

THESIS

EXTERNAL DOSE ASSESSMENT IN THE UKRAINE FOLLOWING THE  
CHERNOBYL ACCIDENT

Submitted by

Remi Jordan Lesartre Frazier

Department of Environmental and Radiological Health Sciences

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Spring 2014

Master's Committee:

Advisor: Thomas Borak

Alex Brandl  
John Harton

## ABSTRACT

### EXTERNAL DOSE RECONSTRUCTION IN THE UKRAINE FOLLOWING THE CHERNOBYL ACCIDENT

While the physiological effects of radiation exposure have been well characterized in general, it remains unclear what the relationship is between large-scale radiological events and psychosocial behavior outcomes in individuals or populations. To investigate this, the National Science Foundation funded a research project in 2008 at the University of Colorado in collaboration with Colorado State University to expand the knowledge of complex interactions between radiation exposure, perception of risk, and psychosocial behavior outcomes by modeling outcomes for a representative sample of the population of the Ukraine which had been exposed to radiocontaminant materials released by the reactor accident at Chernobyl on 26 April 1986.

In service of this project, a methodology (based substantially on previously published models specific to the Chernobyl disaster and the Ukrainian population) was developed for daily cumulative effective external dose and dose rate assessment for individuals in the Ukraine as a result of the Chernobyl disaster. A software platform was designed and produced to estimate effective external dose and dose rate for individuals based on their age, occupation, and location of residence on each day between 26 April 1986 and 31 December 2009. A methodology was developed to transform published  $^{137}\text{Cs}$  soil deposition contour maps from the *Comprehensive Atlas of Caesium Deposition on Europe after the Chernobyl Accident* into a geospatial database to access these data as a radiological source term.

Cumulative effective external dose and dose rate were computed for each individual in a 703-member cohort of Ukrainians randomly selected to be representative of the population of the country as a whole. Error was estimated for the resulting individual dose

and dose rate values with Monte Carlo simulations. Distributions of input parameters for the dose assessment methodology were compared to computed dose and dose rate estimates to determine which parameters were driving the computed results.

The mean external effective dose for all individuals in the cohort due to exposure to radiocontamination from the Chernobyl accident between 26 April 1986 and 31 December 2009 was found to be 1.2 mSv; the geometric mean was 0.84 mSv with a geometric standard deviation of 2.1. The mean value is well below the mean external effective dose expected due to typical background radiation (which in the United States over this time period would be 12.0 mSv).

Sensitivity analysis suggests that the greatest driver of the distribution of individual dose estimates is lack of specific information about the daily behavior of each individual, specifically the portion of time each individual spent indoors (and shielded from radionuclides deposited on the soil) versus outdoors (and unshielded).

# TABLE OF CONTENTS

ABSTRACT .....	ii
TABLE OF CONTENTS .....	iv
LIST OF TABLES .....	viii
LIST OF FIGURES .....	x
1 INTRODUCTION .....	1
2 BACKGROUND .....	4
2.1 The disaster at Chernobyl .....	4
2.2 Previous efforts to characterize the disaster’s impact .....	6
2.2.1 Production of the ATLAS .....	6
2.2.2 Availability of <sup>137</sup> Cs source term raw data .....	8
2.3 Motivation for our study .....	9
2.3.1 Nuclear disasters and psychosocial outcomes .....	9
2.4 Structure of our study .....	10
2.4.1 Project oversight, personnel, and responsibilities .....	10
2.4.2 Project workflow .....	11
2.4.3 Goals of the survey instrument .....	13
2.4.4 Format of the survey dataset .....	14
2.4.5 Indeterminate ordering of residence/occupation records .....	16
2.5 Terminology specific to this document .....	17
2.5.1 Time .....	17
2.5.2 Location .....	17
2.5.3 Radiation physics .....	17
2.5.4 Statistics .....	18
3 MATERIALS AND METHODS .....	20

3.1	Executive summary .....	20
3.2	Methods employed.....	22
3.2.1	Constructing and populating the dose reconstruction database.....	22
3.2.2	Transforming radiocontamination maps into a spatial database .....	23
3.2.3	Computing effective dose rate at an arbitrary time $t$ .....	28
3.2.3.1	Determining demographic information .....	29
3.2.3.2	Recovering initial indicator source term from geodatabase.....	29
3.2.3.3	Finding an effective indicator source term .....	29
3.2.3.4	Finding kerma in air at 1 m from all radionuclides.....	32
3.2.3.5	Finding uncorrected effective dose rate .....	36
3.2.3.6	Finding corrected effective dose rate .....	41
3.2.4	Computing cumulative effective dose.....	42
3.2.5	Accommodating indeterminate residence/occupation record ordering.....	44
3.2.6	Computing summary statistics .....	45
3.2.7	Quality management system.....	46
3.2.7.1	Source data validation .....	46
3.2.7.2	Quality assurance.....	47
3.2.7.3	Quality control .....	48
3.3	Materials used .....	53
4	RESULTS AND DISCUSSIONS .....	56
4.1	Characteristics of the data produced by the survey instrument. ....	56
4.1.1	Demographic distribution .....	56
4.1.1.1	Gender distribution .....	56
4.1.1.2	Distribution of age and occupation.....	57

4.1.2	Number of unique residences .....	59
4.1.3	Number of listed residence/occupation changes .....	60
4.1.4	Number of possible residence/occupation orderings .....	60
4.2	Results of dose reconstruction.....	61
4.2.1	Summary of results.....	61
4.2.2	Change in distribution of dose in the cohort over time .....	63
4.2.3	Comparison to background radiation.....	64
4.2.4	Normality of the distribution of dose .....	65
4.2.5	Dynamics of cumulative dose distribution over time .....	68
4.3	Outcomes for individuals.....	69
4.3.1	Sample outcomes for individuals with known residence ordering.....	69
4.3.2	Sample outcomes for individuals with unknown residence ordering.....	72
4.4	Uncertainty and sensitivity analysis.....	73
4.4.1	Uncertainty analysis for dose and dose rates.....	73
4.4.2	Outcomes of Monte Carlo simulations.....	78
4.4.3	Input parameter sensitivity.....	82
4.4.4	Limitations of assumptions.....	83
4.5	Results of quality control tests.....	84
4.5.1	Comparison to boundary conditions .....	84
4.5.2	Comparison to test cases / expected outcomes.....	85
4.5.3	Examination of extrema.....	86
4.6	Factors affecting cumulative dose and dose rate.....	87
4.6.1	Contribution from outliers .....	87
5	CONCLUSIONS.....	91
5.1	Conclusions from this work.....	91
5.2	Contribution of external dose to total dose .....	93

5.3	Future directions.....	93
	REFERENCES.....	96
APPENDIX A.	Selected plates from the ATLAS .....	99
APPENDIX B.	Analysis of Publicly Available Soil Sampling Data.....	107
APPENDIX C.	Equirectangular transformation of <sup>137</sup> Cs deposition density maps .....	130
APPENDIX D.	Protocol for map transformation process and geodatabase .....	135
APPENDIX E.	Monte Carlo and sensitivity analysis system design.....	138
APPENDIX F.	Sample extract/transact/load processing checklist .....	150
APPENDIX G.	Verification of extrema .....	153
APPENDIX H.	Previously published estimates for dose for similar cohorts.....	155

## LIST OF TABLES

Table 3-1: Age-based dose conversion factors .....	39
Table 3-2: Behavior factors due to age-profession group.....	42
Table 3-3: Descriptions of test cases.....	51
Table 4-1: Summary of statistics for the entire cohort.....	62
Table 4-2: Summary of error estimates for selected cases.....	79
Table B-1: Simple random sampling values for pilot study .....	110
Table B-2: Statistics of simple random sampling for pilot study.....	110
Table B-3: Statistics of set of all samples for pilot study .....	110
Table B-4: Values for Kruskal-Wallis test for pilot study.....	111
Table B-5: Ranks for Kruskal-Wallis test for pilot study.....	111
Table B-6: Simple random sampling values for 1200 km study.....	113
Table B-7: Statistics of simple random sampling for 1200 km study.....	113
Table B-8: Statistics for set of values for 1200 km study .....	113
Table B-9: Values for Kruskal-Wallis test for 1200 km study.....	114
Table B-10: Ranks for Kruskal-Wallis test for 1200 km study.....	114
Table B-11: Kruskal-Wallis correction for ties for 1200 km study.....	115
Table B-12: Description of strata.....	117
Table B-13: Strata weights.....	118
Table B-14: Statistics of strata .....	119
Table B-15: Intermediate values for stratification .....	120
Table B-16: Values for computing optimum sampling for strata.....	121
Table B-17: Optimum number of samples.....	122
Table B-18: Expense for each stratum .....	123
Table B-19: Values for D’Agostino’s test for log-transformed full data set.....	125

Table B-20: Values for D’Agostino’s test for log-transformed censored data set .....128

## LIST OF FIGURES

Figure 2-1: Sample <sup>137</sup> Cs deposition density map from the ATLAS .....	7
Figure 3-1: Plate distribution and sample plates from the ATLAS .....	25
Figure 3-2: Effective fraction of source term over time .....	32
Figure 3-3: Kerma in air .....	36
Figure 3-4: Effective dose conversion factor from kerma in air .....	38
Figure 3-5: Effective dose in the individual from kerma in air .....	39
Figure 3-6: Possible ranges for cumulative dose and dose rate .....	49
Figure 3-7: <sup>137</sup> Cs deposition density near Zhitomir, Kiev, and Chernobyl .....	50
Figure 3-8: Dose and dose rate for test cases in various locations .....	51
Figure 3-9: Dose and dose rate for test cases with various behavior factors .....	52
Figure 4-1: Distribution of age in the Ukraine compared to the cohort .....	58
Figure 4-2: Distribution of age-occupation categories over time .....	59
Figure 4-3: Distribution of total number of unique residences .....	60
Figure 4-4: Distribution of possible residence orders .....	61
Figure 4-5: Cumulative distribution of effective dose .....	63
Figure 4-6: Boxplot of cumulative dose distributions at yearly intervals .....	64
Figure 4-7: Comparison of Average Cumulative Effective Dose to Background .....	65
Figure 4-8: Lognormality of dose distribution .....	67
Figure 4-9: Doses received by end of first year versus end of study .....	68
Figure 4-10: Ranks of dose received in first year compared to study period .....	70
Figure 4-11: Example results for individuals with known residence orders .....	71
Figure 4-12: Doses for sample cases with indeterminate record orderings .....	72
Figure 4-13: Estimates of dose from Monte Carlo simulations .....	78
Figure 4-14: Results of Monte Carlo analysis for sample individual – Case 4 .....	80

Figure 4-15: Results of Monte Carlo analysis for sample individual – Case 8 .....	81
Figure 4-16: Dose in cohort compared to boundary conditions .....	85
Figure 4-17: Cohort dose distribution compared to expected outcomes .....	86
Figure 4-18: Cohort effective dose compared to extrema.....	87
Figure 4-19: Average dose rate across the entire cohort.....	88
Figure 4-20: Average dose rate (with and without extreme outlier).....	89
Figure 5-1: Comparison of results from this study to other published estimates .....	92
Figure A-1: <sup>137</sup> Cs deposition – Europe .....	100
Figure A-2: <sup>137</sup> Cs deposition – Poland.....	101
Figure A-3: <sup>137</sup> Cs deposition – Belarus.....	102
Figure A-4: <sup>137</sup> Cs deposition – Ukraine.....	103
Figure A-5: <sup>137</sup> Cs deposition – Northwestern Russia.....	104
Figure A-6: <sup>137</sup> Cs deposition – Western Central Russia .....	105
Figure A-7: <sup>137</sup> Cs deposition – Southwestern Russia.....	106
Figure B-1: Geographical stratification of Europe .....	117
Figure B-2: Normal probability plot for full data set.....	124
Figure B-3: Lognormal probability plot for full data set .....	124
Figure B-4: Normal probability plot for censored data set.....	127
Figure B-5: Lognormal probability plot for censored data set.....	127
Figure C-1: Oblique azimuthal projection map of the Ukraine .....	131
Figure C-2: Oblique azimuthal projection map of the Ukraine, data stripped.....	132
Figure C-3: Equirectangular projection map of the Ukraine.....	133
Figure C-4: Equirectangular projection map of the Ukraine, deposition only .....	134
Figure E-1: Results of Monte Carlo analysis of trivial system .....	142
Figure E-2: Results of Monte Carlo analysis of time-independent system.....	143
Figure E-3: Results of Monte Carlo analysis of time-dependent system.....	145

Figure E-4: Results of Monte Carlo analysis of system with resampling.....147  
Figure E-5: Results of Monte Carlo analysis of system with summation..... 149

## 1 INTRODUCTION

Since the termination of atmospheric weapons testing after the ratification of the Limited Test Ban Treaty in 1963, there have been only two radiological events with truly global scale implications: The nuclear plant accident and subsequent releases of radionuclides to the environment at the Chernobyl Nuclear Power Plant on 26 April 1986, and the tsunami-induced nuclear accident at Fukushima on 11 March 2011.

Shortly before the Fukushima incident, the National Science Foundation funded a joint project (NSF grant #HSD 0826983)[1] between the Natural Hazards Center at the University of Colorado (PI - Dr. RoseMarie Perez Foster) and the department of Environmental Health and Radiological Sciences at Colorado State University (co-PI/senior radiation physicist - Dr. Thomas Borak), with statistical analysis services provided by New York University (senior statistician - Dr. Robert Yaffee). The NSF gave this project the mission of investigating the long-term psychological and social effects of the Chernobyl accident on the population of the Ukraine by examining a representative cohort of individuals living in the Ukraine. Specifically, the project's mission included a mandate to examine connections between reconstructed radiological dose, perception of risk, and the following pathologies: Anxiety, depression, and post-traumatic stress disorder (PTSD).

The team at the University of Colorado was tasked with producing and managing the epidemiological data necessary for this investigation. The role of Colorado State University was to determine and execute a method for dose reconstruction for members of the cohort. The senior statistician at New York University was charged with joining the dose data with behavioral and physiological pathologies in the cohort and investigating any possible relationships.

After the collaborators defined research approaches, an epidemiological survey instrument was produced, survey-takers were trained, and a cohort of Ukrainian

respondents interrogated between 2009 and 2011. A pilot study of 281 individuals was conducted in 2009, the results of which informed the methodologies adopted by all three working groups. The survey instrument was completed by all 703 members of the cohort by the end of 2011, at which time the data analysis process began [2].

At the time of the 2009 pilot study, the CSU team determined that funding and data availability limited any dose reconstruction capability to external dose only. It was decided that a complete dose reconstruction (incorporating estimates for inhalation and ingestion) would be left to the next iteration of the project. The data collected during the complete study through 2011 should enable computation of internal dose at some point in the future.

The external dose reconstruction efforts were undertaken by the author of this document under the supervision of Dr. Tom Borak. The methodologies and resultant outcomes described herein pertain to dose reconstruction started at the end of 2011 for the entire 703-member cohort. At the time of this writing, the final results for the project (namely, the outcome of the senior statistician's work) are not available.

The methods described here are specific to the Chernobyl disaster and the Ukrainian experience thereafter. In theory, the underlying approaches could be applied to any radiological incident with large-scale dispersal of radiocontaminants, but would require considerable adjustment and event-specific data collection.

Also presented in this document is an effective (if not highly accurate) method of recovering radiation source term information from previously published radiocontaminant isoline maps from *Atlas of Caesium deposition on Europe after the Chernobyl Accident* [3]. The CSU working group used this as a method of last resort after foreign governments declined to release original source term data for radiocontaminant inventories in soil in nations that had formerly been members of the Soviet Union.

It is the author's sincere hope that the work conducted in this project will contribute to an understanding of the long-term psychological and social impact of nuclear disasters, and thereby aid public health authorities in mitigating their profound impact on society.

## 2 BACKGROUND

### *2.1 The disaster at Chernobyl*

On 26 April 1986, a serious accident occurred at an RBMK-1000 graphite-moderated nuclear reactor in the Chernobyl power plant in the Ukraine. A series of operator errors led to two explosions that destroyed part of the core of the reactor and the roof of the reactor building. Portions of the fuel core and contaminated reactor building were sent as high as 1 km into the air from the blast [4]. The larger debris was deposited in the area of the reactor, with small particles of dust and smoke carried considerable distances. The remaining, now un-cooled fuel was exposed to the atmosphere, where it began reacting to release radioactive gases and particulates. Compounding matters, a fire started in the wreck of the reactor building, eventually spreading to the graphite moderator for the core and becoming extremely hot. This produced an ongoing plume of heat and smoke which contained a large quantity of radioactive material. The accident continued in this active state for ten days [5], during which time the plume eventually reached an altitude sufficient to allow the injection of radioactive material into the jet stream. Material from the plume was primarily deposited across Europe and Western Russia, although traces were found throughout the Northern Hemisphere.

The graphite fire significantly complicated the Soviet response to this accident. At the time, there was little expertise in fighting this type of fire [4], resulting in a response strategy which may have prolonged it – compound dropped from helicopters over the fire may have acted as a thermal insulator, helping to keep the graphite burning and eventually allowing the core to melt through its lower shielding and into the facility's basement. This area was still flooded from the original firefighting efforts; when the corium met the water there, steam was created causing an additional release of radionuclides.

By 9 May 1986, the fire had been extinguished, and by 24 May 1986, temporary containment had been re-established through the construction of a new concrete slab beneath the core [4]. A longer-term solution (the Chernobyl Nuclear Power Plant Sarcophagus) was completed by the end of 1987, and construction of a permanent containment facility was started in 2010.

On the order of  $\sim 10^{19}$  becquerels of radioactivity were released from Chernobyl in total [4], of which  $\sim 10^{18}$  becquerels were deposited across Europe. The overall composition of deposition included a wide range of radionuclides initially present in the core inventory, from short-lived  $^{132}\text{Te}$  (half-life of 78 hours) to long-lived  $^{239}\text{Pu}$  (half-life of 24,400 years). Exposure to these radionuclides at sufficient concentrations can result in significant health effects to humans. Pathways include external exposure hazards (primarily from gamma-emitting radionuclides), internal exposure hazards from ingestion of radionuclides introduced to the food chain, and inhalation of airborne contamination. Generally, internal exposure to a given radiation source produces a higher risk than external exposure to the same source. The population of Europe – and especially the populations of the Ukraine, Belarus and Western Russia surrounding the reactor site – were potentially exposed to radiation from Chernobyl in sufficient quantities to produce noticeably elevated cancer risk [4, 5].

Thyroid cancer is the most common pathology observed in individuals affected by the Chernobyl disaster [6]. Of all the radionuclides released by the accident,  $^{131}\text{I}$  was most likely to produce noticeable health impacts due to internal exposure [6]. When  $^{131}\text{I}$  was deposited after the disaster, it remained on the leaves of plants which were eaten by both humans and animals used as food sources by humans (such as cows) [5]. As an analogue of non-radioactive iodine,  $^{131}\text{I}$  moves readily through the trophic levels of the human food chain, and can be concentrated in food products such as milk and meat. Upon ingestion by a human,  $^{131}\text{I}$  concentrates in the thyroid gland.

Health outcomes as a result of internal exposure to  $^{131}\text{I}$  include thyroid cancer [7], autoimmune disease [8] possibly resulting in Graves' disease and hyperthyroidism [9, 10]. Risk of thyroid cancer is considered to be inversely proportional to age at the time of exposure, with a nearly negligible increase in cancer risk after the age of 20 [7]. While thyroid cancer risk increases with increasing exposure, very high intake of  $^{131}\text{I}$  is associated with less cancer risk, as the thyroid tissue is killed entirely before it can become oncogenic [7, 9, 10]. Thus at higher doses, non-cancerous outcomes become more likely.

Many of the radionuclides released by the accident contribute to external radiation exposure. Of these,  $^{137}\text{Cs}$  is most typically used as an indicator of radioactive contamination from the accident. It is an isotope of cesium that does not occur in nature, has a long half-life (30 years), and has a readily detectible gamma emission at 661.7 keV when it decays.  $^{137}\text{Cs}$  was present in the overall release in such large proportion that it was readily traceable across Europe. Numerous models were constructed by various research institutions to relate the deposition density and/or concentration of  $^{137}\text{Cs}$  in soil to other useful information, including the deposition density and/or concentration of all other radionuclides, soil mixing, and external exposure rates.

## ***2.2 Previous efforts to characterize the disaster's impact***

### **2.2.1 Production of the ATLAS**

The contamination of Europe by the Chernobyl accident was characterized through a massive transnational study between 1992 and 1995, leading to the production of the *Comprehensive Atlas of Caesium Deposition on Europe after the Chernobyl Accident* [3]. We will henceforth refer to this document as the ATLAS. Various nations employed teams of technicians and physicists to estimate  $^{137}\text{Cs}$  deposition using a range of methods including soil sampling, ground-based gamma spectrometry, and airborne gamma spectrometers. Each nation aggregated the data resulting from their efforts and reported  $^{137}\text{Cs}$  deposition

estimates to a central organizing body managed by authorities in the European Community, in collaboration with the Institute of Global Climate and Ecology (Russia), Minchernobyl (Ukraine), and the Republic Center of Radiation and Environment Monitoring (Belarus).

About 400,000 total  $^{137}\text{Cs}$  deposition data points were reported in this manner (although the actual number of measurements taken was considerably higher – each data point may represent the results of multiple measurements at a location). These data were then kriged, and the results presented as color-filled isolines representing  $^{137}\text{Cs}$  deposition density in soil, normalized to the day of the accident (examples shown below in *Figure 2-1: Sample  $^{137}\text{Cs}$  deposition density map from the ATLAS* and in *APPENDIX A*).



**Figure 2-1: Sample  $^{137}\text{Cs}$  deposition density map from the ATLAS**

The maps are available as vector graphics with multiple layers of information stored in Adobe PDF files, including contour lines representing intervals of  $^{137}\text{Cs}$  deposition. The ATLAS includes overlapping maps at several different levels of resolution. One plate shows the entire continent of Europe, while other plates show regional detail, and some plates show details of hotspots. Plates with a larger geographical extent contain less detailed  $^{137}\text{Cs}$  deposition density information; whereas hotspot plates show more geographical detail and

display more detailed  $^{137}\text{Cs}$  deposition density information than regional plates, which are in turn more detailed than the continental plate.

### 2.2.2 Availability of $^{137}\text{Cs}$ source term raw data

At the time of the nuclear disaster at Chernobyl, the Ukraine was still a Soviet state. In 1986, the Cold War was approaching its nadir, and despite Gorbachev's policies of perestroika (restructuring) and glasnost (openness) throughout the USSR, the Ukraine remained stubbornly resistant to change [11]. Even after the post-Cold War breakup of the USSR and subsequent separation of Eastern European nations, a resistance to openness with the West partially informs the posture of some national institutions to this day.

This became relevant to our project when our efforts were stalled at the first step of the dose reconstruction approach in an attempt to determine a source term ( $^{137}\text{Cs}$  deposition density in soil) for use in this project. Retrospective dose reconstruction typically requires that the source of radiation be well characterized. The work that led to the production of the ATLAS was the only comprehensive effort to characterize radiocontamination across Europe (and the Ukraine).

While  $^{137}\text{Cs}$  deposition density isolines were published in the ATLAS in 1998, the raw data provided by the European Commission [12] appeared to be either incomplete or inaccurate – particularly with data missing from Eastern European, formerly Soviet regions (see *APPENDIX B - Analysis of Publicly Available Soil Sampling Data*

). The authors attempted – without success – to acquire the complete source term data from the World Health Organization, the European Commission Joint Research Center, and directly from the original custodians of data from the Ukraine at the Ukrainian Radiation Protection Laboratory.

Based on careful reading of previously published reports, we determined that the source term data for the Ukraine existed only in this last institution, that production of  $^{137}\text{Cs}$

deposition density isolines for the ATLAS was executed on-site and the results (but not the data) transferred to the European Commission [3], and that the source term data had never been propagated to higher authorities in the Ukraine or the European Union [13].

### ***2.3 Motivation for our study***

#### **2.3.1 Nuclear disasters and psychosocial outcomes**

While the physiological effects of radiation exposure have been well characterized in general, it remains unclear what the relationship is between large-scale radiological events and psychosocial behavior outcomes. It has been observed that survivors of such events may change their behavior in significant ways, delaying or forgoing pregnancy, abusing substances, and overusing medical services [14, 15]. In some cases, serious mental health disturbances have been observed. It is well understood that an individual's perception of risk is a strong predictor of physical health, mental health, and psychosocial behavior outcomes. It is not clear what connection (if any) there is between radiation dose and the perception of risk in an individual or population, or if there is any direct connection between radiation dose and mental or psychosocial outcomes.

Because mental and psychosocial outcomes are key elements of the long-term impact of a radiological disaster, there would be great utility in a model that accurately predicts these outcomes. To this end, the NSF funded a research project in 2008 at the University of Colorado in collaboration with Colorado State University and New York University to expand the knowledge of complex interactions between radiation exposure, perception of risk, and psychosocial behavior outcomes by modeling outcomes for a representative sample of the population of the Ukraine which had been exposed to radiocontaminant materials released by Chernobyl.

## ***2.4 Structure of our study***

### **2.4.1 Project oversight, personnel, and responsibilities**

Overall project design, coordination, and oversight was provided by the principal investigator -- Dr. RoseMarie Perez Foster, henceforth referred to as the PI -- at the Natural Hazards Center (the NHC) at the University of Colorado. Additional subject matter expertise at the NHC was provided by Dr. Kathleen Tierney, a disaster sociologist; Ukraine-specific expertise was provided by Dr. Victor Chtenguelov at the Health Ministry of the Ukraine in his role as a science advisor to the project.

Radiation dose reconstruction was conducted by the dose reconstruction team at the Department of Environmental Health and Radiological Sciences at Colorado State University. The team was supervised by Dr. Tom Borak in his role as the project's co-principal investigator and senior radiation physicist. He shall be referred to hereafter as the co-PI in the context of the overall management of the project, and as the senior radiation physicist in the context of dose reconstruction. The dose reconstruction team included the senior radiation physicist, this author (Remi Frazier), and occasional support for specific issues from subject matter experts within CSU's Environmental Health and Radiological Sciences department and in the private sector.

Statistical analysis was undertaken by the senior statistician (Dr. Robert Yaffee) at New York University's Silver School of Social Work.

The study relies heavily on data collected with a questionnaire administered to study participants in the Ukraine. This questionnaire, which determined standardized mental health measures, will hereafter be referred to as the survey instrument.

Data collection in the Ukraine was overseen by the Ukrainian project director (Victor Chtenguelov) at the Kiev Academy of Labor and Social Relations. The survey instrument was administered to respondents in interviews by six field interviewers (each at the MS or

PhD level) trained by the Ukrainian project director and staff from the University of Colorado.

Ukrainian participants in the study shall be referred to individually as respondents, subjects, or individuals. The complete group of respondents will be referred to hereafter as the cohort.

Data for the survey instrument was initially received and stored in software written and managed by Vovici, a Maryland-based corporation. At this time, each individual was assigned a unique identifier, an integer that we will refer to as a Subject ID number when we need to reference specific individuals and outcomes.

Data cleaning, format transformations, language translations (between Ukrainian and English), and dissemination of data throughout the project were all handled by the project data management staff at the University of Colorado, under the direct supervision of the PI. The collection of responses to questions in the survey instrument shall be referred to hereafter as the survey dataset.

#### **2.4.2 Project workflow**

The PI collaborated with Vovici to build software that permitted entry of survey instrument responses directly into laptop computers by field interviewers during interviews. Field interviewers were trained to administer the survey instrument by the PI, the project data management team, and the Ukrainian field director.

A cohort minimum size of 700 individuals was established by the PI, Co-PI, and senior statistician. Study participants were recruited using a randomizing algorithm designed by the senior statistician to select and dial phones in the Kiev and Zhitomir oblasts of the Ukraine until more than 700 respondents were recruited. Respondents were

questioned in person by field interviewers in 90-minute interviews and were paid \$20 USD equivalent for their participation.

After finishing each interview, field interviewers uploaded the data from the survey instrument to a central database administered by Vovici. After the entire cohort had been interviewed, Vovici supplied the database to the project data management staff.

The project data management staff then cleaned the data set, inspected it for internal consistency, verified that no data values were outside of allowable ranges, and that all data values necessary for dose reconstruction and statistical analysis were present for each subject. If necessary, the project data management staff required that the Ukrainian project director and/or field interviewers correct missing or inaccurate data.

A pilot study was conducted between 2008 and 2009, wherein 100 respondents were interviewed, , and statistical models were generated by the senior statistician. Lessons learned from this process were used by the project data management staff to improve the survey instrument, by the dose reconstruction team to test and correct the methods used for dose reconstruction, and by the senior statistician to adjust statistical analysis methods.

The full study was conducted between 2009 and 2011. The final total number of respondents was 803; the data for 100 of these was disqualified and censored by the project data management staff for various reasons, leaving a final cohort size of 703.

Once a complete data set was available, the project data management staff anonymized the data and provided it to the dose reconstruction team and the senior statistician. When additional flaws in the data set were identified at this point by either the dose reconstruction team or the senior statistician, they were corrected by the project data management staff, and a new data set was delivered to both the dose reconstruction team and the senior statistician.

The dose reconstruction team used the data set to produce daily cumulative dose and dose rate reconstruction for each respondent, as well as statistics for the cohort as a whole.

Once dose reconstruction had been finished and validated, cumulative dose for each respondent at the end of each calendar year was reported to the senior statistician for time series analysis, and cumulative dose for each respondent at three points in time were reported to the senior statistician for panel analysis. Summary statistics for the cohort were reported to the PI for review.

The senior statistician then incorporated dose reconstruction data into models and algorithms for time series analysis, autoregression, spatial analysis, panel analysis, and other approaches to investigate connections between variables of interest.

Upon conclusion of the statistical analysis, the PI, co-PI, and senior statistician will collaborate to identify meaningful results to report to the National Science Foundation and to consider any outcomes which might warrant publication in peer-reviewed journals.

### **2.4.3 Goals of the survey instrument**

The survey instrument was created in collaboration between the PI, the Co-PI/senior radiation physicist, and the senior statistician. The survey instrument was comprised of questions designed to provide demographic information about respondents, as well as information necessary to complete the following profiles for each respondent:

- A demographic profile, including information summarizing basic demographic information such as gender, birthdate, et cetera;
- A radiation exposure profile, providing all the information needed to permit internal and external dose reconstruction in accordance with international standards and previously published literature regarding dose reconstruction for individuals affected by Chernobyl;
- A health profile, including information regarding prevalence of medical illness, medical services utilization, mental health functioning, and behavioral health patterns;

- Profiles for standardized mental health measures, including:
  - The Nottingham Health Profile: A standard mental health instrument for assessment of perceived health problems;
  - The Basic Symptom Inventory: An instrument which identifies self-reported clinically relevant psychological symptoms in adults;
  - The Coping Strategy Indicator: An instrument which tests for strategies that characterize coping behavior in individuals;
  - The Civilian Mississippi Post-Traumatic Stress Disorder Scale: An instrument that measures general distress and test items “anchored” to a specific traumatological event.

All of these profiles have been validated as effective for use in Ukrainian populations [16, 17].

#### **2.4.4 Format of the survey dataset**

The survey instrument gathered data from individuals in the cohort pertaining to their demographics, behavior, and medical histories for the period of time between 26 April 1986 and 31 December 2009. We will refer to this interval of time as the study period.

The architecture of the data collection software was produced by Vovici; this in turn dictated the structure of the survey instrument and the survey dataset. The survey instrument was constructed such that data collected was organized into two sections: A general demographic overview of each individual (including name, birthdate, etc.) followed by their responses to the survey instrument. The individual’s responses to questions included in the radiation exposure profile of the survey instrument were further segregated into four sub-intervals within the study period. We shall refer to these sub-intervals as time periods. The four time periods were defined by the survey instrument as follows:

- Time Period 1 (which we shall sometimes refer to as TP1) encompassed the time between 26 April 1986 and 30 June 1986 (finest resolution – 1 day)
- Time Period 2 (which we shall sometimes refer to as TP2) encompassed the time between 1 July 1986 and 31 December 1986 (finest resolution – 2 weeks)
- Time Period 3 (which we shall sometimes refer to as TP3) encompassed the time between 1 January 1987 and 31 December 1990 (finest resolution – 1 month)
- Time Period 4 (which we shall sometimes refer to as TP4) encompassed the time between 1 January 1991 and 31 December 2009 (finest resolution – 6 months)

A database record was created within the survey dataset for each occupation and/or location of residence indicated by the respondent. We will refer to these records as residence/occupation records. Residence/occupation records contained the following information:

- Duration of time encompassed by the record (as amount of elapsed time in years, months, and days – not as a start and stop date)
- Location of residence (listed as either a latitude/longitude coordinate, or if that was unknown, a settlement name)
- Occupation (selected from one of three possible categories: indoor worker, outdoor worker, or pensioner)
- Responses to all health inventory questions (not detailed in this document)
- Responses to all food intake questions (not used by the external dose reconstruction detailed in this document)

The survey instrument was restricted in that data collected for an individual was limited to no more than four total residence/occupation records per time period, where the sum of durations of time listed across all residence/occupation records with a time period

was equal to the length of that time period. There could therefore be no less than four total residence/occupation records listed for an individual (one per time period), and no more than sixteen (four per time period).

#### **2.4.5 Indeterminate ordering of residence/occupation records**

Proper ordering of the survey dataset is key to producing a meaningful dose reconstruction: The accumulated dose for an individual who lived close to the reactor and evacuated shortly after the accident would be significantly different than the accumulated dose for an individual who lived far from the reactor, but moved close to it sometime after the accident.

Well after all the survey instruments had been administered to all individuals in the cohort, it was determined that the data survey instrument was not guaranteed to store residence/occupation records in order within each time period.

Examination of the survey dataset by the dose reconstruction team showed that numerous respondents listed their residence as a location outside the Ukraine at the time of the interview. This was suspicious, since all participants in the study were recruited and interviewed in the Ukraine. Further investigation determined that while the survey instrument successfully enforced ordering of the four time periods within the larger study period, it did not enforce ordering of residence/occupation records within each time period. The field interviewers were also not trained to record residence/occupation changes in temporal order. This conspired with the policy of the survey instrument software to store residence/occupation records with time duration rather than start/stop times to produce indeterminate orderings.

Due to the limited budget of the project at the time this discovery was made (a full year after the last survey instrument was completed), correction of the full survey dataset was not possible. Since the majority of dose would be received during the first year after

the accident at Chernobyl, we determined that establishing residence/occupation record orders for Time Period 1 and Time Period 2 was of highest priority. It was clear that some correction was necessary, so a compromise between accuracy and budget was made: Respondents with more than one residence/occupation record in either Time Period 1 or Time Period 2 were contacted to verify the orderings of this information. The dose reconstruction team was then responsible for accommodating indeterminate ordering during Time Period 3 and Time Period 4 in the dose reconstruction methodology.

## ***2.5 Terminology specific to this document***

### **2.5.1 Time**

Throughout this document, when we are considering an arbitrary point in time after the accident (in order to describe our models, or note cumulative effective dose or effective dose rate at a specific point in time, etc.), we will typically refer to the point in time as the day of interest, and express this symbolically with the variable  $[t]$ , where  $t$  is the amount of time (in days) between the date of the Chernobyl accident and the day of interest.

### **2.5.2 Location**

When we need to refer to subnational entities in the Ukraine, we will adopt the following terminology: Oblasts are the largest administrative divisions of the Ukraine, comparable to American states. Raions are the next smallest administrative divisions, comparable to American counties. Settlements may be cities, towns, villages, or even small independent clusters of dwellings.

### **2.5.3 Radiation physics**

The following terminology specific to the field of radiation physics will be used in this document:

A source term is a measure of radioactive contamination, used in radiation physics to refer generally to the source of radiation most relevant to a line of inquiry. For our study, the general source term is radioactive material released by the Chernobyl accident and deposited across Europe.

Activity is the decay rate of a radioactive substance, and gives the number of decays per unit time. Activity has units of decays per unit time, often given as becquerels (1 Bq = 1 decay per second).

Deposition density is used in this document as the abundance of a radionuclide per unit area, and refers to the amount of a material deposited in an area by the Chernobyl accident. Deposition density has units of activity per unit area, and is given herein as kBq/m<sup>2</sup> (these units are chosen to match the data reported in the ATLAS).

Kerma is the total kinetic energy of all charged particles released per unit mass by indirectly ionizing radiation (neutrons and gamma rays) passing through material. Kerma has units of energy per unit mass, often given as the quantity “gray” (1 Gy=1 J/kg).

Absorbed dose is the total kinetic energy deposited per unit mass by ionizing radiation passing through material. Absorbed dose has units of energy per unit mass, often given as (1 Gy=1 J/kg).

Effective dose is a stochastic measure of health risk to an individual from energy deposited by ionizing radiation. Effective dose is typically determined from absorbed dose by correcting for radiation and tissue types, has units of energy per unit mass, and is described with the quantity “sievert” (1 Sv=1 J/kg).

#### **2.5.4 Statistics**

We will sometimes refer to a p-value when discussing the results of statistical tests. Given a statistical test which provides a null hypothesis  $H_0$  (and possibly an alternative hypothesis  $H_A$ ) and a test statistic derived in some fashion from a sampling of a population, a

p-value gives the probability of observing the test statistic if the null hypothesis in question is true. (In other words, it gives the probability of observing the test statistic as the result of random chance.)

## 3 MATERIALS AND METHODS

### *3.1 Executive summary*

We created an algorithm to determine the effective dose rate from external radiation at an arbitrary time after the Chernobyl accident for any individual in the cohort. Our models do not incorporate estimates for internal dose (due to ingestion or inhalation) at this time. Our methodology was generated by combining  $^{137}\text{Cs}$  source term data extracted from the ATLAS with previously published models for soil mixing in the Ukraine, kerma in air from radionuclides in soil, effective dose to a person from kerma in air, and exposure reduction from shielding and individual behavior. Where necessary, we relied on previously published models, selecting those that had a high degree of specificity for the population and geography of the Ukraine. This general approach is well-supported for the computation of external dose for individuals in the Ukraine/Belarus/Russian regions impacted by Chernobyl [18].

We constructed a relational database to store data from the survey dataset and to execute our computations. This database will be referred to as the dose reconstruction database. We developed a method to transform  $^{137}\text{Cs}$  deposition density map plates from the ATLAS and load them into a spatial database, which we will refer to as the source term spatial database.

Effective dose rate (written as  $\dot{D}(t)$ ) from external radiation hazards for each individual, for a point in time  $t$  days after the accident was computed through the following process (summarized here and described in detail later in this section):

1. Demographic information (including age, residence location, and occupation) was found for the individual in question on the day of interest by querying the dose reconstruction database.

2. An *initial indicator source term* (written as  $\boxed{C_0}$ ) was defined as  $^{137}\text{Cs}$  deposition density in soil at the individual's residence location by querying the source term spatial database.
3. The *effective indicator source term* (written as  $\boxed{C_{eff}(t)}$ ) for the day of interest was determined by computing radiological decay of the initial indicator source term and using previously published models to account for radionuclide migration in soil and associated gamma-ray attenuation between the time of the accident and the day of interest.
4. *Kerma rate in air from all radionuclides* (written as  $\boxed{\dot{K}(t)}$ ) at the individual's residence location was found by using previously published models relating kerma in air from all radionuclides to the effective indicator source term, accounting for the decay of all radionuclides between the time of the accident and the day of interest.
5. *Uncorrected effective dose rate* (which we refer to as  $\boxed{\dot{D}(t)}$ ) in the whole body was estimated for the individual in question on the day of interest, using previously published models relating radiocontamination in the Ukraine, kerma in air, and the individual's body composition at their current age. This gave an estimate of whole-body uniformly exposed effective dose.
6. *Effective dose rate* (written as  $\boxed{\dot{D}(t)}$ ) for the individual in question was determined, accounting for reductions in radiation exposure due to shielding imparted by building composition, time spent indoors or outdoors, and the individual's typical behavior. This gave an estimate of whole-body, uniformly exposed effective dose rate.

We computed the dose rate for each individual in the study for each day in the study period, and stored the results in the dose reconstruction database. Cumulative effective

dose for each individual for any day of interest in the study period was computed by summing the effective dose rate for each day between the day of the accident and the day of interest.

For individuals with indeterminate residence/occupation record orderings (as described in *Section 2.4.5 - Indeterminate ordering of residence/occupation records*), we computed effective dose rate for each day of the study for all possible residence/occupation orderings and took the mean of these possibilities as our estimate for effective dose rate for that individual.

We determined statistics for the cohort as a whole by finding characteristic values (min, max, mean, quartiles, arithmetic and geometric variance) for cumulative effective dose estimates at the end of each year in the study period. We tested for kurtosis and skewness to investigate the contribution of outliers to our results, and stratified the cohort using several qualitative characteristics to look for driving subgroups within the population sampled by the cohort. Finally, we implemented a rigorous quality management system to verify and validate our results, requiring well-behaved input data from the survey dataset, repeating our software development process and computations, comparing our results to boundary conditions and predicted outcomes, and manually examining extrema.

## ***3.2 Methods employed***

### **3.2.1 Constructing and populating the dose reconstruction database**

A relational database using a MSSQL engine was designed and built to store data from the survey dataset and execute the dose reconstruction process. This will be referred to as the dose reconstruction database.

An extract, transform, and load process (written in Visual Basic, SQL Server Integration Services, and Transact-SQL) was developed to extract data from the survey dataset and load it into the dose reconstruction database.

A daily dose record was created in the dose reconstruction database for each individual in the cohort, for each day in the study period. Effective dose rate, cumulative effective dose, and associated statistics were computed and stored for each daily dose record for each individual.

This is not the most computationally conservative method for executing external dose reconstruction for a cohort of this size: It would require significantly less computation time to build code in a more efficient language (such as C++ or Java) to compute external dose. The dose reconstruction team elected instead to design a system as generally as possible, with the intent to incorporate internal dose reconstruction at some point in the future. Due to dependence on interconnected systems (such as food distribution networks and radionuclide migration through food chains), internal dose reconstruction promises to be a much more convoluted process than external dose reconstruction – a problem demanding the flexibility of a relational database.

### **3.2.2 Transforming radiocontamination maps into a spatial database**

A spatial database using an ESRI ArcGIS engine was constructed to manipulate and store radiation source term information, predicated on information presented in the ATLAS. We will refer to this database as the source term spatial database. Source term values were recorded in the dose reconstruction database for each residence location listed in the survey dataset by interrogating the source term spatial database (using an ODBC bridge between the two databases).

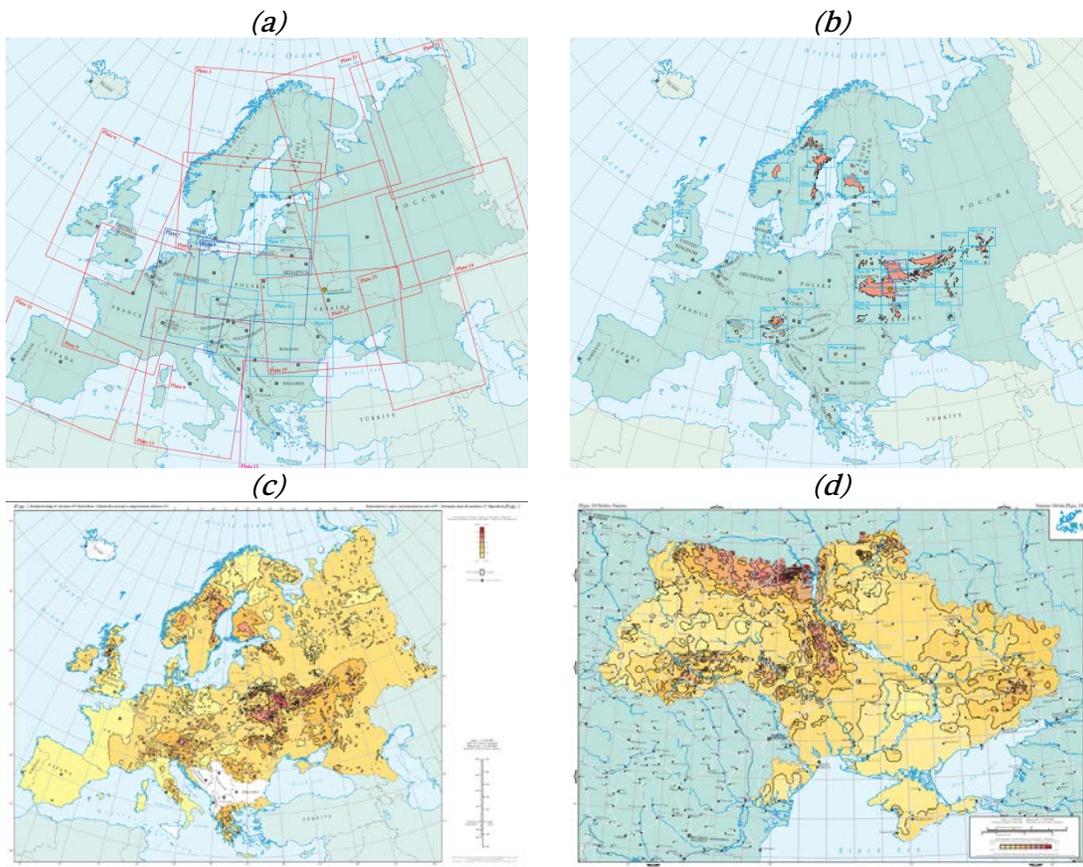
The source term for this study is the deposition density of all  $\gamma$ -emitting radionuclides in soil across Europe due to the Chernobyl accident. This source term was characterized by an extensive transnational soil sampling project between 1992 and 1995, which resulted in the production of the ATLAS as described in *Section 2.2.1 - Production of the ATLAS*. The soil sampling project used the presence of  $^{137}\text{Cs}$  as an indicator for the presence of other radionuclides in soil. Numerous subsequent publications rely on this convention, including the studies upon which much of our reconstruction efforts are predicated [19, 20].

The results of the soil sampling project were reported in the ATLAS as a series of full-color isoline maps showing the distribution of  $^{137}\text{Cs}$  deposition density in soil normalized to the day of the accident. As described in *Section 2.2.2 - Availability of  $^{137}\text{Cs}$  source term raw data*, the soil sampling source data used to produce the ATLAS was unavailable for our study. This led us to develop a process for estimating  $^{137}\text{Cs}$  deposition density at any location based on  $^{137}\text{Cs}$  deposition density isolines shown published plates of the ATLAS. Given an arbitrary location, we took the mean of the lower and upper bound for  $^{137}\text{Cs}$  deposition density contours as shown in the ATLAS. For the purposes of error estimation, we assumed a uniform distribution of possible  $^{137}\text{Cs}$  deposition density valued between the lower and upper bound of the published contours. While this is far from ideal, it was the best available solution at the time of our project.

In total, the ATLAS presents sixty-five plates, of which fifty-six show estimates for  $^{137}\text{Cs}$  deposition density data as color-filled isolines. The exact mechanism of interpolation utilized to produce these isolines is discussed at length in the ATLAS. In general, an inverse distance weighting function was used which estimates data values at unmeasured locations based on data values at measured locations, weighting by number of samples gathered and proximity to the unmeasured location. It is important to note that this process is a procedure which inherently reduces the fidelity of the data presented by smoothing local

minima and maxima. The interpolation method was used to approximate a field of  $^{137}\text{Cs}$  deposition density estimates across Europe. In order to produce the published map plates of the ATLAS, isolines in this field were defined (at apparently arbitrary intervals) and formatted for publication.

Each plate shows  $^{137}\text{Cs}$  deposition density in a subsection of Europe: One plate shows the totality of the continent, twenty-three plates show smaller regions encompassing countries (or portions of western Russia), and the remaining thirty-three plates show details of  $^{137}\text{Cs}$  hotspots (*Figure 3-1: Plate distribution and sample plates from the ATLAS*).



**Figure 3-1: Plate distribution and sample plates from the ATLAS**

(a) Distribution of plates showing regional detail; (b) Distribution of plates showing hotspot detail; (c)  $^{137}\text{Cs}$  deposition density map of the continent of Europe; (d)  $^{137}\text{Cs}$  deposition density map of the Ukraine

The maps themselves are shown in a Lambert azimuthal projection of the globe – this is an equal-area projection from a sphere (the surface of the Earth) to a plane (the published

maps). Each plate also includes a legend that relates the colors of contours shown on the map with an estimate for the upper and lower bound for  $^{137}\text{Cs}$  deposition density. Geographical information, labels, and latitude/longitude lines are overlaid on each plate. The ATLAS is available in printed format or as a set of electronic documents containing each published plate in a vector graphics format.

We obtained map plates from the ATLAS in the vector graphics format as Adobe PDF files. We observed that the published map plates include a great deal of information (text, political boundaries, etc.) which is not directly relevant to  $^{137}\text{Cs}$  deposition density data. We also noted that each of the published maps includes a grid of intersecting latitude/longitude lines (this grid is properly called the *conjugate graticule*). Because the conjugate graticule is labeled, we were able to associate the coordinates of the intersections of the conjugate graticule in each published plate with a latitude and longitude on the surface of the earth. This allowed us to approximate a transformation from each published plate to the surface of the earth. This transformation is an approximation of the inverse of the original Lambert projection. We applied this transformation to the color-filled isolines from each plate and loaded the transformed isolines into a geodatabase as follows:

1. Published map plates from the ATLAS were downloaded from the EC website in a vector graphics format. The vector graphic objects in each plate's file were categorized by type, separating shapes representing the conjugate graticule, shapes representing  $^{137}\text{Cs}$  deposition density isolines and colors, shapes representing the  $^{137}\text{Cs}$  deposition density legend, and discarding all other shapes (using Adobe Acrobat and Illustrator). The retained vector graphics for each plate were exported as three 150DPI, 24-bit CMYK bitmap files: One containing the conjugate graticule, one containing the color-filled isolines giving  $^{137}\text{Cs}$  deposition density, and one containing the  $^{137}\text{Cs}$  deposition density legend.

2. The bitmap files showing the conjugate graticule and the  $^{137}\text{Cs}$  deposition density isolines were loaded into a geographic information systems database (using ESRI's ArcInfo software suite). The conjugate graticule information and the isoline information for each plate were associated with a standard coordinate system such that any point in either set of information could be addressed with two coordinates. Because the two sets of information were originally overlaid, and had been exported from the original vector graphics files consistently, each plate's standard coordinate system correctly addressed points in either the conjugate graticule or the isoline data sets.
3. A second-order polynomial transformation was defined for each plate by associating a latitude/longitude coordinate of each intersection of the conjugate graticule. This transformation was then applied to the matching  $^{137}\text{Cs}$  deposition density isolines. In all cases, the  $R^2$  value for each transformation was greater than 0.95. Using tools available in the geographical information system, the hexadecimal color code for any location on any map plate could then be accessed using latitude/longitude coordinates.
4. The  $^{137}\text{Cs}$  deposition density color legend was exported from each plate's vector graphics file. The listed  $^{137}\text{Cs}$  deposition density bounds were noted, and a hexadecimal RGB color code for each listed deposition density bound was determined (using Adobe Photoshop). The values for hexadecimal color and matching deposition density bounds for each plate were stored in a lookup table in the dose reconstruction database. A color-to- deposition density lookup table was then defined for each plate, which matched hexadecimal color codes with minimum and maximum  $^{137}\text{Cs}$  deposition density estimates.

We then wrote software that used the GIS database to estimate  $^{137}\text{Cs}$  deposition density at an arbitrary location. Given a latitude/longitude coordinate pair, the software interrogated the GIS database to recover the hexadecimal color code for  $^{137}\text{Cs}$  deposition density where the latitude/longitude coordinate pair appeared shown on any plate in the ATLAS. When data from multiple plates was found, the software utilized only the color data from the plate with the highest degree of spatial resolution (prioritizing hotspot plates over regional plates, and prioritizing regional plates over the European continental plate). The software then used the color-to- $^{137}\text{Cs}$ - deposition density lookup table in the dose reconstruction database to find the lower and upper bounds of  $^{137}\text{Cs}$  deposition density as shown on the published plate in the ATLAS, and returned these two values as the  $^{137}\text{Cs}$  source term at the location of interest.

If no data were available from any of the plates from the ATLAS, an extrapolated value was returned by the software: An extrapolated deposition density range of 1-2 kBq/m<sup>2</sup> was returned for data points south or west of Europe (equivalent to the deposition density shown in the ATLAS in southernmost or westernmost Europe). An extrapolated deposition density range of 2-4 kBq/m<sup>2</sup> was returned for data points north or east of Europe (equivalent to the deposition density shown in the ATLAS in northernmost or easternmost Europe).

For dose reconstruction, the mean of the minimum and maximum  $^{137}\text{Cs}$  deposition density estimate shown in the ATLAS (as returned by the source term recovery software) was used as the initial indicator source term.

### **3.2.3 Computing effective dose rate at an arbitrary time $t$**

We employed the following process to determine the effective dose rate for an arbitrary individual on an arbitrary day of interest after the accident. Throughout this description, we will use the variable  $\boxed{t}$  to refer to the time elapsed (in days) since the day of the accident at Chernobyl.

### 3.2.3.1 Determining demographic information

We first interrogate the dose reconstruction database to recover the age, occupation, and location of residence listed for the individual on the day of interest. We will refer to the age of the respondent (in years) on the day of interest with the variable  $\boxed{age}$ .

### 3.2.3.2 Recovering initial indicator source term from geodatabase

We then recover the deposition density of  $^{137}\text{Cs}$  in soil normalized to the time of the accident at the individual's location of residence on the day of interest from the source term spatial database. We will refer to this as the *initial indicator source term*, and express it symbolically as  $\boxed{C_0}$ . The initial indicator source term has units of  $\text{kBq}/\text{m}^2$ .

### 3.2.3.3 Finding an effective indicator source term

We next find the effective deposition density of  $^{137}\text{Cs}$  at the time of interest, accounting for both radiological decay and attenuation due to vertical migration of cesium in soil. We will call this value the *effective indicator source term*, and refer to it symbolically as  $\boxed{C_{eff}(t)}$ . To determine the effective indicator source term from the initial indicator source term, we created a time-based model combining the radiological decay of  $^{137}\text{Cs}$  with a previously published Ukraine-specific soil migration and attenuation model [20]. This model is employed as a unitless time-based function which gives the effective fraction of  $^{137}\text{Cs}$  remaining at the time of interest. We will call this function the *effective decay function*, expressed symbolically as  $\boxed{r_{eff}(t)}$ . We then define  $\boxed{C_{eff}(t) = r_{eff}(t) \cdot C_0}$ .

The effective decay function is found by multiplying the *radiological decay function* (which we shall refer to symbolically as  $\boxed{r_{137\text{Cs}}(t)}$ ) with the *soil migration and attenuation function* (which we shall refer to symbolically as  $\boxed{r_{soil}(t)}$ ). We therefore define

$$\boxed{r_{eff}(t) = r_{soil}(t) \cdot r_{137\text{Cs}}(t)}.$$

The radiological decay function is well-characterized, using the half-life of  $^{137}\text{Cs}$  ( $11018.3 \pm 9.5$  d [21]) with the familiar equation for radiological decay:

$$r_{137\text{Cs}} = e^{-\frac{\ln(2)}{T_{1/2, \text{Cs}}}t} = e^{-\frac{\ln(2)}{11018.3}t}$$

The motivation for the soil migration and attenuation function is described in detail in [20], and is summarized below:

Radiological material migrates in the soil as a result of normal environmental patterns (rain, seepage, etc.). In the absence of human influence (farming, plowing, remediation efforts, etc.), this migration is generally down into the soil, resulting in greater attenuation due to soil and less gamma ray exposure above ground.

A mechanism was published in 2002 which models this type of attenuation of radionuclides in the Ukraine [20]. Approximately 400 soil core samples (of 14-20 cm depth) were taken at different points in time between 1988 and 1999 in non-remediated and undisturbed areas in the Ukraine. The core samples were cross-sectioned into 1-2 cm slices, and their radiological load determined by gamma spectroscopy using germanium detectors. The migration of  $^{137}\text{Cs}$  into the soil at each time point was determined, and the migration rate was determined over time [20].

The resultant estimation for the distribution of  $^{137}\text{Cs}$  in soil volumes over time was then used to compute the expected attenuation of gamma radiation, using previously published MCNP simulations of photon interactions [22-24]. This model does not incorporate adjustments for attenuation of different gamma energies from radionuclides other than  $^{137}\text{Cs}$ , or of differing soil migration patterns for radionuclides other than  $^{137}\text{Cs}$ .

Parameters for this model in different zones of the Ukraine were determined and given in [20]; the model takes the form of the following equation, with input parameters  $C$ ,  $A$ ,  $T1$ , and  $T2$  determined from the experimental data:

$$r_{soil}(t) = C(Ae^{-(\log(2)/T1)t} + (1 - A)e^{-(\log(2)/T2)t})$$

The following values for each of the four input parameters were taken from [20] for “Reference Ukrainian” areas (described as suitable in general for regions in the Ukraine). Error for the parameter  $A$  is taken as the average of error listed for experimental values in [20]; error for the remaining parameters is not listed, and is assumed here to be on the order of 10% of the listed mean value.

$$C = 0.82 \pm 0.082$$

$$A = 0.4 \pm 0.5$$

$$T1 = 1.5 \text{ y} \pm .15 \text{ y}$$

$$T2 = 50 \text{ y} \pm 5 \text{ y}$$

This produces the following equation for the effective fraction remaining:

$$r_{soil}(t) = 0.82(0.4e^{-0.00126t} + 0.6e^{-0.0000383t})$$

Note that while the radiological decay function is normalized to 1, at  $t=0$ , the leading coefficient in the soil migration and attenuation function (and thus the combined model as well) is such that at  $t=0$ , the effective fraction of the source term is 0.82. This is due to initial migration of  $^{137}\text{Cs}$  into the soil at the time of deposition [20].

Combining the soil migration and attenuation function and the radiological decay function by multiplication produces the following equation, which gives the value of the

effective decay function at any point in time (see *Figure 3-2: Effective fraction of source term over time*):

$$r_{eff}(t) = r_{soil}(t) \cdot r_{137Cs}(t)$$

$$r_{eff}(t) = (0.82(0.4e^{-0.00126t} + 0.6e^{-0.0000383t})) \left( e^{-\frac{\ln(2)}{11018.3}t} \right)$$

$$r_{eff}(t) = 0.328e^{-0.00133t} + 0.492e^{-0.000101t}$$

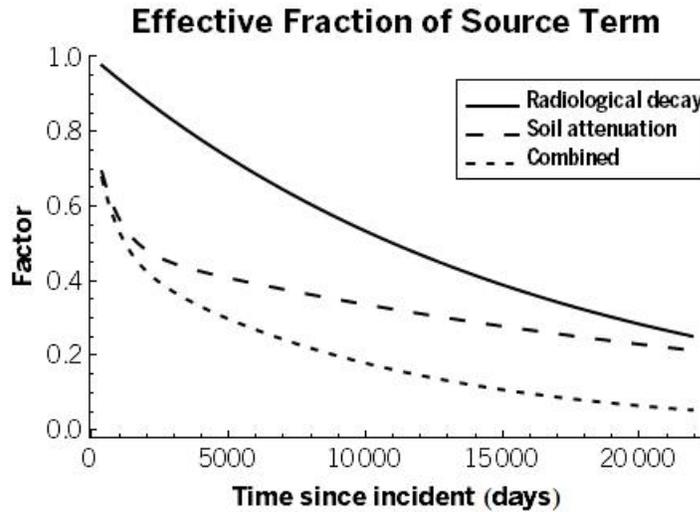


Figure 3-2: Effective fraction of source term over time

The effective indicator source term is then expressed as:

$$C_{eff}(t) = r_{eff}(t) \cdot C_0$$

$$C_{eff}(t) = (0.328e^{-0.00133t} + 0.492e^{-0.000101t}) \cdot C_0$$

### 3.2.3.4 Finding kerma in air at 1 m from all radionuclides

We next find the kerma rate in air, at a height of 1 m from the ground due to gamma-ray emissions from all radionuclides deposited by the Chernobyl accident. We will call this value *kerma rate in air* and express it symbolically as  $\boxed{\dot{K}(t)}$ . To determine kerma rate in air from the effective indicator source term, we use a time-based model described which provides a piecewise function, normalized to  $^{137}\text{Cs}$  deposition density in soil, which converts

from  $^{137}\text{Cs}$  deposition density in soil to kerma rate in air. We will express this conversion symbolically as  $\boxed{\dot{E}(t)}$ . We therefore define  $\boxed{\dot{K}(t) = \dot{E}(t) \cdot C_{eff}(t)}$ . The motivation for  $\dot{E}(t)$  is described in detail in [19], and summarized below.

Experimental data collected at intervals between 1986 and 1989 measured kerma rates in air at 1 m above the ground at locations in the Ukraine. These measurements were compared to measurements of  $^{137}\text{Cs}$  deposition density in soil taken at the same time. Curve-fits of the data described in [19] yielded a three-part piecewise function that relates  $^{137}\text{Cs}$  deposition density in soil to kerma rate in air at 1 m from all radionuclides, expressed as follows:

$$\dot{E}(t) = \begin{cases} \frac{1}{30} A_0 \int_{30}^{370} A_1 t^{-B_1}, & 0 \leq t \leq 30 \\ A_1 t^{-B_1}, & 30 < t \leq 365 \\ A_2 e^{-B_2 t} + A_3 e^{-B_3 t}, & 365 < t \end{cases} \quad (\text{pGy/s})/(\text{kBq/m}^2)$$

The power function for the interval  $30 < t \leq 365$  and the sum of exponential functions for times after the first year ( $365 < t$ ) were found using the method of least squares [19] to fit the collected air kerma and  $^{137}\text{Cs}$  soil deposition data. Fitting also imposed the constraints that the two functions meet smoothly at the joining point (i.e., their values and first derivatives are equal when  $t = 365$ ).

The integral employed for the interval  $0 < t \leq 30$  requires additional justification. Soil sampling data were unavailable for the first month after the Chernobyl disaster. Because kerma in air during that time period was heavily influenced by air contamination and the debris plume, an approximation was made for the value of  $\dot{E}(t)$  during the first thirty days. A proportionality constant which relates total kerma in air measured in the first month ( $t < 30$ ) to total kerma in air measured in the subsequent eleven months ( $30 < t \leq 370$ ) was found. We will refer to this proportionality constant as  $\boxed{A_0}$ , and write a

relationship between the total kerma in air in the first thirty days and total kerma in air in the subsequent eleven months as follows:

$$\int_0^{30} \dot{E}(t)dt = A_0 \int_{30}^{370} \dot{E}(t)dt , \text{ where } A_0 \text{ is a constant.}$$

We elected to approximate kerma in air for this period as roughly constant, yielding the following form for  $\dot{E}(t)$  during the first thirty days:

$$\dot{E}(t) = \frac{1}{30} \cdot A_0 \int_{30}^{370} \dot{E}(t)dt, 0 \leq t \leq 30$$

Note that we are assuming a steady-state equilibrium in kerma rate during the first thirty days, where additional release from the ongoing graphite fire at the Chernobyl site and radiocontaminant deposition from the atmosphere occur at the same rate as radionuclide decay and soil migration, producing a constant source term. We are aware that this is a significant assumption and possible source of error in our methodology, but lacking additional data about air kerma rate during this time are at a loss for alternative approaches.

The constant  $A_0$  was found in as the ratio between the kerma in air measured in the first month compared to the kerma in air measured in the next eleven months. Data from daily air kerma measurements (corrected for background) during the first twelve months after the accident at eight different sites across the Ukraine were used to determine a value for this previously published ratio [19]. This publication provides the following value for the parameter  $A_0$ :

$$A_0 = 0.56 \pm 0.11$$

Mean values for each of the other input parameters are listed in [19], but no error estimate is listed for these values. In order to estimate error for the remaining terms, we repeated a least-squares fitting method described in the publication. We were able to replicate the results for the values for  $A_1$  and  $B_1$  in this fashion, yielding estimates for standard error. We could not replicate the results for the values for  $A_2$ ,  $B_2$ ,  $A_3$ , or  $B_3$ . Lacking a better alternative, we have accepted the published results for these four values, and assumed that the standard error for these values is on the same order of magnitude as those we computed for  $A_1$  and  $B_1$  (that is, on the order of 10% of the mean value):

$$A_1 = 188 \pm 33.7$$

$$B_1 = 0.98 \pm 0.043$$

$$A_2 = 2800 \pm 280$$

$$B_2 = 0.02958 \pm 0.002958$$

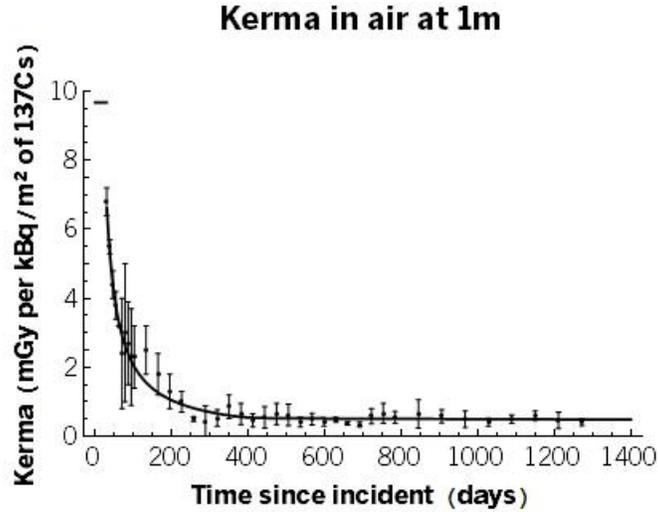
$$A_3 = 0.54 \pm 0.054$$

$$B_3 = 0.0000769 \pm 0.00000769$$

Applying these values to  $\dot{E}(t)$  yields the following form (also shown in *Figure 3-3*:

*Kerma in air*):

$$\dot{E}(t) = \begin{cases} 9.68, & 0 \leq t \leq 30 \\ 188t^{-0.98}, & 30 < t \leq 365 \\ 2800e^{-0.02958t} + 0.54e^{-0.0000769t}, & 365 < t \end{cases}$$



**Figure 3-3: Kerma in air**

*Kerma in air, normalized to  $^{137}\text{Cs}$  deposition in soil, as a function of time ( $\dot{E}(t)$ ). Superimposed are mean values and standard deviations of the soil sampling data which  $\dot{E}(t)$  is predicated upon.*

Having defined and justified  $\dot{E}(t)$  as described above, we can then express kerma in air for our methodology as follows:

$$\dot{K}(t) = \dot{E}(t) \cdot C_{eff}(t)$$

$$\dot{K}(t) = \begin{cases} 9.68 \cdot C_{eff}(t), & 0 \leq t \leq 30 \\ 188t^{-0.98} \cdot C_{eff}(t), & 30 < t \leq 370 \\ (2800e^{-0.02958t} + 0.54e^{-0.0000769t}) \cdot C_{eff}(t), & 370 < t \end{cases} \quad (\text{pGy/s})$$

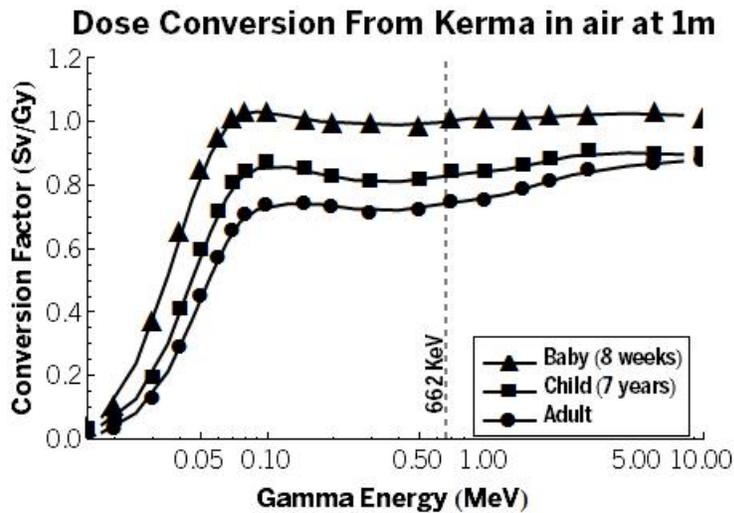
### 3.2.3.5 Finding uncorrected effective dose rate

We next find the effective dose rate, uncorrected for any factors (such as shielding) which would reduce an individual's exposure. We will call this the *uncorrected effective dose rate*, and express it symbolically as  $\dot{D}(t)$ . To determine the uncorrected effective dose rate from kerma rate in air, we use a model that approximates the effective dose to an individual from kerma in air based on the age of the individual.

We will express this conversion symbolically as  $\boxed{X(t)}$ . We can then define  $\boxed{\dot{D}(t) = X(t) \cdot \dot{K}(t)}$ . The motivation for  $X(t)$  is described at length in a succession of publications [25-28], indicated as suitable for Ukrainian populations [20], and is summarized below:

Software phantoms and MCNP were employed by a number of studies investigating the relationship between effective dose rate in an individual and kerma in air [26, 27]. This work assumed a height of 1 m above an infinite plane, with the source at a depth in the ground corresponding to a mass per area of 0.5 g/cm<sup>2</sup>. The 0.5 g/cm<sup>2</sup> configuration is expected to reasonably approximated contamination from the Chernobyl accident [20], but since it does not accommodate migration of radionuclides below this depth, it likely overestimates dose slightly.

Absorbed dose to each organ for different gamma energies was computed, and from this a conversion found to determine whole-body uniformly exposed effective dose from air kerma at 1m above the ground for standardized phantoms of different sizes, approximating three ages (baby of 8 weeks, child of 7 years, and an ICRP-standard adult). The results of this work are shown in *Figure 3-4: Effective dose conversion factor from kerma in air* [26, 28] (next page).



**Figure 3-4: Effective dose conversion factor from kerma in air**

*Conversion from kerma in air to whole-body uniformly exposed effective dose at 1m above an infinite plane source with the source at a depth in the ground corresponding to a mass per area of 0.5 g/cm<sup>2</sup>. Data for data points taken from [26, 28]; lines show a moving average between adjacent data points. The characteristic gamma energy for the 662keV emission of <sup>137</sup>Cs is emphasized.*

The phantoms were simulated to stand vertically on the ground for the purpose of computing absorbed dose to individual organs. It should be noted that this model is therefore limited in that it does not accommodate varying body sizes or compositions between individuals at a given age. This model also did not accommodate gender differences, which might produce additional adjustments to the dose conversion factors due to the location of female breast and male gonad tissue [25]. Furthermore, the model makes the assumption that individuals are standing on the ground at all times (which is unlikely to be the case for infants). Additional modeling using the phantoms in prone, face-up or face-down positions produced dose conversion factors 15-25% smaller than the standing positions [25].

Lacking specific information about the portion of time a typical Ukrainian spends prone versus standing, we elected to adopt the dose factors for standing individuals for our methodology. While these have been previously confirmed as suitable for the Ukrainian population [20], we recognize that this may produce an overestimation of effective dose.

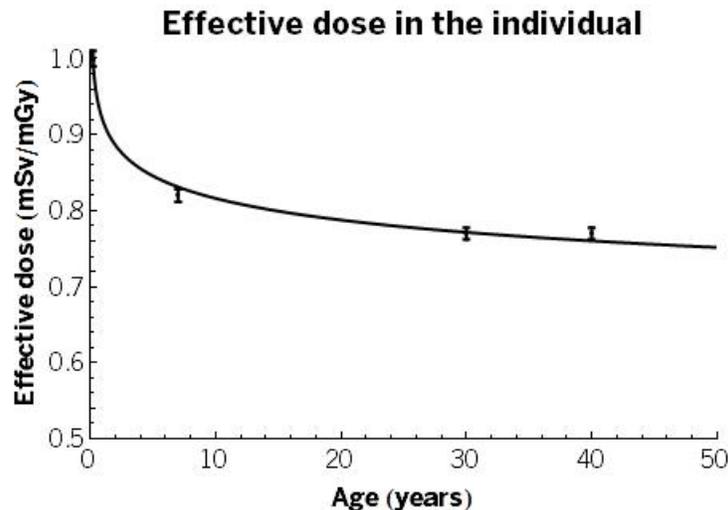
Estimates for the dose conversion factor from the gamma emission of  $^{137}\text{Cs}$  (662 keV) were taken from publications describing this model, and are shown in *Table 3-1: Age-based dose conversion factors*. We have made the assumption that the MCNP modeling used to generate these values utilized sufficient iterations to produce minimal error, that the distribution of this error is normal, and we have assumed that error is on the order of 1% for each value.

**Table 3-1: Age-based dose conversion factors**

*Dose factors from kerma in air at 1m above an infinite plane source of 662 keV gamma rays corresponding to a depth in soil of 0.5g/cm<sup>2</sup> for standing individuals. Data from [19, 26-28].*

<b>Age</b>	<b>Conversion Factor (Sv/Gy)</b>
0.2	1 ± 0.01
7.0	0.82 ± 0.0082
30.0	0.77 ± 0.0077
40.0	0.77 ± 0.0077

To produce an estimate of the effective dose conversion factor, we used the method of least squares to find a power curve fit of these data (shown in *Figure 3-5: Effective dose in the individual from kerma in air*).



**Figure 3-5: Effective dose in the individual from kerma in air**

The curve fit produced the following function ( $p < 0.01$  for both fitted parameters) which we used to estimate the effective dose in individuals from air kerma based on their age:

$X(t) = \alpha \cdot \left( age_0 + \frac{t}{365.25} \right)^{-\beta}$ , where  $\alpha = 0.917 \pm 0.00613$ ,  $\beta = 0.0507 \pm 0.00238$ , and  $age_0$  gives the attained age of the individual at the time of the accident.

$$X(t) = 0.917 \cdot \left( age_0 + \frac{t}{365.25} \right)^{-0.0507}$$

Our methodology does not fully capture the complexity of the real situation, where gamma rays of various energies are contributing to effective dose: We assume that kerma in air at 1m is produced exclusively by gamma radiation from  $^{137}\text{Cs}$ . In actuality, the kerma in air is produced from the gamma emissions of the entire inventory of radionuclides deposited in soil from the Chernobyl accident. The majority of the gamma-emitting radiocontaminants have relatively high energies ( $>500$  keV), and would therefore produce very similar dose conversion factors using this method (refer to *Figure 3-4: Effective dose conversion factor from kerma in air* above). Since the emissions from  $^{137}\text{Cs}$  are expected to produce the majority of dose, we believe that the error introduced by this approximation is insignificant.

With  $X(t)$  defined as above, we can express  $\dot{D}(t)$  as follows:

$$\dot{D}(t) = X(t) \cdot \dot{K}(t)$$

$$\dot{D}(t) = 0.917 \cdot \left( age_0 + \frac{t}{365.25} \right)^{-0.0507} \cdot \dot{K}(t)$$

### 3.2.3.6 Finding corrected effective dose rate

Finally, we find the effective dose rate, including any adjustments that would reduce an individual's exposure (such as shielding). We will call this the *effective dose rate*, and express it symbolically as  $\dot{D}(t)$ . To determine effective dose rate from the uncorrected effective dose rate, we employ a set of factors based on the individual's age and occupation at the time of interest. These factors approximate the reduction in exposure due to typical Ukrainian building materials and the portion of time spend indoors and shielded, outdoors in areas where soil was remediated, and outdoors in areas where soil was not remediated. We will refer to these factors as behavior factors, and express them symbolically as  $f_{age,occ}$ . We can then define  $\dot{D}(t) = f_{age,occ} \cdot \check{D}(t)$ . The motivation for  $f_{age,occ}$  is described in detail in [19], and summarized here:

Research in the Ukraine comparing indoor and outdoor dose rates (measured with handheld survey monitors) determined that for typical Ukrainian residences of brick and wood, shielding was such that the dose rate inside buildings was reduced to just 7% of the dose rate at undisturbed outdoors sites [19]. Furthermore, efforts within settlements to remediate contaminated areas such as yards, gardens, streets, and playgrounds produced a reduction of contamination such that the dose rate at these locations was reduced to 50% of the dose rate at otherwise undisturbed outdoors sites.

A population survey published in [29] and revisited in [19] and [20] established estimates for the fraction of time Ukrainians in different occupations and different ages spent indoors or out, and if outdoors, whether the time was spend in remediated locations.

The responses to this population survey were combined with the dose reduction factors to produce a set of behavior factors by which overall dose was reduced for typical individuals based on their occupation and age. We have adopted these factors as mean values for typical behavior, distributed lognormally based on experimental data collected

between 1987 and 1990 as provided in [19] (*Table 3-2: Behavior factors due to age-profession group*).

**Table 3-2: Behavior factors due to age-profession group**

*Values of age-profession behavior factors of Ukrainian population from model alongside geometric standard deviation ( $\sigma_g$ ) and arithmetic standard deviation ( $\sigma_a$ ) of experimentally determined values. [19, 20].*

Age-professional group		$f_{age,occ}$	$\sigma_g$	$\sigma_a$
Children	$\leq 7$ y	0.13	1.58	0.457
Adolescents to young adults	8-17 y	0.18	1.76	0.565
Adults	Employees	0.26	1.71	0.536
	Agricultural workers	0.38	1.52	0.419
	Pensioners	0.26	2.07	0.728

We set  $f_{age,occ}$  for the individual in question examining their listed occupation and age on the day of interest. If their age is such that they fit into either the “Children” or “Adolescent” group, we assign them a matching behavior factor based on the age on the day of interest. Otherwise, we assign them the behavior factor that matches their listed occupation. Thus, the final form of effective dose rate can be written as:

$$\dot{D}(t) = f_{age,occ} \cdot \dot{\tilde{D}}(t)$$

Expanded to show all the subsidiary models described above, effective dose rate can be written as:

$$\dot{D}(t) = f_{age,occ} \cdot X(t) \cdot \dot{E}(t) \cdot r_{soil}(t) \cdot r_{137Cs}(t) \cdot C_0 \quad (\text{pSv})$$

### 3.2.4 Computing cumulative effective dose

The cumulative effective dose, which we will refer to symbolically as  $\boxed{D(t)}$ , is determined from the effective dose rate by summing the effective dose rate for all previous

days. The cumulative dose and dose rate (in Sv/day) for any individual can be recovered for any day in the study period.

Consider an interval of time within the study period between the day of the Chernobyl accident ( $t=0$ ) and a point in time  $T$  days after the disaster. Cumulative effective dose  $D(T)$  for the interval of time between  $t=0$  and  $t=T$  is most exactly found by integration:

$$D(T) = \int_0^T \dot{D}(t) dt$$

The full piecewise function for  $\dot{D}(t)$ , however, cannot be analytically integrated using our current mathematical tools (an exercise left to the reader). Numeric integration would be suitable for our purposes, but executing this for each individual in the cohort was determined to require more computation time than was absolutely necessary. We elected instead to compute the cumulative effective dose by summation, using subintervals of one day each ( $\Delta t = 1 \text{ d} = 86400 \text{ s}$ ):

$$D(T) = \sum_{t=0}^T 86400 \cdot \dot{D}(t) \Delta t \quad (\text{pSv})$$

$$D(T) = \sum_{t=0}^T 86400 \cdot f_{age,occ} \cdot X(t) \cdot \dot{E}(t) \cdot r_{eff}(t) \cdot C_0 \Delta t \quad (\text{pSv})$$

Note that the coefficient value that appears here (86400) is employed to effect the units conversion from seconds to days.

It is worthwhile to note that this produces a slight overestimation of the cumulative effective dose: We are taking the upper sum (in daily increments) of the complete function for effective dose rate rather than the numeric integral. We anticipate that this will contribute very little error to the system as a whole (since we are principally interested in potential dose effects over years rather than days), but reduces our computation time considerably.

### 3.2.5 Accommodating indeterminate residence/occupation record ordering

Recall from *Section 2.4.5 - Indeterminate ordering of residence/occupation records* that the survey data set did not necessarily correctly order residence/occupation records during Time Periods 3 or 4. In the cases where an individual listed a move or occupation change during these time periods, we computed the effective dose rate for all possible orders of the listed residence/occupation records. We then reported the mean of all of the possible outcomes from these computations. This results in a new definition of  $\dot{D}(t)$ , where  $N$  is defined as the set of all possible residence/occupation record orderings:

$$\dot{D}(t) = \frac{1}{|N|} \sum_{n \in N} \dot{D}_n(t)$$

Note that in cases with known residence/occupation record orderings (i.e., where  $|N|=1$ ), this reduces to the same form as the equation for  $\dot{D}(T)$  in the previous section. Thus, when computing the cumulative effective dose between the date of the accident ( $t=0$ ) and an arbitrary time  $T$  later, for an individual with  $N$  possible orderings of residence/occupation records during the interval  $0 \leq t \leq T$ , the cumulative effective dose is computed as follows:

$$\begin{aligned} \dot{D}(T) &= \frac{1}{|N|} \sum_{n \in N} \dot{D}_n(T) \\ D(T) &= \sum_{t=0}^T 86400 \cdot \dot{D}(t) \Delta t = \frac{1}{|N|} \sum_{t=0}^T \sum_{n \in N} 86400 \cdot \dot{D}_n(T) \Delta t \end{aligned}$$

It is this form for computing the cumulative effective dose that was used to produce the final reported cumulative effective dose estimates for individuals in the cohort.

### 3.2.6 Computing summary statistics

Summary statistics were computed for the entire cohort, for each day of the study. These included the mean and median cumulative effective dose, the average effective dose rate, the arithmetic standard deviation, the interquartile range, and the geometric mean and standard deviation.

We reported these statistics to the senior statistician for the last day of each calendar year in the study, beginning with 31 December 1986 and ending with 31 December 2009. The selection of these time points was chosen at the request of the senior statistician, who employed these data for time series analysis to investigate connections between cumulative effective dose and health outcomes.

Three time points were selected for further investigation: 31 December 1986, 31 December 1996, and 31 December 2009. The selection of these time points was chosen at the request of the senior statistician, who had identified these as appropriate time points to use in panel data analysis for connections between cumulative effective dose and health outcomes. At these three time points, we conducted additional analysis: We attempted to fit distributions to the cohort data at these time points; we investigated the contribution of outliers to the distribution of cumulative effective doses in the cohort by computing the skewness and Pearson's kurtosis; and we explored the dynamics of dose distribution over time by examining dose rank changes in the cohort and Q-Q plots of dose distribution between each time point.

To explore extreme outcomes in the cohort's dose rate, we stratified the cohort data and looked for evidence that a small subgroup of individuals was contributing disproportionately to cumulative effective dose or effective dose rate in the cohort. We also compared dose distributions in various strata to test whether some strata were drawn from different populations than others.

### **3.2.7 Quality management system**

To confirm that the dose reconstruction process – including the project workflow, software development, and computation system – was producing accurate results, the following data validation, quality assurance, and quality procedures were followed.

#### **3.2.7.1 Source data validation**

To validate the data, we enforced several rules to ensure source data accuracy. These data validation rules were enforced prior to accepting any data for processing:

- Listed locations of residences were checked to verify that no residences were shown in patently implausible locations (i.e., in the middle of oceans).
- Listed birthdates were checked to ensure that they were no later than the date of the accident.
- The total amount of time listed for residency in each time period was checked to verify that it exactly covered the time period.
- In order to ensure confidentiality, we ensured that all identifying respondent data were removed from the data set before it was delivered to the dose reconstruction team. Listed Subject ID numbers were retained to allow the senior statistician to match the dose reconstruction output data with their data sets.

If any of these rules was broken for any individual in the survey dataset, we rejected that version of the survey dataset and returned it to the project data management staff for correction. Upon receipt of a corrected survey dataset, we repeated the source data validation steps above.

This iterative process produced a survey dataset that met all the above criteria. We did not attempt to further verify the accuracy of data provided (i.e., whether listed latitude/longitude coordinates matched the settlements provided by individuals).

### **3.2.7.2 Quality assurance**

To verify that our methodology and systems architecture was capable of executing dose reconstruction, we conducted the following tests before executing dose reconstruction.

#### *3.2.7.2.1 Analysis software validation*

The computation code was completely refactored on two different occasions by the author using the original specifications and design documents in order to confirm that some unnoticed programmer error had not produced a computation error. In all cases, the outcomes were identical.

#### *3.2.7.2.2 Technical replication of analysis*

To validate that software and hardware idiosyncrasies were not affecting our outcomes, we ran dose reconstruction for an early version of the survey dataset on three different hardware and software platforms. Dose reconstruction was computed for the survey dataset from the survey instrument on three different database servers employing different CPU chipsets (a four core Pentium, a six-core Pentium, and an eight-core AMD), different operating system versions (Microsoft Windows XP, Windows 7, and Windows 7 64bit), and different database engine versions (Microsoft SQL Server 2005 and 2008R2). In all cases, the outcomes were identical to within machine precision, finding no differences before the 14<sup>th</sup> position after the decimal place in any computed value.

### *3.2.7.2.3 Comparison of summation method to numeric integration*

The cumulative effective dose was computed in two industrially available software packages (MathCAD and Mathematica) for ten randomly selected individuals in an early version of the survey data set. The code for these computations was based on different specifications documents than the code used in the dose reconstruction system, and progressed along two fronts in each software package.

Cumulative effective dose was first determined by using the summation method as described in this paper, and compared to the outcome from the dose reconstruction system. We would expect these to be very slightly different depending on how each of the three systems in question handles numeric data types, but not at any level of significance. The outcome from both MathCAD and Mathematica in all ten cases was very close to the outcome from our dose reconstruction system (no differences before the 4<sup>th</sup> decimal place in any computed value).

Cumulative effective dose was also determined by numeric integration in both MathCAD and Mathematica (i.e., replacing summations with numeric integrals). This method would be expected to produce slightly lower values than the summation method described in this paper, since the summation method we used employs an upper sum of the effective dose rate in pSv/s to determine effective dose rate in mSv/day. The outcome from both MathCAD and Mathematica in all ten cases was in fact slightly lower than the outcome from our dose reconstruction system. The outcomes from MathCAD and Mathematica were found to be in close agreement with one another.

### **3.2.7.3 Quality control**

To verify that our estimates for dose were reasonable and accurate, we compared the output of our reconstruction system to boundary conditions and test cases. We also

manually examined extreme outcomes to convince ourselves that there was a reasonable explanation for their values.

### 3.2.7.3.1 Comparison to boundary conditions

Based on the minimum and maximum source term values shown in the ATLAS, we computed absolute minimum and maximum values possible for cumulative effective dose and dose rate from our methodology (see *Figure 3-6: Possible ranges for cumulative dose and dose rate*).

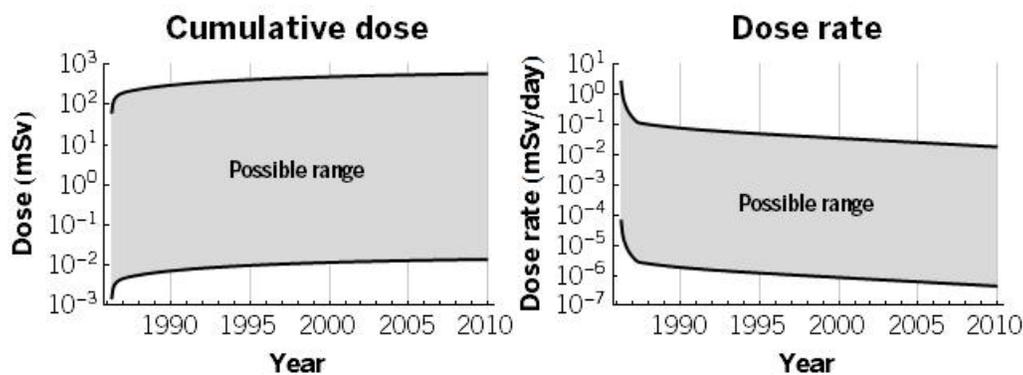
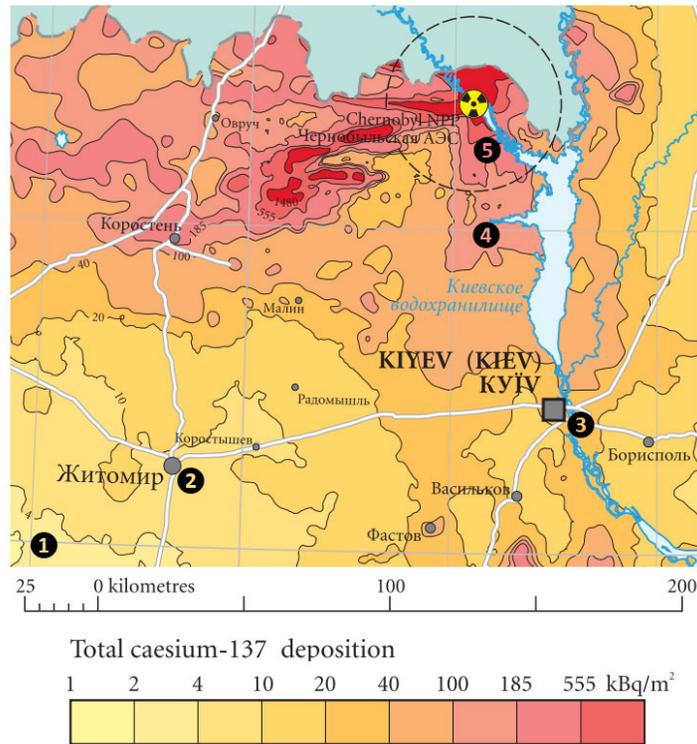


Figure 3-6: Possible ranges for cumulative dose and dose rate

After completing dose reconstruction for the entire cohort, we examined the outcomes from each individual to confirm that none of our dose estimates exceeded these boundary conditions.

### 3.2.7.3.2 Comparison to test cases

To produce guidelines for the expected outcome of dose reconstruction for individuals, dose was first computed for test cases. From the ATLAS plates, we can observe that  $^{137}\text{Cs}$  deposition density contours in the Kiev and Zhitomir oblasts range between 4  $\text{kBq}/\text{m}^2$  and 100  $\text{kBq}/\text{m}^2$  (see *Figure 3-7:  $^{137}\text{Cs}$  deposition density near Zhitomir, Kiev, and Chernobyl*, next page).



**Figure 3-7: <sup>137</sup>Cs deposition density near Zhitomir, Kiev, and Chernobyl**  
*This figure was extracted from the ATLAS [30]; locations of Test Cases 1-5 overlaid*

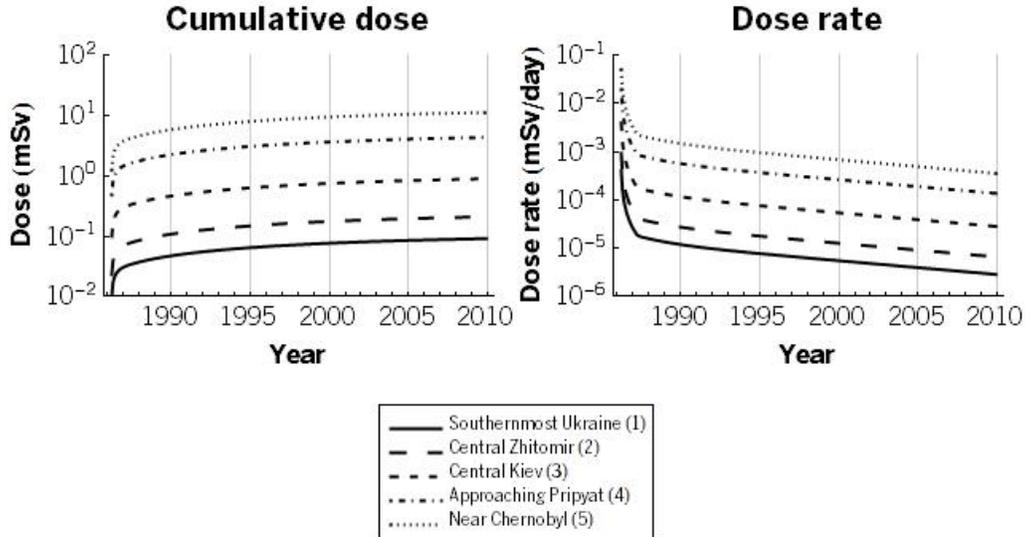
We generated five test cases using different source term values that we anticipated would be typical of the cohort: We first constructed a test case representing individuals residing in southern or western Ukraine with low exposure to radiocontamination (Test Case 1). We then produced test cases representing individuals residing in the center of Zhitomir (Test Case 2) and individuals residing in the center of Kiev (Test Case 3). Finally, we produced two cases representing individuals in the most highly contaminated regions that did not become exclusion zones (Test Cases 4 and 5). (See *Table 3-3: Descriptions of test cases*, next page)

**Table 3-3: Descriptions of test cases**

*Test cases in varying locations (with constant age-occupation group)*

Test Case #	Location	Age in 1986	Occupation	Source term
1	Southernmost Ukraine	Adult	Employee	2-4 kBq/m <sup>2</sup>
2	Central Zhitomir	Adult	Employee	4-10 kBq/m <sup>2</sup>
3	Central Kiev	Adult	Employee	20-40 kBq/m <sup>2</sup>
4	Approaching Pripjat	Adult	Employee	100-185 kBq/m <sup>2</sup>
5	Near Chernobyl	Adult	Employee	185-555 kBq/m <sup>2</sup>

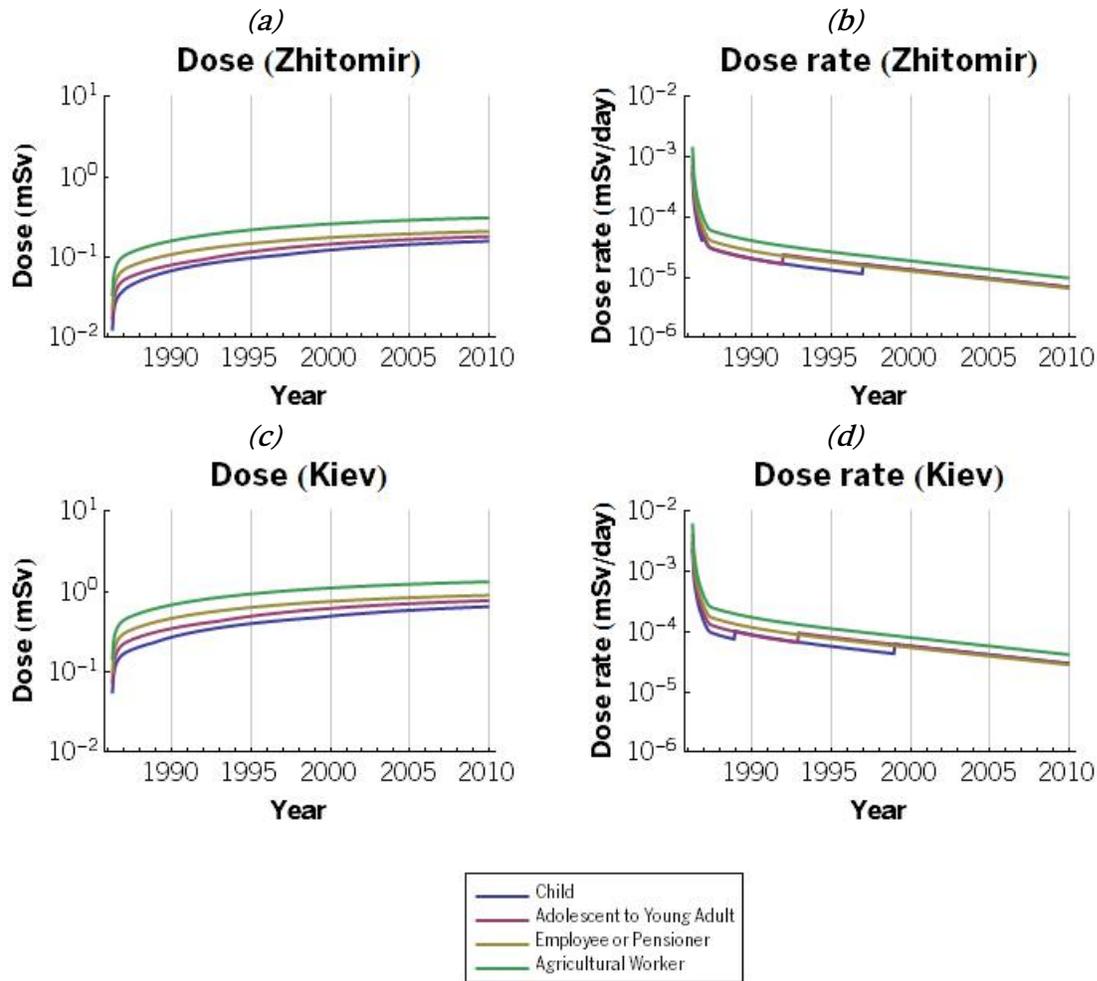
We anticipate that most of the dose outcomes in the reconstruction effort for the cohort should fall within the bounds of cumulative dose established by the test cases. In particular, we expect that the preponderance of the distribution of real cumulative effective dose should lie between a minimum of the dose computed here for Test Case 1 and a maximum of the dose computed here for Test Case 5. We further anticipate that most of the dose should lie near the doses computed for Zhitomir and Kiev, and that the cohort average dose would likely fall between the outcomes for Test Case 2 and Test Case 4. The results for these cases are shown in *Figure 3-8: Dose and dose rate for test cases in various locations*.



**Figure 3-8: Dose and dose rate for test cases in various locations**

We generated additional test cases representing each of the five possible age-occupation categories (children, adolescents to young adults, indoor workers, agricultural workers, and pensioners) and computed their dose twice – once as if they lived in the center

of Zhitomir for the study period (Test Cases 6-10), and once as if they lived in the center of Kiev for the study period (Test Cases 11-15). Cumulative dose and dose rate plots for these test cases are shown in *Figure 3-9: Dose and dose rate for test cases with various behavior factors*.



**Figure 3-9: Dose and dose rate for test cases with various behavior factors**  
 (a) Cumulative dose for subjects in Zhitomir ( $4\text{-}10\text{ kBq/m}^2$ ); (b) dose rate for subjects in Zhitomir ( $4\text{-}10\text{ kBq/m}^2$ ); (c) cumulative dose for subjects in Kiev ( $20\text{-}40\text{ kBq/m}^2$ ); (d) dose rate for subjects in Kiev ( $20\text{-}40\text{ kBq/m}^2$ )

Note that discontinuities appear in the dose rate charts for some test cases. These represent shifts in age/occupation categories and associated changes in behavior factors as individuals age from the “Child” category into the “Adolescent to young adult” category, or from the “Adolescent to young adult” category into the “Employee” category.

### *3.2.7.3.3 Manual examination of extrema*

Extreme outcomes in the final reported data were examined to verify that we could justify why some individuals were found to have significantly higher or lower cumulative dose than the rest of the population. Responses to the survey instrument for the five individuals who were determined to have received the most cumulative dose were examined, and a justification sought for why those individuals may have received significantly more dose than the rest of the population. This was repeated for the five individuals who were determined to have received the least cumulative dose.

### ***3.3 Materials used***

The following software and hardware technologies were used for dose reconstruction:

Data reported in this document (including charts) are the result of the entire dose reconstruction process being completed on a Hewlett-Packard HPE-560Z workstation with a six-core AMD Phenom II X6 1090T processor running at 3.20 Ghz, employing 16 GB of 1333 MHz SDRAM. Data were stored on a 300 GB Western Digital WD3000 SATA hard drive. The operating system running at the time of computation was Windows Professional 7 x64 SP1.

The survey dataset was provided to the dose reconstruction team both as an SPSS data file and as a Microsoft Excel 2007 file. The Microsoft Excel file was loaded into Microsoft Excel 2010 and exported as a tab-delimited text file.

The dose reconstruction database was constructed on a Microsoft SQL Server 2008R2 SP1 (SQL Server 10.50.2550) engine. The extract-transact-load process to recover information from the survey dataset text file was written in Visual Basic 6.0, employed a SSIS/DTSX import package, and utilized Transact-SQL statements to manipulate loaded data.

$^{137}\text{Cs}$  deposition density plates from the ATLAS were extracted from Adobe Acrobat PDF files downloaded from the EC website on 9 May 2011 [12]. Security was removed from the PDF files with Adobe Acrobat X Pro (v10.1.6). The PDFs were then loaded into Adobe Illustrator CS5 (v15.1.0), vector graphic shapes were categorized into layers, and layers of interest were exported as 150DPI 24 bit CMYK bitmap files.

The source term spatial database was constructed in an ESRI ArcGIS 10 spatial database engine, employing the ArcMap and ArcGlobe platforms (v10.0.0.2414).  $^{137}\text{Cs}$  deposition density layers (conjugate graticule and  $^{137}\text{Cs}$  color-filled isolines from each plate) were loaded from the bitmap files into ArcMap as raster layers and rectified. A unique session was created for each map plate. The inverse of the Lambert oblique azimuthal transformation was defined in ArcMap by picking at least 50% of the intersections on a given plate's conjugate graticule, associating latitude/longitude coordinates with each intersection, and generating a transform using tools in ArcMap. The transformation was saved, and then applied to the  $^{137}\text{Cs}$  color-filled isoline layer. An ArcGlobe session was then created, and each transformed  $^{137}\text{Cs}$  color-filled isoline layer was loaded and labeled.

Hexadecimal color-to- $^{137}\text{Cs}$  deposition density lookup tables were generated by loading the  $^{137}\text{Cs}$  deposition density legend from each plate (which had been exported as a bitmap file) into Adobe Photoshop CS5x64 (v12.1) and examining the colors displayed for each  $^{137}\text{Cs}$  deposition density range.

An ODBC connection was established between the dose reconstruction database and the source term spatial database using drivers built in to ArcGIS 10. An export table was created in the dose reconstruction database containing all latitude/longitude coordinates of interest, which was then imported into the source term spatial database over the ODBC connection. The ArcGlobe *Extract Multi Values To Points* tool was used to find the hexadecimal color for each coordinate on each loaded  $^{137}\text{Cs}$  color-filled isoline layer. A null value was returned if no data was found for a plate. A list of the coordinates and their

matching hexadecimal colors were exported to a text file and loaded into the dose reconstruction database using an SSIS/DTSX import script.

The computations described in *Section 3.2 - Methods employed* were executed in the dose reconstruction database using a series of Transact-SQL scripts. In accordance with current industry best practices, a separate reporting database was constructed on the MSSQL server instance to store the dose reconstruction results which are reported in this document and other publications generated by the project.

Simple statistical analysis (minima, maxima, means, medians, quantiles, arithmetic variance, geometric means/variances) was executed in the reporting database using Transact-SQL scripts.

More complex analyses (skewness, kurtosis, T-tests, Kruskal-Wallis tests) were executed in Mathematica 8 (v8.0.4.0 64 bit), making use of a read-only JDBC connection between the reporting database and Mathematica. The charts throughout this document were generated in Mathematica 8. Comparisons between summation and numeric integration were executed in Mathematica and MathCAD v14.0. The Monte Carlo simulation and rank correlation system described in APPENDIX E was developed and executed in Mathematica.

## 4 RESULTS AND DISCUSSIONS

### ***4.1 Characteristics of the data produced by the survey instrument.***

Between 2008 and 2011, individuals were called in the Kiev and Zhitomir oblasts of the Ukraine to solicit participation in the study. Approximately 6000 calls were placed, yielding an initial cohort of 803 respondents. These individuals were contacted by field interviewers by phone and in person to produce responses to the full battery of questions in the survey instrument. Of the completed surveys produced by the initial study group, 100 were disqualified and censored by the project data management staff, leaving 703 individuals with sufficient information to enable dose reconstruction.

#### **4.1.1 Demographic distribution**

##### **4.1.1.1 Gender distribution**

We would expect the gender makeup of our cohort to approximate the gender makeup of the population of the Ukraine.

The 703-member cohort was comprised of 339 men and 363 women (48.2% men). This was compared to the demographics of the Ukraine listed by the UN/WHO, which found that men made up 46.1% of the population as a whole, and 44.0% of the sub-population of individuals in the same age range as those which were allowed to be selected for the cohort (those of age 25 and older) [31]. Because we sampled from the population without replacement, we would expect our probability distribution function to be hypergeometric (assuming no bias).

Since we know the true distribution of gender in the Ukrainian population, we can readily compute the expectation value of our sample:

$$\mu = \frac{nN}{m+n} = (0.440) \cdot 702 = 309$$

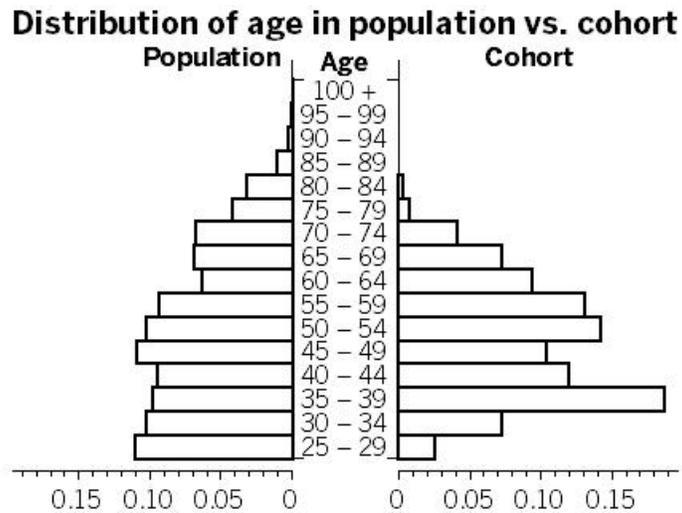
We computed the standard deviation of the expectation value using the probability mass function of the hypergeometric distribution, accounting for variance and covariance [32]:

$$\sigma = \sqrt{\frac{mnN(m+n-N)}{(m+n)^2(m+n-1)}} = 13.2$$

The actual distribution of male subjects in our cohort (339) is therefore more than two standard deviations greater than the expected value (309±13.2). This leads us to suspect that the cohort may have sampled from the population with a bias, as ~95% of truly random samplings of this size from the population should have fallen within two standard deviations from the expected value.

#### **4.1.1.2 Distribution of age and occupation**

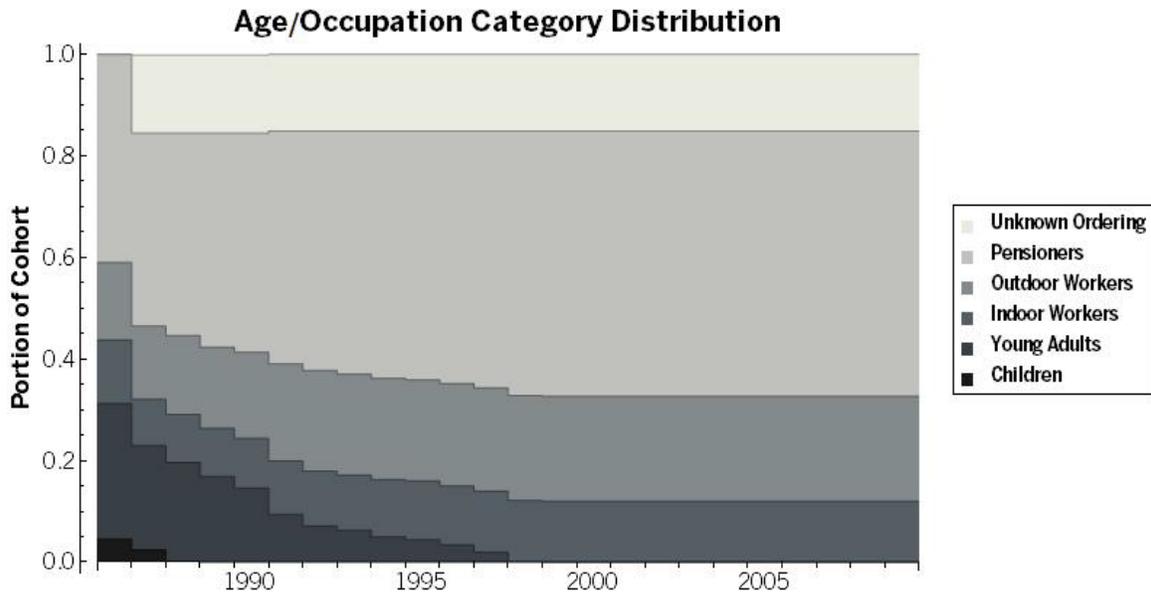
We obtained data for the distribution of ages in the overall Ukrainian population from a 2009 UN estimate [31]. Recalling that inclusion in the cohort required that individuals be alive at the time of the accident, we adjusted the listed distribution by censoring all data points below 25 years of age. We then compared the distribution of ages in our cohort (in 2009) to those in the Ukrainian population (shown in *Figure 4-1: Distribution of age in the Ukraine compared to the cohort.*)



**Figure 4-1: Distribution of age in the Ukraine compared to the cohort**  
*Comparison of the distribution of attained age in the Ukraine by the end of 2009 compared to the distribution of attained age in the cohort by the end of 2009.*

Visual inspection of this figure leads us to suspect that there may have been a slight bias in the selection of the cohort data: Individuals under the age of 35 are under-represented in the cohort. While the uncorrected effective dose in these individuals should generally be slightly higher than in older individuals (due to the monotonically decreasing function for effective dose), the corrected effective dose should be slightly lower (due to greater dose reduction due to behavior factors). We therefore anticipate that the average cumulative effective dose for the cohort will slightly overestimate the average cumulative effective dose for the population.

To gain a sense of occupation group changes in the cohort, we examined the distribution of age-occupation categories over time (*Figure 4-2: Distribution of age-occupation categories over time, next page*).



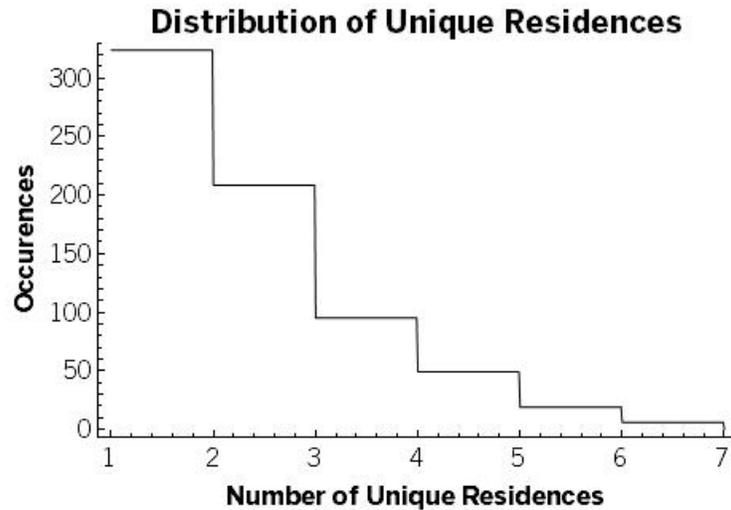
**Figure 4-2: Distribution of age-occupation categories over time**

This figure highlights the number of individuals with unknown residence/occupation orderings, and also points to another attribute of the data set, which was unexpected: A significant portion of the cohort listed their occupation as a pensioner – indicating that much of the cohort was either retired or unemployed.

#### **4.1.2 Number of unique residences**

The 703 individuals in the data set listed a total of 1372 residence locations. Degeneracy in the latitude/longitude coordinates for these locations was such that there were only 576 unique residence locations listed across the entire data set. This is a result of insufficient specificity in the survey instrument (as described in *Section 2.4.5 - Indeterminate ordering of residence/occupation records*) – respondents or field interviewers may have only recorded a settlement, without an exact latitude/longitude coordinate for a respondent’s address. The project data management staff resolved this by using the latitude/longitude coordinates of the center of the listed city or settlement (as described in *Section 2.4.2 - Project workflow*). As a result, individuals living in close

proximity to one another could be listed with the same coordinates. The maximum number of unique locations listed for an individual did not exceed 7. (See *Figure 4-3: Distribution of total number of unique residences.*)



**Figure 4-3: Distribution of total number of unique residences**

#### **4.1.3 Number of listed residence/occupation changes**

284 of the respondents indicated that they did not change their residence or occupation during the study period, while 149 indicated that they moved only once. The remaining 269 listed two or more residences.

#### **4.1.4 Number of possible residence/occupation orderings**

165 respondents indicated residence location or occupation changes in Time Period 3 and/or 4: 108 individuals indicated changes during Time Period 3, 106 indicated changes in Time Period 4, and 49 indicated changes during both time periods. The data for these individuals were therefore susceptible to the residence ordering problem described in *Section 2.4.5* and *Section 3.2.5*. Taking into account all of the possible residence/occupation record orderings for these individuals, dose reconstruction was computed for a total of 1700 cases, and the mean of all possible orderings reported for each individual. In most cases

(126), if the residence ordering problem occurred, it produced 2 possible orderings in Time Period 3 and/or 4, yielding either 2 or 4 total possible orderings for the entire study period. Some cases had significantly more ordering possibilities, including the two maximal cases that each yielded 144 possible orderings. (See *Figure 4-4: Distribution of possible residence orders* below.)

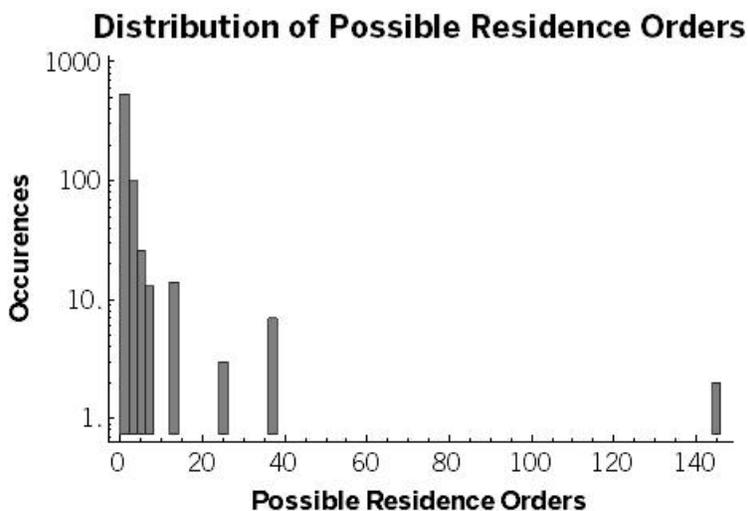


Figure 4-4: Distribution of possible residence orders

## 4.2 Results of dose reconstruction

### 4.2.1 Summary of results

The final version of the survey dataset was delivered to the dose reconstruction team by the project data management staff on 21 July 2011. We completed dose reconstruction computation for this data set on 5 August 2011. The high level of granularity in the computation (multiple passes of daily doses and dose rates for each individual as described in *Section 3.2.1 - Constructing and populating the dose reconstruction database*) produced an analysis database approximately 11 GB in size. The reporting database (used to produce

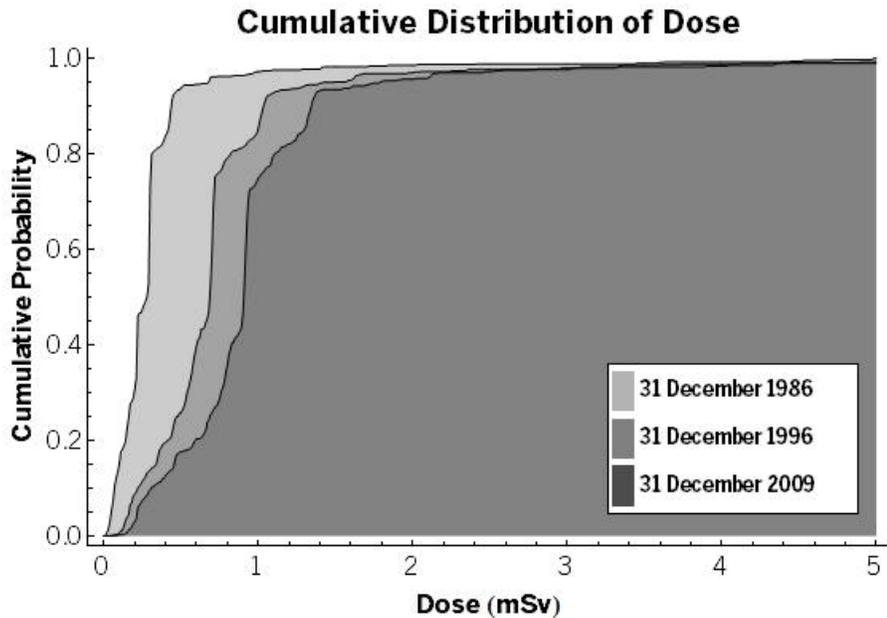
charts and formatted data for the statistics team as well as this document) finished at just over 2 GB in size.

The results of computations for the cohort are shown at three dates in *Table 4-1: Summary of statistics for the entire cohort* below. The dates for these three intervals are significant because the cumulative dose for individuals and the average cumulative dose for the cohort at these dates were used as inputs to a panel analysis by the senior statistician.

**Table 4-1: Summary of statistics for the entire cohort**  
**Cumulative whole-body uniformly**  
**exposed effective dose (mSv)**

Statistic	Cumulative whole-body uniformly exposed effective dose (mSv)		
	12/31/1986	12/31/1996	12/31/2009
Lowest value of External Dose received by an individual	0.0074	0.036	0.047
Largest value of External Dose received by an individual	28.0	30.0	31.0
2.5-97.5% inter-quantile range of External Dose received by the sample	0.037 - 1.4	0.14 - 3.4	0.19 - 4.4
Average value of External Dose received by the sample	0.38	0.93	1.2
Standard Deviation of External Dose received by the sample	1.2	2.0	2.2
Median value of External Dose received by the sample	0.28	0.69	0.91
Geometric Standard Deviation of External Dose received by the sample	2.3	2.1	2.1
Geometric mean of External Dose received by the sample	0.23	0.61	0.84
Skewness	17.5	10.2	8.71
Pearson's kurtosis	374	133	97.5
Estimated Average value of external dose from natural background (terrestrial plus cosmic) in the United States, corrected for behavior and shielding from typical building materials (NCRP160) [33]	0.33	5.3	12.0

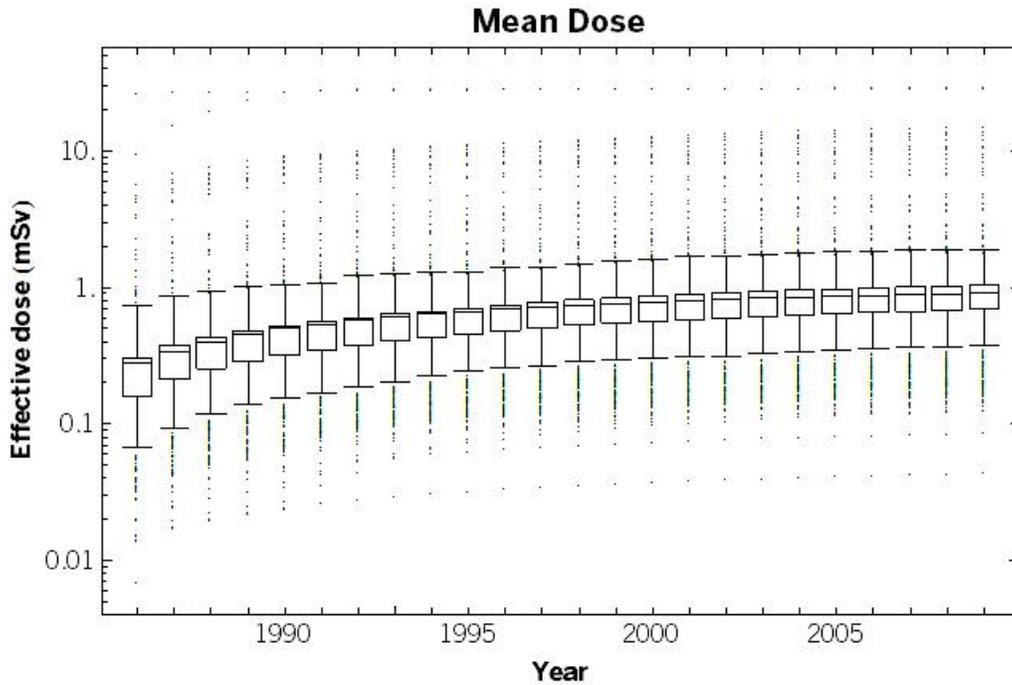
Visual examination of the cumulative distributions of doses in the cohort at these three time points (*Figure 4-5: Cumulative distribution of effective dose*) shows a number of rugged features. These are the results of the discretized nature of the  $^{137}\text{Cs}$  deposition density data used for the initial indicator source term.



**Figure 4-5: Cumulative distribution of effective dose**  
*Distribution of whole-body uniformly exposed effective dose for members of the cohort on three dates of interest.*

#### 4.2.2 Change in distribution of dose in the cohort over time

The results for computations for the cohort are shown at yearly intervals in *Figure 4-6: Boxplot of cumulative dose distributions at yearly intervals*. This boxplot reveals that the distribution of doses in the cohort has a long tail.



**Figure 4-6: Boxplot of cumulative dose distributions at yearly intervals**

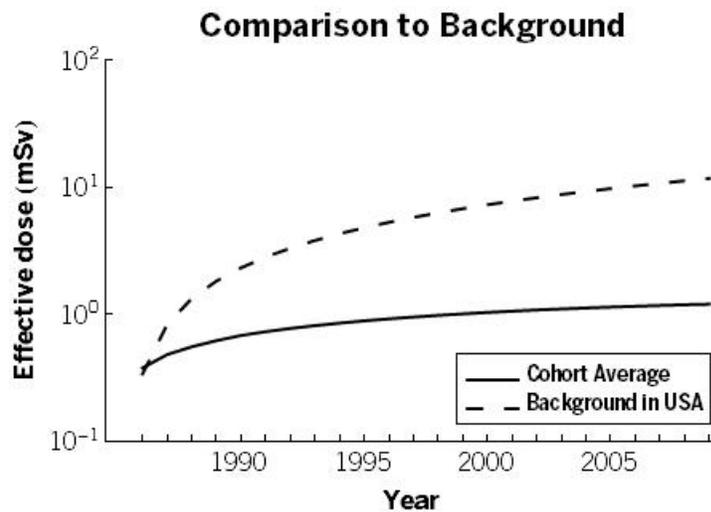
*Distribution of whole-body uniformly exposed effective dose for members of the cohort at yearly intervals (on the last day of each year in the study period). Boxes show 25% and 75% quartiles; lines dividing boxes show medians; fences show 1.5 IQR beyond the closest quartile; dots show outliers.*

These yearly calculations of cumulative dose for individuals were used for time series analysis by the senior statistician. Thus, *Table 4-1*, *Figure 4-5*, and *Figure 4-6* not only summarize the cohort results, but also represent the principal analysis deliverables of the dose reconstruction team’s contribution to the larger research project.

#### 4.2.3 Comparison to background radiation

It is particularly valuable to consider the dose received by members of the cohort due to the Chernobyl disaster in the context of the dose received due to background radiation. We compared the average cumulative effective dose which we computed for members of the cohort with the average cumulative effective dose due to natural background radiation from cosmic and terrestrial sources (see *Figure 4-7: Comparison of Average Cumulative Effective Dose to Background*). Note that this comparison uses background radiation levels in the

United States due to a lack of available data on background radiation in the Ukraine which does not include contamination from the Chernobyl disaster. The comparison dose include approximations for shielding due to behavior and building construction materials, but these are also representative of the population in the United States rather than the Ukraine. Thus the background comparison shown here may be slightly different than the actual background in the Ukraine, but is still suitable to act as a rough approximation.



**Figure 4-7: Comparison of Average Cumulative Effective Dose to Background**  
*Whole-body, uniformly exposed dose, comparing the cohort average cumulative dose with the dose from background radiation (terrestrial and cosmic) to the average member of the population of the United States (per ICRP160)[33]*

External doses received by individuals in the cohort from the Chernobyl disaster are less than the dose received from natural background. By the end of the study period, this disparity has grown such that the average cumulative dose received from Chernobyl is just a tenth of the dose received from natural background.

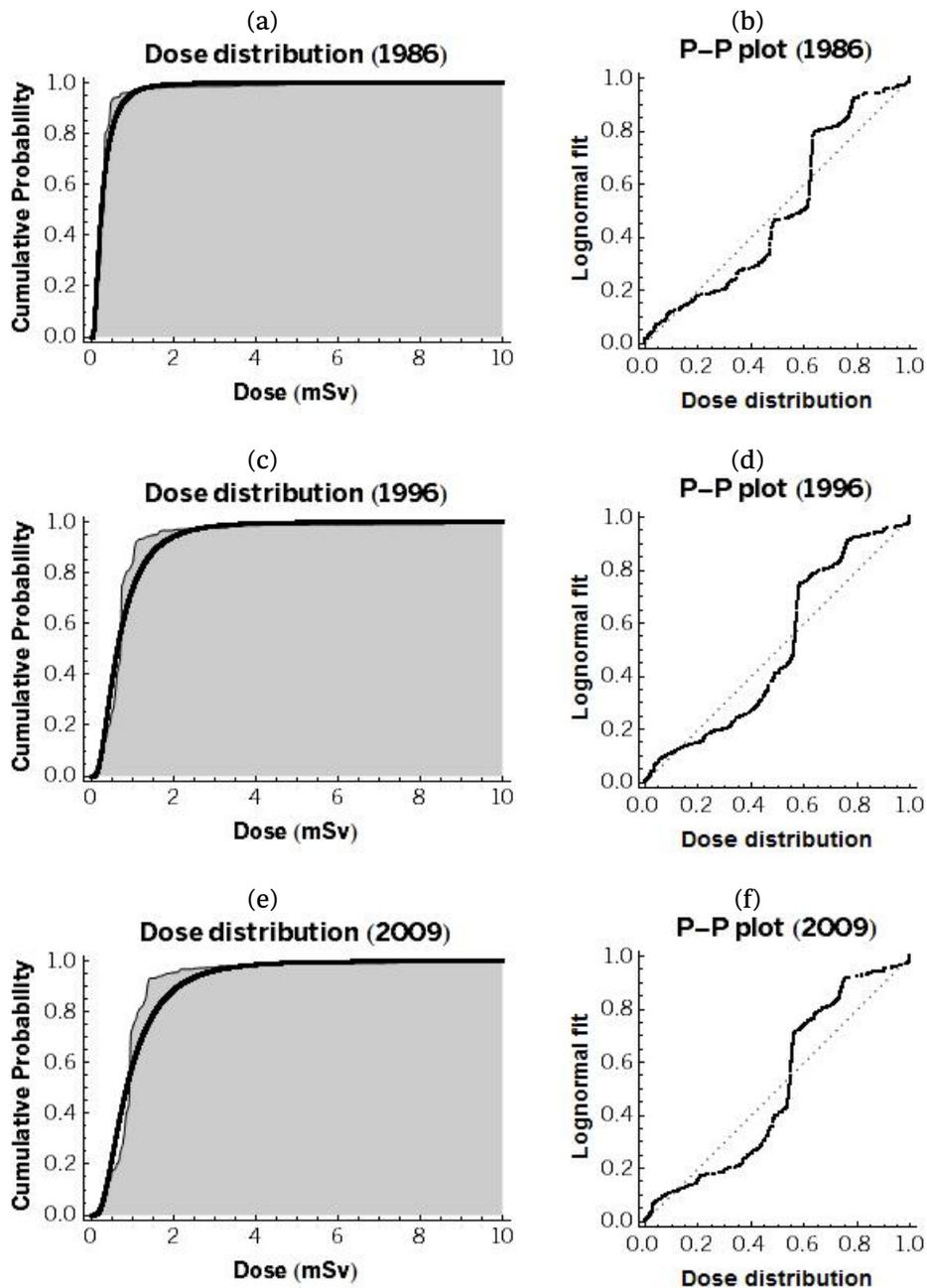
#### 4.2.4 Normality of the distribution of dose

We would expect the distribution of doses in a cohort exposed to a source term with a continuous probability distribution be a continuous probability distribution of some sort. Assuming that the dispersal of radioactive material in the Chernobyl accident can be

approximated with Gaussian plumes (resulting in ground level deposition proportional to the inverse of instantaneous wind speed) we expect that in general, deposition across Europe can be approximated as lognormal [34-39], with the possibility of significant skewness and kurtosis [40]. We would therefore expect that if our cohort was drawn from the population without bias with respect to the source term, the distribution of doses computed should approximate a lognormal distribution.

To investigate this, we attempted to fit a lognormal distribution to the distribution of cumulative effective doses in the cohort at three time points (see *Figure 4-8: Lognormality of dose distribution*, next page).

Note that the lognormal fit(s) shown in those figures seem to approximate the overall shape of the cumulative distribution, but not sufficiently to state that the dose distribution is actually lognormal. Excursions from the diagonal in probability plots suggested that the cumulative dose distributions are not lognormal. We executed statistical tests for each of these distributions, and rejected lognormality at the  $p < 0.01$  level via the Anderson-Darling, Cramér Von-Mises, and Pearson  $\chi^2$  tests.



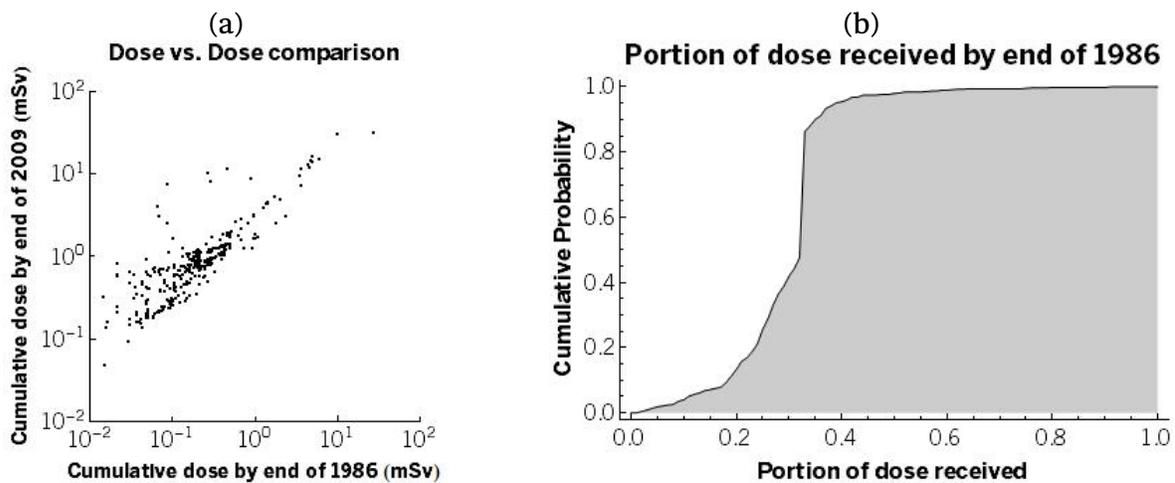
**Figure 4-8: Lognormality of dose distribution**

(a) Distribution of cumulative effective dose by 31 December 1986 with lognormal fit superimposed; (b) probability-probability plot comparing dose distribution by 31 December 1986 with lognormal fit; (c) distribution of cumulative effective dose by 31 December 1996 with lognormal fit superimposed; (d) probability-probability plot comparing dose distribution by 31 December 1996 with lognormal fit; (e) distribution of cumulative effective dose by 31 December 2009 with lognormal fit superimposed; (f) probability-probability plot comparing dose distribution by 31 December 2009 with lognormal fit

Despite the visual appeal of lognormal fits which seem to closely follow the distribution of doses, we are unable to confidently state that the cohort was drawn from the population without some sort of bias with respect to the source term. Due to the discretized nature of the  $^{137}\text{Cs}$  deposition density data extracted from the ATLAS, the source term we are using does not follow a smooth distribution – so we are unable to determine whether the lack of lognormality in the dose distribution is due to bias, methodology, or a different distribution of the source term than expected (i.e., not lognormal).

#### 4.2.5 Dynamics of cumulative dose distribution over time

To gain a sense of the changes in the cohort’s dose distribution over time, shifts in individual cumulative dose and dose rankings within the cohort were examined at different times. Cumulative dose was compared at two time points – the end of the first year on 12/31/1986 versus the end of the study on 12/31/2009. (See *Figure 4-9: Doses received by end of first year versus end of study.*)



**Figure 4-9: Doses received by end of first year versus end of study**

(a) Comparison of doses received by individuals by 31 December 1986 compared to doses received by 31 December 2009; (b) Portion of total cumulative dose received by 31 December 2009 which was imparted by 31 December 1986.

This showed that the majority of the cohort received more than 65% of their dose after the first year. Approximately 10% of the cohort received more than 35% of their dose by

the end of 1986, while just 1% of the cohort received more than 60% of their dose by the end of 1986. Only one individual received more than 80% of their dose by the end of 1986.

Comparisons of ordered dose ranks at different time points reveals that some individuals had profound changes in their cumulative dose relative to the cohort (see *Figure 4-10: Ranks of dose received in first year compared to study period*, next page). For example, 8 individuals jumped from quite low (below 200) in the ordered dose ranking to quite high (above 650) between 1986 and 2009.

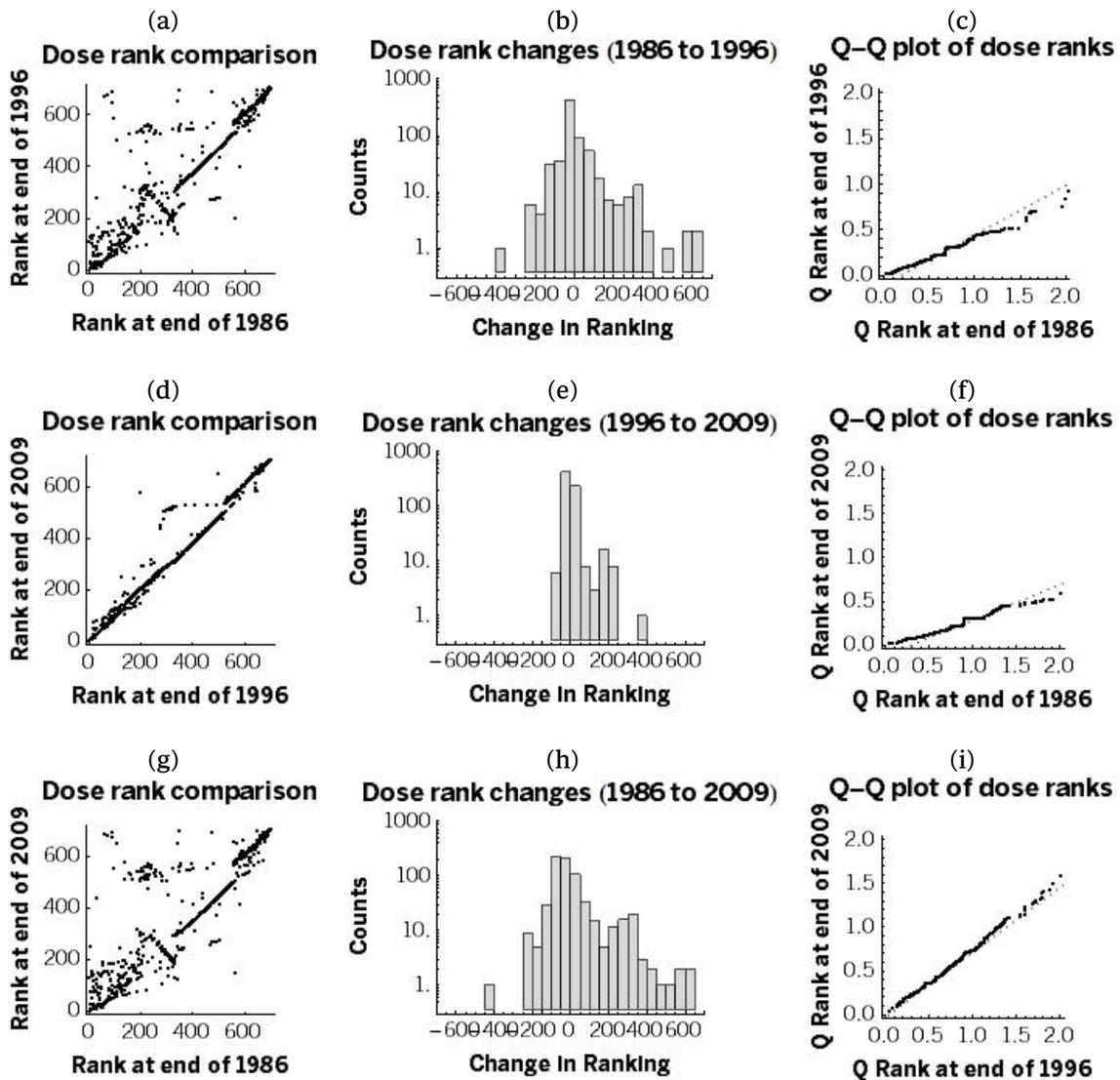
### ***4.3 Outcomes for individuals***

#### **4.3.1 Sample outcomes for individuals with known residence ordering**

Examination of the cumulative dose and dose rate outcomes for individual patients demonstrates that our computation accurately tracks residence and occupation changes as expected:

- Individuals who resided in highly contaminated areas are calculated as receiving significantly more dose than those who lived in areas of low contamination.
- Individuals who moved between areas with different source terms show sudden changes in their dose rates at the time of each move.

Individuals with changes in their age/occupation category show sudden changes in their dose rates at the time of each change.



**Figure 4-10: Ranks of dose received in first year compared to study period**

(a) Rank vs. rank scatterplot of cumulative effective dose received by 31 December 1986 vs. 31 December 2009;

(b) histogram showing distribution of rank changes between 31 December 1986 and 31 December 2009;

(c) Q-Q plot comparing quantile rank of cumulative effective dose received by 31 December 1986 vs. 31 December 2009 with linear fit shown for comparison;

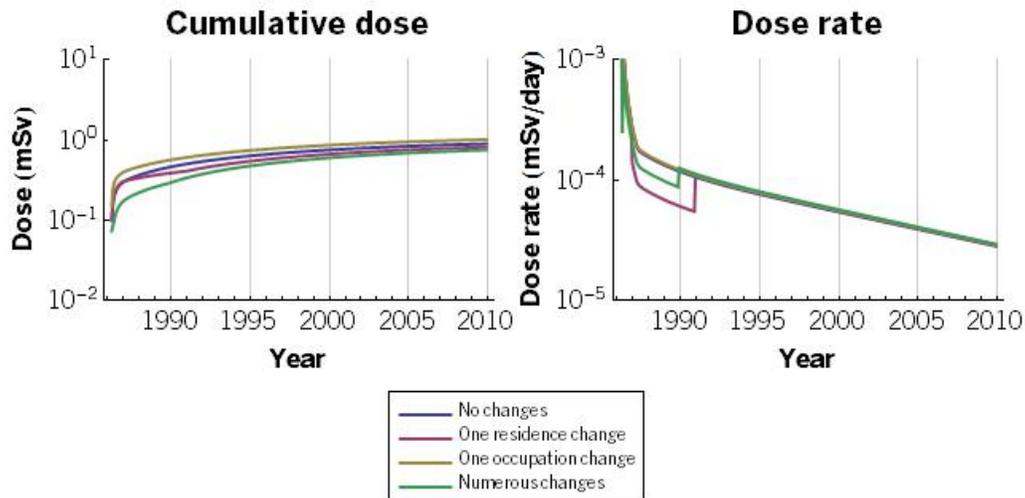
(d-f) same information shown for 31 December 1996 compared to 31 December 2009;

(g-i) same information shown for 31 December 1986 compared to 31 December 2009.

To illustrate residence and occupation change tracking, the dose was determined for four different subjects with known residence order:

- Subject 682, an indoor worker who lived in a single residence in Kiev throughout the study period and did not change occupation;
- Subject 2, an indoor worker who moved once during the study period, but did not change occupation;
- Subject 66, who did not move but changed occupation once during the study period;
- Subject 767, who moved and/or changed occupation a total of seven times during the study period.

Results are shown in *Figure 4-11: Example results for individuals with known residence orders*.



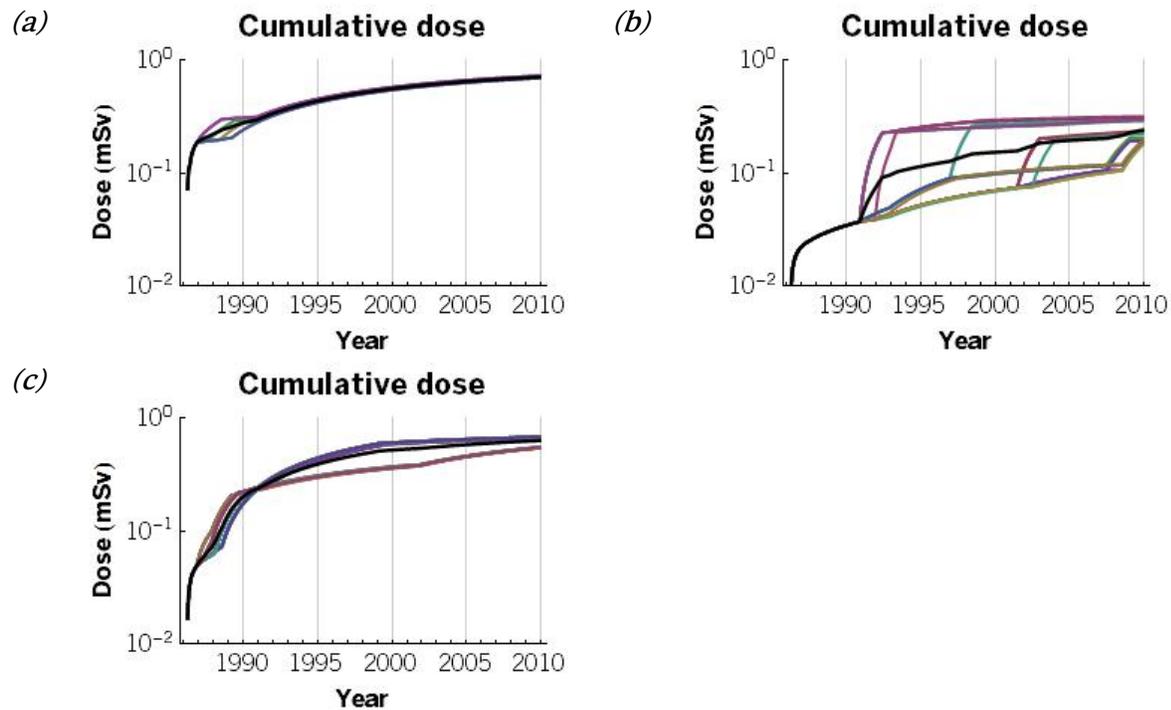
**Figure 4-11: Example results for individuals with known residence orders**

For the three examples that include residence or occupation changes, discontinuities appear in the dose rate charts – in some cases, these lead to profound increases or decreases in dose rates. These discontinuities are expected – when individuals move from one contaminated area to another, or change their occupation (and therefore the portion of time

they spend outdoors), we would anticipate a concomitant and abrupt shift in their dose rates.

#### 4.3.2 Sample outcomes for individuals with unknown residence ordering

It was also informative to examine the dose reconstructions for individuals with significant numbers of residence/occupation records listed in Time Periods 3 and 4 (that is, individuals subject to residence order permutations as described in *Section 3.2.5*) *Figure 4-12* shows sample cumulative dose for individuals with moves or occupation changes in Time Period 3 only, Time Period 4 only, and both time periods.



**Figure 4-12: Doses for sample cases with indeterminate record orderings**

*Colored lines show the cumulative dose computed for each possible ordering; thick black line shows the mean of all possible cumulative doses. (a) Example of an individual with unknown residence/occupation orderings in Time Period 3 only; (b) example of an individual with unknown residence/occupation orderings in Time Period 4 only; (c) Example of an individual with unknown residence/occupation orderings in both Time Period 3 and Time Period 4.*

## 4.4 Uncertainty and sensitivity analysis

### 4.4.1 Uncertainty analysis for dose and dose rates

In order to find the uncertainty in our estimates for dose, we employed Monte Carlo simulation software developed for this purpose (see *APPENDIX E - Monte Carlo and sensitivity analysis system design*). We used the Monte Carlo software to estimate the distribution of dose rate and cumulative dose for each day of the study for thirteen cases similar to individuals in the cohort. We selected cases which typified the cohort by choosing a range of different occupations, ages, source term deposition densities at residence locations, and changes in residence locations and/or occupations (see *Table 4-2* below).

**Table 4-2: Case studies for error estimation and sensitivity analysis**

Case	Description
1	Low deposition range (1-2 kBq/m <sup>2</sup> ); indoor worker
2	Medium deposition range (10-20 kBq/m <sup>2</sup> ); child at time of accident
3	Medium deposition range (10-20 kBq/m <sup>2</sup> ); adolescent at time of accident
4	Medium deposition range (10-20 kBq/m <sup>2</sup> ); indoor worker
5	Medium deposition range (10-20 kBq/m <sup>2</sup> ); pensioner at time of accident
6	Medium deposition range (10-20 kBq/m <sup>2</sup> ); outdoor worker at time of accident
7	High deposition range (100-185 kBq/m <sup>2</sup> ); indoor worker
8	Numerous moves and occupation changes
9	Move from high (555-1480 kBq/m <sup>2</sup> ) to low (2-4 kBq/m <sup>2</sup> ) deposition early (day 100)
10	Move from high (555-1480 kBq/m <sup>2</sup> ) to low (2-4 kBq/m <sup>2</sup> ) deposition late (day 1000)
11	Move from low (2-4 kBq/m <sup>2</sup> ) to high (555-1480 kBq/m <sup>2</sup> ) deposition late (day 1000)
12	Move from low (2-4 kBq/m <sup>2</sup> ) to high (555-1480 kBq/m <sup>2</sup> ) deposition early (day 100)
13	Very high deposition range (555-1480 kBq/m <sup>2</sup> ); indoor worker

We loaded our model for determining dose rate (described in *Section 3.2.3*) into the Monte Carlo software. The complete symbolic function for this model is revisited below:

$$\dot{D}(t) = 86400 \cdot f_{age,occ} \cdot X(age) \cdot \dot{E}(t) \cdot r_{soil}(t) \cdot r_{137Cs}(t) \cdot C_0, \text{ where:}$$

- $f_{age,occ}$  is based on listed occupation/age and is lognormally distributed;
- $X(age) = \alpha \cdot \left( age_0 + \frac{t}{365.25} \right)^{-\beta}$ , where  $\alpha$  and  $\beta$  are normally distributed;
- $\dot{E}(t) = \begin{cases} \frac{1}{30} A_0 \int_{30}^{370} A_1 t^{-B_1}, & 0 \leq t \leq 30 \\ A_1 t^{-B_1}, & 30 < t \leq 365 \\ A_2 e^{-B_2 t} + A_3 e^{-B_3 t}, & 365 < t \end{cases}$ , where each  $A_i, B_i$  is normally distributed;
- $r_{soil}(t) = C \left( A e^{-\left(\frac{\log(2)}{T_1}\right)t} + (1 - A) e^{-\left(\frac{\log(2)}{T_2}\right)t} \right)$ , where  $A, C, T_1$  and  $T_2$  are each normally distributed;
- $r_{137Cs}(t) = e^{-\frac{\ln(2)}{T_{1/2,Cs}} t}$ , with  $T_{1/2,Cs}$  normally distributed;
- $C_0$  is a value found from the ATLAS.

We defined the distributions for each dependent parameter listed above ( $\alpha, \beta, A_0$ , et cetera) using the values listed in *Section 3.2.3* and summarized in *Table 4-3* (next page).

**Table 4-3: Values used for Monte Carlo input parameter distributions**

Parameter	Distribution type	$\mu$	$\sigma$	Units	Source and assumptions
$f_{children}$	lognormal	0.13	0.457	n/a	From published values
$f_{adolescents}$	lognormal	0.18	0.565	n/a	From published values
$f_{indoor}$	lognormal	0.26	0.536	n/a	From published values
$f_{outdoor}$	lognormal	0.38	0.419	n/a	From published values
$f_{pensioner}$	lognormal	0.26	0.728	n/a	From published values
$\alpha$	normal	0.917	0.00613	$Sv/Gy$	Power curve fit of published data; made assumption of error in published MCNP values on the order of 1%
$\beta$	normal	0.0507	0.00238	n/a	
$age_0$	uniform	n/a	n/a	y	
$A_0$	normal	0.56	0.11	n/a	From published values
$A_1$	normal	188	33.7	$\frac{pGy/s}{kBq/m^2}$	Repeated published curve fit
$A_2$	normal	2800	280	$\frac{pGy/s}{kBq/m^2}$	Made assumption of normal distribution with CV=0.1
$A_3$	normal	0.54	0.054	$\frac{pGy/s}{kBq/m^2}$	Made assumption of normal dist. with CV=0.1
$B_1$	normal	0.98	0.043	n/a	Repeated published curve fit
$B_2$	normal	0.02958	0.002958	$d^{-1}$	Made assumption of normal dist. with CV=0.1
$B_3$	normal	0.0000769	0.0000769	$d^{-1}$	Made assumption of normal dist. with CV=0.1
$A$	normal	0.4	0.05	n/a	From published values
$C$	normal	0.82	0.082	n/a	Made assumption of normal dist. with CV=0.1
$T_1$	normal	1.5	0.15	y	Made assumption of normal dist. with CV=0.1
$T_2$	normal	50	5	y	Made assumption of normal dist with CV=0.1
$T_{1/2,Cs}$	normal	11018.3	9.5	d	From published values
$C_0$	uniform	n/a	n/a	$kBq/m^2$	Bounds set at the lower and upper value of $^{137}Cs$ deposition from ATLAS

We defined the following distributions based on the specifics of the case of interest:

- Attained age at the time of the accident ( $age_0$ ) was taken as a uniform distribution of  $\pm 6$  months around the listed age of the case of interest. Use of a uniform distribution here rather than an exact value was due to previously identified issues with reliability of individual age data in the survey dataset.
- $^{137}\text{Cs}$  deposition in soil ( $C_0$ ) was taken as a uniform distribution between the lower and upper bound shown in the contour maps of the ATLAS for the residence location(s) of the case of interest.

We will refer to the computation of cumulative dose and dose rate for each day of the study period (from 26 April 1986 to 31 December 2009) for a given set of input parameters as a history. For each case study, we computed  $10^3$  histories, randomly sampling once from the input distribution prior to computation of each history.

We elected to compute  $10^3$  histories for each case due to the computation time necessary for this process. To validate the suitability of examining  $10^3$  histories (rather than a larger number), we executed  $10^4$  iterations for several cases. The results from the higher number of iterations was comparable to those for  $10^3$  iterations, with differences on the order of 1% for all outcomes.

Some of our cases which involved changes in the  $^{137}\text{Cs}$  soil deposition source term (representing a change in residence location) or age/occupation category at specific time points. For these cases, the changing parameters were resampled at those specific time points, but all other parameters were not resampled. An example of the algorithmic process flow is shown in *Table 4-4* (next page).

**Table 4-4: Example of Monte Carlo process flow**

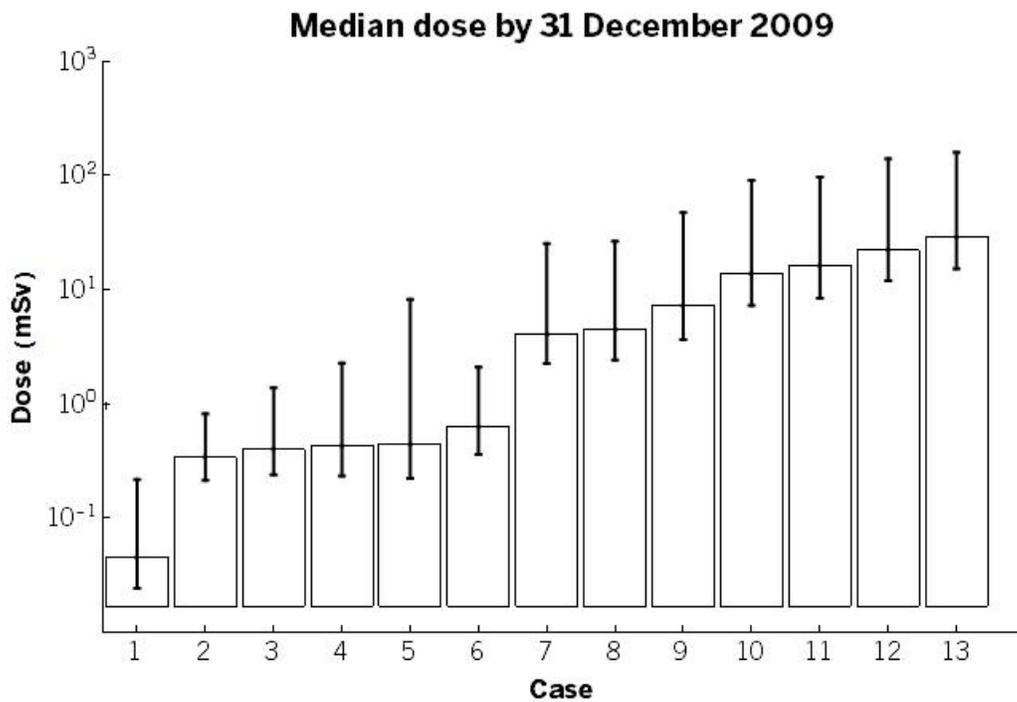
Case 9 presents an individual who moved from a residence with a high  $^{137}\text{Cs}$  deposition density (555-1480 kBq/m<sup>2</sup>) to one with a low  $^{137}\text{Cs}$  deposition density (2-4 kBq/m<sup>2</sup>) on the 100<sup>th</sup> day after the accident.

For this case, our Monte Carlo algorithm proceeded as follows:

- 1) The distributions for all input parameters were sampled  $10^3$  times, including the distribution for  $C_0$ , which is initially set as a uniform distribution with a lower bound of 555 kBq/m<sup>2</sup> and an upper bound of 1480 kBq/m<sup>2</sup>.
- 2) The values of these input parameter samplings were used to compute histories ( $10^3$  in total) between Day 1 and Day 100 as follows: For an arbitrary day  $t$ , the first history was computed by selecting the first value from each list of input parameter samplings. These values (all parameters and the value for  $t$ ) were input into the dose assessment model, and dose and dose rate for the first history for day  $t$  determined. Then the second history was computed by selecting the second value from each list of input parameter samplings, inputting these values into the dose assessment model, and computing dose for the second history for day  $t$  determined. This was repeated for all  $10^3$  histories for each day between Day 1 and Day 100.
- 3) Before computing histories for Day 100 and beyond, the system paused and redefined the distribution for  $C_0$ , setting its lower bound to 2 Kbk/m<sup>2</sup> and its upper bound to 4 kBq/m<sup>2</sup> and drawing  $10^3$  samples from the new distribution. No other distributions were resampled.
- 4) Now employing the new input parameter sampling for  $C_0$ , but changing nothing else, computation of histories continued, determining dose and dose rate for each history between Day 100 and the final day of the study period (Day 8650) using the method described in (2) above.
- 5) All histories ( $10^3$  in total) computed in this manner for this case were returned as output.

#### 4.4.2 Outcomes of Monte Carlo simulations

The distribution of computed effective dose and dose rates were found to be lognormally distributed for all cases at each time point (using a Wilson U<sup>2</sup> test for lognormality at  $p \leq 0.05$ ). The span of cumulative dose for all thirteen cases at the end of the study are shown below in *Figure 4-13*.



**Figure 4-13: Estimates of dose from Monte Carlo simulations**

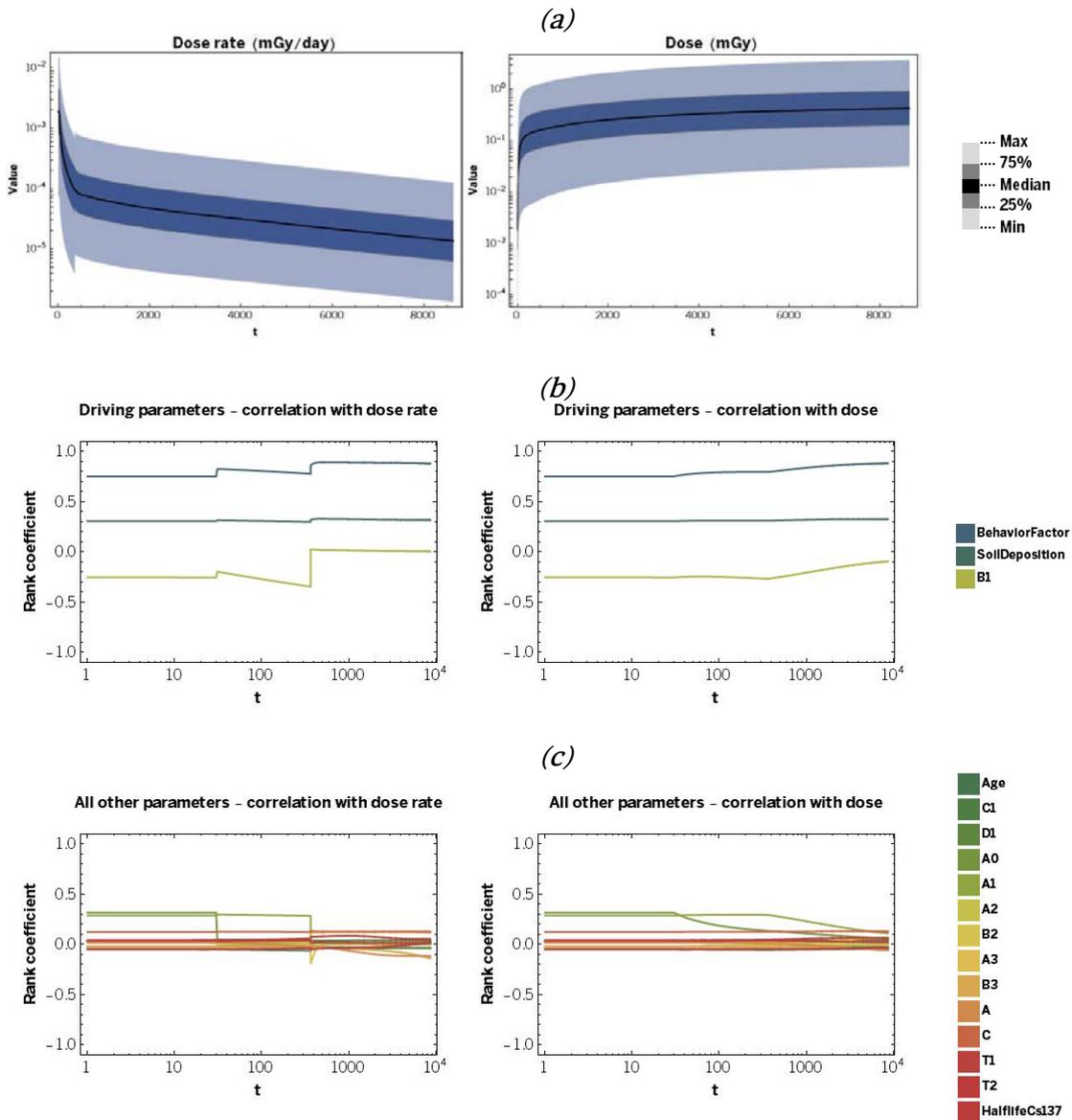
*Median dose for 13 case studies are shown; error bars show 5%-95% confidence interval for each case. Complete case descriptions are given in Table 4-5.*

In all cases, the coefficient of variation of the sample was approximately 1, and slightly less than 1 for most cases. The 5% and 95% quantiles of the distributions of our estimates for both cumulative effective dose and dose rate fell within one order of magnitude of the computed mean for these values for all cases except for Case 5. This case study explores the “pensioner” subcategory of the behavior factor input parameter. The broader distribution of dose for this case demonstrates the influence of greater uncertainty in the measurement of this input parameter compared to other age/occupation categories.

Outcomes for all thirteen cases at the same three time points of interest shown in *Table 4-1* are shown in *Table 4-5*; examples of the complete results for two cases are shown in *Figure 4-14* and *Figure 4-15* (next two pages).

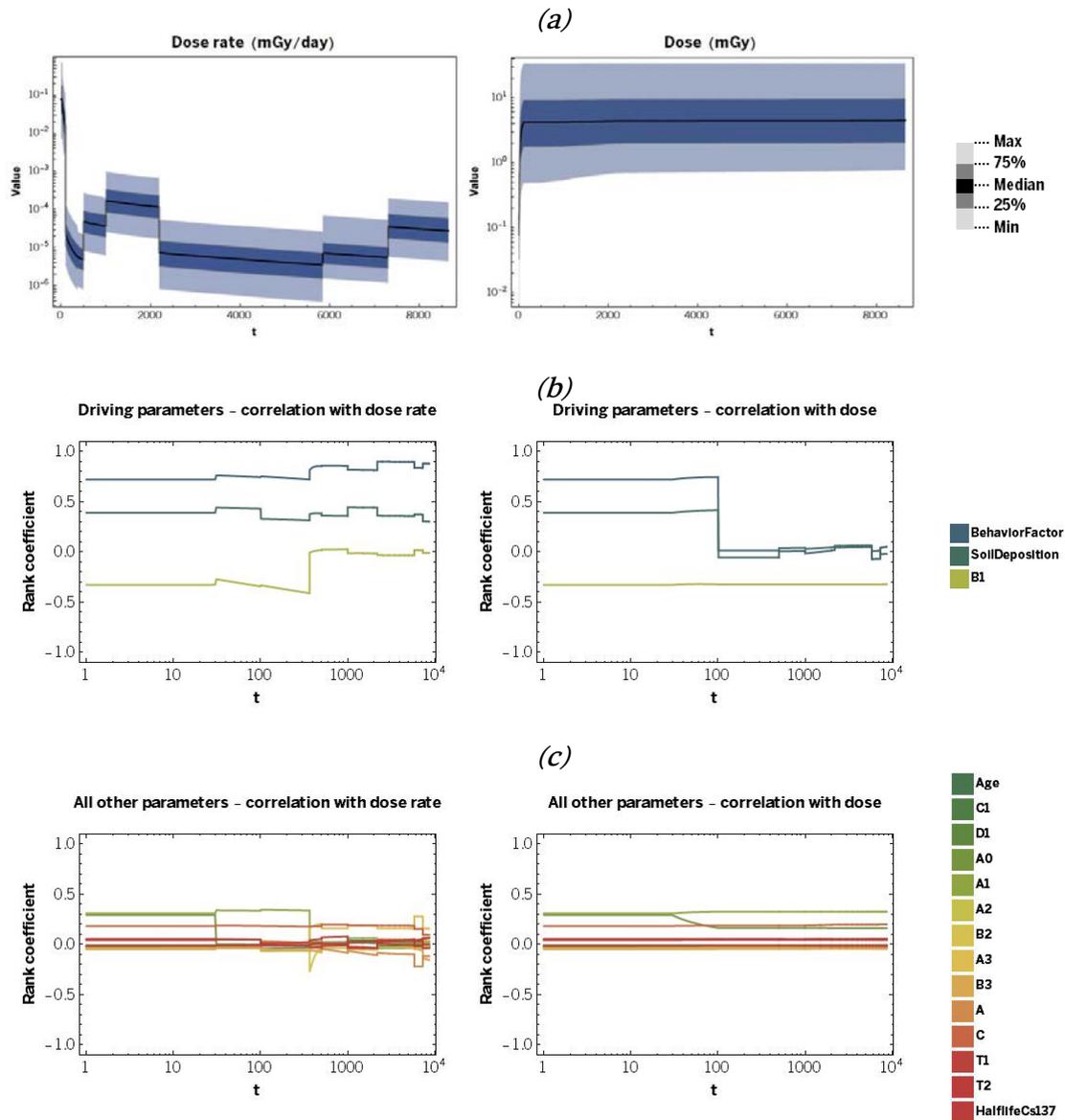
**Table 4-5: Estimated mean and coefficient of variation for selected cases**

Case	Description	12/31/1986		12/31/1996		12/31/2009	
		Mean (Median)	CV	Mean (Median)	CV	Mean (Median)	CV
1	Low deposition range (1-2 kBq/m <sup>2</sup> ); indoor worker	0.017 (0.015)	0.66	0.041 (0.035)	0.60	0.053 (0.045)	0.60
2	Medium deposition range (10-20 kBq/m <sup>2</sup> ); child at time of accident	0.102 (0.084)	0.65	0.249 (0.227)	0.48	0.364 (0.335)	0.40
3	Medium deposition range (10-20 kBq/m <sup>2</sup> ); adolescent at time of accident	0.132 (0.104)	0.78	0.315 (0.268)	0.64	0.449 (0.394)	0.51
4	Medium deposition range (10-20 kBq/m <sup>2</sup> ); indoor worker	0.167 (0.137)	0.71	0.393 (0.333)	0.67	0.514 (0.430)	0.67
5	Medium deposition range (10-20 kBq/m <sup>2</sup> ); pensioner at time of accident	0.201 (0.143)	1.0	0.469 (0.345)	0.96	0.612 (0.448)	0.95
6	Medium deposition range (10-20 kBq/m <sup>2</sup> ); outdoor worker at time of accident	0.232 (0.201)	0.61	0.546 (0.481)	0.54	0.714 (0.634)	0.54
7	High deposition range (100-185 kBq/m <sup>2</sup> ); indoor worker	1.645 (1.315)	0.73	3.853 (3.108)	0.68	5.034 (4.089)	0.68
8	Numerous moves and occupation changes	5.034 (4.173)	0.71	5.265 (4.401)	0.68	5.336 (4.467)	0.67
9	Move from high (555-1480 kBq/m <sup>2</sup> ) to low (2-4 kBq/m <sup>2</sup> ) deposition early (day 100)	8.901 (7.185)	0.80	8.949 (7.225)	0.79	8.974 (7.252)	0.79
10	Move from high (555-1480 kBq/m <sup>2</sup> ) to low (2-4 kBq/m <sup>2</sup> ) deposition late (day 1000)	11.572 (9.483)	0.73	16.762 (13.873)	0.70	16.787 (13.903)	0.70
11	Move from low (2-4 kBq/m <sup>2</sup> ) to high (555-1480 kBq/m <sup>2</sup> ) deposition late (day 1000)	0.034 (0.028)	0.69	10.961 (9.074)	0.65	19.584 (16.164)	0.65
12	Move from low (2-4 kBq/m <sup>2</sup> ) to high (555-1480 kBq/m <sup>2</sup> ) deposition early (day 100)	2.904 (2.307)	0.78	18.602 (15.221)	0.70	26.992 (22.201)	0.70
13	Very high deposition range (555-1480 kBq/m <sup>2</sup> ); indoor worker	11.487 (9.285)	0.72	26.791 (22.157)	0.66	34.927 (29.152)	0.65



**Figure 4-14: Results of Monte Carlo analysis for sample individual – Case 4**

*Case: Medium deposition range (10-20 kBq/m<sup>2</sup>); indoor worker. (a) plot of values computed for output dose rate and dose distributions, showing minimum, maximum, and mean computed values over time with 10%-90% quantile range highlighted; (b) results of computation of Spearman rank correlation coefficient for dose rate and dose with behavior factors, soil deposition, and a coefficient from the air kerma model (B1); (c) results of computation of Spearman rank correlation coefficient for dose rate and dose with all other input parameters.*



**Figure 4-15: Results of Monte Carlo analysis for sample individual – Case 8**

*Case: Numerous moves and occupation changes. (a) plot of values computed for output dose rate and dose distributions, showing minimum, maximum, and mean computed values over time with 10%-90% quantile range highlighted; (b) results of computation of Spearman rank correlation coefficient for dose rate and dose with behavior factors, soil deposition, and a coefficient from the air kerma model (B1); (c) results of computation of Spearman rank correlation coefficient for dose rate and dose with all other input parameters.*

#### 4.4.3 Input parameter sensitivity

To determine which input parameters most influence the output of our dose assessment model, we completed a sensitivity analysis on the results of our Monte Carlo simulations described above.

We chose Spearman's rank correlation coefficient as our measure of association between the distributions of input parameters to the distribution of computed dose and dose rate. This nonparametric method provides a measure of monotonic association between two distributions: A positive value indicates that increases in one distribution correlate to increases in the other distribution, while a negative value indicates that increases in one distribution correlate to decreases in the other distribution. Values for the coefficient itself range between -1 and +1, where a value of -1 indicates perfect monotonic association between increases in the values of one distribution and decreases in the other distribution, and a value of +1 indicates perfect monotonic association between increases in both distributions. A value of 0 indicates no association between changes in values in both distributions.

We examined the results from the Monte Carlo simulations for the cases described above in *Table 4-5*; example outcomes are shown above in *Figure 4-14* and *Figure 4-15*. For every day of the study period, we determined the Spearman rank correlation coefficient for each input parameter in comparison to both dose and dose rate.

The three input parameters found to be most significantly correlated to cumulative dose at all time points in all cases which did not involve a residence location or occupation change were: Behavior factors, soil deposition, and one coefficient (symbolically referred to as  $B_1$ ) from the air kerma conversion model described in *Section 3.2.3.4*. The parameter  $B_1$  is a value which controls the reduction in air kerma rate in the first year after the Chernobyl accident, when both short- and long-lived radionuclides contributed to dose.

For cases which did involve a residence location or occupation change, the input parameters associated with this change (soil deposition if a location change; behavior factor if an occupation change) cannot be meaningfully correlated with cumulative dose, as there is not a clear choice of which of the possible choices of input distributions for that parameter to use for comparison. In these cases, however, the same three input parameters are more correlated with dose rates than other input parameters for the first 365 days of the study period. After this point in time, the coefficient  $B_1$  is no longer used in the dose rate assessment methodology, and is therefore no longer correlated with dose rate (although it does remain correlated with cumulative dose).

We thus conclude that the input parameters most strongly correlated with the distribution of estimates for cumulative dose and dose rate are those for behavior factors, soil deposition, and one parameter ( $B_1$ ) from the model for air kerma conversion. Of these three, soil deposition is typically the least closely correlated. This indicates that the distribution of dose using our method is being driven more by uncertainty in individual behavior than it is by our methodology for source term determination.

#### **4.4.4 Limitations of assumptions**

The following assumptions, while necessary to enable error estimation for our dose assessment methodology, are the most likely source of ambiguity in the overall error estimation process:

Note that a number of the distributions listed in *Table 4-3* are for parameters which had no published error estimate. For these cases, we have assumed that the standard error for these parameters is similar to that for other normally distributed parameters in our model with known error. Observing that the coefficients of variation for parameters with known distributions are generally on the order of 0.1, we defined distributions for most

parameters with unknown error as normally distributed around the mean of the published value, with a coefficient of variation of 0.1. The exception to this involved parameters for our absorbed dose conversion model, which were drawn from results of MCNP simulations. For these parameters, which were determined through a curve fit of published values, we assumed a high level of accuracy and defined normal distributions around published means with coefficients of variation of 0.01 prior to curve fitting the values.

Note that we have also assumed that  $^{137}\text{Cs}$  deposition density is uniformly distributed between lower and upper contour bounds shown in the ATLAS. This does not accommodate the complexity of information underlying those published contours, which are actually isolines of a smooth surface of interpolated  $^{137}\text{Cs}$  deposition density data. The distribution of  $^{137}\text{Cs}$  deposition density between isolines is not necessarily uniform. Furthermore, these isolines do not provide real bounds of  $^{137}\text{Cs}$  deposition density, but rather present bounds for estimates produced by interpolation of collected soil sampling data (as described in the ATLAS). Our assumption of uniform distribution is one of necessity, but is another likely source of inaccuracy in the overall error estimation process.

## ***4.5 Results of quality control tests***

### **4.5.1 Comparison to boundary conditions**

We compared the results of dose reconstruction for the entire cohort to the boundary conditions of minimum and maximum possible outcomes as discussed in *Section 3.2.7.3.1 - Comparison to boundary conditions*). We found that all cases fell within the range of possible outcomes, indicating that our reconstruction had produced allowable outcomes. (See *Figure 4-16: Dose in cohort compared to boundary conditions*.)

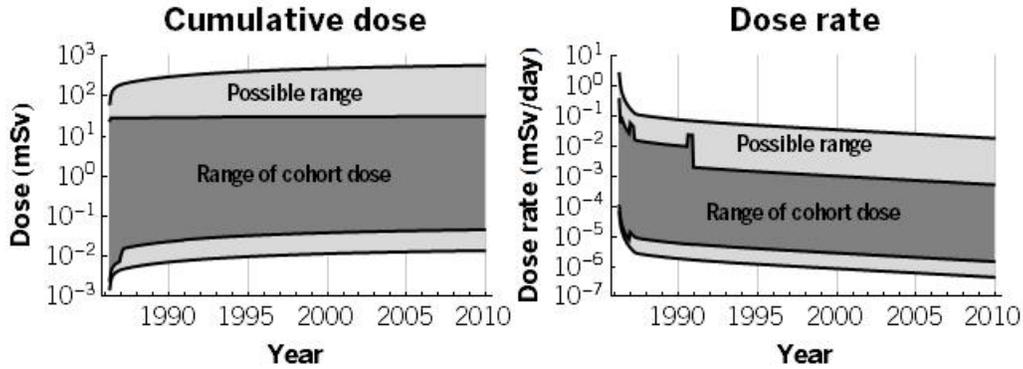
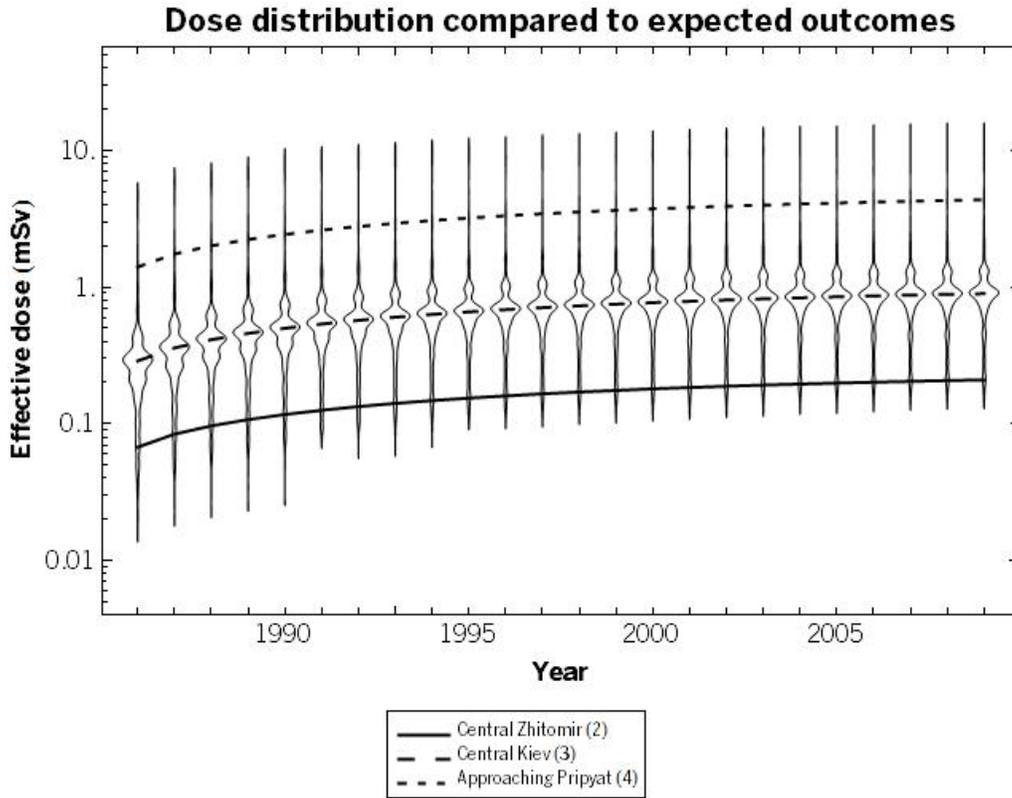


Figure 4-16: Dose in cohort compared to boundary conditions

#### 4.5.2 Comparison to test cases / expected outcomes

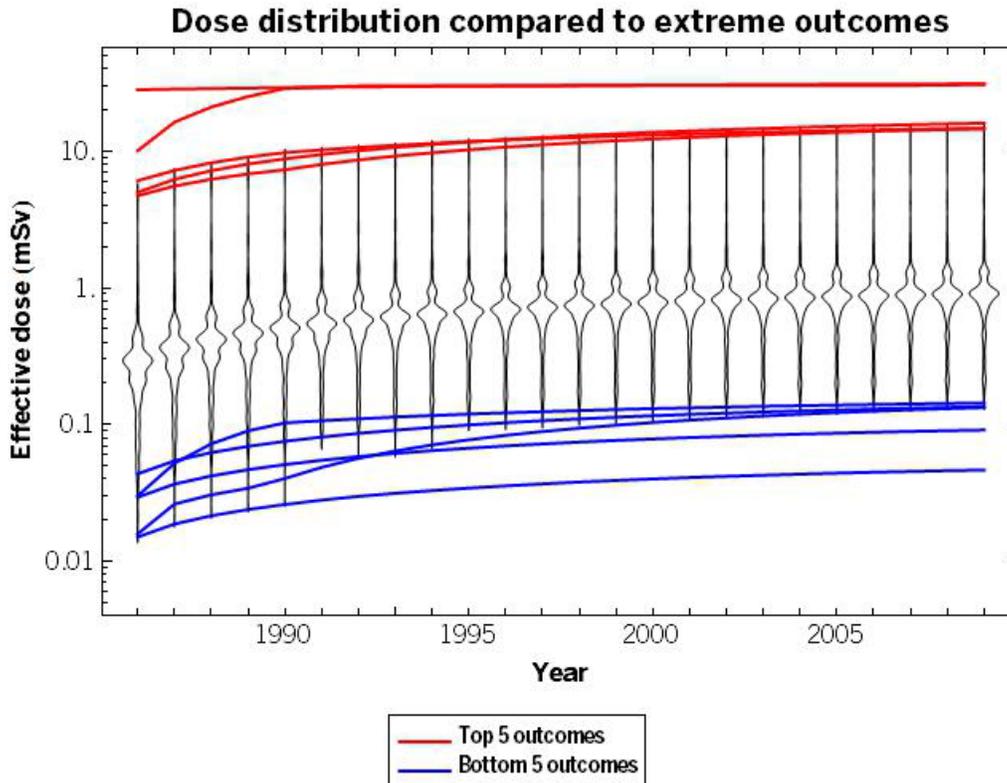
We compared the results of our dose reconstruction for the entire cohort to the expected range of outcomes as discussed in *Section 3.2.7.3.2 - Comparison to test cases*. We observed that the distribution of cumulative effective doses computed for the survey dataset general fell within the range of our prediction that most outcomes would lie between the test case in Zhitomir (with a source term of  $7 \text{ kBq/m}^2$ ) and the test case for the most contaminated areas of Kiev (with a source term of  $142 \text{ kBq/m}^2$ ). The centers of our distributions lie almost exactly along our prediction (the test case in for central Kiev, with a source term of  $25 \text{ kBq/m}^2$ ). (See *Figure 4-17: Cohort dose distribution compared to expected outcomes*) We conclude that overall, our reconstruction produced reasonable outcomes.



**Figure 4-17: Cohort dose distribution compared to expected outcomes**  
*Violin plot showing smoothed distributions of cumulative whole-body, uniformly exposed effective doses by the end of each calendar year, with expected outcomes overlaid.*

#### 4.5.3 Examination of extrema

We examined data for the individuals with the ten most extreme outcomes of cumulative effective dose by the end of the study: The five who received the lowest doses, and the five who received the highest doses.



**Figure 4-18: Cohort effective dose compared to extrema**

*Violin plot showing smoothed distributions of cumulative effective doses by the end of each calendar year, with extreme outcomes overlaid.*

In all cases, we were able to justify why those individuals would have received significantly more or less dose than the rest of the cohort. For detailed descriptions of the justifications found for each individual, see *APPENDIX G - Verification of extrema*.

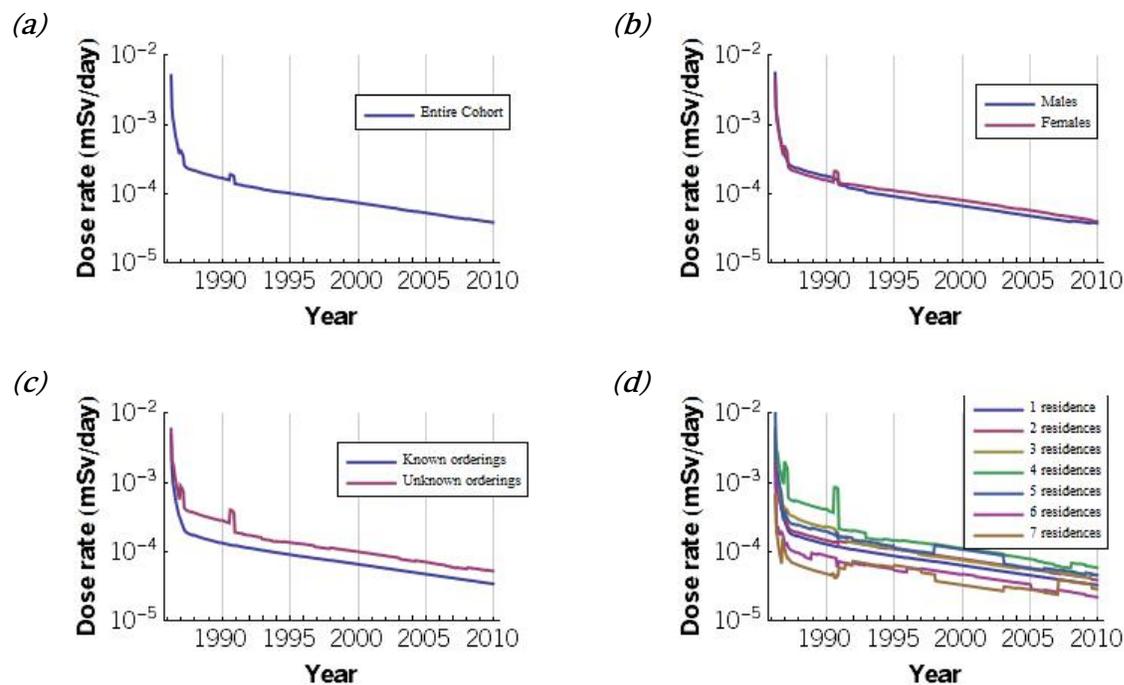
## ***4.6 Factors affecting cumulative dose and dose rate***

### **4.6.1 Contribution from outliers**

Based on inspection of the distribution of the cohort's cumulative effective doses (as shown in *Figure 4-6: Boxplot of cumulative dose distributions at yearly intervals*) we conclude that there was a high likelihood that outliers were contributing disproportionately to the cohort's average cumulative effective dose rate. The high values for skewness and

kurtosis shown in Table 4-1: Summary of statistics for the entire cohort confirmed our suspicions, leading us to investigate individual effective dose rates to see if a small number of extrema were driving the cohort's average effective dose rate.

We stratified the cohort and examined the average dose rates for different strata. *Figure 4-19: Average dose rate across the entire cohort* (next page) shows the average dose rate for individuals in the cohort, stratified by the following criteria: Known vs. unknown residence/occupation record ordering, gender, and number of residences. It is apparent that there are some irregular artifacts in the data (since we would anticipate that the average dose rate should be generally monotonic, and instead see some very distinct peaks), which warrant additional investigation.

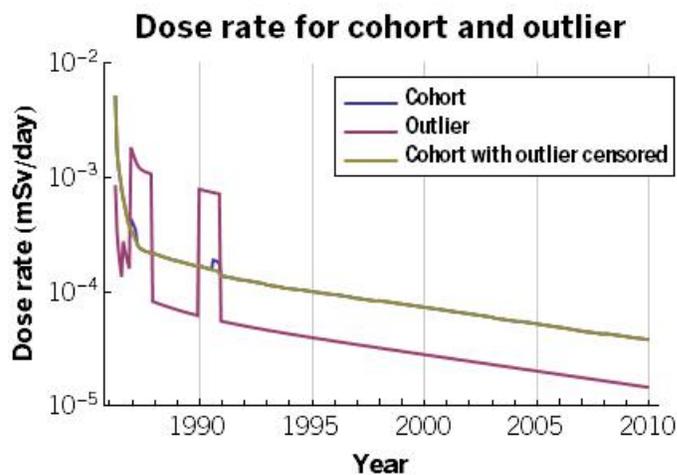


**Figure 4-19: Average dose rate across the entire cohort**

(a) Unstratified; (b) stratified by gender; (c) stratified by known vs. unknown residence/occupation record ordering; (d) stratified by number of residences

Using the stratification criteria for the figures listed above, we can hypothesize that if the dose rates for certain individuals are producing these artifacts, they must be female, have been subject to permutation analysis, and had 4 total residences. Using these clues, the

data set was interrogated to produce dose rate charts for all 21 individuals in the survey dataset matching this description. From this, we are able to determine that the artifacts are indeed produced by exactly one individual's dose rate, which is at times higher than the cohort average by two orders of magnitude -- sufficient to produce significant jumps in the cohort averages. *Figure 4-20: Average dose rate (with and without extreme outlier)* shows the average cohort dose rate with and without censoring this outlier.



**Figure 4-20: Average dose rate (with and without extreme outlier)**

Further investigation shows that this individual's profound dose rate change is due to that individual moving into a highly contaminated region in Belarus for a period of time in TP3.

The double peak of this individual's dose rate was determined to be caused by our approach to resolving the residence/occupation record ordering problem (as described in *Section 3.2.5 - Accommodating indeterminate residence/occupation record ordering*). Because we were unable to determine the correct ordering for this individual's residence/occupation records in TP3, effective dose was computed twice for this individual: Once for the possibility that they moved into the contaminated location at the beginning of Time Period 3 and then moved out after a period of time, and once for the possibility that

they resided outside of Belarus for the beginning of TP3 and then moved into the contaminated location at the end of the TP3. Each of these two cases shows an effective dose rate peak at different times; the average of the two outcomes retains both peaks.

Through this, we observe that individuals who both receive significant dose and have a large number of moves can produce significant noise in the outcome.

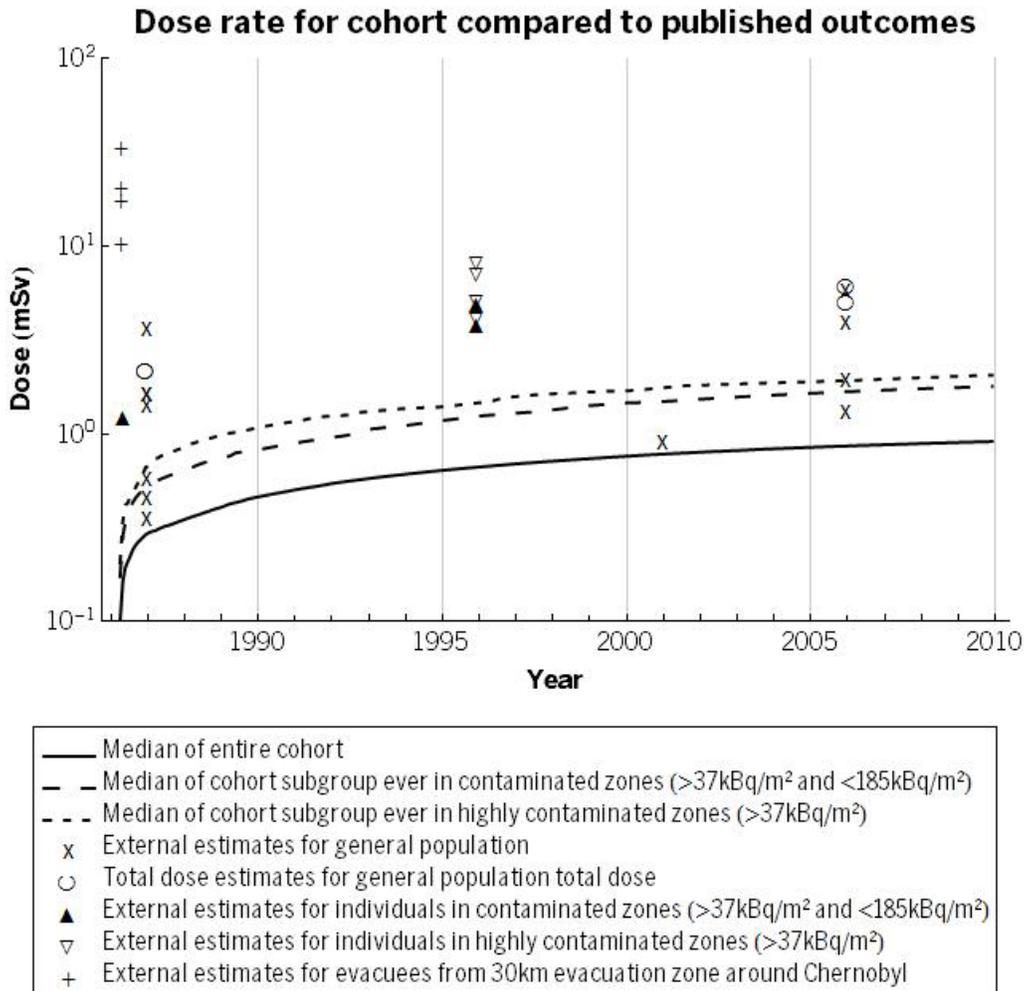
## 5 CONCLUSIONS

### *5.1 Conclusions from this work*

The reconstruction effort undertaken in this project successfully produced a rough estimate for external dose for members of the study cohort. Only a very few individuals exceeded the average background dose from external sources over the entire study period, and none of these accumulated a dose sufficient that we would expect to produce medical pathologies.

We compared the results of our efforts with the results from other published studies that estimated cumulative doses for various groups in the Ukraine from the Chernobyl accident. We included estimates for groups of individuals which should have received considerably more dose than the cohort (evacuees from the 30 km exclusion zone around the site), total dose estimates (including internal dose), and general population estimates. The results are shown in *Figure 5-1: Comparison of results from this study to other published* (next page); source data for comparisons shown is given in *APPENDIX H - Previously published estimates for dose for similar cohorts*.

The previously published estimates shown here for the general population of the Ukraine are within two geometric standard deviations of the median of the results for our cohort using the methodology described in this document. Many of these studies used substantially different methodologies; in particular, some estimates made shortly after the accident are consistently higher than estimates made with more contemporary models and data.



**Figure 5-1: Comparison of results from this study to other published estimates**

We were encouraged to observe that the one estimate for dose to the general population of the Ukraine made using many of the same approaches we employed was in close concordance with our results. That study suggests an average cumulative dose of 0.89 mSv for the time between the day of the accident and 31 December 2000; our estimate for the median during this same time period was 0.79 mSv (with a mean of 1.04 mSv).

We were further encouraged to note that our estimates for individuals exposed to the most contaminated areas (where <sup>137</sup>Cs deposition is greater than 37 kBq/m<sup>2</sup>) are comparable to previously published estimates for individuals in these areas.

## ***5.2 Contribution of external dose to total dose***

Other published reconstruction efforts and references have suggested that in general, internal dose for individuals in contaminated regions of the Ukraine may approximately equal external dose [6, 41, 42]. Organ-specific dose to the thyroid may be significantly higher, perhaps as much as 10 times the external dose on average [18]. We would not expect external dose to be a predictor of internal dose, since external exposure cannot reasonably be expected to be a good indicator of intake hazards (particularly by way of ingestion of foodstuffs, which were distributed across nations or continents). We are unable to state any conclusions about the contribution of effective dose from external sources to effective dose from all sources for our cohort. Further study would be necessary to determine if total effective dose correlates in some way to health effects (physical or psychological/behavioral).

## ***5.3 Future directions***

The author recognizes that there are many approaches that could be taken to produce a dose reconstruction for the Ukrainian population. While the methods described in this document are well-supported by previous publications, it is the opinion of the author that these methods must be improved if they are to be used in the future.

We recognize that our method for source term determination (acquiring the mean value of discretized data shown in the published contour maps of <sup>137</sup>Cs deposition of the ATLAS) is less than ideal. However, if we had the opportunity to improve only one portion of our methodology, we would be best served by producing more refined models encompassing the behavior of members of our cohort, specifically incorporating the portion of time they spent indoors versus outdoors.

The overall approach used here would be most profoundly improved by the following changes:

1. A more comprehensive understanding of the portion of time individuals spend indoors (and shielded somewhat from soil-deposited radiocontaminants) versus outdoors (and fully exposed) on each day of the study period.
2. Full access to the source data used to produce  $^{137}\text{Cs}$  deposition density estimates – the largest source of uncorrectable error in this study is produced by the use of discretized contour maps for source terms rather than kriged surfaces produced from raw soil sampling data.
3. Better refinement of residence location coordinates: The coordinates listed in the data set provided for analysis are not highly specific – when exact latitude/longitude coordinates were not given by a respondent, the coordinates were found by looking up latitude/longitude coordinates for the cities and settlements listed. Better location specificity (by determining more exact coordinates for addresses) would yield a more accurate source term. In very large cities (i.e., Kiev) or settlements adjacent to hot spots, this might result in significant changes to dose reconstruction outcomes.
4. Correction of residence/occupation record orderings in time: A significant flaw in the current reconstruction method is the inability to order the residence locations and associated occupation categories after 1/1/1987. The permutation analysis method used in the current reconstruction system is a workaround, which should be disposed of if at all possible.
5. Incorporation of better specificity for subsidiary models: The previously published models utilized by this dose reconstruction process could be employed with greater accuracy. For example, values for topsoil mixing have been published for multiple regions in the Ukraine [20]. These could be incorporated in the model described in

*Section 3.2.3.3 - Finding an effective indicator source term*, we employed the most general reference values to avoid overcomplicating our methodology. Similar improvements could be made in some of the other models by incorporating slightly more refined versions of the models already used in our process: Accommodating seasonal changes in age-occupation coefficients [20], kerma-to-dose conversions for different genders [25], and possibly other improvements heretofore unidentified.

Finally, the author strongly recommends that any future efforts at dose reconstruction using this system be conducted with a strong emphasis on the role of professionalized data management, well-described data analysis platform architecture, and well-defined workflows with clear chains of custody for project data. The ultimate goal of the epidemiological dosimetrist is a universal platform for dose estimation and reconstruction. Until contemporary researchers adopt best practices for data availability, systems architecture, and workflows, this is a technology that will remain aspirational rather than achievable.

## REFERENCES

1. Foster, R.M., T. Borak, and K. Tierney, *Modeling nuclear disaster risk: The effects of perceived risk and radiation exposure on post-Chernobyl psychosocial and health behavior outcomes in Ukrainian residents*, 2008, National Science Foundation.
2. Perez Foster, R., et al., *Long-Term Effects of External Radiation Exposure and Perceived Risk on Post-Chernobyl Psychosocial and Health Behavior Outcomes in the Ukrainian Population: A PRELIMINARY REPORT*, 2012.
3. De Cort, M., et al., *Atlas of Caesium Deposition on Europe after the Chernobyl Accident*. 1998, Luxembourg: Office for Official Publications of the European Communities.
4. NEA, *Chernobyl: Assessment of Radiological and Health Impacts*. 2002.
5. UNSCEAR, *Sources and Effects of Ionizing Radiation*, 2008.
6. Bennett, B., M. Repacho, and Z. Carr, *Health Effects of the Chernobyl Accident and Special Health Care Programmes*, 2006: Geneva.
7. NAP/NAS, *Exposure of the American People to Iodine-131 from Nevada Nuclear-Bomb Tests: Review of the National Cancer Institute Report and Public Health Implications* 1999: The National Academies Press.
8. Vykhovanets, E.V., et al., *131I dose-dependent thyroid autoimmune disorders in children living around Chernobyl*. Clin Immunol Immunopathol, 1997. 84(3): p. 251-9.
9. Maxon, H. and E. Saenger, *Biological effects of radioiodines on the human thyroid gland*, in *The Thyroid*, L. Braverman and R. Utiger, Editors. 1996, Lippincott-Raven. p. 342-351.
10. Cooper, D., *Treatment of thyrotoxicosis*, in *The Thyroid*, L. Braverman and R. Utiger, Editors. 1991, J.B. Lippincott Co. p. 887-916.
11. Magocsi, P.R., *History of Ukraine: The Land and Its Peoples*. 2010, Toronto: University of Toronto Press.
12. Group, R.E.M., *Chernobyl air concentration and deposition data*, 1998, European Commission Joint Research Center.
13. Shigematsu, I., *THE INTERNATIONAL CHERNOBYL PROJECT AN OVERVIEW*, 1991.
14. Havenaar, J.M., et al., *Perception of risk and subjective health among victims of the Chernobyl disaster*. Social Science & Medicine, 2003. 56(3): p. 569-572.
15. Yamada, M. and S. Izumi, *Psychiatric sequelae in atomic bomb survivors in Hiroshima and Nagasaki two decades after the explosions*. Social Psychiatry and Psychiatric Epidemiology, 2002. 37(9): p. 409-415.
16. Foster, R. and M. Goldstein, *Chernobyl Disaster Sequelae in Recent Immigrants to the United States from the former Soviet Union (FSU)*. Journal of Immigrant and Minority Health, 2007. 9(2): p. 115-124.
17. Foster, R.P., *The Long-Term Mental Health Effects of Nuclear Trauma in Recent Russian Immigrants in the United States*. American Journal of Orthopsychiatry, 2002. 72(4): p. 492-504.
18. Bouville, A., et al., *Radiation dosimetry for highly contaminated Belarusian, Russian and Ukrainian populations, and for less contaminated populations in Europe*. Health Physics, 2007. 93(5): p. 487-501.
19. Likhtariov, I., et al., *Effective doses due to external irradiation from the Chernobyl accident for different population groups of Ukraine*. Health Physics, 1996. 70(1): p. 87-98.

20. Likhtarev, I.A., et al., *Chernobyl accident: Retrospective and prospective estimates of external dose of the population of Ukraine*. Health Physics, 2002. 82(3): p. 290-303.
21. Unterweger, M.P., et al., *Radionuclide Half-life Measurements*, 2010, NIST: NIST Website - <http://www.nist.gov/pml/data/halflife-html.cfm>.
22. Jacob, P., et al., *Organ doses from radionuclides on the ground. Part I. Simple time dependences*. Health Physics, 1988. 54(6): p. 617-33.
23. Jacob, P., H.G. Paretzke, and H. Rosenbaum, *Organ doses from radionuclides on the ground. Part II. Non-trivial time dependences*. Health Physics, 1988. 55(1): p. 37-49.
24. Saito, K. and P. Jacob, *Gamma Ray Fields in the Air Due to Sources in the Ground*. Radiation Protection Dosimetry, 1995. 58(1): p. 29-45.
25. Saito, K., et al., *Calculation of organ doses from environmental gamma rays using human phantoms and Monte Carlo methods. Part 1: Monoenergetic sources and natural radionuclides in the ground*. 1990: Gesellschaft für Strahlen- und Umweltforschung.
26. Saito, K., N. Petoussi-Henss, and M. Zankl, *Calculation of the effective dose and its variation from environmental gamma ray sources*. Health Physics, 1998. 74(6): p. 698-706.
27. Petoussi, N., et al., *Organ Doses for Foetuses, Babies, Children and Adults from Environmental Gamma Rays*. Radiation Protection Dosimetry, 1991. 37(1): p. 31-41.
28. Zankl, M., N. Petoussi, and G. Drexler, *Effective dose and effective dose equivalent--the impact of the new ICRP definition for external photon irradiation*. Health Physics, 1992. 62(5): p. 395-399.
29. Savkin, M.N., et al., *Investigation of the radiation and hygiene situation at settlements Jovtnavoye, Grezlya, and Novo-Markovka of the Poleskiy district*. 1989, Moscow: Institute of Biophysics.
30. De Cort, M., et al., *Plate 19 - Ukrainian map of caesium-137 deposition*, in *Atlas on the caesium deposition across Europe after the Chernobyl accident* 1998, Office for Official Publications of the European Communities: Luxembourg.
31. Division, U.N.S., *Demographic Yearbook, Table 7: Population by age, sex and urban/rural residence: latest available year, 2000-2009*, 2010.
32. Weisstein, E.W. *Hypergeometric Distribution*.
33. *Ionizing radiation exposure of the population of the United States*. NCRP report 2009, Bethesda, MD: National Council on Radiation Protection and Measurements.
34. Limpert, E., W. Stahel, and M. Abbt, *Log-normal Distributions across the Sciences: Keys and Clues*. BioScience, 2001. 51(5): p. 341-352.
35. Koch, A.L., *The logarithm in biology 1. Mechanisms generating the log-normal distribution exactly*. Journal of Theoretical Biology, 1966. 12(2): p. 276-290.
36. Koch, A.L., *The logarithm in biology: II. Distributions simulating the log-normal*. Journal of Theoretical Biology, 1969. 23(2): p. 251-268.
37. Bernarie, M.M., *The Validity of the Log-normal Distribution of Pollutant Concentrations (Paper SU-18D)*, in *2nd International Clean Air Conference* 1971.
38. Blackwood, L., *The lognormal distribution, environmental data, and radiological monitoring*. Environmental Monitoring and Assessment, 1992. 21(3): p. 193-210.
39. Crow E. L. and S. K., *Lognormal Distributions: Theory and Applications*. 1988, New York: Marcel Dekker, Inc.
40. Singh, A.K., A. Singh, and M. Engelhardt, *The Lognormal Distribution in Environmental Applications* 1997, Office of Research and Development.

41. Baloga, V.I., et al., *20 years after Chernobyl catastrophe: Future outlook*, 2006, All-Ukrainian Research Institute of Population and Territories Civil Defence from Technogenic and Natural Emergencies.
42. Likhtarev, I.A., et al., *Internal exposure from the ingestion of foods contaminated by 137Cs after the Chernobyl accident--report 2. Ingestion doses of the rural population of Ukraine up to 12 y after the accident (1986-1997)*. Health Phys, 2000. 79(4): p. 341-57.
43. Gilbert, R.O., *Statistical methods for environmental pollution monitoring* 1987.
44. Loftis, J., *Statistics for Environmental Modeling (lecture notes)*, 2010.
45. Wolfram, S., *Mathematica 8 - System Documentation*, 2010.
46. Cureton, E.E., *The average spearman rank criterion correlation when ties are present*. Psychometrika, 1958. 23(3): p. 271-272.
47. UNSCEAR, *Sources and Effects of Ionizing Radiation*, 2000.
48. Cardis, E., et al. *Estimated long term health effects of the Chernobyl accident*. in *Joint EU, IAEA and WHO International Conference One Decade after Chernobyl*. 1996.
49. Hatch, M., et al., *The Chernobyl Disaster: Cancer following the Accident at the Chernobyl Nuclear Power Plant*. Epidemiologic Reviews, 2005. 27(1): p. 56-66.
50. Drozdovitch, V., et al., *Radiation exposure to the population of Europe following the Chernobyl accident*. Radiat Prot Dosimetry, 2007. 123(4): p. 515-28.
51. Pukkala, E., et al., *Breast cancer in Belarus and Ukraine after the Chernobyl accident*. Int J Cancer, 2006. 119(3): p. 651-8.
52. UNSCEAR, *Sources and Effects of Ionizing Radiation*, 1988.
53. DOE, *Health and Environmental Consequences of the Chernobyl Nuclear Power Plant Accident*. 1989.

**APPENDIX A. Selected plates from the ATLAS**



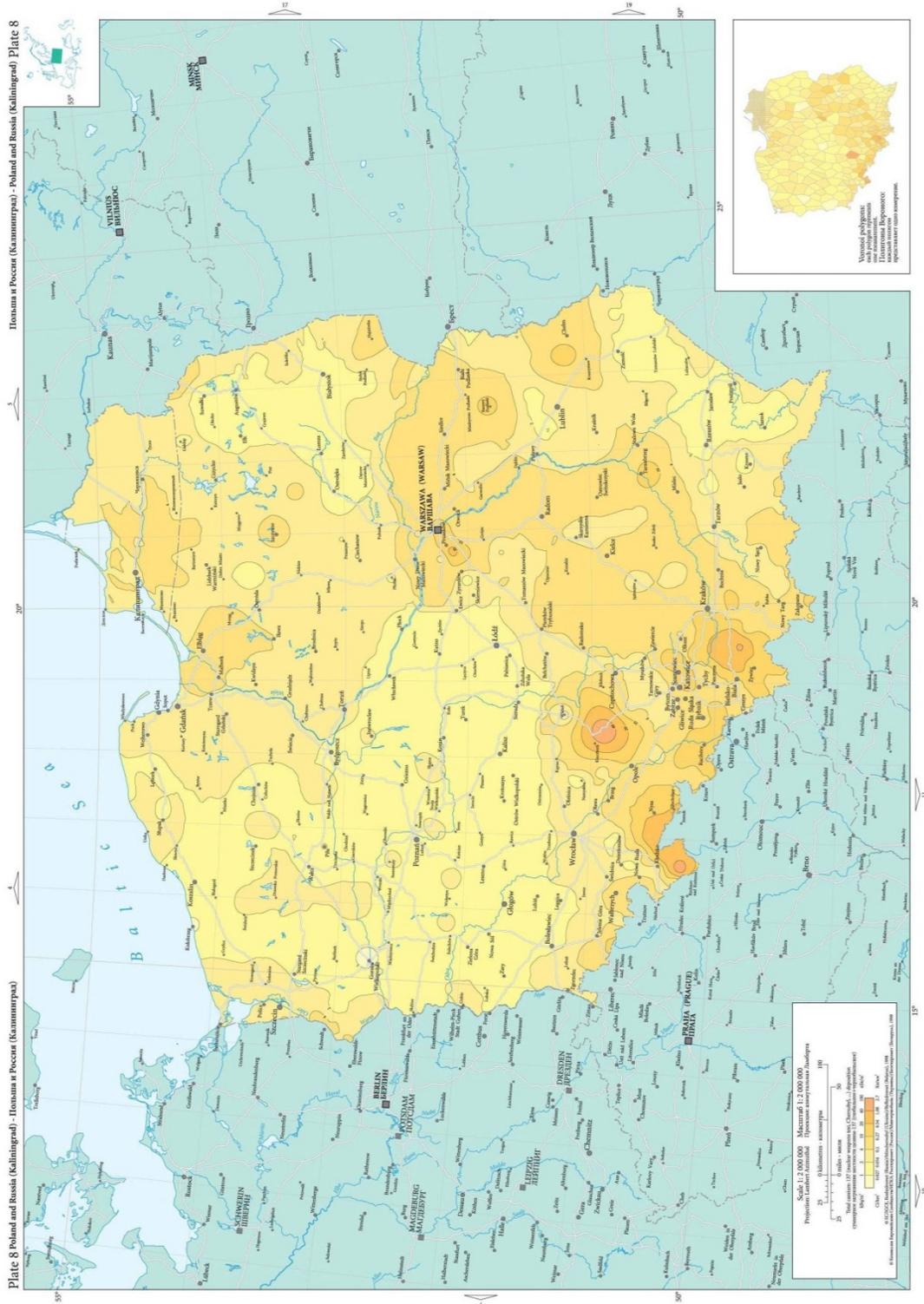


Figure A-2:  $^{137}\text{Cs}$  deposition – Poland



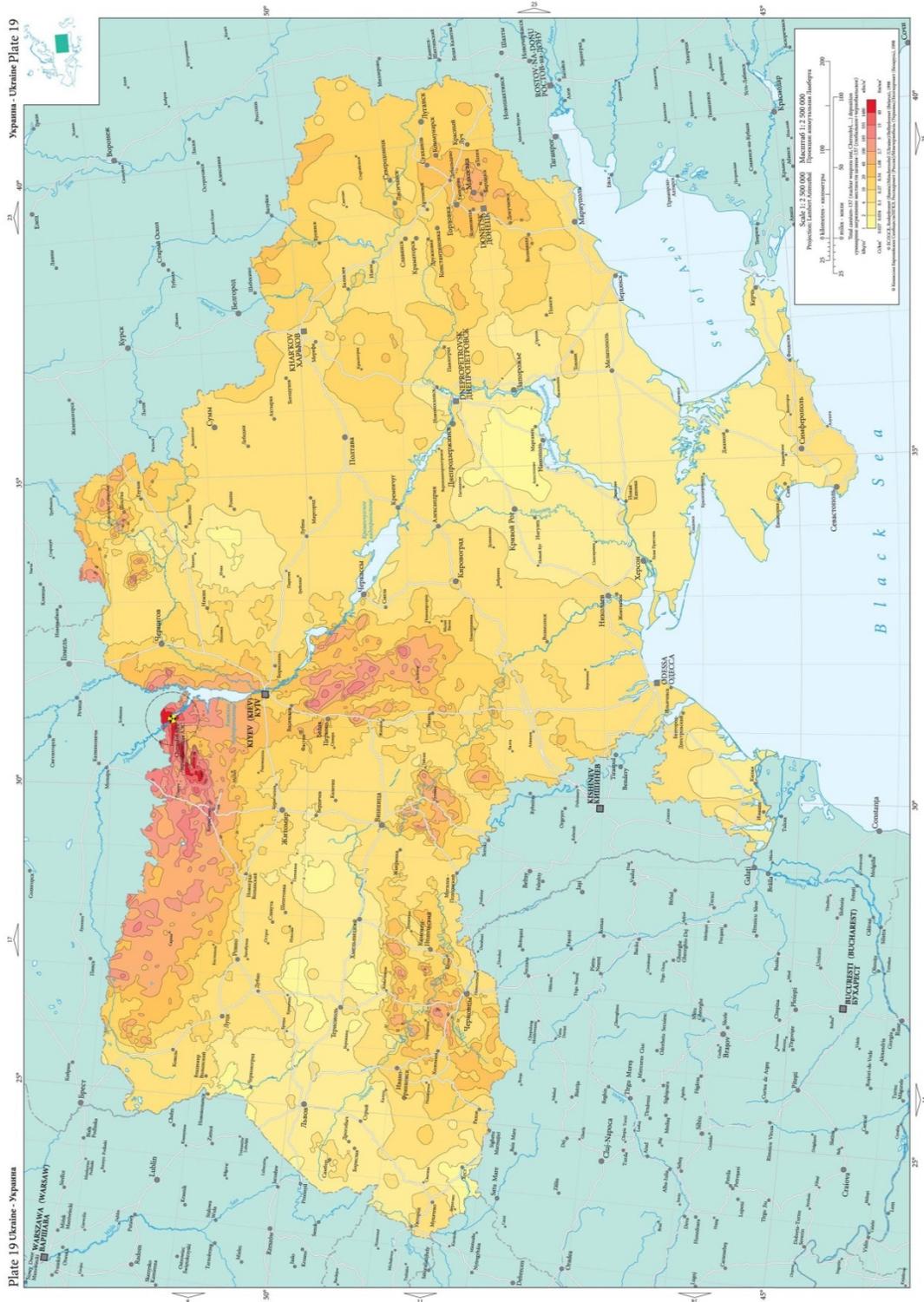


Figure A-4:  $^{137}\text{Cs}$  deposition - Ukraine

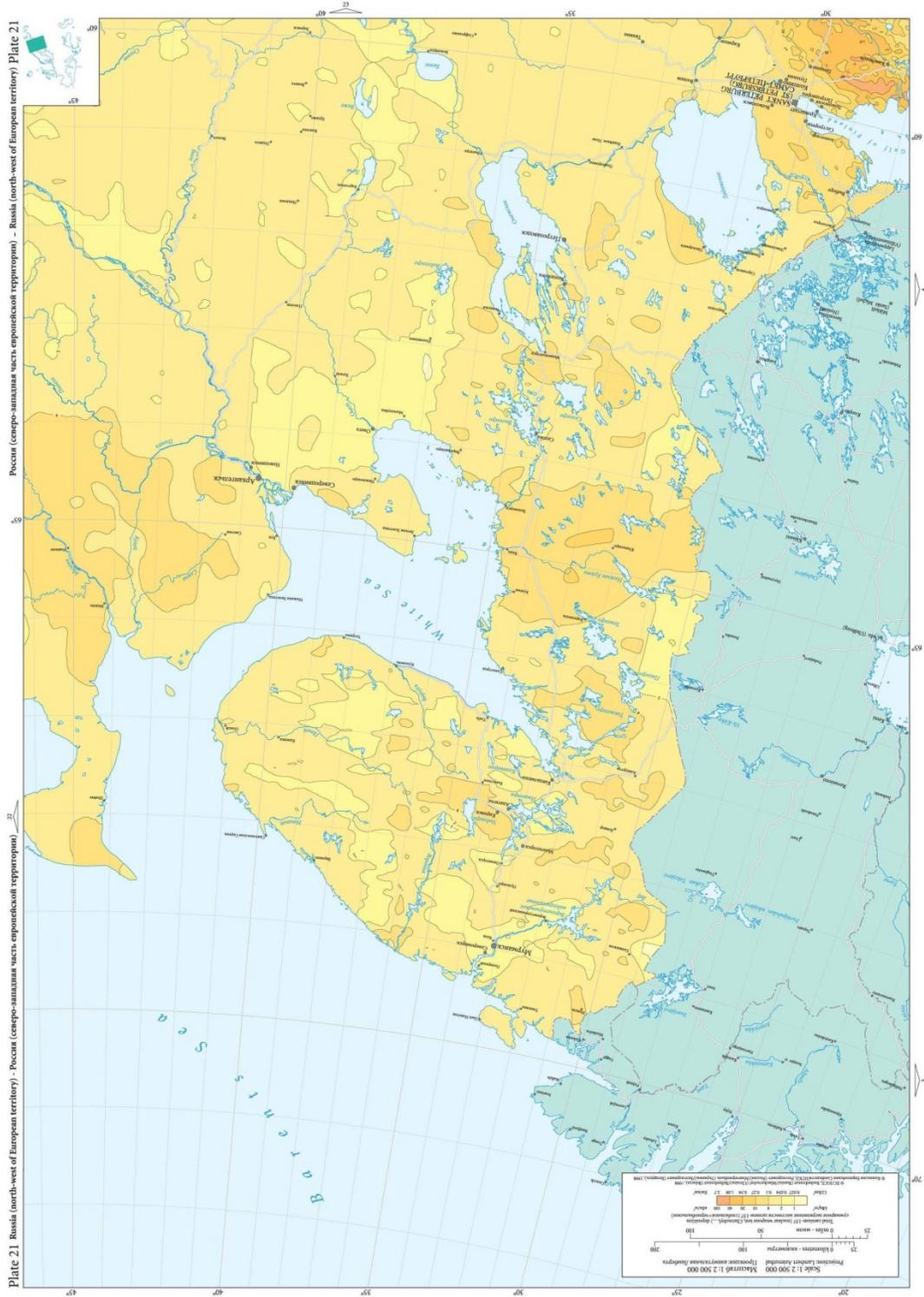


Figure A-5: <sup>137</sup>Cs deposition - Northwestern Russia

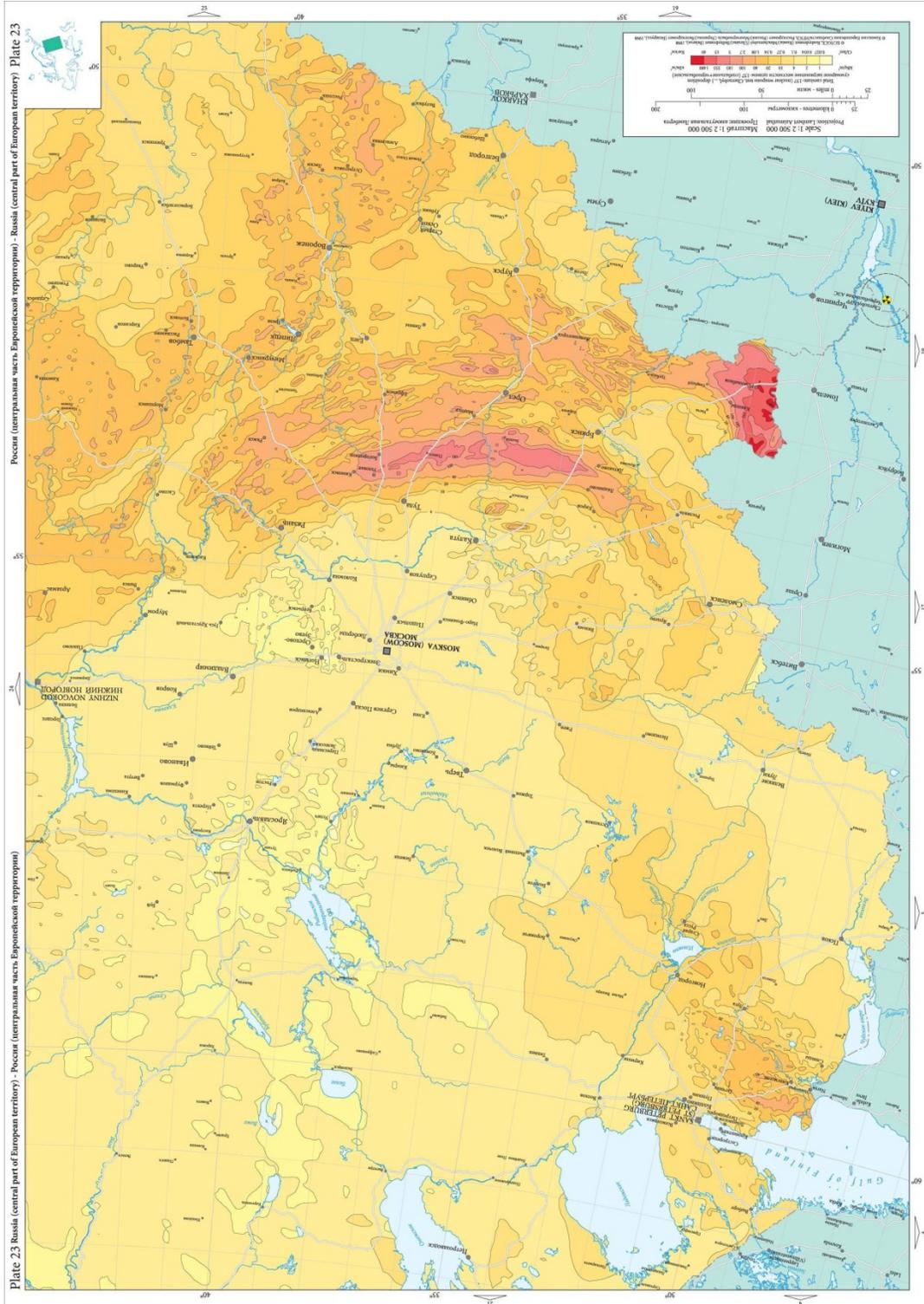


Figure A-6:  $^{137}\text{Cs}$  deposition – Western Central Russia



Figure A-7: <sup>137</sup>Cs deposition – Southwestern Russia

**APPENDIX B. Analysis of Publicly Available Soil Sampling Data**

## APPENDIX B

### ***B.1 Introduction***

Prior to the dose reconstruction team's decision to utilize  $^{137}\text{Cs}$  deposition density plates published in the ATLAS to determine the initial indicator source term, we searched for the original  $^{137}\text{Cs}$  deposition density data which was used to construct the ATLAS map plates. Formal requests from the various institutions involved in this project were filed with various national authorities in Europe without success; personal appeals to individuals who may have had access to the complete body of source data were also unsuccessful. We were able to acquire a database of  $^{137}\text{Cs}$  soil deposition density measurements from the EC website as a downloadable dataset. We were suspicious that this was not a complete data set: Some of the data points listed seemed arbitrary on casual inspection, and there seemed to be altogether too few data points listed for former Soviet states, where the majority of the sampling was conducted.

The author therefore executed a series of statistical analyses designed to conclusively determine if the data set was incomplete or inaccurate.

### ***B.2 Statistical population of interest***

The statistical population of interest in this case is a set of  $^{137}\text{Cs}$  soil deposition density measurements that were made by various national agencies throughout Europe and Western Russia. It is important to emphasize that these measurements were made by individual countries and then reported to the European Commission for aggregation into a single data set. Due to the geopolitics of the region, this meant that different Western European states accounted for much of the data collected, while former states of the USSR were responsible for the remainder. We would expect that the highest deposition densities

of  $^{137}\text{Cs}$  would be closest to the reactor, around the intersection of the border of the Ukraine, Belarus, and Western Russia.

These measurements determined the deposition density of  $^{137}\text{Cs}$  in the top layer of soil immediately following the accident. For the purpose of this discussion, whenever sampling data are needed, they will be drawn (without any selection preference) from these measurements. In principle, this should serve the same function as actually taking the samples – the raw data set should represent samples actually taken after the reactor accident.

The available data set contains thousands of data points, many more than is needed to complete the pilot studies and small-scale follow-up described in this paper.

### ***B.3 Pilot Study (Simple Random Sampling)***

In order to make an initial determination of the scope of contamination from the accident, a two-part pilot study was conducted.

#### **B.3.1 Samples drawn from across Europe**

##### **B.3.1.1 Simple random sampling**

In the first part of the study, five groups of five samples were collected from random locations across Europe (*Table B-1*). The mean, variance and standard deviation of each five-sample group was determined (*Table B-2*). Comparison of the results showed substantial differences in the variance. Characteristics of the set of all samples were also determined (*Table B-3*).

**Table B-1: Simple random sampling values for pilot study**

Sample	Group				
	1	2	3	4	5
1	2.30	11.46	11.17	28.37	0.01
2	2.90	6.78	11.43	6.25	9.28
3	19.00	0.10	15.03	28.40	1.09
4	20.00	0.50	39.50	1.40	2.50
5	39.00	0.04	1.50	37.00	4.13

**Table B-2: Statistics of simple random sampling for pilot study**

	Group				
	1	2	3	4	5
Mean	16.64	3.77	15.73	20.28	3.40
Variance	227.79	26.57	201.83	241.05	13.19
Standard Deviation	15.09	5.15	14.21	15.53	3.63

**Table B-3: Statistics of set of all samples for pilot study**

<b>Mean</b>	<b>11.96</b>
<b>Variance</b>	<b>169.57</b>
<b>Standard Deviation</b>	<b>13.02</b>

From this information, it is unclear whether  $^{137}\text{Cs}$  deposition was distributed across Europe in a systematic fashion. Heuristically, we would expect substantial hot spots and uneven distribution of  $^{137}\text{Cs}$ .

### B.3.1.2 Kruskal-Wallis Test

To test the groups of samples for systematic distribution, the Kruskal-Wallis test was conducted. The hypotheses in this case are:

- Null hypothesis ( $H_0$ ) - The populations from which the groups of data were drawn have the same mean.
- Alternative hypothesis ( $H_A$ ) - At least one population has a mean larger or smaller than one other population. [43]

The test was conducted at the  $\alpha=0.10$  level. Ranked values for each  $x_{ij}$  sample were found, and then the ranks for each group were summed to find the total rank for each group ( $R_j$ ). (See *Table B-4* and *Table B-5*)

**Table B-4: Values for Kruskal-Wallis test for pilot study**  
Group (*j*)

Sample <i>i</i>	1		2		3	
	Value $x_{ij}$	Rank $R_{ij}$	Value $x_{ij}$	Rank $R_{ij}$	Value $x_{ij}$	Rank $R_{ij}$
1	2.30	8	11.46	17	11.17	15
2	2.90	10	6.78	13	11.43	16
3	19.00	19	0.10	3	15.03	18
4	20.00	20	0.50	4	39.50	25
5	39.00	24	0.04	2	1.50	7

Group (*j*) [continued]

Sample <i>i</i>	4		5	
	Value $x_{ij}$	Rank $R_{ij}$	Value $x_{ij}$	Rank $R_{ij}$
1	28.37	21	0.01	1
2	6.25	12	9.28	14
3	28.40	22	1.09	5
4	1.40	6	2.50	9
5	37.00	23	4.13	11

**Table B-5: Ranks for Kruskal-Wallis test for pilot study**

Group <i>j</i>	Number of Samples $n_j$	Total Rank $R_j$	$R_j^2/n_j$
1	5	81	1312.2
2	5	39	304.2
3	5	81	1312.2
4	5	84	1411.2
5	5	40	320

Since this set of samples had no tied values, the Kruskal-Wallis statistic is not corrected for ties, and is found by:

$$K_w = \left[ \frac{12}{m(m+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} \right] - 3(m+1) \quad [43]$$

The values for this equation are:

- $m=25$

- $\Sigma(R_j^2/n_j)= 4659.8$
- $K_w=8.03$

The  $1-\alpha$  quantile of the chi-square distribution was found from Table A19 in [43]. A total of five groups were compared, yielding 4 degrees of freedom. For the  $\alpha=0.10$  level, this produced  $\chi_{1-\alpha,k-1}^2 = 7.78$ . Because  $K_w > \chi_{1-\alpha,k-1}^2$ , we can reject the  $H_0$  and accept  $H_A$ . We therefore conclude at the  $\alpha=0.10$  level that at least one population from which we drew our data has a different mean than at least one other population. The reason for this cannot be immediately determined, but we might further hypothesize that this might indicate the presence of substantial  $^{137}\text{Cs}$  hotspots or systematic errors in some of the collected data.

### **B.3.2 Samples drawn from within 1200 km distance from Chernobyl**

It has been suggested in the relevant literature that the majority of  $^{137}\text{Cs}$  was deposited within 1000 km of the reactor site after the accident. The region of interest for this paper was therefore reduced a 1200 km radius around the reactor site. This would include the suggested  $^{137}\text{Cs}$  deposition zone of 1000 km radius, along with an additional 200 km to account for variance in the published model, which suggested 1000 km. The selection of 1200 km (as opposed to 1100 km, 1300 km, or another value) was arbitrary. The same analysis that was applied to all of Europe was then done on this restricted region.

The source data set lists  $^{137}\text{Cs}$  deposition density samples and associated longitude/latitude coordinates. To convert this to distance from the Chernobyl reactor, the haversine formula was utilized.

#### **B.3.2.1 Simple random sampling**

Five groups of five samples were collected from random locations no more than 1200 km away from the reactor site (*Table B-6*). The mean, variance and standard deviation of

each five-sample group was determined (*Table B-7*). Comparison of the results showed substantial differences in the variance. Characteristics of the set of all samples were also determined (*Table B-8*)

**Table B-6: Simple random sampling values for 1200 km study**

Sample	Group				
	1	2	3	4	5
1	3.14	1.70	1.10	4.00	4.54
2	0.70	7.19	31.00	20.00	8.40
3	0.60	160.00	19.00	160.00	1.10
4	2.40	74.00	37.00	70.00	19.00
5	3.74	55.00	55.00	3.01	2.49

**Table B-7: Statistics of simple random sampling for 1200 km study**

	Group				
	1	2	3	4	5
Mean	2.12	59.58	28.62	51.40	7.11
Variance	2.02	4101.90	405.42	4428.53	51.78
Standard Deviation	1.42	64.05	20.14	66.55	7.20

**Table B-8: Statistics for set of values for 1200 km study**

<b>Mean</b>	<b>29.76</b>
<b>Variance</b>	2047.50
<b>Standard Deviation</b>	45.25

From this information, it is unclear whether  $^{137}\text{Cs}$  deposition was distributed within the 1200 km zone in a systematic fashion. Again, we would expect substantial hot spots and uneven distribution of  $^{137}\text{Cs}$ . It is interesting to compare results for the 1200 km zone to those for all of Europe: The mean for the 1200 km zone is higher than that for all of Europe, which intuitively is what we would expect to see. However, the standard deviation is so high that we cannot conclusively state this.

### B.3.2.2 Kruskal-Wallis Test

Because of the apparent disparity between the means and standard deviations of the groups of samples within the 1200 km zone, the Kruskal-Wallis test was also applied to this data. The hypotheses in this case are the same as before:

- Null hypothesis ( $H_0$ ) - The populations from which the groups of data were drawn have the same mean.
- Alternative hypothesis ( $H_A$ ) - At least one population has a mean larger or smaller than one other population. [43]

The test was conducted at the  $\alpha=0.10$  level. Ranked values for each  $x_{ij}$  sample were found, and then the ranks for each group were summed to find the total rank for each group ( $R_j$ ). (See *Table B-9* and *Table B-10*.)

**Table B-9: Values for Kruskal-Wallis test for 1200 km study**

		Group ( $j$ )					
		1		2		3	
Sample		Rank		Rank		Rank	
$i$	$x_{ij}$	$R_{ij}$	$x_{ij}$	$R_{ij}$	$x_{ij}$	$R_{ij}$	
1	3.14	9	1.70	5	1.10	3	
2	0.70	2	7.19	13	31.00	18	
3	0.60	1	160.00	24	19.00	15	
4	2.40	6	74.00	23	37.00	19	
5	3.74	10	55.00	20	55.00	21	

Group ( $j$ ) [continued]

		Group ( $j$ ) [continued]			
		4		5	
Sample		Rank		Rank	
$i$	$x_{ij}$	$R_{ij}$	$x_{ij}$	$R_{ij}$	
1	4.00	11	4.54	12	
2	20.00	17	8.40	14	
3	160.00	25	1.10	4	
4	70.00	22	19.00	16	
5	3.01	8	2.49	7	

**Table B-10: Ranks for Kruskal-Wallis test for 1200 km study**

Group	Number of Samples	Total Rank	$R_j^2/n_j$
$j$	$n_j$	$R_j$	
1	5	28	156.8
2	5	85	1445
3	5	76	1155.2
4	5	83	1377.8
5	5	53	561.8

By inspection, we see that this set of samples has several tied values: (11,23), (13,24), (10,15), and (8,18).

We can therefore calculate the  $K_w$  statistic as before, but we will need to apply a correction factor to account for ties. Calculation of the uncorrected  $K_w$  factor yields:

- $m=25$
- $\sum (R_j^2/n_j)=4696.6$
- $K_w=8.71$

The Kruskal-Wallis statistic must then be corrected for ties. The corrected Kruskal-Wallis statistic is given by:

$$K'_w = \frac{K_w}{1 - \frac{1}{m(m^2-1)} \sum_{j=1}^g t_j(t_j^2 - 1)} \quad [43]$$

Note that [43] states that the sum should be only over the set of all groups, which contain a tie. The inclusion of untied groups, however, will not change the result of the sum – untied groups will add nothing to the sum (as the  $t_j$  value for untied groups is 0).

The correction factor was determined for the four pairs of ties (see *Table B-11*).

**Table B-11: Kruskal-Wallis correction for ties for 1200 km study**

Group j	Number of Ties t <sub>j</sub>	t <sub>j</sub> (t <sub>j</sub> <sup>2</sup> - 1)
1	0	0
2	2	6
3	3	24
4	1	0
5	2	6

The correction factor then is:

$$\text{Correction Factor} = 1 - \frac{1}{m(m^2 - 1)} \sum_{j=1}^g t_j(t_j^2 - 1)$$

Application of the correction factor yields the following:

- Correction factor = 0.94
- $K'_w = 9.26$

The  $1-\alpha$  quantile of the chi-square distribution is the same as before:  $\chi_{1-\alpha, k-1}^2 = 7.78$  for the  $\alpha=0.10$ . Because  $K_w > \chi_{1-\alpha, k-1}^2$ , we can reject the  $H_0$  and accept  $H_A$ . We therefore conclude at the  $\alpha=0.10$  level that at least one population from which we drew our data has a different mean than at least one other population.

As before, the reason for this cannot be immediately determined, but might serve to reinforce our hypothesis that there are substantial  $^{137}\text{Cs}$  hotspots within the 1200 km zone, or that systematic errors indeed exist in some of the collected data. It seems likely that a more refined sampling methodology is called for in this situation.

#### ***B.4 Follow-up Study (Stratified Sampling)***

##### **B.4.1 Stratified Sampling Design – strata within 1200 km distance from Chernobyl**

In an attempt to produce a more refined analysis of  $^{137}\text{Cs}$  deposition within the 1200 km-radius zone around the reactor site, a stratified sampling scheme was developed. The zone was divided into four concentric rings, with outer radii of 150 km, 300 km, 600 km, and 1200 km. These annuli were further segmented into northwest, northeast, southeast, and southwest quadrants (see *Figure B-1: Geographical stratification of Europe*). This produced a total of 16 strata (detailed in *Table B-12*).



*Location of the 1200 km zone and strata*



*Expanded view of the strata*

**Figure B-1: Geographical stratification of Europe**

**Table B-12: Description of strata**

Distance	Northwest Quadrant	Northeast Quadrant	Southeast Quadrant	Southwest Quadrant
<150km	Stratum 1	Stratum 2	Stratum 3	Stratum 4
150km-300km	Stratum 5	Stratum 6	Stratum 7	Stratum 8
300km-600km	Stratum 9	Stratum 10	Stratum 11	Stratum 12
600km-1200km	Stratum 13	Stratum 14	Stratum 15	Stratum 16

The Black Sea and the Sea of Azov make up the majority of Stratum 15. Because of this (and because of a paucity of sampling availability in that stratum), it was censored. It appears in the tables below only for the sake of completeness, but no data from this area were utilized in any calculations.

#### **B.4.1.1 Results of Statistical Analysis**

The proportional weight for each stratum is determined by the ratio of the area of each stratum to the total area of all strata. With the exception of the censored Stratum 15, the area of any bodies of water in the zone is disregarded.

The total area of all strata is given by:

$$A_T = \pi R^2 - A_{15} = 4523893 \text{ km}^2 - 848230 \text{ km}^2 = 3675663 \text{ km}^2$$

Where:

- $R = 1200 \text{ km}$  is the radius of the entire zone
- $A_{15} = \pi[(1200 \text{ km})^2 - (600 \text{ km})^2]$  is the area of the censored Stratum 15

150 samples were drawn randomly from the 15 uncensored strata. The number of samples from each stratum, strata areas, and strata weights are given below in *Table B-13*).

**Table B-13: Strata weights**

Stratum h	Number of samples $n_h$	Stratum Area $A_h$	Stratum Weight $W_h$
1	5	17671	0.0048
2	3	17671	0.0048
3	2	17671	0.0048
4	4	17671	0.0048
5	12	53014	0.0144
6	8	53014	0.0144
7	4	53014	0.0144
8	9	53014	0.0144
9	18	212058	0.0577
10	30	212058	0.0577
11	4	212058	0.0577
12	7	212058	0.0577
13	14	848230	0.2308
14	3	848230	0.2308
15	-----	CENSORED -----	
16	27	848230	0.2308
TOTAL	150	3675663	1.0000

Characterization of the data requires determining the following for the samples taken from each strata:

- Mean:

$$\bar{x}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi} \quad [43]$$

- Standard deviation (as usual)

- Estimation of variance:

$$s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)^2 \quad [43]$$

- Standard error:

$$SE(\bar{x}_h) = \sqrt{\frac{s_h^2(1 - f_h)}{n_h}} = \sqrt{\frac{s_h^2}{n_h}} \quad [44]$$

The results for each stratum are given below in *Table B-14*.

**Table B-14: Statistics of strata**

Stratum h	Mean $\bar{x}_h$	Estimation of Variance $s_h^2$	Standard Deviation $\sigma_h$	Standard Error $SE_h$
1	108.00	1370.00	37.01	16.55
2	123.33	1633.33	40.41	23.33
3	20.00	2.00	1.41	1.00
4	65.50	1587.00	39.84	19.92
5	84.08	2962.63	54.43	15.71
6	180.88	5458.98	73.88	26.12
7	22.75	30.25	5.50	2.75
8	39.00	591.00	24.31	8.10
9	27.44	317.67	17.82	4.20
10	68.43	4845.43	69.61	12.71
11	19.25	0.25	0.50	0.25
12	27.03	211.06	14.53	5.49
13	7.07	65.02	8.06	2.16
14	31.00	432.00	20.78	12.00
15	----- CENSORED -----			
16	2.84	9.05	3.01	0.58

**Important:** Please note that the two sample values for Stratum 3 were identical (both were 20 kBq/m<sup>2</sup>), yielding an estimation of variance, standard deviation, and standard error of 0. In order to better illustrate the stratified sampling technique, the values for the two samples in this stratum were manually changed by the author. One sample was set to 19 kBq/m<sup>2</sup>, the other was set to 21 kBq/m<sup>2</sup>. This induced no change in the mean, but did produce non-zero values for  $s_h^2$ ,  $\sigma_h$ , and  $SE_h$ . Approximation of the mean and standard error for the population is found as follows.

- Mean:

$$\bar{x}_{st} = \sum_{h=1}^L W_h \bar{x}_h \quad [43]$$

- Standard Error:

$$SE(\bar{x}_{st}) = \sqrt{\sum_{h=1}^L W_h^2 \frac{s_h^2(1 - f_h)}{n_h}} = \sqrt{\sum_{h=1}^L W_h^2 \frac{s_h^2}{n_h}} \quad [44]$$

The values for the functions to be summed for each stratum are given in *Table B-15* (next page). Applying the equations for determining the estimation of the mean and the standard error yields an estimation for the mean of  $x_{st} = 23.88$  and a estimate for the standard error of  $SE(x_{st})=2.98$ .

**Table B-15: Intermediate values for stratification**

Stratum h	$W_h \cdot x_h$	$W_h^2 \cdot s_h^2/n_h$
1	0.519	0.00633
2	0.593	0.01258
3	0.096	0.00002
4	0.315	0.00917
5	1.213	0.05136
6	2.609	0.14195
7	0.328	0.00157
8	0.563	0.01366
9	1.583	0.05874
10	3.948	0.53758
11	1.111	0.00021
12	1.560	0.10035
13	1.631	0.24734
14	7.154	7.66864
15	----- CENSORED -----	
16	0.656	0.01785
TOTAL	23.878	8.86736

#### B.4.2 Optimum Sampling Design

It is likely that the accuracy of our estimation for the mean and error could theoretically be dramatically improved through further sampling. This section will determine optimum sampling design for hypothetical additional sampling under two limiting conditions: Cost and margin of error.

### B.4.2.1 Cost as a limiting factor

While the exact cost for taking each sample in the European Commission’s data set is difficult to estimate, we can propose a general model to represent the difficulty of acquiring any given sample. We presuppose that there was sufficient financial and political capital to take 3000 total samples in the developed Western European nations (i.e.,  $C=3000$ , with no overhead, so  $c_0=0$ ). We further suppose that it was twice as difficult to take comparable samples in the underdeveloped former Soviet states. We therefore normalize  $c_h$  to 1 for the strata which are comprised of Western European states, and set  $c_h=2$  in the strata which are comprised of former Soviet states. We find the optimum total number of samples via:

$$n = \frac{C - c_0}{\sum_{h=1}^L w_h c_h} \quad [44]$$

where

$$w_h = \frac{W_h \sigma_h / \sqrt{c_h}}{\sum_{h=1}^L W_h \sigma_h / \sqrt{c_h}} \quad [44]$$

The values necessary for this calculation are given below in *Table B-16*.

**Table B-16: Values for computing optimum sampling for strata**

Stratum h	Cost $c_h$	$W_h \cdot \sigma_h / \sqrt{c_h}$	$W_h$	$W_h \cdot c_h$
1	2	0.1258	0.0102	0.0203
2	2	0.1374	0.0111	0.0222
3	2	0.0048	0.0004	0.0008
4	2	0.1354	0.0109	0.0219
5	2	0.5551	0.0448	0.0896
6	2	0.7535	0.0608	0.1217
7	2	0.0561	0.0045	0.0091
8	2	0.2479	0.0200	0.0400
9	2	0.7271	0.0587	0.1174
10	2	2.8397	0.2292	0.4585
11	2	0.0204	0.0016	0.0033
12	1	0.8381	0.0677	0.0677
13	1	1.8608	0.1502	0.1502
14	2	3.3916	0.2738	0.5476
15	----- CENSORED -----			
16	1	0.6941	0.0560	0.0560
TOTAL		12.3880	1.0000	1.7261

The total number of samples is then found through the equation above as  $n=1738$ . We can then determine the optimum number of samples for each strata by calculating  $n_h=nW_h$  (*Table B-17*).

**Table B-17: Optimum number of samples**

Stratum h	Number $n_h$	Stratum h	Number $n_h$
1	17.65	9	102.01
2	19.28	10	398.40
3	0.67	11	2.86
4	19.00	12	117.59
5	77.88	13	261.07
6	105.72	14	475.84
7	7.87	15	n/a
8	34.78	16	97.39
		TOTAL	1738.03

#### B.4.2.2 Pre-specified margin of error as a limiting factor

Suppose that cost is not a factor; instead, we require an extremely high level of accuracy: We demand that our estimate is within  $d=0.5$  kBq/m<sup>2</sup> at the  $\alpha=0.01$  level. We can determine the total number of samples required by:

$$n = \frac{Z_{1-\alpha/2}^2 \cdot \sum W_h s_h^2 / d^2}{1 + Z_{1-\alpha/2}^2 \cdot \sum W_h s_h^2 / d^2 N} \quad [43]$$

We find  $Z_{.995}=2.58$  from Table A1 in [43], can then use the equation above to determine the requisite value for  $n$  as  $n=5979$ . Using the values for  $w_h$  from the previous section, we can also determine the number of samples needed from each stratum. *Table B-18* (next page) contains these values, as well as the total expense for each stratum and the values of  $W_h s_h^2$  used to determine  $n$  above.

**Table B-18: Expense for each stratum**

Stratum h	$W_h \cdot s^2_h$	Number $n_h = nW_h$	Expense for this stratum $n_h \cdot C_h$
1	6.59	60.73	121.46
2	7.85	66.31	132.62
3	0.01	2.32	4.64
4	7.63	65.36	130.73
5	42.73	267.92	535.84
6	78.74	363.68	727.36
7	0.44	27.07	54.14
8	8.52	119.66	239.32
9	18.33	350.92	701.85
10	279.54	1370.53	2741.06
11	0.01	9.84	19.69
12	12.18	404.52	404.52
13	15.01	898.11	898.11
14	99.69	1636.91	3273.82
15	-----	CENSORED	-----
16	2.09	335.02	335.02
TOTAL	579.35	5978.91	10320.17

Summing the costs for all strata gives us the total cost incurred by requiring such a high degree of accuracy as  $C=10320$ .

***B.5 Characterizing the Data Set***

It is useful to determine whether our data set is easily characterized (i.e., whether or not it might be normal/lognormal). We might guess that it should be lognormal, but can easily check this using probability plotting.

**B.5.1 Complete Data Set -- Normality/Lognormality**

**B.5.1.1 Normal Probability Plot**

We start with the 150-sample data set used for the stratified sampling scheme in the previous section. This data were ordered from smallest to largest and then ranked. Each ranked datum was then associated with its Z-value; a plot of the data values against their associated Z-value was then made (see *Figure B-2: Normal probability plot for full data set*).

A linear curve was fitted to the data to check for normality. The poor fit here implies that this data is not normally distributed.

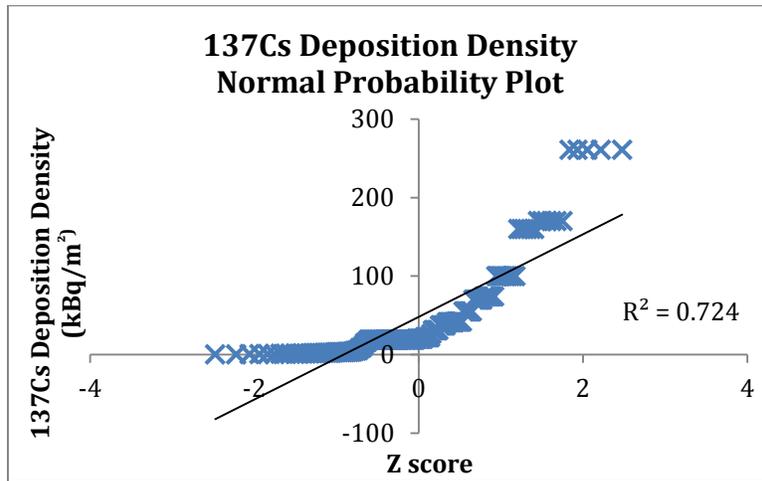


Figure B-2: Normal probability plot for full data set

### B.5.1.2 Lognormal Probability Plot

To check for lognormality, the data set was transformed by applying the natural logarithm to each data point. Data points were then ranked, and ranks associated with their Z scores. The data was then plotted against associated Z scores. A linear curve was fitted to the data to check for lognormality (see *Figure B-3: Lognormal probability plot for full data set*) (next page).

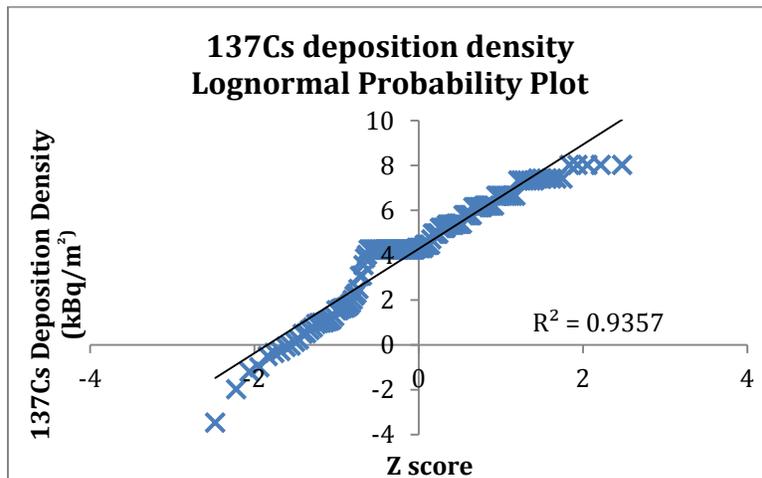


Figure B-3: Lognormal probability plot for full data set

The  $R^2$  value of 0.9357 hints that the data might be lognormal, but it is insufficient to provide a strong level of confidence. Of particular note is the pronounced “shelf” visible in the plotted data between  $Z=-1$  and  $Z=0$ . Further assurance is needed to test for lognormality.

### B.5.1.3 D’Agnostino’s Test

To rigorously test for lognormality, D’Agnostino’s test was executed against the log-transformed data at the  $\alpha=0.05$  level. Here the hypotheses tested are:

- Null hypothesis ( $H_0$ ) - The log-transformed data distribution is normal.
- Alternative hypothesis ( $H_A$ ) - The log-transformed data distribution is not normal.

[43]

Quantiles of D’Agnostino’s test for  $\alpha=0.05$ , found from Table A8 in [43] are  $Y_{0.025} = 2.452$  and  $Y_{0.975} = 1.423$ . D’Agnostino’s test requires computing the following values [43]:

$$D = \frac{\sum_{i=1}^n [i = \frac{1}{2}(n + 1)] x_i}{n^2 s}$$

$$s = \sqrt{\left[ \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right]}$$

$$Y = \frac{D - 0.28209479}{0.02998598/\sqrt{n}}$$

Values for D’Agnostino’s test for log-transformed data are shown in *Table B-19*.

**Table B-19: Values for D’Agnostino’s test for log-transformed full data set**

<b>n</b>	<b>150</b>
<b>Mean</b>	1.288
<b>s</b>	0.703
$\Sigma[(x_i - \bar{x})^2]$	74.213
$\Sigma[(i - \frac{1}{2}(n+1))x_i]$	4327.446
<b>D</b>	0.273
<b>Y</b>	-3.537

Because  $Y$  is less than  $Y_{0.025}$ , we reject lognormality at the 95% confidence level.

### **B.5.2 Censored Data Set -- Normality/Lognormality**

Serious consideration of the “shelf” seen in the previous plot produced the following question: Perhaps the sample data is systematically flawed somehow. Rigorous investigation of the entire data set available from the European Commission produced the following revelation: All of the sample data provided from formerly Soviet nations were suspiciously consistent. Every data point from the USSR was rounded to an integer value, while the preponderance of data points from Western European nations was present in a decimal form. Furthermore, there were only 14 unique <sup>137</sup>Cs deposition values amongst the 109 data points located in the USSR. Most compellingly, where the sample data from Western Europe were provided from apparently random latitude/longitude coordinates, the sample data provided from the USSR were aligned on an extremely regular grid: One sample every 0.5-degree of longitude and 1.0 degree of latitude.

To test for inaccurate reporting of the Soviet data, these data points were removed from the data set. The total number of samples within the area covered by the stratified sampling scheme was reduced from 150 to 42 by censoring the Soviet data. D’Agostino’s test was then repeated.

#### **B.5.2.1 Normal Probability Plot**

Repeating the same method as before, a normal probability plot was produced for the censored data. In this case, the R<sup>2</sup> value of 0.5894 was much lower than when including the Soviet data. The censored data are clearly not normally distributed. (see *Figure B-4: Normal probability plot for censored data set*, next page).

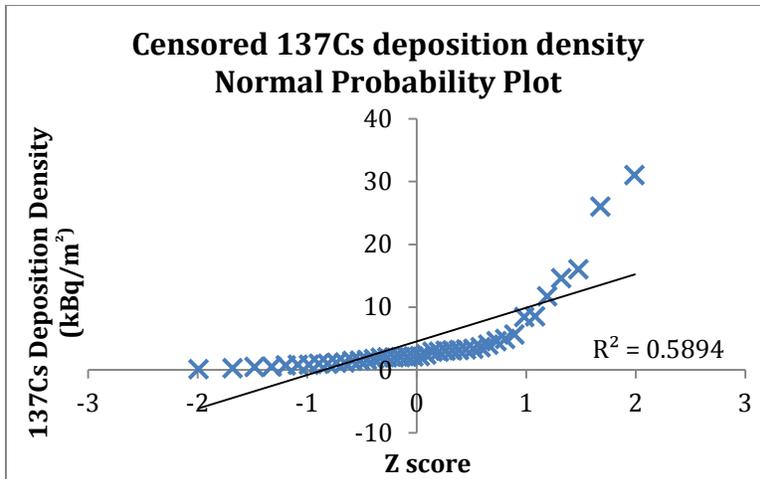


Figure B-4: Normal probability plot for censored data set

### B.5.2.2 Lognormal Probability Plot

The censored data set was then transformed by taking the natural logarithm of each data point, matched with Z scores, and plotted as before (see *Figure B-5: Lognormal probability plot for censored data set*).

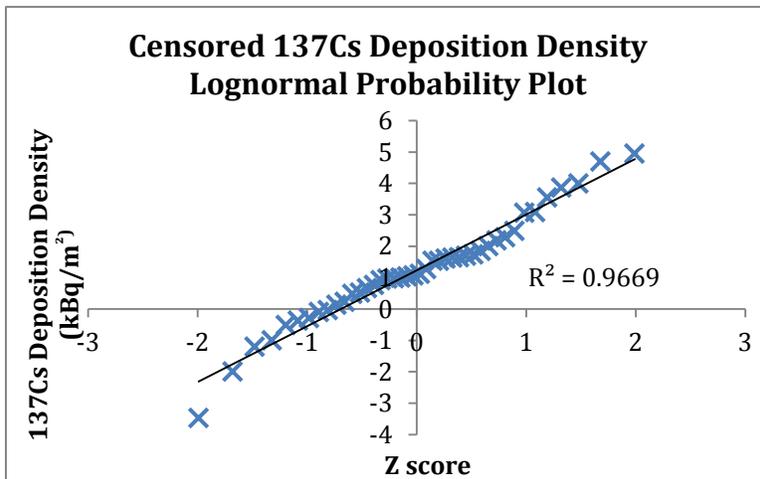


Figure B-5: Lognormal probability plot for censored data set

### B.5.2.3 D’Agostino’s Test

Utilizing the same method as before, D’Agostino’s test was conducted at the  $\alpha=0.05$  level on the censored, transformed data. The results are shown in *Table B-20*. In this case, Y

is greater than  $Y_{0.025}$  and is less than  $Y_{0.975}$ . We cannot reject lognormality at the 95% confidence level, and tentatively accept it.

**Table B-20: Values for D’Agnostino’s test for log-transformed censored data set**

<b>n</b>	<b>42</b>
<b>Mean</b>	1.229
<b>s</b>	1.671
$\Sigma[(x_i - \bar{x})^2]$	117.297
$\Sigma[(i - \frac{1}{2}(n+1))x_i]$	807.401
<b>D</b>	0.274
<b>Y</b>	-1.774

## ***B.6 Conclusion***

Simple random sampling for an issue as broad and complex as the deposition of radionuclides from the Chernobyl disaster was shown to lack sufficient sophistication for properly monitoring the event. Furthermore, while stratified sampling yielded a better qualitative picture of the environmental impact of the accident, it fails to produce effective quantitative information. The best approach to this scenario would likely be a rigorous two- or three-stage sampling scheme combined with a kriging tool.

Despite the desire for more quantitative rigor, it is interesting to note the following: Estimation of a total inventory of  $^{137}\text{Cs}$  deposition from the mean deposition value found through the stratified sampling technique yields the following:

$$I = \bar{x}_{st} \cdot A = 23.88\text{kBq/m}^2 \cdot 3675663\text{km}^2 = 8.77 \cdot 10^{13}\text{kBq} = 87.7\text{PBq}$$

This compares extremely favorably to the Nuclear Energy Agency’s seminal work *Chernobyl: Assessment of Radiological and Health Impacts* [4], which suggests a total release of ~85PBq.

While this analysis was produced as a retrospective study, utilizing extant data, it has produced an interesting result. Discrepancies in the source data led to analysis of the entire data set, which may have produced a revelation: At the  $\alpha=0.05$  level, the censored data set appears to follow a lognormal distribution, while the uncensored data set does not. That,

combined with the fact that the data from former Soviet countries bears several strong heuristic indications of inaccuracy, leads the author to conclude that the data provided to the European Commission for these locations are inaccurate and unusable.

APPENDIX C. Equirectangular transformation of  $^{137}\text{Cs}$  deposition density maps

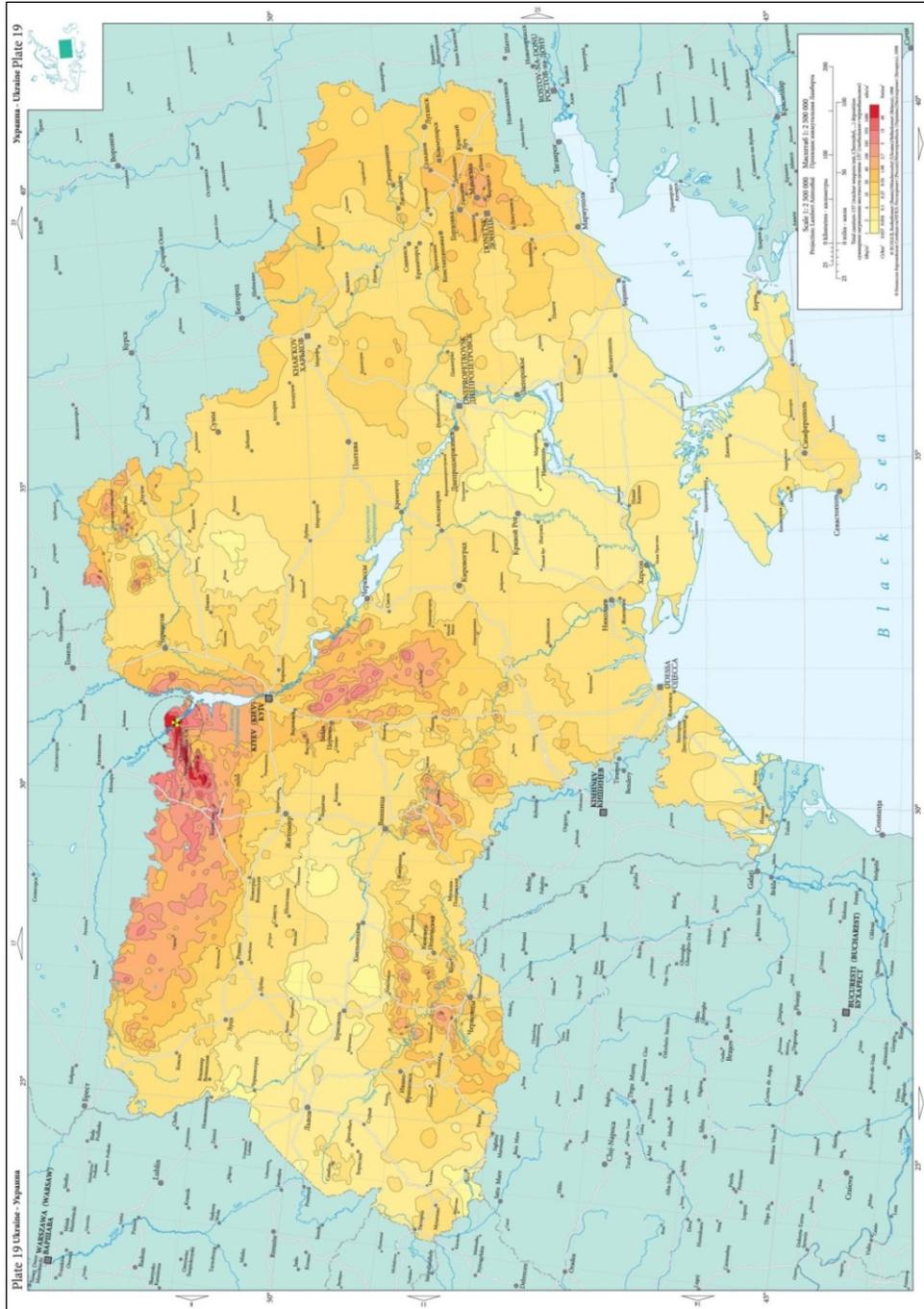


Figure C-1: Oblique azimuthal projection map of the Ukraine

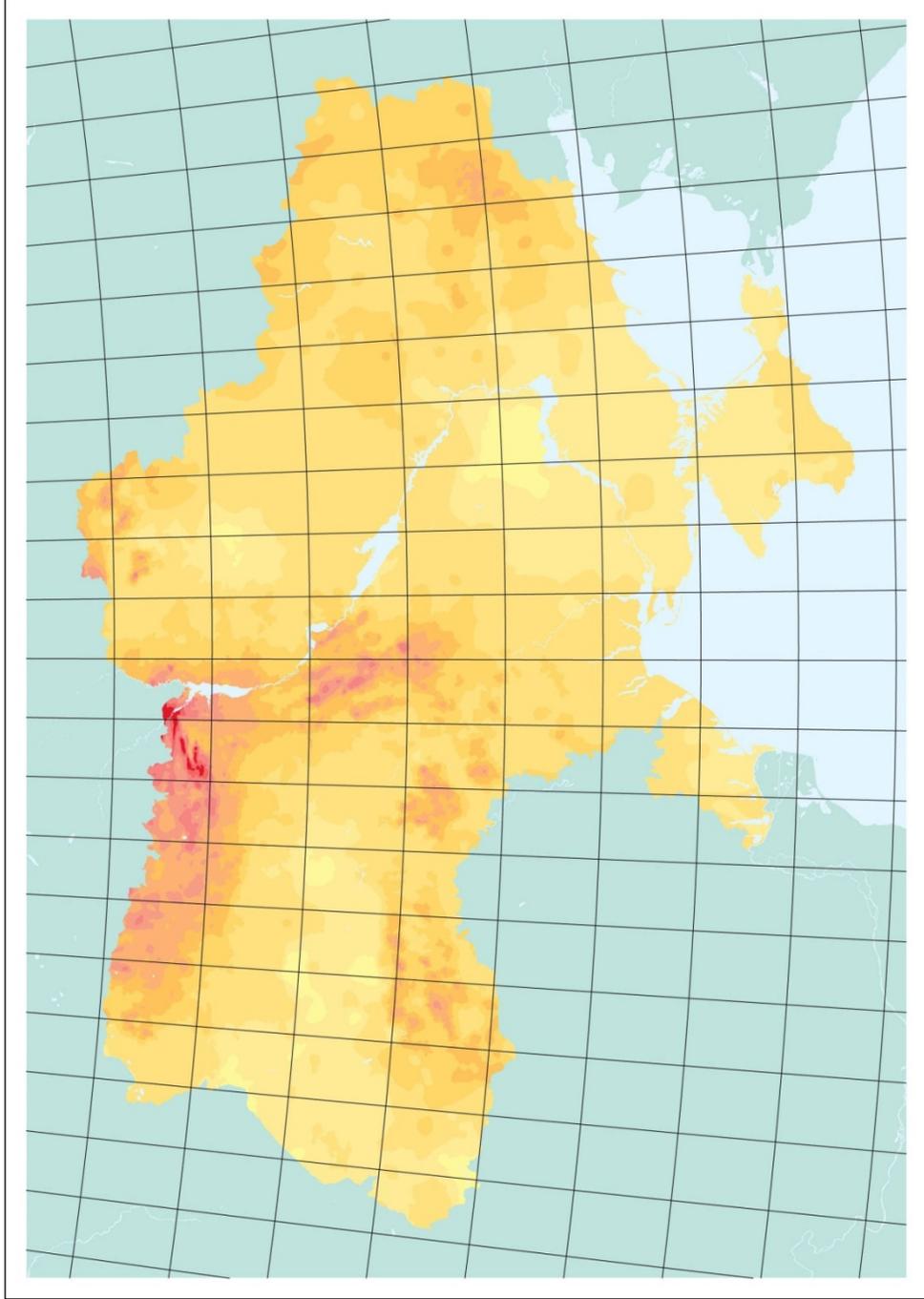
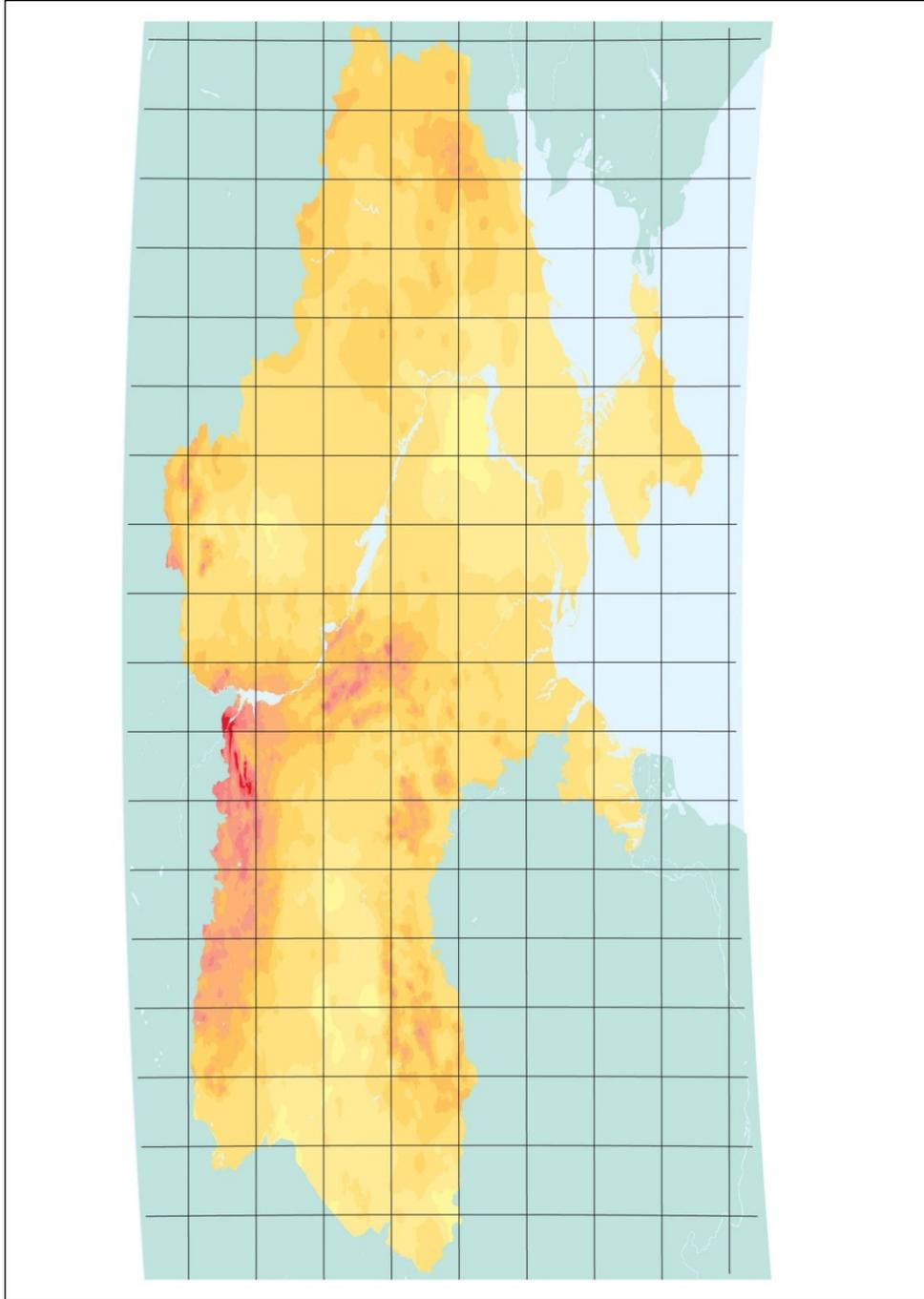
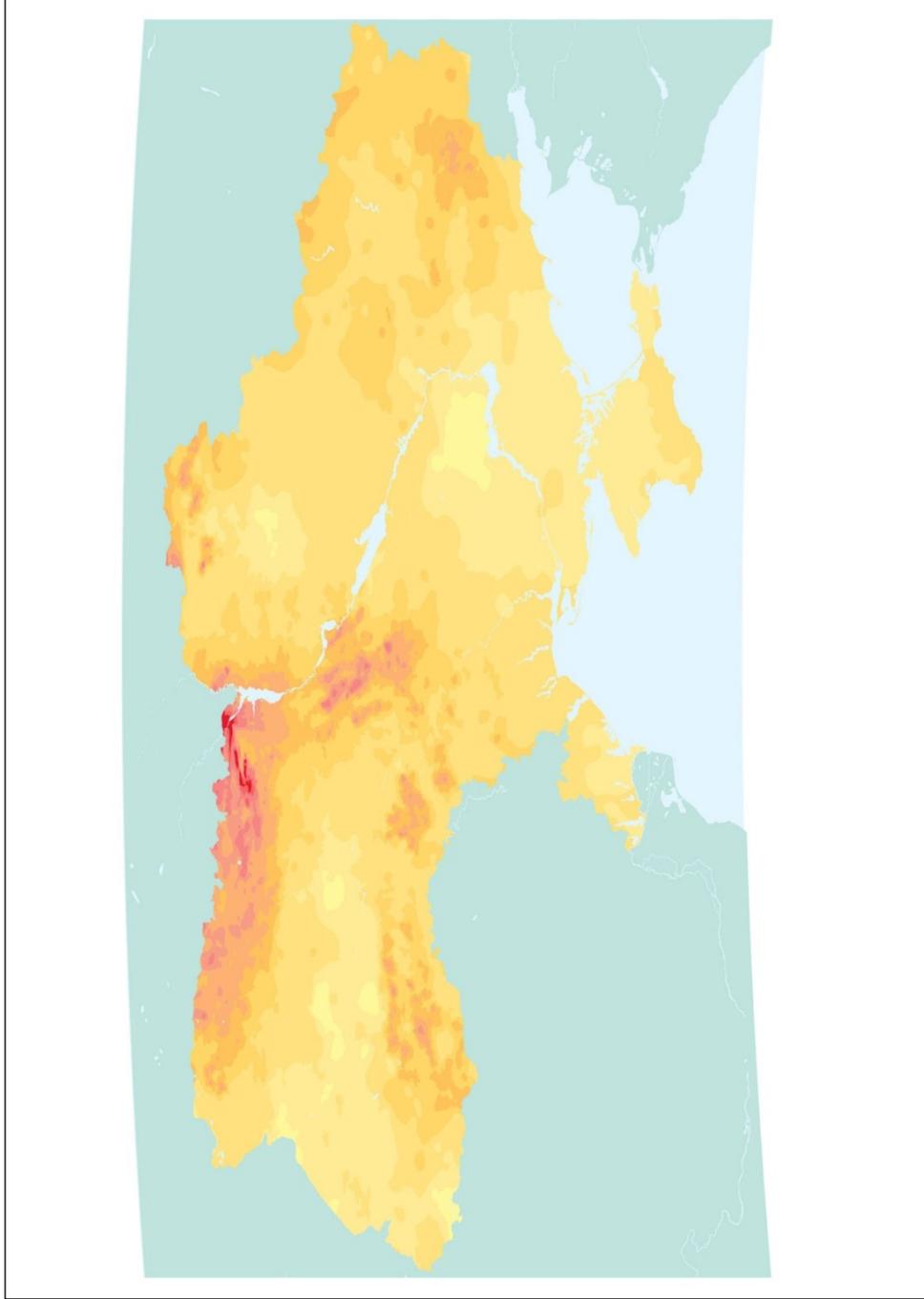


Figure C-2: Oblique azimuthal projection map of the Ukraine, data stripped



**Figure C-3: Equiarectangular projection map of the Ukraine**



**Figure C-4: Equiarectangular projection map of the Ukraine, deposition only**

APPENDIX D. Protocol for map transformation process and geodatabase

For this project, the following specific techniques were used to implement the process described above for loading PDF map plates from the ATLAS into a spatial database:

1. Published map plates from the ATLAS were available in a write-protected Adobe PDF format. Write protection was removed.
2. The vector graphics in each plate's Adobe PDF file were organized into layers in Adobe Acrobat, including one layer for the conjugate graticule, one layer for the color-filled  $^{137}\text{Cs}$  isolines, and one layer for the  $^{137}\text{Cs}$  deposition density color legend.
3. The layers containing the conjugate graticule and the color-filled  $^{137}\text{Cs}$  isolines were exported as separate 150DPI, 24-bit CMYK bitmap files.
4. These bitmap files were loaded into ESRI's ArcGIS geodatabase as *raster* layers. An ArcMap session was created for each map plate. The conjugate graticule and color-filled  $^{137}\text{Cs}$  isoline raster of each map were loaded into its ArcMap session. A second-order polynomial transformation was defined using the conjugate graticule raster by picking 50% of the intersections on the graticule and assigning a latitude and longitude coordinate based on the labels displayed in the original published map plate. The transformation was saved and then applied to the respective  $^{137}\text{Cs}$  isoline raster. In all cases, the  $R^2$  value for each transformation was greater than 0.95.
5. An ArcGlobe session was created to contain the entire set of transformed map plates. Each transformed  $^{137}\text{Cs}$  isoline raster was loaded into this session.
6. The  $^{137}\text{Cs}$  deposition density color legend was exported from each plate's Adobe PDF file. The listed  $^{137}\text{Cs}$  deposition density bounds were noted, and Adobe Photoshop was used to find the hexadecimal color for each listed deposition density bound. These values were stored in a different lookup table for each plate.
7. A table containing the latitude/longitude coordinates for all unique locations listed in the survey dataset was loaded into the ArcGlobe session. The ESRI *Extract Multi*

*Values To Points* tool was used to extract the color shown on every plate for each location.

8. The  $^{137}\text{Cs}$  color-to- deposition-density lookup table was used to determine the  $^{137}\text{Cs}$  deposition density range for each location based on the extracted color.
9. The previously discussed hierarchy for plate preference was used to select values from hotspot data first, then from regional data if no hotspot data were available, and then from continental data if no regional data were available.
10. For the final data set, only the following map plates were loaded to the geodatabase (listed in order of precedence):
  - i. Plate 60 – Chernobyl Hotspot
  - ii. Plate 19 – Ukraine
  - iii. Plate 17 – Belarus
  - iv. Plate 23 – Western Russia
  - v. Plate 1 – Europe
11. If no data were available from any of the plates from the ATLAS, an extrapolated value was used:
  - a. An extrapolated deposition density range of 1-2 kBq/m<sup>2</sup> was used for data points south or west of Europe (equivalent to the deposition density shown in the ATLAS in southernmost or westernmost Europe).
  - b. An extrapolated deposition density range of 2-4 kBq/m<sup>2</sup> was used for data points north or east of Europe (equivalent to the deposition density shown in the ATLAS in northernmost or easternmost Europe).
12. For computation, the mean of the minimum and maximum  $^{137}\text{Cs}$  deposition density estimate shown in the ATLAS is used as the initial source term.

**APPENDIX E. Monte Carlo and sensitivity analysis system design**

## APPENDIX E

### ***E.1 Introduction***

In order to determine a general sense of the distribution of computed dose and dose rates using our methodology, we executed Monte Carlo simulations against our system using distributions for every input parameter (i.e., each experimentally derived value described in *Section 3.2.3 - Computing effective dose rate at an arbitrary time t*). We also computed rank correlation coefficients for each input distribution to determine which (if any) input parameters were correlated to the distribution of our estimates for dose and dose rate.

This appendix details the design and quality assurance methods used to produce and validate our Monte Carlo simulation and rank correlation computation code.

### ***E.2 Monte Carlo system design***

We developed software predicated on the Mathematica analysis engine to execute Monte Carlo experiments against arbitrary real-valued functions. The software accepts as inputs a symbolic function, distributions for each variable, and a variety of parameters defining computation rules. The software then randomly samples the distribution for each input parameter and computes the value of the function based on these samplings. This process is iterated as many times as desired, and the software then returns the resulting distribution of computed values of the function as an output. Basic statistics (min/max/mean/median/standard deviation/quantiles) are computed for the output.

The software is also capable of executing computations with an independent variable. In this mode, one of the input variables for the function of interest is identified as an independent variable and assigned a lower and upper bound of possible values as well as a step size. The distribution of possible values for the function is then computed as

described above for every possible value of the independent variable, starting at the lower bound and incrementing by the step size until the upper bound is reached. The output from the software is then the distribution and basic statistics for the distribution of the function for each of these possible values of the independent variable. For our purposes, this independent variable will be assumed to be representative of time, and shall be referred to as time henceforth; the possible values of the independent variable will be referred to as time points. The software also computes the running total of values of the function (thus replicating the summation process which is used in our dose assessment methodology).

When computing with an independent variable, the software can compute either with or without resampling from input distributions. When utilizing the resampling capability, the input distributions are resampled at each possible time point prior to computing the distribution of values of the function. When executed without the resampling capability, the input distributions are sampled only once, prior to any computation of values of the function.

### ***E.3 Sensitivity analysis methodology***

For the purpose of sensitivity analysis, the software also finds the Spearman rank correlation between the values sampled from each input distribution and the distribution of values determined for the function (and for the running total of the function). To effect this, the software determines the Spearman rank correlation coefficient  $\rho$  for each time point, comparing each input distribution with the output distribution at this time point. We elected to utilize the Spearman rank correlation method rather than the Kendall rank correlation method because it is more suitable to large distributions and significantly less computationally expensive.

The Spearman rank correlation coefficient provides a measure of monotonic association between two distributions. A positive value indicates that increases in one

distribution correlate to increases in the other distribution, while a negative value indicates that increases in one distribution correlate to decreases in the other distribution. Values for the coefficient itself range between -1 and +1, where a value of -1 indicates perfect monotonic association between positive values of one distribution and negative values in the other distribution, and a value of +1 indicates perfect monotonic association between positive values in both distributions. A value of 0 indicates no association between values in both distributions. The coefficient is defined as follows [45]:

$$\rho = \frac{((n^3-n)/6 - T_x - T_y - \sum r_i^2)}{\sqrt{((n^3-n)/6 - 2T_x)((n^3-n)/6 - 2T_y)}}, \text{ where:}$$

- $n$  is the number of elements in the distributions
- $r_i$  is the rank difference between  $x_i$  and  $y_i$
- $T_x$  and  $T_y$  is the correction term for ties. For an arbitrary distribution  $\alpha$ , this term is defined as  $T_\alpha = \sum(t^3 - t)$ , where  $t$  gives the number of elements of  $\alpha$  having equal ranks. [46]

#### ***E.4 Proof of capability***

We tested the software by computing results for various systems of relationships between statistical distributions and examining the results. The results of five of these tests are summarized below; in all cases our Monte Carlo software produced results which we found acceptable.

### E.4.1 Time-independent trivial case

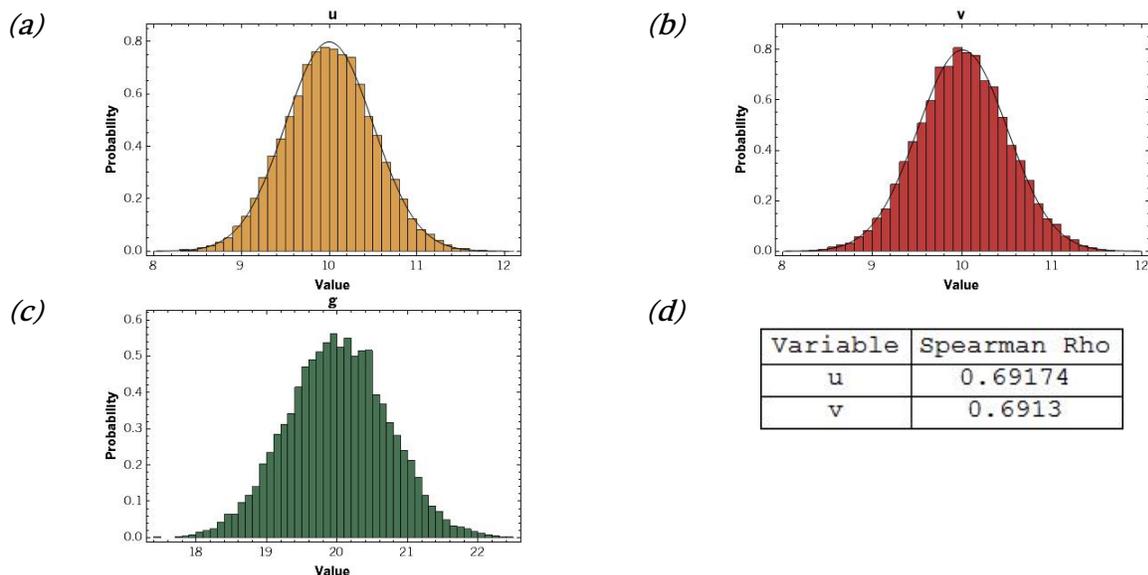
**System:**  $g(u,v) = u+v$

**Input parameter distributions:**  $u, v$  normally distributed:  $\mu=10, \sigma=0.5$

**Iterations:**  $10^4$  iterations

We first selected a very simple case: A system with an output determined by simple summation of two variables. The two variables were assigned identical two normal distributions, and the results of the output computed for  $10^4$  iterations.

For such a system, we would expect the output distribution to be normally distributed around a mean equal to the sum of the means of the input distributions. On initial inspection, we would anticipate that when comparing input distributions to the output distribution, the rank correlation coefficient should be roughly the same for each parameter. Inspection of the results of Monte Carlo analysis (see *Figure E-1: Results of Monte Carlo analysis of trivial system*) shows that both the output distribution and the values for the rank correlation coefficients match our expectations.



**Figure E-1: Results of Monte Carlo analysis of trivial system**

*(a,b) – histograms of samplings from input distributions with the original input distribution overlaid; (c) histogram of values computed for output distribution; (d) results of computation of Spearman rank correlation coefficient, comparing output distribution to each input distribution.*

### E.4.2 Time-independent case

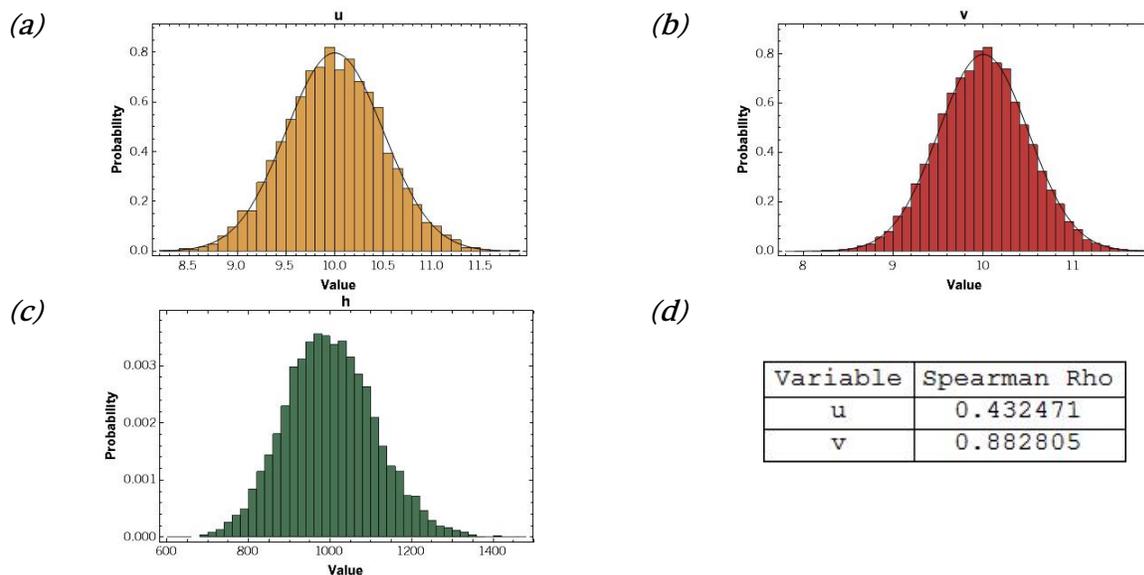
**System:**  $h(u,v) = u v^2$

**Input parameter distributions:**  $u, v$  normally distributed:  $\mu=10, \sigma=0.5$

**Iterations:**  $10^4$  iterations

We next selected a more complex case: A system with an output determined by the multiplication of one variable with the square of another. The two variables were assigned identical two normal distributions, and the results of the output computed for  $10^4$  iterations.

For such a system, we expect the output distribution to be normally distributed around a mean equal to the mean of the first input distribution ( $u$ ) multiplied by the square of the second ( $v$ ) (ie,  $\sim 1000$ ). On initial inspection, we anticipate that when comparing input distributions to the output distribution, the rank correlation coefficient for the squared parameter ( $u$ ) should be significantly higher than the coefficient for the other parameter ( $v$ ). Inspection of the results of Monte Carlo analysis (see *Figure E-2*) shows that both the output distribution and the values for the rank correlation coefficients match our expectations.



**Figure E-2: Results of Monte Carlo analysis of time-independent system**

*(a,b) – histograms of samplings from input distributions with the original input distribution overlaid; (c) histogram of values computed for output distribution; (d) results of computation of Spearman rank correlation coefficient, comparing output distribution to each input distribution.*

### E.4.3 Time-dependent case

**System:**  $m(x,y,t) = xt + \text{Log}(yt)$

**Input parameter distributions:**  $x, y$  normally distributed;  $\mu=1, \sigma=.1$

**Range of independent variable:**  $t$  runs from 0.1 to 10.0 in steps of 0.01

**Iterations:**  $10^4$  iterations at each time point

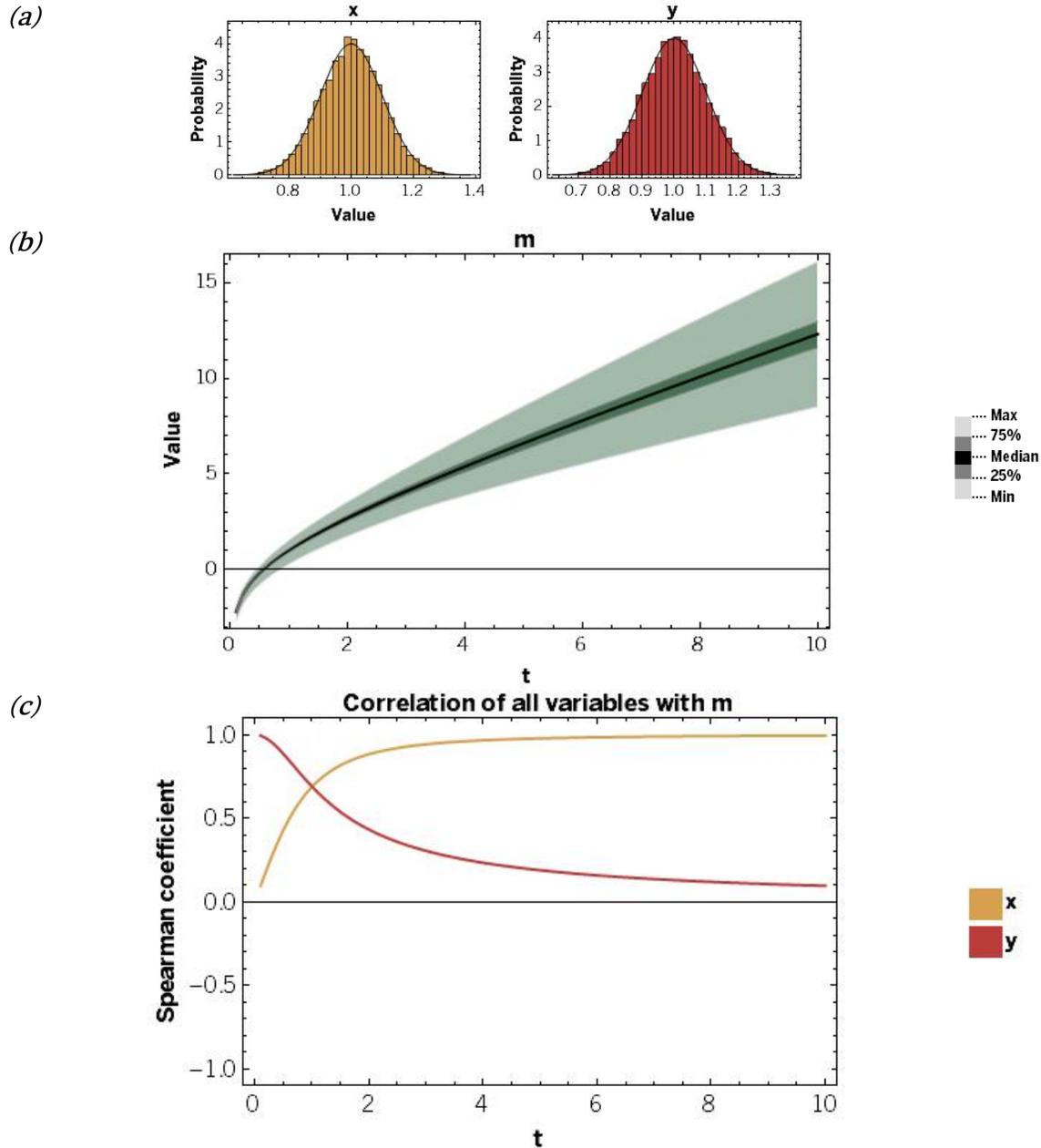
**Resampling scheme:** Input distributions are sampled once at beginning

In order to examine the use of our software with time-dependent systems, we then selected a case which involved an independent variable: A system with an output determined by two input distributions as well as a single independent variable, where each input distribution is involved in different ways with the independent variable (as shown above).

The two variables were assigned identical two normal distributions, and the results of the output computed for  $10^4$  iterations. For this system, we anticipate that the output will be such that the width of the distribution will increase along with the independent variable.

Furthermore, upon observing that for any arbitrary value  $t$ ,  $\frac{d(\log(t))}{dt} > \frac{d(t)}{dt}$ ,  $0 < t < 1$ , while  $\frac{d(\log(t))}{dt} < \frac{d(t)}{dt}$ ,  $1 < t$ , we can make an intuitive leap about the behavior of the rank correlations of the distributions of  $x$  and  $y$  when compared to  $t$ . We suspect that when comparing input distributions to the output distribution, the log-transformed input distribution ( $y$ ) might dominate early before being overtaken by the multiplicative input distribution ( $x$ ), probably switching around 1.

Inspection of the results of Monte Carlo analysis (see *Figure E-3: Results of Monte Carlo analysis of time-dependent system*, next page) shows that both the output distribution and the values for the rank correlation coefficients match our expectations.



**Figure E-3: Results of Monte Carlo analysis of time-dependent system**

(a) – plots of samplings from input distributions, showing minimum, maximum, and mean computed values over time with interquartile range highlighted; (b) plot of values computed for output distribution, showing minimum, maximum, and mean computed values over time with interquartile range highlighted; (c) results of computation of Spearman rank correlation coefficient, comparing output distribution to each input distribution over time.

#### E.4.4 Time-dependent case with resampling

**System:**  $m(x,y,t) = xt + \text{Log}(yt)$

**Input parameter distributions:**  $x, y$  normally distributed;  $\mu=1, \sigma=.1$

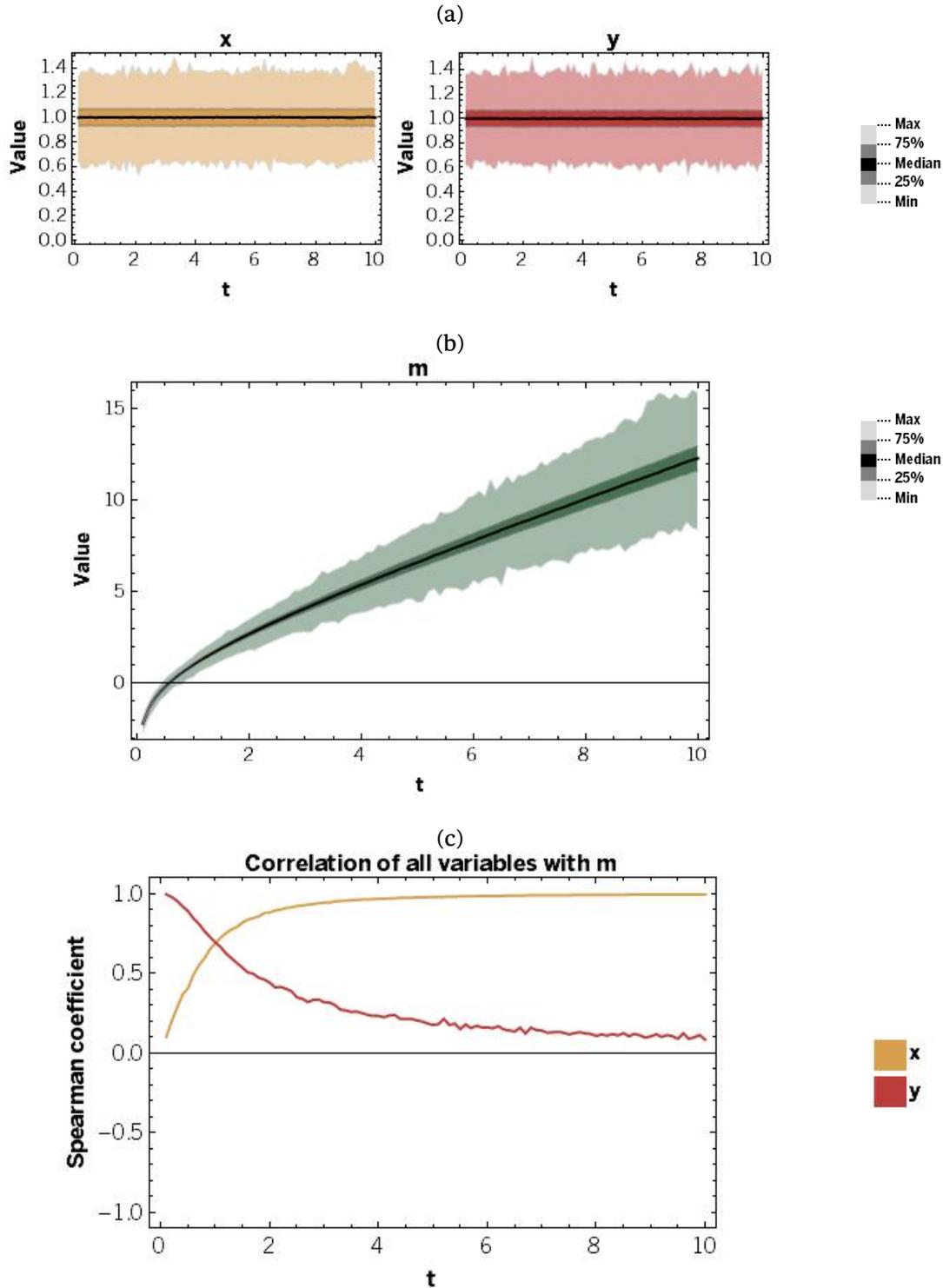
**Range of independent variable:**  $t$  runs from 0.1 to 10.0 in steps of 0.01

**Iterations:**  $10^4$  iterations at each time point

**Resampling scheme:** Input distributions are resampled at every time point

We investigated the use of our software with time-dependent systems utilizing the resampling scheme wherein the distributions for input parameters were resampled at every time point, rather than only once prior to computation. The same system as described in the previous case was utilized for this. For this system, we anticipated that the results from this would generally be comparable to the previous case.

Inspection of the results of Monte Carlo analysis (see *Figure E-4: Results of Monte Carlo analysis of system with resampling*) shows that the output from our software matches our expectations. We are encouraged somewhat to see clear evidence of the “noisy” resampling method in these results.



**Figure E-4: Results of Monte Carlo analysis of system with resampling**

(a) – plots of samplings from input distributions, showing minimum, maximum, and mean computed values over time with interquartile range highlighted; (b) plot of values computed for output distribution, showing minimum, maximum, and mean computed values over time with interquartile range highlighted; (c) results of computation of Spearman rank correlation coefficient, comparing output distribution to each input distribution over time.

#### E.4.5 Time-dependent case with summation and resampling

**System:**  $n(x,y,z,t) = (xt + \text{Log}(yt))e^{-zt}$

**Input parameter distributions:** x, y, z normally distributed;  $\mu=1, \sigma=.1$

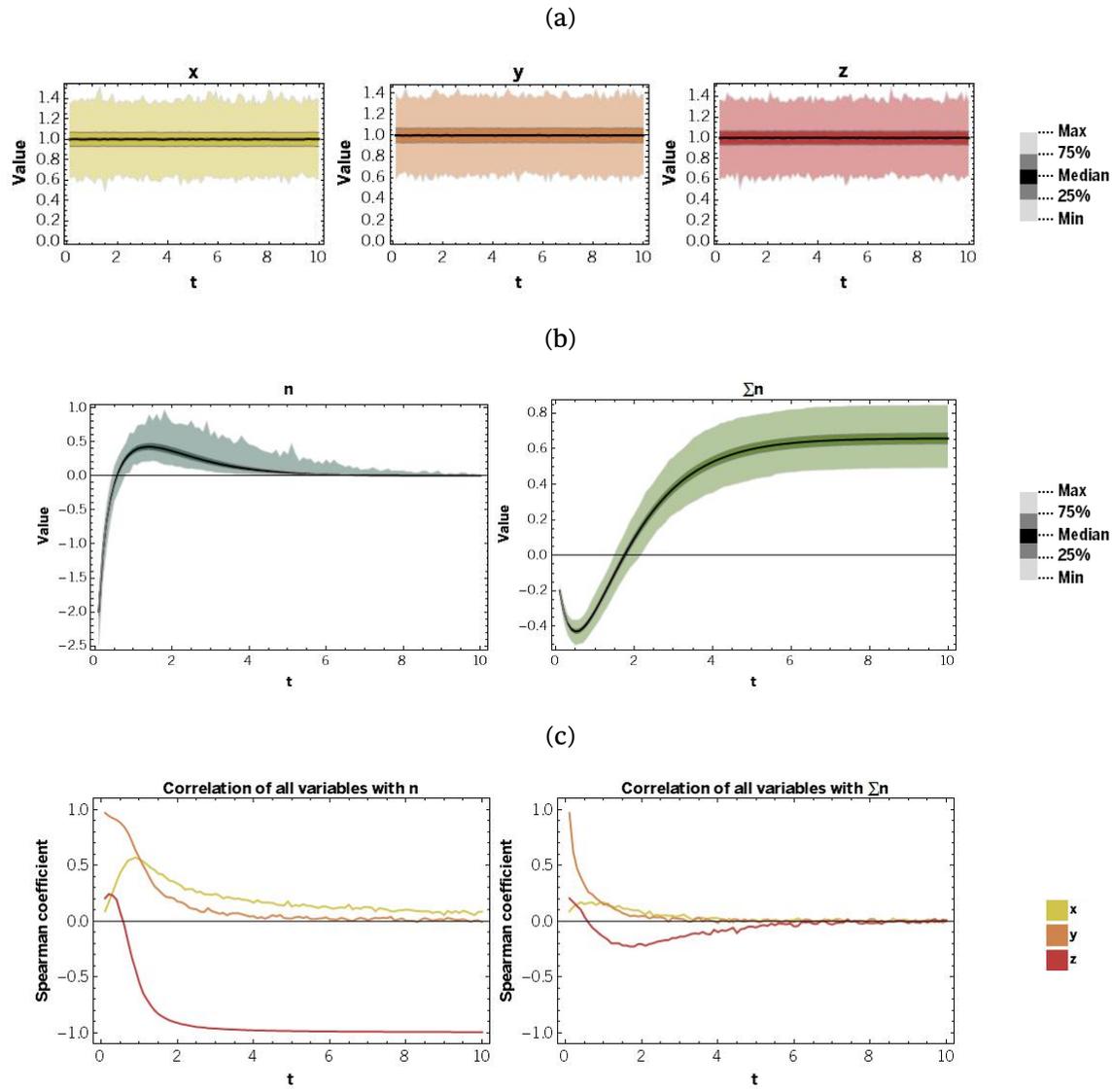
**Range of independent variable:** t runs from 0.1 to 10.0 in steps of 0.01

**Iterations:**  $10^4$  iterations at each time point

**Resampling scheme:** Input distributions are resampled at every time point

Finally, we investigated the use of our software with time-dependent systems utilizing both resampling scheme and computing a running total for our output. A similar system as described in the previous case was utilized for this, with the addition of an exponential component designed to mimic the decay processes present in our dose assessment methodology. While generally we still expect y to dominate at first, this system is complex enough that it is not immediately clear which input distributions might be closely correlated to the output distribution as time increases. It is unclear what effect summation will have.

Inspection of the results of analysis (see *Figure E-5: Results of Monte Carlo analysis of system with summation*) shows that in this case, our expectations are initially correct, but that the exponentiated input distribution becomes closely correlated with the output distribution. Additionally, we see that as time passes, no input distribution maintains correlation with the running total of the output distribution. This is due to the use of the resampling scheme – because the input parameters are resampled at each time point, and contributing less and less to the running total as time goes on, they become less correlated to the sum of the output. When repeated without resampling, this is not the case.



**Figure E-5: Results of Monte Carlo analysis of system with summation**

*(a) – plots of samplings from input distributions, showing minimum, maximum, and mean computed values over time with interquartile range highlighted; (b) plot of values computed for output distribution and running total of output distributions, showing minimum, maximum, and mean computed values over time with interquartile range highlighted; (c) results of computation of Spearman rank correlation coefficient, comparing output distribution and running total of output distributions to each input distribution over time.*

**APPENDIX F. Sample extract/transact/load processing checklist**

<b>FILENAME</b>	rawparts2_3_final_masha_june10.sav
<b>File Date</b>	6/11/2010
<b>Survey Date</b>	5/24/2010
<b>Pass:</b>	2
<b>Notes</b>	Filename renamed to append " - corrected 3" for clarity. Survey data is unclear, 5/24/2010 is the date in the filename.

		Survey Subset (ID 1..5)	Full Survey	Notes	
0	<b>EXPORT FROM SPSS &gt;&gt; CSV</b>	n/a	SUCCESSFUL	File was in different format than last file, contained approximately 600 fewer columns, none pertinent to external dose calculation	
	a.	Verify SPSS data format	n/a	Successful	
	b.	Amend DATEBORN field	n/a	Successful	DATEBORN field in SPSS file is in DD-MMM-YYYY format instead of MM/DD/YYYY format and must be corrected in SPSS.
	c.	Export data from SPSS >> CSV	n/a	Successful	Recall that some fields in SPSS contain commas, thus output from SPSS file must be in CSV format with field delimiters of some sort.
	d.	Copy data from SPSS export location (C:\cherdata.csv) to data import process source directory	n/a	Successful	Recall that SPSS has difficulty with "deep" directory structures and long filenames, thus SPSS output is restricted to the root directory for this output.
	e.	Create subset of data for import testing using subject ID=(1..5).	n/a	Skipped	
f.	Change import script target files to match current file name if needed	Skipped	Skipped		

		Survey Subset (ID 1..5)	Full Survey	Notes	
1	<b>DATA REFORMATTING</b>	Skipped	SUCCESSFUL		
	a.	Find Columns	Skipped	Successful	
	b.	Reduce survey to radiation data alone	Skipped	Successful	This also done in the SPSS export, but this step also creates the working file format.
	c.	Fill fields with placeholder markers ((",", ".", " ", "...")) with empty values.	Skipped	Successful	Numerous fields necessary for dose calculation contain "... " or ". " instead of data, these will be extended from the previous row.
	d.	Fix column errors	Skipped	Successful	MNTH3R1, MNTH3R4 are mislabelled as DAY3R1, DAY3R4. This is a known/expected problem.
	e.	Denormalize from Subjects >> Residences	Skipped	Successful	
	f.	Concatenate columns	Skipped	Successful	
	h.	Remove subjects with blank fields	Skipped	Successful	
	i.	Manually copy output from "good data" and "bad data" files into Excel file for data review, check to verify no "bad data" records	Skipped	Successful	
	j.	Manually copy output to Excel file "CherDB Import.XLS"	Skipped	Successful	

		Survey Subset (ID 1..5)	Full Survey	Notes	
2	<b>CREATE DATABASE: CHERDB</b>	Skipped	SUCCESSFUL		
	a.	Create CherDB database (manual)	Skipped	Successful	
	b.	Build structure	Skipped	Successful	Use "Create CherDB.sql" unless troubleshooting individual tables/import process, in which case use "Create CherDB Barebones.sql"
	c.	Import data from Excel file "CherDB Import.XLS" using DTSX import script	Skipped	Successful	
	d.	Populate lookup tables	Skipped	Successful	

			Survey Subset (ID 1..5)	Full Survey	Notes
<b>3</b>	<b>MAPPING ACQUISITION</b>		Skipped	SUCCESSFUL	
	a.	Export lat/long data from table LookupLatLongPreliminary to tab-delimited file "A - LatLong Coordinates.DAT" (manual)	Skipped	Successful	
	b.	Convert lat/long >> x/y	Skipped	Successful	Working note: Variable names inside this script should be adjusted. Also note that this script has been amended to return negative values if appropriate
	c.	Collect x/y colors	Skipped	Successful	Note: X/Y values that are out-of-bounds compared to the pixel-image map are rejected and returned with a value of "0"
	d.	Convert colors to CS137	Skipped	Successful	
	e.	Copy output to Excel file "CherDB cesium import.xls"	Skipped	Successful	Note that column names must be preserved from the Excel file template in order to match the titles of column names in the import script; these columns are titled differently in the CS137 acquisition process
	f.	Import lat/long/CS137 to CHERDB from file "CherDB cesium import.xls"	Skipped	Successful	
	g.	Populate lat/long lookup table	Skipped	Successful	
	h.	Acquire off-grid CS137 (manual)	n/a	Successful	Note that most values here fall into the [2..10] range of min/max Cs137 values; this is based on the European overview plate (Plate 01), which has different ranges than the Ukraine plate. Generally speaking, areas South and West of Europe are extrapolated to the [1..2] range, while areas North and East of Europe are extrapolated to the [2..10] range.
	i.	Import manual CS137 data points to database	n/a	Successful	
j.	Populate lat/long lookup table with manually acquired CS137 data		Successful		
			Survey Subset (ID 1..5)	Full Survey	Notes
<b>4</b>	<b>DATA ANALYSIS</b>		Skipped	SUCCESSFUL	
	a.	Join identity fields/lookups	Skipped	Successful	
	b.	Populate subject, residence tables	Skipped	Successful	
	c.	Create denormalized-by-day tables	Skipped	Successful	
	d.	Calculate dose	Skipped	Successful	
	e.	Calculate cumulative dose	Skipped	Successful	
			Survey Subset (ID 1..5)	Full Survey	Notes
<b>5</b>	<b>REPORTING</b>		Skipped	SUCCESSFUL	
	a.	Generate cumulative dose at yearly intervals	Skipped	Successful	
b.	Generate cumulative dose at three "wave" intervals	Skipped	Successful		

APPENDIX G. Verification of extrema

Rank	Subject ID	Dose by 2009 (mSv)	Justification
1	105	0.05	Subject resided in negligible-radiation locations for the duration of the study, with minimum starting <sup>137</sup> Cs deposition of 1 kBq/m <sup>2</sup>
2	93	0.09	Subject resided in negligible-radiation locations for the duration of the study, with minimum starting <sup>137</sup> Cs deposition of 2 kBq/m <sup>2</sup>
3	265	0.13	Subject resided in negligible-radiation locations for the duration of the study, with minimum starting <sup>137</sup> Cs deposition of 2 kBq/m <sup>2</sup>
4	131	0.14	Subject resided in negligible-radiation locations for the duration of the study, with minimum starting <sup>137</sup> Cs deposition of 2 kBq/m <sup>2</sup>
5	270	0.14	Subject resided in negligible-radiation locations for the duration of the study, with minimum starting <sup>137</sup> Cs deposition of 2 kBq/m <sup>2</sup>
699	617	14.5	Subject resided in a relatively low-radiation location for the first eight months of the study period, where minimum starting <sup>137</sup> Cs deposition was 10 kBq/m <sup>2</sup> . On 2 January 1987, the subject relocated to a higher radiation location with minimum starting <sup>137</sup> Cs deposition of 185 kBq/m <sup>2</sup> and resided there for the duration of the study period.
700	614	14.7	Subject resided in a medium-radiation location for the entire duration of the study, with starting <sup>137</sup> Cs deposition at 100 kBq/m <sup>2</sup>
701	612	16.0	Subject resided in a medium-radiation location for the almost the entire duration of the study, with starting <sup>137</sup> Cs deposition at 100 kBq/m <sup>2</sup> . The subject spent 24 days in a low-radiation location, with starting <sup>137</sup> Cs deposition of 4 kBq/m <sup>2</sup> . This individual worked in an occupation that spent considerable time outdoors (occupation/dose factor of 0.38 compared to 0.24) for the first seven months after the reactor accident, and for four additional years of the study period.
702	690	30.6	Subject resided in a medium-radiation location for the first seven years of the study period, with starting <sup>137</sup> Cs deposition at 185 kBq/m <sup>2</sup> . The subject spent the remainder of the study period in a low-radiation location, with starting <sup>137</sup> Cs deposition of 10 kBq/m <sup>2</sup> .
703	767	30.9	Subject resided in a very hot spot in Belarus for the first seven months after the reactor accident; then lived in the same location for another ten years intermittently over the course of the study period.

APPENDIX H. Previously published estimates for dose for similar cohorts

Date	Estimate	Units	External or Internal+External?	Description	Ref
12/31/1995	5	mSv	External	Ukraine population in contaminated areas (>37kBq/m2)	[6]
4/30/1986	20	mSv	External	Ukraine individuals in immediately evacuated areas	[6, 18, 47]
12/31/1986	185	mGy	External	Ukrainian operations workers	[48, 49]
4/30/1986	10.1	mSv	External	Evacuees from Pripyat	[41]
12/31/1986	1.6	mSv	External	Ukraine population in raions of Kiev which were subject to focused measurements	[41]
12/31/1986	2.1	mSv	Internal+External	Ukraine population in raions of Zhitomir	[41]
12/31/1986	2.1	mSv	Internal+External	Ukraine population in raions of Kiev	[41]
12/31/1986	3.6	mSv	External	Ukraine population in raions of Zhitomir which were subject to focused measurements	[41]
12/31/2005	4.9	mSv	Internal+External	Ukraine population in raions of Kiev	[41]
12/31/2005	5.9	mSv	Internal+External	Ukraine population in raions of Kiev which were subject to focused measurements	[41]
12/31/1986	0.45	mSv	External	Ukraine population in Kiev city; maybe contaminated areas	[41, 50, 51]
12/31/1986	0.57	mSv	External	Ukraine population outside of Kiev, Zhitomir, Rivno, and Chernihiv; maybe contaminated areas	[41, 50, 51]
12/31/1986	1.4	mSv	External	Ukraine population in Kiev oblast; maybe contaminated areas	[41, 50, 51]
12/31/1986	1.6	mSv	External	Ukraine population in Zhitomir oblast; maybe contaminated areas	[41, 50, 51]
12/31/2005	1.3	mSv	External	Ukraine population in Kiev city; maybe contaminated areas	[41, 50, 51]
12/31/2005	1.9	mSv	External	Ukraine population outside of Kiev, Zhitomir, Rivno, and Chernihiv; maybe contaminated areas	[41, 50, 51]
12/31/2005	3.9	mSv	External	Ukraine population in Kiev oblast; maybe contaminated areas	[41, 50, 51]
12/31/2005	5.7	mSv	External	Ukraine population in Zhitomir oblast; maybe contaminated areas	[41, 50, 51]
12/31/1986	0.35	mSv	External	Ukraine population in rural areas	[20]
12/31/2000	0.89	mSv	External	Ukraine population in rural areas	[20]
4/30/1986	17	mSv	External	Evacuees from Pripyat	[52]
5/26/1986	1.2	mSv	External	USSR population in most contaminated regions (Belarus, Ukraine, Western Russia)	[52]
12/31/1995	3.7	mSv	External	Ukraine population in Kiev oblast; in contaminated areas >37kBq/m2 but <185kBq/m2	[47]
12/31/1995	4	mSv	External	Ukraine population in Kiev oblast; in contaminated areas >37kBq/m2	[47]
12/31/1995	4.8	mSv	External	Ukraine population in Zhitomir oblast; in contaminated areas >37kBq/m2 but <185kBq/m2	[47]
12/31/1995	7	mSv	External	Ukraine population in Zhitomir oblast; in contaminated areas >37kBq/m2	[47]
12/31/1995	8	mSv	External	Ukraine population in contaminated areas (>37kBq/m2)	[18, 47]
4/30/1986	33	mSv	External	Evacuees from Pripyat	[53]