A Discussion of Completeness in Co-Exposure Models

White Paper

National Institute for Occupational Safety and Health

January 26, 2023

Thomas LaBone
Oak Ridge Associated Universities Team

Reviewed by Tim Taulbee, John Cardarelli, and Lara Hughes
Division of Compensation Analysis and Support

This is a working document prepared by NIOSH's Division of Compensation Analysis and Support (DCAS) or its contractor for use in discussions with the ABRWH or its Working Groups or Subcommittees. Draft, preliminary, interim, and White Paper documents are not final NIOSH or ABRWH (or their technical support and review contractors) positions unless specifically marked as such. This document represents preliminary positions taken on technical issues prepared by NIOSH or its contractor. NOTICE: This report has been reviewed to identify and redact any information that is protected by the *Privacy Act 5 USC §552a* and has been cleared for distribution.

1. INTRODUCTION

The Advisory Board on Radiation and Worker Health (ABRWH) has raised data completeness issues multiple times over the past few years during both Special Exposure Cohort evaluations and co-exposure model discussions. Completeness is one of the six primary data quality dimensions defined by the Data Management Association [Wang and Strong 1996] and is discussed in Section 2.2 of DCAS-IG-006 [NIOSH 2020]. During a recent ABRWH meeting[1] the topic was discussed again, and it was proposed that a report be prepared summarizing the current issues involving data completeness and proposing a path forward or an approach to address this issue (being quantitative if possible). Completeness is a complex issue because:

- The term "completeness" (and its complement "missingness") has multiple meanings, which often creates confusion during discussions of the subject.
- Judging the completeness of a given dataset cannot be accomplished by examining the dataset alone; ancillary information must be incorporated into the evaluation.
- There are degrees of completeness, and one must specify "how complete is complete enough" for the task at hand.

Both the necessary ancillary information and the degree of completeness desired are specific to each dataset and application.

Completeness, with a focus on the completeness of datasets[2] used in the construction of co-exposure models, is the subject of this report. The completeness of a dataset in the context of co-exposure models refers to two different concepts: (i) what proportion of the data generated by a site is in the dataset provided (data completeness), and (ii) what proportion of the workers who should have been monitored, especially the most highly exposed workers, were in fact monitored (monitoring completeness). The discussion begins in Section 2 with a review of what a co-exposure model is, how it is constructed, and how it is used. Examples based on known datasets are given in Section 3 to illustrate how a co-exposure model responds to datasets with missingness caused by random and non-random mechanisms. The implications of "data missingness" and "monitoring missingness" are discussed in more detail in Sections 4 and 5. Section 6 discusses why compliance with a regulatory monitoring program cannot be used by itself to judge the completeness of a dataset used to construct a co-exposure model. Section 7 discusses the absence of quantitative criteria for completeness in epidemiology, and Section 8 discusses why stratification cannot fix data missingness or monitoring missingness. The discussion and the main conclusions are summarized in Section 9.

[1] See Appendix A.
[2] Individual monitoring only -- the use of general area air monitoring is not addressed.

This is a working document prepared by NIOSH's Division of Compensation Analysis and Support (DCAS) or its contractor for use in discussions with the ABRWH or its Working Groups or Subcommittees. Draft, preliminary, interim, and White Paper documents are not final NIOSH or ABRWH (or their technical support and review contractors) positions unless specifically marked as such. This document represents preliminary positions taken on technical issues prepared by NIOSH or its contractor. NOTICE: This report has been reviewed to identify and redact any information that is protected by the *Privacy Act 5 USC §552a* and has been cleared for distribution.

2. CO-EXPOSURE MODELS

Some details of what a co-exposure model is and how it works were given in ORAUT-RPRT-102, *Assessment of Los Alamos National Laboratory plutonium bioassay programs 1996 to 2001* [Oak Ridge Associated Universities (ORAU) Team (ORAUT) 2021]. Here that discussion will be expanded, focusing on the completeness of co-exposure datasets. The goal of a co-exposure study as used in the Energy Employees Occupational Illness Compensation Program Act (EEOICPA) program is to estimate the probability distribution of external doses or intake rates in a target population [Coggon et al. 1997], which consists of all workers exposed to a given radioactive material or external radiation in a given year during the course of work (Exposed Workers in Figure 1). All members of the target population who were monitored are part of the study population (Monitored-Exposed in Figure 1). A second population of monitored workers who are unexposed (Monitored-Unexposed in Figure 1) is also part of the study population because exposed workers who are monitored cannot be differentiated from unexposed workers who are monitored. The distribution of external doses or intake rates in the study population (or a sample of it – the study sample) is referred to as a co-exposure model. Finally, the co-exposure model is used to estimate external doses or intake rates to workers who are assumed to be exposed and are unmonitored (Unmonitored-Exposed in Figure 1).

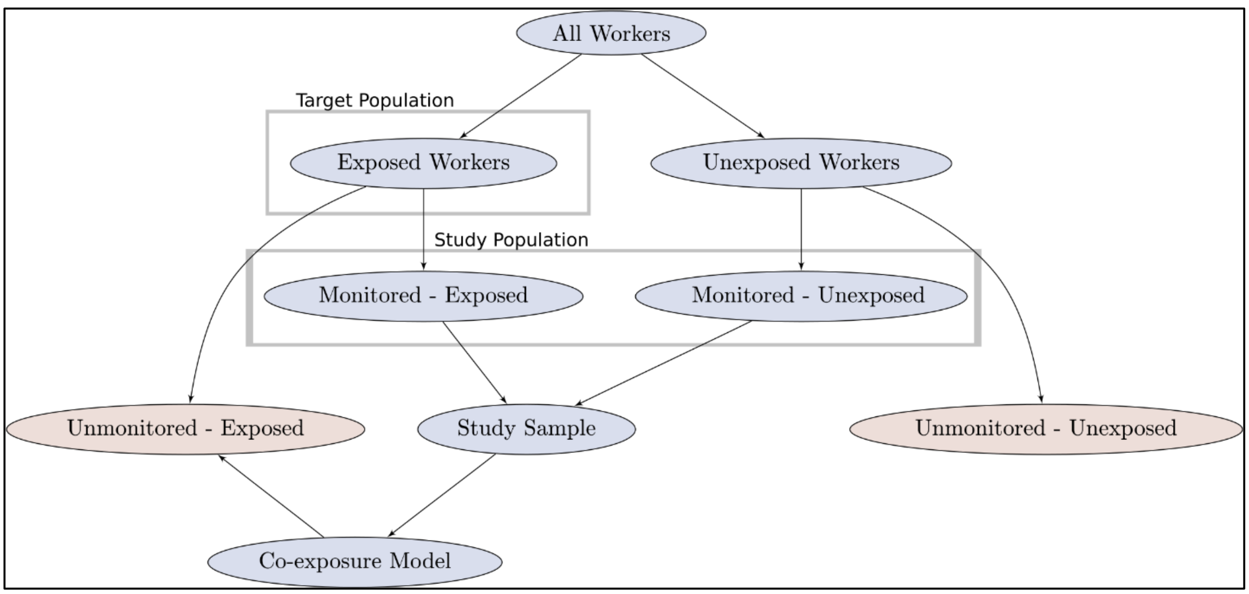

Figure 1. Co-exposure flowchart.

3. EXAMPLE CO-EXPOSURE MODELS

Co-exposure models generated from four known datasets [ORAUT 2023] will be used as examples to further the discussion. These are simple datasets for single year, with each record consisting of two fields: a unique worker identifier and an annual dose. In practice the dosimetric information could be bioassay instead of annual doses and additional fields like work location

This is a working document prepared by NIOSH's Division of Compensation Analysis and Support (DCAS) or its contractor for use in discussions with the ABRWH or its Working Groups or Subcommittees. Draft, preliminary, interim, and White Paper documents are not final NIOSH or ABRWH (or their technical support and review contractors) positions unless specifically marked as such. This document represents preliminary positions taken on technical issues prepared by NIOSH or its contractor. NOTICE: This report has been reviewed to identify and redact any information that is protected by the *Privacy Act 5 USC §552a* and has been cleared for distribution.

and occupation might be present. Missing data might be individual fields in a record or the entire record. Missing records are much more difficult to deal with compared to missing fields because:

- You will know with absolute certainty when a field is missing and there is a vast body of literature devoted to dealing with it, e.g., Soley-Bori [2013], but
- In practice, you can never be certain whether a record is missing and the methods available to deal with missing records are much more limited.

3.1 Complete Dataset

Assume a worker population consists of three strata: low-dose workers, medium-dose workers, and high-dose workers. The number of workers and the true distribution of doses in each stratum, which were selected to create clear examples, are given in Table 1. The geometric mean (GM) and geometric standard deviation (GSD) for each of the three lognormal distributions are given. A realization of this theoretical model is the population shown in Figure 2, where a random dose is drawn for each worker in the population from the theoretical lognormal distribution of their stratum. A single lognormal distribution is fit to the resulting dataset to give the distribution of the population.

Table 1. True dose distributions for worker population with three different categories of workers: low-dose, medium-dose, high-dose. The GM is in units of mrem.

Dose	Number	GM	GSD
Low	500	50	2.5
Medium	300	100	2.5
High	100	500	2.5

This is a working document prepared by NIOSH's Division of Compensation Analysis and Support (DCAS) or its contractor for use in discussions with the ABRWH or its Working Groups or Subcommittees. Draft, preliminary, interim, and White Paper documents are not final NIOSH or ABRWH (or their technical support and review contractors) positions unless specifically marked as such. This document represents preliminary positions taken on technical issues prepared by NIOSH or its contractor. NOTICE: This report has been reviewed to identify and redact any information that is protected by the *Privacy Act 5 USC §552a* and has been cleared for distribution.

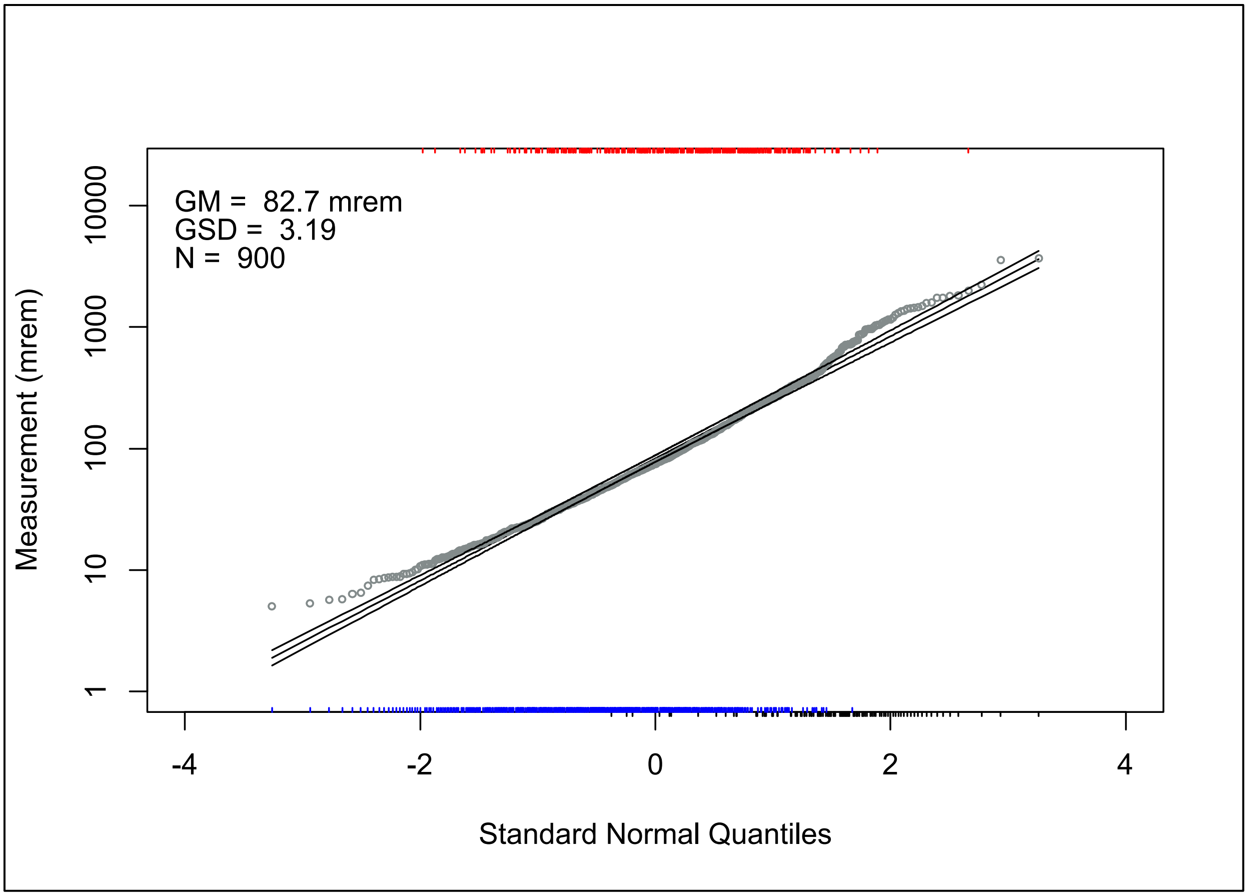

Figure 2. Distribution of doses from population of 900 workers defined in Table 1. The upper and lower lines are the 95% confidence band on the line of best fit, which is the center line. The red rug plot[3] on the top indicates the z-score of the workers from the middle stratum, the blue rug plot on the bottom indicates the z-score for workers from the low stratum, and the black rug plot on the bottom indicates the z-score for the workers from the high stratum.

Note there are no doses missing from the co-exposure model in Figure 2; it is complete in every way. In practice doses are often missing. For example, worker doses can be missing because:

- Some workers who should have been monitored according to site procedures were not monitored, and their doses are therefore not available for inclusion in the co-exposure model, which is referred to as "monitoring missingness."

- Some of the worker doses generated by the site have not been captured and therefore are not available for inclusion in the co-exposure model, which is referred to as "data missingness."

3 https://en.wikipedia.org/wiki/Rug_plot.

This is a working document prepared by NIOSH's Division of Compensation Analysis and Support (DCAS) or its contractor for use in discussions with the ABRWH or its Working Groups or Subcommittees. Draft, preliminary, interim, and White Paper documents are not final NIOSH or ABRWH (or their technical support and review contractors) positions unless specifically marked as such. This document represents preliminary positions taken on technical issues prepared by NIOSH or its contractor. NOTICE: This report has been reviewed to identify and redact any information that is protected by the *Privacy Act 5 USC §552a* and has been cleared for distribution.

Rather than missingness, its complement is usually discussed: the completeness of a dataset. Note that when someone refers to completeness of a dataset he or she may be referring to the monitoring completeness, data completeness, or both. When this discussion states that a dataset is complete without modification, it means it is complete both in the sense of monitoring and data.

3.2 Records Missing at Random

Continuing the example, in Figure 3 only the data from 100 workers are available from the original 900 workers, meaning that data from 800 workers are missing. For this example, the monitored workers were selected at random across the strata (Random column of Table 2). In other words, the mechanism for the missingness is random. A random sample of the workers will give a co-exposure model that is an unbiased estimate[4] of the full co-exposure model, but with greater uncertainties in the model parameters. This can be seen by comparing the 95% confidence band in Figure 2 to the 95% confidence band in Figure 3. Such co-exposure models are referred to as being representative. In this example the random sample has doses from 11% of the workers, yet gives an unbiased estimate of the true co-exposure model. This illustrates that the number of workers in a sample or the proportion of workers monitored is not by itself a conclusive measure of completeness. On the contrary, the mechanism of missingness (e.g., is it random?) is almost always more important than the actual number of results missing.

Another point to make is that the reason for the missingness, i.e., the workers were not monitored or their data are not available, is irrelevant when the missingness is random. When the missingness is not random, we must consider data completeness and monitoring completeness separately. In the remainder of this section the behavior of the example dataset is first studied for two cases of non-random monitoring missingness and then some general comments are made on the issue of non-random data missingness.

4 The term "unbiased" means that if this experiment was repeated a large number of times the mean of the GM and GSD would equal the population GM and GSD.

This is a working document prepared by NIOSH's Division of Compensation Analysis and Support (DCAS) or its contractor for use in discussions with the ABRWH or its Working Groups or Subcommittees. Draft, preliminary, interim, and White Paper documents are not final NIOSH or ABRWH (or their technical support and review contractors) positions unless specifically marked as such. This document represents preliminary positions taken on technical issues prepared by NIOSH or its contractor. NOTICE: This report has been reviewed to identify and redact any information that is protected by the *Privacy Act 5 USC §552a* and has been cleared for distribution.

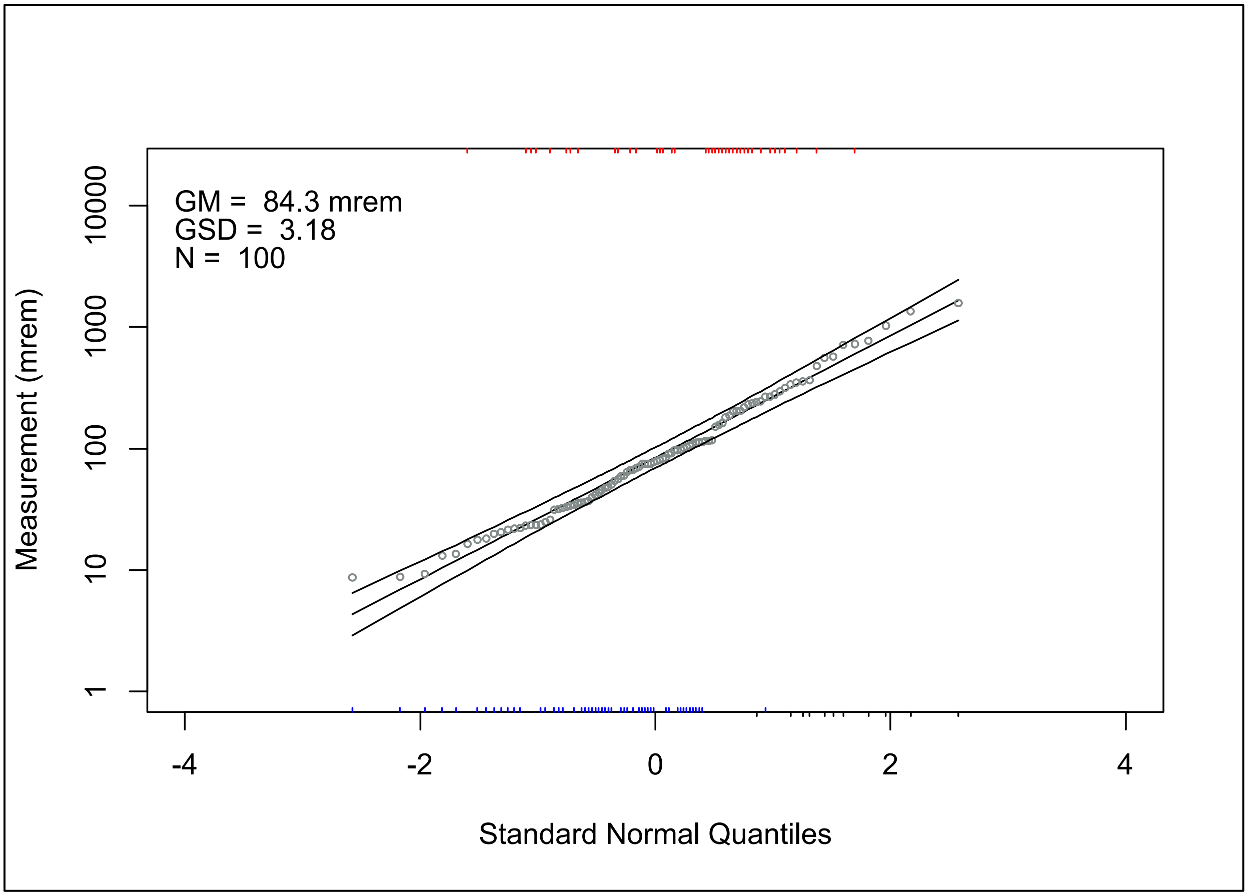

Figure 3. Random sample of 100 doses taken from 900 doses (11% of population monitored) shown in Figure 2. The upper and lower lines are the 95% confidence band for the center line, which is the line of best fit. The red rug plot on the top indicates the z-score of the workers from the middle stratum, the blue rug plot on the bottom indicates the z-score for workers from the low stratum, and the black rug plot on the bottom indicates the z-score for the workers from the high stratum.

Table 2. Number of workers in dose categories in each of the four co-exposure datasets. The number of workers in each stratum of the Biased-High and Biased-Low datasets was arbitrarily selected from the Complete dataset to reflect the condition being discussed.

Dose category	**Complete dataset**	**Random sample**	**Biased-high sample**	**Biased-low sample**
Low	500	50	250	450
Medium	300	39	150	250
High	100	11	100	10
Total	900	100	500	710
Percent complete	100	11	56	79

This is a working document prepared by NIOSH's Division of Compensation Analysis and Support (DCAS) or its contractor for use in discussions with the ABRWH or its Working Groups or Subcommittees. Draft, preliminary, interim, and White Paper documents are not final NIOSH or ABRWH (or their technical support and review contractors) positions unless specifically marked as such. This document represents preliminary positions taken on technical issues prepared by NIOSH or its contractor. NOTICE: This report has been reviewed to identify and redact any information that is protected by the *Privacy Act 5 USC §552a* and has been cleared for distribution.

3.3 Records from Low-Dose and Medium-Dose Strata are Preferentially Missing (Biased High)

The assumption of data missing at random discussed in the previous section is a very strong assumption that is seldom realized in radiation protection monitoring programs. A more common pattern of monitoring missingness is to monitor workers from the high-dose strata more completely than workers from the other strata. This pattern is a result of radiation protection programs focusing on compliance with dose limits, which means that more attention is given to properly monitoring the workers in the high-dose stratum. This is both intuitively obvious and empirically verified when looking in detail at the exposure potential of unmonitored workers [Richardson et al. 1999, p. 6], for example. An example of this type of missingness pattern is shown in Figure 4, which shows the co-exposure model resulting from the systematic exclusion of low- and medium-dose workers from the monitoring program (Biased-High column of Table 2). This illustrates that if data are preferentially missing from low-dose and medium-dose workers, the resulting co-exposure model will be biased high (i.e., the lognormal distribution will be shifted to the right -- compare the GM in Figure 2 to the GM in Figure 4) and therefore be claimant favorable.

This is a working document prepared by NIOSH's Division of Compensation Analysis and Support (DCAS) or its contractor for use in discussions with the ABRWH or its Working Groups or Subcommittees. Draft, preliminary, interim, and White Paper documents are not final NIOSH or ABRWH (or their technical support and review contractors) positions unless specifically marked as such. This document represents preliminary positions taken on technical issues prepared by NIOSH or its contractor. NOTICE: This report has been reviewed to identify and redact any information that is protected by the *Privacy Act 5 USC §552a* and has been cleared for distribution.

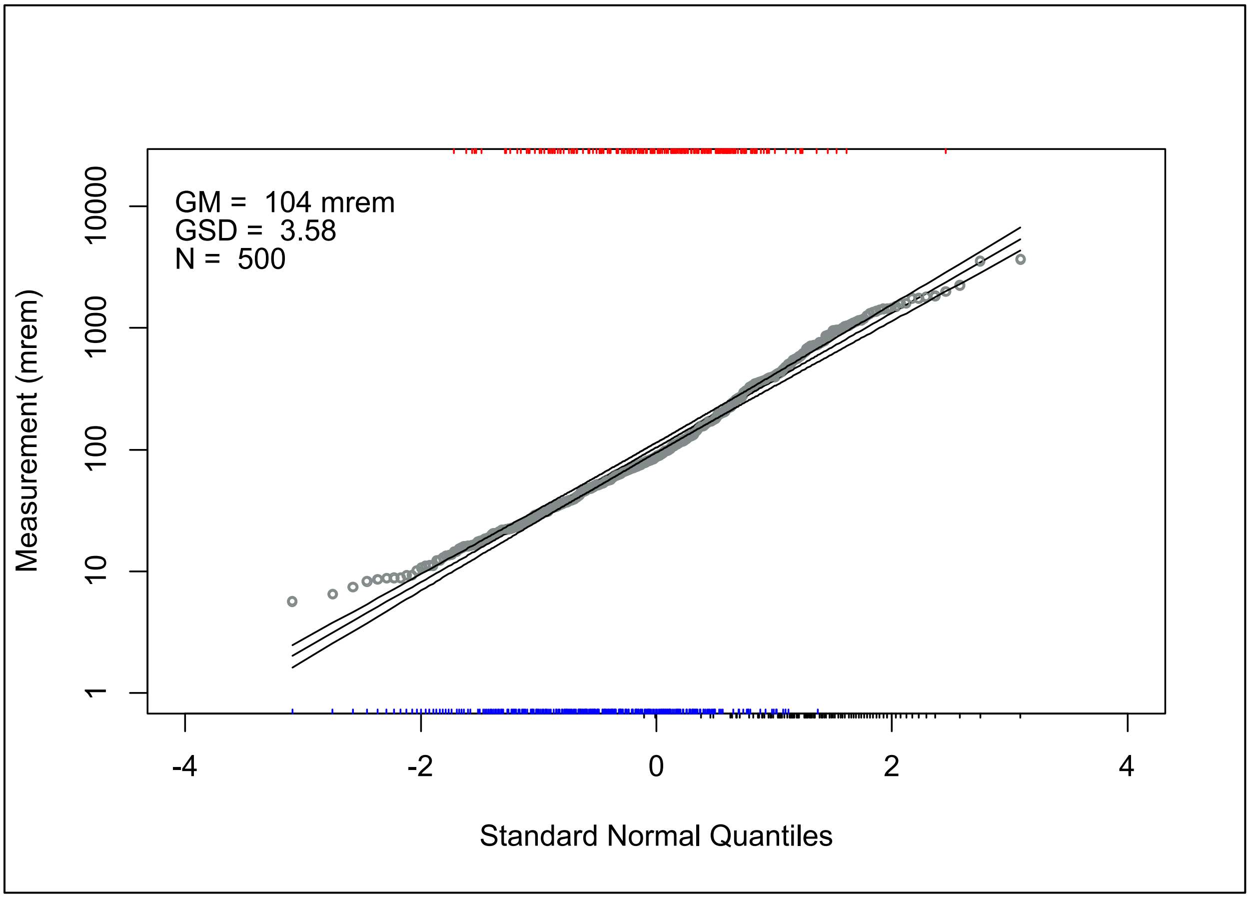

Figure 4. Sample of 500 doses taken from 900 doses (56% of population monitored) shown in Figure 2, systematically excluding workers with low and medium doses (see the Biased-High column in Table 2). The upper and lower lines are the 95% confidence band for the center line which is the line of best fit. The red rug plot on the top indicates the z-score of the workers from the middle stratum, the blue rug plot on the bottom indicates the z-score for workers from the low stratum, and the black rug plot on the bottom indicates the z-score for the workers from the high stratum.

In this case the use of only 56% of the total data resulted in a co-exposure model that is biased high. Since radiation protection programs tend to preferentially monitor the highest exposed workers, most co-exposure models are expected to fall into this category.

3.4 Records from High-Dose Stratum are Preferentially Missing (Biased Low)

At the opposite extreme is the case where a significant proportion of the high-dose workers is missing from the sample because they were not monitored. This is illustrated in Figure 5, where the co-exposure model resulting from the systematic exclusion of high-dose workers is given (Biased-Low column of Table 2). This illustrates that if data are preferentially missing from high-dose workers, the resulting co-exposure model will be biased low, i.e., the lognormal

This is a working document prepared by NIOSH's Division of Compensation Analysis and Support (DCAS) or its contractor for use in discussions with the ABRWH or its Working Groups or Subcommittees. Draft, preliminary, interim, and White Paper documents are not final NIOSH or ABRWH (or their technical support and review contractors) positions unless specifically marked as such. This document represents preliminary positions taken on technical issues prepared by NIOSH or its contractor. NOTICE: This report has been reviewed to identify and redact any information that is protected by the *Privacy Act 5 USC §552a* and has been cleared for distribution.

distribution will be shifted to the left (compare the GM in Figure 2 to the GM in Figure 5). In this unlikely scenario the data can't be used to construct a representative or bounding co-exposure model.

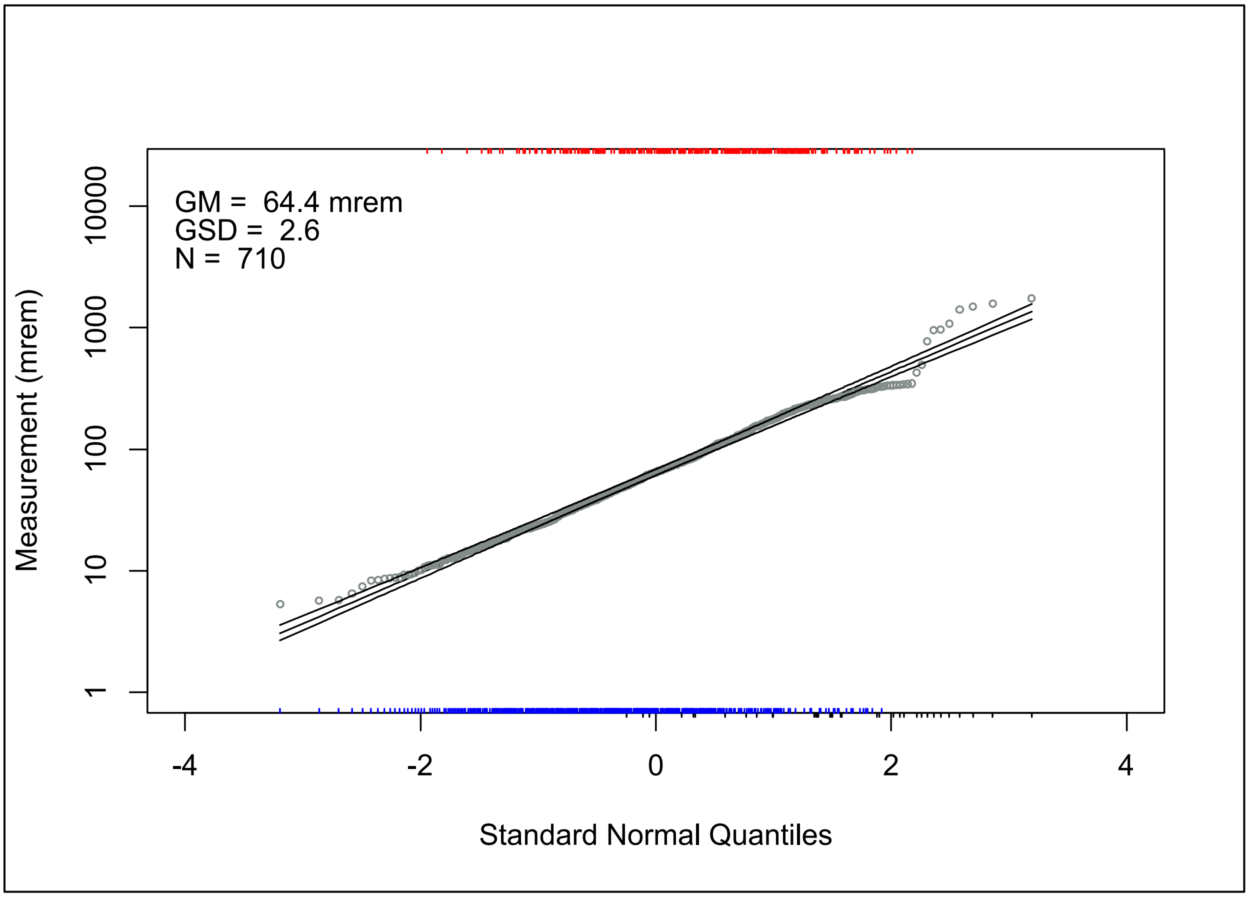

Figure 5. Sample of 710 doses taken from 900 doses (79% of population monitored) shown in Figure 2, systematically excluding workers with high doses (see the Biased-Low column in Table 2). The upper and lower lines are the 95% confidence band for the center line, which is the line of best fit. The red rug plot on the top indicates the z-score of the workers from the middle stratum, the blue rug plot on the bottom indicates the z-score for workers from the low stratum, and the black rug plot on the bottom indicates the z-score for the workers from the high stratum.

Table 3. Summary of estimated parameters from co-exposure models. The GM is in units of mrem.

Parameter	GM	GSD
Complete	82.7	3.19
Random	84.3	3.18
Biased high	104	3.58
Biased low	64.4	2.6

This is a working document prepared by NIOSH's Division of Compensation Analysis and Support (DCAS) or its contractor for use in discussions with the ABRWH or its Working Groups or Subcommittees. Draft, preliminary, interim, and White Paper documents are not final NIOSH or ABRWH (or their technical support and review contractors) positions unless specifically marked as such. This document represents preliminary positions taken on technical issues prepared by NIOSH or its contractor. NOTICE: This report has been reviewed to identify and redact any information that is protected by the *Privacy Act 5 USC §552a* and has been cleared for distribution.

In this case, the use of 79% of the total data resulted in a co-exposure model that was biased low, because the highest exposed workers were preferentially missing from the dataset. This example further illustrates that the mechanism of data missing can be more important than the percentages.

3.5 Relevance to Co-Exposure Models

These examples illustrate a number of key concepts:

- When performing dose reconstructions in the EEOICPA program, all of the workers in the target population do not have to be monitored to construct a co-exposure model. In fact, if all the workers were monitored (and all the data were available) there would be no need for a co-exposure model.
- The absolute number or proportion of workers monitored is not by itself an adequate measure of the ability to construct a co-exposure model from that dataset, i.e., the mechanism of missingness (e.g., are the data missing at random) must also be considered.
- Rather than focusing on the number or proportion of workers monitored, determine if the radiological monitoring program preferentially missed doses from the most highly exposed workers in the target population.
- A bounding co-exposure model can be constructed if the doses from the most highly exposed workers were not preferentially missed (Section 3.1, 3.2, 3.3), regardless of the number or proportion of workers monitored. In fact, in theory a bounding co-exposure model can be constructed from a single dose: that of the most highly exposed worker.
- A bounding co-exposure model cannot be constructed if the doses from the most highly exposed workers were preferentially missed (Section 3.4), regardless of the number or proportion of workers monitored.

Who are the most highly exposed workers? In a functional radiation protection program under routine conditions, the most highly exposed workers tend to be those who:

- Routinely work in close proximity to radiation sources and unencapsulated radioactive material. These work locations can be identified by postings like Airborne Radioactivity Area and High Radiation Area.
- Perform work under job-specific radiation work permits that specify the use of Personal Protective Equipment, establish hold points, specify required bioassay and external dosimetry, etc.
- Use protective equipment like respirators (full face, plastic suit, etc.) and have received specialized training in the use of such equipment.

This is a working document prepared by NIOSH's Division of Compensation Analysis and Support (DCAS) or its contractor for use in discussions with the ABRWH or its Working Groups or Subcommittees. Draft, preliminary, interim, and White Paper documents are not final NIOSH or ABRWH (or their technical support and review contractors) positions unless specifically marked as such. This document represents preliminary positions taken on technical issues prepared by NIOSH or its contractor. NOTICE: This report has been reviewed to identify and redact any information that is protected by the *Privacy Act 5 USC §552a* and has been cleared for distribution.

- Are trained in proper procedures for donning and doffing protective clothing, detecting and controlling radioactive contamination, and minimizing the time required to complete a specific task.

With the data from a significant portion of these workers, a representative or bounding co-exposure model can be constructed. Again, note that all of the most highly exposed workers need not have been monitored, just a significant portion of them.

The discussion thus far has focused on monitoring completeness. Similar conclusions can be reached when considering data completeness in the examples, with the important exception that no pattern of data missingness is more likely than another. In other words, given that workers were monitored and there are flaws in the recordkeeping system that result in gaps in the data, any conceivable pattern of data missingness is possible. This difference in monitoring and data missingness mechanisms means the two types of missingness must be assessed using different methods.

4. MORE ON DATA COMPLETENESS

The discussion in the previous section begs the question of how to determine if there is enough data from the most highly exposed workers to generate a representative or bounding co-exposure model. To facilitate the discussion, as a thought experiment consider a modern[5] radiation protection program (e.g., as of the date of this report) at a facility where the workers handle unencapsulated radioactive material and the work is performed under an electronic radiological work permit (RWP) system. Before workers can successfully sign into an RWP they must have the proper training and qualifications registered in the system. The dates workers sign into the RWP and the associated bioassay programs required by the RWP are recorded, and at the end of the routine monitoring period requests are made for the appropriate bioassay based on all the RWPs they signed into for the monitoring period. In theory this system would monitor every worker that needed to be monitored and a co-exposure model is unnecessary. However, even in this hypothetically perfect system some workers might not be properly monitored because (i) they failed to participate in the prescribed bioassay program, (ii) the bioassay requirements on the RWP were in retrospect misspecified by the radiological control personnel, or (iii) some of the data were not in the dataset provided by the site. This means a co-exposure model is needed.

To obtain the bioassay data for the co-exposure model, we request all the bioassay results assigned to workers for the period of interest. Subsequently the data is received, but is it all the data generated by the site? We can ask the site personnel how they know that they have provided all the data. The site personnel might respond by saying they have provided all the data in their recordkeeping system, which is not necessarily the same as what was requested. How does one ascertain that all the data in the recordkeeping system is all the data that was generated? In the

5 For the sake of consistency, the "modern" era will be assumed to be when the site was formally in compliance with Federal Rule 10 C.F.R. § 835 [2022]. Note that many important aspects of the modern program may have been implemented years or decades prior to that time.

This is a working document prepared by NIOSH's Division of Compensation Analysis and Support (DCAS) or its contractor for use in discussions with the ABRWH or its Working Groups or Subcommittees. Draft, preliminary, interim, and White Paper documents are not final NIOSH or ABRWH (or their technical support and review contractors) positions unless specifically marked as such. This document represents preliminary positions taken on technical issues prepared by NIOSH or its contractor. NOTICE: This report has been reviewed to identify and redact any information that is protected by the *Privacy Act 5 USC §552a* and has been cleared for distribution.

modern monitoring program described in the thought experiment, presume that: (i) the records system would have been designed according to industry standards, (ii) the software would be implemented, tested, and documented, (iii) procedures would have been implemented and used with the system, (iv) all these documents are available, and (v) individuals involved with the recordkeeping system are available to answer questions. In other words, it would be practical to vet the modern recordkeeping system to ensure it is reasonable to presume that all the data is indeed in the system. The key idea is to vet the process that generates the data, but in practice (and perhaps in principle) it is not possible to know with absolute certainty that the data are complete.

After vetting the data generation process, internal and external consistency checks of the data can be performed, e.g., looking for unexplained gaps in the data, comparing the roster of monitored workers to the worker roster obtained from human resources, or comparing the data in the site dataset to data in the NIOSH Office of Compensation Analysis and Support (OCAS) Claims Tracking System (NOCTS)[6]. The idea is to look for evidence of missing data that would cast doubt on the presumption of data completeness and prompt one to assess the nature of the missing data. For example, are the data missing completely at random or is there a pattern to the missing data that would indicate a systematic problem? One can go back to the site and attempt to obtain the missing data, but the final analysis must decide if the data provided by the site personnel are complete enough to generate a representative or bounding co-exposure model for unmonitored workers. It would be convenient to establish a quantitative limit for adequate data completeness, e.g., 80% of the data is needed to declare that it is “complete enough” to make a co-exposure model, but in essence this is a decision-based qualitative preponderance of evidence argument.

The discussion thus far has assumed working with modern data loaded on a modern computer recordkeeping system. Older data are likely to have been originally loaded into a different recordkeeping system by different people using different software and procedures and were later migrated to the current recordkeeping system. Data collected before computer recordkeeping systems became commonplace might be in hardcopy form only. The same comparisons and vetting procedures discussed above should be applied to hardcopy and migrated datasets as much as possible.

5. MORE ON MONITORING COMPLETENESS

In the modern radiation protection program of the thought experiment in Section 4, monitoring completeness is a function of the prescribed program being the right program for the hazards encountered during the job and the workers complying with the prescribed monitoring program. Details concerning the workers who did not participate in the monitoring program should be

[6] The test described in ORAUT-RPRT-0086 compares the dataset provided by the site for the co-exposure model to the data provided by the site for dose reconstructions. We are asking for the same data twice and comparing them for consistency (another one of the six primary data quality dimensions defined by the Data Management Association). If one dataset (i.e., NOCTS) is assumed to be complete, then one can say inconsistencies imply incompleteness and we can estimate the degree of incompleteness.

This is a working document prepared by NIOSH’s Division of Compensation Analysis and Support (DCAS) or its contractor for use in discussions with the ABRWH or its Working Groups or Subcommittees. Draft, preliminary, interim, and White Paper documents are not final NIOSH or ABRWH (or their technical support and review contractors) positions unless specifically marked as such. This document represents preliminary positions taken on technical issues prepared by NIOSH or its contractor. NOTICE: This report has been reviewed to identify and redact any information that is protected by the *Privacy Act 5 USC §552a* and has been cleared for distribution.

readily available from the modern monitoring program. This information can be analyzed and decisions reached as to whether these workers constitute a significant portion of the workers with the highest exposure potential. It is presumed that the modern radiation protection program at a major Department of Energy (DOE) facility has the wherewithal to identify significant hazards on the job and ensure that appropriate monitoring is offered to workers with the highest exposure potential. A realistic assessment of whether the right monitoring program was prescribed in the past requires vetting of the radiation protection program to some degree. Such an evaluation might consist of a description of the program and its capabilities, the expressed intent to monitor workers, and evidence describing the extent to which workers with the highest exposure potential may not have been monitored. Thus, deciding if the data are complete enough to construct a bounding co-exposure model is once again a preponderance of evidence exercise that cannot be associated with a single statistic for all cases.

6. REGULATORY COMPLIANCE AND COMPLETENESS

Essentially all of the internal and external dose data used to construct co-exposure models were collected to demonstrate compliance with occupational dose limits. At first thought, regulatory compliance with the monitoring program might be considered to provide evidence of data and monitoring completeness. However, compliance with regulations established for limiting dose to workers is an entirely different endeavor than constructing a co-exposure model with the intent of reconstructing dose to a claimant. The goal of monitoring for compliance with regulations is to demonstrate that no workers exceeded the occupational dose limits. On the other hand, the goal of monitoring for generating co-exposure models is to obtain a representative or bounding sample of the monitored workers. Thus, regulatory compliance is not necessarily related to the ability to construct co-exposure models and cannot be used by itself to judge the monitoring completeness of a dataset. In other words, compliance with the regulations does not by itself prove that one can construct a co-exposure model, and noncompliance with the regulations by itself does not prove that one cannot.

7. EPIDEMIOLOGY AND COMPLETENESS

In the past 15 years the construction of co-exposure models in EEOICPA has become more sophisticated, adopting many methods used in radiation epidemiology. For this reason, it is of interest to explore how the topic of completeness is handled in those types of studies. In radiation epidemiology studies the researchers are concerned about the completeness of the data used because it directly impacts the validity of their studies. Papers like those of Richardson et al. [1999; 2006] show that efforts are made to ensure data completeness. Nevertheless, quantitative criteria for “complete enough” are not offered in these papers, and a cursory review of the literature supports the conclusion that such criteria are not generally available.

In the development of a co-exposure model considerable efforts are required to ensure that the workers with the highest doses were monitored and are adequately represented in the model. In

This is a working document prepared by NIOSH’s Division of Compensation Analysis and Support (DCAS) or its contractor for use in discussions with the ABRWH or its Working Groups or Subcommittees. Draft, preliminary, interim, and White Paper documents are not final NIOSH or ABRWH (or their technical support and review contractors) positions unless specifically marked as such. This document represents preliminary positions taken on technical issues prepared by NIOSH or its contractor. NOTICE: This report has been reviewed to identify and redact any information that is protected by the *Privacy Act 5 USC §552a* and has been cleared for distribution.

epidemiology this monitoring completeness is often handled in a less rigorous fashion. For example:

- Richardson et al. [2006, p. 5 described monitoring completeness of the dataset used for the epidemiological study at the Savannah River Site:

  Not all workers were included in the Site's radiation dosimetry program in all years; however, since the start of operations, Site policy has been to monitor external radiation exposure for workers who entered a controlled area.

- Monitoring completeness for the Rocketdyne epidemiological study was described by the National Council on Radiation Protection and Measurements (NCRP) [2018, p. 106]:

  A worker whose folder contained no indication of potential exposure to radionuclides or radiation was assumed not to be a radiation worker.

Epidemiologic studies can proceed (e.g., to define risk estimates from exposure to radiation) if some workers were not monitored, even if the workers with the highest exposure potential are missing. In essence, the study is defined so that conclusions can be drawn from the available data. For example, an epidemiologic study could be based solely on the doses from white male workers, excluding doses from females and other races. This is not an option for co-exposure modeling, i.e., we must be concerned with workers who were not monitored because the purpose of the model is to estimate doses to these workers. Thus, epidemiologic methodology does not appear to offer us quantitative guidelines on what degree of completeness is acceptable or how to deal with the thorny problem of monitoring completeness.

8. STRATIFICATION IN CO-EXPOSURE MODELS

The workers in the target population (see Figure 1) inevitably perform a variety of tasks with different exposure potentials. If there is a logical way to a priori group the target population so that workers in each group receive similar doses, each group can be given its own co-exposure model. The stratification of the target population can result in more accurate co-exposure models if the exposure potential of the groups (strata) was indeed significantly different [Lohr 2010, p. 46] and workers were accurately assigned to the appropriate strata. In epidemiology, stratification of the dataset is also used to reduce confounding[7] [Rothman et al. 2008, p. 70]. Note that:

- In general, stratification of the target population when there are in fact no significant differences in exposure potential will produce less accurate co-exposure models because

[7] For example, when studying the relationship between exposure of uranium miners to radon and lung cancer, smoking tobacco is a confounding factor because it also causes lung cancer and miners tend to smoke. To address the confounding factor the population of miners would be stratified by whether they smoked.

This is a working document prepared by NIOSH's Division of Compensation Analysis and Support (DCAS) or its contractor for use in discussions with the ABRWH or its Working Groups or Subcommittees. Draft, preliminary, interim, and White Paper documents are not final NIOSH or ABRWH (or their technical support and review contractors) positions unless specifically marked as such. This document represents preliminary positions taken on technical issues prepared by NIOSH or its contractor. NOTICE: This report has been reviewed to identify and redact any information that is protected by the *Privacy Act 5 USC §552a* and has been cleared for distribution.

of the reduction in the sample size of each stratum. In other words, the specification of the strata must be meaningful.

- For stratification to be effective, workers must be accurately assigned to the appropriate strata in the development of the co-exposure model (which can be a lengthy and complex task) and in the application of the co-exposure model to a specific worker.

- Stratified co-exposure models will result in some workers being assigned more dose and some workers being assigned less dose when compared to the doses assigned from an unstratified co-exposure model. This fact is often unappreciated.

- Stratification cannot be used to correct for data missingness or monitoring missingness. More specifically, stratification cannot be used to deal with target populations where a particular group of workers is assumed to be the most highly exposed and whose data are missing.

As discussed previously, radiation protection programs like those at major DOE facilities do not monitor workers at random. Rather, workers who a priori are deemed to have the highest potential for exposure are preferentially monitored. The workers most likely to be monitored and therefore form the basis of a bounding co-exposure model would be the ones with the higher doses. This means that workers most likely to be unmonitored and therefore most likely to require a dose assigned from a co-exposure model would be receiving lower doses. Thus, a bounding unstratified co-exposure model is likely to (i) assign higher doses to the low-dose workers than would be assigned from a stratified co-exposure model (thus being claimant favorable), and (ii) assign lower doses to the high-dose workers than they would be assigned from a stratified co-exposure model—but those workers are most likely to be monitored and will not need the co-exposure model. Thus, if the radiation protection program at a facility was mature and functional, stratification is not needed to construct a bounding co-exposure model and can result in less accurate predictions of dose compared to unstratified co-exposure models.

9. SUMMARY

The completeness of a dataset in the context of co-exposure models refers to the proportion of the data generated by a site that is in the dataset provided (data completeness) and whether the workers who should have been monitored were in fact monitored (monitoring completeness). This report began with a discussion of how co-exposure models are constructed and used and how the two different types of completeness, data and monitoring, impact the model. Co-exposure models generated from simulated data were used as examples to help illustrate the key concepts. The main conclusions of this discussion are:

- Making a bounding co-exposure model does not require all the data, just a significant portion of the data from the most highly exposed workers.

This is a working document prepared by NIOSH's Division of Compensation Analysis and Support (DCAS) or its contractor for use in discussions with the ABRWH or its Working Groups or Subcommittees. Draft, preliminary, interim, and White Paper documents are not final NIOSH or ABRWH (or their technical support and review contractors) positions unless specifically marked as such. This document represents preliminary positions taken on technical issues prepared by NIOSH or its contractor. NOTICE: This report has been reviewed to identify and redact any information that is protected by the *Privacy Act 5 USC §552a* and has been cleared for distribution.

- Even the best radiation protection monitoring programs are seldom 100% effective. However, if a radiation protection program is working properly, monitoring missingness is most likely to affect workers with low doses, whereas data missingness can exhibit practically any pattern.

- To demonstrate that the dataset is "complete enough" to construct a co-exposure model one must, to some degree, vet the radiation protection programs that generated the data and the recordkeeping systems that store and report the data.

- After vetting the programs and starting under the presumption that the data are complete, a limited number of internal and external checks of the datasets can be performed to look for signs of significant data missingness or monitoring missingness.

- It may not be feasible to establish universally applicable, technically based quantitative limits for a dataset being complete enough. This is primarily a qualitative decision based on the preponderance of evidence.

- Regulatory compliance with a monitoring program or lack thereof cannot be used by itself to decide if a dataset is complete enough to construct an acceptable co-exposure model.

- Stratification of datasets cannot be used to correct for data missingness or monitoring missingness and can be very time consuming and resource intensive to perform. Furthermore, stratification of the datasets will likely not provide any significant benefit to the unmonitored worker to whom the co-exposure model is applied.

REFERENCES

10 C.F.R. 835, [2022]. Occupational radiation protection. [SRDB Ref ID: 195609]

Coggon D, Barker D, Rose G [1997]. What is epidemiology? In: Epidemiology for the uninitiated. 4th ed. London: BMJ Publishing Group. July 1. [SRDB Ref ID: 184328]

Lohr SL [2010]. Sampling: design and analysis. 2nd ed. Boston, MA: Brooks/Cole. [SRDB Ref ID: 139758]

NCRP [2018]. NCRP Report 178 - Deriving organ doses and their uncertainty for epidemiologic studies (with a focus on the one million U.S. workers and veterans study of low-dose radiation health effects). Bethesda, MD: National Council on Radiation Protection and Measurements. NCRP Report No. 178, November 20. [SRDB Ref ID: 193706]

NIOSH [2020]. Criteria for the evaluation and use of co-exposure datasets. Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention,

This is a working document prepared by NIOSH's Division of Compensation Analysis and Support (DCAS) or its contractor for use in discussions with the ABRWH or its Working Groups or Subcommittees. Draft, preliminary, interim, and White Paper documents are not final NIOSH or ABRWH (or their technical support and review contractors) positions unless specifically marked as such. This document represents preliminary positions taken on technical issues prepared by NIOSH or its contractor. NOTICE: This report has been reviewed to identify and redact any information that is protected by the *Privacy Act 5 USC §552a* and has been cleared for distribution.

National Institute for Occupational Safety and Health. DCAS-IG-006 Rev. 0, March 6. [SRDB Ref ID: 179686]

NIOSH [2022]. Centers for Disease Control National Institute for Occupational Safety and Health Advisory Board on Radiation and Worker Health Sandia National Laboratories Working Group Monday, April 11, 2022. Transcript. Washington, DC: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, Advisory Board on Radiation and Worker Health. [SRDB Ref ID: 194772]

ORAUT [2021]. Assessment of Los Alamos National Laboratory plutonium bioassay programs 1996 to 2001. Oak Ridge, TN: Oak Ridge Associated Universities Team. ORAUT-RPRT-0102 Rev 00, December 2. [SRDB Ref ID: 187245]

ORAUT [2023]. R code: plots for completeness white paper. Zip file. Oak Ridge, TN: Oak Ridge Associated Universities Team. January 17. [SRDB Ref ID: 195576]

Richardson DB, Wing S, Daniels RD [2006]. Evaluation of external radiation dosimetry records at the Savannah River Site, 1951–1989. J Expo Sci Environ Epidemiol 17(1):2–12. [SRDB Ref ID: 46009]

Richardson D, Wing S, Watson J, Wolf S [1999]. Missing annual external radiation dosimetry data among Hanford workers. J Expo Sci Environ Epidemiol 9(6):575–585. [SRDB Ref ID: 193549]

Rothman KJ, Greenland S, Lash TL [2008]. Modern epidemiology. 3rd ed. Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins. [SRDB Ref ID: 193705]

Soley-Bori M [2013]. Dealing with missing data: key assumptions and methods for applied analysis. Boston, MA: Boston University School of Public Health. Technical Report No. 4, May 6. [SRDB Ref ID: 194781]

Wang RY, Strong DM [1996]. Beyond accuracy: what data quality means to data consumers. J Manag Inf Syst 12(4):5–33. [SRDB Ref ID: 194780]

This is a working document prepared by NIOSH's Division of Compensation Analysis and Support (DCAS) or its contractor for use in discussions with the ABRWH or its Working Groups or Subcommittees. Draft, preliminary, interim, and White Paper documents are not final NIOSH or ABRWH (or their technical support and review contractors) positions unless specifically marked as such. This document represents preliminary positions taken on technical issues prepared by NIOSH or its contractor. NOTICE: This report has been reviewed to identify and redact any information that is protected by the *Privacy Act 5 USC §552a* and has been cleared for distribution.

APPENDIX A: ADVISORY BOARD WORKING GROUP TRANSCRIPTS

Page 50 of April 11, 2022 transcripts of ABRWH Sandia National Laboratories Working Group [NIOSH 2022]:

Member Roessler: SC&A has done a very detailed and thoughtful evaluation of this, and it's my impression that SC&A is pretty much in agreement with that conclusion. However, there's this thing that's always hanging over one of these discussions, and it's the data completeness. This kind of brings me then into my second comment. Bob mentioned that, I think expressed a desire for having a quantitative way to look at completeness, and boy that would really be helpful, because these discussions that take place in other, on other Sites too, are quite frustrating because SC&A will say well, we don't feel that there's data completeness, but it's left sort of vague. It's very hard for NIOSH to answer that and to come back then without some specific direction on, as to how to answer that question. Of course quantitative, if there a quantitative way, then that would make it much easier. So I guess my thought at this point is that the second part really doesn't affect the decision today, but it makes me think that we need -- it's an overreaching discussion, a separate discussion that affects other Sites. It seems like there should be maybe not for this Work Group but a special group within the Board to look at this particular question, as Bob brings it up. So that I think is kind of expresses my thoughts at this point in time.

This is a working document prepared by NIOSH's Division of Compensation Analysis and Support (DCAS) or its contractor for use in discussions with the ABRWH or its Working Groups or Subcommittees. Draft, preliminary, interim, and White Paper documents are not final NIOSH or ABRWH (or their technical support and review contractors) positions unless specifically marked as such. This document represents preliminary positions taken on technical issues prepared by NIOSH or its contractor. NOTICE: This report has been reviewed to identify and redact any information that is protected by the *Privacy Act 5 USC §552a* and has been cleared for distribution.