Response to SC&A report "Review of the Methods for Estimating Neptunium-237 Intakes at the Savannah River Site for the Period 1972-2007"

Prepared By Mike Mahathy, Matt Arno, and Elizabeth Brackett
Reviewed by Tim Taulbee, DCAS
1/9/2014

Finding 1: SC&A has concluded that NIOSH's method for comparing the measurements of two sets of workers requires the monitoring protocols of the two sets of workers to be the same. NIOSH has not established that there were protocols for whole-body count (WBC) monitoring of either non-construction trades workers (NCWs) or construction trades workers (CTWs) using WBCs during the 1972–1990 period, and if so, whether they were comparable. It appears unlikely that either group (as a whole) was routinely counted, except for fast scans in the mid-to late 1980s, which are not relevant to neptunium dose reconstruction.

Response:
From NIOSH's "Response to SC&A Comments on ORAUT-RPRT-0053":

The worker monitoring data were collected in the past to demonstrate compliance with the applicable occupational dose limits that were in place at the time. We are provided these *retrospective* data and are asked to perform statistical analyses on the data to answer questions being asked in the EEOICPA program today. We did not have the opportunity to select the workers and monitoring programs needed to ensure that we could develop definitive answers to these questions.

In most cases, we believe that the radiation protection staff who worked in the facilities when these retrospective data were collected made a concerted effort to monitor all individuals who they felt had a likelihood of receiving significant intakes. The monitored workers were intended to be a *census* rather than a *sample* (random or otherwise). Revision 4 of DPSOL 193-302, dated March 1971 (SRDB 126998) specifies annual chest counting for workers in the 221-H Sample Aisle and B-Line, 772-F, 221-F B-Line, JB-Line, 234-F and select personnel in the 221-F Sample Aisle and 773-A. During the era, chest and whole body counts were conducted simultaneously with the 40-cm arc, the difference between the two being the assumed location of the radionuclide(s) of interest. Neptunium urinalysis was performed when requested by Health Physics for CTWs and NCWs.

The goal of the coworker model is to estimate intakes for workers who were inadvertently not monitored. The implicit assumptions of the current coworker models are that unmonitored workers

- are more likely to be the workers who had lower potential for significant intakes than they are to be workers who had higher potential for significant intakes, and
- are not monitored *completely at random*.

If valid, these two assumptions ensure that any intakes assigned to unmonitored workers are conservative. In cases where the two assumptions given above may not be valid, the current practice is to assign the 95th percentile intake to minimize the chances of underestimating the dose to the worker.

Finding 2: It is not possible with the data available so far to establish that the same protocol for bioassay sampling was used for CTWs and NCWs in the 1991–2007 period. Establishing that equivalence is a necessary (though not sufficient) condition for the application of a coworker model based on "all-worker" data to CTWs.

Response:
There is no statistical requirement that all workers be on the same monitoring program in order to use the data to develop a coworker model as long as the monitoring programs adequately characterize all significant intakes. Further, most sites had *graded* monitoring programs where the frequency and types of bioassay performed were based on the likelihood of the workers having a significant intake of radioactive material[1]. Even today this is standard radiation protection practice, so we would expect the bioassay (i.e., sampling) protocols to be different for different groups of workers. Revision 5 of DPSOL 193-302, dated September 1, 1971 (SRDB 124941) specifies bioassay sampling frequencies for Construction Division workers and notes that an annual urine sample was requested from each employee. Specific sampling frequencies are provided for fission products (annual) and plutonium (triennial) and notes that frequencies for other radionuclides are as specified by Health Physics in the Construction Job Plans. These same sampling frequencies were used at least into the 1990's. See SRDB 45958.

Finding 3: NIOSH has not demonstrated that 30 samples in each comparison group (CTWs and NCWs) would be sufficient to simultaneously maintain low levels (for instance, less than 10% for each) of Type 1 and Type 2 errors in determining whether CTW and NCW sample distributions are the same (or not). The issue is moot in the case of neptunium only because the minimum number of 30 samples for the comparison to be valid is not available for CTWs in any year in the 1972–1990 period.

---

[1] Graded monitoring is also common in external dosimetry programs.

Response:
While this finding addresses Type 2 errors, Type 2 errors are directly related to the concept of power, so the following power discussion from NIOSH's "Response to SC&A Comments on ORAUT-RPRT-0053" is relevant:

The *a priori* power of a statistical test is usually considered during the design phase of the data collection procedure (e.g., the experiment or survey) so that the information collected is adequate to answer the questions being asked. In coworker modeling, we are presented with a predetermined dataset and cannot collect more, so it is not possible to perform an *a priori* power calculation. Even though an *a priori* power analysis could not be performed, we did make efforts to select the most powerful tests available. Based on our research the Peto-Prentice is the most powerful test available that can be used for comparing two groups with left-censored lognormal data, and while the power of this test will vary depending on the actual data used, there is no better statistical test that can be used for this purpose.

To perform an *a priori* power analysis, an acceptable level of power $1 - \beta$ has to be defined. To define β we must first define the *size of the effect*[2] that we want to detect, i.e., the size of the effect that is of *practical significance*. If we could define practical significance (we tried and were unsuccessful), we would perform an *equivalence test* [Streiner 2003], which tells us if the difference in the two groups is of practical significance, rather than a *null-hypothesis test*, which tells us if the difference in the two groups is of statistical significance.

The broad issue of practical significance is being addressed by the ABRWH SEC Workgroup and an evaluation is under way to address this issue generically.

Finding 4: NIOSH's method for concluding that CTW measurements were drawn from the same distribution as those for NCWs does not meet the minimum technical requirements for 30 samples from each of the groups being compared for the 1972–1990 period. There were fewer than 30 total CTW samples in each year in the 1972–1990 period for which NIOSH reported data; the number of above minimum detectable activity (MDA) results ranged from just 6 to a high of 17 for the same period. As a result, the coworker model is based on a comparison that does not satisfy the minimum data adequacy criteria.

Response:
From RPRT-0053, "Analysis of Stratified Coworker Datasets":

[2] The magnitude of the difference between the two groups.

As a general guideline, the minimum sample size used for coworker modeling is 30 individuals (i.e., 30 OPOS results) in a given period[3]. This minimum[4] can be relaxed if, in the judgment of the statistician performing the analysis, the uncertainty in the resulting parameter estimates is not excessive. The minimum sample size applies to a *sample of the monitored workers* and does not apply to the *population of monitored workers*; for example, if the entire population of monitored workers at a given site in a given year is 25, the coworker model derived from this population is valid.

Since the general guideline of 30 OPOS results was not met for CTW in any period between 1972 and 1990, and because the analysis of interest was statistical tests, analysis was done based on footnote 4 from RPRT-0053 (see above). In any year where there were ~15 OPOS results for CTW and less than 80% censoring, statistical tests were performed at the discretion of the statistician.

Finding 5: The coworker model is dominated by the routine monitoring data of NCWs and may not adequately represent the exposure experience of at least some job types among CTWs.

Response:
From "Response to SC&A Comments on ORAUT-RPRT-0053":
We agree with the statement "*If the unmonitored workers are from a different population, the applicability of a coworker model derived from monitored coworkers would be in question."* This relates back to the issue of stating why unmonitored workers were not monitored. In the development of coworker models we assume that either:

- unmonitored individuals are members of the monitored population who were not monitored completely at random, or
- unmonitored individuals were unmonitored because they had no potential for exposure to radioactive materials.

In the first case we have the right model and in the second a conservative model. One can also theorize that these assumptions are wrong and that perhaps unmonitored workers were highly exposed and intentionally not monitored because of this. This fundamental and largely unstated difference in assumptions probably needs to be discussed and eventually resolved.

[3] Data from multiple years (usually no more than 3) can be combined to achieve this minimum if the conditions in the workplace are reasonably constant over the period in question.
[4] The U.S. Environmental Protection Agency (Singh, Armbya, and Singh 2010, p. 27) discusses minimum sample size required for performing statistical tests on censored datasets and recommends ~15 results per sample (stratum) as a minimum. Here we are estimating parameters from the data, so the default minimum has been increased to 30.

Finding 6: The NIOSH coworker model for CTW exposure to neptunium for the period 1972–1990 would often lead to results that are very claimant unfavorable.

Response:
From "Response to SC&A Comments on ORAUT-RPRT-0053":
The concept of "claimant favorability" expressed in this comment is confusing. To illustrate the problem, assume Group A and Group B have very different dose distributions, with A higher than B, and we combine the doses from the two groups to form Group C. The dose distribution for C will overestimate the dose to B (be "claimant favorable") and underestimate the dose to A (not be "claimant favorable"). But, we typically assume that workers with higher doses are less likely to be unmonitored, so this is not considered a problem in practice. Now, if we stratify the model, the doses to both A and B will be more accurate. However, the dose to B will be lower and less "claimant favorable" and the dose to A will be higher and more "claimant favorable" than the combined dose distribution. So, in principle, no coworker model(s) can be "claimant favorable" to all strata in the model at the same time.

From the late 1960s through 1977 SRS performed routine plutonium analysis on workers in the HB Line and 235-F. DPSOL- 47-137 (SRDB 129514) states that neptunium analyses were performed when requested by area Health Physics but neptunium had never been detected without at least an equal amount of plutonium. Examination of Work Technical Reports prior to 1978 showed that Pu-238 contamination in Np-237 feed for the billet line was at least equal to the activity of Np-237. In January 1978, workers in the 235-F billet line were added to annual neptunium bioassay while other workers in 235-F and HB Line remained on plutonium bioassay (SRDB 45958). Fabrication of neptunium billets and Pu-238 fuel forms at SRS ceased in December 1983 though the 235-F facilities were left in "enhanced production readiness" status from 1985 through 1990 with maintenance activities being performed by just three operators (SRDB 120782). In 1989 all workers in HB Line and 235-F were placed on routine, annual neptunium bioassay (SRDB 45958) though no work with neptunium was performed until 2004 when surplus neptunium fuel tubes were processed in the HB Line (SRDB 124003; 124025; 128540). Construction workers were not involved in operations at HB Line or 235-F when neptunium was being processed but were bioassayed as requested by Health Physics in Construction Job Plans (SRDB 129514; 45958).

Finding 7: A statistically valid analysis comparing neptunium exposure potential of CTW job types or of relative CTW exposure potential in various Savannah River Site (SRS) areas is not possible for the 1972–1990 period due to insufficient data.

Response:
NIOSH agrees, that there is not much data. Also, see response to *Finding 8*.

Finding 8: A statistically valid analysis comparing neptunium exposure potential of subgroups of CTWs or of relative CTW exposure potential in various SRS areas is not possible at the present time for the 1991–2007 period since the data for that period have not been separated into CTW and NCW categories.

Response:
From NIOSH's "Response to SC&A Comments on ORAUT-RPRT-0053":

One needs to be careful in the analysis of data to avoid what has been called "data dredging." One definition[5] of "data dredging" involves the practice of using the same set of data to both

- form a hypothesis to test, and
- subsequently test the hypothesis.

In other words, if NIOSH does not think it is appropriate to randomly sift through data looking for significant differences and then use the same data to test for those differences, we are bound to find significance more often than we should. Data dredging is strongly frowned on because the probability of falsely identifying significant differences is not readily quantified or controlled. Thus, any conclusions reached after data dredging are of dubious statistical or practical value. The preferable approach is to identify all strata to be tested prior to looking at the data and then test those strata for significant differences.

During the 1990s and after, when bioassay data are stored in the HPRED database, stratification of the bioassay data into CTW and NCW is possible for the more common radionuclides for which bioassay data were obtained. This can be done if desired.

Finding 9: NIOSH has given preference to results in the I-131 region obtained using 40-cm arc geometry over results obtained using chest geometries to calculate Np-237 body burdens and intakes, in the period when results were reported in units of count rate (first reporting format, described in Section 5.1.1). NIOSH did not explain why in-vivo results obtained from the 40-cm

[5] http://en.wikipedia.org/wiki/Data_dredging

arc geometry were preferred over results obtained using chest counts geometry to derive Np-237 intakes.

Response:

The preference is to use results obtained from the I-131 region of the spectrum rather than from a specific geometry. The in-vivo measurements identified as Np-237 are based on the 230-290 keV region of interest as shown in Figure 1 of the SC&A report. Np-237 has five gamma rays in this region with a summed abundance of 0.08%. Pa-233 has an additional two gamma rays with a summed abundance of 0.06%. In contrast, Pa-233 has 5 gamma rays with a summed abundance of 51.8% in the 300-400 keV region of interest. This abundance is two orders of magnitude greater, which allows for MDAs two orders of magnitude smaller. The gamma ray abundance from Np-237 in the 230-290 keV region of interest is simply too low to allow for reasonable use of this data.

SRBD 10931, *A History of Personnel Radiation Dosimetry at the Savannah River Site*, pages 62 and 63 describe the 40 cm arc geometry. Later pages describe the stretcher geometry. Page 62 has a figure of the NaI detector used in the 40 cm arc geometry. As can be seen, this detector is centered over the chest. Both “chest” and “whole body” counts were thus obtained simultaneously. The difference lay in the calibration factors used and the assumed location in the body of the radionuclides of interest. It was common to report both “chest” counts and “whole body” counts on the same form. Later forms distinguish between the two, but earlier forms do not necessarily do so.

Finding 10: For the years in which in-vivo results files present gross, background, net, CALC and DIFF values for each region of interest (ROI) (second reporting format, analyzed in Section 5.1.2), NIOSH assumes that either a 40-cm arc or a stretcher geometry was used to obtain those results. SC&A did an extensive review of worker data for this period to check this assumption. This is important, because conversion factors and the equations that NIOSH derived to quantify Np-237 are based on the assumption that a certain geometry was used. They may not be valid if the assumption is incorrect in some part of the data that NIOSH has compiled to make neptunium body burden estimates. While the in-vivo monitoring results in this second format (Section 5.1.2), analyzed by SC&A, do not specify which geometry was used, all of the reporting formats in all results reviewed indicate that they were obtained from chest monitoring. SC&A recognizes that only a small fraction of in-vivo results were analyzed, but they included results from a random sample of 30 workers (sample 3) and from 44 workers in areas where Np-237 contamination could occur (samples 1 and 2). The analysis of this data points to the conclusion that most workers were monitored using the chest geometry. NIOSH should address the implications of this finding for its estimation method.

Response:
SRBD 10931, *A History of Personnel Radiation Dosimetry at the Savannah River Site,* pages 62 and 63 describe the 40 cm arc geometry. Later pages describe the stretcher geometry. Page 62 has a figure of the NaI detector used in the 40 cm arc geometry. As can be seen, this detector is centered over the chest. Both "chest" and "whole body" counts were thus obtained simultaneously. The difference lies in the calibration factors used and the assumed location in the body of the radionuclides of interest. It was common to report both "chest" counts and "whole body" counts on the same form. Later forms distinguish between the two, but earlier forms do not necessarily do so.

Finding 11: During the 1980–1989 time period, NIOSH proposes to use the net counts in the Cr-51 region to derive Np-237 body burdens. All in-vivo files using the third reporting format as described in Section 5.1.3 during this time period also report chest count results. SC&A's review included an in-vivo result specifically referring to Np-237 in relation to an inhalation intake in September 1988. The MDA of Np-237 was calculated using chest counts and the 86.5 keV Np-237 photon. SC&A does not agree with NIOSH's assessment that 86.5 keV photons are unsuitable for estimating Np-237. They can be and are so used.

Response:
The third reporting format, IVCR forms, clearly distinguishes between chest counts and whole body counts and report both on the same form. It is apparent that both types of counts were performed.

The Np-237 MDAs referenced in Section 5.1.3 of SC&A's report are from 1993 when germanium detectors were in use. It is inappropriate to apply peak discrimination and quantification capabilities of germanium detectors to those obtained when using NaI or phoswich detectors.

Finding 12: NIOSH has suggested the use of WBC results not explicitly geared to quantify Np-237 to build a coworker model for Np-237. SC&A's review shows that workers in areas where exposure to Np-237 was possible were not always monitored through in-vivo bioassay. SC&A even found examples of workers with accidental intakes of Np-237 who did not have in-vivo counts after detection of an intake.

Response:
The basic premise of having and using coworker intake rates is that not all workers were monitored that should have been monitored by modern standards. Therefore, it is expected that

there were workers in areas where exposure to Np-237 was possible that were not monitored. That is why Np-237 coworker intakes are evaluated. SRS performed monitoring for Np-237 based on the professional judgment of the Health Physicists and noted that Np-237 had never been detected without an equal amount of plutonium (SRDB 126998, 45958).

Finding 13: NIOSH's assumption of equilibrium between Pa-233 and Np-237 in making its estimates of Np-237 body burdens is likely to be claimant unfavorable in some or many cases.

Response:
The coworker studies are based on the premise of chronic intakes occurring for at least a year's duration. Chronic intakes are used as an integrated approximation of the intakes that may occur from multiple acute intakes. Calculation of intakes are based on bioassay dates in the middle of the year, 6 months after commencement of the intake. If a worker were to chronically inhale freshly-separated Np-237 (i.e., no Pa-233 ingrowth has occurred), after six months of such inhalation, the Pa-233 would be at 72% of equilibrium with the Np-237 in the worker's body. At one year, the equilibrium ratio is 84%, and it is 91% at two years. The calculated coworker intakes are based on multi-year intakes. Even assuming 72% equilibrium represents less than a 40% underestimate in the magnitude of the Np-237 intake, much less than the minimum factor of 3 uncertainty applied to coworker intakes. This is based on the conservative assumption that the intake was always fresh Np-237; any ingrowth of Pa-233 prior to intake would increase the degree of equilibrium. In actuality, most Np-237 intakes are more likely to be aged material characteristic of surface contamination on objects that may be aged weeks, months, or longer. Even though Np-237 billets were normally processed approximately 25 days after production, some occasions were longer and residual contamination in the processing line would be expected to be considerably older than the age of the billets. Under these conditions, assumption of equilibrium is reasonable and any uncertainty is accounted for with the minimum factor of 3 uncertainty applied.

Finding 14: NIOSH has stated that the age of Np-237 in an employee's body at the time of the in-vivo count could be determined, if needed, but has not demonstrated that this can be done. SC&A's review of in-vivo data indicates that this may be difficult or impossible in many or most cases since the times and places of neptunium exposure are generally not expressly indicated in in-vivo records showing I-131 and Cr-51 counts.

Response:
The precise age of the Np-237 in an employee's body is not needed if sufficient assurance can be obtained that it has aged sufficiently to establish a reasonable equilibrium with Pa-233. As

discussed above, the coworker studies are based on the premise of chronic intakes occurring for at least a year's duration. Chronic intakes are used as an integrated approximation of the intakes that may occur from multiple acute intakes. Calculation of intakes are based on bioassay dates in the middle of the year, 6 months after commencement of the intake. If a worker were to chronically inhale freshly-separated Np-237 (i.e., no Pa-233 ingrowth has occurred), after six months of such inhalation, the Pa-233 would be at 72% of equilibrium with the Np-237 in the worker's body. At one year, the equilibrium ratio is 84%, and it is 91% at two years. The calculated coworker intakes are based on multi-year intakes. Even assuming 72% equilibrium represents less than a 40% underestimate in the magnitude of the Np-237 intake, much less than the minimum factor of 3 uncertainty applied to coworker intakes. This is based on the conservative assumption that the intake was always fresh Np-237; any ingrowth of Pa-233 prior to intake would increase the degree of equilibrium. In actuality, most Np-237 intakes are more likely to be aged material characteristic of surface contamination on objects that may be aged weeks, months, or longer. Even though Np-237 billets were normally processed approximately 25 days after production, some occasions were longer and residual contamination in the processing line would be expected to be considerably older than the age of the billets. Under these conditions, assumption of equilibrium is reasonable and any uncertainty is accounted for with the minimum factor of 3 uncertainty applied.

Finding 15: The distributions on which the coworker models are likely based are a mixture of results with varying errors of an undetermined degree built into them due to lack of knowledge of the disequilibrium between Np-237 and Pa-233. The means and geometric standard deviations (GSDs), based on an assumption of equilibrium, have indeterminate errors, which would make the body burden estimates as they stand at present not statistically meaningful. An approach that takes the Np-237/Pa-233 disequilibria into account is needed to determine a scientifically valid distribution for neptunium coworker models based on in-vivo data; that is, for a valid coworker model for the 1972–1990 period.

Response:

As discussed in the response to Finding 13, quantification of the exact degree of disequilibrium is not needed because the amount of disequilibrium is limited, making the assumption of equilibrium reasonable. Therefore, the approach that has been taken is valid.

Finding 16: Assigning Np-237 doses to workers who had I-131/Cr-51 in-vivo counts but no exposure potential to Np-237 is scientifically unreasonable, even though it may provide a claimant-favorable result.

Response:
Np-237 doses are only assigned to those workers with some potential for intake. Whether a worker had potential exposure to I-131 or Cr-51 is not relevant. Workers are not monitored specifically for I-131 or Cr-51. They are monitored for these radionuclides by whole body counts. The proper question is whether workers were monitored by whole body counts for any reason. As mentioned in SC&A's report one paragraph prior to this finding, many of the workers SC&A analyzed did work in areas where "contamination [sic] to Np-237 was possible." Many people included in a coworker study may have been designated as having some potential for intakes but in reality had none; their results are included in the studies because there is no way to distinguish between them when the results are <MDA.

Finding 17: NIOSH has not described how it would estimate the dose of workers who had exposure potential for Np-237, when there are no relevant WBC data available.

Response:
The discussion leading to this finding does not support the finding. As mentioned in SC&A's report two paragraphs prior to this finding, many of the workers with WBCs that SC&A analyzed did work in areas where "contamination [sic] to Np-237 was possible."

Finding 18: There are sharp discontinuities between the intakes estimated using in-vivo data and urinalysis, with the former being much higher than the latter. NIOSH has not explained whether these discontinuities resulted from sudden changes in actual exposure conditions. If not, the discontinuities may indicate a problem with NIOSH's approach of using WBC data to estimate neptunium body burdens and intake rates as described in ORAUT-RPRT-0056 (ORAUT 2012a), Equations 1 through 4.

Response:
SC&A's observation is acknowledged. Such discontinuities occur frequently whenever bioassay methods change, whether from in-vivo to in-vitro methods or a change in the in-vitro or in-vivo method used, especially when the change results in a large change in the MDA. Changes in the MDA of the analytical method used are known to result in changes in the calculated intakes even when no change in actual exposure conditions is suspected.

In this case, the discontinuities do indicate that the in-vivo-based intake rates are overestimates of the actual intake rate and that this overestimate is due to the higher MDAs associated with the in-vivo counts. The fact that it is an overestimate does not de facto rule it out as a reasonable estimation of the intakes. Although the intake rates based on the in-vivo data are higher than those based on the in-vitro data, the resulting annual doses are still low for most organs. The committed equivalent doses from the Np-237 intakes due to chronic intakes from 1970 through 1979 (those based on WBC) are given in the table below. Most doses are a few rem or less except for bone surface and red bone marrow. During an actual dose reconstruction, these doses would be even lower due to the shorter commitment period.

Organ	50-Year Equivalent Dose (rem)
Adrenals	3.54E-01
Urinary Bladder	3.52E-01
Breast	3.52E-01
Heart Wall	3.53E-01
Kidneys	9.58E-01
Liver	4.28E+00
Muscle	3.53E-01
Ovaries	3.58E+00
Pancreas	3.53E-01
Testes	3.64E+00
Thyroid	3.52E-01
R.B.M.	1.03E+01
Bone Surface	2.68E+02
Stomach	3.52E-01
S.I.	3.53E-01
U.L.I.	3.56E-01
L.L.I.	3.62E-01
Skin	3.52E-01
Spleen	3.52E-01
Thymus	3.52E-01
Uterus	3.52E-01
Lung	6.92E+00
ET2	5.36E+00
LN(ET)	5.69E-01
LN(TH)	1.36E+00
Esophagus	3.52E-01

Finding 19: There are sharp discontinuities between the body burdens calculated using different equations, as described in ORAUT 2012a, with the body burden being much higher when equation (1) was used. The body burdens calculated using equation 4 are 5 to 10 times lower than the ones calculated using equations 2 and 3. Unless NIOSH can show that the discontinuities are related to exposure changes, the discontinuities are too large for the WBC estimation of Np-237 as proposed by NIOSH to be acceptable.

Response:
A factor of ten drop in calculated intake rate is not uncommon in the coworker studies. As stated in the response to Finding 18, when the MDA of the bioassay method used decreases, the calculated intake rate typically drops as well and by a commensurate amount. In this case, the whole body counting methods improved with time with more and better detectors, leading to a decrease in the MDAs for most radionuclides. The "discontinuities" noted are not remarkable.