A Unified Probabilistic Framework for Cancer Risk Assessment

Christoph M. Rheinberger

BACKGROUND: Cancer risk models applied in regulatory contexts are often deterministic. Probabilistic approaches to model exposure to carcinogens, individual and population risk effects and socio-economic impacts are well suited to represent the uncertainty and variability inherent to assessing the health impacts of regulating carcinogens.

OBJECTIVES: I propose a unified framework for probabilistic cancer risk assessment in the context of chemicals regulation.

METHODS: The framework consists of four distinct parts: (i) a probabilistic exposure model that takes into account variability in individual exposure to the substance of concern; (ii) a probabilistic dose-response model that accounts for differences in individual cancer susceptibility; (iii) an impact assessment model that quantifies individuals’ excess lifetime cancer risk; and (iv) a valuation model that values changes in quality-adjusted life expectancy based on workers’ willingness-to-pay for cancer risk reduction and aggregates individual valuations across the population at risk.

RESULTS: I illustrate the approach with data from French workers exposed to hexavalent chromium. The results suggest that limiting the time-weighted average exposure to the national binding occupational exposure limit (BOEL) of 1 µg/m$^3$ per 8h-shift across hard chrome platers in France reduces the statistical worker’s excess lifetime risk of developing fatal and non-fatal lung cancer by 6.7E-2 and 1.5E-3, respectively. In a simulated cohort of 1,000 workers, roughly 33% benefit from the introduction of the BOEL. On the aggregate, the risk reduction corresponds to approximately 12.2 life years gained and 2.7 quality-impaired life years avoided distributed. Assuming a social discount rate of 4%, the corresponding welfare value would be about €3m.

CONCLUSIONS: Although some implementation challenges remain, this unified probabilistic framework for cancer risk assessment is hoped to provide a more complete and transparent characterization of regulatory impacts and support better-informed risk management decisions.

Introduction

Many environmental and health regulations are aimed at reducing cancer risk. Over the last 30 years, methodologies for assessing the impacts of such regulations have constantly improved. Yet there is still a disciplinary wedge between risk assessors charged with establishing the risk reduction potential of a proposed regulatory action and economists charged with determining whether the benefits of this action are proportionate to its costs.

1 Address correspondence to C.M. Rheinberger, Risk Management Directorate, European Chemicals Agency, Annankatu 18, P.O. Box 400, 00121 Helsinki, Finland. Telephone: +358-9-6861-8847. E-mail: christoph.rheinberger@echa.europa.eu. This paper has been prepared for the Harvard Center for Risk Analysis “Risk Assessment, Economic Evaluation, and Decisions” workshop, September 26-27 2019. The views expressed in the paper do not represent official views and positions of the European Chemicals Agency. The author has no actual or potential competing financial interests. He thanks Raymond Vincent for sharing the INRS data.
Following Kaplan and Garrick (1981), risk assessors typically seek to determine ‘what can happen’ under a reasonable worst-case assumption. Economists, instead, are interested in the expected impact of one vs. another course of action. Statistically speaking the two disciplines are thus focused on different moments of the distribution of outcomes.

In this paper, I propose to overcome this disciplinary wedge by introducing a unified probabilistic framework for cancer risk assessment. The proposed framework consists of four distinct parts: (i) a probabilistic exposure model that determines how many individuals are exposed at what level of exposure to a hazardous substance; (ii) a probabilistic dose-response relationship that accounts for differences in individual cancer susceptibility; (iii) an impact assessment model that quantifies workers’ excess lifetime cancer risk; and (iv) a valuation model that values changes in quality-adjusted life expectancy based on workers’ willingness-to-pay (WTP) for cancer risk reduction and aggregates individual valuations across the population at risk.

Parts (i) and (ii) provide answers to both questions: what can happen under reasonable worst-case assumptions, and what is expected to happen across a population of exposed individuals with idiosyncratic characteristics. Parts (iii) and (iv) bundle these heterogeneous effects as well as any uncertainties surrounding the modeling parameters of parts (i) and (ii) in one valuation figure and provide for a host of sensitivity analyses. Conceptually, the paper builds on previous work by Evans et al. (1994), Rice et al. (2010), Rheinberger and Hammitt (2012; 2014), and Chiu and colleagues (2015; 2018), among others.

To illustrate the framework, I use the case of exposure to hexavalent chromium (Cr(VI) for short), which is a potent carcinogen associated with increased lung cancer risk among workers in certain industries. More than 640,000 workers in North America are assumed to be frequently exposed to airborne Cr(VI), while the estimated number of Cr(VI)-exposed workers in the EU is approximately 780,000 (Proctor et al. 2016). In short, Cr(VI) is a ubiquitous industrial chemical for which both high-quality occupational exposure measurements (Vincent et al. 2015) and various dose-response relationships (Goldbohm et al. 2006; Wilbur et al. 2012) and unit risk factors (URFs; Haney et al. 2014; Procter et al. 2016) are readily available. I wish to stress that the framework is fully independent of the substance and could with some modifications be used for other health endpoints than cancer.

---

2 Since September 2017, the use of Cr(VI) compounds in the EU requires special authorization by the European Commission pursuant to Regulation (EC) No 1907/2006; see Georgiou et al. (2018) for details.
Methods

I start by presenting a generic model of cancer risk assessment (Figure 1), which I subsequently populate with four probabilistic submodels; one that determines the exposure levels to which workers across an industry are exposed, one that seeks to understand what being exposed to a constant or varying dose of the substance means in terms of excess lifetime cancer risk, one that converts risk into quality-adjusted life years lost, and one that stacks them into a life-cycle consumption model to obtain the welfare changes brought about by regulatory-induced exposure reductions. Below I am going to discuss each of the submodels in some detail.

—INSERT FIGURE 1—

Exposure model. For the purpose of setting up a probabilistic exposure model, I define a population of $i = 1, ..., N$ workers in the industry of interest. A random worker is exposed to the substance of concern—here Cr(VI)—at a time-weighted average (TWA) daily dose of $X$. While individual exposure is unknown, the empirical distribution of exposure prevailing in the industry of interest can be measured or probabilistically modeled (see McNally et al. 2014). The expected exposure, that is the average dose the statistical worker is exposed to, is given by $\mathbb{E}[X] = \int_{0}^{\infty} x f(x) dx$, where $f(x)$ is the PDF of exposure levels across the relevant industry, and $F(x)$ denotes the corresponding CDF.

Now, consider a regulatory action that establishes an occupational exposure limit (OEL) value or similar maximum daily dose $\delta$. If $\delta$ is enforced, it will result in a limited expected exposure $\mathbb{E}[X \wedge \delta]$ across the industry:

$$\mathbb{E}[X \wedge \delta] = \int_{0}^{\delta} x f(x) dx + \int_{\delta}^{\infty} \delta f(x) dx.$$  \[1\]

Integration by parts allows me to rewrite the first integral in Eq. [1] as:

$$\int_{0}^{\delta} x f(x) dx = \delta F(\delta) - \int_{0}^{\delta} F(x) dx.$$  \[2\]

The second integral in Eq. [1] may be replaced by:

$$\int_{\delta}^{\infty} \delta f(x) dx = \delta \int_{\delta}^{\infty} f(x) dx = \delta - \delta F(\delta).$$  \[3\]
By combining these two expressions, one obtains an alternative expression for the limited expected exposure $\mathbb{E}[X \wedge \delta] = \int_0^\delta 1 - F(x)\,dx$. Based on this definition, we are now ready to ask what is the overall exposure reduction brought about by enforcing $\delta$.

Two general cases (with many variants) may be considered.

(i) Firms may comply by introducing technical risk management measures if, and only if, the exposure is otherwise above the limit value $\delta$. In this case, the attributable exposure reduction is given by $\Delta X_\delta = \mathbb{E}[X] - \mathbb{E}[X \wedge \delta]$, where $\Delta X_{\delta=0} = \mathbb{E}[X]$ is the special case in which the substance of concern would no longer be used;

(ii) Firms may introduce risk management measures even if they are already complying with the limit value $\delta$. In that case, the analysis becomes somewhat more complex as this is equivalent to a shift of the cumulative exposure distribution from $F_0(x)$ before to $F_1(x)$ after the regulatory action, so that $F_1(x)$ is entirely to the left of $F_0(x)$.

This modeling approach is an obvious simplification and many complications may arise in reality. Some warrant further discussion. The model assumes, for instance, that the distribution of exposure across all $N$ workers is known even if individual worker $i$’s exposure is unknown. My presumption here is that exposure to the statistical worker is measurable or can be reasonably well modeled. However, exposure measurement campaigns are notoriously noisy even if they follow an established sampling protocol. To gauge the statistical uncertainty that comes with noisy measurement, it may be useful to replace $F(x)$ by $F(x + \epsilon)$ with $\epsilon \sim N(0, \sigma_X)$. Accounting for measurement error in this manner will not affect the expected exposure reduction, but will result in realistic confidence bounds around the modelled CDF that can be simulated via bootstrapping procedures (Cullen and Frey 1999).\(^3\)

So far, I have assumed that an equal number of workers are exposed at each workplace. This assumption is convenient from an analytical point as it implies that the exposure CDF can be directly elicited from measured data. It is, however, not very realistic because workplaces vary in terms of prevailing exposure levels in systematic ways; e.g., if larger firms invest more in safety than smaller firms, the latter can be expected to have systematically higher exposure levels than the former. Ideally such correlations are

\(^3\) The idea here is to repeatedly draw samples (with replacement) from the empirical distribution to obtain confidence bounds. For practical implementation, I have found the R package fitdistrplus to be extremely useful.
accounted for in \( F(x) \). A formal way of doing so is by assuming that \( i = 1, \ldots, N \) workers are employed by \( j = 1, \ldots, K \) firms. Now, assume firm size varies across the industry. Workplace-specific exposure \( x_j \) needs then to be weighted by the workplace-specific number of exposed workers per firm \( N_j \), resulting in the modified distribution \( F(x_j) \) in which \( x_j \) is observed \( N_j \) times. (More fine-grained exposure data permits further differentiation.)

A last complication I mention here is the fact that workers’ exposure may not be constant over any given period of time. This may be because they specific tasks change from shift to shift or from one job to another, or because the firm decides to implement additional risk management measures or is faced with an economic downturn that curbs production and thereby exposure. Such variability in cumulative exposure is difficult to model without reference to a worker’s lifetime dose. I return to this issue in the impact assessment model presented below, which converts cumulative exposure into excess lifetime cancer risk.

Dose-response model. Excess lifetime risk of cancer mortality in workers exposed to hazardous substances such as Cr(VI) is typically modeled using so-called unit risk factors (URFs). URFs are explain the difference in the probability of dying from cancer if exposed, \( m_1 \), and the baseline probability of dying from cancer if not exposed, \( m_0 \). Any such probability is necessarily age-specific and contingent upon surviving to age \( \tau \). Background mortality of an individual aged \( \tau \) may thus be split into cancer and non-cancer causes, i.e. 

\[
\begin{align*}
\text{m}_0(\tau) &= \text{m}_C(\tau) + \text{m}_{\text{NC}}(\tau).
\end{align*}
\]

Similarly, an exposed worker of age \( \tau \) bears the mortality risk \( \text{m}_1(\tau) = \text{m}_0(\tau) + \text{m}_x(\tau) \), where \( \text{m}_x(\tau) = \text{URF} \times \kappa \times x(\tau) \) denotes the contribution to cancer mortality from exposure \( x(\tau) \) in year \( \tau \), and \( \kappa \) converts the URF into an annual risk contribution factor.\(^4\) Moreover, it is convenient to introduce the cumulative mortality probability \( M(\tau) = \int_a^\tau m(t) \, dt \), where \( a \) may stand for either age at enrolment or age at regulatory intervention, and the corresponding survival function \( S(\tau) = \exp(-M(\tau)) \).\(^5\) This allows us to formally define the URF as:

\[
\text{URF} = \frac{\int_a^\tau \text{m}_1(t) \, dt - \int_a^\tau \text{m}_0(t) \, dt}{\int_a^\tau x(t) \, dt} = \frac{\int_a^\tau \text{m}_x(t) \, dt}{\int_a^\tau x(t) \, dt}.
\]

\(^4\) Obviously, this conversion hinges on low-dose linearity, see Crawford and Wilson (1996) for a discussion of this assumption.

\(^5\) As we are interested in cancer caused by occupational exposure, we may safely ignore baseline mortality risk before job enrolment.
In practice, URFs are almost always used in a deterministic fashion. However, they are prone to statistical uncertainty just like any other factor entering quantitative cancer risk assessment (Chiu and Slob 2015). One may account for such uncertainty by randomization of the relevant URF. A straightforward way to do so assumes that the cancer potency factor \( \hat{\theta} \) of the proportional hazard model based on which the URF was originally estimated can be simulated by taking random draws from the distribution \( f(\theta) \sim N(\hat{\theta}, \hat{SE}_\theta) \), where \( \hat{SE}_\theta \) denotes the estimated standard error of \( \hat{\theta} \). Keeping the coefficients of confounding factors such as age at enrolment and smoking status fixed, the simulated distribution of \( \theta \) can then be plugged into the standard formula that converts cancer potency into URF (see e.g. TECQ 2015). This procedure yields a distribution of possible URF values reflecting the variability in people’s susceptibility to a hazardous substance.

Impact assessment model. Building on the primitives of the dose-response model, the impact assessment model I use here is a competing risks model with two causes of deaths—cancer and other causes. Obviously, a worker can only die from occupational cancer if they have not prematurely died from another cause. I assume that cancer is associated with exposure to Cr(VI), whereas noncancer mortality is not. Following Vaeth and Pierce (1990), I denote the probability of surviving until age \( \tau \), conditional on being alive at age \( a \), by \( S_0(\tau; a) \). This conditional survival probability may be obtained as \( S_0(\tau; a) = S_0(\tau)/S_0(a) = \exp(-[M(\tau) - M(a)]) \). Accordingly, lifetime cancer (LR) risk for someone alive at age \( \tau \) can be obtained as the integral of the product of the survival function and the background cancer mortality rate over all ages \( t > \tau \):

\[
LR_0(\tau) = \int_a^\infty m_C(t)S(t; a)dt, \tag{5}
\]

where \( LR_0(\tau) \) is the proportion of all deaths among the unexposed that are due to cancer. Now, consider the impact of exposure on a cohort of workers exposed at age \( a \). By assumption noncancer mortality in that cohort is the same as in the reference population. Let \( m_1(\tau; a) = m_0(\tau; a) + m_x(\tau; a) \) stand for the mortality rate at age \( \tau \) for someone whose exposure onset was at age \( a \). As Vaeth and Pierce (1990) point out, \( m_1(\tau; a) \) might be a rather complicated function of the dose received, the age, the age at exposure, the time since exposure, and the background cancer mortality rate \( m_C(\tau) \). Yet when calculating

---

6 In the application I add further complexity by considering a cohort with known age distribution \( \phi(a) \).
excess lifetime risk (ELR) of cancer mortality, we do not need to be overly concerned with this function per se. This is because the problem to consider involves the comparison of cancer risk in two populations, an exposed vs an unexposed one, or—conceptionally very similar—the comparison of cancer risk in one population under two different exposure scenarios.

For this purpose, I recast the above definition of \( m_1(\tau:a) \) as relative risk model \( m_1(\tau:a) = [1 + r(a)]m_0(\tau:a) \), where \( r(a) \) captures the relative contribution of excess cancer risk to the overall mortality of an individual exposed at age \( a \). The survival function, \( S_1(\tau:a) \) for individuals first exposed at age \( a \) can now be obtained as \( S_1(\tau:a) = S_0(\tau:a)exp \{ -r(a)[M_0(\tau) - M_0(a)] \} \), where the second term on the left-hand side of the equality entails that a fraction of exposed workers will not live up to age \( \tau \) because of their exposure history. LR for someone exposed at age \( a \) is thus equal to the relative increment in the proportion of all eventual cancer deaths \( r_C(a) \) in a population of individuals exposed at age \( a \) becomes:

\[
LR_1(a, r_C(a)) = \int_a^\infty [1 + r_C(a)]m_C(t:a)S_1(t:a)dt.
\]  

Based on Eqs. [5] and [6], one may now formally define the excess lifetime cancer risk (ELR) born by a worker who has been exposed from age \( a \) onward as the difference between his lifetime cancer risk and the lifetime cancer risk of an unexposed individual:

\[
ELR(a, r(a)) = LR_1(a, r(a)) - LR_0(a).
\]  

A related measure that is even more relevant for assessing the impacts of cancer risk regulation is the loss of life expectancy (LLE) for individuals exposed at age \( a \). LLE is defined as the difference between the expected life expectancy of individuals exposed at age \( a \) and those unexposed (and alive at age \( a \)). Formally, LLE is given by:

\[
LLE(a, r(a)) = \int_a^\infty S_1(t:a)dt - \int_a^\infty S_0(t:a)dt.
\]  

Both ELR and LLE measure individual outcomes that can be calibrated by fitting Gompertz functions to cohort life table data (Rheinberger and Hammitt 2014) and combining these with either a deterministic URF or a probabilistic URF.\(^7\) Analogous measures for a cohort

---

\(^7\) This implies fitting the age-dependent baseline mortality rate in the reference population assuming that mortality can be modeled by
of workers with different ages at onset of exposure can be obtained as a weighted average of age-specific measures with weights proportional to the age distribution in question (Vaeth and Pierce 1990).

Valuation model. The valuation model quantifies in monetary terms the expected welfare implications of a regulatory intervention such as the establishment of a BOEL value. One simplistic approach would be to multiply LLE obtained from Eq. [8] by a deterministic value per statistical life-year. However, this would not result in a valid measure of welfare for at least two reasons. First, it would ignore that some life years would be lost sooner and others later in life. Hence, Eq. [8] needs to be amended with a proper discounting regime in order to account for temporal aspects. Second, and perhaps more important, LLE cannot reflect that not every cancer disease will lead to premature death. For example, according to the American Cancer Society (www.cancer.org), 23 out of 100 US patients diagnosed with non-small cell lung cancer will be alive 5 years after diagnosis. Yet any form of cancer disease will have both a physical and psychological toll. Rheinberger and Hammitt (2014) combined a life-cycle consumption model with age-specific disease survival rates to value the changes in relative health quality caused by a reduction in the risk of non-fatal health effects of exposure to substance of concern.

Based on the above insights, I extend Eq. [8] to capture both temporal aspects and the effect of a change in the risk of a non-fatal cancer disease on quality of life. (Rheinberger and Hammitt (2014) provide theoretical details.) Specifically, I build on the idea that a cancer disease developed at any age $\tau \geq a$ may be conceptualized as a random binary event that can be survived with the conditional survival probability $\pi(\tau)$. This probability may be obtained from the age-specific incidence and mortality rates associated with a specific type of cancer. The reduction in the risk of a non-fatal cancer disease, $\Delta p(\tau)$, then equals the product of the odds of cancer survival and the regulation-induced reduction in cancer mortality risk. 

---

8 The risks of non-fatal and fatal cancer are related through the following relationship: $\Pr(\text{Survive Cancer}) = \Pr(\text{Death|Cancer}) \times \Pr(\text{Cancer}) \times \Pr(\text{Survive Cancer|Cancer}) / \Pr(\text{Death|Cancer})$, or equivalently, $p(\tau) = m_C(\tau) \frac{\pi(\tau)}{1-\pi(\tau)}$. Hence, $\Delta p(\tau) = \Delta m_C(\tau) \frac{\pi(\tau)}{1-\pi(\tau)}$. 

---
Any survived cancer disease entails a temporary or permanent loss in health quality. I capture this loss by the relative measure \( \Delta q(\tau) \) such that, together, \( \Delta p(\tau) \) and \( \Delta q(\tau) \) define the regulatory change in the health decay rate \( \Delta h(\tau) = \Delta p(\tau) \Delta q(\tau) \). The regulatory change in life quality is thus given by the integral \( \Lambda(\tau, a) = \int_a^\infty \Delta h(t) \, dt \), and the relative effect of a change in non-fatal cancer risk on life quality may be calculated as:

\[
h(\tau) = \log \left[ \frac{1}{1 + \Lambda(\tau, a)} \right].\tag{9}\]

Equipped with these expressions, we may now revisit Eq. [8]. In particular, the welfare value of reductions in cancer mortality corresponds to:

\[
V_m(e) = \int_e^\infty [v(t) + y(t) - c(t)] \exp(-\rho(t-a)) [S_1(t;a) - S_0(t;a)] \, dt,\tag{10a}
\]

and the corresponding welfare value of cancer morbidity may be expressed as:

\[
V_h(e) = \int_e^\infty [v(t) \, h(t)] \exp(-\rho(t-e)) [S_1(t:a) - S_0(t:a)] \, dt.\tag{10b}
\]

In Eqs. [10a] and [10b], there are four parameters that have not been introduced so far. I use \( \rho \) to denote the social discount rate, \( v(t) \) is the age-specific value per statistical life-year, \( y(t) - c(t) \) is net income earned in year \( t \). All of these parameters can be calibrated based on existing studies. (Rheinberger and Hammitt 2014) provide details on the calibration.)

The presented valuation model is extremely flexible. It permits replacing the exponential discounting regime by a (quasi-) hyperbolic one; plugging in a constant value per statistical life-year, studying either temporary or permanent health effects of non-fatal cancer disease, and accounting for any length of exposure a worker might be subject to. With some modifications one may also accommodate for latency periods between exposure and possible disease onset (Vaeth and Pierce 1990), for non-constant exposure levels over the period of exposure, for exposure cessation at retirement age (Rheinberger and Hammitt 2012), and for suspected sampling bias.

---

9 In Rheinberger and Hammitt (2014), we used EQ-5D data from the U.S. Medical Expenditure Panel Survey to econometrically pin down age-specific effects of disease on life quality assuming these are permanent. Other methods are conceivable including ones that could better account for temporary “blips” in health deterioration.
Results

In this section, I am going to illustrate the merits of the probabilistic framework for assessing cancer risk using the case of Cr(VI) exposure of chrome platers in France and the introduction of a national BOEL value of 1 µg Cr(VI)/m³ over an 8h-shift. To populate the exposure model, I use data from a measurement campaign conducted by the French National Institute for Health and Safety (INRS) between 2010-2013. (See Vincent et al. (2015) for a detailed description of both the measurement protocol and the data.) Figure 2 displays the cumulative exposure distribution for hard chrome plating activities (97 measurements made in 12 hard chrome shops) as well as a bootstrapped confidence interval. The mean exposure level in this sample was 1.60 µg Cr(VI)/m³ and, as the red line in Figure 2 illustrates, around 33% of the measurements were actually above the BOEL.

The question I explore here in order to illustrate the framework proposed is what theoretical risk reduction could be achieved (both in terms of ELR of an average-aged worker and LLE across the French hard chrome sector) if the BOEL was fully enforced such that measurements above 1 µg Cr(VI)/m³ would no longer occur (or would be limited to no more than 15 min in compliance with the STEL of 5 µg Cr(VI)/m³). For simplicity, I will make several assumptions. First, I shall assume that without full enforcement non-compliant firms would continue to operate at the exposure levels measured in the INRS measurement campaign. Second, I maintain the assumption that compliant firms have already implemented risk management measures to stay below the BOEL and will therefore not take further action. For the purpose of simulating I consider a cohort of 1,000 synthetic workers (which seems a plausible if not confirmed assumption on the number of actual workers in French hard chrome shops). Drawing from a uniform distribution, I generate an age at intervention and then trace workers through the remaining exposure years until retirement at age 65.

In order to assess the impacts the enforcement of the BOEL might have on this synthetic cohort, I apply the URF for airborne Cr(VI) exposure that Haney et al. (2014) recently

---

10 Since 1 July 2014, France has a BOEL of 1 µg Cr(VI)/m³ for an 8h-shift and a 15-min short-term exposure limit (STEL) of 5 µg Cr(VI)/m³.
developed based on observed lung cancer rates in chromate production workers in Painesville, Ohio and Baltimore, Maryland. Combining estimates of ELR of fatal lung cancer in both occupational cohorts Haney et al. obtain a URF of 2.28E-3 per µg Cr(VI)/m³ with a weighted standard error of 3.03E-4. Based on the weighted standard error I then simulate the URF for 1,000 synthetic workers (Figure 3).

—IINSERT FIGURE 3—I

I use the 2013 life table from Eurostat¹¹ for the male population in the EU28 to fit a Gompertz mortality function \( \hat{m}_0(\tau) \) with parameter estimates \( \hat{\alpha}_0 = 3.45 \times 10^{-5} \) and \( \hat{\beta}_0 = 9.46 \times 10^{-2} \). Similarly, the lung cancer mortality function \( \hat{m}_C(\tau) \) in the male population in the EU28 can be fitted based on GLOBOCAN data, yielding parameter estimates \( \hat{\alpha}_C = 2.71 \times 10^{-6} \) and \( \hat{\beta}_C = 1.01 \times 10^{-1} \).

It remains to obtain an empirical estimate of the relative risk factor \( r(a) \). I exploit that

\[
 r(a) = \frac{m_a(\tau,a)}{m_0(\tau,a)}.
\]

Combining \( \hat{m}_0(\tau) \), \( \hat{m}_C(\tau) \), and \( \hat{m}_0(\tau;a) \) according to the relationships described by Eqs. [4-8] allows pinning down the ELR for a worker whose exposure onset is at age \( a \) (taken to be between 35 and 65) and lasts until retirement age of 65. Workers who had been exposed to Cr(VI) during their worklife remain to bear a higher lung cancer risk for the reminder of their lives. I account for this fact by adding an exposure cessation period starting with retirement and lasting up to a maximum age of 85. Also I account for possible non-fatal lung cancer disease using a constant health deterioration of roughly one third as found by Nafees et al. (2008).

The cancer risk metrics established above are summarized in Table 1 for those 33% of the synthetic workers who would benefit from the full enforcement of the BOEL. As can be seen, the risk reductions achievable over a lifetime are substantial (up to 6.7E-2 for fatal lung cancer and 1.5E-2 for non-fatal lung cancer), highlighting the carcinogenicity of Cr(VI). Within the synthetic cohort this corresponds to 12.2 full life years and 2.7 QALYs that could be gained from the introduction of the BOEL, respectively.

—IINSERT TABLE 1—I

The valuation of cancer mortality risk is based on a recent study by Alberini and Ščasný (2018) who surveyed respondents in four EU countries and obtained a preferred value per statistical life (VSL) estimate of €2.2m. Knowing the age of an individual, the VSL can be converted into a value per statistical life-year (VSLY) by applying standard annuitization techniques. As per Eq. [10b], the valuation of cancer morbidity risk is also based on the resulting VSLY; hence, the welfare value of reductions in non-fatal lung cancer on health quality will also be based on the VSLY, but will be downscaled by the age-specific odds of having a non-fatal rather than a fatal lung cancer. The health decay rate may be obtained following the approach outlined in Rheinberger and Hammitt (2014) or, as I assume here, may be based on a constant health deterioration caused by non-fatal lung cancer.

Table 2 reports the estimated welfare value of enforcing the BOEL given the prevailing exposure levels at the time of the INRS measurement campaign. If one assumes a social discount rate of $\rho = 4\%$ as suggested by the European Commission’s Better Regulation guidelines, the simulation results suggest that the central value of reducing excess mortality and morbidity within a cohort of 1,000 synthetic workers amounts to €3.0m (95% CI: €22.2k-€19.5m). Another valuable welfare measure can be obtained by trailing each worker only for one year of exposure and then sent them into “early retirement”. That way, one may capture the average welfare value of introducing the BOEL per exposure year—an amount which may be readily compared with the annual cost of complying with and enforcing the BOEL, respectively. The central estimate of the respective value amounts to €0.4m (95% CI: €6k-€3.7m).

—INSERT TABLE 2—

Conclusion

Assessing the impacts of regulatory actions to curb exposure to carcinogenic substances is often done in a deterministic way. While this may convey the illusion of certainty to the outside world, it cannot do away with the fact that many factors in such impact assessments are inherently uncertain. As Gray and Cohen (2012) conclude, bringing these uncertainties into the open is the only means to enable an honest debate about the appropriateness of a course of action in a specific situation. It also seems a useful way to reconcile risk assessors and economists.
As stated in the introduction, the questions the two disciplines try to answer are often fundamentally different, or as I like to joke, economists are from Mars, risk assessors are from Venus. Having distributions of specific outcomes should speak to both disciplines, as it allows addressing what is most likely to happen and what might happen under precautionary assumptions. Hence moving toward probabilistic approaches seems a promising avenue for advancing cancer risk assessment.

The framework I have outlined and illustrated in this paper is very flexible and can be adopted to explore a host of assumptions and can be adapted to special settings, e.g. where exposures across workplaces are positively correlated or where workers are exposed to carcinogenic substances for a short period only (“exposure blips”). Other extensions can be readily introduced as long as they link to excess lifetime risk. This flexibility comes at the cost of additional complexity and, indeed, many practitioners of regulatory impact assessments are reluctant to adopt models and modeling tools they conceive as black boxes. Yet I wish to stress that the framework does not rely on any other information than what has typically been used in deterministic assessments. Instead of codifying specific assumptions through constant values, the framework embraces uncertainties as valuable outcome of a regulatory impact assessment.
References


Figures

Fig. 1. Schematic representation of the unified probabilistic framework for cancer risk assessment. Panel A represents the cumulative distribution function of 8h-TWA exposure to Cr(VI); Panel B shows the possible exposure paths for two statistical workers (solid and dashed line) whose exposure starts at time $e$ and ceases at time $T$; Panel C indicates that reduction in worklife exposure to Cr(VI) reduces mortality risk (from the solid to the dotted line); Panel D converts changes in mortality risk into changes in life expectancy (from the solid to the dotted line).
Fig. 2. Cumulative exposure distribution across French chrome platers. Black dots show actual measurements; the blue dashed line represents the BOEL of 1 \( \mu g \) Cr(VI)/m\(^3\) (8h-TWA); the red solid line represents a fitted lognormal distribution (GM = 0.575, GSD = 4.197); the red dashed lines represent 95%-CI (GM\(_{UB} = 0.795\), GM\(_{LB} = 4.411\), GSD\(_{UB} = 5.047\), GSD\(_{LB} = 3.435\)).
Fig. 3. Simulated distribution of the Cr(VI)-URF for 1,000 synthetic individuals. The dashed line indicates the mean URF of 0.00228 per µg Cr(VI)/m³ (8h-TWA) found by Haney et al. (2014); the grey area delimits the 95 confidence interval around the mean URF.
Tab. 1. Summary of cancer risk reduction caused by the BOEL of 1 µg Cr(VI)/m³ (8h-TWA).

<table>
<thead>
<tr>
<th>Cancer risk metric</th>
<th>Mean</th>
<th>Median</th>
<th>2.5-% CI</th>
<th>97.5-% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduction in excess lifetime risk of fatal lung cancer</td>
<td>6.70E-2</td>
<td>2.16E-2</td>
<td>2.52E-4</td>
<td>4.33E-1</td>
</tr>
<tr>
<td>Reduction in excess lifetime risk of non-fatal lung cancer</td>
<td>1.46E-2</td>
<td>4.71E-3</td>
<td>5.49E-5</td>
<td>9.42E-2</td>
</tr>
<tr>
<td>Gain in life expectancy</td>
<td>12.22</td>
<td>3.31</td>
<td>0.008</td>
<td>86.44</td>
</tr>
<tr>
<td>Avoidance of quality-impaired life years</td>
<td>2.66</td>
<td>0.72</td>
<td>0.002</td>
<td>18.81</td>
</tr>
</tbody>
</table>
Tab. 2. Summary of welfare gain associated with a BOEL of 1 µg Cr(VI) (8h-TWA).

<table>
<thead>
<tr>
<th>Welfare value</th>
<th>Mean</th>
<th>Median</th>
<th>2.5-% CI</th>
<th>97.5-% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Permanent exposure reduction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reduction in fatal lung cancer risk across the cohort</td>
<td>€2.71m</td>
<td>€1.02m</td>
<td>€0.02m</td>
<td>€17.66m</td>
</tr>
<tr>
<td>Reduction in non-fatal lung cancer risk across the cohort</td>
<td>€0.28m</td>
<td>€0.08m</td>
<td>€2.0k</td>
<td>€1.88m</td>
</tr>
<tr>
<td><strong>1-year exposure reduction at age 50</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reduction in fatal lung cancer risk across the cohort</td>
<td>€0.25m</td>
<td>€0.11m</td>
<td>€3.6k</td>
<td>€2.79m</td>
</tr>
<tr>
<td>Reduction in non-fatal lung cancer risk across the cohort</td>
<td>€0.14m</td>
<td>€63.8k</td>
<td>€2.4k</td>
<td>€0.87m</td>
</tr>
</tbody>
</table>
Appendix: R code for replicating the simulation

###Simulations for "A Unified Probabilistic Framework for Cancer Risk Assessment"
##to be presented at HCRA workshop on risk assessment, economic evaluation, and decisions
##Sept-12-2019
library(fitdistrplus)
library(mvtnorm)

FDATA <- read.csv("~/Dropbox/Harvard_WS/FDATA.csv", comment.char="#")#read in exposure measurement data from Vincent et al. 2015
attach(FDATA)
x=exposure[activity=="Hard chrome plating"]

##Exposure modeling
par(mfrow=c(1,1))
plot(dist(x,demp=T,histo=T,breaks=10,col=2)#check
descdist(x,boot = 1000)
?descdist
z.l=fitdist(x,"lnorm")#alternative fit: z=fitdist(x, "gamma");check w/ gofstat(z)
z.w=fitdist(x,"weibull")
b.l=bootdist(z.l,niter=1e4,"nonparam");summary(b.l)
b.w=bootdist(z.w,niter=1e4,"nonparam");summary(b.w)
CIcdfplot(b.w,CI.level=.95, CI.output = "quantile",main=""
lines(ecdf(sim_expo),col=4)
##use resulting fit to simulate exposure
n=1e4#no of synthetic workers
meanlog_l=-0.8240365;sdlog_l=1.2353691
meanlog_mu=-0.5392783;sdlog_mu=1.4236578
meanlog_u=-0.2542975;sdlog_u=1.6114272
meanlog_sd=round((meanlog_mu-meanlog_l)/1.96,3)#double check by quantile(rnorm(10000,sdlog_c,sdlog_sd),p=c(.025,.975))
sdlog_sd=round((sdlog_mu-sdlog_l)/1.96,3)#dito
sim_meanlog=rnorm(n,meanlog_mu,meanlog_sd)
sim_sdlog=rnorm(n,sdlog_mu,sdlog_sd)
corr=1#assumes that lognormal distribution parameters are fully correlated
sim=mvrnorm(n,mu=c(meanlog_mu,sdlog_mu),Sigma=matrix(c(meanlog_sd^2,corr*meanlog_sd*sdlog_sd,corr*meanlog_sd*sdlog_sd,sdlog_sd^2),2,2))
set.seed(12)
sim_expo=pmin(rlnorm(n,sim[,1],sim[,2]),25)# EU wide BOEL
quantile(sim_expo,p=c(.025,.975))
red_expo=pmax(sim_expo-1,0)#reduction in exposure after enforcement of BOEL
lim_expo=pmin(sim_expo,1)#exposure after enforcement of BOEL
plot(ecdf(sim_expo),col=1)
lines(ecdf(lim_expo),col=2)

###########
##URF modelling
sim_age=round(runif(n,20,65),1)#uniform distribution assumed, may be modified to any cohort
mu_urf=2.28e-3#Gompertz shape parameter fit based on EU28 baseline risk (2013, male)
beta_urf=9.459e-2#Gompertz scale parameter fit based on EU28 baseline risk (2013, male)
alpha_c=2.71e-6#Gompertz shape parameter fit based on globocan 2012 date for male lung cancer;
beta_c=1.01e-1#Gompertz scale parameter fit based on globocan 2012 date for male lung cancer;
a=rep(50,n)#sim_age#age at intervention
RET=51 #65 #age at retirement
l=10#latency period based on HSE report, currently not used
MIN=35# minimum age of cancer
MAX=85# maximum age of cancer
t=0.04# social discount rate based on EC’s Better regulation guidelines
urf=sim_urf
arf=urf/45
QALYLOSS=1-0.653# see Nafees++2008

S_TABLE=matrix(data=NA,nrow=n,ncol=5)
for (i in 1:n){
    s_0<-function(t){
        pmin((exp(-alpha_0/beta_0)*(exp(beta_0*t)-exp(beta_0*a[i])))-
            sim_arf[i]*(t-max(MIN,a[i])*sim_expo[i]),1) #Survivor
    }
    r_0<-function(t){
        pmin((exp(-alpha_0/beta_0)*(exp(beta_0*t)-exp(beta_0*a[i])))-
            sim_arf[i]*(RET-max(MIN,a[i])*sim_expo[i]),1) #Survivor
    }
    s_1<-function(t){
        pmin((exp(-alpha_0/beta_0)*(exp(beta_0*t)-exp(beta_0*a[i])))-
            sim_arf[i]*(t-max(MIN,a[i])*lim_expo[i]),1) #Survivor
    }
    r_1<-function(t){
        pmin((exp(-alpha_0/beta_0)*(exp(beta_0*t)-exp(beta_0*a[i])))-
            sim_arf[i]*(RET-max(MIN,a[i])*lim_expo[i]),1) #Survivor
    }
    S_0<-integrate(s_0, lower = max(MIN,a[i]), upper=RET, subdivisions=1e4)$value
    R_0<-integrate(r_0, lower = RET, upper=MAX, subdivisions=1e4)$value
    S_1<-integrate(s_1, lower = max(MIN,a[i]), upper=RET, subdivisions=1e4)$value
    R_1<-integrate(r_1, lower = RET, upper=MAX, subdivisions=1e4)$value
    S_TABLE[i,1]<-S_0
    S_TABLE[i,2]<-R_0
    S_TABLE[i,3]<-S_1
    S_TABLE[i,4]<-R_1
    S_TABLE[i,5]<-S_1+R_1-S_0-R_1
}
summary(S_TABLE[,5])
sum(S_TABLE[,5])/n
(1-QALYLOSS)*(.25/.75)*sum(S_TABLE[,5])/n
Y=S_TABLE[,5]
Z=Y*Y-1E-6
summary(.1*length(Z)*Z);quantile(.1*length(Z)*Z,p=c(0.025,.975))
summary(.1*length(Z)*(1-QALYLOSS)*(.25/.75)*Z);quantile(.1*length(Z)*(1-QALYLOSS)*(.25/.75)*Z,p=c(0.025,.975))

ELR_TABLE=matrix(data=NA,nrow=n,ncol=5)
for (i in 1:n){
    s_0<-function(t){
        pmin((alpha_c*exp(beta_c*t)+sim_arf[i]*(t-max(MIN,a[i])*sim_expo[i]))-
            (exp(-alpha_0/beta_0)*(exp(beta_0*t)-exp(beta_0*a[i])))-
            sim_arf[i]*(t-max(MIN,a[i])*sim_expo[i]),1) #Survivor
    }
    r_0<-function(t){
        pmin((alpha_c*exp(beta_c*t)+sim_arf[i]*(RET-max(MIN,a[i])*sim_expo[i]))-
            (exp(-alpha_0/beta_0)*(exp(beta_0*t)-exp(beta_0*a[i])))-
            sim_arf[i]*(RET-max(MIN,a[i])*sim_expo[i]),1) #Survivor
    }
    s_1<-function(t){
        pmin((alpha_c*exp(beta_c*t)+sim_arf[i]*(t-max(MIN,a[i])*lim_expo[i]))-
            (exp(-alpha_0/beta_0)*(exp(beta_0*t)-exp(beta_0*a[i])))-
            sim_arf[i]*(t-max(MIN,a[i])*lim_expo[i]),1) #Survivor
    }
    r_1<-function(t){
        pmin((alpha_c*exp(beta_c*t)+sim_arf[i]*(RET-max(MIN,a[i])*lim_expo[i]))-
            (exp(-alpha_0/beta_0)*(exp(beta_0*t)-exp(beta_0*a[i])))-
            sim_arf[i]*(RET-max(MIN,a[i])*lim_expo[i]),1) #Survivor
    }
LE_0 <- integrate(s_0, lower = max(MIN,a[i]), upper=RET, subdivisions=1e4)$value
LEC_0 <- integrate(r_0, lower = RET, upper=MAX, subdivisions=1e4)$value
LE_1 <- integrate(s_1, lower = max(MIN,a[i]), upper=RET, subdivisions=1e4)$value
LEC_1 <- integrate(r_1, lower = RET, upper=MAX, subdivisions=1e4)$value
ELR_TABLE[i,1] <- LE_0
ELR_TABLE[i,2] <- LEC_0
ELR_TABLE[i,3] <- LE_1
ELR_TABLE[i,4] <- LEC_1
ELR_TABLE[i,5] <- (LEC_0 - LEC_1) + (LE_0 - LE_1)
summary(ELR_TABLE[,5])*(.25/.75)
quantile(ELR_TABLE[,5], p = c(.65,.975))
Y=ELR_TABLE[,5]
Z=Y[Y>1E-6]
summary(Z);quantile(Z, p = c(0.025,.975))
summary((1-QALYLOSS)*(.25/.75)*Z);quantile((1-QALYLOSS)*(.25/.75)*Z, p = c(0.025,.975))

# Mortality valuation
M_TABLE=matrix(data=NA, nrow=n, ncol=3)
for (i in 1:n)
{ m_s_0 <- function(t){
  exp(-r*(t-a[i]))*discount
  (2.2/7.8) * (-1.53E-3*t^5 + 5.21E-1*t^4 - 6.43E+1*t^3 + 3.4E+3*t^2 - 7.13E+4*t + 0.748e+6)*v(t)
  (exp(-(-alpha_0/beta_0)*(exp(beta_0*t)-exp(beta_0*max(MIN,a[i]))))-
  sim_arf[i]*(t-max(MIN,a[i]))*sim_expo[i])#Survivor among Cr6 workers before change
}
m_r_0 <- function(t){
  exp(-r*(t-a[i]))*discount
  (2.2/7.8) * (-1.53E-3*t^5 + 5.21E-1*t^4 - 6.43E+1*t^3 + 3.4E+3*t^2 - 7.13E+4*t + 0.748e+6)*v(t)
  (exp(-(-alpha_0/beta_0)*(exp(beta_0*t)-exp(beta_0*max(MIN,a[i]))))-
  sim_arf[i]*(RET-max(MIN,a[i]))*lim_expo[i])#Survivor among Cr6 workers before change
}
m_s_1 <- function(t){
  exp(-r*(t-a[i]))*discount
  (2.2/7.8) * (-1.53E-3*t^5 + 5.21E-1*t^4 - 6.43E+1*t^3 + 3.4E+3*t^2 - 7.13E+4*t + 0.748e+6)*v(t)
  (exp(-(-alpha_0/beta_0)*(exp(beta_0*t)-exp(beta_0*max(MIN,a[i]))))-
  sim_arf[i]*(t-max(MIN,a[i]))*lim_expo[i])#Survivor among Cr6 workers before change
}
m_r_1 <- function(t){
  exp(-r*(t-a[i]))*discount
  (2.2/7.8) * (-1.53E-3*t^5 + 5.21E-1*t^4 - 6.43E+1*t^3 + 3.4E+3*t^2 - 7.13E+4*t + 0.748e+6)*v(t)
  (exp(-(-alpha_0/beta_0)*(exp(beta_0*t)-exp(beta_0*max(MIN,a[i]))))-
  sim_arf[i]*(RET-max(MIN,a[i]))*lim_expo[i])#Survivor among Cr6 workers before change
}
M_0 <- integrate(m_s_0, lower = max(MIN,a[i]), upper=RET, subdivisions=1e4)$value+integrate(m_r_0, lower = RET, upper=MAX, subdivisions=1e4)$value
M_1 <- integrate(m_s_1, lower = max(MIN,a[i]), upper=RET, subdivisions=1e4)$value+integrate(m_r_1, lower = RET, upper=MAX, subdivisions=1e4)$value
M_TABLE[i,1] <- M_0
M_TABLE[i,2] <- M_1
M_TABLE[i,3] <- M_1-M_0
}
summary(M_TABLE[,3])
VSLY=sum(M_TABLE[,3])/sum(S_TABLE[,5]);VSLY
Y=M_TABLE[,3]
Z=Y[Y>1]
1000*summary(Z);1000*quantile(Z, p = c(0.025,.975))

# Morbidity valuation
H_TABLE=matrix(data=NA, nrow=n, ncol=3)
for (i in 1:n){
\[ h_s_0 = \text{function}(t) \{
  \text{pmax}(\exp(-r*(t-a[i]))*\text{discount} \times (2.2/7.8)(-1.53E-3*t^5 + 5.21E-1*t^4 - 6.43E+1*t^3 + 3.4E+3*t^2 - 7.13E+4*t + 0.748e+6) - \#v(t) - 1.47e-2*t^4 + 4.05E-1*t^3 - 4.15e+2*t^2 + 1.9e+4*t - 1.81E5)*\#y1(t) + (3.49e-2*t^4 - 7.24*t^3 + 4.65E+2*t^2 - 9.74e+3*t + 1.48e+5)))*\#c1(t) \}
\]

\[ h_r_0 = \text{function}(t) \{
  \text{pmax}(\exp(-r*(t-a[i]))*\text{discount} \times (2.2/7.8)(-1.53E-3*t^5 + 5.21E-1*t^4 - 6.43E+1*t^3 + 3.4E+3*t^2 - 7.13E+4*t + 0.748e+6) - \#v(t) - 1.47e-2*t^4 + 4.05E-1*t^3 - 4.15e+2*t^2 + 1.9e+4*t - 1.81E5)*\#y1(t) + (3.49e-2*t^4 - 7.24*t^3 + 4.65E+2*t^2 - 9.74e+3*t + 1.48e+5)))*\#c1(t) \}
\]

\[ h_s_1 = \text{function}(t) \{
  \text{pmax}(\exp(-r*(t-a[i]))*\text{discount} \times (2.2/7.8)(-1.53E-3*t^5 + 5.21E-1*t^4 - 6.43E+1*t^3 + 3.4E+3*t^2 - 7.13E+4*t + 0.748e+6) - \#v(t) - 1.47e-2*t^4 + 4.05E-1*t^3 - 4.15e+2*t^2 + 1.9e+4*t - 1.81E5)*\#y1(t) + (3.49e-2*t^4 - 7.24*t^3 + 4.65E+2*t^2 - 9.74e+3*t + 1.48e+5)))*\#c1(t) \}
\]

\[ h_r_1 = \text{function}(t) \{
  \text{pmax}(\exp(-r*(t-a[i]))*\text{discount} \times (2.2/7.8)(-1.53E-3*t^5 + 5.21E-1*t^4 - 6.43E+1*t^3 + 3.4E+3*t^2 - 7.13E+4*t + 0.748e+6) - \#v(t) - 1.47e-2*t^4 + 4.05E-1*t^3 - 4.15e+2*t^2 + 1.9e+4*t - 1.81E5)*\#y1(t) + (3.49e-2*t^4 - 7.24*t^3 + 4.65E+2*t^2 - 9.74e+3*t + 1.48e+5)))*\#c1(t) \}
\]

\[ H_0 = \text{integrate}(h_s_0, \text{lower} = \text{max}(\text{MIN},a[i]), \text{upper} = 65, \text{subdivisions} = 1e4)*\text{value} + \text{integrate}(h_r_0, \text{lower} = 65, \text{upper} = \text{MAX}, \text{subdivisions} = 1e4)*\text{value} \]

\[ H_1 = \text{integrate}(h_s_1, \text{lower} = \text{max}(\text{MIN},a[i]), \text{upper} = 65, \text{subdivisions} = 1e4)*\text{value} + \text{integrate}(h_r_1, \text{lower} = 65, \text{upper} = \text{MAX}, \text{subdivisions} = 1e4)*\text{value} \]

\[ \text{H_TABLE}[i,1] = H_0 \]
\[ \text{H_TABLE}[i,2] = H_1 \]
\[ \text{H_TABLE}[i,3] = H_1 - H_0 \]

\[ \text{Y} = \text{H_TABLE}[3] \]
\[ Z = Y[1] \]
\[ 1000*\text{summary}(Z) ; 1000*\text{quantile}(Z, p = c(0.025, .975)) \]