ESPUM
L'ÉCOLE DE SANTÉ PUBLIQUE
DE L'UNIVERSITÉ DE MONTRÉAL

MODERN INDUSTRIAL HYGIENE STATISTICS

Jérôme Lavoué

June 9th, 2022

Université de Montréal et du monde.

COPYRIGHT

Université de Montréal

ESPUM

Jérôme Lavoué

- Engineer (industrial chemistry)
- M.Sc Toxicology
- Ph.D occupational hygiene
- Professor at University of Montreal
- Research in exposure assessment sciences

Université de Montréal

ESPUM

Content overview

1. Exposure variability and the lognormal distribution
2. Challenges in uncertainty management
3. The European approach
4. The contribution of Bayesian statistics

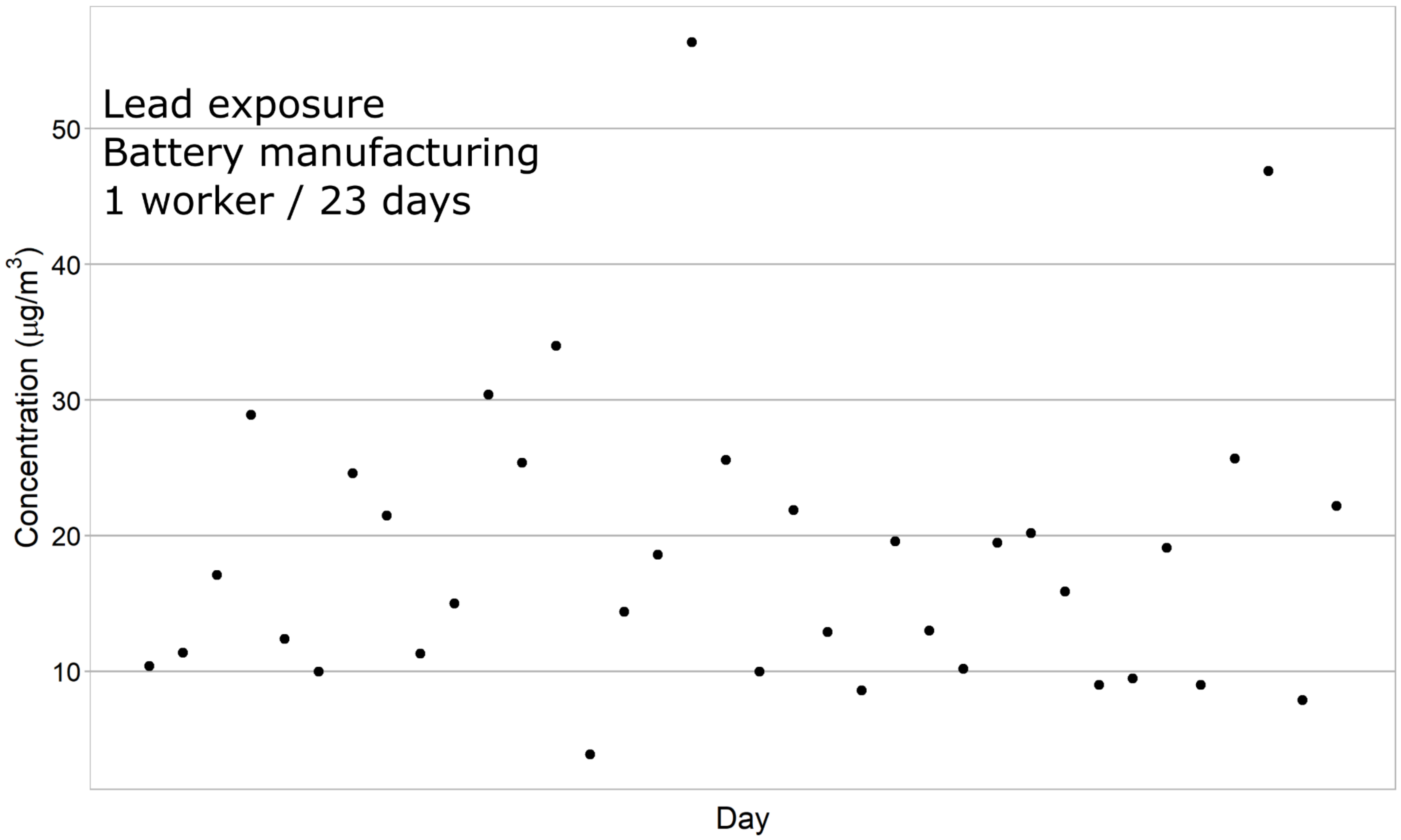
EXPOSURE VARIABILITY IN THE WORPLACE

«Never measure a second time, because you will find something different…»

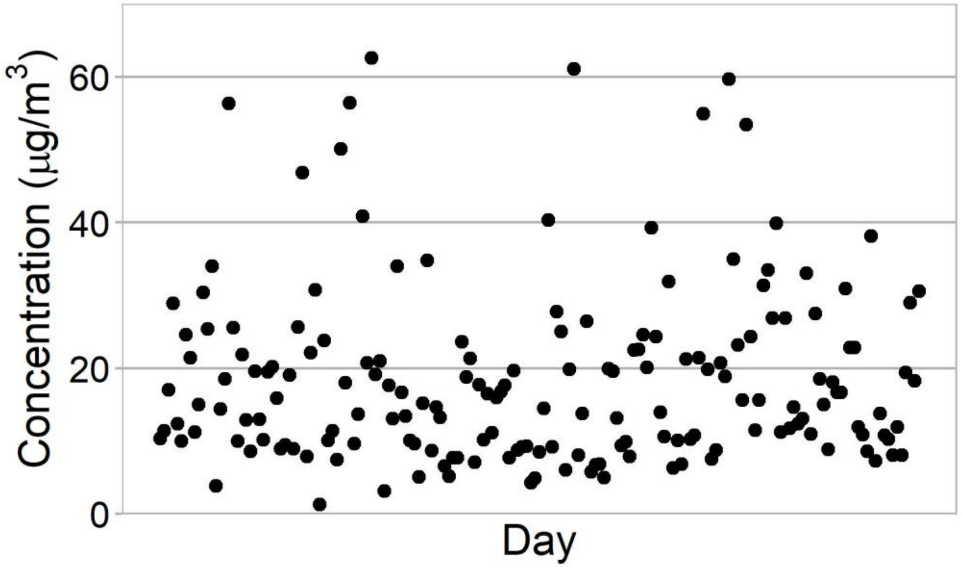
HÄKEN WESTBERG, 2001

X2001 conference, Gothenburg, Sweden

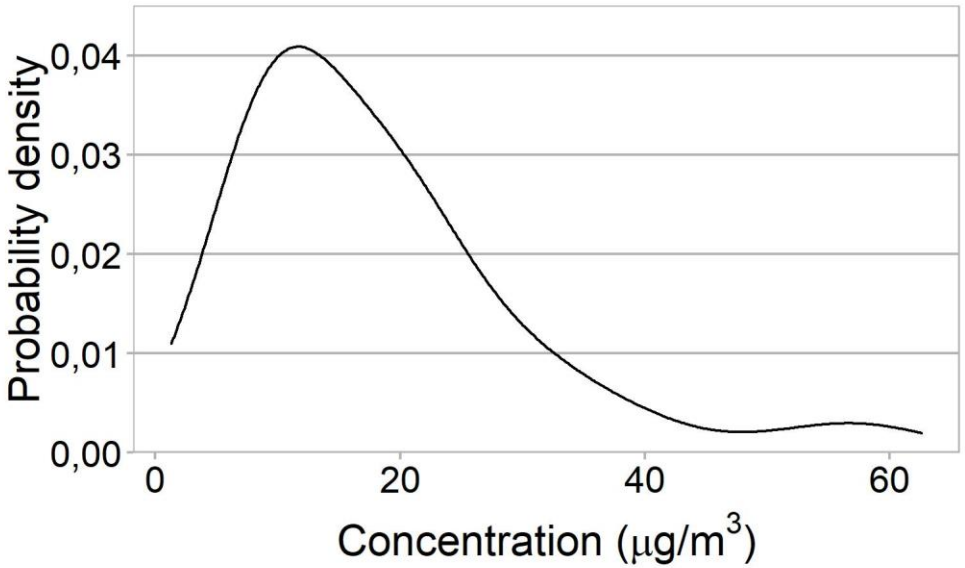
Occupational exposure varies across time and space



A MODEL FOR EXPOSURE VARIABILITY : THE LOGNORMAL DISTRIBUTION

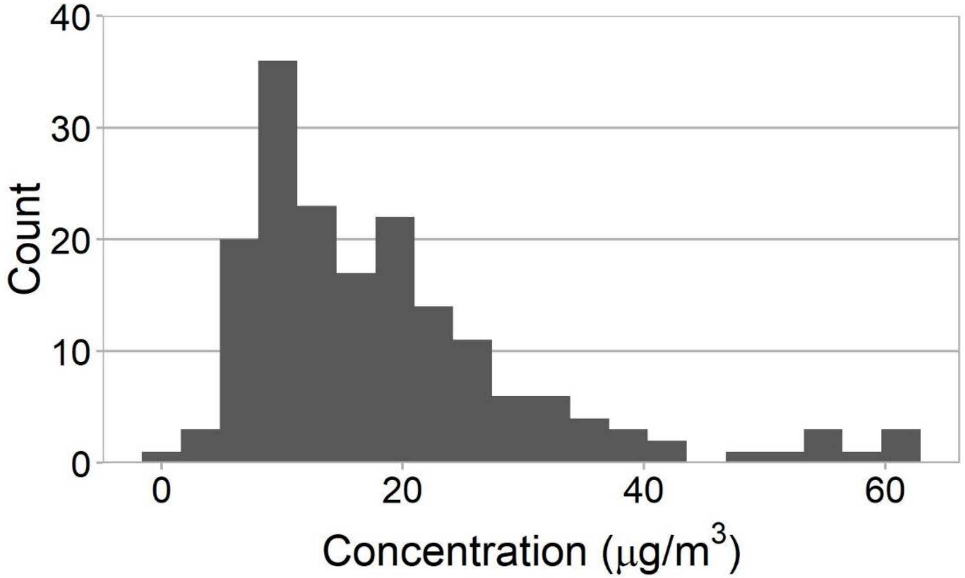


Simple scatterplot



Probability density curve

177 lead exposure measurements on six workers performing the same tasks for one year in the gasoline additive manufacturing industry



Histogram

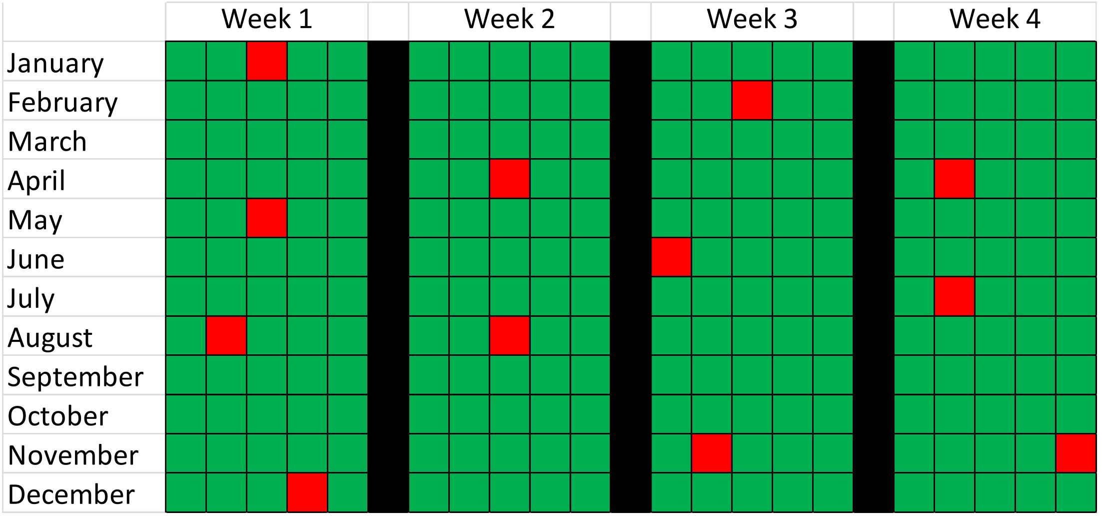
Université de Montréal

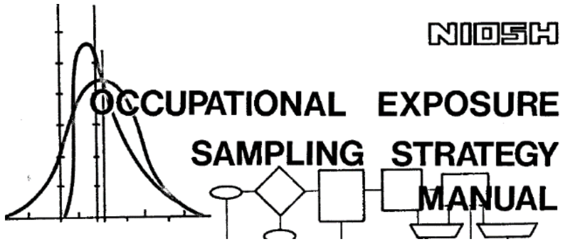
ESPUM

WHICH PARAMETER OF THE LOGNORMAL FOR RISK ASSESSMENT ? – CURRENT CONSENSUS

It is impossible to demonstrate exceedance fraction=0 without measuring every day-worker.

In practice : We'll try to demonstrate that exceedance fraction is <5%

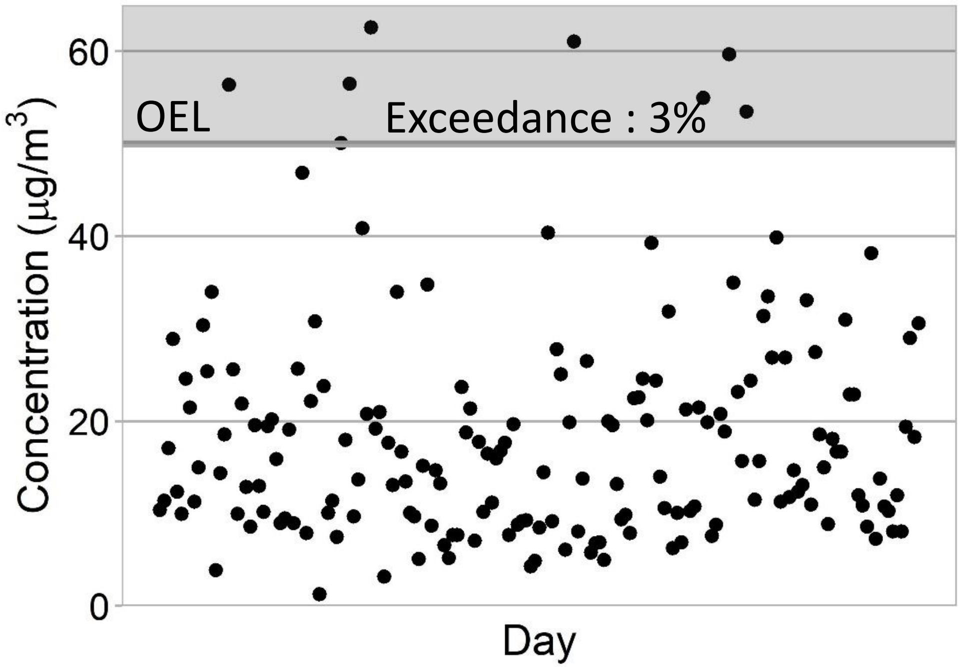




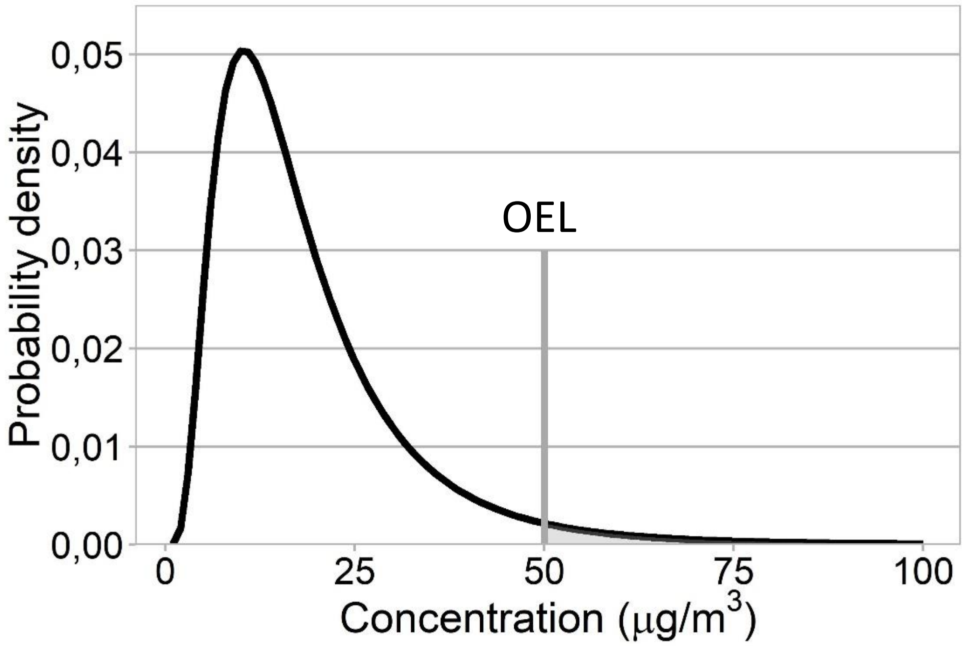
EXCEEDANCE FRACTION

Overexposure : exceedance fraction (F) ≥ 5%

Sequential plot



Density curve

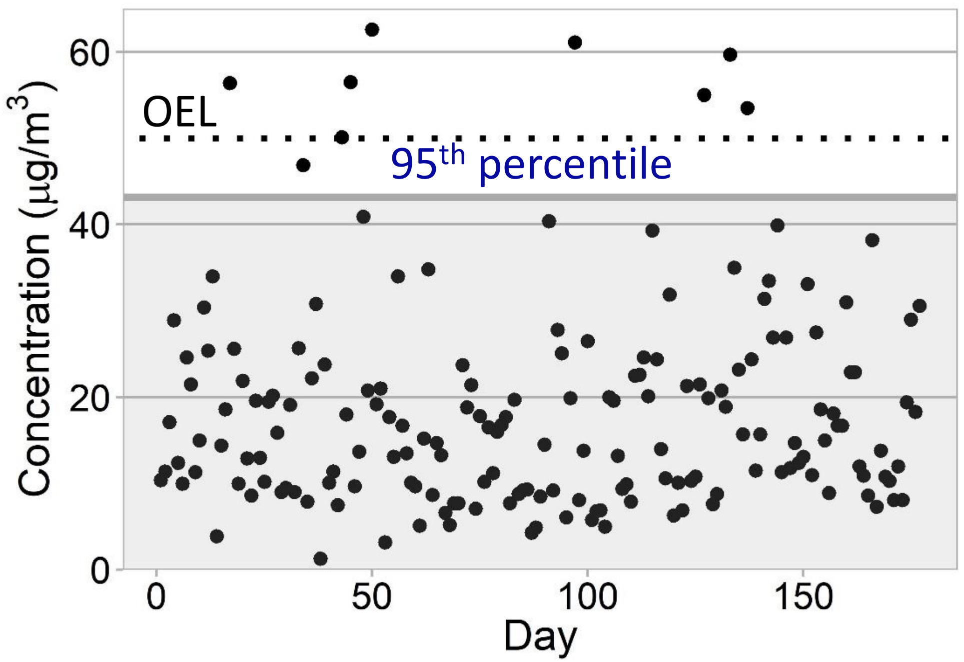


Lead exposure measurements in the battery manufacturing industry (n=177)

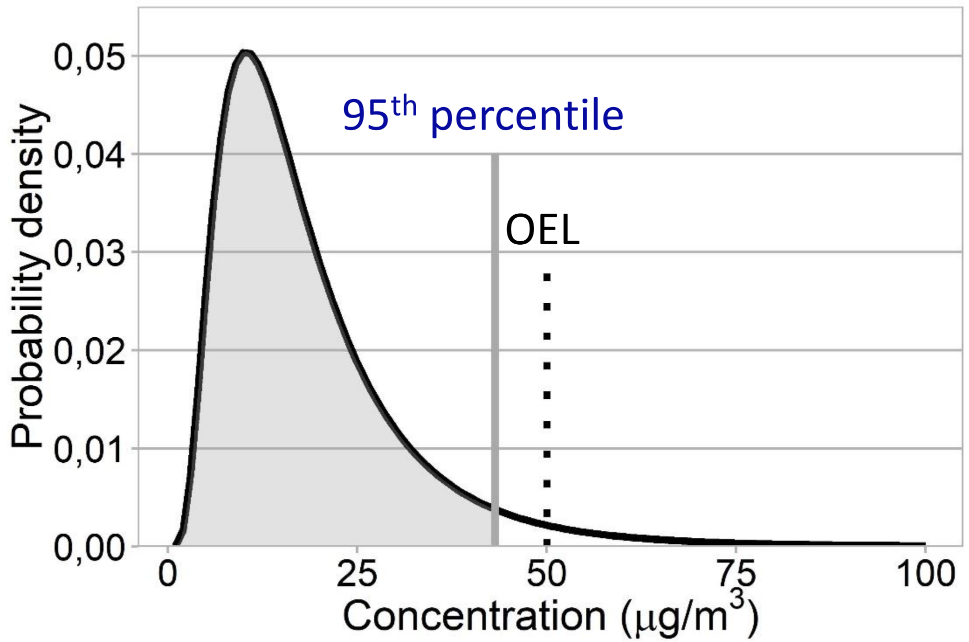
95TH PERCENTILE

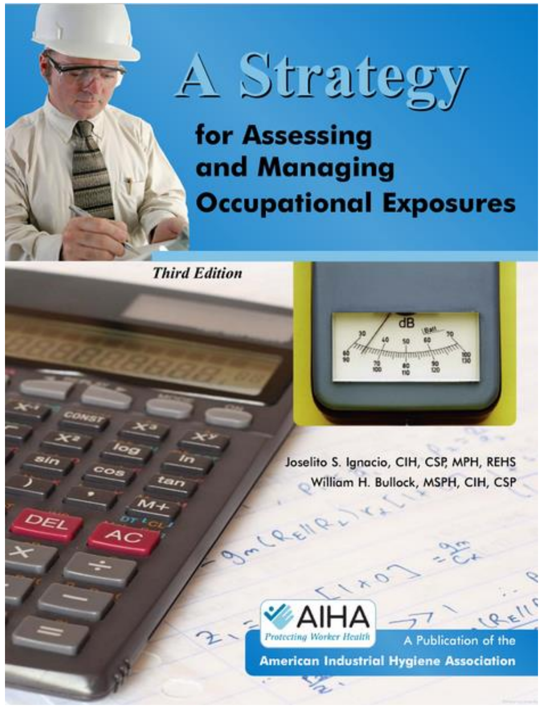
Overexposure : 95th percentile > OEL

Sequential graph



Density curve





Lead exposure measurements in the battery manufacturing industry (n=177)

Université de Montréal

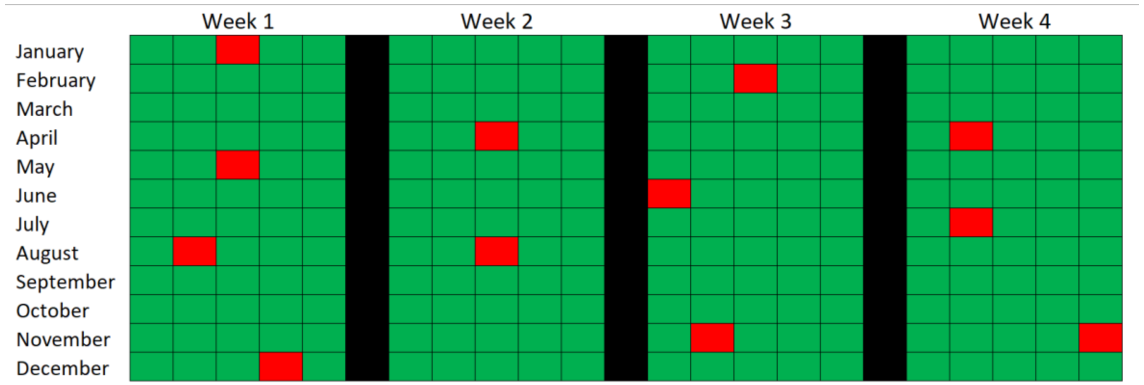
ESPUM

Challenges in uncertainty management

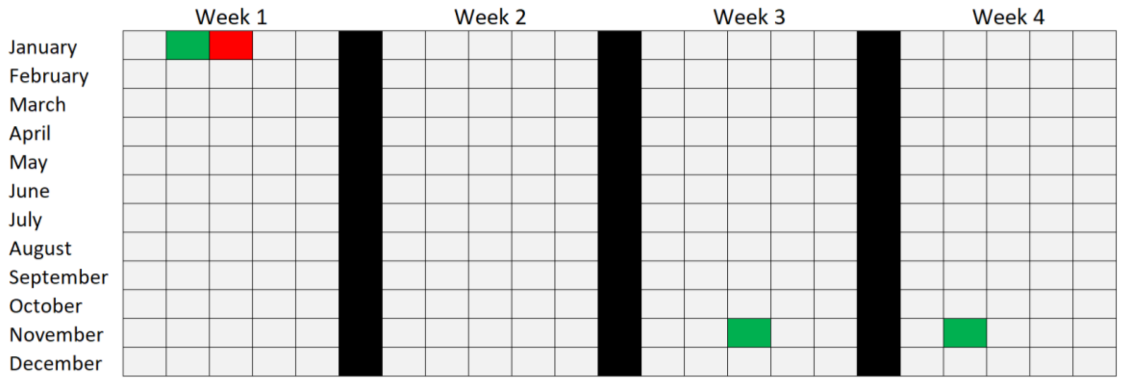
«Never estimate the 95th percentile a second time, because you will find something different…»

WHY ARE OUR ESTIMATES UNCERTAIN ?

We are interested in this

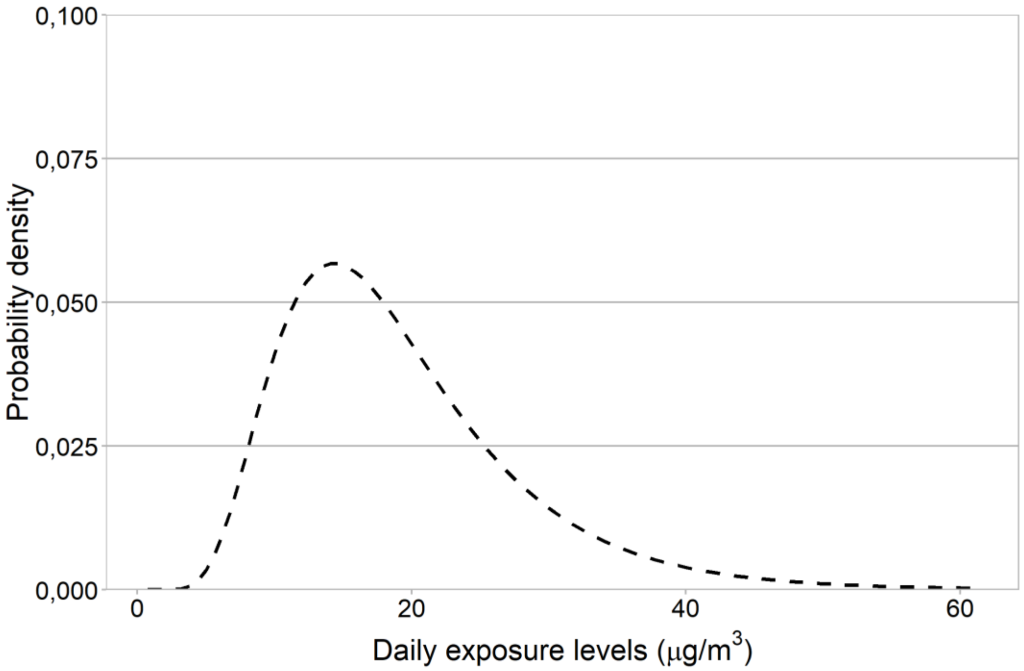


We have this

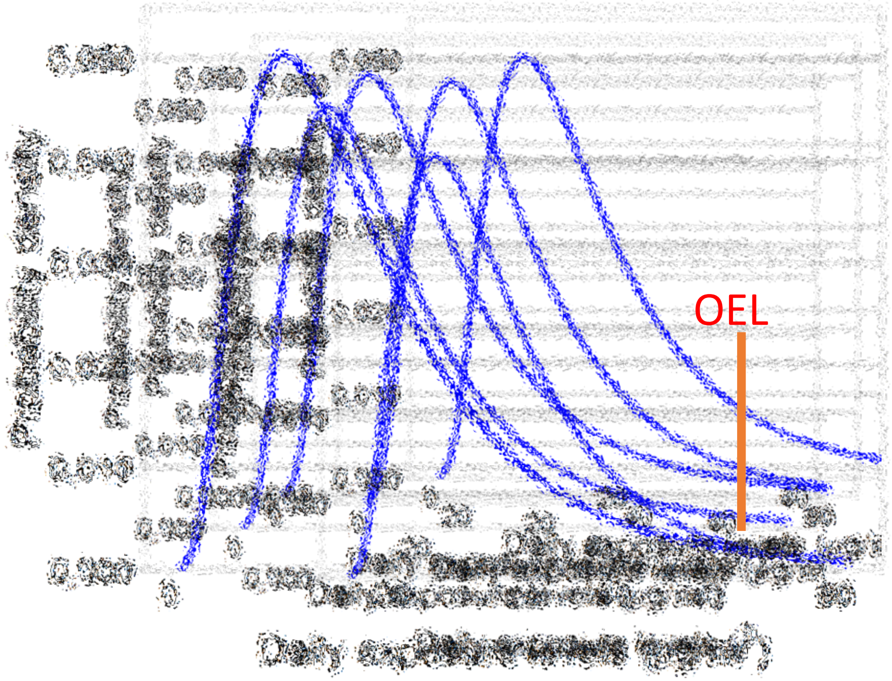


Exceedance fraction in the target population ≠ Exceedance fraction in my sample

WHAT DOES MY OWN SINGLE SURVEY TELL ME ABOUT THE TARGET POPULATION ?



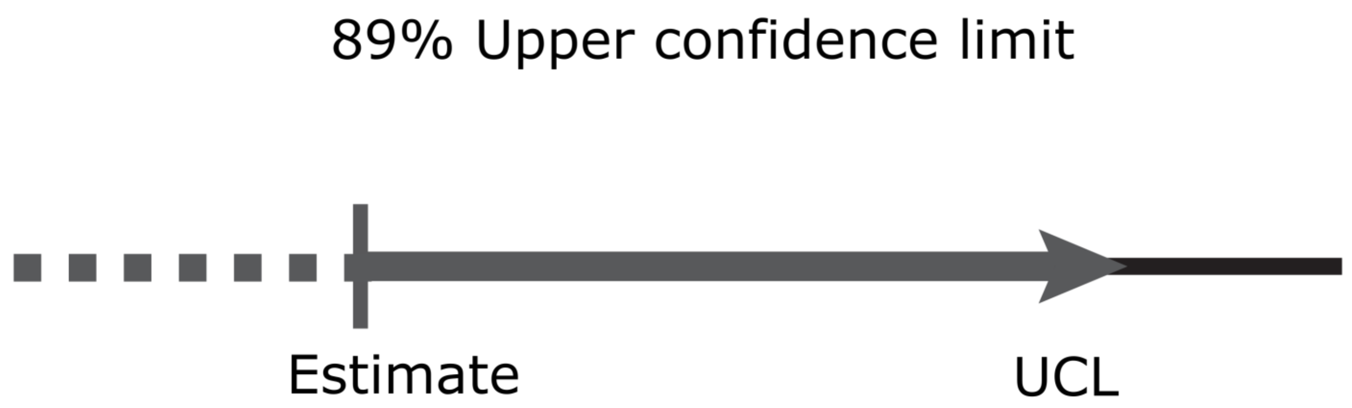
Distribution estimated from my sample GM and GSD



Distribution of exposures in the target population

Traditional approach in IH: Confidence limits

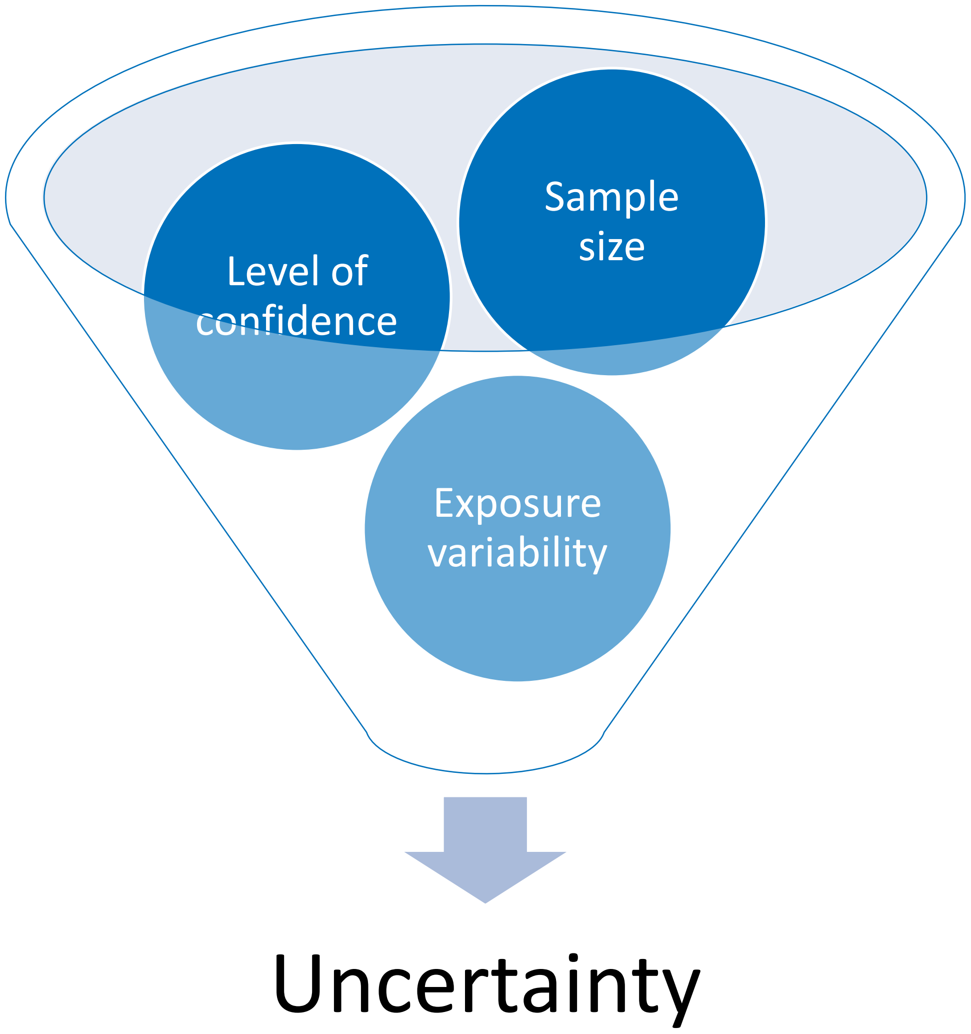
In occupational hygiene, we are more interested in the right (upper) side of the confidence interval (we want to be sure that we are under the OEL). This is called the upper confidence limit (or bound) on the parameter of interest.

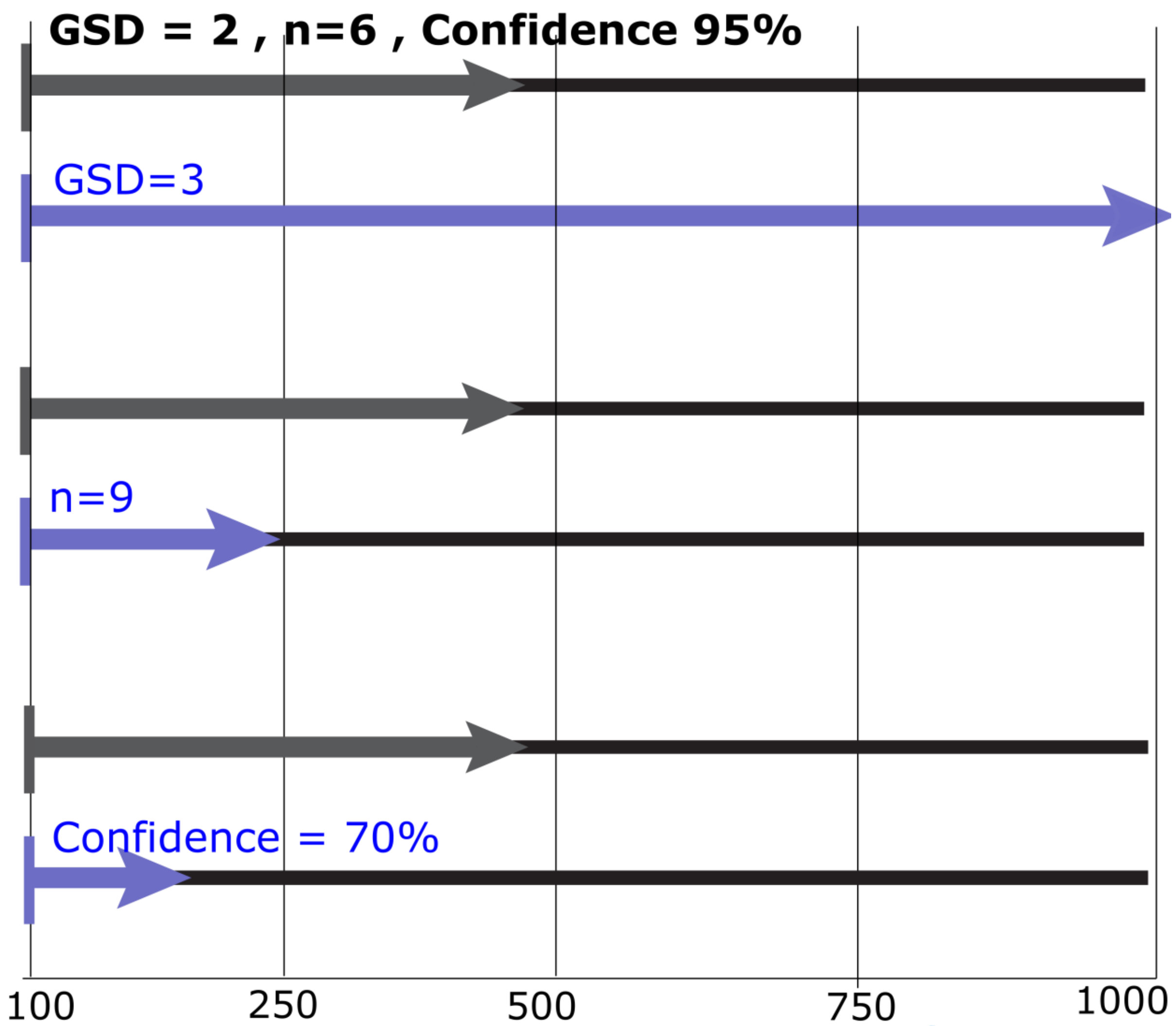


Pragmatic interpretation example : After my study, I am **89%** sure that the 95th percentile of exposure levels in **my target population** is **below** UCL.

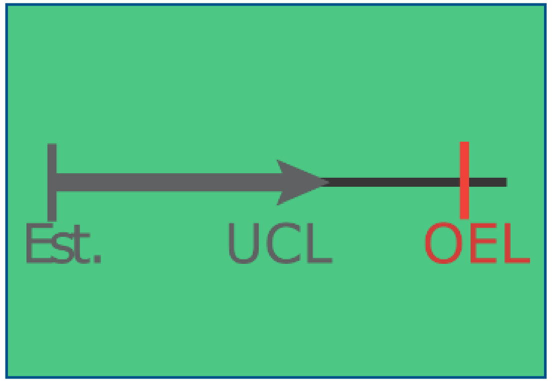
If the UCL is below the OEL : I am 89% sure that there is no overexposure

What drives the distance from the point estimate to the upper confidence limit



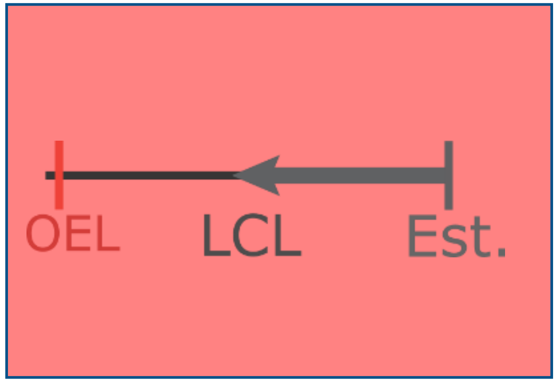


Possible outcomes of an analysis (example with P95 and 95% upper confidence limit)



Situation is acceptable (statistical demonstration)

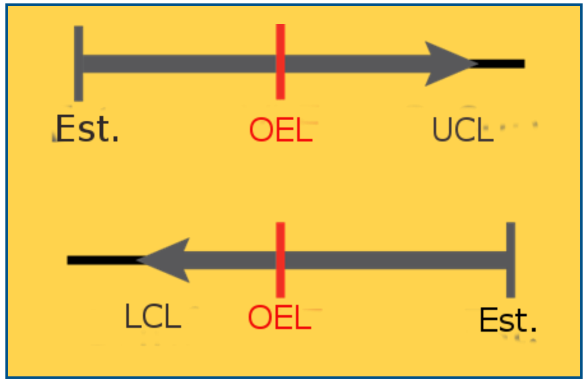
We are 95% sure that there is no overexposure



Situation is unacceptable (statistical demonstration)

We are 95% sure that there is overexposure

Possible outcomes of an analysis



Acceptable ?

Unacceptable ?

We are unsure whether there is overexposure or not. We don't know the probability of overexposure.

Depending on sample size and true exposure variability, when exposure levels are ~ 0.1-0.3*OEL, the orange situation can be the most likely outcome and it is impossible to conclude with any stated certainty.

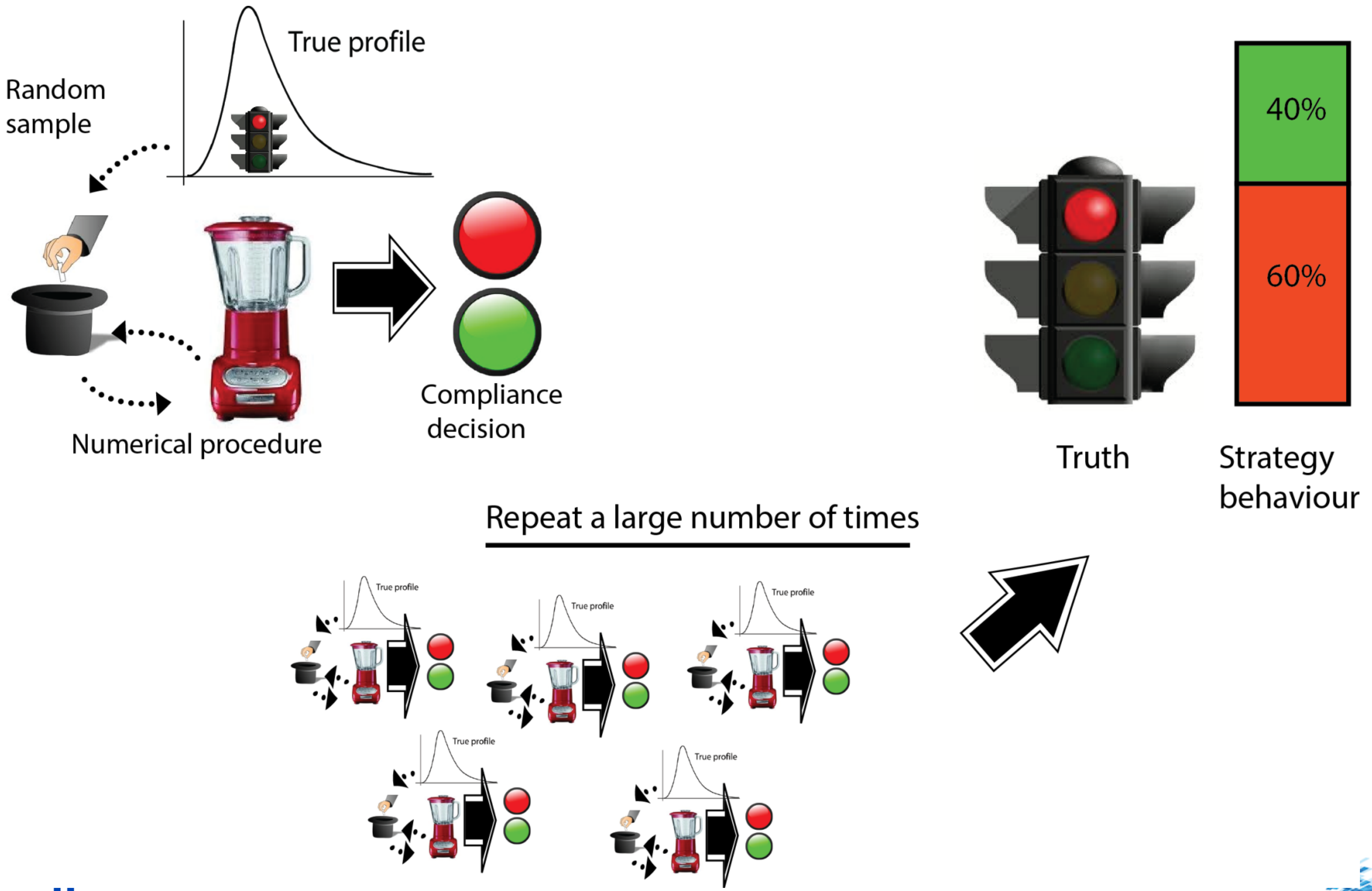
If increasing sample size is not possible, orange can be equaled to red, or to green, to allow making a decision, but at what cost ?

The European approach

Managing uncertainty for decision making : The « French » compromise

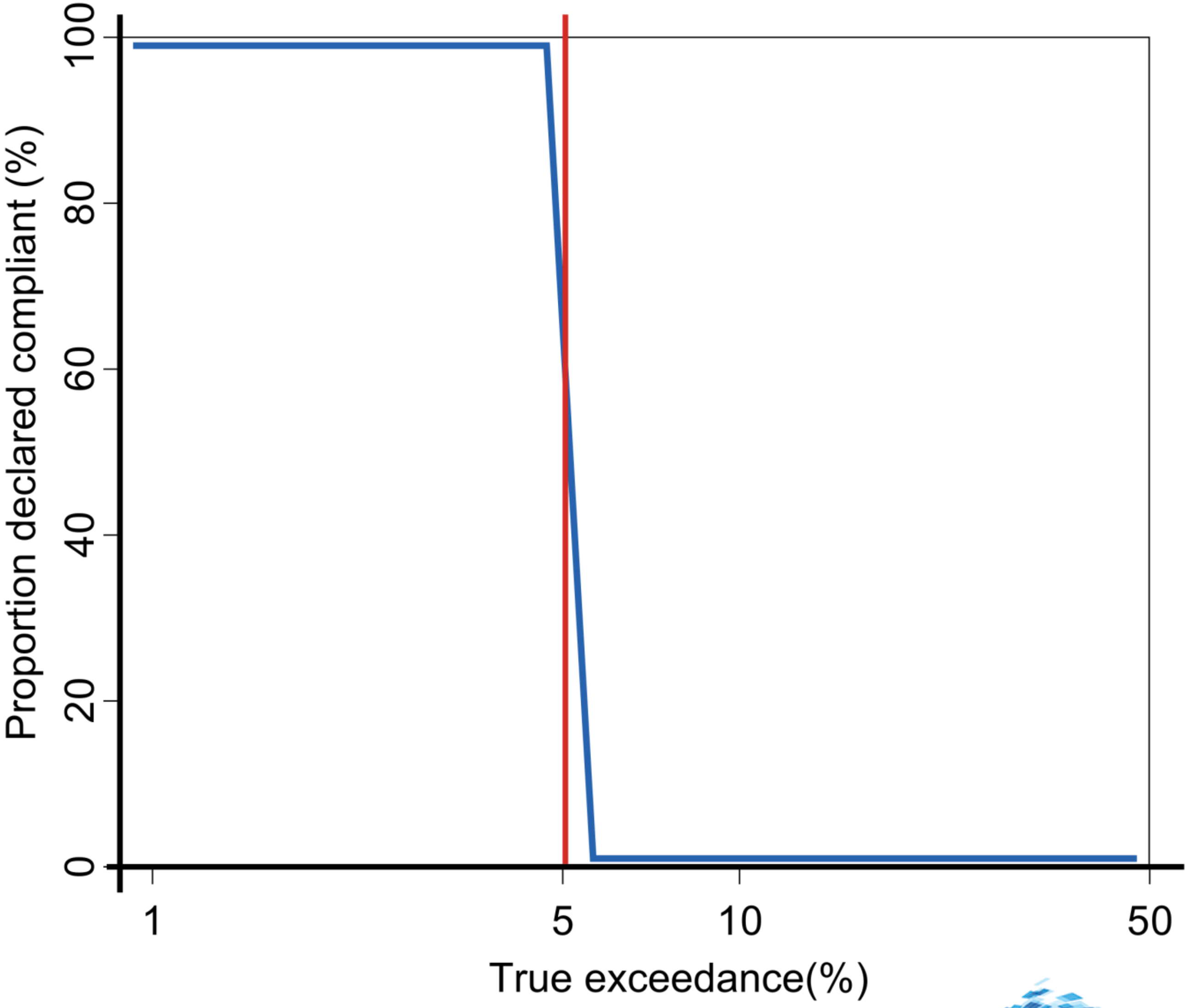
- Requesting 95% confidence that the 95th percentile is <OEL or that exceedance fraction is <5% is too stringent given realistic sample sizes.
- Sample size will probably not increase significantly in the near future, despite it most certainly should.
- First proposed by France, then adopted by the British and Dutch occupational societies and the recent European guideline, a pragmatic proposal is to relax the statistical stringency by using 70% confidence instead of 95%.
- This more lenient criterion will lead to more diagnosis errors, but they will be shared equally between not protecting the workers / inducing futile prevention costs.

Some simulations to shed light on the « French » compromise



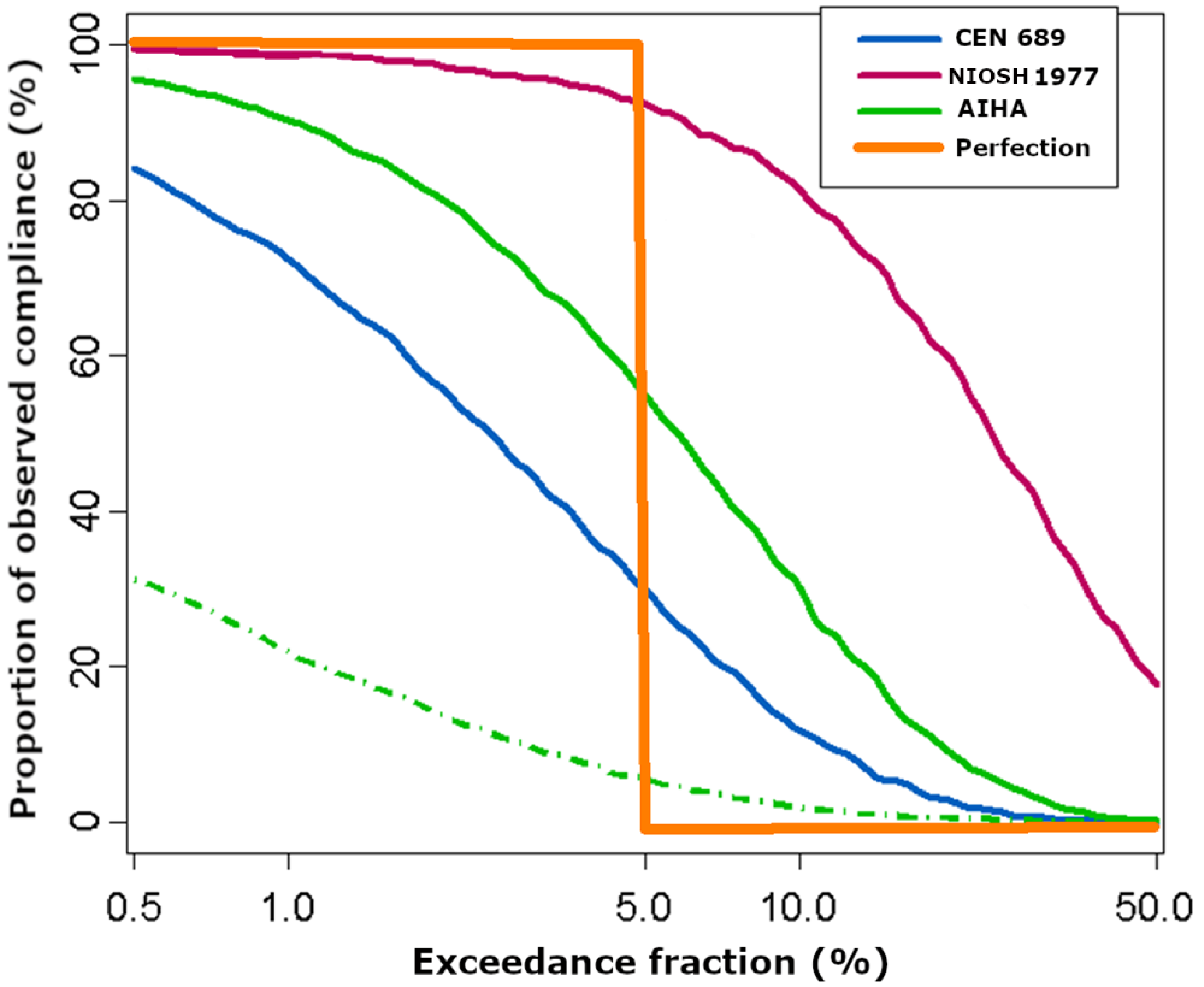
Some simulations to shed light on the « French » compromise

The ideal performance curve



Some simulations to shed light on the « French » compromise

Real performance curves



Université de Montréal

THE FRENCH REGULATION

Arrêté du 15 décembre 2009 relatif aux contrôles techniques des valeurs limites d'exposition professionnelle sur les lieux de travail et aux conditions d'accréditation des organismes chargés des contrôles

Evaluer de façon représentative l'exposition professionnelle aux polluants est une tâche difficile. Les procédés et produits industriels sont très nombreux. Chaque phase de fabrication peut correspondre à des rythmes de production différents, faire intervenir une grande variété d'agents chimiques et donc présenter des conditions d'exposition spécifiques. La distance du poste de travail par rapport aux sources d'émission et les paramètres tels que l'intensité de l'émission, la ventilation, les variations météorologiques et saisonnières peuvent avoir également une influence très marquée. Cette variabilité spatiale et temporelle des conditions d'exposition est encore renforcée par celle des pratiques individuelles et du geste professionnel lui-même.

C'est pourquoi, quelques mesurages réalisés sur un seul jour ou dans une période trop restreinte fournissent un aperçu insuffisant de la variabilité réelle de l'exposition individuelle. Il apparaît indispensable d'établir une

A l'issue des trois campagnes de mesures, le diagnostic de dépassement de la VLEP 8 heures est établi à partir de l'analyse statistique de l'ensemble des mesures d'exposition réalisées : au minimum neuf par GEH. Le diagnostic de dépassement de la VLEP 8 heures est établi lorsque, sous hypothèse d'une distribution log-normale des expositions, la borne supérieure de l'intervalle de confiance à 70 % de la probabilité de dépassement de la VLEP 8 heures est supérieure à 5 % (Pr [IC 70 %] >5 %).

MANAGING UNCERTAINTY : 70% UCL VS 95% UCL

95% UCL < OEL – **acceptable**

70% UCL < OEL – **tolerable**, **assuming the SEG has a required monitoring plan**

70% UCL > OEL– **problematic**, particularly if the SEG has no monitoring plan.

Université de Montréal

ESPUM

The contribution o f Bayesian statistics

Bayesian statistics in one slide

A certain quantity is of interest : e.g. the proportion of exposures above the OEL in a group of workers

The « I know nothing » option leads to results similar to classical stats

What you know / think in advance

+

What the data says

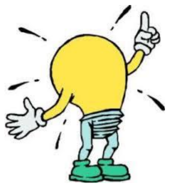
→

What you know given the data and your prior belief/knowledge

[0 - 100%]

[14 - 34%]

This is very powerful, and dangerous too



[14 - 34%]

[0 - 50%]

[16 - 22%]

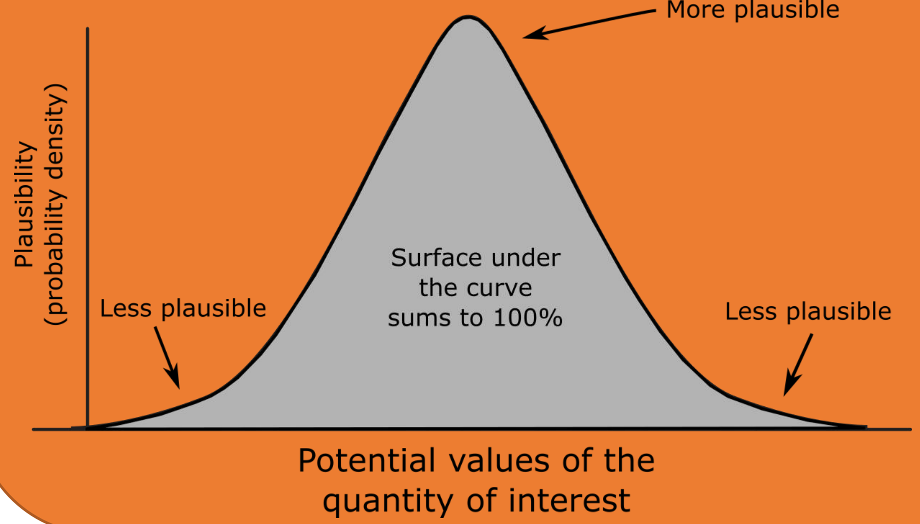
Université de Montréal

ESPUM

BAYESIAN INFERENCE : GENERAL PRINCIPLE

Prior uncertainty distribution about a parameter p

What we know before data



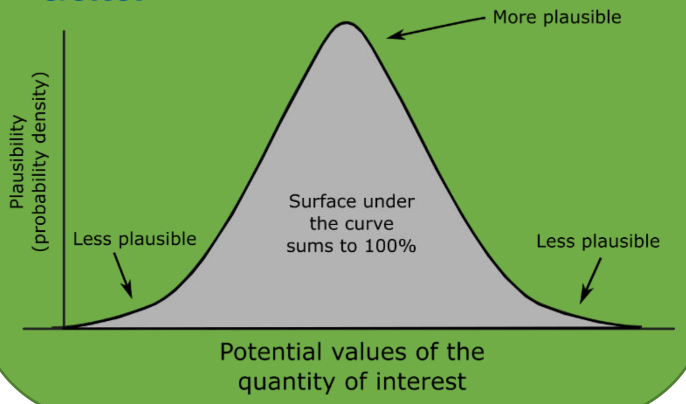
DATA

Likelihood function for P

given a statistical model, measures the likelihood of observing the data across the universe of parameter values selected in the prior

Posterior uncertainty distribution about P

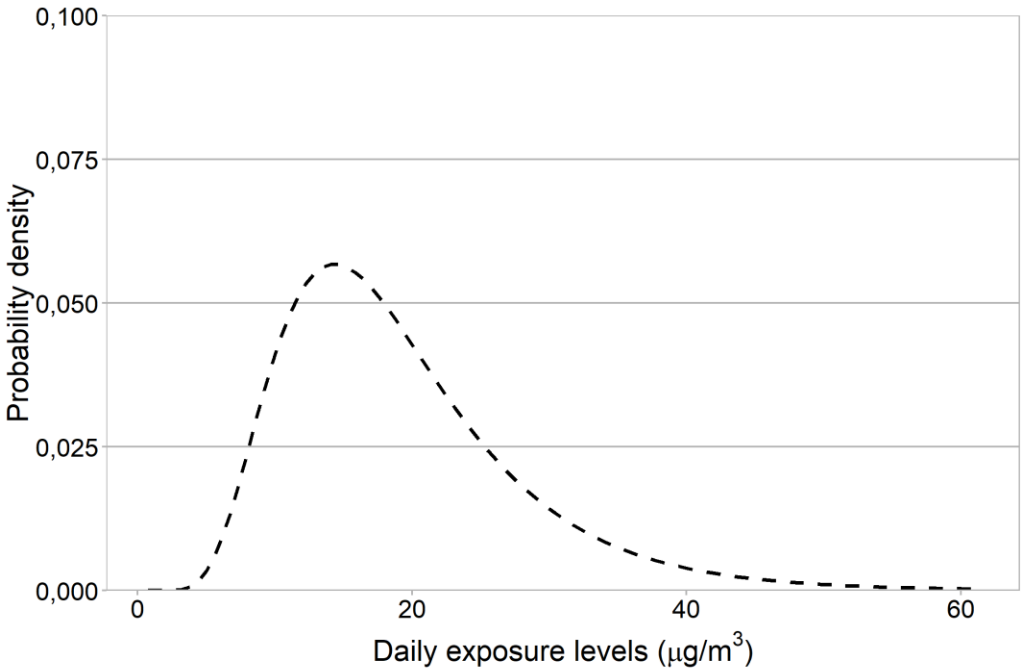
What we know after data



Université de Montréal

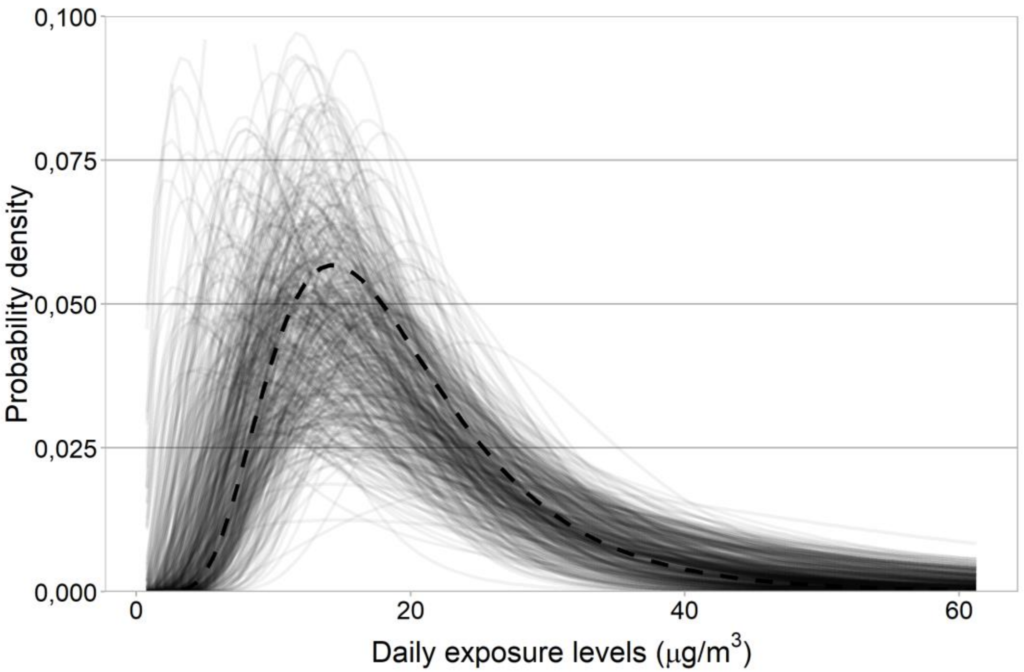
ESPUM

COMING BACK TO EXTRAPOLATING MY SINGLE SURVEY



Distribution estimated from my sample GM and GSD

Bayesian stats



1000 plausible true distributions

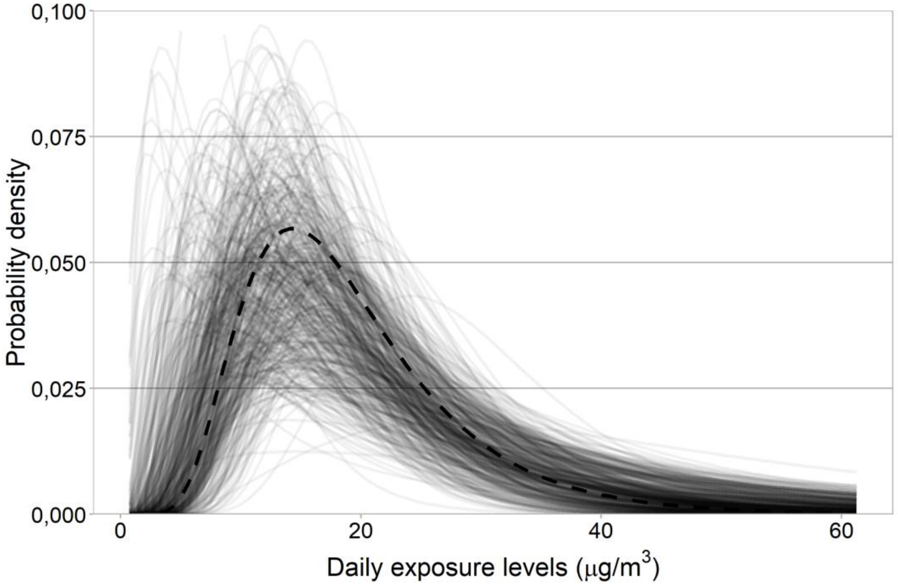
Based on assumptions (lognormal distribution) and the study design (6 measures randomly drawn from the population), Bayesian statistics provide a way to generate a picture of the possibilities for the target population given our own observations (here 1000 possibilities illustrated)

Université de Montréal

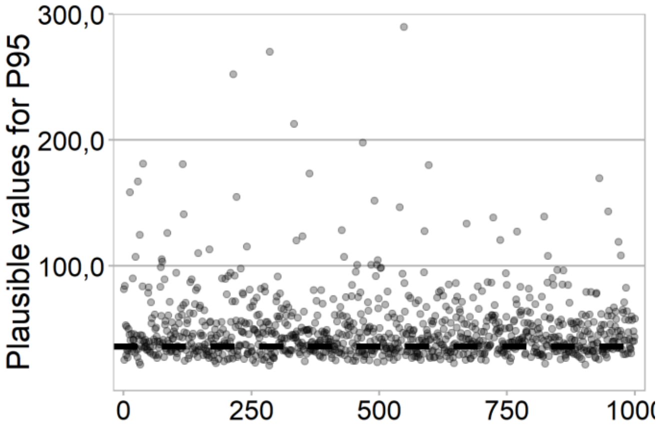
ESPUM

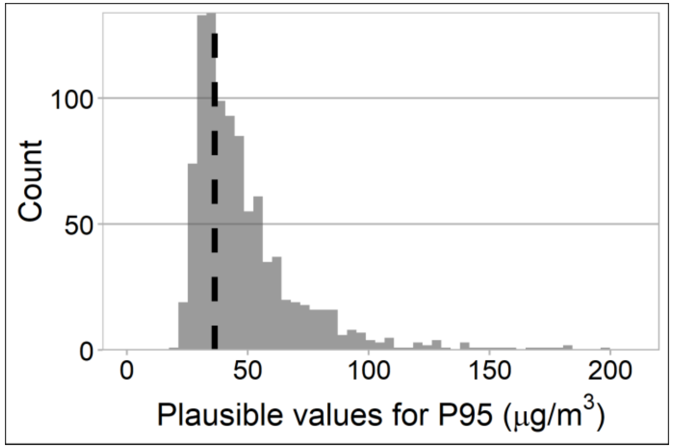
The distribution of uncertainty for the 95th percentile

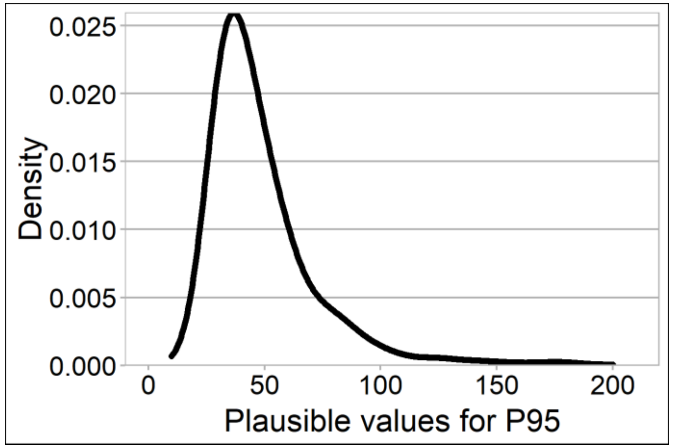
The set of possibilities for the distribution can be translated into a set of possibilities the 95th percentile.



1000 plausible exposure distributions

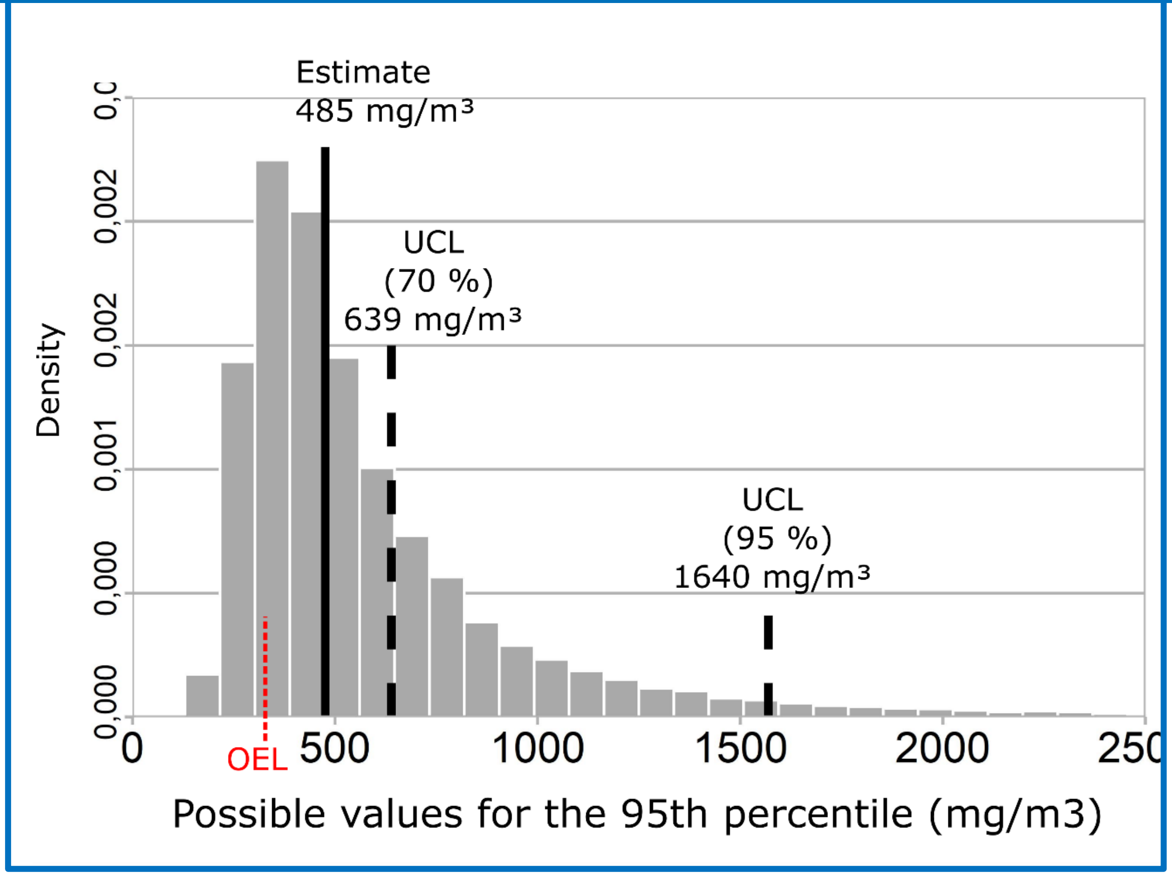




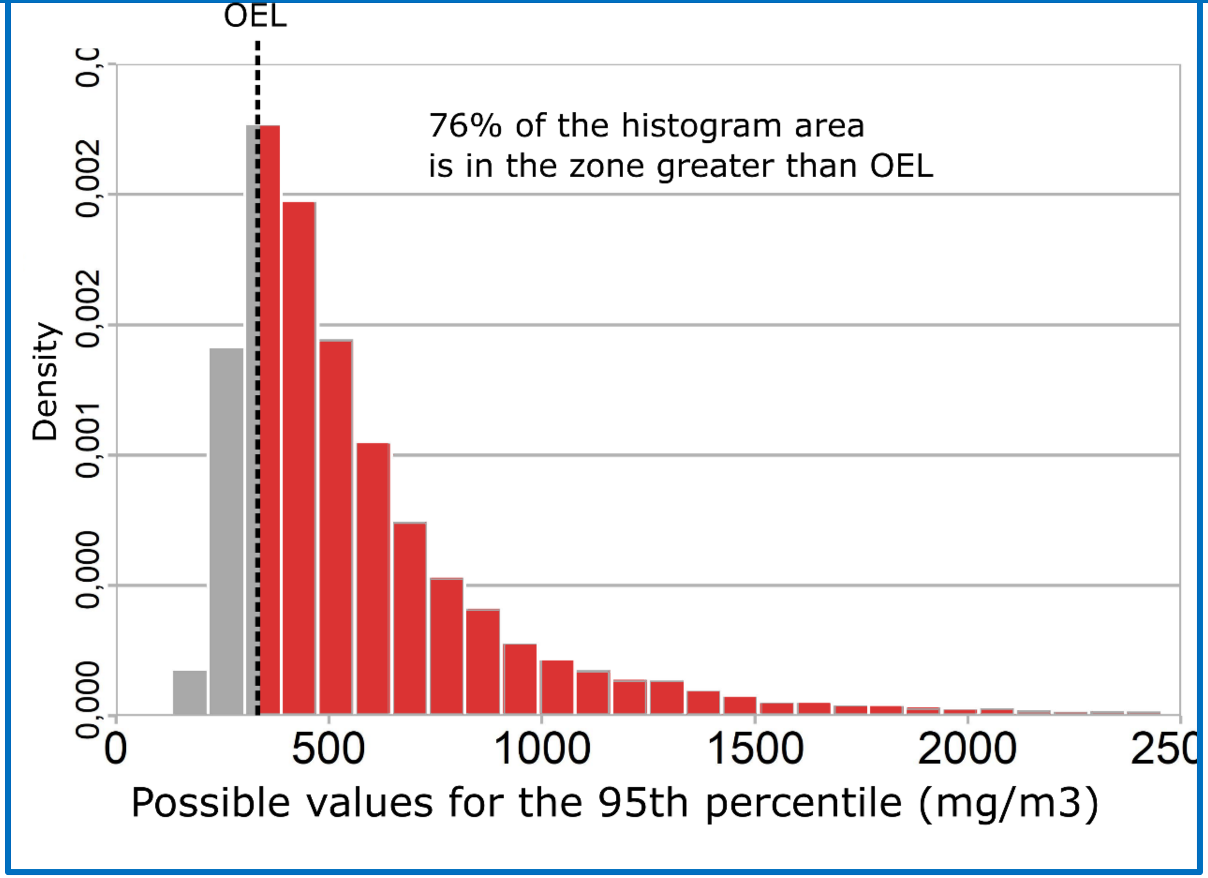


1000 plausible P95 values

An example run from Expostats : probability of overexposure vs confidence limits



Traditional interpretation



Probabilistic interpretation

Université de Montréal

ESPUM

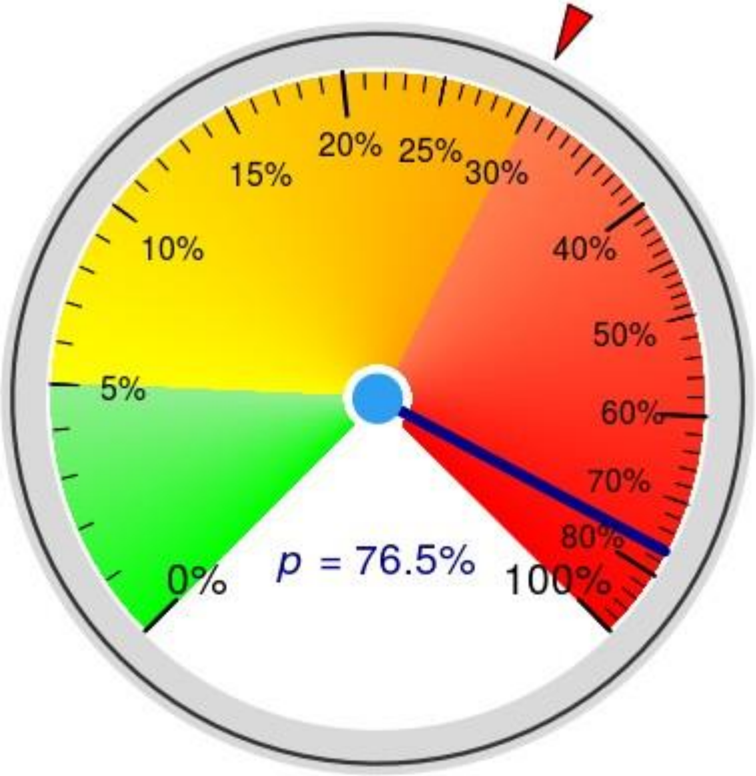
A PROBABILISTIC DECISION FRAMEWORK PROPOSED IN WWW.EXPOSTATS.CA

OEL = 350 mg/m3

Overexposure defined as : **95th percentile ≥ OEL**

Probability that this criterion is met : 76.5%

This, which we call overexposure risk is the probability that exposure is unacceptable according to the set criteria, given our model and the data.



What threshold should set us in motion ?

(overexposure risk threshold)

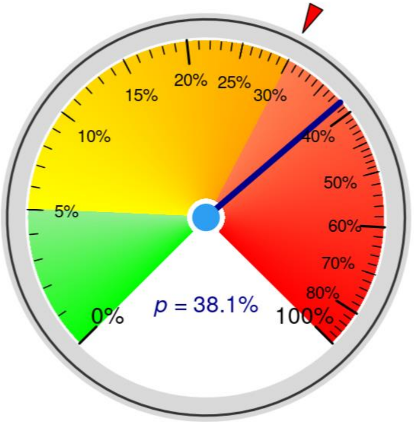
P <5% – **acceptable**

P < 30% – **tolerable**, **assuming the SEG has a required monitoring plan**

P > 30%– **problematic**, particularly if the SEG has no monitoring plan.

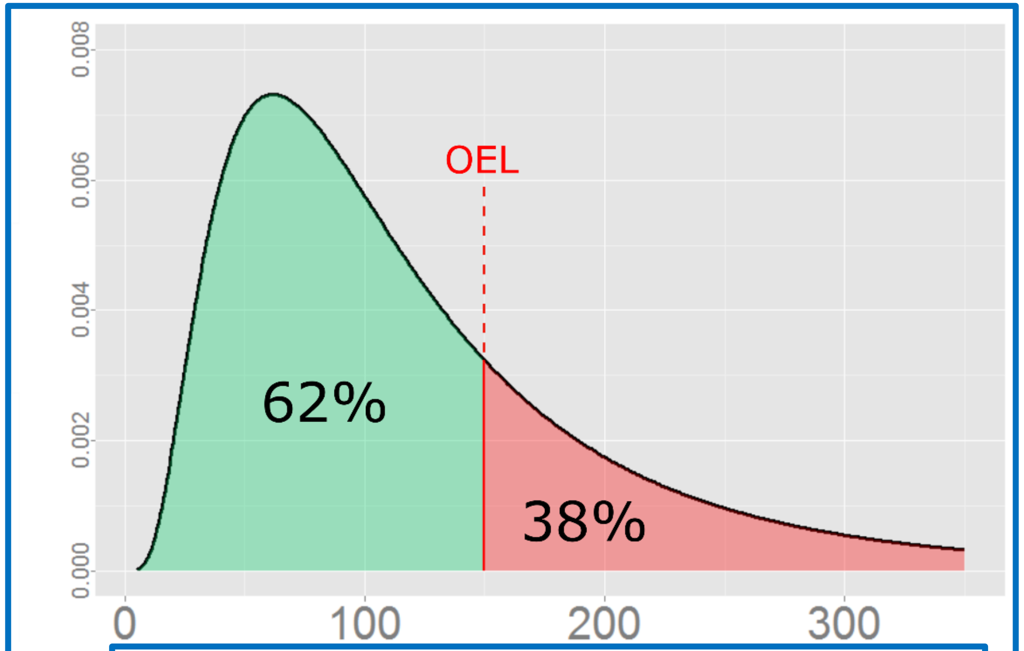
UNCERTAINTY DISTRIBUTION – RISK BANDS

As we have seen, the uncertainty distribution can be used to estimate the probability that the 95th percentile is above the OEL



This is interesting because it answers the question: what is the probability that the exposure situation is acceptable given the data at hand ?

Uncertainty distribution for the 95th percentile



Possible values for the 95th percentile

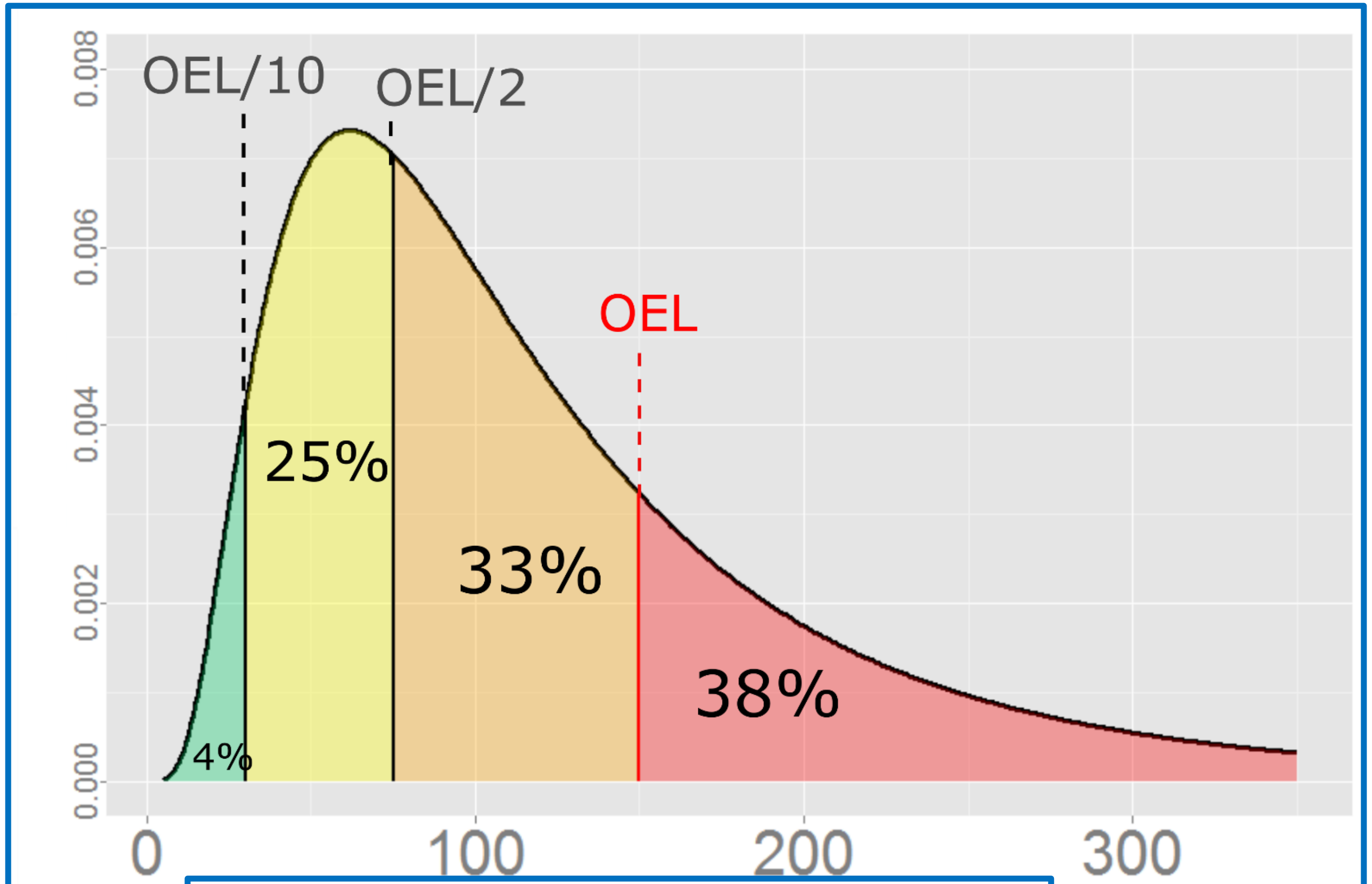
In this example: 38% chances that 95th percentile is >OEL

i.e. 38% chances that the exposure situation corresponds to overexposure

UNCERTAINTY DISTRIBUTION – RISK BANDS

Instead of dividing the uncertainty distribution into two zones, it is possible to divide it into several bands related to the OEL:

Uncertainty distribution for the 95th percentile



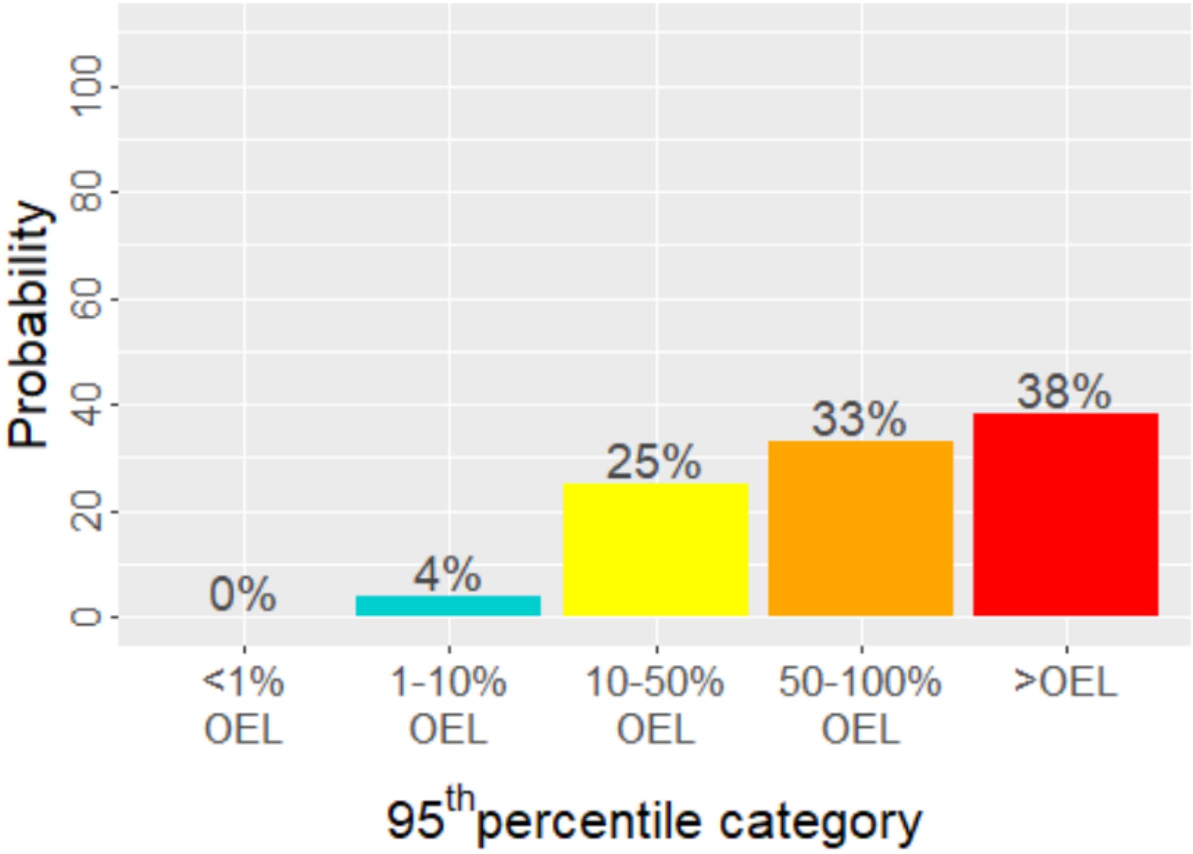
Possible values for the 95th percentile

Université de Montréal

ESPUM

UNCERTAINTY DISTRIBUTION – RISK BANDS (AIHA EXPOSURE CONTROL CATEGORIES)

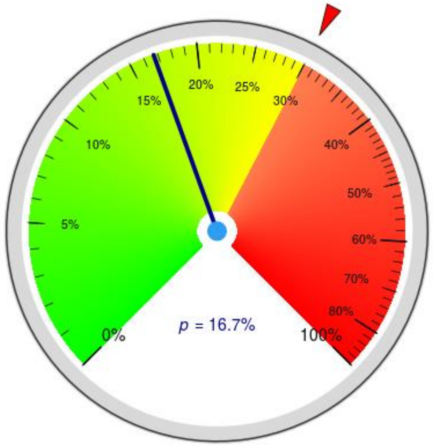
	*SEG Exposure Risk Rating***
(green)	0 (<1% of OEL)
(blue)	1 (<10% of OEL)
(yellow)	2 (10-50% of OEL)
(orange)	3 (50-100% of OEL)
(red)	4+ (>100% of OEL, Multiples of OEL; e.g., based on respirator APFs)



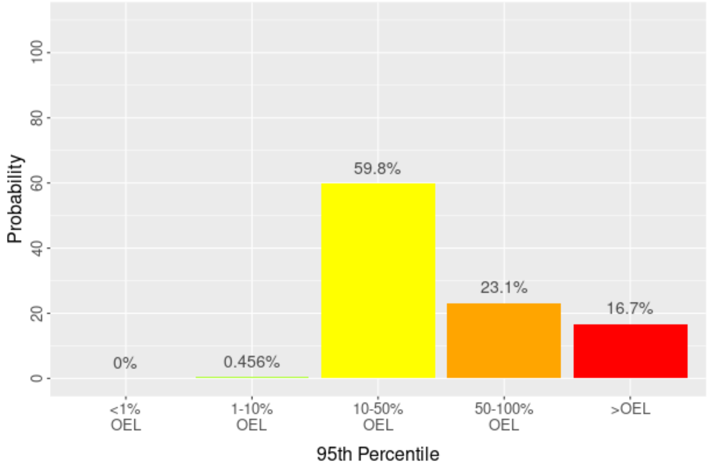
SWITCHING BETWEEN THE VARIOUS APPROACHES TO UNCERTAINTY MANAGEMENT

Traditional confidence limits

Probability of overexposure



Probability of the AIHA risk bands



The 70% upper confidence limit on the 95th percentile is smaller than the OEL

=

Overexposure risk is smaller than 30%

=

The probability of a category 4 is smaller than 30%

WWW.EXPOSTATS.CA

EDITOR'S CHOICE

Expostats: A Bayesian Toolkit to Aid the Interpretation of Occupational Exposure Measurements FREE

Jérôme Lavoué ✉, Lawrence Joseph, Peter Knott, Hugh Davies, France Labrèche, Frédéric Clerc, Gautier Mater, Tracy Kirkham

Annals of Work Exposures and Health, Volume 63, Issue 3, April 2019, Pages 267–279, https://doi.org/10.1093/annweh/wxy100

Published: 14 December 2018 **Article history** ▾

Université de Montréal

AIHA Risk Assessment Tools

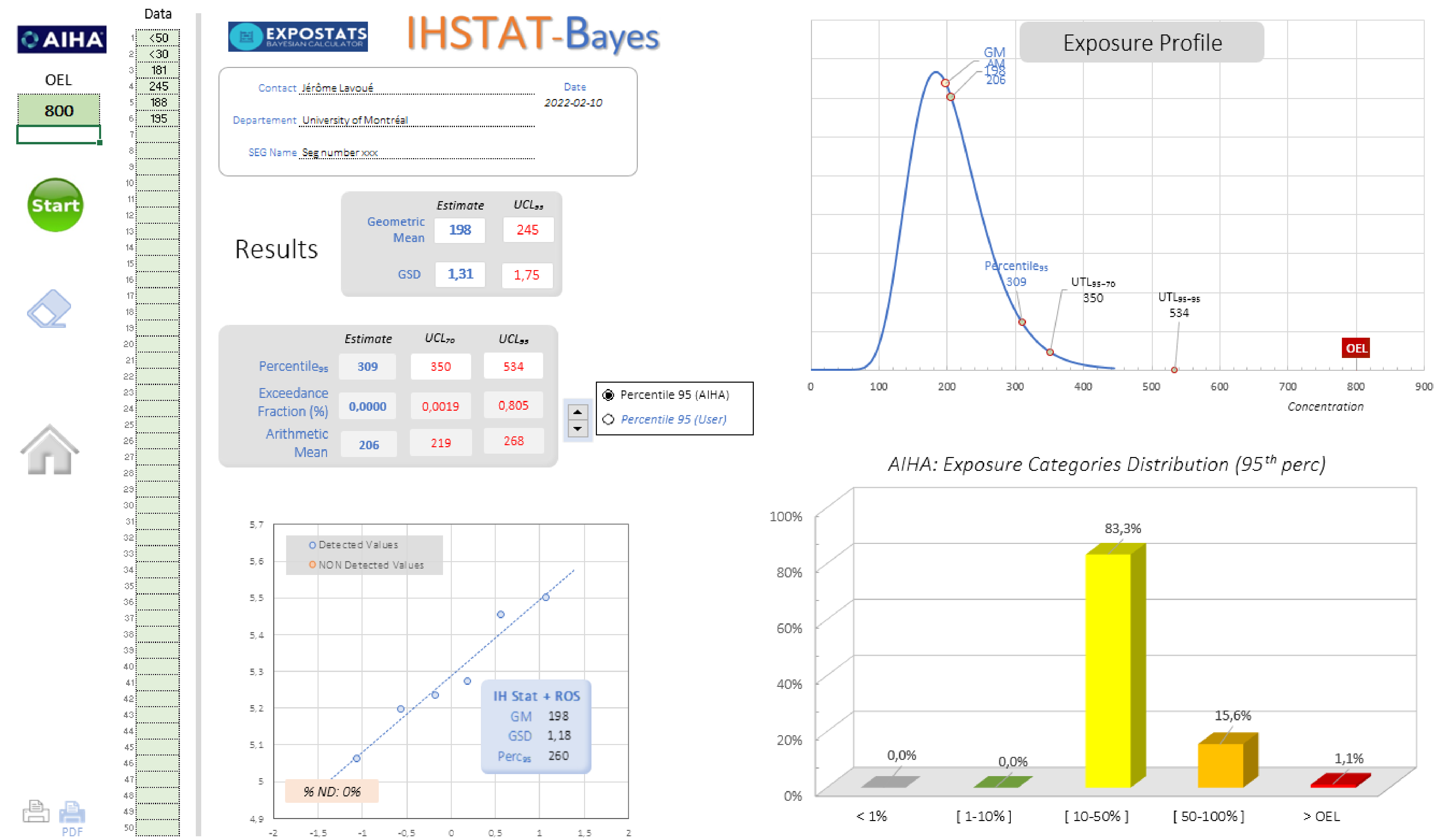
STEP 5: Refining/Validating the Exposure Assessment

IHDA-AIHA is a version of the IH Data Analyst program (www.easinc.co) designed for EHS students and professionals taking classes or professional development courses on the analysis and interpretation of occupational exposure measurements. The program calculates the standard descriptive and decision statistics recommended by the AIHA, and includes goodness-of-fit procedures and graphs, several methods for analyzing datasets containing non-detects, and Bayesian Decision Charts, which were designed to assist in the selection of the most appropriate AIHA exposure control category. Those using IHDA-AIHA for paid work on behalf of their client or employer are asked to show their appreciation to AIHA and the program's author by donating $100 USD to the American Industrial Hygiene Foundation scholarship program for each installed copy.

Expostats is a free toolbox of web applications from the University of Montréal for the interpretation of industrial hygiene measurements using the lognormal distribution. Tool1, 2, and 3 from Expostats answer the following questions, respectively: is my similar exposure group overexposed? Is my similar exposure group homogenous? What factors are associated with exposure levels in my dataset? All calculations are performed using Bayesian statistics and results are presented in the intuitive probabilistic form. An offline version is also available. Expostats is presented in an article from the *Annals of Work Exposure and Health* which was awarded the 2021 Bedford prize.

IHSTAT™ is an Excel application that calculates various exposure statistics, performs goodness of fit tests, and graphs exposure data. Multiple languages are available. The book "*A Strategy for Assessing and Managing Occupational Exposures*" is intended to accompany this tool. It provides critical detail on the use and interpretation of the various statistical outputs.

IHSTAT_BAYES (SUMMER/FALL 2022)



BAYESIAN STATS : FEATURES OF INTEREST IN INDUSTRIAL HYGIENE

Informative priors

Potential sources of information include:

Prior expert assessment
Tier ½ models
Mathematical fugitive emission models
Occupational exposure databases
Literature review
Past data

Informative priors are not mandatory

Uncertainty management - communication

It is now possible to ask questions in the form : what are the chances that ?

Technical challenges difficult for frequentist statistics

- Treatment of non detects

Bayesian statistics Vs Frequentist statistics in industrial hygiene

Both use the same distributional assumptions

Both need a "random" sample from the target population

Both can be used to derive confidence / credible limits for our parameters of interest

Both will lead to **similar results** when the priors are uninformative

BUT : Bayesian permits prior information

Bayesian handles censored data way better

Bayesian allows improved communication of uncertainty

BUT : Bayesian cannot be computed easily

Université de Montréal

ESPUM

THANK YOU !