Mortality Comparisons 'At a Glance': A Mortality Concentration Curve and Decomposition Analysis for India

John Creedy
Victoria University of Wellington, Wellington, New Zealand
S. Subramanian
*Independent Scholar; formerly, Madras Institute of Development Studies,
Chennai, India*

Abstract

This paper uses the concept of the Mortality Concentration Curve (*M*-Curve), which plots the cumulative proportion of deaths against the corresponding cumulative proportion of the population (arranged in ascending order of age), and associated measures, to examine mortality experience in India. A feature of the *M*-curve is that it can be combined with an explicit value judgement (an aversion to early deaths) in order to make welfare-loss comparisons. Empirical comparisons over time, and between regions and genders, are made. Furthermore, in order to provide additional perspective, selective results for the UK and New Zealand are reported. It is also shown how the *M*-curve concept can be used to separate the contributions to overall mortality of changes over time (or differences between population groups) to the population age distribution and age-specific mortality rates.

Keywords. Mortality Curve, Mortality-inefficiency measure, Crude Death Rate, Lorenz Curve, Gini Coefficient, Age-distribution of population, Age-specific death rates, *M*-Curve comparisons, Decomposition: age and fatality effects, Decomposition: mean and dispersion effects
JEL Classification. D30, D63, I31, J17

1 Introduction

The crude death rate (CDR) within a country or region, defined as a weighted sum of age-specific death rates with weights given by the age-group population shares, is known to be inadequate for comparison purposes. It does not provide a clear summary of mortality experience, in view of the complicating role of the age distribution (which itself is influenced by earlier mortality characteristics). Furthermore, as a 'purely statistical' phenomenon, it is not possible to use the CDR to make judgements about whether mortality has 'improved' or is 'preferred' in some sense. In the context of income inequality comparisons, statistical indices of dispersion

have been replaced by measures that have a solid foundation in explicitly-stated value judgements. This means that independent judges can at least appreciate whether differing views arise because those judges hold different value judgements. Furthermore, it is possible to determine whether a change is widely judged to be an improvement, even if value judgements were to differ to some extent among judges.

In the present context, the question therefore arises of how explicit value judgements can be used in determining whether changes in mortality are regarded as an 'improvement' or 'worsening', such that a hypothetical judge can be said to prefer one situation over another. This question has been considered by Creedy and Subramanian (2022a) who, taking inspiration from the famous Lorenz curve and associated inequality literature, introduce the concepts of mortality and generalised mortality curves, or M- and GM-curves respectively. They show how the introduction of a value judgement – in the form of an 'aversion to early deaths' – allows normative comparisons using a 'loss function' defined in terms of the CDR and an 'inefficiency' measure, I_M. The latter, expressed in terms of an area in the M-curve diagram, reflects the 'wastefulness' of early deaths (the loss of 'life years'). The ability to make normative comparisons is a significant advantage, in addition to the ability of the M-curve to reveal 'at a glance' the main mortality characteristics of a group.

The purpose of the present paper is to use this new apparatus to examine mortality experience in India. Comparisons over time, and between regions and genders, are made. Furthermore, in order to provide additional perspective, selective results for the UK and New Zealand are reported. It is also shown how the M-curve concept can be used to separate the contributions to overall mortality of changes over time (or differences between population groups) to the population age distribution and age-specific mortality rates. Rather than involving the use of a standard age distribution, two additional artificial M-curves are used in which each relevant age distribution is matched with its 'opposite' set of mortality rates. This leads to a more formal decomposition analysis of changes, or differences, in the CDR following the general approach advocated by Shorrocks (2013).

Section 2, based on Creedy and Subramanian (2022a), briefly explains the M-curve and associated concepts. A number of Indian comparisons are then made in Section 3. Section 4 decomposes changes in the CRD into two components, those of the population and age-specific mortality rates. A further decomposition, this time of changes in the welfare, or loss function, is examined in Section 5. Conclusions are in Section 6.

For readers who may not be familiar with some of the demographic concepts used in this paper, and their relationship to elements of the income-inequality literature, Appendix 1 provides a brief glossary of relevant terms.

2 The Mortality Concentration Curve

The mortality concentration curve, or *M*-curve, is obtained by plotting the cumulative proportion of total deaths against the corresponding cumulative proportion of the population, as follows. Suppose a population is divided into K age-groups, the groups being indexed in ascending order of age, by $j = 1, ..., K$. Let p_j be the proportion of the total population, and q_j the proportion of all deaths, accounted for by members of the j th age-group, $j=1,...,K$. Then, for every $j \in \{1, ..., K\}$, the cumulative proportion of the population of age not exceeding the upper limit of the jth age-group is given by $P_j \equiv \sum_{i=1}^{j} p_i$; and the cumulative proportion of deaths accounted for by those of age not exceeding the upper limit of the jth age-group is given by $Q_j \equiv \sum_{i=1}^{j} q_i$.

An example is given in Fig. 1, showing the extent to which deaths are concentrated among the aged members of the population. If the age-specific mortality rate is identical for all ages, the curve follows the upward sloping diagonal line, whatever the form of the age distribution.

Because the ranking is by age rather than age-specific mortality, the curve need not be convex and always below the diagonal. High infant mortality can cause the curve to begin concave and above the diagonal, eventually moving below the diagonal and becoming convex. Mortality and population data are usually available for a number of age groups, so that in practice the curve consists of a number of piece-wise linear segments.

Statements about a particular *M*-curve representing a 'preferred outcome', when compared with another curve, must be based on explicitly-stated value
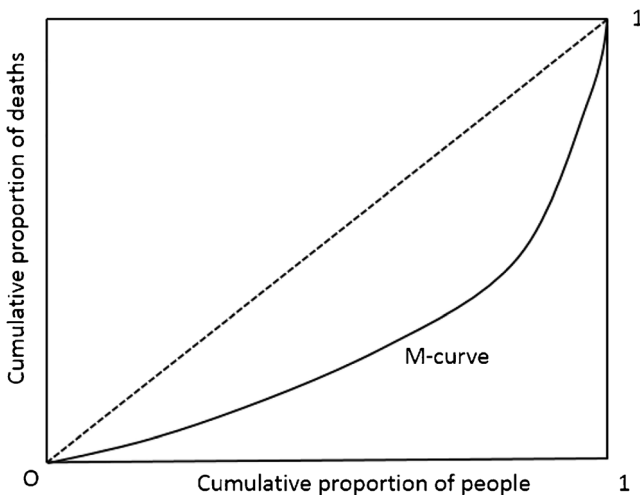

Figure 1: A Hypothetical *M*-Curve

judgements about mortality.[1] The simple judgement that greater loss is attached to a death, the lower the age at which it occurs, is explored here (assuming there are no other relevant individual characteristics). Hence, the worst situation is the unsustainable one in which deaths all take place in the youngest age-cohort. The 'best' distribution is one in which all deaths occur at the biological maximum age. These corresponding M-curves are respectively denoted M_W, which follows the left-hand side and top of the box, and M_B, which follows the base of the box and the right-hand side. (That is, the M_W curve is obtained as the graph of the function $M_W(x) = 1 \ \forall \ x \in [0, 1]$, and the M_B curve as the graph of the function $M_B(x) = 0 \ \forall \ x \in [0, 1)$ and $M_B(1) = 1$.) Early deaths are regarded as a 'waste of life-years': hence the term 'inefficiency' is used. A quantitative measure of mortality inefficiency, I_M, corresponding to the Gini inequality measure, may be defined as the area between the observed M-curve and the M_B rectangle.[2] The computation of I_M can be done using the standard trapezoidal approximation. Thus, for P_j and Q_j, for $j=1,...,K$ (with $P_0 = Q_0 = 0$, and $P_K = Q_K = 1$), as defined above, inefficiency, I_M, is the sum of the areas of a number of trapeziums, given by $I_M = (1/2) \sum_{i=1}^{K} (P_i - P_{i-1}) (Q_i + Q_{i-1})$: this expression is familiar from the computation of the Gini coefficient of inequality employing grouped data on the distribution of income.

   If one country has a mortality curve that is everywhere closer to M_B than that of another country, the former unequivocally displays less inefficiency. If the CDRs of two countries are identical, the value judgement discussed above implies that the former country is preferred to the other. If CDRs differ, an explicit trade-off is involved in making overall judgements. For income distribution comparisons, a similar problem arises using the Lorenz curve. Shorrocks (1983) showed that when arithmetic mean incomes of the two distributions differ, the appropriate concept is that of the 'Generalised Lorenz (GL) curve', in which the values on the vertical axis of the Lorenz curve are multiplied by arithmetic mean,μ. The Gini measures was initially defined in terms of areas within the Lorenz curve. However, it also arises from the adoption of a 'social

---

[1] In the Lorenz curve context, the statement that one curve has less inequality than that of another distribution if it is everywhere closer to the upward-sloping diagonal 'line of equality' is founded on the value judgement summarised by the 'Principle of Transfers'. This states that a transfer from a richer to a poorer person, which does not affect their ranks, is an improvement: such a transfer necessarily moves the Lorenz curve closer to the diagonal.

[2] The Gini inequality measure 'normalises' the relevant area by dividing by the area contained within the extremes of inequality and equality. In the present context this area is unity.

welfare function' expressed as the Borda rank-order-weighted sum of incomes. [3] This is combined with the class of inequality measures defined as the proportional difference between μ and an 'equally distributed equivalent' income (the equal income giving the same social welfare as the actual distribution). This gives rise to an 'abbreviated' welfare function, $W = \mu(1 - G)$, which is itself equal to the equally distributed equivalent income, and makes the trade-off between 'equity and efficiency' explicit. [4]

In the present context, a loss function is needed, where loss is captured by both the crude death rate, D, and the inefficiency of the distribution of deaths, I_M. An abbreviated loss function is thus $D^* = D(1 + I_M)$: the loss is an increasing function of each of its arguments, D and $(1 + I_M)$. When I_M is zero, the loss is simply D. A Generalised Mortality curve, or GM-curve, can therefore be derived from the M-curve, by first shifting the M-curve up by the crude death rate, D, and then scaling the M-curve by multiplying by D. An example is shown in Fig. 2, where the area under the curve is a sum of A and B. Given the definition of I_M, area A is equal to DI_M, while Area B is equal to D. The sum of the two areas is thus $D(1 + I_M)$, which is the abbreviated loss, D^*. [5]

3   Selected M-Curve Comparisons for India

This section uses the M-curve approach to examine a number of Indian comparisons. Further context is added by making comparisons with selected UK and NZ data. [6] The examples demonstrate the additional insights provided by the M-curve methods, as well as the immediacy of results. First, Table 1 reports values of the CDR, D, the inefficiency measure, I_M, and the welfare loss, D^*, for each population group considered. References to these measures are made in the following discussion of associated M-curves.

---

[3] The 'Borda' rule, widely employed in social choice theory, is a rule for aggregating individual rankings of alternatives into a collective ranking by assigning a score to each alternative which is the sum of that alternative's ranks in all individuals' orderings of the alternative. The rule has also been employed in deriving real-valued measures of inequality, such as the Gini coefficient, from an underlying social welfare function which is a weighted sum of individual incomes, the weights being the 'reverse' ranks of the incomes (so that a lower income receives a higher weight, thus rendering the welfare function sensitive to 'equity' considerations); see, for example, Sen (1973).

[4] To link this to the Generalised Lorenz curve, it is necessary only to recognise that the area under the Lorenz curve is $(1 - G)/2$, so that the area under the Generalized Lorenz curve is that of the Lorenz curve scaled by μ, and is thus $\mu(1 - G)/2$: the above abbreviated welfare function is simply twice the area under the GL curve (Bishop *et al.*, 2009).

[5] This also suggests, by analogy with the income distribution context, that the area under the Generalized M-curve reflects an 'optimally-distributed equivalent death rate', just as the area under the Generalised Lorenz curve is the equally-distributed equivalent income.

[6] The data sources for all the tables and figures are listed in Appendix 2.
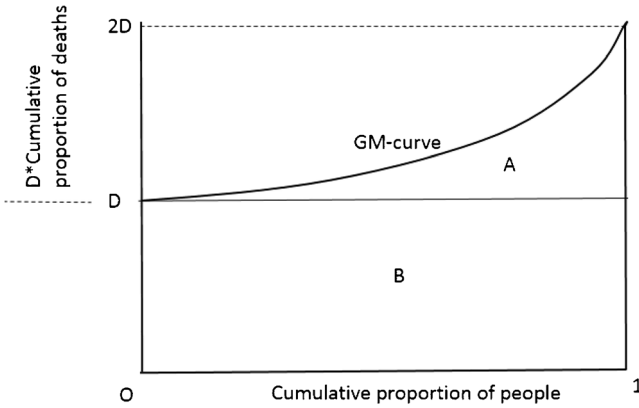
Figure 2: A Generalised Mortality Curve

Figure 3 shows curves, almost sixty years apart, for Uttar Pradesh (UP) in 1953 and 2001. In terms of human development indicators, UP is one of the most 'backward' States in the Indian Union: it is one of the infamous 'BIMARU' States—Bihar, Madhya Pradesh, Rajasthan and UP, 'Bimaru' in

Table 1: Summary Measures of Crude Death Rates, Inefficiency and the Loss Function

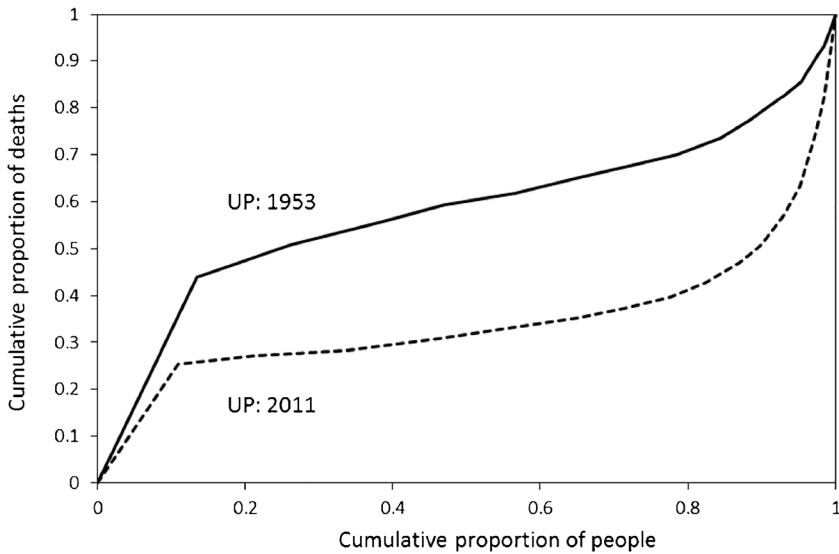| Population Group	D	I_M	D^*
UP 1953	21.10	0.578	35.30
UP 2011	7.72	0.333	10.29
Kerala 2011	7.10	0.123	7.97
India 2011: M	7.77	0.253	9.73
India 2011: F	6.29	0.297	8.16
India 2011: Rural	7.68	0.285	9.87
India 2011: Urban	5.77	0.227	7.08
UK 1951	12.40	0.149	14.30
NZ 2019: M	7.58	0.110	8.41
NZ 2019: F	7.01	0.089	7.63
NZ 2019: All	7.30	0.118	8.16
India 2011: All	7.10	0.268	9.00

Figure 3: *M*-Curves for Uttar Pradesh 1953 and 2001

Hindi meaning 'sick. Nevertheless, in the period from 1953 to 2011, the State, post-Independence, has registered a substantial improvement. The CDR has been brought down to a third of its 1953 level, and inefficiency, as measured by I_M, has also been reduced, from 0.58 to a value of 0.33, although the latter remains high. The improvement is clearly reflected in an *M*-curve for 2011 which dominates – is everywhere lower than – the 1953 *M*-curve.

It is instructive to compare one of the most backward States, UP, with the most advanced State (in terms of Human Development Indicators), Kerala. The *M*-curves are shown in Fig. 4. While the CDR for UP, at 7.72 per thousand population, is higher than for Kerala, at 7.10, the gap is not vast. However, the welfare loss values, D^*, for the two states display a greater divergence than do the D values. This is because of much greater inefficiency in UP ($I_M = 0.33$) than in Kerala ($I_M = 0.12$). Kerala's *M*-curve, compared to UP's, displays a relatively small initial concave hump and is thereafter convex and relatively close to the base of the unit square. This is a clear reflection of Kerala's relatively high investment in literacy, public health, and child-centred welfare measures.

In the preceding pairs of comparisons, the welfare loss, or D^*, values displayed wider divergence than the CDR values, D. However, consider comparisons between males and females in India for 2011, for which the *M*-curves are shown in Fig. 5. In this comparison, the D^* value has a narrower divergence than the D value. This is not really a 'good thing' from the point of view of

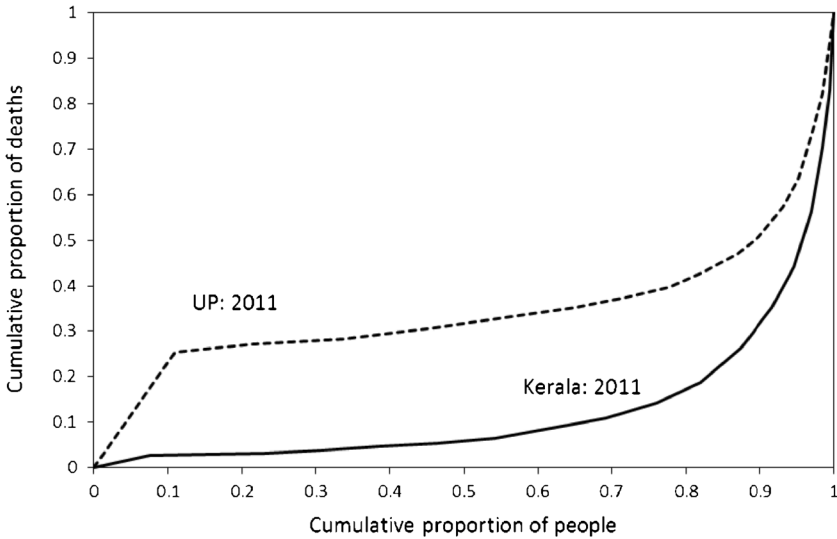*J. Creedy and S. Subramanian*

Figure 4: *M*-Curves for Uttar Pradesh and Kerala: 2011

what one might consider to be the norm of relative female advantage in a developed demographic regime. Specifically, in India in 2011, the female I_M coefficient of 0.297 is *larger* than the male I_M coefficient of 0.253. In Fig. 5, the female *M*-curve initially lies above the male *M*-curve, and the former
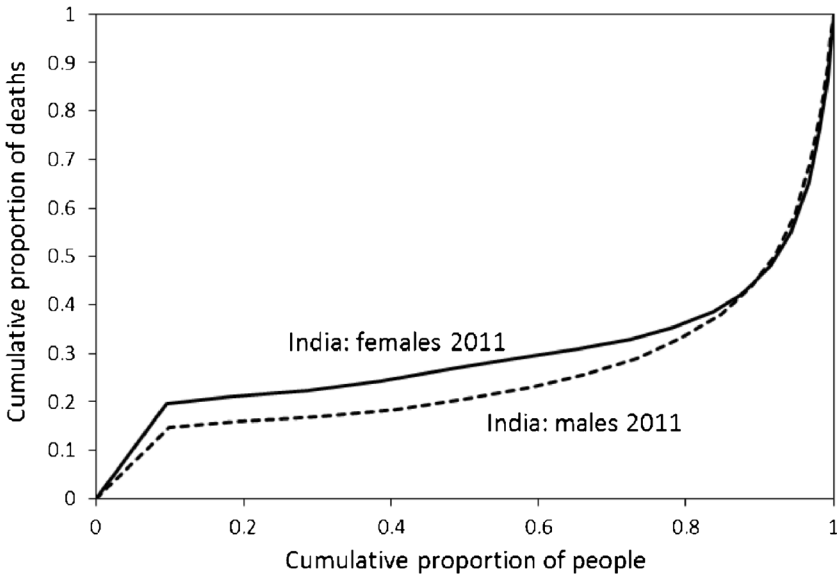

Figure 5: *M*-Curves for Indian Males and Females: 2011

intersects the latter in about the 45-49 age group, and thereafter lies below the male curve.

This property is a reflection of the relative neglect of the female child in India, and of female lives lost to maternal mortality. For women who have survived these contingencies, the natural biological advantage in longevity of women asserts itself, coupled with life-style related deaths among males in middle-age and thereafter, arising from, for example, smoking and alcohol. To help place these gender comparisons in perspective, the nature of the female and male M-curves for India may be contrasted with those for New Zealand. These are presented in Fig. 6, which shows that the female M-curve 'dominates' the male M-curve over its full length.

Figure 7 displays M-curves for rural and urban areas in India in 2011, and clearly demonstrates the 'dominance' of urban areas, for which the curve lies everywhere below that of rural areas. These characteristics may be said (at least qualifiedly) to support Lipton's thesis of 'urban bias', such that, '[t]he most important class conflict in the poor countries of the world today is ... between the rural and the urban classes' (Lipton (1977, p. 1)).

The observed mortality comparisons revealed by Fig. 7 and Table 1 reflect several aspects of the welfare divide between the rural and urban areas of the country. For instance, in 2011 the headcount ratio of money-metric poverty was
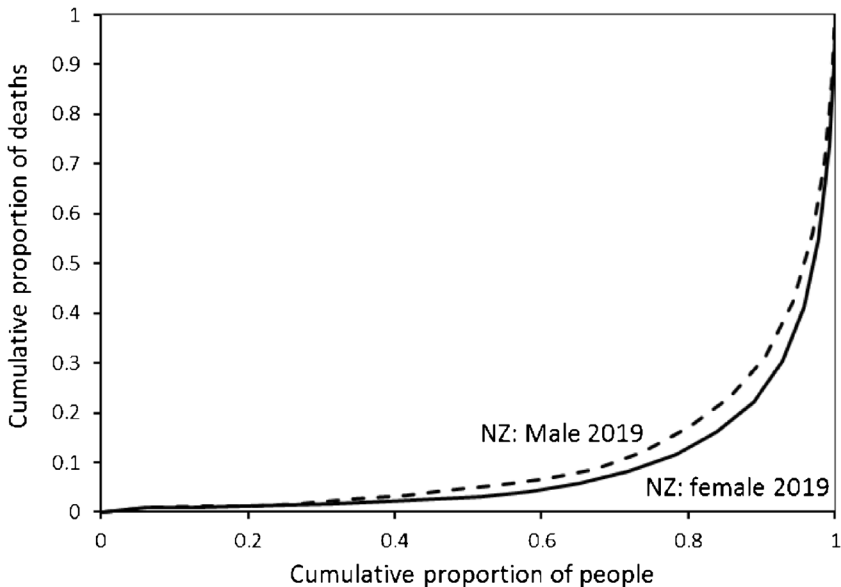

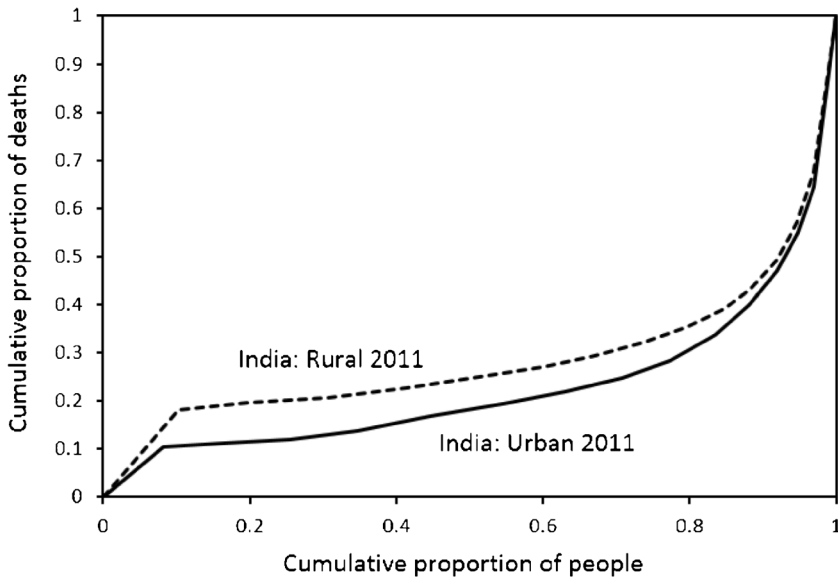Figure 6: M-Curves for New Zealand Males and Females in 2019

Figure 7: *M*-Curves for India 2011: Rural and Urban Areas

25.7 per cent for rural India and 13.7 per cent for urban India.[7] The literacy rates were 69.9 and 85 per cent respectively in rural and urban areas (Census 2011[8]). The Scheduled Castes and Tribes, the historically most disadvantaged and impoverished groups of the Indian population, accounted for 29.8 and 15.4 per cent respectively of the rural and urban populations (Census 2011).[9]

National Family Health Survey data for 2019-20 furnish a series of relevant contrasts between rural and urban India.[10] Thus, for example, 64.9 and 81.5 per cent of rural and urban populations respectively lived in households that used an improved sanitation facility. The proportions of households using clean fuel for cooking were 43.2 and 89.7 per cent for rural and urban populations respectively. The proportion of institutional births was 86.7 per cent for rural areas and 93.8 per cent for urban areas. The incidence of under-5 stunted children was 37.3 per cent and 30.1 per cent respectively in rural and urban

[7] Press Information Bureau (Government of India, Planning Commission, July 2013): 'Poverty Estimates for 2011-12'. http://www.niti.gov.in/sites/default/files/2020-05/press-note-poverty-2011-12-23-08-16.pdf

[8] Census of India 2011: Rural Urban Distribution of Population (Provisional Population Tables), Ministry of Home Affairs. https://www.censusindia.gov.in/2011-prov-results/paper2/data_files/india/Rural_urban_2011.pdf

[9] Census of India 2011: Release of Primary Census Abstract Data Highlights Scheduled Caste (SC) and Scheduled Tribe (ST), Ministry of Home Affairs, 30th April 2013. https://www.indiaspend.com/wp-content/uploads/2020/09/INDIA_CENSUS_ABSTRACT-2011-Data_on_SC-STs.pdf

[10] National Family Health Survey (NFHS-5) 2019-20: Compendium of Fact Sheets, Key Indicators, India and 14 States/UTs (Phase II), Ministry of Health and Family Welfare, Government of India. National Family Health Survey (NFHS-5) (rchiips.org).

India, with corresponding figures of 33.8 and 27.3 per cent respectively for under-5 underweight children. Health facilities are markedly less available in the rural than in the urban areas. As Anand and Fan (2016, p. 9) observe, in 2001, '[o]f all health workers, 59.2% were in urban areas, where 27.8% of the population resides, and 40.8% were in rural areas, where 72.2% of the population resides. The ratio of urban density to rural density for doctors was 3.8, for nurses and midwives 4.0, and for dentists 9.9'.

To place the UP mortality in an international context, *M*-curves are shown in Fig. 8 for UP and the UK in 1953 and 1951 respectively. India, which attained Independence in 1947, was a very recently decolonized country in 1953. A relatively young population—one in which the aged account for a relatively small proportion of the total population—is expected to have a relatively low aggregate death rate. In UP in 1953, the 75+ age cohort accounted for only 0.6 per cent of the total population, while the corresponding figure for the UK in 1953 was nearly six times higher, at 3.5 per cent. Even so, the CDR for UP, at 21.1 per thousand population, is 1.7 times the CDR for the UK, at 12.4 per thousand.

The high CDR for UP is mainly because of the massive level of child (under-5) mortality, especially infant (under-1) mortality. This is aided by poverty, under-nutrition, the uncontrolled spread of diseases like malaria, cholera, smallpox, kala-azar and TB, and the poor status of public health facilities and public hygiene and sanitation, accompanied by low levels of awareness and literacy. If
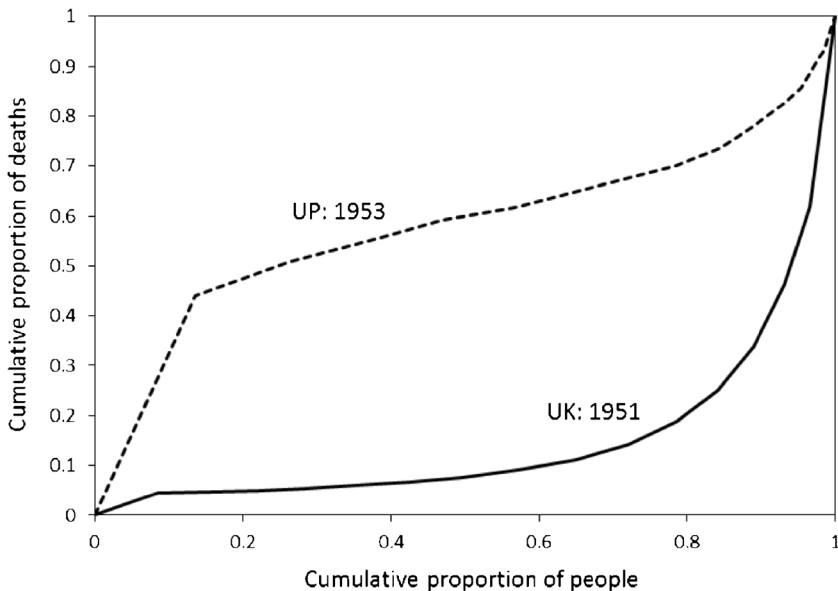

Figure 8: *M*-Curves for Uttar Pradesh 1953 and UK 1951

the CDR in immediately post-colonial UP is high, the wastage or inefficiency in the distribution of death across age is also huge: the value of the I_M coefficient, at 0.58, is a very large number. The *M*-curve for UP lies everywhere above the 45-degree line, and is roughly concave over its entire range. The contrast between the *M*-curves of Fig. 8 is easily discernible.

A further international comparison of interest is that between India in 2011 and New Zealand in 2019. This is an example where the generalised *M*-curves, or *GM*-curves, are particularly useful. These are shown in Fig. 9. The CDR for India in 2011, of 7.1 per thousand, is actually lower than that of NZ in 2019, of 7.3. Hence the *GM*-curve for India begins and ends slightly below that of NZ. Yet the higher Indian inefficiency, of $I_M = 0.268$, compared with that of NZ, of $I_M = 0.118$, means that the former *GM*-curve lies above that of NZ over most of its length. The value of the NZ loss function, $D^* = D(1+I_M)$, reflected in the area below its generalised *M*-curve, is lower than that of India: 8.16 compared with 9.13. This is a further example of the potentially misleading nature of the CDR.

4   Decompositions of CDR Differences

This section returns to the basic feature of the crude death rate, CDR, that it depends on the population age distribution as well as age-specific mortality
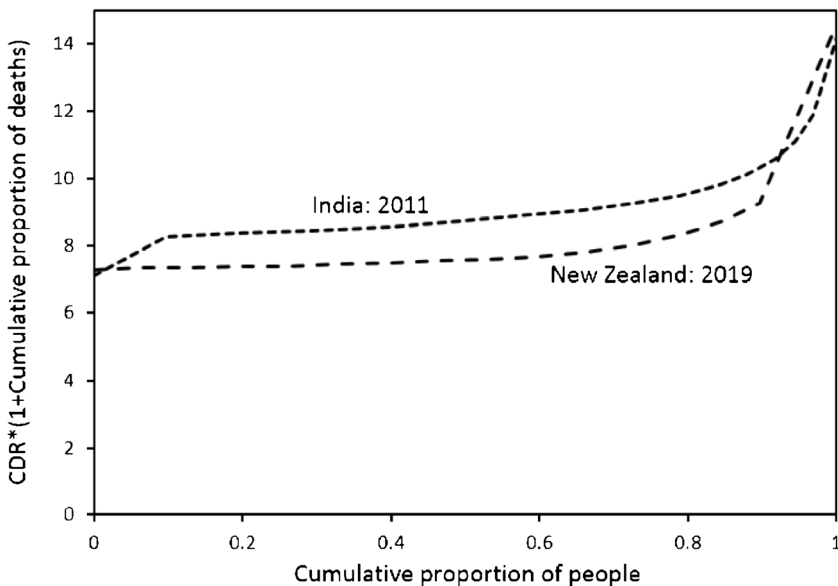

Figure 9: Generalised *M*-Curves for India and New Zealand

rates. The CDR, denoted D, is a weighted sum of age-specific death rates (ASDRs), the weights being the age-specific population shares. The CDR thus conflates the age-profile of mortality and the age-structure of the population in a single real number, which tends to obscure the role played by each of these factors in determining the overall pattern of mortality. The M-curve can be used to get a sense of the contribution of each of these factors to aggregate mortality, in the following way.

The relationship between two M-curves, M_1 and M_2, can be thought of as the product of two counterfactual exercises. In the first exercise, the two curves are derived by using the actual age-specific death rates (ASDRs) of the two societies and a presumed hypothetical shared population age structure. In the second exercise the two curves are derived using the actual population age structures of the two societies and a presumed hypothetical shared pattern of ASDRs. The first pair of M-curve comparisons isolate the difference attributable to the differing ASDRs, and the second pair isolate the difference attributable to the differing age structures. The comparison of the actual M-curves is some combination of these counterfactual ASDR and age structure comparisons.

This is illustrated in Fig. 10 by the M-curves for India in 1971 and 2011. This figure shows, 'at a glance', that the change in the population age structure over the period had little effect on the M-curves. The improvement in the age-specific mortality rates, in particular the reduced inefficiency of deaths over the period, reflected in the reduction in I_M, contributed substantially to the overall change in the M-curve.

This type of comparison also lies behind a Shorrocks-Shapley form of decomposition of a change in the crude death rate over time, or between geographical areas.[11] Suppose the CDRs in two different periods or regions are denoted D_1 and D_2, and the population age distributions are the column vectors, p_1 and p_2, with elements measuring proportions in each age group. The age-specific death rates are given by vectors, Θ_1 and Θ_2. It can be seen that, where the prime indicates transposition:

$$D_2 - D_1 = [p'_2 \Theta_2 - p'_1 \Theta_2] + [p'_1 \Theta_2 - p'_1 \Theta_1] \tag{1}$$

---

[11] See Shorrocks (2013). For applications to the decomposition of Covid-19 fatalities, see Dudel *et al.* (2020) and Philip et al. (2021).
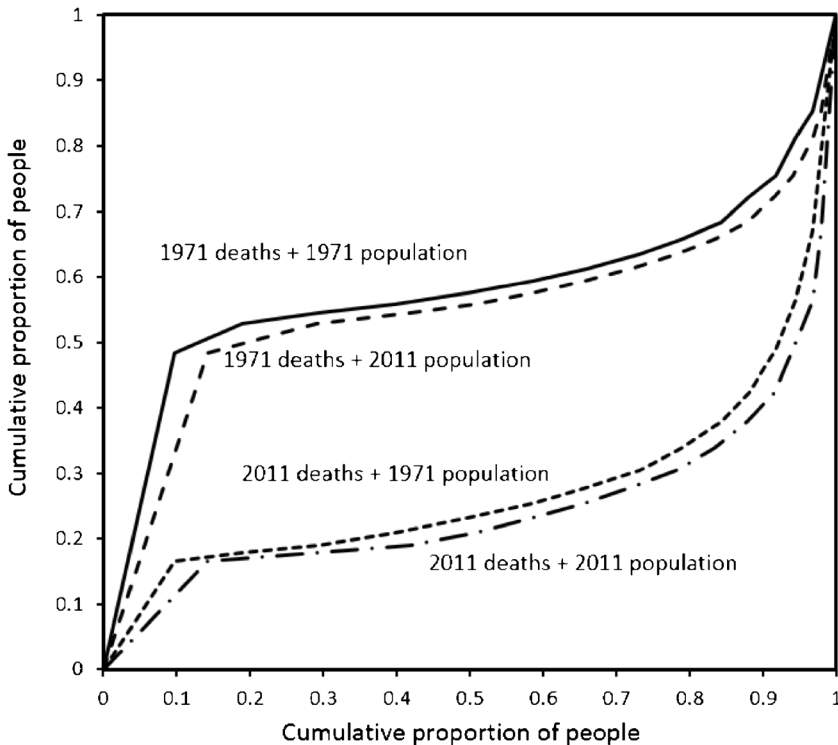
Figure 10: M-Curve Decomposition of CDR for India: 1971 to 2011

Alternatively, the difference can be written as:

$$D_2 - D_1 = [p'_2\Theta_1 - p'_1\Theta_1] + [p'_2\Theta_2 - p'_2\Theta_1] \qquad (2)$$

In each case, the first term in square brackets measures a change arising from differences in population distributions, while the second term measure the contribution of changes in age-specific mortality rates. These two effects can be referred to respectively as A-effects and F-effects (with A and F referring to age and fatality). There are two ways of expressing the decomposition because, for example, the A-effect can be measured holding age-specific death rates constant at either the initial or final period values. As there is no presumption in favour of either form, the results reported here are obtained by taking arithmetic means of the appropriate two terms from (1) and (2).

Table 2 reports the results of such a decomposition for a number of mortality comparisons. In all but one case – the exception being the difference between rural and urban areas in India in 2011 – the sign of the A-effect is

Table 2: Decompositions of CDR Differences

Pop 1	Pop 2	$\Delta D =$ $D_2 - D_1$	Pop age structure	Death Rates
UP 1953	UP 2011	-13.3800	0.35846	-13.73846
UP 1953	UK 1951	-8.7000	6.39692	-15.09692
Kerala 2011	UP 2011	0.6200	-2.95110	3.57110
India 2011: M	India 2011: F	-1.4800	0.29448	-1.77448
India 2011: Rural	India 2011: Urban	-1.9100	-0.00872	-1.90128
NZ 2019: M	NZ 2019: F	-0.5712	1.87424	-2.44546
NZ 2019: All	India 2011: All	-0.2000	-5.46467	5.26467

opposite to that of the F-effect. While the F-effect, on its own, exerts an upward pressure on the aggregate CDR, the A-effect exerts a downward pressure. The A-effect, that is, dampens the F-effect, but does not 'swamp' it. In the comparison between India 2011 and NZ 2019, the signs of the A- and F-effects are again opposing, but here the F-effect is dwarfed by the A-effect.

There is a large difference between NZ's and India's ASDRs (in fact the NZ profile comfortably dominates the Indian profile), but that difference is neutralised by the large difference in the two countries' age distributions. India's poor record of death rates is compensated – indeed overcompensated – by its being a younger population. Hence, there are two differences of roughly equal magnitude, pulling in opposite directions, with the population effect just edging out the mortality effect. This is a further example of why and how the CDR, read on its own, can be a misleading indicator.

5   Decompositions of Welfare Loss Differences

This section turns from decompositions of the CDR, D, to decompositions of the welfare loss, D^*, associated with deaths and their inefficiency, as defined above. The decomposition can be carried out by dividing the discrete change in D^* into a number of terms, as follows. First, write $D^* = DL$, where $L = 1 + I$. The difference between two regimes can be written as

$$\Delta D^* \equiv D_2^* - D_1^*$$
$$= (1/2)\left(D_2^* - D_1^*\right) + (1/2)\left(D_2^* - D_1^*\right)$$
$$= (1/2)(D_2 L_2 - D_1 L_1) + (1/2)(D_2 L_2 - D_1 L_1)$$
$$= (1/2)(D_2 L_2 - D_1 L_2 + D_2 L_1 - D_1 L_1) + (1/2)(D_2 L_2 + D_1 L_2 - D_2 L_1 - D_1 L_1) \quad (3)$$
$$= (1/2)[D_2(L_1 + L_2) - D_1(L_1 + L_2)] + (1/2)[D_2(L_2 - L_1) + D_1(L_2 - L_1)]$$
$$= (1/2)(L_1 + L_2)(D_2 - D_1) + (1/2)(D_1 + D_2)(L_2 - L_1)$$
$$= [(1/2)(L_1 + L_2)\Delta D] + [(1/2)(D_1 + D_2)\Delta L].$$

The first term in square brackets in (3) can be called the 'mean (or CDR)' contribution to the difference in the D^* values. The second term may be called the 'dispersion (or inefficiency)' contribution. That is, the variation in D^* is decomposed into a variation in the average mortality level (the mean effect), and a variation in the dispersion term (the 'inefficiency effect').

Using the values reported earlier in Table 1, these decompositions are shown in Table 3. In cases 1, 2, 3, 5 and 6, both the mean and the dispersion effects have the same sign: they are mutually reinforcing. In cases 1, 2, 5 and 6, the mean effect dominates. However, this dominance is relatively less in case 1 (the UP-UK comparison), where the inefficiency effect is substantial. In such a case the dispersion effect has a non-trivial role to play in explaining the mortality differential. In Cases 4 and 7, the mean and dispersion effects are mutually opposing, but in opposing ways. In the Male/Female comparison for India (Case 4), the superior performance of females in terms of age-specific death rates is somewhat neutralized by their greater inefficiency. This can be contrasted with the Male/Female comparison for NZ (case 5), where both the contributory effects are positive, with the mean effect dominating. This is similar to the pattern for the rural and urban areas of India in 2011 (Case 6).

The case of India/NZ (Case 7) is interesting: here, as in Case 4, the two effects have opposing signs, but unlike Case 4, it is the mean effect that is negative. India's CDR is actually *lower* than NZ's, but the inefficiency of the age distribution of deaths in India is so high that it swamps the mean effect and causes the welfare loss measure D^* to reverse the ranking by D. This is a further example of a case where going solely by the mean (CDR) without taking stock of the dispersion in reckoning mortality differentials could be misleading.

6   Conclusions

This paper has provided a number of mortality analyses for India, involving comparisons over time, between States, regions, and genders. To provide further perspective, selective comparisons with other countries were made.

Table 3: Decompositions of Welfare Loss Differences

Case	Comparison	$\Delta D^* =$ $D_2^* - D_1^*$	$\Delta D =$ $D_2 - D_1$	$\Delta L =$ $L_2 - L_1$	Mean Effect	Dispersion Effect
1	1=UK	19	8.7	0.4286	11.86	7.14
	2=UP				(62.42)	(37.58)
2	1=UP 2011	23	13.4	0.2449	19.47	3.53
	2=UP 1953				(84.65)	(15.35)
3	1=Kerala	2.3	0.6	0.2101	0.74	1.56
	2=UP				(32.17)	(67.83)
4	1=India F	1.57	1.48	-0.0444	1.88	-0.32
	2=India M				(119.75)	(-19.75)
5	1=NZ F	0.78	0.57	0.021	0.63	0.15
	2=NZ M				(80.77)	(19.23)
6	1=India Urban	2.79	1.91	0.0582	2.4	0.39
	2=India Rural				(86.02)	(13.98)
7	1=NZ	1.64	-0.21	0.2598	-0.24	1.88
	2=India				(-14.63)	(114.63)

Note: Percentages are given in parentheses below each effect

Instead of relying on 'crude' or 'age-standardised' death rates, comparisons were made using the concept of the mortality concentration curve, or M-curve, which plots the cumulative proportion of total deaths against the corresponding cumulative proportion of people, when individuals are ranked in ascending order by age. Associated 'inefficiency' measures, I_M, were also obtained in terms of the area underneath the M-curve: it measures the 'distance', as an area measure, between the curve and that which would arise if all individuals were to die at a biological maximum age. Furthermore, by introducing a value judgement, expressed in terms of an 'aversion to early deaths', a loss function provides a welfare ranking of mortality. This allows for the possibility that one population could have a lower CDR than a second population, but the latter has a sufficiently lower 'inefficiency' that the welfare ranking is the opposite of the CDR ranking.

For example, it was found that, for Uttar Pradesh (one of the poorest regions of India) over the period 1953 to 2011, the CDR fell by 63 per cent. In addition, there was a 40 per cent reduction in the 'inefficiency' of deaths,

resulting in a substantial drop in the welfare loss, of 71 per cent, over the period. Nevertheless, the high infant mortality continued in 2011 to produce a concave M-curve over its early ranges. The region compares unfavourably with Kerala, for which 'inefficiency' is nearly one third that of Uttar Pradesh. Perhaps surprisingly, India in 2011 had a slightly lower CDR than New Zealand in 2019, yet the much lower 'inefficiency' in NZ means that their rankings are reversed when using the 'loss function'. The lower CDR in India arose partly from its 'younger' age distribution: there were relatively fewer individuals in the higher-age high-mortality groups.

A startling contrast was found when comparing Uttar Pradesh in 1953 and the UK in 1951. While UP had a CDR of 1.75 times that of the UK, the 'inefficiency', as measured by I_M, was almost four times higher: the contrast is immediately apparent when the two M-curves are plotted together.

Two decomposition methods were used to isolate separate contributions to differences in the CDR, and the welfare loss. First, changes in the CDR were decomposed using a Shorrocks-type approach to produce two components, relating to age-specific mortality (a fatality or 'F-effect') and the population age distribution (an age or 'A-effect'). In all cases, except for the comparison between urban and rural regions of India, these two effects worked in opposite directions. The F-effect tends to increase the CDR, while the A-effect reduces the CDR. Second, changes in the welfare loss, measured using the loss function expressed as the product of the CDR and 1 plus I_M, were decomposed into 'mean' and 'dispersion' effects (in terms of D and I_M respectively). These effect were sometimes found to be self-reinforcing, while for other comparisons they worked in opposite directions.

It is suggested that the measures used here provide fresh insights into mortality experience in India. In addition, they illustrate the value of the new analytical approach via the M-curve which, like the Lorenz curve in the context of inequality, provides an instant visual contrast between, say, two time periods or regions. Furthermore, the approach allows normative, as well as purely statistical, comparisons to be made. This is achieved by introducing an explicit value judgement – an aversion to early deaths – along with a measure of inefficiency which combine to provide welfare loss comparisons.

Declarations

Conflict of Interests   The authors have NO conflict of interests to declare.

Appendix 1: Glossary of Some Relevant Terms and Concepts

Let $A = \{a_1, ..., a_j, ..., a_K\}$ be a set of K mutually exclusive and exhaustive age-groups into which a population is divided, with the groups indexed in ascending order of age; that is, $a_1 < a_2 < ... < a_K$. Let n be the size of the total population, and n_j the size of the jth age-group, so that $n = \sum_{j=1}^{K} n_j$. Let d be the total number of deaths occurring over the reference period (usually a calendar year), and d_j the number of deaths occurring in the jth age-group, $j = 1, ..., K$, so that $d = \sum_{j=1}^{K} d_j$. The proportion of the total population is $p_j \equiv n_j/n$, and $q_j \equiv d_j/d$ is the proportion of all deaths, in the jth age-group, $j = 1, ..., K$.

1.   *Age-Distributions of Populations and Deaths*

The age-distribution is the K-tuple $\{(a_1, p_1), ..., (a_j, p_j), ...(a_K, p_K)\}$, and the age-distribution of deaths is the K-tuple $\{(a_1, q_1), ..., (a_j, q_j), ...(a_K, q_K)\}$.

2.   *Age-Specific Death Rate (ASDR)*

The death-rate of the jth age-group is the proportion of the population in the group that die; that is, $D_j \equiv d_j/n_j$.

3.   *Crude Death Rate (CDR)*

The crude death rate is the proportion of a population that dies over the reference period, $D \equiv d/n$. This is a weighted sum of the age-specific death rates, the weights being the relevant population shares: $D = \sum_{j=1}^{K} p_j D_j$.

4.   *Income Distribution*

Let $X = \{x_1, ..., x_j, ..., x_K\}$ be a set of K mutually exclusive and exhaustive income-classes into which a population is divided, with the classes indexed in ascending order of income, that is, $x_1 < x_2 < ... < x_K$. Let n be the size of the total population, and n_j the size of the population in the jth income-class, so that $n = \sum_{j=1}^{K} n_j$. The proportion of the total population in the jth income-class is $\pi_j \equiv n_j/n$. Then, an *income distribution* is defined by the K-tuple $\{(x_1, \pi_1), ..., (x_j, \pi_j), ..., (x_K, \pi_K)\}$. The cumulative proportion of the population with incomes not exceeding the upper limit of the jth income-class is $\Pi_j \equiv \sum_{i=1}^{j} \pi_i, j = 1, ..., K$.

5.   **The Lorenz Curve**

Let $\overline{x}_j$ be the average income in the jth income-class, and $\overline{x}$ the average income of the entire population (so that $\overline{x} = \sum_{j=1}^{K} \pi_j \, \overline{x}_j$). The share in total income of members of the jth income-class is given by $\lambda_j = \pi_j \, \overline{x}_j / \overline{x}$. The *cumulative income share* of those with incomes that do not exceed the upper limit of the jth income-class is designated by Λ_j, $j = 1, ..., K$. The Lorenz curve plots the cumulative income share against the cumulative population share, when income is arranged in ascending order. A typical ordinate of the Lorenz curve is given by: $\Lambda_j (X; \Pi_j) \equiv \sum_{i=1}^{j} \lambda_i$. Set $\Pi_0 = \Lambda_0 \equiv 0$ and, by definition, $\Pi_K = \Lambda_K = 1$. A piece-wise linear Lorenz curve is obtained by connecting, with straight lines, the points (Π_0, Λ_0), (Π_1, Λ_1), ..., (Π_j, Λ_j), ..., (Π_K, Λ_K) within the unit square. An equal distribution of income has a Lorenz curve for which $\Lambda_j = \Pi_j$ for all j: this defines the 'line of perfect equality'.

6.   **The Gini Coefficient of Inequality**

The *Gini inequality measure G* can be derived from the Lorenz curve as the area enclosed by the Lorenz curve and line of perfect equality, as a proportion of the area below the diagonal (which is one-half). For a piece-wise linear Lorenz curve of the type described above, the relevant areas can be computed as the sums of a set of trapeziums, and the well-known 'trapezoidal approximation formula' for the Gini coefficient is given by: $G = 1 - \sum_{j=1}^{K} (\Pi_j - \Pi_{j-1}) (\Lambda_j + \Lambda_{j-1})$; see, for example, Fellman (2012).

Appendix 2: Data Sources

Information on the age-distribution of population and of deaths for each of the population groups considered in the text has been derived from the following data sources. Tables providing the coordinates of the *M*-curve and *GM*-curve for each of the comparisons examined in the text are provided in Creedy and Subramanian (2022b).

Uttar Pradesh 1953

Census of India: Paper No.1, 1955: Sample Census of Births and Deaths – 1953-54: Table 44 (p.53) – Age-Specific Death-Rates and Percentage Distribution of Total Population by Age.

Uttar Pradesh 2011

Sample Registration System Statistical Report 2011: Table 1 - Percent distribution of estimated population by age-group, sex and residence, 2011, Uttar Pradesh; and Table 8 - Age-specific death rate by sex and residence, 2011, Uttar Pradesh. Census of India Website : SRS Statistical Report (censusindia. gov.in)

Kerala 2011

Sample Registration System Statistical Report 2011: Table 1 - Percent distribution of estimated population by age-group, sex and residence, 2011, Kerala; and Table 8 - Age-specific death rate by sex and residence, 2011, Kerala. Census of India Website : SRS Statistical Report (censusindia.gov.in)

India 2011 (Males)

Sample Registration System Statistical Report 2011: Table 1 - Percent distribution of estimated population by age-group, sex and residence, 2011, India, Males; and Table 8 - Age-specific death rate by sex and residence, 2011, India, Males. Census of India Website : SRS Statistical Report (censusindia. gov.in)

India 2011 (Females)

Sample Registration System Statistical Report 2011: Table 1 - Percent distribution of estimated population by age-group, sex and residence, 2011, India, Females; and Table 8 - Age-specific death rate by sex and residence, 2011, India, Females. Census of India Website : SRS Statistical Report (censusindia.gov.in)

India 2011 (Rural) and India 2011 (Urban)

Sample Registration System Statistical Report 2011: Table 1 - Percent distribution of estimated population by age-group, sex and residence, 2011. Census of India Website : SRS Statistical Report (censusindia.gov.in)

Census of India Compendium of India's Fertility and Mortality Indicators : Table 8 – Age-specific mortality rate by sex and residence from 1991 to 2013 at interval of 5 years (Table T-8A, 2011). Census of India Website : Compendium of India's Fertility and Mortality Indicators ,1971 - 2013 (censusindia.gov.in)

UK 1951 (England and Wales, 1951)

Census of India: Paper No.1, 1955: Sample Census of Births and Deaths – 1953-54: Table 44 (p.53) – Age-Specific Death-Rates and Percentage Distribution of Total Population by Age.

New Zealand 2019 (Males), New Zealand 2019 (Females), and New Zealand 2019 (All)

For New Zealand it is necessary to go to each of the following web sites and use the on-line data selection facility to select and then download the required tables.

https://www.stats.govt.nz/topics/births-and-deaths

https://www.stats.govt.nz/topics/population

India 2011 (All)

Sample Registration System Statistical Report 2011: Table 1 - Percent distribution of estimated population by age-group, sex and residence, 2011, India, Total; and Table 8 - Age-specific death rate by sex and residence, 2011, India, Total. Census of India Website : SRS Statistical Report (censusindia.gov.in)

References

ANAND, S. and FAN, V. (2016) *The Health Work Force in India.* World Health Organization Human Resources for Health Observer Series, No.16. 16058health_workforce_India.pdf (who.int)

BISHOP, J.A., CHAKRABORTI, S. and THISTLE, P. (2009) An Asymptotically Distribution Free Test for Sen's Welfare Index. *Oxford Bulletin of Economics and Statistics*, 52, 105-113.

CREEDY, J. and SUBRAMANIAN, S (2022a) Mortality Comparisons and Age: a New Mortality Curve. *Victoria University of Wellington Working Papers in Public Finance.* No. 03/2022.

CREEDY, J. and SUBRAMANIAN, S (2022b) Mortality Comparisons 'At a Glance': A Mortality Concentration Curve and Decomposition Analysis for India. *Victoria University of Wellington Working Papers in Public Finance.* No. 06/2022.

DUDEL, C., RIFFE, T., ACOSTA, E., VAN RAALTE A.A., STROZZA, C. and MYRSKYLA, M. (2020) Monitoring Trends and Differences in COVID-19 Case Fatality Rates using Decomposition Methods: Contributions of age structure and age-specific fatality. *medRxiv.* https://doi.org/10.1101/2020.03.31.20048397.

FELLMAN, J. (2012) Estimation of Gini Coefficients Using Lorenz Curves, *Journal of Statistical and Econometric Methods*, 1, 31-38 .

LIPTON, M. (1977) *Why Poor People Stay Poor: Urban Bias in World Development.* Harvard University Press: Cambridge, Mass.

PHILIP, M., RAY, D. and SUBRAMANIAN, S. (2021) Decoding India's Low Covid-19 Case Fatality Rate. *Journal of Human Development and Capabilities*, 22, 27-51.

SEN, A. (1973) *On Economic Inequality.* Oxford: Clarendon.

SHORROCKS, A. F. (1983) Ranking Income Distributions. *Economica*, 50, 3-17.

SHORROCKS, A. F. (2013) Decomposition Procedures for Distributional Analysis: A Unified Framework Based on the Shapley Value. *The Journal of Economic Inequality*, 11, 99-126.

JOHN CREEDY
VICTORIA UNIVERSITY OF WELLINGTON,
WELLINGTON, NEW ZEALAND

S. SUBRAMANIAN
INDEPENDENT SCHOLAR; FORMERLY, MADRAS
INSTITUTE OF DEVELOPMENT STUDIES, CHENNAI,
INDIA