Draft

ADVISORY BOARD ON
RADIATION AND WORKER HEALTH

National Institute for Occupational Safety and Health

SC&A'S EVALUATION OF ORAUT-RPRT-0086, REVISION 00, "INTERNAL DOSIMETRY COWORKER DATA COMPLETENESS TEST"

Contract No. 211-2014-58081
SCA-TR-2018-PR001, Revision 0

Prepared by

Ron Buchanan, PhD, CHP
Harry Chmelynski, PhD

SC&A, Inc.
2200 Wilson Boulevard, Suite 300
Arlington, VA 22201-3324

Saliant, Inc.
5579 Catholic Church Road
Jefferson, Maryland 21755

January 2018

SC&A, INC.: ***Technical Support for the Advisory Board on Radiation and Worker Health Review of NIOSH Dose Reconstruction Program***

DOCUMENT TITLE:	SC&A's Evaluation of ORAUT-RPRT-0086, Revision 00, "Internal Dosimetry Coworker Data Completeness Test"
DOCUMENT NUMBER/ DESCRIPTION:	SCA-TR-2018-PR001
REVISION NO.:	0 (Draft)
SUPERSEDES:	N/A
EFFECTIVE DATE:	January 12, 2018
TASK MANAGER	Ron Buchanan, PhD, CHP [signature on file]
PROJECT MANAGER:	John Stiver, MS, CHP [signature on file]
DOCUMENT REVIEWER(S):	Kathy Behling [signature on file] John Stiver, MS, CHP [signature on file]

Record of Revisions

Revision Number	Effective Date	Description of Revision
0 (Draft)	01/12/2018	Initial issue

TABLE OF CONTENTS

Abbreviations and Acronyms 4

1 Introduction and Background 5

2 Overview of ORAUT-RPRT-0086 5

3 SC&A'S Evaluation of ORAUT-RPRT-0086 7

3.1 Findings 7

3.2 Observations 7

4 Summary and Conclusions 8

5 References 8

ABBREVIATIONS AND ACRONYMS

ABRWH	Advisory Board on Radiation and Worker Health
AQL	acceptable quality level
α	alpha
β	beta
CI	confidence interval
DCAS	Division of Compensation Analysis and Support
LTPD	lot tolerance percent defective
$\bar{M}$	the mean number of results in the n clusters
M_i	number of results (records) in ith cluster
n	sample size (number of clusters)
N	number of clusters (claims) in population
NIOSH	National Institute for Occupational Safety and Health
NOCTS	NIOSH-DCAS Claims Tracking System
ORAUT	Oak Ridge Associated Universities Team
RPRT	report
SE	standard error
y_i	proportion of missing data in ith cluster
$\hat{\bar{y}}_r$	the proportion (or fraction) of the data that is missing in the sample

1 INTRODUCTION AND BACKGROUND

In October 2017, the Advisory Board on Radiation and Worker Health tasked SC&A with a technical review of ORAUT-RPRT-0086, *Internal Dosimetry Coworker Data Completeness Test*, Revision 00, issued September 18, 2017 (NIOSH 2017, referred to as "RPRT-0086"). In RPRT-0086, the National Institute for Occupational Safety and Health (NIOSH) presents statistical methods that analysts can (1) use to select a sample of people from the NIOSH-DCAS Claims Tracking System (NOCTS) records and (2) use the data to calculate an interval estimate of the proportion of missing data for the population of all monitored workers at a site. These methods are generic because many of the details are site dependent. The specific procedures will have to be tailored to the site and documented in a site-specific report.

This report presents SC&A's evaluation of RPRT-0086.

2 OVERVIEW OF ORAUT-RPRT-0086

RPRT-0086 is a detailed document and, for evaluation purposes, it is advantageous to provide a brief outline of its contents, as follows:

- **Terms** – The following terms are used in RPRT-0086:
  - NOCTS dataset – dosimetry data in the form of electronic copies of the hardcopy records; these data are not usually in a computer-readable form
  - Source dataset – the dataset from the site, usually in the form of copies of hardcopy documents, or the NOCTS dataset for claimants
  - Original dataset – the electronic dataset usually transcribed from a source dataset in computer-readable form
  - Sampling frame – the set of data that has entries (in RPRT-0086 this is the claimants and their associated bioassay data) in both the source dataset and the original datasets
  - Cluster – a claim number with its associated bioassay results when using NOCTS as the source dataset
- **Summary of Steps** – NIOSH recommended processes to estimate the fraction (referred to in RPRT-0086 as the proportion) of the missing data in the original dataset compared to the source dataset. Section 9.0 (pages 17–18) of RPRT-0086 briefly outlines the steps of these processes. To simplify the overall process, SC&A has summarized the steps recommended in RPRT-0086 to estimate the fraction of missing data in the electronic dataset.
  1) Assemble the sampling frame – This consists of comparing the entries in the source dataset (RPRT-0086 used claim numbers from the NOCTS records) to the corresponding entries in the original (i.e., electronic) dataset. The sampling frame consists of those entries (e.g., claims) with data in both datasets.

2) Obtain the list of entries – In the case of using NOCTS records as the source dataset, obtain a list of the claim numbers and bioassay records for each claim that is present in the sampling frame.

3) Determine the required sample size (n) – The procedures recommended in ORAUT-RPRT-0078, *Technical Basis for Sampling Plan* (NIOSH 2016, referred to as "RPRT-0078"), are adapted in RPRT-0086 with some modifications as discussed in Section 5.0 (pages 8–15). The modifications are necessary because each claim (i.e., cluster) in NOCTS has a unique number of bioassay data results instead of simply a spreadsheet of entries as used in RPRT-0078. The required sample size will, in part, be determined by the choice of the following parameters, as stated in RPRT-0086 (pages 8–9):

   - *AQL = 0.025; the acceptable error rate or acceptable quality level (AQL), which is the percentage of defects at which the consumer is willing to accept the lot as good.*
   - *LTPD = 0.05; the unacceptable error rate or lot tolerance percent defective (LTPD), which is the upper limit of the percentage of defects in a lot that the consumer is willing to accept.*
   - *α = 0.025; the producer's risk, which is the probability that a good lot containing defects equal to AQL will be rejected on the basis of sample data.*
   - *β = 0.025; the consumer's risk, which is the probability that a bad lot containing defects equal to LTPD will be accepted on the basis of sample data.*

   A computer simulation is outlined on page 10 of RPRT-0086 to assist in the construction of a distribution plot (as illustrated in Figure 5-2 of RPRT-0086, page 11) for deriving an approximately sample size for a given set of parameters.

   The required sample size is further refined in Section 5.2 (pages 12–13) of RPRT-0086 by use of regression analyses (as illustrated in Figure 5-3 of RPRT-0086, page 13) and an operational characteristic curve (as illustrated in Figure 5-4 of RPRT-0086, page 13). The required sample size is also restricted by other considerations as discussed in Section 5.3 (pages 14–15), including the statistical consideration that the sample size must be 30 or greater. Additionally, compensating for the granularity of missing data must be considered. The procedure outlined to this point assumes that the missing data are distributed randomly among the claims. However, the missing data may be compressed into only a few claims. Therefore, Section 7.0 of RPRT-0086 (pages 15–17) provides a method to address the possible granularity in the missing data. The construction of a sequential plot to aid selection of the appropriate sample size is illustrated in Figure 7-1 of RPRT-0086 (page 17).

4) Determine the proportion of data missing – The proportion (or fraction) of the data that is missing in the sample ($\hat{\bar{y}}_r$) of the original dataset is estimated by using Equation (4-1) of RPRT-0086 (page 7):

$$\hat{\bar{y}}_r = \frac{\sum_{i=1}^{n} M_i \bar{y}_i}{\sum_{i=1}^{n} M_i}$$

where

n = number of clusters (claims) in sample
M_i = number of results (records) in ith cluster
y_i = proportion of missing data in ith cluster
$\bar{y}_i$ = mean result (proportion of records missing) for the ith cluster

The standard error is given by Equation (4-2) of RPRT-0086 (page 7):

$$SE(\hat{\bar{y}}_r) = \sqrt{\left(1 - \frac{n}{N}\right)\left(\frac{1}{n\bar{M}^2}\right)\left(\frac{1}{n-1}\right)\sum_{i-1}^{n} M_i^2(\bar{y}_i - \hat{\bar{y}}_r)^2}$$

where

N = number of clusters in population
$\bar{M}$ = the mean number of results in the n clusters.

5) Determine the 95% confidence interval (CI) for the proportion of data missing by using Equation (4-3) of RPRT-0086 (page 8):

$$95\%\ \text{CI} = \hat{\bar{y}}_r \pm t_{0.975,df} SE(\hat{\bar{y}}_r)$$

where the degrees of freedom in the Student's t quantile equals the number of clusters minus 1 (i.e., $df = n - 1$). The CI is discussed further in Section 6.0 of RPRT-0086.

3 SC&A'S EVALUATION OF ORAUT-RPRT-0086

The following is a summary of SC&A's evaluation of RPRT-0086.

3.1 FINDINGS

SC&A did not identify any specific findings about the generic approach and statistical methods used in RPRT-0086 to perform a completeness test for internal dosimetry coworker data. To determine the usability of the process recommended in RPRT-0086, site-specific procedures will need to be developed and evaluated.

3.2 OBSERVATIONS

SC&A did have the following observations that may require discussion or clarification concerning RPRT-0086:

- **Observation 1:** SC&A would like to emphasize that some of the parameters used in RPRT-0086 are variable, and the selection of their values will determine the required sample sizes and may affect the outcome of the analyses. Some of the parameters that are variable are:
  - Parameter 1: Producer's risk α (NIOSH recommended the use of 0.025, i.e., 2.5%)
  - Parameter 2: Consumer's risk β (NIOSH recommended the use of 0.025, i.e., 2.5%)
  - Parameter 3: Acceptable error rate (AQL) (NIOSH recommended the use of 0.025, i.e., 2.5%)
  - Parameter 4: Unacceptable error rate (or LTPD) (NIOSH recommended the use of 0.05, i.e., 5%)

- **Observation 2:** The term "original dataset" is used throughout the document to refer to the computer-readable dataset in electronic form that has been transcribed from the hardcopy records, or from the electronic reproduction of the hard copies. The term "original" generally refers to origin or first, such as the hardcopy records. Therefore, a different term for the electronic dataset would be less confusing when reading RPRT-0086.

- **Observation 3:** The last paragraph on page 11 states:

  > *This process is illustrated in Figure 5-3, where, for example, the critical values for* n *= 25 are those shown in Figure 5-2.*

  According to the caption for Figure 5-2, *n* = 24, not 25.

4 SUMMARY AND CONCLUSIONS

SC&A did not identify any specific findings about the generic approach and statistical methods used in RPRT-0086 to perform a completeness test for internal dosimetry coworker data. To determine the usability of the process recommended in RPRT-0086, site-specific procedures will need to be developed and evaluated.

SC&A did identify several areas that could use discussion or clarification; Section 3.2 lists these as Observations 1–3.

5 REFERENCES

NIOSH 2016. *Technical Basis for Sampling Plan*, ORAUT-RPRT-0078, Revision 00, National Institute for Occupational Safety and Health, Cincinnati, Ohio. June 17, 2016.

NIOSH 2017. *Internal Dosimetry Coworker Data Completeness Test*, ORAUT-RPRT-0086, Revision 00, National Institute for Occupational Safety and Health, Cincinnati, Ohio. September 18, 2017.