

Discussie is de kern van wetenschap

Een meta-review van drie Gezondheidsraad-adviezen als reactie op de publicatie van het advies blootstelling aan stoffen bij ijzer- en staalgieten



Ira Helsloot
Jaap Hanekamp

Auteurs

Ira Helsloot (Crisislab en Radboud Universiteit Nijmegen)

Jaap C. Hanekamp (University College Roosevelt en University of Massachusetts Amherst, USA)

Augustus 2020

Crisislab is de onderzoeksgroep die het onderzoek van de leeropdracht Besturen van Veiligheid van de Radboud Universiteit Nijmegen ondersteunt. De doelstelling van Crisislab is de ontwikkeling en verspreiding van kennis op het domein van crisisbeheersing en veiligheidszorg. Voor Crisislab is een kernactiviteit het verrichten van empirisch gefundeerd onderzoek op het veiligheidsdomein, omdat momenteel feiten vaak ontbreken bij beleidsvorming en discussies op het terrein van het besturen van veiligheid. Op basis van dit onderzoek adviseren we overheden en bedrijven om tot redelijk en proportioneel veiligheidsbeleid te komen. De oefeningen en trainingen die wij verzorgen zijn gericht op het realistisch leren omgaan met crisismechanismen en met de veerkrachtige samenleving.

Crisislab
Dashorsterweg 1
3927 CN Renswoude
www.crisislab.nl

Abstract

Deze methodologische review van drie recente adviezen van de Gezondheidsraad laat zien dat de Gezondheidsraad zich in deze adviezen niet houdt aan haar eigen leidraad voor de classificatie van carcinogene stoffen (2010), zich niet houdt aan algemene wetenschappelijke normen voor goed onderzoek en resultaten niet op een transparante wijze en in de juiste context presenteert waarmee een afgewogen bestuurlijk besluit ernstig wordt bemoeilijkt.

Samenvattend:

- In tegenspraak met de eigen leidraad is onduidelijk op wat voor gronden onderzoeken worden geselecteerd als basis voor de Gezondheidsraadsadviezen en wordt te weinig rekening gehouden met confounding factors waarvan roken de belangrijkste is.
- In tegenspraak met de gangbare wetenschappelijke praktijk is bijvoorbeeld de afwezigheid van een gevoeligheidsanalyse en het gebruik van lifetables bij het bepalen van het risico van blootstelling aan een stof en het gebruik van correlatie als bewijs voor causaliteit.
- Tenslotte is de Gezondheidsraad, conform haar eigen leidraad, bewust conservatief in haar schattingen van het effect van blootstelling aan stoffen hetgeen de waarde van de adviezen voor de praktijk sterk vermindert.

1. Inleiding

De Gezondheidsraad (GR) adviseert over de classificatie en normering van de blootstelling aan gevaarlijke stoffen. Meer precies geeft de Gezondheidsraad een advies aan het ministerie van Sociale Zaken en Werkgelegenheid (SZW) en aan de (subcommissie Grenswaarden stoffen op de werkplek van de) Sociaal Economische Raad (SER). De SER adviseert vervolgens ook de minister van SZW.

De praktijk is dat het advies van de GR niet meer ter discussie staat en de inhoudelijke basis is voor de belangrijke besluitvorming over de omgang met de blootstelling aan gevaarlijke stoffen. Dit legt een zware verantwoordelijkheid bij de GR voor de wetenschappelijke precisie en bruikbaarheid voor bestuurlijke besluitvorming van haar adviezen.

Over de kwaliteit van de adviezen van de GR bestaat echter twijfel bij verschillende deskundigen die inhoudelijk hebben gereageerd op diverse conceptadviezen van de GR *zonder* dat dit tot een voor hen inhoudelijke, begrijpelijk en toetsbaar antwoord heeft geleid.¹ Ook de auteurs van deze meta-review is dat ‘overkomen’: het antwoord van de GR op een review door Crisislab van het conceptadvies over de blootstelling aan stoffen bij ijzer- en staalgieterijen is in de context van het wetenschappelijk discours onbegrijpelijk.² Wat ons echter vooral verontrust maar ook agiteert, is de stellingname van de GR:

Er zal geen gedachtewisseling over standpunten plaatsvinden. De Gezondheidsraad wil elke schijn voorkomen dat belanghebbenden – op niet wetenschappelijke gronden – invloed kunnen uitoefenen op de totstandkoming van het advies.³

Met deze positie sluit de GR zich af voor wat fundamenteel wetenschappelijk is: een open discours over feitenkwesties, analyse(methoden), theorievorming en conclusies. Tevens diskwalificeert de GR daarmee haar eigen werk omdat inhoudelijke reflectie van buitenaf niet wordt gehonoreerd.

Om wel aan dit wetenschappelijke discours bij te dragen, hebben de auteurs daarom een methodologische review uitgevoerd van drie recente rapporten (adviezen) van de Gezondheidsraad over de blootstelling aan stoffen:

- *Blootstelling aan chroom CrVI*, nr. 2016/13,
- *Dieselmotoremissie*, nr. 2019/02 en

¹ Vijf gesprekken met andere wetenschappelijke reviewers van Gezondheidsraadadviezen. Het is mogelijk veelzeggend dat niemand met naam genoemd wil worden.

² Deze review van het conceptadvies over de blootstelling aan stoffen bij ijzer- en staalgieterijen is uitgevoerd op verzoek van en met subsidiering door de Koninklijke Metaalunie.

³ Reactie Gezondheidsraad op o.a. commentaar Crisislab, d.d. 20 februari 2020.

- *Emissies van ijzer- en staalgieten* (Occupational exposure during iron and steel founding - Evaluation of the carcinogenicity and genotoxicity), nr. 2020/02.

Om inzicht te krijgen in de methodische aanpak van de GR toetsen wij de aanpak van de GR aan de (meest recente) wetenschappelijke inzichten over (de effecten van) blootstelling aan stoffen. De vraag over de kwaliteit van bewijsvoering voor de causale relatie tussen blootstelling en ziekte bij blootgestelden staat daarbij centraal. Het gebruik van adequate statistische methoden in bijvoorbeeld epidemiologische studies is daarvan een belangrijk onderdeel.

Leeswijzer

In deze review gaan we op de volgende terugkerende problematische aspecten in die wij geïdentificeerd hebben in de drie adviezen van de GR:

- De GR houdt zich niet aan haar eigen *Leidraad classificatie carcinogene stoffen* uit 2010, zoals het feit dat onvoldoende aandacht wordt geschonken aan confounding factors zoals roken.
- De GR houdt zich niet aan algemeen geldende wetenschappelijke normen voor goed onderzoek, zoals het feit dat de GR correlatie als (deel)bewijs voor causaliteit beschouwt.
- De GR presenteert resultaten niet in een transparante vorm en in de juiste context zodat een afgewogen bestuurlijk besluit kan worden genomen.

Wij zullen in deze review vooral voorbeelden uit het *Blootstelling aan Chroom CrVI*-advies gebruiken ter illustratie van onze punten. Dit heeft twee redenen: het *Dieselmotoremissie*-advies is (terecht) terzijde gelegd door de SER als niet proportioneel en uitvoerbaar (we komen daar nog op terug) en voor het advies over ijzer- en staalgieten geldt dat we een uitgebreidere review als bijlage hebben opgenomen in vergelijking met de korte reviews van de twee andere adviezen.

2. De GR volgt de eigen leidraad niet consequent

De GR heeft in 2010 een *Leidraad classificatie carcinogene stoffen* uitgebracht waarin de methodiek van beoordeling en een aantal kwaliteitscriteria voor te gebruiken literatuur zijn samengebracht.⁴ Zo schrijft de GR:

Voldoende bewijs voor carcinogeniteit. Er wordt een causale relatie verwacht tussen blootstelling aan een agens en het ontstaan van kanker. Ofwel, er is een positieve associatie vastgesteld in meer dan één humane studie tussen blootstelling en kanker waarbij toeval, bias en confounding redelijkerwijs zijn uitgesloten.

In dezelfde leidraad wordt in bijlage H1 de kwaliteitscriteria voor cohortonderzoek gegeven:

- a. Is er een duidelijke specifieke a priori hypothese onderzocht?
- b. Is het samenstellen van de blootgestelde groep zodanig gedaan dat het cohort aan het begin van follow-up qua ziektekansen vergelijkbaar mag worden verondersteld met de referentie groep? Dus, zouden beide groepen eenzelfde kankerincidentiepatroon vertonen als de te onderzoeken stof niet kankerverwekkend zou zijn (het healthy-worker effect is geen reden tot afwijzing)?
- c. Is de ziektestatus op een vergelijkbare wijze vastgesteld voor wat betreft de blootgestelde groep en de referentiegroep?
- d. Is de volledigheid en wijze waarop de follow-up is uitgevoerd betrouwbaar?
- e. Is de statistische analyse adequaat, waarbij voor leeftijdsverschillen, duur en periode van follow-up is gecorrigeerd?
- f. Is de invloed van versturende variabelen, die in staat zouden zijn de waargenomen verhoging in gezondheidseffecten te veroorzaken, op adequate wijze onder controle gebracht?

Aan criteria a, b, c, d en e moet altijd zijn voldaan. Voor criterium f geldt dat als er substantiële gedocumenteerde co-expositie is aan een stof die eenzelfde type carcinogeen effect heeft, aannemelijk dient te worden gemaakt dat deze co-expositie niet de oorzaak van het waargenomen effect is.

Wij constateren dat in geen van de drie gereviewde adviezen de GR consequent over deze kwaliteitscriteria rapporteert over de cohortonderzoeken die zij als basis ziet voor haar oordeel.

⁴ Wij hebben serieuze kritiek op de leidraad zelf (zie hoofdstuk 4) maar in dit hoofdstuk kijken we 'slechts' of de GR ten minste zelf haar eigen leidraad volgt.

Voor een van de belangrijkste confounding factors waar de GR dus rekening mee *moet* houden volgens haar eigen leidraad, dat is het aspect roken, hebben wij de onderliggende literatuur zelf beschouwd. Onze conclusie (zie de bijlagen van dit rapport) is dat hiermee zeker niet consequent goed is omgegaan in die onderliggende literatuur. Sommige van de auteurs van de artikelen die door de GR geciteerd worden melden dit ook zelf maar toch heeft de GR hier op zijn beurt geen melding van gemaakt.

Als voorbeeld nemen we het *Blootstelling aan Chroom CrVI*-advies van de GR. In dit advies is de onderstaande passage op pagina 22 de enige uitleg van de GR waarom de gebruikte literatuur de 'juiste' zou zijn:

Deze vier cohorten hebben betrekking op werknemers in de chromaatproductie, laten een verhoogd risico op longkanker zien (met uitzondering van de studie van Luippold et al. (2005)), sluiten roken als oorzaak van longkanker uit en grijpen terug op goed gedocumenteerde blootstellingsgegevens.⁵

Wat de GR hier doet is opschrijven dat zij zich aansluit bij de keuze van Seidler et al. in hun overzichtsartikel uit 2013.⁶ Eveneens neemt de GR het artikel van Seidler et al. als enige bron voor hun advies, dat wil zeggen dat de GR rechtstreeks de uitkomsten van Seidler et al. overneemt zonder een zichtbare eigen controle daarvan. Bijzonder is dan dat Seidler et al. zelf wel kanttekeningen maken bij de vijf studies die zij selecteren voor hun berekeningen (en die eigenlijk op twee datasets van plants in Baltimore en Paine gebaseerd zijn). Zo schrijven zij over de dataset van Paine: 'Smoking status was only known for a minority of the study population'. Voor de Baltimore dataset geldt volgens Seidler et al. dat 'Smoking status (yes/no) at the beginning of employment was known for over 90 % of the study subjects'. Omdat roken een veel sterkere invloed heeft op de incidentie van longkanker dan blootstelling aan chroom6 hebben keuzes over het rookgedrag van de betrokken werknemers een zeer grote invloed op de uitkomsten van de berekeningen. Noch in het advies van de GR, noch in Seidler et al, noch in de onderliggende studies treffen we daar op z'n minst een (gevoeligheids)analyse van aan.

⁵ Luippold, R. S., Mundt, K. A., Dell, L. D., & Birk, T. (2005). Low-level hexavalent chromium exposure and rate of mortality among US chromate production employees. *Journal of occupational and environmental medicine*, 47(4), 381-385.

⁶ A. Seidler, S. Jähnichen, J. Hegewald, A. Fishta, O. Krug, L. Rüter, C. Strik, E. Hallier & S. Straube (2013). Systematic review and quantification of respiratory cancer risk for occupational exposure to hexavalent chromium. *Int Arch Occup Environ Health* 86:943-955.

3. De GR wijkt af van goed wetenschappelijk gebruik

Er zijn enkele breed aanvaarde wetenschappelijke uitgangspunten voor onderzoek dat moet leiden tot risicoschatting van stoffen die verdacht worden van een carcinogeen effect. Deze zijn bijvoorbeeld samengevat in een artikel van Goldbohm et al. uit 2006 waaraan een gedurende jaren vooraanstaand lid van de GR heeft meegeschreven:⁷

'It is generally recognized that human, epidemiological data, if available, are preferred as the starting point for quantitative risk analysis above the use of data from animal studies. Although methods to obtain proper risk estimates from epidemiological data are available, several impediments prevent their widespread application. These impediments include unfamiliarity with epidemiological methods and the lack of a structured and transparent approach. We described a framework to conduct quantitative cancer risk assessment based on epidemiological studies in a structured, transparent, and reproducible manner. Important features of the process include a weight-of-the-evidence approach, estimation of the optimal exposure-risk function by fitting a regression model to the epidemiological data, estimation of uncertainty introduced by potential biases and missing information in the epidemiological studies, and calculation of excess lifetime risk through a life table to take into account competing risks. Sensitivity analyses are a useful tool to obtain insight into the impact of assumptions made and the variability of the underlying data. The framework is sufficiently flexible to allow many types of data, ranging from published, sometimes incomplete data to detailed individual data, while maintaining an optimal result, i.e., a state-of-the-art risk estimate with confidence intervals, based on all available evidence of sufficient quality.'

Cruciaal element voor elk wetenschappelijk risicoadvies, of het nu over de blootstelling aan stoffen of het instortingsrisico door geïnduceerde aardbevingen gaat, is *reproduceerbaarheid* door het hanteren van een wel beschreven analysekader en een gevoeligheidsanalyse. In die zin kan het bovenstaande citaat ook niet verbazen maar is het des te opvallender dat de GR in de drie beschouwde adviezen deze aanpak niet hanteert.

We merken op dat het genoemde Goldbohm et al. artikel zelf al het beschreven framework toepast op het *Blootstelling aan Chroom CrVI*-advies zodat de GR dat ook had kunnen doen of overnemen. Het artikel benoemt een aantal onzekerheden (zoals de omgang met de confounding factor roken dat een ordegrrootte verschil in de uitkomsten kan uitmaken) en vraagt om nader onderzoek naar de onzekerheden in de diverse stappen in de methode.

⁷ Goldbohm, R.A., Tielemans E.L., Heederik, D., Rubingh, C. M., Dekkers, S., Willems M.I., & Kroese E.D. (2006). Risk estimation for carcinogens based on epidemiological data: A structured approach, illustrated by an example on chromium. *Regulatory Toxicology and Pharmacology*, 44 (3): 294–310.

Het GR-rapport verwijst wel naar het Goldbohm et al. artikel in een overzichtstabel⁸ maar zegt niets over de geïdentificeerde onzekerheden.

Dat de vraag naar gevoeligheidsanalyses cruciaal is voor een wetenschappelijk valide advies op basis van de berekende risico's is voor iedereen zichtbaar in bijvoorbeeld onderstaand diagram (figuur 1) uit het eerder aangehaalde artikel van Seidler et al.:

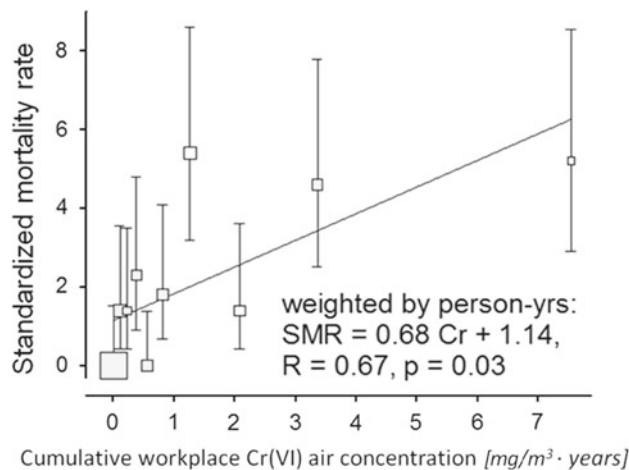


Fig. 5 Cumulative Cr(VI) exposure and standardized mortality rate (SMR) based on the data from Crump et al. (2003). Linear regression was performed weighted by person-years. SMRs are illustrated with confidence intervals (95 % CI). The size of the symbol is proportional to the number of person-years within the respective categories

Figuur 1: voorbeeld gevoeligheidsanalyse uit Seidler et al.

Deze figuur laat in een oogopslag zien dat de onzekerheden in de risicoberekening per groep van aan een bepaalde concentratie blootgestelde werknemers (zeer) groot is. De **aanname** dat blootstelling versus risico de vorm van een rechte lijn zou hebben leidt tot een risiconorm. Bedenk dat de rechte lijn, het toegepaste (gefitte) model, 'niet gebeurd' is; de data 'zijn gebeurd'.

Het voorbeeld is typerend voor een cruciaal probleem voor de kwaliteit van de GR-bewijsvoering voor de causale relatie tussen blootstelling en ziekte bij blootgestelden. Epidemiologische studies waarop de GR haar adviezen baseert kunnen per definitie geen causale verbanden leggen tussen blootstellingsniveaus en morbiditeit/mortaliteit; alleen correlaties kunnen zichtbaar worden gemaakt. In de adviezen wordt de vermeende sterkte van correlaties met een bepaalde significantie (p-waarde) gerapporteerd. Echter, gevonden significanties (het vinden van bepaalde p-waarden kleiner dan een van tevoren gekozen waarde, meestal 0,05) zeggen niets over het wel of niet waar zijn van welke hypothese dan ook.

⁸ De overzichtstabel laat zien dat de uitkomst van de meeste risico-inschattingen voor chroom6 op elkaar lijken. Dit kan natuurlijk niet erg verbazen omdat dezelfde 4 - 6 bronartikelen gebruikt worden die in essentie op de twee Amerikaanse en een Duitse dataset gebaseerd zijn.

Het onjuiste gebruik van significantie als bewijs voor causaliteit is statistici al jaren een doorn in het oog. De statistische gemeenschap hamert er sinds jaar en dag op dat toetsen op significantie onbruikbaar is voor het wel of niet bewezen achten van een bepaalde hypothese. In 2016 kwam deze ergernis tot uitbarsting in de *ASA's Statement on p-Values: Context, Process, and Purpose* van de American Statistical Association uit 2016 (met al bijna 3.000 citaties). Deze laat geen onduidelijkheid bestaan over de beperkte waarde van de p-waarde voor het aantonen van causaliteit: 'A p-value does not provide a good measure of evidence regarding a model or hypothesis.' En dus 'Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.'⁹

Zoals Hurlbert et al. (2019) opmerken:¹⁰

There is now wide agreement among many statisticians who have studied the issue that for reporting of statistical tests yielding p-values it is illogical and inappropriate to dichotomize the p-scale and describe results as "significant" and "nonsignificant." Authors are strongly discouraged from continuing this never justified practice that originated from confusions in the early history of modern statistics.

Bovenstaande heeft evident verstrekende consequenties voor het gebruik door de GR van epidemiologische studies voor het afleiden van gezondheidsstandaarden. Er is veel meer argumentatie nodig dan een simpele toets op significantie om tot een advies te komen (met vergaande maatschappelijke gevolgen).

Specifiek voor het *Blootstelling aan Chroom CrVI*-advies wijzen we op een ander bekend statistisch probleem dat Seidler et al. signaleren als 'remarkable' maar zonder statistische duiding: Gibb et al. en Park et al. bestuderen dezelfde Baltimore-dataset maar komen tot verschillende conclusies omdat Gibb et al. de werknemers in 4 groepen onderverdelen en Park et al. dezelfde werknemers in 5 groepen.^{11 12} De duiding hier zou moeten zijn, als onderdeel van een gevoeligheidsanalyse, dat als de keuze van het aantal subgroepen zo'n grote invloed heeft op de resultaten, dat daarmee de uitgevoerde statistische analyse niet stabiel genoeg is om definitieve conclusies aan te verbinden.

Het verweer van de GR in haar reactie op een kortere versie van bovenstaand commentaar van Crisislab (we verwezen naar een boek van Briggs hierover¹³) op het conceptadvies *Emissies van ijzer- en staalgieten* is dat deze handelingswijze conform a) de Bradford-Hill principes zijn die een gangbare epidemiologische praktijk zouden zijn en b) volgens de Globally Harmonized System criteria (GHS) zou zijn:

⁹ Wasserstein R.L. (2016). ASA Statement on Statistical Significance and P-Values. *The American Statistician*, 70(2), 129-133.

¹⁰ Hurlbert, S. H., Levine, R. A., & Utts, J. (2019). Coup de grâce for a tough old bull: "Statistically significant" expires. *The American Statistician*, 73(sup1), 352-357.

¹¹ Gibb, H. J., Lees, P. S., Pinsky, P. F., & Rooney, B. C. (2000). Lung cancer among workers in chromium chemical production. *American journal of industrial medicine*, 38(2), 115-126.

¹² Hier moet Park et al. maar die kan ik niet vinden

¹³ Briggs, W. (2016). *Uncertainty – The soul of modelling, probability and statistics*. Springer, Switzerland.

Proof of causation. Concerning the book by Briggs, the Subcommittee refers to the widely used Bradford-Hill criteria for causation in establishing epidemiological evidence of a potential causal relationship. In addition, the more observational studies show an association, the more likely that there is a causal relationship. Instead of using the term 'potential causal relationship', the Subcommittee prefers the use of the term 'association'. When there is evidence for a strong association, according to the Globally Harmonized System criteria, the substance or in this case the emission, should be classified as 'known to be carcinogenic to humans'.

Bron: Reactie GR op commentaar Crisislab op conceptadvies *Emissies van ijzer- en staalgieten*.

Het verweer vinden wij op beide punten ronduit zwak.

Het beroep op de Bradford-Hill criteria – In 1965 heeft de Engelse statisticus Austin Bradford Hill negen criteria voorgesteld om te bepalen of epidemiologische data op een causale relatie tussen oorzaak en gevolg duiden.¹⁴ Hij gebruikte deze criteria om het effect van roken op longkanker vast te stellen. De lijst van negen criteria van Bradford-Hill uit 1965:

- Strength (effect size): A small association does not mean that there is not a causal effect, though the larger the association, the more likely that it is causal.
- Consistency (reproducibility): Consistent findings observed by different persons in different places with different samples strengthens the likelihood of an effect.
- Specificity: Causation is likely if there is a very specific population at a specific site and disease with no other likely explanation. The more specific an association between a factor and an effect is, the bigger the probability of a causal relationship
- Temporality: The effect has to occur after the cause (and if there is an expected delay between the cause and expected effect, then the effect must occur after that delay).
- Biological gradient (dose-response relationship): Greater exposure should generally lead to greater incidence of the effect.
- Plausibility: A plausible mechanism between cause and effect is helpful.
- Coherence: Coherence between epidemiological and laboratory findings increases the likelihood of an effect.
- Experiment: Ideally there should be experimental evidence.
- Analogy: The use of analogies or similarities between the observed association and any other associations.

¹⁴ Hill, A.B. (1965). The Environment and Disease: Association or Causation? *Proceedings of the Royal Society of Medicine*. 58 (5): 295–300.

Het debat over de al dan niet causale relatie tussen data en werkelijkheid heeft echter niet stilgestaan sinds 1965. Al sinds decennia is binnen de epidemiologie onderkend dat de Bradford-Hill criteria niet voldoen om causaliteit met voldoende zekerheid aan te tonen.¹⁵ Overigens faalt het beroep op de Bradford-Hill criteria door de GR omdat zij nergens in haar rapport deze criteria *navolgbaar* toepast.

Het beroep op het GHS – Opgemerkt moet worden dat de bijlage GHS waarin de wijze van hazard assessment expliciet wordt beschreven als niet verplichtend zodat de eigen verantwoordelijkheid voor de GR overeind blijft staan en dat ook de GHS alleen spreekt in kwalitatieve termen als ‘voldoende’.

<u>CATEGORY 1:</u>	Known or presumed human carcinogens The classification of a substance as a Category 1 carcinogen is done on the basis of epidemiological and/or animal data. This classification is further distinguished on the basis of whether the evidence for classification is largely from human data (Category 1A) or from animal data (Category 1B):
Category 1A:	Known to have carcinogenic potential for humans. Classification in this category is largely based on human evidence.
Category 1B:	Presumed to have carcinogenic potential for humans. Classification in this category is largely based on animal evidence. The classification of a substance in Category 1A and 1B is based on strength of evidence together with weight of evidence considerations (See paragraph A.6.2.5). Such evidence may be derived from: - human studies that establish a causal relationship between human exposure to a substance and the development of cancer (known human carcinogen); or - animal experiments for which there is sufficient evidence to demonstrate animal carcinogenicity (presumed human carcinogen). In addition, on a case by case basis, scientific judgment may warrant a decision of presumed human carcinogenicity derived from studies showing limited evidence of carcinogenicity in humans together with limited evidence of carcinogenicity in experimental animals.

Bron: Non-mandatory Appendix F Part A for further guidance regarding hazard classification for carcinogenicity.

Interessant in dat verband is overigens dat de GHS zelf spreekt over een aantoonbaar *causaal* verband wanneer het gaat om *voldoende* bewijs voor carcinogeniteit:

¹⁵ Zie bijvoorbeeld: Schünemann, H., Hill, S., Guyatt, G. et al. (2011). The GRADE approach and Bradford Hill's criteria for causation. *Journal of Epidemiology & Community Health*. 65 (5): 392–95; Höfler, M. (2005). The Bradford Hill considerations on causality: a counterfactual perspective? *Emerging Themes in Epidemiology*. 2 (1), 1 - 9; Howick, J., Glasziou, P., & Aronson, J. K. (2009). The evolution of evidence hierarchies: what can Bradford Hill's 'guidelines for causation' contribute? *Journal of the Royal Society of Medicine*. 102 (5); Glass, T. A., Goodman, S. N., Hernán, M. A., & Samet, J.M. (2013). Causal inference in public health. *Annual Review of Public Health*. 34: Potischman, N., & Weed, D. L. (1999). Causal criteria in nutritional epidemiology. *American Journal of Clinical Nutrition*. 69 (6): 1309S–14S; Phillips, C. V., & Goodman, K. J. (2006). Causal criteria and counterfactuals; nothing more (or less) than scientific common sense. *Emerging Themes in Epidemiology*. 3. (1): 1-7.

Carcinogenicity in humans:

The evidence relevant to carcinogenicity from studies in humans is classified into one of the following categories:

- (a) **Sufficient evidence of carcinogenicity:** A causal relationship has been established between exposure to the agent and human cancer. That is, a positive relationship has been observed between the exposure and cancer in studies in which chance, bias and confounding could be ruled out with reasonable confidence.
- (b) **Limited evidence of carcinogenicity:** A positive association has been observed between exposure to the agent and cancer for which a causal interpretation is considered by the Working Group to be credible, but chance, bias or confounding could not be ruled out with reasonable confidence.

Bron: Non-mandatory Appendix F Part A for further guidance regarding hazard classification for carcinogenicity.

Kortom, de GR moet nog steeds zelf volgens de meest recente wetenschappelijke inzichten kunnen uitleggen wanneer zij epidemiologische studies van voldoende kwaliteit acht om als basis voor haar eigen inschatting te dienen dat er een causaal verband zou kunnen worden verwacht.

Dit alles staat nog los van het feit dat hazard assessments als zodanig wetenschappelijk achterhaald zijn. De GR rept daarover met geen woord. Een potentieel gevaar geeft geen enkel uitsluitsel over het daadwerkelijk risico van het specifieke agens. Onder de hazard systematiek vallen bananenschillen in dezelfde categorie als auto's. Beide kunnen leiden tot ongelukken en letsel, maar het risico daarop is fors verschillend. Bananenschillen zullen zelden leiden tot (ernstig) letsel bij personen, terwijl als gevolg van auto-ongelukken jaarlijks vele doden zijn te betreuren. Zoals Boobis et al. (2016) opmerken:¹⁶

The EU in its Classification and Labelling Guidelines (ECHA, 2012) has implemented the UN GHS categorization system which is very similar using strength of evidence involving only the enumeration of tumors in human and animal studies and determination of their level of statistical significance. [...] Thus a system set up half a century ago based on an overly simplistic concept as an initial attempt to address the disease burden of cancer has found its way into classification schemes for chemicals and also into some downstream risk management processes. [...] The pressure to replace a chemical which had been identified as a potential human carcinogen was immense, even on the basis of animal studies, and the very act of categorizing a chemical in this way leads to a stigma which would often result in major changes in its use, including withdrawal, whereas risk-based assessments show that there is reasonable certainty no harm will result from its use.

¹⁶ Boobis, A. R., Cohen, S. M., Dellarco, V. L., Doe, J. E., Fenner-Crisp, P. A., Moretto, A., ... & Wolf, D. C. (2016). Classification schemes for carcinogenicity based on hazard-identification have become outmoded and serve neither science nor society. *Regulatory Toxicology and Pharmacology*, 82, 158-166.

4. Adviezen moeten proportionele bestuurlijke besluitvorming faciliteren

‘Goed’ overheidsveiligheidsbeleid kenmerkt zich door een proportionele, integrale blik zo heeft de regering meerdermalen betoogd. Nieuw veiligheidsbeleid moet gebaseerd zijn op een afweging tussen de maatschappelijke kosten en baten van beleid.¹⁷ Dat geldt dus ook voor overheidsbeleid op het gebied van bescherming van werkenden tegen blootstelling aan stoffen.

Berekening van de kosten van arbeidsomstandighedenbeleid is ingewikkelder dan simpelweg de kosten van maatregelen tegen de blootstelling aan stoffen. Zo moet worden meegewogen dat het überhaupt hebben van een baan een enorme baat is. Zeker voor lager opgeleide mensen scheelt het statistisch zo maar een tiental gezonde levensjaren wanneer men werk heeft.

Berekening van de baten van stoffenbeleid is echter zo mogelijk nog complexer omdat we nog beperkt zicht hebben op de gezondheidsschade van veel stoffen. Waar er over de risico's van blootstelling aan hoge concentraties van veel stoffen geen discussie bestaat, bestaat die discussie wel als het gaat om omvang, aard en de al dan niet cumulatieve effecten van blootstelling aan kleine hoeveelheden van stoffen. Deze worden theoretisch bepaald op basis van modelmatige aannames en keuzes.

De wetenschappelijke opgave is om de maatschappelijke kosten-baten analyse, inclusief alle onzekerheden, zo helder mogelijk op de bestuurlijke tafel te krijgen.

Vanuit dit perspectief zien wij een aantal problemen met de methode van de GR.

4.1 Problemen met de GR Leidraad classificatie carcinogene stoffen: bewust conservatisme in beoordeling of iets kankerverwekkend is

Wij hebben een fundamenteel probleem met het beoordelingsschema uit de leidraad classificatie carcinogene stoffen: dit beoordelingsschema kiest het schijnbare ‘zekere voor het onzekere’. Dit klinkt als prudent adviseren maar is dat alleen vanuit het smalle perspectief van de blootstelling aan stoffen. Het bredere perspectief van proportioneel beleid is daarmee verloren gegaan.

¹⁷ Meest recent in het kabinetsstandpunt ‘Bestuurlijk balanceren met risico's en verantwoordelijkheden’ uit 2015.

Het beoordelingsschema ziet eruit als volgt:

Epidemiologie Dierexperimenten ^a	++	+	?	-
	++	+	?	-
++	1A	1A	1B	2
+	1A	2	2	4
?	1A	2	3	4
-	1A	3	3	4

Waarin op de assen de volgende beoordeling van carcinogeniteit in mensen en dieren staat:

Beoordeling resultaten in epidemiologisch onderzoek

- ++ *Voldoende bewijs voor carcinogeniteit.* Er wordt een causale relatie verwacht tussen blootstelling aan een agens en het ontstaan van kanker. Ofwel, er is een positieve associatie vastgesteld in meer dan één humane studie tussen blootstelling en kanker waarbij toeval, *bias* en *confounding* redelijkerwijs zijn uitgesloten.
- + *Beperkt bewijs voor carcinogeniteit.* Er is een positieve associatie vastgesteld in humane studies tussen blootstelling en kanker. Echter, het kan niet met zekerheid worden uitgesloten dat toeval, *bias* en *confounding* een belangrijke rol spelen.
- ? *Er zijn geen (dan wel onvoldoende) gegevens* om een uitspraak te kunnen doen over een associatie tussen blootstelling aan een agens en kanker.
- *Voldoende bewijs voor het afwezig zijn van carcinogeniteit* in meer dan één humane studie.

Beoordeling resultaten in dierexperimenteel onderzoek

- ++ *Voldoende bewijs voor carcinogeniteit.* Er is een causale relatie vastgesteld tussen blootstelling aan een agens en kwaadaardige tumoren in:
 - a twee of meerdere diersoorten, of
 - b in twee of meerdere onafhankelijke studies in één diersoort.
- + *Beperkt bewijs voor carcinogeniteit.* De dierexperimentele gegevens suggereren de aanwezigheid van een carcinogeen effect, maar een definitieve conclusie kan (nog) niet worden getrokken omdat:
 - a er slechts gegevens beschikbaar zijn uit één dierexperiment;
 - b er meerdere positieve dierstudies beschikbaar zijn, maar er twijfels bestaan over de kwaliteit van de studies qua opzet en interpretatie;
 - c de beschikbare studies van goede kwaliteit zijn maar de resultaten geen eenduidig beeld geven, ofwel de gegevens niet eenduidig zijn te interpreteren (bijvoorbeeld een verhoging van het aantal goedaardige tumoren of van tumoren met een hoge achtergrond incidentie).
- ? *Er zijn geen (dan wel onvoldoende) gegevens* om een uitspraak te kunnen doen over een associatie tussen blootstelling aan een agens en kanker.
- *Voldoende bewijs voor het afwezig zijn van carcinogeniteit* in dierexperimentele studies.

Zichtbaar is daarmee dat slechts als er hard bewijs is dat een stof niet carcinogeen is, de stof ook als 4 'het is niet waarschijnlijk dat de stof kankerverwekkend is' wordt beschouwd. In een situatie van 'twijfels over de kwaliteit van de onderzoeken', wordt een stof nog steeds als 1A -'de stof is kankerverwekkend voor de mens'- beschouwd.

4.2 *Problemen met de GR Leidraad classificatie carcinogene stoffen: geen afwegingskader voor omgang met tegengestelde resultaten*

In zekere zin in het verlengde van ons vorige probleem is de afwezigheid van een afwegingskader voor de omgang met de telkens terugkomende situatie waarin sommige artikelen die schijnbaar aan alle kwaliteitseisen van de GR voldoen geen correlatie vinden tussen blootstelling en kanker en sommige andere wel.

Wij schreven daarover in ons review van het conceptadvies over de blootstelling aan stoffen bij ijzer- en staalgieten:

Overviewing the Committee's work on these studies, it is surprising that no formal risk assessment framework is offered. No methodology is utilised other than a summarising of studies which, as we have shown, in some cases lack precision, followed by an expert judgement that overall these studies show a genotoxic effect. This judgement therefore cannot be considered to be scientifically valid.

Dit probleem is in meerdere commentaren van verschillende deskundigen op conceptadviezen gegeven maar nooit expliciet geadresseerd door de GR.

Alternatieve studies waarin een voorstel voor een grenswaarde is gedaan, zijn niet meegenomen in het advies, noch meegewogen. Binnen alle mogelijke benaderingen heeft de Gezondheidsraad gekozen voor die benadering die tot de hoogste risicoschatting leidt.¹⁸

De GR-methode lijkt te zijn om ervoor te kiezen slechts die artikelen mee te nemen die een positieve correlatie vinden tussen blootstelling en kanker en dus de andere artikelen simpelweg 'voor de zekerheid' te negeren.

De GR staat overigens niet alleen bij deze aanpak. Als voorbeeld kan de veelheid van risico-adviezen voor chroom6 gelden. In deze adviezen wordt telkens de analyse van Luippold et al. (2003) genoemd als een studie met voldoende kwaliteit.¹⁹ Luippold et al. laten echter geen relatie zien tussen blootstelling aan chroom6 en meer longkanker. De studie van Luippold et al. wordt vervolgens zonder nadere uitleg overal (ook bij de GR) buiten beschouwing gelaten.

4.3 *Problemen met de GR Leidraad classificatie carcinogene stoffen: uit zekerheid voor stochastisch kiezen*

Een klassiek probleem is de bepaling of de blootstelling aan de stof stochastisch of niet-stochastisch (verdeeld) is.

- Stochastisch wil hier zeggen dat de kans op kanker lineair samenhangt met de blootstelling aan een stof. Dus een tweemaal hogere blootstelling betekent een tweemaal hogere kans op kanker.
- Bij een niet-stochastische verdeling kan bijvoorbeeld rekening worden gehouden met een drempelwaarde waaronder een stof helemaal geen nadelige effecten meer heeft. Van veel stoffen is een dergelijke drempelwaarde bekend zoals voor vitamines die schadelijk zijn in hogere dosis maar onschadelijk en zelfs noodzakelijk in kleine hoeveelheden. Voor in hoge dosis carcinogene stoffen als zuurstof geldt dat het lichaam een herstelmechanisme heeft om de schade veroorzaakt door lage dosis te repareren.

Reactieve vormen van zuurstof (ROS) worden in normale omstandigheden in grote hoeveelheden gevormd door stofwisselingsprocessen en ontstekingsreacties in het lichaam. Deze kunnen aanleiding geven tot oxidatieve DNA-schade (oxidatieve stress). Ames en Gold (1991) hebben geschat dat in elke cel van de rat – onder steady state omstandigheden – ongeveer 1×10^6 (één miljoen) oxidatieve DNA-adducten aanwezig zijn en dat er dagelijks ongeveer 1×10^5 nieuwe oxidatieve DNA-adducten worden gevormd.²⁰ Op cellulair niveau bestaat er echter een krachtig antioxidant afweersysteem dat ROS onschadelijk maakt, alsmede een DNA-herstelsysteem, dat ervoor zorgt dat ROS-geïnduceerde schade aan het DNA snel wordt hersteld. Dankzij deze uiterst efficiënte afweer- en herstelsystemen kan de mens toch leven in een zuurstofrijke omgeving.

¹⁸ Jongeneel, F.J. (2018). *Commentaar conceptrapport DME*. IndusTox.

¹⁹ Luippold, R. S., Mundt, K. A., Austin, R. P., Liebig, E., Panko, J., Crump, C., & Proctor, D. (2003). Lung cancer mortality among chromate production workers. *Occupational and environmental medicine*, 60(6), 451-457.

²⁰ Ames, B. N., & Gold, L. S. (1991). Endogenous mutagens and the causes of aging and cancer. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 250(1-2), 3-16.

De GR heeft er bewust voor gekozen om in haar adviezen uit zekerheid voor een stochastisch model te kiezen. In de *Leidraad classificatie carcinogene stoffen* uit alweer 2010 zegt de GR daarover:

Aan de lineariteit bij lage blootstelling liggen enkele veronderstellingen ten grondslag. Een eerste is dat het optreden van DNA-schade een stochastisch proces is, waarbij wel of geen “DNA-schade” de toestandsvariabele is, en de kans of waarschijnlijkheid van het optreden daarvan door blootstelling aan een kankerverwekkende stof, door toeval wordt bepaald en niet door de mate van “DNA-schade”. Een tweede is dat er geen drempelwaarde bestaat waaronder een stof die treffers veroorzaakt als niet-werkzaam beschouwd kan worden, met andere woorden er bestaat geen dosis waarbij de kans op een relevant effect gelijk is aan nul. De one-hit kinetiek bij lage blootstelling betekent dat blootstelling, hoe laag ook, altijd een verhoogd risico op kanker met zich meebrengt. Uit veiligheidsoverwegingen kan daarom het beste een referentiewaarde worden afgeleid.

Overigens treden continu DNA-veranderingen op door genotoxische carcinogenen die van nature aanwezig zijn in de omgeving en de voeding, of die ontstaan door normale stofwisselingsprocessen en ontstekingsreacties in het lichaam, zoals verschillende reactieve zuurstofradicalen (achtergrondtreffers). Veel van de DNA-schade wordt door efficiënt werkende DNA-herstelenzymen weer gecorrigeerd. Wat de risico-evaluatie betreft beoordeelde de commissie tot nu toe alleen het vermogen van stoffen om DNA-schade te kunnen veroorzaken, ongeacht het vermogen van het lichaam om deze schade te kunnen herstellen.

Onder het kopje ontwikkelingen op het gebied van carcinogene werkingsmechanismen schrijft de GR:

Sinds het vorige rapport uit 1996 van de Gezondheidsraad over de beoordeling van de carcinogeniteit (en werkingsmechanismen) van chemische stoffen heeft een enorme ontwikkeling plaats gevonden in de inzichten over de relevante mechanismen op celbiologisch en moleculairbiologisch gebied, maar dit heeft voorsnog niet geleid tot een principiële andere werkwijze voor de risicoschatting betreffende carcinogene verbindingen.

De commissie voorziet wel dat in de toekomst op basis van die verworven inzichten met meer zekerheid onderscheid gemaakt kan worden in de te verwachten blootstellingsresponsrelatie van potentieel carcinogene stoffen voor de mens, en dat wellicht zelfs veilige blootstellingsniveaus voor stochastisch genotoxische carcinogenen voor de mens zijn te definiëren.

Zo kunnen verdedigingsmechanismen en achtergrondeffecten verklaren waarom pas bij hogere blootstelling de door stochastisch genotoxische verbindingen ontstane DNA-schade en mutaties zichtbaar worden. [...] Zo kan:

- een schadelijk metaboliet worden weggevangen, bijvoorbeeld de zuurstofradicalen door antioxidanten en radicaalscavengers
- via biotransformatie een carcinogene stof omgezet worden in onschadelijke metabolieten
- herstel van DNA plaatsvinden door bepaalde enzymen
- de celdeling bij schade vertraagd worden, zodat er meer tijd is voor DNA-herstel
- het afweersysteem de tumorcellen en voorlopers daarvan herkennen en onschadelijk maken.

Uiteindelijk bepaalt de balans tussen blootstelling en verdedigingsmechanismen de kans op DNA-schade en het ontstaan van een tumorcel; overbelasting van deze verdedigingsmechanismen versterkt de effecten van een (genotoxisch) carcinogene stof.

Opvallend is dat de GR zelfs de genotoxiciteit van stoffen die dezelfde tussenproducten (ROS) opleveren die bij lichaamseigen processen ontstaan, beoordeelt met de lineaire modelering waarvan bekend is dat die dan niet correct is:

Ook bepaalde lichaamsvreemde carcinogene stoffen kunnen ROS produceren en het vermoeden bestaat dat dit in sommige gevallen het mechanisme is waarmee ze – zij het bij hoge blootstelling – kanker veroorzaken zoals aangetoond is in proefdierexperimenten. Dat geen toename van tumoren wordt gezien bij lage blootstelling kan worden verklaard doordat de hoeveelheid ROS die wordt gevormd door de carcinogene stof, verwaarloosbaar was ten opzichte van de hoeveelheid reeds gevormde zuurstofradicalen door normale cellulaire processen. [...] De kans op DNA-schade neemt dus niet toe bij lage blootstelling en dit betekent dat er sprake is van een drempel waar beneden geen noemenswaardige effecten waarneembaar zijn. Voorbeelden van carcinogene stoffen die ROS genereren zijn pyrocatechol en cadmium. Het zal in de praktijk echter niet eenvoudig zijn om zo'n drempel voor een dergelijke stof aan te tonen, omdat daarvoor gegevens nodig zijn over de status van de oxidatieve stress op cellulair niveau in relatie tot de tumorincidentie als functie van de blootstelling aan de stof binnen één dierexperimenteel onderzoek. Dit is nodig om vast te stellen of er een causaal verband is tussen die oxidatieve schade en de carcinogeniteit, en of er inderdaad een experimenteel aantoonbare drempel is. Indien dergelijke gegevens niet beschikbaar zijn, zal de commissie genotoxische carcinogenen die ROS genereren beoordelen als stochastisch genotoxische carcinogenen vanwege de genotoxiciteit van de gevormde ROS. Dit impliceert dat de methodiek voor de risicoschatting de lineaire extrapolatie is.

In het *Blootstelling aan Chroom CrVI*-advies schrijft de GR in dezelfde lijn:

De commissie merkt op dat chroom VI-verbindingen op meerdere manieren schade kunnen toebrengen aan het DNA en concludeert dat de onderliggende mechanismen van genotoxiciteit voornamelijk non-stochastisch zijn, maar echter ook stochastisch kunnen zijn. In het geval van stochastisch genotoxische mechanismen wordt aangenomen dat er kanker kan optreden bij elk blootstellingsniveau, terwijl in geval van non-stochastische mechanismen veilige blootstellingsdrempels kunnen bestaan. Voorzichtigheidshalve, zoals aangegeven in haar leidraad, kiest de commissie bij haar risicoberekening voor het 'worst case'-scenario en neemt het stochastisch genotoxisch mechanisme als uitgangspunt. Dit impliceert dat lineaire extrapolatie wordt toegepast voor de berekening van risicogetallen: de concentraties in de lucht die samenhangen met een extra kans op kanker van 4 per 1.000 en 4 per 100.000 door beroepsmatige blootstelling aan stoffen. Deze risicogetallen vormen vervolgens de *wetenschappelijke* [cursivering door reviewers] basis voor de door de minister vast te stellen grenswaarde.

Ons centrale probleem is daarmee dat de GR een 'voorzichtige' -in feite beleidsmatige- keuze maakt en deze presenteert als wetenschappelijk. Dat dit onterecht is maken Ricci & Sammis ondubbelzinnig duidelijk:²¹

Attitude towards risk should not be included in cancer risk analysis, but neutrality should. If it were, risk aversion or proneness should be calculated along with the risk-neutral information, on a case-by-case basis. But then: whose individual aversion or proneness is an agency using?

Dat de keuze voor worst-case risicoberekeningen tot verbazingwekkende uitkomsten kan leiden laat het *Dieselmotoremissie*-advies van de GR zien: het berekende risiconiveau waarbij 1 op de 10.000 blootgestelde medewerkers na blootstelling gedurende 8 uur per dag longkanker oploopt na 40 jaar werk is $1 \mu\text{g}/\text{m}^3$. Dit is ongeveer het achtergrondniveau voor Nederland volgens de GR zelf.

Een simpele berekening, uitgaande van een lineair verband zoals de GR die zelf hanteert, laat daarmee zien dat onder 17 miljoen Nederlanders er ongeveer 14.000 longkankerincidenties per jaar zouden moeten zijn (nog geen rekening houdend met kwetsbare groepen). Feitelijk zijn er 'maar' 13.000 gevallen van longkanker in Nederland, waarvan volgens de gebruikelijke modellen 90% veroorzaakt wordt door roken. Het berekende risiconiveau als gevolg van blootstelling aan dieselmotoremissie lijkt daarmee tenminste een ordegrootte te hoog te zijn.

²¹ Ricci, P.F. & Sammis, I.R. (2012). The Role of Linear and Nonlinear Dose-Response Models in Public Decision-Making, *Dose-Response*, 10:203–208.

5. Tot slot

Het moet ons van het hart hoezeer wij het betreuren dat de Gezondheidsraad naar aanleiding van het commentaar op bijvoorbeeld het conceptadvies over ijzer- en staalgieten stelt dat:

‘Commentaren over procedures, haalbaarheid, gevoerde of te voeren beleid, en uitvoering (beschermende maatregelen) worden niet beantwoord, omdat deze onderwerpen de Gezondheidsraad niet aangaan en buiten de reikwijdte van het advies vallen.’

Daarmee sluit de GR zich niet alleen af van wat de basis voor wetenschap is in onze ogen, namelijk een open discussie over feiten, analyse(methoden) en conclusies, maar geeft zij ook aan dat adviezen van de Gezondheidsraad eigenlijk van geen waarde moeten worden geacht voor de samenleving.

Bijlage 1: Comments on the choice of literature used for ‘Diesel Engine Exhaust - Health-based recommended occupational exposure limit’ by the Health Council of the Netherlands

In deze bijlage geven we enkele punten van commentaar op de literatuur die de GR gebruikt om haar advies te onderbouwen over de normering van de blootstelling aan dieselmotorenemissie.

1. Introduction

The Dutch Expert Committee on Occupational Safety (DECOS), in their study of DEE (Health Council, 2019), deemed only four studies qualified to derive HBC-OCRVs, as these studies purportedly include quantitative exposure-response data. In their own words (p. 29):

‘Numerous cohort and case-control studies are available on lung (and bladder) cancer among workers exposed to DEE. However, only four studies included quantitative exposure-response data, which could be used to derive HBC-OCRVs (Steenland et al. (1998); Garshick et al. (2012); Silverman et al. (2012) and Möhner et al. (2013). These include two nested case-control studies and two retrospective cohort studies on lung cancer mortality, in which DEE exposures (REC) are estimated retrospectively. Two studies were performed in the trucking industry and two in the mining industry. All four studies concerned exposure to DEE from engines powered with petroleum-diesel fuels, without the newest technologies to reduce emission of particulates.’

And (p. 31):

‘Three of the four studies showed statistically significant positive associations and trends between cumulative REC exposure and lung cancer mortality in the trucking and mining industry, in which workers were mainly exposed to DEE (Steenland et al. (1998); Garshick et al. (2012); Silverman et al. (2012)). No association was found by Möhner et al. (2013).’

The focus is on respirable elemental carbon particles (RECP) in DEE per cubic metre (m³) of air. Elemental carbon particles is regarded as a surrogate (marker) for diesel exhaust.

2. Steenland et al. (1998); Garshick et al. (2012); Silverman et al. (2012) and Möhner et al. (2013)

As these four studies form the basis of the HBC-OCRVs, we need to take a look at the individual studies as to comprehend the strength of the health-based occupational exposure limits as proposed by the Health Council. We will also mention those aspects that the Health Council chooses to ignore.

Let's start with Möhner et al. (2013), as this study among the four used by the Health Council gave *no evidence* for an association between respirable elemental carbon (REC) exposure and lung cancer risk in potash miners. The authors did correct for smoking, although that was not done extensively. The older men who were 'measured' badly, had higher lung cancer cases, both because of age, and probably smoking (see Tables 1 and 2 in Möhner et al. (2013)).

Another interesting aspect of this study is the cumulative exposure the authors calculated for the miners. Nonetheless, cumulative exposure obviously correlates with age, and age correlates with cancer, so exposure is a to all intents and purposes a proxy for age. The authors tried to fix this with case-control matching: 'Five controls were matched to each lung cancer case, i.e. those five controls whose date of birth was closest to that of the case were selected.'

The summary of this study can be found in Table 9. The odds ratios are small and the CI's are extremely large, meaning the uncertainty is substantial, and even larger than published, considering the above. Carrying through uncertainty in the whole analysis-process is essential but rarely done.

Silverman et al. (2012) contains a major problem known as the epidemiologist fallacy: the authors never got actual measures of exposures, only crude estimates (see p. 857). That means that the correlation between found cancer cases and actual exposure in principle cannot be known.

The authors conducted a nested case-control study in a cohort of 12,315 workers in eight non-metal mining facilities. Among those workers, 217 deaths were identified with lung cancer. Even if every single one of these cases was caused by diesel, the rate would be 0.018. That is quite low. Obviously, not every case will be caused by diesel exhaust, so the actual rate must be even lower than this.

The authors selected each case subject up to four control subjects, which is low and will exaggerate the overall cancer risk. Ideally, which of course is not easy to do, one would use the same ratio as found in real life.

Another problem emerges as for 'average REC intensity and cumulative REC exposure, we evaluated lag intervals by excluding exposure occurring 0, 3, ..., 25 years (by 2-year intervals) before the death/reference date and compared changes in model deviance to a model that omitted REC exposure. The optimal lag interval (ie, the largest improvement in model fit) occurred for a lag between 13-17 years for average REC intensity and 15 years for cumulative REC exposure ...'. This means that Silverman et al. did multiple testing, and as the authors use small p-values as to define significance, searching for the smallest p-value is bound to exaggerate the evidence. Finally, none of their models in Figure 1 fit very well.

Garshick et al. (2012) contains the same major problem as Silverman et al. 2012: no actual exposure of workers was determined. What the authors did was model historical trends in ambient terminal EC based on historical trends in the coefficient of haze available for 1971 through 2000, a measurement of particulate matter based on optical density (p. 1302). This of course is a very crude proxy of exposure. Moreover, the uncertainty in the proxy itself is not carried through.

In total 31,135 male workers employed in the unionized U.S. trucking industry in 1985 were scrutinised. Through 2000, there were 779 lung cancer cases (734 underlying cause). As with Silverman et al. the rate is quite small: 0.024. Again, not every case will be caused by diesel exhaust, so the actual rate must be even lower than this. Also, by their own admission, the authors *included* deaths with primary lung cancer as cases, even if lung cancer was not the underlying cause of death. So, that means that hazard was further exaggerated.

Steenland et al. (1998) again does not provide actual workers-exposure to DEE. As Steenland et al. remark: 'Diesel exhaust exposure was estimated based on a 1990 industrial hygiene survey. Past exposures were estimated assuming that they were a function of 1) the number of heavy duty trucks on the road, 2) the particulate emissions (grams/mile) of diesel engines over time, and 3) leaks from trucks' exhaust systems for long-haul drivers.' However, the authors do not account for the uncertainty in the estimates.

Interestingly, Figure 2 and 3, in the authors' paper, show a clear decline in emissions over the years depicted. Updated to 2019, diesel exhaust would even be reduced further. So even if high levels of DEE would have caused cancer, that would not be the case right now as we don't have high exhaust levels any more. Also, odds ratios that the authors reported have a relatively large confidence interval.

References

Boobis, A. R., Cohen, S. M., Dellarco, V.L., Doe, J. E., Fenner-Crisp, P. A., Moretto., A. Pastoor, T.P., Schoeny, R. S., Seed, J. G. & Wolf, D. C. (2016). Classification schemes for carcinogenicity based on hazard- identification have become outmoded and serve neither science nor society. *Regulatory Toxicology and Pharmacology* 82: 158–166.

Briggs, W. (2016). *Uncertainty – The soul of modelling, probability and statistics*. Springer, Switzerland.

Centers for Disease Control and Prevention (CDC) 1999. Achievements in Public Health, 1900-1999: Tobacco Use -- United States, 1900-1999. *Morbidity and Mortality Weekly Report* 48(43): 986–993. See <https://www.cdc.gov/mmwr/preview/mmwrhtml/mm4843a2.htm> (last accessed on the 18th of February 2019).

Garshick, E., Laden, F., Hart, J. E., Davis, M.E., Eisen, E.A., & Smith, T.J. (2012). Lung cancer and elemental carbon exposure in trucking industry workers. *Environmental Health Perspectives* 120(9), 1301-1306.

Gibb, H. J., Lees, P.S., Pinsky, P.F., & Rooney, B.C. (2000). Lung cancer among workers in chromium chemical production. *American Journal of Industrial Medicine* 38(2), 115-126.

Mohner, M., Kersten N., & Gellissen, J. (2013). Diesel motor exhaust and lung cancer mortality: reanalysis of a cohort study in potash miners. *European Journal of Epidemiology* 28(2), 159-68.

The Health Council of the Netherlands, Subcommittee on the Classification of carcinogenic substances of the Dutch Expert Committee on Occupational Safety. 2010. *Guideline to the classification of carcinogenic compounds. Guide for classifying compounds in terms of their carcinogenic properties, and for assessing their genotoxicity*. Report No. A10/07E, The Hague.

The Health Council of the Netherlands (2019). *Diesel Engine Exhaust. Health-based recommended occupational exposure limit*. Report No. 2019/02, The Hague, the Netherlands.

Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006. 2008. *Official Journal of the European Union* L 353: 1–1355.

Ricci, P.F., & Sammis, I.R. (2012). Regulatory-Science: Biphasic Cancer Models or the LNT-Not Just a Matter of Biology! *Dose-Response*, 10(2):120-154.

Silverman, D.T., Samanic, C.M., Lubin J.H., Blair, A.E., Stewart, P.A., Vermeulen, R., (2012). The Diesel Exhaust in Miners study: a nested case-control study of lung cancer and diesel exhaust. *Journal of the National Cancer Institute* 104(11): 855-868.

Steenland, K., Deddens, J., & Stayner, L. (1998). Diesel exhaust and lung cancer in the trucking industry: exposure-response analyses and risk assessment. *American Journal of Industrial Medicine* 34(3): 220-228.

Bijlage 2: Comments on the choice of literature used for ‘Blootstelling aan chroom CrV’ by the Health Council of the Netherlands

In deze bijlage geven we enkele punten van commentaar op de literatuur die de GR gebruikt om haar advies te onderbouwen over de normering van de blootstelling aan chroom-6 verbindingen.

1. Introduction

The Committee on CrVI compounds selected only few studies (employee cohort studies) to assess exposure and disease risks. The following are selected: Hayes et al. (1979), Braver et al. (1985), Gibb et al. (2000), Mancuso (1975), Mancuso (1997), Luippold et al. (2003), Luippold et al. (2005), Birk et al. (2006) and Mundt et al. (2002).

The Health Council Committee remarks that all studies, except Luippold et al. (2005) show and increased risk of lung cancer (the endpoint of choice), exclude smoking, and exposure data are well-documented. Let's have a look.

2. Hayes et al. (1979), Braver et al. (1985), Mancuso et al. (1997) Gibb et al. (2000), Luippold et al. (2003), Luippold et al. (2005), Birk et al. (2006)

Hayes et al. (1979) was designed to determine whether employment in a modernized factory had resulted in a reduction in the risk of lung cancer as compared to the risk for workers employed, at approximately the same period in time, in an old production facility. Cigarette smoking history data were abstracted from the plant medical records, it was found to be of insufficient quality to allow analysis. Comparison with the Baltimore population makes for a rather poor control situation, as factory and city populations vastly differ.

Braver et al. (1985) remark that comparison to the ‘concurrent exposure data for the study plant were abstracted from the records of a local health department. The usual air concentration of hexavalent chromium was estimated as 413 micrograms per cubic meter ($\mu\text{g}/\text{m}^3$) during 1945-1949.’ This is thus a proxy for actual exposure, which was never measured. Also, reliable cigarette smoking data were not available for the study cohort members. The observed increase in SMRs with increasing duration of exposure suggests, according to the authors a dose-response relationship for chromium-induced cancer, or increased age or more inhaled tobacco smoke.

Mancuso et al. (1997) studied men hired between 1931-1937 working at a chromate plant. Nothing is mentioned on smoking and no ‘correction’ is proposed or carried out. Also, all employees were all pre-war hired. Living conditions were vastly different than they are

now. Also, they lived through the war, and we have no idea under what conditions they worked. This study cannot give us any insight into the risks of Cr VI.

Gibb et al. (2000) based their study on Hayes et al. (1979). Exposure of Cr VI was not determined but a proxy was given in an estimation for the duration of their employment at the chromate production facility, job titles and approximately 70,000 contemporary measurements of airborne hexavalent chromium concentration spanning the study period. However, the uncertainty of this exposure estimation was *not* carried through (see e.g. table VI). The proposed dose-response could also be explained as lung cancer – age. This should be considered as smoking habits were high. The vast majority of the lung cancer cases occurred among cigarette smokers (116) vs. non-smokers (4).

Luippold et al. (2003) propose to have found a threshold effect, as “risk is significantly increased only at exposure levels over 1.05 mg/m³-years”. Smoking information was scantily available (35%). The exposure assessment was based on some 800 air sampling measurements from 21 industrial hygiene surveys describing airborne concentrations of speciated Cr(VI), encompassing the years 1943 to 1971. So, no actual exposure measurements are on hand. Most lung cancers are found in the oldest pre-war birth groups. Again, those people have vastly different lives than we do nowadays.

Luippold et al. (2005) implicitly state that Cr VI is actually healthy for human beings. All standardized mortality ratios (SMR) are in favour of CR VI being healthy. *That is, if you take significance (p-values) and CIs seriously, that would be the conclusion.* That shows the weakness of all other papers, which *do* take significance and CIs seriously as indicative of lung cancer and CR VI exposure.

In *Birk et al. (2006)*, table 2 summarises the findings of the authors. All causes of death for these workers had a 95%CI < 1, which basically points at the fact that these workers were healthier than everybody else. This holds even for ischemic heart disease and accidents. This could be artefacts of the parameter or it really could be that chromium exposure of some undefined level is healthy. The good part of this paper they made a good proxy of exposure, which is measuring urine-samples for chromium levels as a biomarker for Cr VI exposure.

References

- Braver, E. R., Infante, P., & Chu, K. (1985). An analysis of lung cancer risk from exposure to hexavalent chromium. *Teratogenesis, carcinogenesis, and mutagenesis*, 5(5), 365-378.
- Gibb, H. J., Lees, P. S., Pinsky, P. F., & Rooney, B. C. (2000). Lung cancer among workers in chromium chemical production. *American journal of industrial medicine*, 38(2), 115-126.
- Luippold, R. S., Mundt, K. A., Austin, R. P., Liebig, E., Panko, J., Crump, C., ... & Proctor, D. (2003). Lung cancer mortality among chromate production workers. *Occupational and environmental medicine*, 60(6), 451-457.

Luippold, R. S., Mundt, K. A., Dell, L. D., & Birk, T. (2005). Low-level hexavalent chromium exposure and rate of mortality among US chromate production employees. *Journal of occupational and environmental medicine*, 47(4), 381-385.

Bijlage 3: Reviewing ‘Occupational exposure during iron and steel founding’ by the Health Council of the Netherlands

In deze bijlage geven we een review van het advies van de GR over de blootstelling aan de dampen die vrijkomen bij staal- en ijzergieten. Dit is een herziene versie van onze review²² van het eerdere concept waar door de GR niets significant is veranderd maar waar zij wel op gereageerd heeft in een brief bij haar definitieve advies.²³

1. Introduction

The Subcommittee on the Classification of carcinogenic substances of the Dutch Expert Committee on Occupational Safety of *The Health Council of the Netherlands*, henceforth referred to as the Subcommittee, assessed whether occupational exposure during iron and steel founding may induce genotoxic effects and may cause cancer.

The following literature was taken into account by the Subcommittee: literature on genotoxicity as in mutagenicity in vitro (8 articles); genotoxicity in humans as gene mutation essays (1 study), other genotoxic tests (3 studies), DNA adduct formation (4 studies). The Subcommittee considered 37 epidemiological studies of various kinds.

On the available data, the Subcommittee recommends ‘classifying occupational exposure during iron and steel founding as a germ cell mutagen in category 2 (“Substances which cause concern for humans owing to the possibility that they may induce heritable mutations in the germ cells of humans”). The mutagenicity is caused by a stochastic genotoxic mechanism. The committee concludes that exposure during iron and steel founding is carcinogenic to humans, and recommends classifying the exposure in category 1A (“known to have carcinogenic potential for humans”).’

2. Categorising chemicals

The Health Council of the Netherlands states in their report ‘Guideline to the classification of carcinogenic compounds’ (2010) that category 1A contain compounds that are known to be carcinogenic to man. This is defined as a ‘compound is classified as “carcinogenic to man” if there is sufficient evidence from epidemiological studies to support the existence of a causal relationship between human exposure and the development of cancer in those who were exposed to the compound in question. In addition, there is likely to be a causal relationship between exposure and effect. In some cases, a compound for which there is

²² Helsloot, I., & Hanekamp, J.C. (2019). *Reviewing ‘Occupational exposure during iron and steel founding’ by the Health Council of the Netherlands*.

²³ Gezondheidsraad (2020). *Reactie op commentaar conceptadvies Emissies tijdens ijzer- en staalgieten*.

only limited evidence from epidemiological studies to support a relationship between exposure and the development of cancer can still be placed in this category if the studies in question are complemented by sufficient evidence from animal studies to establish the existence of such a relationship.’ (p. 30)

Regulation (EC) No 1272/2008 reserves category 2 of hazard categories for germ cell mutagens, comprising of 4 categories 1, 1A, 1B, and 2, for those substances that cause concern for humans owing to the possibility that they may induce heritable mutations in the germ cells of humans. This category 2 classification based on: (i) positive evidence obtained from experiments in mammals and/or (ii) in some cases from *in vitro* experiments obtained from somatic cell mutagenicity tests *in vivo*, in mammals or other *in vivo* somatic cell genotoxicity tests which are supported by positive results from *in vitro* mutagenicity assays. Substances which are positive in *in vitro* mammalian mutagenicity assays, and which also show chemical structure activity relationship to known germ cell mutagens, shall be considered for classification as Category 2 mutagens.

3. Analysis - epidemiology

First we will reflect on the epidemiological studies discussed by the Subcommittee.

Our central observation is that all 4 meta-analyses the Subcommittee reports on do not consider smoking habits. Overall, 20 of the 37 epidemiological studies do not adjust for smoking, which makes their results difficult if not impossible to interpret. Interestingly, one of the 4 meta-analyses (again, not reporting on smoking habits) remarks that no associations were found between PAH (polycyclic aromatic hydrocarbons) exposure and lymphatic and haematopoietic neoplasms (Alicandro et al., 2016).

Looking more precise into the literature gathered by the Subcommittee, we find several inconsistencies between the interpretation by the subcommittee and what the literature actually says:

- *Gibson et al.* (1977) is reported as having found a ‘positive association with lung cancer’ for foundry workers. However, the authors reported a ‘serious drawback’ as to the ‘lack of smoking histories in the groups’. The ‘only relevant data available was a random sample of 100 foundry workers, taken in 1976. showing that 58% smoked cigarettes’, whereas ‘foundry group consisted of those who had worked at least five years in the foundry over their employment prior to 1967 (439 men)’. This is important, as a ten year difference between the sampling of smoking in 1976 and the analysis of the foundry group earlier than 1967 most likely will result in an underestimation of the influence of smoking. Smoking among men declined considerably in the second half of the 20th century (CDC, 1999).
- *Becher et al.* (1989) is reported as having found a ‘positive association for lung cancer in longest exposed worker population’. Becher et al. however, showed no such thing. They found ‘tobacco smoking ... to be the strongest risk factor, with “relative risks”

obviously increasing “with the amount of cigarette consumption’. Becher et al.’s calculated excess relative risks remain small with only >30 years of foundry work calculated to be above 2 (2.66), which is considerably smaller than the relative risks of smoking.

- *Austin et al.* (1997) is reported as having found a ‘positive association with lung cancer in workers handling materia’, as remarked by the Committee. However, Austin et al. in fact report that ‘there is a tendency toward positive findings for lung cancer among foundry workers, but the excesses are only small to moderate. The studies did not adjust for potential confounding by the single most important cause of lung cancer, cigarette smoking. Thus, some of the excess lung cancer may be attributable to an increased prevalence of cigarette use among foundry workers.’ In line therewith, Moulin et al. (2020) in their large study, reported that their ‘study failed to detect any relationship between lung cancer and exposure to iron, chromium, nickel and/or their compounds.’ The relative risk for exposure to PAH was calculated by Moulin et al. (2020) to be 1.46.²⁴

Some of the literature gathered by the Subcommittee clearly states that no significant or even positive effects of foundry work were found. Firth et al. (1999) found that ‘small increases in risk for several causes of death among foundry and heavy engineering workers; however, these increases were small and the possible effects of smoking and other lifestyle factors could not be excluded.’ Hoshuyama et al. (2006) found a healthy worker effect in their large retrospective study. They do report a positive association with lung cancer when combined PAH-exposure with dusts. However, as Hoshuyama et al. did not correct for smoking, this result is difficult to interpret, as PAH is found abundantly in cigarette smoke (see e.g. Vu et al., 2015).

In our view therefore the Subcommittee rightly reports on the 26 cohort studies that the ‘majority of the studies did not find associations between the cumulative exposure and cancer mortality.’ (p. 20). Moreover, ‘in most cohort studies data on smoking habits were not collected or reported’. (p. 19) Also, actual exposure to workers was never measured in the considered studies; even data on exposure levels to which the workers might have been exposed is missing in most studies.

The Subcommittee reports on the 7 case-control studies that ‘workers’ exposure is assessed by job titles, work areas and duration of employment, rather than by measuring exposure levels.’ (p. 20) This is misleading, as in none of the studies considered actual exposure is measured. Moreover, in 2 studies ‘no associations were found for lung cancer or bladder cancer and working in iron and steel foundries.’ (p. 20)

Overviewing the Subcommittee’s work on these studies, it is surprising that no formal risk assessment framework is offered. No methodology is utilised other than a summarising of

²⁴ Moulin, J. J. Clavel, T. Roy, D. Dananché, B. Marquis, N. Févotte J. & Fontana J. M. (2020). Risk of lung cancer in workers producing stainless steel and metallic alloys, *International Archives of Occupational and Environmental Health* 73(171–180).

studies which, as we have shown, in some cases lack precision, followed by an expert judgement that, overall, these studies show a genotoxic effect. The scientific validity of this inference is questionable.

We observe one other flaw in all studies. In none of the studies, exposure to workers is actually measured, making reported associations inherently weak. This flaw is the more serious because the studies therefore give no insight in the effects of the present-day exposure of Dutch foundry workers that is arguably lower than the exposure to all kind of substances decades ago.

Overall, we would like to point to the fact that epidemiological data are fraught with imprecisions and 'controlling' for covariates is notoriously difficult and always incomplete. In this specific case, as said, smoking, as an important variable, is in the majority of studies not taken into consideration. Generally, epidemiological reported associations can never be considered as proof of causation (Briggs, 2016). This general principle alone stands in stark contrast with the Committee's decision to rank 'exposure during iron and steel founding' as 'known to have carcinogenic potential for humans' (p. 7). This decision clearly cannot be defended based on the studies considered.

4. Analysis – animal and cell studies

Here we will reflect on the cell and animal studies discussed by the Subcommittee. Humfrey et al. (1996), the only known animal study, carried out an animal experiment on the carcinogenicity of extracts of aerosols collected from the emission of iron and steel founding. The Subcommittee, however, noted that the chosen exposure route is irrelevant for the human situation.

On the subject of genotoxicity (Annex B), the Subcommittee summarises 8 *in vitro* mutagenicity studies. All are Ames tests. Four of these 8 studies reported no data on cytotoxicity, which makes these studies difficult to interpret. One study, not part of the 4 previously mentioned, used urinary samples from Canadian steel foundry workers to test for reverse mutations (Ames). However, the results focus on smoking habits, making the results difficult to interpret. Humfrey et al. (1996) does not present results on positive and negative controls and reports no data on their statistical analysis. Overall, all 8 studies have methodological or reporting flaws that make them less than suitable for interpretation of their data.

On the subject of other genotoxicity testing in humans (Annex C), Perera et al. (1994), for instance, studied benzo[a]pyrene (B[a]P) exposure in workers in an iron foundry and DNA mutations. B[a]P is a well-known carcinogen from the group of PAHs found in foods and cigarette smoke. No significant mutations were found. Kubiak et al. (1999) compared frequencies of micronuclei (MN; a biomarker of chromosome breakage and/or whole chromosome loss) in cytokinesis-blocked lymphocytes between high- and low-exposure

steel foundry workers and found no significant difference. Two other studies generally lacked data to interpret the results properly.

On DNA-adduct formation, 4 papers are summarised, on which the Committee remarks that “moderate to clear increases in DNA-adduct formation was observed, in two of these studies reaching statistical significance.” (p. 17) However, all these studies are small, that is a small numbers of workers were sampled for blood cells. Two studies did not include a reference group, and showed considerable individual variability. Also, the Committee did not consider smoking as a source of PAH-adducts beyond these 4 studies. For instance, Reddy et al. (1991) in Table 1 show no difference between DNA-adduct of smokers and non-smokers, which seems unlikely (see e.g. Godschalk et al., 1998). Also, DAN-adduct formation is not as such indicative of mutagenicity. Kobets and Williams (2019) remark that “not every chemical adduct produced in DNA is potentially miscoding, since some are formed at sites on bases that are not involved in base pairing. Also, not every interaction with DNA involves a susceptible site for mutation in critical genes. ... Additionally, adduct levels at 1 in 10^9 nucleotides, which represents about 12 adducts per somatic cell, are extremely small compared to that of endogenous DNA lesions per cell, estimated to be in the order of 10^4 to 10^6 per cell, much of which is oxidative damage resulting from cellular metabolism, e.g., ROS.”

Overviewing the Committee’s work on animal and cell studies, it is, again, surprising that no formal risk assessment is offered. No methodology is utilised here other than a summarising of studies followed by an expert judgement.

It might therefore come as a surprise that the Committee ‘recommends classifying exposure during iron and steel founding as a germ cell mutagen in category 2 (“Substances which cause concern for humans owing to the possibility that they may induce heritable mutations in the germ cells of humans”).’ This in fact is the lowest (weakest) class in the ‘Hazard categories for germ cell mutagens’ as found in Regulation (EC) No 1272/2008 (p. 101). Parenthetically, this hazard classification does not contain a category for chemicals for which insufficient evidence is available (see The Health Council of the Netherlands, 2010, p. 29, 30) such as is the case here, we would argue. The recommended hazard classification 2 does not amount to an actual understanding of the risks involved (see e.g. Boobis et al., 2016).

5. Conclusions

The above leaves no other conclusion then that the draft-study of the Subcommittee on the Classification of carcinogenic substances of the Dutch Expert Committee on Occupational Safety of *The Health Council of the Netherlands* that occupational exposure during iron and steel founding may induce genotoxic effects is seriously flawed in its assessment.

We find that this assessment is inadequate as it is to a large extent based on epidemiological studies that generally did not take into account smoking habits, which is

known for its carcinogenic and genotoxic effects related to many different chemicals found in tobacco smoke. In the same vein, no analytical frame work is presented that explains why the Subcommittee does not come the conclusion from the cell and animal studies that no hard indication is present that the exposure induces heritable mutations in the germ cells of humans.

We propose that new epidemiological studies should be undertaken that relate *actual measured exposure levels of foundry workers* and genotoxic and mutagenic effects, under the condition that new, carefully formulated, *in vitro* studies bear out inducible heritable mutations.

References

- Boobis, A. R., Cohen, S. M., Dellarco, V. L., Doe, J. E., Fenner-Crisp, P. A., Moretto, A., Pastoor, T. P., Schoeny, R. S., Seed, J. G. & Wolf, D. C. (2016). Classification schemes for carcinogenicity based on hazard- identification have become outmoded and serve neither science nor society. *Regulatory Toxicology and Pharmacology*, 82, 158–166.
- Gibson, F., Cox, G., Downie, J., & Radik, J. (1977). A mutation affecting a second component of the F0 portion of the magnesium ion-stimulated adenosine triphosphatase of *Escherichia coli* K12. The uncC424 allele. *Biochemical Journal*, 164(1), 193-198.
- Firth, H. M., Elwood, J. M., Cox, B., & Herbison, G. P. (1999). Historical cohort study of a New Zealand foundry and heavy engineering plant. *Occupational and environmental medicine*, 56(2), 134-138.
- Hoshuyama, T., Pan, G., Tanaka, C., Feng, Y., Liu, T., Liu, L., ... & Takahashi, K. (2006). Mortality of iron-steel workers in Anshanl China: A retrospective cohort study. *International journal of occupational and environmental health*, 12(3), 193-202.
- Humfrey, C. D. N., Levy, L. S., & Faux, S. P. (1996). Potential carcinogenicity of foundry fumes: a comparative in vivo-in vitro study. *Food and Chemical Toxicology*, 34(11-12), 1103-1111.
- Kobets, T., & Williams, G. M. (2019). Review of the evidence for thresholds for DNA-Reactive and epigenetic experimental chemical carcinogens. *Chemico-biological interactions*, 301, 88-111.