Avoidable burden of disease: conceptual and methodological issues in substance abuse epidemiology

JÜRGEN REHM,^{1,2,3} BENJAMIN TAYLOR,^{1,2} JAYADEEP PATRA,^{1,4} GERHARD GMEL^{1,5,6}

1 Centre for Addiction and Mental Health, Toronto, Ontario Canada

2 Department of Public Health Sciences, University of Toronto, Canada

3 Research Institute for Public Health and Addiction, Zurich

4 Department of Human Development and Applied Psychology, University of Toronto, Canada

5 Swiss Institute for the Prevention of Alcohol and Drug Problems, Lausanne, Switzerland

6 Alcohol Treatment Center, Lausanne University Hospital, Switzerland

---

Abstract

Determining the proportion of avoidable disease burden attributable to substance use is important for both policy development and intervention implementation. Current epidemiological theory has in principle provided a method to estimate avoidable burden of disease and the available statistical tools can provide first rough estimates. The method described in this paper, and its statistical procedures, are exemplified to estimate avoidable burden of tobacco-related disease in Canada. However, further effort is needed to find solutions in the methodological details, namely exposure measurement, risk factor multidimensionality, estimation of changes in exposure distribution over time, and estimation of risk relationships from multiple exposures changing over time with multiple endpoints (causal webs). The impetus to begin refining methods to obtain better starting points for estimating avoidable burden of disease is obvious and should be carried through in order to see real changes through evidence-based policy and intervention. Copyright © 2006 John Wiley & Sons, Ltd.

Key words: substance use, burden of disease, risk factor, distribution of exposure, avoidable burden, tobacco, Canada

---

Introduction

The social costs of substance abuse in developed countries are high (Single et al., 1998; Andlin-Sobocki, 2004; Andlin-Sobocki and Rehm, 2005) and mainly related (over 85%) to costs of health care, law enforcement and losses in productivity (Harwood et al., 1998; Collins and Lapsley, 2002). Mortality and morbidity attributable to the use of alcohol, tobacco and illegal drugs are the main underlying components of substance-use related social costs (Single et al., 1998; Andlin-Sobocki, 2004; Andlin-Sobocki and Rehm, 2005). Knowing the overall attributable costs of substance use or knowing the overall substance-use attrib-

utable disease burden has been found helpful in practice but it is not clear what proportion of the costs or the burden could be changed in principle. This changeable part has been labelled the 'avoidable cost' or 'avoidable burden' (World Health Organization, 2002).

The first step in estimating avoidable burden is conceptualizing attributable burden of disease – the burden of a given disease in a given population that is identified as due to a specific past and present exposure of one or multiple risk factors (for a more formal definition of 'attributable' see below). Burden of disease is used here in the tradition of the Global Burden of Disease (GBD) study as comprising fatal and non-fatal health

outcomes, in the GBD measured by disability-adjusted life years (DALYs) (Murray and Lopez, 1996, 1997). The DALY is a health gap measure that extends the idea of potential years of life lost due to premature death, to include equivalent years of 'healthy' life lost to poor health or disability (Murray and Lopez, 1996, 1997). Information on both mortality and non-fatal health outcomes is thus combined into a single measure of population health. To calculate DALYs for a disease or injury cause in a population, the years of life lost due to premature mortality from the cause are added to the number of years lost due to disability from incident cases of the disease or injury. We will later discuss how independent the proposed approach is with respect to burden definition.

Thus defined attributable burden of disease could, in principle, be reduced or eliminated if the causative exposure is reduced or eliminated. The stress here is on 'in principle', as attributable burden is conceptualized whether or not such a reduction is achievable in practice. Thus, it represents the maximal change in incidence of an outcome that can occur if the risk factor were to be removed – not the change that would occur if the risk factor were to be removed. In fact, for a maximal reduction, it must be assumed that the risk factor is a causal risk factor, that the risk factor can be fully manipulated and that changing the risk factor would not change other causal risk or protective factors.

Based on the conceptualization of attributable burden specified above, it is then possible to introduce the term 'avoidable burden of disease'. This term denotes the proportion of disease burden that can be reduced by changing the current exposure distribution to an alternative, more favourable, exposure distribution. Clearly, the size of the avoidable burden caused by a given risk factor will always be smaller than or, at best, equal to the burden attributable to that risk factor. The aim of this article is to discuss a method to estimate avoidable burden specifically for substance use as a risk factor, with special emphasis on methodological problems and potential solutions. We are using the term 'substance use' rather than abuse because it has been shown that a large portion of the burden attributable to substances is caused by use only – by individuals who do not meet the gold standard definitions of substance dependence or abuse (Rehm, 2003). An example would be an alcohol-attributable death caused by a person who drank four glasses of wine with a friend, drove a car and caused an accident because of alcohol-

induced impairment, even though this person usually did not consume more than one glass per day. This person may not fulfil any criterion of dependence or abuse found in the International Classification of Diseases, Tenth Revision (ICD-10) (World Health Organization, 1992–4) or the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV-R) (American Psychiatric Association, 1994).

From the classical to the distributional approach in epidemiology

Two different approaches to the epidemiology of risk factor attribution can be broadly distinguished: the classical epidemiological approach where, on the basis of a 2×2 table, disease risk can be estimated with respect to exposure and non-exposure of various populations to a single specific risk factor. Based on this categorically derived relative risk and the prevalence, the attributable burden can be derived, resulting in an estimate for a population that has *already* been exposed – focusing on the past. The main counterfactual scenario (Maldonado and Greenland, 2002) in this approach asks the question: 'What would have happened if no exposure had occurred?'

More modern developments of epidemiology ask the question: 'What would happen if risk factor distributions shifted to different counterfactual scenarios?' (Murray and Lopez, 1999). The modern approach not only looks at distributional shifts at one time but also takes the future time dimension into consideration and thus is able to predict future developments.

The current contribution will focus on the second approach. It will thus conceptualize avoidable burden by specifying alternative scenarios for risk factor distributions, including potential future distributions. It is important to look at the impact of a risk factor over time, not just an epidemiological snapshot of today's attributable burden, in order to estimate fully the contributions of acute and chronic disease burden that might be avoided in the future. This is an especially important factor to consider for policy development. We will also consider potential interactions and competing risks between influencing factors.

Please note that the proposed framework will be conceptualized independently from the measure for burden of disease used, whether it is mortality, morbidity, or any summary measure (such as the traditionally used DALY – cf. Murray and Lopez, 1997), so that it can be universally applied. The next section of this

paper will discuss the conceptual basis for estimating avoidable burden in the context of the issues outlined above.

An epidemiological model to estimate avoidable burden

Figure 1 illustrates the conceptual model of the difference between attributable and avoidable burden, cited from the epidemiological model of Murray and colleagues (Murray et al., 2003).

Let us consider the disease burden at time T_0 (that portion which is attributable to prior exposure), denoted by the letter a in Figure 1. This is all the burden of disease that can be attributed to prior exposure of the risk factor under consideration before T_0 .

In the example, a general situation is given where the background burden – the burden due to factors other than the risk factor under consideration – is constant over time. This background burden is denoted by the letter b at time T_0 and, because it is constant over time in this specific example, the burden is the same size at all time points. In other situations, of course, the burden attributable to other factors than the risk factor under consideration may fluctuate.

The burden attributable to the risk factor under consideration in the example is increasing constantly over time until time T_0 . An example for this situation

would be the second stage of a tobacco epidemic in developed countries (Lopez et al., 1994). Several different scenarios are shown. Let us discuss three of them. If nothing (for example, intervention, changes in cultural acceptance) happens at time T_0 , the attributable burden will continue to increase linearly. On the other hand, if the exposure is reduced completely, then the attributable burden will decrease until time T_x , when it reaches zero. Thirdly, if there is reduction of exposure, the resulting attributable burden will lie between these two extremes.

Consider tobacco use as an example in a society where prevalence rates have been increasing and would continue to increase if there was no intervention. If some drastic intervention could reduce smoking completely at a certain time point, smoking-related disease burden would not be zero one moment later. Instead, some burden of disease would persist – for example, the burden of disease due to existing tobacco-related lung cancer – and some people may even develop new lung cancers based on their past exposure.

Now consider a reduction of exposure to the risk factor by 50% at time T_0 . In the example, such a reduction would mean that a constant attributable burden would result. This burden is, of course, a mixture of the impact of past exposure (i.e. prior to T_0) plus the impact of the exposure after T_0 .

Figure 1. A conceptual model of attributable and avoidable risk with increasing projected burden (Murray et al., 2003).

Three main components have to be known to estimate this model. They are:

- The relationship between exposure (i.e. the risk factor under consideration) and disease burden (attributable risk) and its temporal trend.
- The amount of burden caused by factors other than the risk factor under consideration and the temporal stability of this burden.
- The specific trajectory of burden reduction after exposure reduction.

Methods to estimate the relationship between a certain exposure distribution and disease burden are well established (for example, odds ratios, relative risks – see Rothman and Greenland, 1998). Similarly, methods to estimate the proportion of burden attributable to the distribution of a certain risk factor were first described by Miettinen (1972) in the early 1970s and by Walter a little while later (Walter, 1976, 1980). They have since been used to generate estimates of the attributable burden for substance use and other risk factors around the world (Ezzati and Lopez, 2000). The key concept used here is that of an attributable fraction (AF) (also called the etiologic fraction). It depends on the exposure distribution (or, in the discrete case, on the prevalence of different exposure categories) and on the relative risk for burden related to the respective exposure levels.

Below are the two formulas for the case of continuous exposure levels and the discrete case, both of which are fully developed and described elsewhere (Walter, 1976, 1980; Eide and Heuch, 2001; Murray et al., 2003). The contribution of a risk factor to disease can be estimated by comparing the burden due to the observed exposure distribution in a population with that from another *distribution* (rather than a single reference level such as non-exposed) as described by the generalized equation shown as Equation 1.

Equation 1

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

where *PIF* is the 'potential impact fraction', a generalized form of the attributable fraction; *RR(x)* is the rela-

tive risk at exposure level *x*, *P(x)* is the population distribution of exposure, *P'(x)* is the counterfactual distribution of exposure, and *m* the maximum exposure level.

As Murray et al. (2003) have noted, this formula can be further generalized to deal with a situation, where the relative risks change in the counterfactual scenario. The corresponding relationship when exposure is described as a discrete variable with *k* levels is given by Equation 2 (Walter, 1976, 1980):

Equation 2

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

where

I = exposure level category

RR_i = relative risk at exposure level *i*

P_i = prevalence of the *i*th category of exposure under the counterfactual scenario

The concept of attributable fraction (or the generalized form of *PIF*), as defined here, can only describe a snapshot at a specific time – it only gives a very rough estimate of attributable burden accumulated by past and present exposure. Attributable fractions that do not include a time dimension are not able to characterize those cases whose occurrence would have been delayed or in part prevented due to exposure reduction (see Greenland and Robins, 1988). As a remedy, Greenland and Robins (1988) recommend the use of etiologic fractions with a time dimension to account for this shortcoming. Time-based measures are discussed in some detail by Murray et al. (2003) in their seminal conceptual framework of the comparative quantification of health risks. However, they have rarely been used in practice and mainly appear in methodological contributions.

The theoretical minimum burden (see Figure 1) and the theoretical minimum risk are trickier to define than the exposure-burden relationship. The *theoretical minimum risk* denotes the exposure distribution that would result in the lowest population burden, irrespective of whether such a distribution is currently attainable in practice (Murray and Lopez, 1999). In the example of Figure 1, the theoretical minimum burden

was achieved by 100% exposure reduction – a theoretical minimum risk at exposure of zero was associated with zero exposure-related burden in the long run. This is not necessarily always the case. Consider alcohol, and assume only one relevant exposure dimension in disease aetiology (see below for a discussion of volume and patterns of drinking as a two-dimensional exposure association), such as average volume of drinking. Minimal risk then occurs at zero (i.e. no drinking at all) for most related diseases such as cancer or traffic injury, but not for some whose risk actually decreases at exposures greater than zero, for example heart disease (Rehm et al., 2003a,b; Rehm et al., 2004). For composite outcome measures, such as all-cause mortality, there is also reason to believe that the level of exposure associated with the minimum burden is greater than zero (for example, some level of drinking) (Rehm et al., 2001; Gmel et al., 2003). The exact value of exposure associated with the minimal burden will depend on the composite measure used (such as mortality or DALYs), the disease distribution in the country or region examined (such as the degree to which IHD mortality is a main contributor or overall mortality), and the predominant drinking pattern (regular and moderate use versus irregular and heavy use). Thus, the theoretical minimum risk measured with average volume only will fluctuate across cultures due to several factors such as drinking patterns or disease distribution.

A theoretical minimum burden with an exposure greater than zero has interesting implications. For alcohol in the above example it means that, even if drinking occurs at the theoretical minimum, there will be an attributable burden for some diseases. For instance, in a society with relatively large ischaemic burden and assuming a theoretical minimum risk for the population occurring at approximately one drink/day, there will be disease burden associated with such an exposure level of moderate drinking, for example for certain gastrointestinal diseases (Taylor et al., 2005) or accidents (Rehm and Gmel, 2003).

The above example is hypothetical for several reasons:

- It assumes all people having the same exposure level (1 drink/day for all people) rather than a distribution of exposure, which is what is usually found in populations (see, for example, Skog, 1985).
- It completely disregards the fact that alcohol has at least two dimensions relevant for disease, average

volume of alcohol consumption and patterns of drinking (Rehm et al., 2003a).

- There is no known intervention, which would result in the above-mentioned distribution of 1 drink/day for all (Babor et al., 2003; Room et al., 2005).

In more general terms, the following points can be made:

- We should always model shifts in risk factor distributions when estimating avoidable burden of disease.
- We should always model all relevant dimensions of a risk factor when estimating avoidable burden of disease.
- We need other counterfactual scenarios in addition to the theoretical minimum risk, when estimating avoidable burden of disease. The question will be where to obtain the counterfactual scenarios.

The trajectory of the burden reduction after changes in exposure is also difficult to define, not only because it should incorporate changes of exposure reduction as well but because it also has to make estimates for the change of relative risks over time of several disease outcomes, potentially both acute (for example, alcohol and deaths from drinking and driving) and chronic (for example, alcohol and chronic pancreatitis). For instance, after a change of smoking status to abstinence, we have to know the relative risk of an ex-smoker after one year after abstaining, two years after abstaining, three year after abstaining and so on. For acute outcomes, the problem is much easier. Once the prevalence of alcohol, tobacco or illicit drugs is reduced, all the acute outcomes (such as injuries) are reduced accordingly. To give an example: while drinking in previous years does affect the cancer risk of people today, even if they started abstaining in between, it does not affect the traffic accident risk.

An example: shifting the smoking distribution in Canada

Below is a simplified example applying distributional shifts in exposure to tobacco consumption in Canada using smoking prevalence data from the 2003 Canadian Community Health Survey and mortality data from the Statistics Canada in 2002. A hypothetical decrease in exposure prevalence of 10% per category

was run – the prevalence of each smoking category was reduced by 10% shifting into the next lower category and the 10% from the lowest smoking category shifted into the former smoker category. To give an example: for men in the second highest smoking category (smoking between 15 and 24 cigarettes/day) the prevalence in this category was 6.55% before the shift. Then, 10% of the of 6.55% or 0.66% went to the next lower smoker category (smoking between one and 14 cigarettes/day) but 0.55% were added coming from the highest smoking category, resulting in a net decrease of 0.11% in this category. All other cells were modelled accordingly. The resulting changes in prevalence are shown in Table 1.

Table 1 shows the gender-specific shift in prevalence of smoking for different categories before and after a 10% decrease in each exposure prevalence category as described above. Overall, smoking prevalence, indepen-

dent of dose, decreased by 2.2% for both genders. The impact that this shift of distribution had on tobacco-attributable disease-specific mortality was modelled based on the mortality data for 2002. The baseline scenario was taken from the second Canadian study on social costs of substance abuse where tobacco-related attributable burden was modelled including information on gender, age and disease specific with a counterfactual scenario of 0 smoking as the level denoting minimal risk (Baliunas et al., 2005). To estimate avoidable burden, the shift in prevalence was again modelled specifically by gender, age and disease, resulting in the reduction of smoking-related mortality summarized in Table 2. The relative risk to denote the relationships between category of exposure and outcome were taken from meta-analyses (for details see Baliunas et al., 2005). Please note that far more than 20 different disease categories had to be modelled in order to arrive at Table 2.

Table 1. Resulting prevalence of tobacco smoking in (percentages) after a 10% shift in exposure distribution in Canada 2003, by gender

Smoking categories	Men		Women	
	Before	After	Before	After
Never	29.12	29.12	40.39	40.39
Former	44.59	45.18	37.80	38.29
Current/occasional	5.90	6.14	4.90	5.34
1–14 cigarettes / day	8.35	8.17	9.28	8.80
15–24 cigarettes / day	6.55	6.44	4.51	4.37
25+ cigarettes / day	5.49	4.94	3.12	2.81
Total	100.00	100.00	100.00	100.00

Table 2. Impact of shift of –10% in exposure to tobacco smoking on disease-specific, tobacco-attributable mortality in Canada, 2002 (number of deaths)

	Lung Cancer		Other Cancers		Cardiovascular Disease		Respiratory Disease		Passive Smoking (lung cancer and IHD)		Total†	
	M	F	M	F	M	F	M	F	M	F	M	F
Before shift	9127	4753	4199	1795	8476	6604	4776	3810	757	528	27563	17655
After shift	9026	4553	4160	1784	8267	6394	4761	3797	683	475	27117	17164
Diff.	101	200	38	12	209	210	15	13	74	53	447	491
Total	301		50		419		28		127		938	

†Totals represent mortality due to all outcomes, not row totals of those selected.

A distributional shift of 10% in the exposure level to tobacco smoke resulted in 938 fewer tobacco-attributable deaths, 2.1% less than before the shift. The largest absolute differences before and after were seen for lung cancer and cardiovascular diseases, and also for deaths due to lung cancer and ischemic heart disease attributable to passive smoking. However, relative differences were most pronounced for the acute outcomes.

The above example illustrates the possibilities of modelling avoidable burden using the framework specified above. It also illustrates some points that still have to be improved. First, smoking-related attributable burden is still modelled using current levels of smoking as an indicator of cumulated past exposure – the temporality of processes is mainly lacking or only crudely modelled by the category of former smokers. The procedure presented is usual in the literature and is the basis of the most commonly used software for calculation of smoking-attributable mortality but it clearly introduces measurement error, which may increase, as smoking behaviour across the lifespan does not seem to be as stable as it used to be. Second, the distributional shift is somewhat arbitrary, as it is not clear why a 10% shift should be the basis for avoidable burden rather than, for example, a 15% shift, nor why the same proportional shift downwards to the next lower category should apply to all smoking categories. For instance, in interventions underlying reductions in smoking prevalence such as taxation, different shifts in smoking prevalence distributions might be seen (for example, changing from the highest smoking level to abstinence). Finally, it is once again a snapshot picture, depicting one change without adequately incorporating the time dimension. Thus, this example can only be seen as a first step in modelling avoidable burden of tobacco.

Discussion

Problems with estimation

1. The theoretical and feasible minimum risk

The level of exposure associated with the minimal attributable burden, the *theoretical minimal risk*, is not an obvious definition. Traditionally, an exposure of zero has been used as the counterfactual scenario, but an exposure of zero is not necessarily associated with minimal burden in the field of substance use (see above),

where certain levels of use, such as light-to-moderate alcohol consumption without heavy drinking episodes, are considered as beneficial for certain disease outcomes. However, when using a non-zero exposure as minimal risk category, and considering several disease outcomes, each of which is associated with different exposure levels for minimal burden, we have the difficulty that the associated combined minimal risk level will depend on the disease distribution in society. For example, the exposure level of alcohol consumption associated with minimal burden of disease will be lower in societies with lower rates of ischaemic mortality for alcohol. In practice, this problem has often been circumvented by using zero exposure as the counterfactual for all substance use and doing sensitivity analyses (see Single et al., 1996, 1999 for example).

However, the theoretical minimum is not what is needed to estimate the avoidable burden in a society. To achieve this portion, Armstrong defined an Arcadian normal as a *feasible minimum* based on disease prevalence in the following way (Armstrong, 1990):

The first approach [of estimating avoidable burden, the author(s)] assumes the existence of some level of disease that might reasonably be achieved if only we knew all that might reasonably be known about the causes of the disease in question and could apply them in practical programs in the community . . . I have termed this level of disease the *Arcadian normal* because it represents the nearest approach that we can make to harmony between humankind and its environment. Arcadia in ancient Greece was a region renowned for the contented pastoral simplicity of its people.

Armstrong operationalized this definition by comparing the lowest disease-specific rates in several similar countries. Thus, for instance, for ischemic heart disease (IHD), the 'healthiest' country in 1990 was France with 76 per 100,000 age-standardized deaths, compared to 231 per 100,000 age-standardized deaths in Australia. With this logic, about 67.1% of the IHD deaths in Australia would be potentially preventable (= avoidable) using France as the Arcadian normal. This method was extended to calculate preventable deaths caused by all major disease categories.

However, there are some problems with the Arcadian normal and thus it does not fit well into our

framework. First, it is based on disease and thus is not exposure specific. Different countries will still have different, less than optimum, exposure profiles that could be changed to lower the specific disease burden, but the Arcadian normal does not recognize these exposure-specific differences: it just covers broad summary outcome measures. Second, the theoretical minimum risk needed in our framework is for burden of disease across different diseases, not for an individual disease category. Summing up across disease categories has the additional problem that the Arcadian normal will stem from different countries (France for IHD, Austria for infectious disease, Sweden for digestive disease and so forth), which means different burdens from risk factors other than the exposure under consideration have to be taken into account. Third, countries are so different with respect to other factors that they are not comparable as regards achievable disease reduction. There might also be other reasons, such as cultural differences (for example, Muslim countries with very low alcohol consumption), which explain different exposure levels, and such 'feasible' exposure levels might not be really feasibly achieved in other countries.

Although the Arcadian normal of Armstrong (1990) cannot work for the proposed model, an exposure-specific feasible minimum of disease burden could be introduced. The concept of the Arcadian normal thus could be transferred to exposure and used to define feasible exposure changes based on the minimum of exposure distributions in similar societies achieved by intervention. It would thus introduce a feasible minimum risk (Murray and Lopez, 1999) – defined as an exposure distribution that has already been achieved in comparable societies. Using a feasible minimum will result in more practically achievable solutions and thus less avoidable burden than if using a purely theoretical minimum. One could also estimate pathways towards the avoidable burden in hypothetical scenarios based on changes to current risk factor exposure making this a powerful statistical tool for policy and intervention development. However, we would still have to struggle with problems of comparability between countries.

Thus it might be better not to formally define any feasible minimum of burden of disease and instead model shifts in risk factor distributions in the direction of the theoretical minimal burden. The first step in such an approach is given in the example for tobacco use in Canada mentioned above.

However, to select which shifts should be modelled and to see how realistic they are it may still be helpful to look at similar countries or historical trends as comparators. Thus, one would first screen plausible developments based on historical trends or the distribution in other countries and then model shifts in risk factor distributions and thus the avoidable burden associated with these shifts. Thus, for the example above, one would look at the distributional changes resulting from intervention packages being used in countries or regions similar to Canada and model avoidable burden accordingly.

2. Determining risk relationships

There are well-defined statistical methods to define attributable burden of certain factors (Walter, 1980 – see above), which have been shown to work well in practice (Single et al., 1998). However, there are still problems to overcome, which will be detailed below.

The first problem is related to exposure measurement. Even for licit substances – for example when trying to assess alcohol consumption in national surveys – it is often estimated that only 60% or even less of the total consumption is actually captured (Midanik, 1988). In other words, the true exposure to alcohol can only be measured with huge measurement error in surveys; survey estimates usually form the basis for prevalence of exposure estimates. Moreover, the prevalence of alcohol consumption at a certain point does not necessarily capture the accumulated exposure, which is causative for disease. As most relative risk estimates for disease were derived from similar imperfect measurements of exposure, the two biases – imperfect measurement of exposure and imperfect measurement of relative risks at these levels of imperfectly measured exposures – might counterbalance each other. On the other hand, they might not do so.

Similar problems exist for tobacco exposure and its relation to disease (Peto et al., 1992). It is inherently difficult to record prevalence of illegal drugs because of their illegal nature (Warner-Smith et al., 2001). Solutions to these types of problems are beyond the scope of this paper and are limitations in the field.

The second problem is that interactions between risk factors and their relationships over time must be accounted for in order to estimate attributable and avoidable burden of disease accurately. Causal webs for specific diseases are complex and multidimensional, making it very difficult to determine major influencing

factors (Murray and Lopez, 1999). Estimations of causal webs have been shown to be possible (Robins et al., 1999; 2004) but they require strong assumptions. There is very little literature on the subject to indicate if such assumptions are realistic and what the consequences might be of violations of these assumptions. Moreover, in such causal webs there is the problem of comorbidity in endpoints.

Consider the case of depression and alcohol dependence as multiple endpoints. Three causal relations are likely to exist between these disorders: heavy alcohol use and dependence may cause depression, depression may lead to heavy alcohol use and dependence, for example as a way to self-medicate their dysphoria, and a third variable, such as genetic vulnerability, may cause both disorders (Rehm et al., 2003b, 2004). This situation often leads to non-recursive models with feedback loops, which may not even be estimable with multiple measurements over time without strong assumptions. On the other hand, if burden or avoidable burden attributable to alcohol is estimated, burden from depression clearly should be one part (Rehm et al., 2004). This is also important from a practical perspective, as depressive symptoms have been reported to disappear in many heavy drinkers or alcohol dependent persons during periods of abstinence, either in the natural course, or as a result of treatment (for example, Brown and Schuckit, 1988; Schuckit et al., 1997).

Thus, future research should not only include risk estimates for the exposure of interest, but also for interactions of this exposure with other exposures or confounders (Taylor and Rehm, 2006). In addition, multiple endpoints and reciprocal causation have to be accounted for by estimating causal webs. To account for changing exposure distributions over time, distributional changes must be modelled. In order to achieve this, we need study cohorts that have exposure distributions similar to the general population distribution instead of strictly defined groups such as British doctors or American nurses.

Conclusion

Current epidemiological theory and statistical techniques provide a novel way to estimate avoidable burden of disease under different exposure scenarios for substance abuse. The example in this paper used tobacco consumption in Canada and this method showed a 2.1% decrease in mortality associated with a 10% down-

ward shift into the next lesser exposure prevalence category. There are a number of details that must be addressed in the future, specifically measurement of exposures individually and accounting for their interactions with other factors, time and multidimensionality. The methods described in this paper and their future development may have important implications for policy development and intervention implementation.

Acknowledgements

The authors wish to thank Health Canada and the Canadian Centre on Substance Abuse for their financial support (Projects: Second Study on Social Costs of Substance Abuse in Canada under the umbrella of the Canadian Centre on Substance Abuse; Estimating avoidable burden of substance-related disease by Health Canada). In addition, we received valuable comments by two anonymous reviewers and the participants of the workshop on *Guidelines for the Estimation of the Avoidable Costs of Substance Use and Abuse* hosted by Health Canada in Ottawa, Canada, 2005. Lastly, thanks to Dolly Baliunas and Lana Popova for their helpful comments in previous drafts of this paper.

References

- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4 edn. Washington DC, American Psychiatric Association, 1994.
- Andlin-Sobocki P. Economic evidence in addiction: a review. *Eur J Health Econ* 2004; 5(Suppl.1): S5–12.
- Andlin-Sobocki P, Rehm J. Cost of addiction in Europe. *Eur J Neuro* 2005; 12(1): 28–33.
- Armstrong BK. Morbidity and mortality in Australia: how much is preventable? In J McNeill, R King, G Jennings, J Powles. (eds) *A Handbook of Preventative Medicine*. Melbourne: Edward Arnold, 1990.
- Babor T, Caetano R, Casswell S, Edwards G, Giesbrecht N, Graham K, Grube J, Gruenewald P, Hill L, Holder H, Homel R, Österberg, E, Rehm J, Room R, Rossow I. *Alcohol: No Ordinary Commodity. Research and Public Policy*. Oxford: Oxford University Press, 2003.
- Baliunas D, Patra J, Rehm J, Taylor B, Popova S, Kaiserman M. Tobacco-attributable Mortality and Potential Years of Life Lost in Canada 2002: Conclusions for Prevention and Policy. Toronto: Centre for Addiction and Mental Health, 2005.
- Brown SA, Schuckit MA. Changes in depression among abstinent alcoholics. *J Stud Alcohol* 1988; 49: 412–17.
- Collins D, Lapsley H. Estimating the Economic Costs of Drug Abuse in Australia. National Campaign Against Drug Abuse Monograph No.49. Canberra: Commonwealth of Australia, 2002, available at http://www.ocat.org/pdf/socialcosts_Australia.pdf, accessed 8 November 2006.

- Eide GE, Heuch I. Attributable fractions: fundamental concepts and their visualization. *Stat Meth Med Res* 2001; 10: 159–93.
- Ezzati M, Lopez AD. Regional, disease specific patterns of smoking-attributable mortality in 2000. *Tob Con* 2004; 13(4): 388–95.
- Gmel G, Gutjahr E, Rehm J. How stable is the risk curve between alcohol and all-cause mortality and what factors influence the shape? A precision-weighted hierarchical meta-analysis. *Eur J Epi* 2003; 18(7): 631–42.
- Greenland S, Robins JM. Conceptual problems in the definition and interpretation of attributable fractions. *Am J Epidemiol* 1988; 128: 1185–97.
- Harwood H, Fountain D, Livermore G. The Economic Costs of Alcohol and Drug Abuse in the United States 1992. Report prepared for the National Institute on Drug Abuse and the National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Department of Health and Human Services. NIH Publication No. 98-4327. Rockville MD: National Institutes of Health, 1998, <http://www.nida.nih.gov/EconomicCosts/Index.html>, accessed 30 October 2006.
- Lopez AD, Collishaw NE, Piha T. A descriptive model of the cigarette epidemic in developed countries. *Tob Control* 1994; 3: 242–7.
- Maldonado G, Greenland S. Estimating causal effects. *Int J Epi* 2002; 31: 422–9.
- Midanik LT. Validity of self-reported alcohol use: a literature review and assessment. *Br J Addict* 1988; 83(9): 1019–30.
- Miettinen OS. Standardization of risk ratios. *Am J Epidemiol* 1972; 96(6): 383–8.
- Murray CJ, Lopez A. Mortality by cause for the eight regions of the world: the global burden of disease study. *Lancet* 1997; 349(9061): 1269–76.
- Murray CJ, Lopez A. On the comparable quantification of health risks: lessons from the Global Burden of Disease Study. *Epidemiol* 1999; 10(5): 594–605.
- Murray C, Lopez A. The Global Burden of Disease. Cambridge MA: Harvard School of Public Health (on behalf of the WHO and World Bank), 1996.
- Murray CJ, Ezzati M, Lopez AD, Rodgers A, VanderHoorn S. Comparative quantification of health risks conceptual framework and methodological issues. *Pop Health Met* 2003; 1(1): 1–20.
- Peto R, Lopez AD, Boreham J, Thun M, Heath C Jr. Mortality from tobacco in developed countries. *Lancet* 1992; 339: 1268–78.
- Rehm J. Suchtmittel und Public Health. *Suchttherapie* 2003; 4: 72–5.
- Rehm J, Gmel G. Alcohol consumption and total mortality/morbidity-definitions and methodological implications. *Best Pract Res Clin Gastroenterol* 2003; 17(4): 497–505.
- Rehm J, Gutjahr E, Gmel G. Alcohol and all-cause mortality: a pooled analysis. *Contemporary Drug Problems* 2001; 28: 337–61.
- Rehm J, Room R, Graham K, Monteiro M., Gmel G, Sempos CT. The relationship of average volume of alcohol consumption and patterns of drinking to burden of disease – an overview. *Addiction* 2003a; 98(10): 1209–28.
- Rehm J, Room R, Monteiro M, Gmel G, Graham K, Rehn N, Sempos CT, Frick U, Jernigan D. Alcohol use. In M Ezzati, AD Lopez, A Rodgers, CJL Murray (eds) *Comparative Quantification of Health Risks. Global and Regional Burden of Disease Attributable to Selected Major Risk Factors. Volume 1*. Geneva: World Health Organization, 2004, pp. 959–1108.
- Rehm J, Sempos CT, Trevisan M. Average volume of alcohol consumption, patterns of drinking and risk of coronary heart disease – a review. *J Cardio Risk* 2003b; 10 (1): 15–20.
- Robins JM, Greenland S, Hu F-C. Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome. *J Am Stat Assoc – Appl Case Studies* 1999; 94: 687–700.
- Robins JM, Hernan M, Siebert U. Effects of multiple interventions. In M Ezzati, AD Lopez, A Rodgers, CJL Murray (eds) *Comparative Quantification of Health Risks. Global and Regional Burden of Disease Attributable to Selected Major Risk Factors. Volume 2*. Geneva: World Health Organization, 2004, pp. 2199–238.
- Room R, Babor T, Rehm J. Alcohol and public health. *Lancet* 2005; 365(9458): 519–30.
- Rothman KJ, Greenland S. *Modern Epidemiology*. Philadelphia: Lippincott-Williams, 1998.
- Schuckit MA, Tipp JE, Bergman M, Reich W, Hesselbrock VM, Smith TL. Comparison of induced and independent major depressive disorders in 2,945 alcoholics. *Am J Psychiatry* 1997; 154(7): 948–57.
- Single E, Robson L, Rehm J, Xie X. Morbidity and mortality attributable to alcohol, tobacco, and illicit drugs in Canada. *Am J Public Health* 1999; 89(3): 385–90.
- Single E, Robson L, Xie X, Rehm J. The Costs of Substance Abuse in Canada. Ottawa: Canadian Centre on Substance Abuse, 1996.
- Single E, Robson L, Xie X, Rehm J. The economic costs of alcohol, tobacco and illicit drugs in Canada 1992. *Addiction* 1998; 7: 991–1006.
- Skog OJ. The collectivity of drinking cultures: a theory of the distribution of alcohol consumption. *Br J Add* 1985; 80(1): 83–99.
- Taylor B, Rehm J. When risk factors combine: the interaction between alcohol and smoking for aerodigestive cancers, coronary heart disease, and traffic and fire injury. *Addict Behav*, in press.
- Taylor B, Rehm J, Gmel G. Moderate alcohol consumption and the gastrointestinal tract. *Digestive Diseases* 2005; 23(3–4): 170–76.

- Walter SD. Prevention of multifactorial disease. *Am J Epidemiol* 1980; 112: 409–16.
- Walter SD. The estimation and interpretation of attributable risk in health research. *Biometrics* 1976; 32: 829–49.
- Warner-Smith M, Lynskey M, Hall W, Monteiro M. Challenges and approaches to estimating mortality attributable to the use of selected illicit drugs. *Eur Addict Res* 2001; 7(3): 104–16.
- World Health Organization. International Classification of Diseases and Related Health Problems, Tenth revision. Geneva: World Health Organization, 1992–4.
- World Health Organization. The World Health Report 2002: Reducing Risks, Promoting Healthy Life. Geneva: World Health Organization, 2002.
- Correspondence:* J. Rehm, 33 Russell Street, Room 2035B, Toronto ON M5S 2S1, Canada.
Telephone (+1) 416-535-8501 x 4495.
Fax (+1) 416-260-4156.
Email: jtrehm@aol.com.