they occur within the same society? The truth is surely that income is related to health where—as within countries—it serves as a marker for position in the national structure of class inequality.

For those who still prefer to believe in the primacy of the direct—over the psychosocially mediated—effects of material living standards on health, and interpret the association between income inequality and health as a reflection of a curvilinear relation between individual income and health (Gravelle, 1998), there is one more major obstacle. Although it once seemed plausible that the curvilinear relation between individual income and health reflected a tendency towards diminishing health returns to increased income, so that any given sum of money made more difference to the health of the poor than the rich, such a pattern cannot explain the findings of this review. If the relation between individual income and health resulted simply from the healthfulness of whatever material standard of living a given income can buy, then these effects would be just as apparent if inequality was measured in small areas as in large areas. According to that interpretation it is the incomes themselves which count—regardless of their social meaning. We would then be left with no explanation of why the overwhelming majority of studies which measure inequality in large areas do report associations whereas only a minority of those using data for small areas do. This confirms our view that it is mistaken to control for individual income in multilevel models because it amounts to controlling the effects of income (and class) inequality for the effects of individual social status. It also seems likely that if we were dealing with the material effects of individual income they would be harder to control away than the effects of class inequality and social differentiation for which we believe income inequality is merely a proxy.

**International analyses**

A group of results which our interpretation does not explain, is a small group of international studies using data from between the later 1980s and the mid 1990s. Although 30 of the 45 international studies are classified as wholly supportive and a further nine as partially supportive, there are nevertheless four classified as unsupportive which cannot be attributed to the use of inappropriate control variables, and so run counter to the interpretation we have advanced in this paper.

During the 1980s and early 1990s, when income differences were widening particularly rapidly in many countries, much of the relation between inequality and mortality among rich countries temporarily disappeared (Gravelle, Wildman, & Sutton, 2002; Judge, 1995; Mellor & Milyo, 2001; Wildman, Gravelle, & Sutton, 2003). What happened varied by age group (Judge, Mulligan, & Benzeval, 1998; Lobmayer & Wilkinson, 2000; Lynch, Smith, & Hillemeyer, 2001). While infant mortality rates remained consistently related to inequality, the relation was entirely lost with death rates among the middle aged and elderly. The relation was clear earlier (Wilkinson, 1992) and has now reappeared (De Vogli, Mistry, & Gnesotto, 2005). It is noticeable that the publication dates of the positive international studies tends to be earlier than the mixed and negative ones.
The three most likely explanations of why the international relationship temporarily disappeared—except with health in the youngest age groups—are: first, that it was affected by the downward shift in the age distribution of relative poverty which took place in many countries. Although relative poverty had been more common among the elderly, it became more common among young families with children (Kangas & Palme, 2000). Second, in one country after another, death rates among older people began an unprecedentedly rapid decline, particularly from cardiovascular mortality. Perhaps as a result of differences in the uptake of both primary and secondary forms of prevention, the timing of the onset of the decline was earlier in some countries than others (Menotti et al., 2003). It may therefore have affected international comparisons of mortality before the decline became general. Interestingly, three of the very few statistically significant negative associations (greater inequality related to better health) reported in any of the studies, were international studies among rich countries of death rates among the elderly. The third possible explanation is that the changes in income distribution may have had lagged effects on mortality, particularly on mortality at later ages (Subramanian & Kawachi, 2004). Although income differences widened particularly rapidly in many countries during the 1980s and 1990s, the relationship between income distribution and infant mortality remained throughout, perhaps because lag times are shortest at youngest ages. Mayer and Sarin (2005) found that neonatal mortality rates are significantly more closely related to current inequality than to inequality 5 years earlier, whereas Subramanian and Kawachi (2004) found the strongest associations with adult health after a lag of 10 or 15 years. Health among adults may then reflect the inequalities of the past.

Mechanism

Low social status and the quality of the social environment are both known to affect health (Berkman & Kawachi, 2000; Marmot & Wilkinson, 1999). Not only are more unequal societies likely to have a bigger problem of low social status, but there is now substantial evidence to suggest that inequality is socially corrosive, leading to more violence, lower levels of trust, and lower social capital (Wilkinson, 2005). Psychosocial factors, many of which are associated with low social status, are known to affect health partly through direct physiological effects of chronic stress (Brunner & Marmot, 1999), and partly through their influence on health related behaviour. Marmot (2004) has argued that low social status is stressful because it reduces people's control over their lives and work. Others have argued that low social status is stressful because people are made to feel looked down on, devalued and inferior (Charlesworth et al., 2004; Wilkinson, 2005). Both suggestions are borne out by a recent review of the most salient stressors affecting cortisol responses (Dickerson & Kemeny, 2005). What matters most are uncontrollable threats to ones social esteem, value and status. As well as explaining the relationship between health and inequality, this approach is concordant with the suggestion that inequality is related to violence because the increased burden of low social status makes more people feel disrespected. Feeling disrespected, put down and humiliated is much the most frequent trigger to violence (Gilligan, 1996; Wilkinson, 2004).

Conclusions

Our interpretation of 168 analyses of the relationship between income inequality and health is that income distribution is related to health where it serves as a measure of the scale of social class differences in a society. In small areas, where income inequality is unlikely to reflect the degree of social stratification in the wider society, it is—as Table 1 shows—less likely to be related to health. The overwhelmingly positive evidence from studies of larger areas suggests that this interpretation is correct. The fact that social stratification is such a fundamental feature of social organisation explains why there are so many socioeconomic factors correlated with inequality. Many will function, like income inequality itself, as other proxies for the extent of social stratification or socioeconomic inequality. Others still may be mediating or pathway variables.

The methods researchers have used to test the hypothesis that greater inequality is associated with poorer population health have reflected many different assumptions about the mechanisms involved. In effect, a whole family of quite different hypotheses about income distribution and health have been tested. The two most important kinds of differences between the tests are those we have discussed: first, the different sizes of areas in which people have thought inequality most likely to be salient; and second, in what are regarded as legitimate control variables. Studies which have analysed data for areas as small as parishes, and controlled for things as closely related to class as education differentials, have helped clarify how income inequality does not work. Similarly, the suggestion that the per cent of the population who are black explains away the income inequality relation at state (if not county) level in the USA, has been regarded by some as a falsification of the inequality hypothesis (Deaton, 2003); but we think it comes closer to being a confirmation of the underlying view that what matters is the extent of social class differentiation. No one suggests that it is blackness itself which matters. Rather it is the social meaning attached
to it—the fact that it serves as a marker for class and attracts class prejudice—which leads both to worse health and to wider income differences. Future tests of the theory that the extent of class inequality is a determinant of population health must test its most plausible form.

One of the most important points to come out of this analysis is that it looks as if there are fundamentally important and measurable differences in the extent of socioeconomic stratification in different societies. While income distribution is a convenient and widely applicable measure, we hope that better ones may be found.

We recognise that some have been reluctant to accept the involvement of psychosocial pathways to ill health and argue that differences in health—such as the social gradients in health or the relation between population health and inequality—are the direct and unmediated effects of exposure to different material circumstances. However, if the argument were to move on to explanations of the range of behavioural outcomes which also show social gradients and relations with inequality, the psychosocial mediation of behaviour is undeniable. So, for instance, in the very well-established relation between violence and inequality, the causal chain must run all the way from the material facts of inequality to the psychosocial effects which lead to violence. The same is presumably true of other behavioural outcomes—such as teenage pregnancy, obesity and trust—which seem to be related to inequality as well as showing social gradients. If psychosocial processes consequent on low social status are recognised as having health effects, then it seems likely that they will also have behavioural consequences. Perhaps there are some pathways common to health and many of the social problems known to be more common in poorer areas. The possibility that we shall arrive at a general theory of social gradients capable of explaining (and policy prescriptions capable of reducing) a wide range of social ills, is obviously a worthwhile objective.

References


Kahn, R. S., Wise, P. H., Kennedy, B. P., Kawachi, I., et al. (2000). State income inequality, household income, and


