
Chapter 25

ESTIMATING ATTRIBUTABLE BURDEN OF DISEASE FROM EXPOSURE AND HAZARD DATA

STEPHEN VANDER HOORN, MAJID EZZATI,
ANTHONY RODGERS, ALAN D. LOPEZ AND
CHRISTOPHER J.L. MURRAY

1. ESTIMATING POPULATION ATTRIBUTABLE FRACTIONS

As described in earlier chapters, the contribution of a risk factor to disease burden (expressed as the fraction of disease or death attributable to the risk factor in a population) is given by the generalized “potential impact fraction” (PIF) in Equation 1a (Drescher and Becher 1997; Eide and Heuch 2001; Walter 1980).

$$PIF = \frac{\int_{x=0}^m RR(x)P(x)dx - \int_{x=0}^m RR(x)P'(x)dx}{\int_{x=0}^m RR(x)P(x)dx} \quad (1a)$$

$RR(x)$: relative risk at exposure level x

$P(x)$: population distribution of exposure

$P'(x)$: counterfactual distribution of exposure, and

m : maximum exposure level

The first and second terms in the numerator of Equation 1a represent the total exposure-weighted risk of disease or mortality in the population under current and counterfactual exposure distributions. The corresponding relationship when exposure is described as a discrete variable with n levels is given by:

$$PIF = \frac{\sum_{i=1}^n P_i RR_i - \sum_{i=1}^n P'_i RR_i}{\sum_{i=1}^n P_i RR_i} \quad (1b)$$

The PIF equation can be used to estimate the population attributable fraction (PAF), defined as the proportional reduction in disease or death that would occur if exposure to the risk factor were reduced to the counterfactual exposure distribution. The remainder of this chapter outlines how data on exposure, hazard and disease burden were combined to derive estimates of attributable disease burden, with estimation of the population attributable fraction as the intermediate step. The application of Equations 1a and 1b in the context of the comparative risk assessment (CRA) project is discussed and several issues regarding its implementation are detailed.

2. ESTIMATING ATTRIBUTABLE MORTALITY AND BURDEN OF DISEASE

For each risk factor–disease outcome pair, PAFs for each of the 224 age, sex, subregion¹ groups were calculated using the relationships in Equation 1, separately for mortality (PAF_M) and incidence (PAF_I) when the relative risks for mortality and incidence were different. For each of these 224 groups, the estimates of mortality (AM_{ij}) and burden of disease (AB_{ij}) from disease j attributable to risk factor i were calculated as below. Burden of disease, reported annually in the annexes of the *World Health Report*, was expressed in disability-adjusted life years (DALYs), with methods and assumptions described elsewhere (Murray and Lopez 1996). Specifically:

$$\begin{aligned} AM_{ij} &= PAF_{M-ij} \times M_j \\ A YLL_{ij} &= PAF_{M-ij} \times YLL_j \\ A YLD_{ij} &= PAF_{I-ij} \times YLD_j \\ AB_{ij} &= A YLL_{ij} + A YLD_{ij} \end{aligned}$$

Where “A” indicates “attributable” and

YLL : years of life lost to premature mortality

YLD : years of life lived with disability due to disease incidence

For those risk factors with insufficient data to estimate a relative risk model (e.g. occupational or alcohol-caused injuries or the effects of lead exposure on blood pressure), disease burden or mortality was estimated using existing registers or corresponding hazard relationships. Estimates were then aggregated across age groups to obtain subregional estimates,

and across subregions to obtain global estimates. The details of this aggregation are described later in this chapter.

3. COUNTERFACTUAL EXPOSURE DISTRIBUTION

The estimates of burden of disease and injuries due to risk factors in the CRA project are based on a counterfactual of theoretical-minimum-risk exposure distribution, defined in chapter 1 and described in individual risk factor chapters. By using the theoretical-minimum-risk exposure distribution, which by definition has a relative risk of 1.0, as the counterfactual exposure distribution or level, Equations 1a and 1b are reduced to:

$$AF = \frac{\int_{x=0}^m RR(x)P(x)dx - 1}{\int_{x=0}^m RR(x)P(x)dx} \tag{2a}$$

and

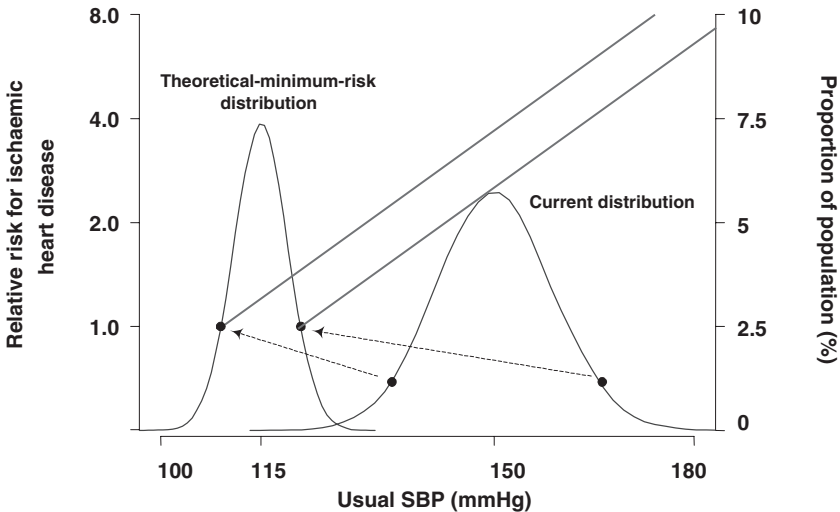
$$AF = \frac{\sum_{i=1}^n P_i(RR_i - 1)}{\sum_{i=1}^n P_i(RR_i - 1) + 1} \tag{2b}$$

3.1 THEORETICAL-MINIMUM-RISK EXPOSURE DISTRIBUTION FOR CONTINUOUS EXPOSURE VARIABLES

The theoretical-minimum-risk exposure distribution for continuous risk factors is itself often a distribution of exposure levels, vs a constant baseline. Figure 25.1, for example, illustrates a scenario for systolic blood pressure (SBP) with typical exposure levels in an older population (mean: 150 mmHg; SD: 9 mmHg) compared with the theoretical-minimum-risk exposure distribution (mean: 115 mmHg; SD: 6 mmHg). The non-zero standard deviation of the theoretical-minimum-risk distribution reflects the reality that there always is some inter-person variability within any given population, even after hypothetical reductions such as that shown in Figure 25.1.

The optimal exposure distribution for a population would overlap precisely with the theoretical-minimum-risk exposure distribution. By definition of theoretical-minimum risk, such a population would be collectively without any increased risk and therefore with zero attributable burden due to the risk factor of interest. Any population containing individuals outside this distribution will then have a population attributable fraction greater than zero and exposure distributions converging on the

Figure 25.1 Theoretical-minimum-risk exposure distribution for continuous risk factors using systolic blood pressure (SBP) as an example



Note: Each point represents a hypothetical individual or small group of individuals in the population. The solid straight lines represent the increasing relative risk, on a log scale, for ischaemic heart disease with increasing SBP.

theoretical minimum will have attributable burden tending towards zero. The risk for any *individual* (or groups of individuals in a narrow range of exposure) in the population would be determined by the difference between her/his current exposure (SBP level) and the SBP level that s/he would have when the population distribution overlaps with the theoretical-minimum-risk exposure distribution.

Estimating the total hazard at the population level can be achieved using a micro-simulation approach in which individuals are drawn randomly from current and theoretical-minimum-risk exposure distributions. For most risk factors, however, individual exposures “track” over relatively long periods of time (Lauer and Clarke 1988; Voors et al. 1979; Wilsgaard et al. 2001). In other words, those with higher/lower exposure levels of a particular risk factor are expected to have higher/lower exposure levels within the theoretical-minimum-risk exposure distribution (see the hypothetical individuals in Figure 25.1). Random (uncorrelated) draws of individuals from current and theoretical-minimum-risk exposure distributions would be inconsistent with the empirical evidence on tracking. Consistent with this evidence, we assumed that the ordering of individuals in the exposure distribution remains unchanged (i.e.

the rank-order correlation of individual exposures equals 1) in the transition to the theoretical-minimum-risk distribution in estimating the PAF.

With correlated rank-ordering of individuals in current and theoretical-minimum-risk exposure distributions, if hazards were a linear function of exposure, then for those risks with symmetric distributions, shifting the population to the theoretical minimum distribution would be computationally equivalent to shifting everyone to the mean exposure of the theoretical-minimum-risk exposure distribution (i.e. the standard deviation of the theoretical-minimum-risk exposure distribution would not change the total hazard). This is because, with a linear hazard function, the changes in hazards for individuals above and below the mean, as a result of changing the standard deviation of the theoretical-minimum-risk exposure distribution, would fully compensate each other.

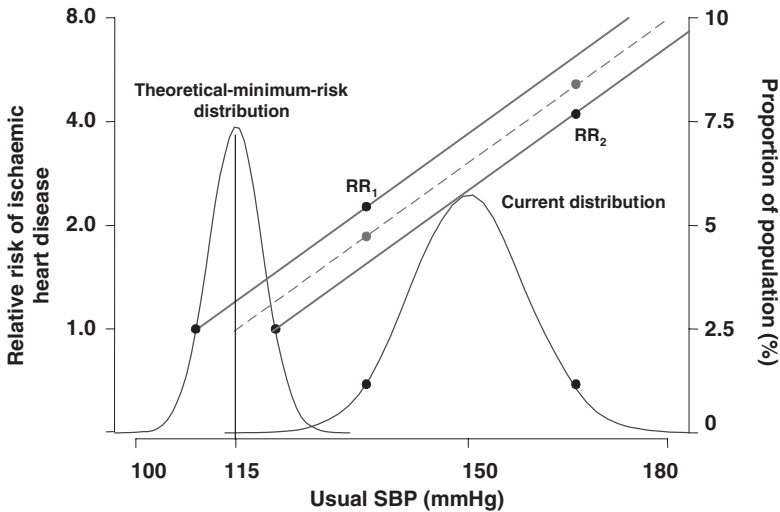
Risk is, however, an exponential function of exposure in most epidemiological models. With an exponential hazard function, when the baseline is the mean of theoretical-minimum-risk exposure distribution, the integrated risk is larger for those above the mean than those below the mean (Figure 25.2), compared to the case of treating theoretical-minimum-risk exposure as a distribution with non-zero standard deviation. The net difference would depend on both the steepness of the risk curve (i.e. increased risk per unit increase in exposure) and the standard deviations of the current and theoretical-minimum-risk exposure distributions.

For computational reasons, we estimated PAFs for continuous risk factors relative to the mean of the theoretical-minimum-risk exposure distribution. In these calculations, the relative risk for any individual in the population with an exposure below the mean of the theoretical-minimum-risk exposure distribution was set to 1.0 (e.g. in Figures 25.1 and 25.2, the lower tail of current blood pressure distribution is inside the theoretical minimum distribution with some individuals already at a level below 115 mmHg. These individuals were assigned a relative risk of 1.0). Sensitivity analysis showed that in the scenarios analysed in the CRA project, global PAFs estimated by integrating risk relative to the mean of the theoretical-minimum-risk exposure distribution were up to 2% larger than those estimated by integrating risk relative to the full distribution. As described above, this is because of the non-linear shape of most hazard functions.

4. AGGREGATION OF ATTRIBUTABLE BURDEN ACROSS AGE, SEX AND SUBREGION

Within each of the 14 subregions, all-age-sex population attributable fractions ($PAF_{subregion}$) were calculated by aggregating attributable burden estimates across the 16 age-sex-specific estimates within the subregion

Figure 25.2 Effect of a non-linear hazard function and choice of baseline on total population risk



Note: With an exponential hazard function, when theoretical-minimum-risk exposure is a distribution with a non-zero standard deviation, those falling above the mean of the current distribution (e.g. 155 mmHg for SBP) contribute more to total population hazard than those below it, relative to the case when the baseline is a constant level (115 mmHg for SBP). In the figure, the solid lines represent the hazard when the theoretical-minimum-risk exposure is a distribution, and the dotted line when a constant baseline is considered. The difference between the two relative risks on the right (RR_2) is larger than those on the left (RR_1). As a result of this imbalanced contribution to hazard, using the mean of theoretical-minimum-risk exposure distribution as baseline in estimating total population hazard would result in slightly larger PAFs than using the complete theoretical minimum distribution.

and then dividing by the total subregional disease burden using the relationship in Equation 3 (the estimates could similarly be aggregated across ages separately for males and females).

$$PAF_{subregion} = \frac{\sum_{age,sex} AB_{subregion,age,sex}}{\sum_{age,sex} B_{subregion,age,sex}} \quad (3)$$

Similarly, for each age-sex group, world attributable fractions ($PAF_{age,sex}$) were calculated by aggregating attributable burden estimates across all the 14 subregion-specific estimates and then dividing by the disease burden for that age-sex group using the relationship in Equation 4.

$$PAF_{age,sex} = \frac{\sum_{subregion=1}^{14} AB_{subregion,age,sex}}{\sum_{subregion=1}^{14} B_{subregion,age,sex}} \quad (4)$$

This is shown in Tables 25.1–25.3 for the case of SBP and ischaemic heart disease (IHD). The non-italic numbers in Table 25.1 are the subregion-age-sex specific PAFs estimated using Equation 2. Next, these fractions were applied to the Global Burden of Disease (GBD) 2000 estimates of disease burden for IHD, shown in Table 25.2, producing the estimates of IHD disease burden attributable to SBP in Table 25.3 (similar estimates could be made for mortality or YLL). Dividing the total attributable burden in any subregion (e.g. 1.548 million DALYs for AMR-A in the highlighted cell) by the total IHD burden for the subregion in the GBD database (3.506 million DALYs for AMR-A in the highlighted cell) gives the all-age-sex subregional PAFs (44% for AMR-A in the highlighted cell, obtained by dividing 1.548 by 3.506).

The all-age-sex PAF estimates for the remaining 13 subregions are also shown in Table 25.1 in italics. Similarly, world PAFs were calculated within each age and sex group, as well as overall, using Equation 4, and are shown in the bottom row of Table 25.1. For example, the world PAF for 60–69 years old males was obtained by dividing the total attributable burden in that age-sex group (4.71 million DALYs) by the total world IHD burden in the GBD database (9.015 million DALYs), giving a PAF of 52%.

Computationally, aggregate PAFs (whether aggregated across age-sex groups or subregions) are equivalent to weighted averages of the subregion-age-sex specific estimates, with weights being the same as the total number of events (i.e. deaths, YLL, or DALYs) for each subregion-age-sex group. In the above example, total subregion-age-sex specific DALYs are the weighting factor. As a result, subregion and world PAFs are weighted more towards the ages and/or subregions that have higher DALYs (rather than those with larger populations). For each risk factor, these separate aggregate PAFs were estimated for deaths, YLL, and DALYs with the corresponding GBD estimates used as the denominator (or weighting factor). As a result, even when the age-sex-subregion-specific PAFs are the same for the three measures, the aggregate ones may differ.

For example, although the subregion-age-sex-specific PAFs are the same for deaths and DALYs in the case of elevated SBP and IHD, dividing the IHD deaths attributable to this risk factor in AMR-A (203 000) by the total IHD deaths in this region (618 000), gives a regional PAF for mortality of 33% in AMR-A. In fact, the smaller subregional PAF for mortality compared to that of DALYs highlights the higher weighting towards the older age PAFs (which are smaller) in the case of

Table 25.1 PAFs for IHD attributable to increased SBP (%), by age, sex and subregion

Subregion	0-4 (years)		5-14 (years)		15-29 (years)		30-44 (years)		45-59 (years)		60-69 (years)		70-79 (years)		≥80 (years)		
	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females	
AFR-D	—	—	—	—	—	—	63	56	70	73	56	63	50	58	16	19	57
AFR-E	—	—	—	—	—	—	51	39	59	61	46	53	40	48	12	15	46
AMR-A	—	—	—	—	—	—	46	20	57	53	49	51	46	48	15	18	44
AMR-B	—	—	—	—	—	—	46	23	61	60	51	57	47	54	14	17	50
AMR-D	—	—	—	—	—	—	50	31	60	58	48	52	43	49	13	15	45
EMR-B	—	—	—	—	—	—	55	57	64	71	53	61	48	56	15	19	55
EMR-D	—	—	—	—	—	—	49	44	61	68	51	61	46	57	14	19	51
EUR-A	—	—	—	—	—	—	66	47	72	70	59	62	53	57	17	19	54
EUR-B	—	—	—	—	—	—	63	49	77	78	64	70	59	66	20	23	64
EUR-C	—	—	—	—	—	—	66	55	74	80	62	73	56	68	18	23	63
SEAR-B	—	—	—	—	—	—	43	39	60	60	51	52	46	48	14	15	46
SEAR-D	—	—	—	—	—	—	33	30	53	49	46	42	42	38	13	11	41
WPR-A	—	—	—	—	—	—	60	40	71	64	59	58	54	53	18	17	52
WPR-B	—	—	—	—	—	—	28	23	51	55	44	51	41	48	12	14	41
World	—	—	—	—	—	—	47	35	61	59	52	54	48	53	15	18	49

— No data.

Table 25.2 GBD 2000 estimates of total disease burden for IHD (000s of DALYs), by age, sex and subregion

Subregion	0-4 (years)		5-14 (years)		15-29 (years)		30-44 (years)		45-59 (years)		60-69 (years)		70-79 (years)		≥80 (years)		Total
	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females	
AFR-D	3	4	6	5	31	48	116	88	242	281	190	189	122	166	37	48	1576
AFR-E	4	5	7	7	49	66	149	88	273	278	200	188	110	148	29	50	1653
AMR-A	1	1	1	0	15	7	214	74	709	260	514	289	470	391	219	340	3506
AMR-B	2	2	3	2	46	21	218	107	594	301	412	279	255	223	69	96	2631
AMR-D	1	2	1	1	10	7	25	11	54	33	42	30	31	24	9	11	294
EMR-B	3	5	6	4	32	17	179	59	411	143	219	128	114	99	26	29	1474
EMR-D	24	24	21	12	90	84	322	170	771	475	522	469	292	328	69	73	3746
EUR-A	1	0	1	7	15	15	181	40	657	146	717	284	671	518	239	389	3882
EUR-B	1	0	1	1	35	15	266	78	734	258	647	440	431	471	93	176	3647
EUR-C	0	0	1	1	78	15	693	127	1708	477	1496	920	847	1223	171	563	8319
SEAR-B	7	3	4	3	98	49	222	123	394	279	318	286	188	188	43	54	2259
SEAR-D	70	52	99	58	283	622	1105	900	3688	2119	2657	2361	1412	1484	278	291	17480
WPR-A	0	0	3	3	10	6	40	10	151	40	135	59	112	85	46	66	765
WPR-B	8	7	14	7	155	77	495	278	1099	665	945	783	682	776	182	340	6513
World	125	107	168	111	946	1049	4225	2154	11484	5755	9015	6704	5737	6127	1510	2526	57743

Table 25.3 Burden of IHD attributable to increased SBP (000s of DALYs), by age, sex and subregion

Subregion	0-4 (years)		5-14 (years)		15-29 (years)		30-44 (years)		45-59 (years)		60-69 (years)		70-79 (years)		≥80 (years)		
	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females	
AFR-D	—	—	—	—	—	—	73	49	168	205	107	119	61	96	6	9	893
AFR-E	—	—	—	—	—	—	76	35	160	171	92	99	45	72	3	8	760
AMR-A	—	—	—	—	—	—	97	15	403	137	251	146	215	189	34	61	1 548
AMR-B	—	—	—	—	—	—	101	25	364	180	211	158	119	120	10	16	1 303
AMR-D	—	—	—	—	—	—	12	3	33	19	20	16	13	12	1	2	1 32
EMR-B	—	—	—	—	—	—	98	34	264	101	116	78	55	56	4	5	811
EMR-D	—	—	—	—	—	—	156	75	473	324	265	285	134	186	10	14	1 922
EUR-A	—	—	—	—	—	—	120	19	476	102	422	175	354	296	41	74	2 079
EUR-B	—	—	—	—	—	—	167	38	565	201	417	308	256	309	18	40	2 320
EUR-C	—	—	—	—	—	—	456	71	1 270	381	927	670	475	826	31	132	5 239
SEAR-B	—	—	—	—	—	—	95	48	237	168	161	149	87	90	6	8	1 050
SEAR-D	—	—	—	—	—	—	367	270	1 972	1 030	1 222	988	595	568	35	33	7 080
WPR-A	—	—	—	—	—	—	24	4	107	26	80	34	60	46	8	12	400
WPR-B	—	—	—	—	—	—	138	65	557	366	420	399	278	370	22	49	2 664
World	—	—	—	—	—	—	1 983	750	7 049	3 412	4 710	3 625	2 746	3 234	229	463	28 201

— No data.

mortality, because greater numbers of IHD deaths occur in these age groups. On the other hand, because deaths at younger ages contribute to larger loss of life (YLL), when DALYs are considered, the contribution of PAFs at younger ages to the all-age-sex PAF becomes greater.

5. EXCEPTIONS TO THE GENERAL ESTIMATION PROCEDURE

The following list briefly describes the major departures from the standard analysis framework which were required so that all risk factors could be adequately assessed within the project. Further details are provided in the individual risk factor chapters.

- Theoretical minima varied by age, sex and subregion for iron deficiency, since this was the haemoglobin distribution that would be observed if iron deficiency were eliminated from each population.
- Theoretical minima varied by age, sex and subregion for lack of contraception as a risk factor to reflect different fertility preferences across populations.
- Fruit and vegetable intakes in any population were truncated at zero. In other words, all individuals falling below zero in the distribution were allocated a value of zero. Sensitivity analyses were also performed to assess the effects of possible skewness in the distribution of fruit and vegetable intake.
- The burden of cardiovascular diseases attributable to lead exposure was estimated by assessing different scenarios of elevated blood pressure due to lead and then estimating the total mediated effect through blood pressure.

6. OTHER METHODOLOGICAL ISSUES

It has been shown that the attributable fraction estimates using the PIF relationship in Equation 1 lead to biased estimates when the relative risk has been adjusted for confounding (Greenland 1984; Greenland and Robins 1988). This bias is in fact a result of the correlation among multiple risks (the risk factor of interest and other risk factors that act as confounders), as well as the diseases affected by them (Ezzati et al. 2003). Accounting for this correlation in the estimation of attributable burden, however, would require the availability of exposure and disease data stratified by the confounding variable(s). In general, such stratified data are not available and therefore reliance on the formula with direct use of the adjusted relative risk estimates was necessary. In the case of positive correlation among risk factors, this would generally result in an underestimation of population attributable fraction.

NOTE

- 1 See preface for an explanation of this term.

REFERENCES

- Drescher K, Becher B (1997) Estimating the generalized impact fraction from case-control data. *Biometrics*, **53**:1170–1176.
- Eide GE, Heuch I (2001) Attributable fractions: fundamental concepts and their visualization. *Statistical Methods in Medical Research*, **10**:159–193.
- Ezzati M, Vander Hoorn S, Rodgers A, Lopez AD, Mathers CD, Murray CJ, Comparative Risk Assessment Collaborating Group (2003) Estimates of global and regional potential health gains from reducing multiple major risk factors. *The Lancet*, **362**:271–280.
- Greenland S (1984) Bias in methods for deriving standardized morbidity ratio and attributable fraction estimates. *Statistics in Medicine*, **3**:131–141.
- Greenland S, Robins JM (1988) Conceptual problems in the definition and interpretation of attributable fractions. *American Journal of Epidemiology*, **128**:1185–1197.
- Lauer R, Clarke W (1988) A longitudinal view of blood pressure during childhood: the Muscatine Study. *Statistics in Medicine*, **7**:47–57.
- Murray CJL, Lopez AD, eds. (1996) *The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020*. Global Burden of Disease and Injury, Vol 1. Harvard School of Public Health on behalf of WHO, Cambridge, MA.
- Voors A, Webber L, Berenson G (1979) Time course studies of blood pressure in children—the Bogalusa Heart Study. *American Journal of Epidemiology*, **109**:320–334.
- Walter SD (1980) Prevention of multifactorial disease. *American Journal of Epidemiology*, **112**:409–416.
- Wilsgaard T, Jacobsen BK, Schirmer H et al. (2001) Tracking of cardiovascular risk factors: the Tromso Study, 1979–1995. *American Journal of Epidemiology*, **154**:418–426.