



Australian Government

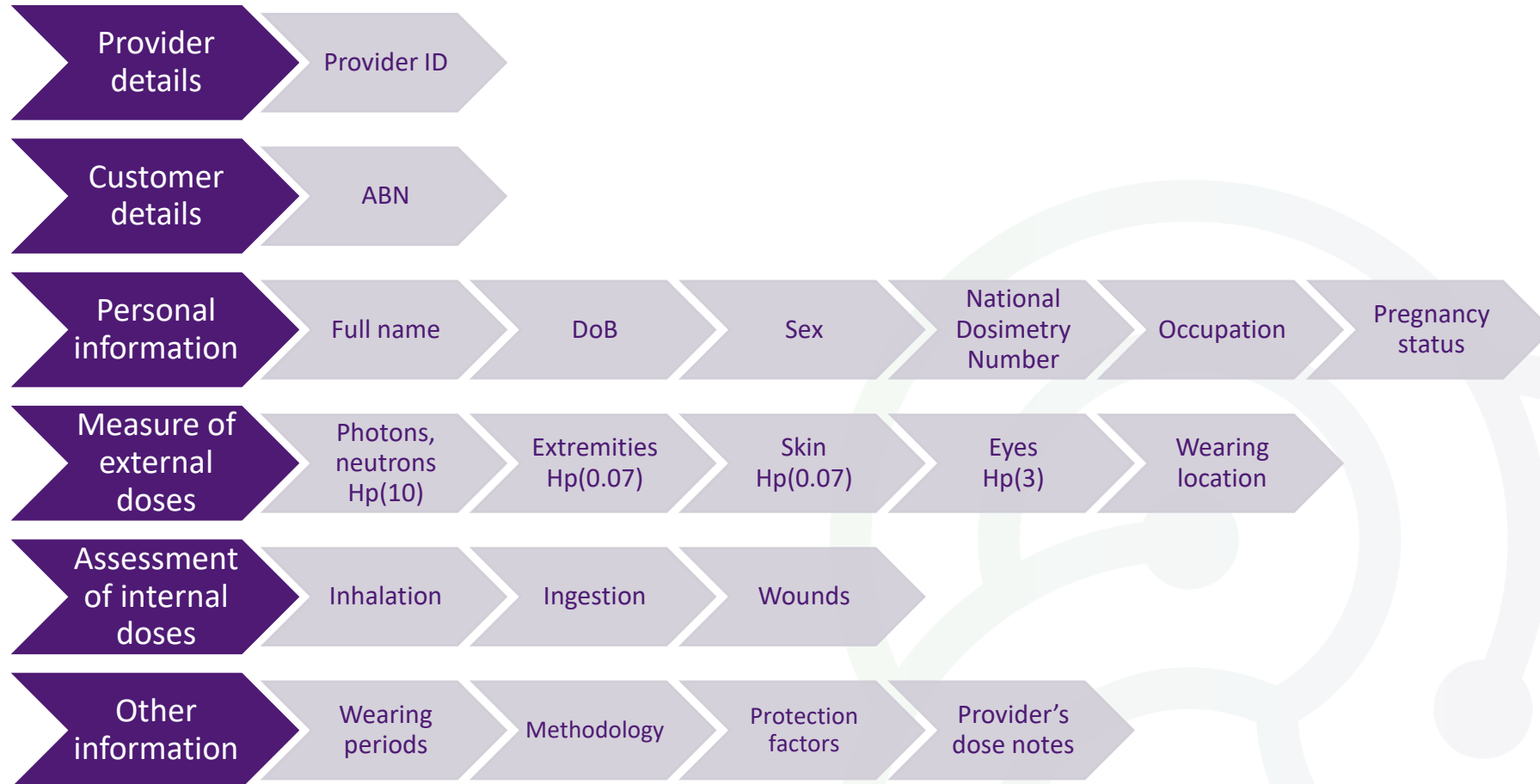
Australian Radiation Protection
and Nuclear Safety Agency



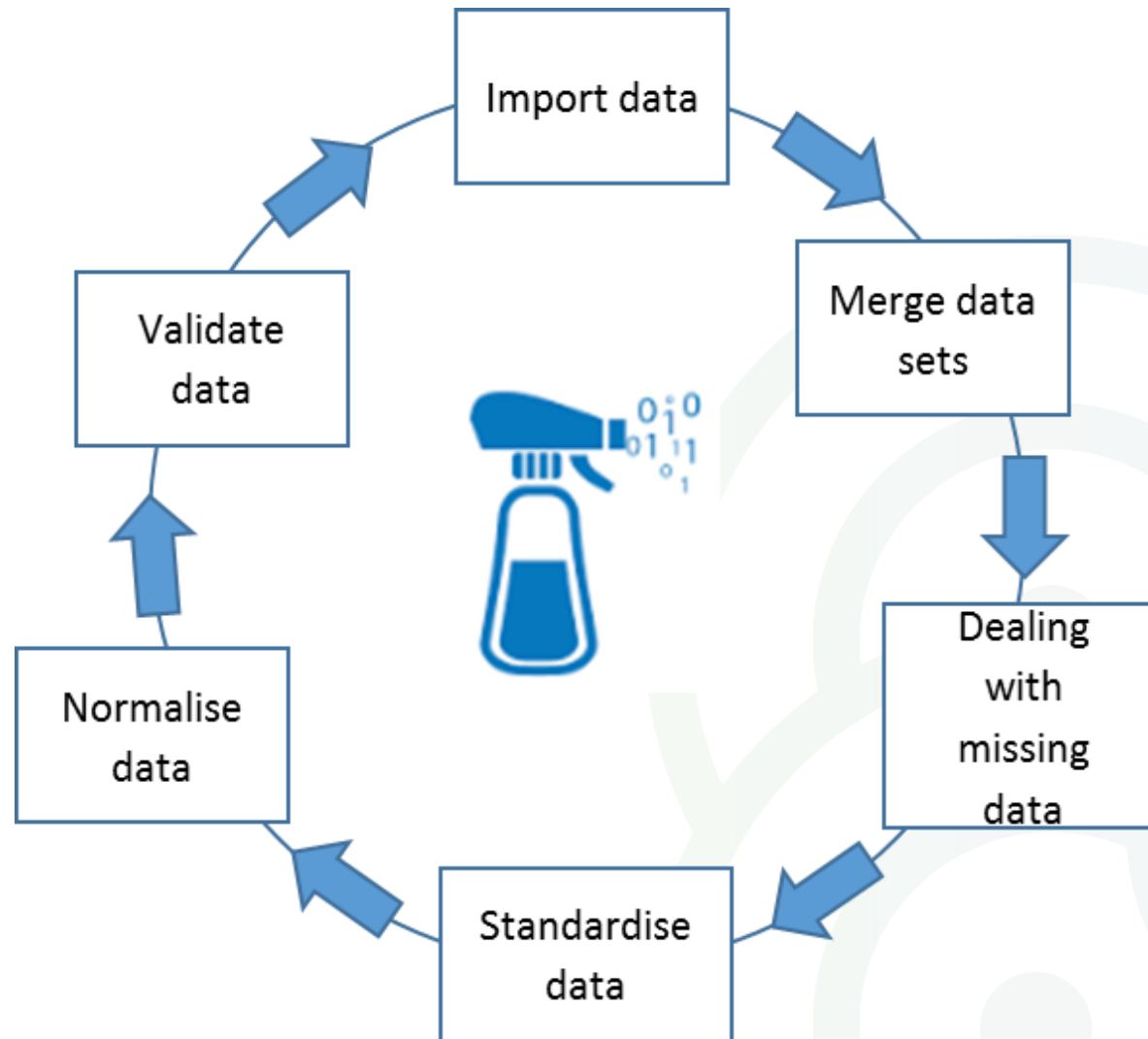
Data Analysis

Cleansing, validating & analysing the ANRDR dataset

Information we collect

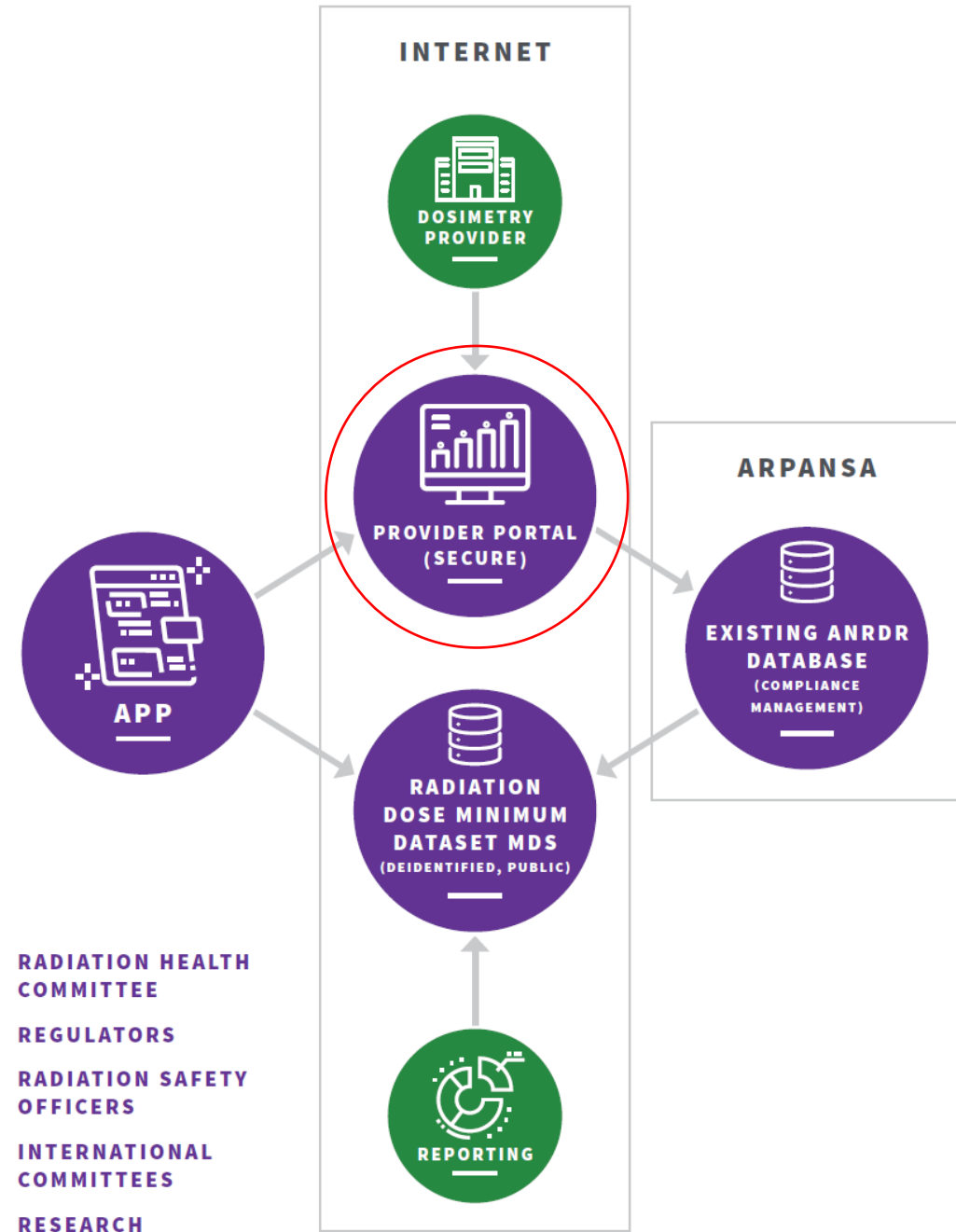


The data validation process



Import data

- Data is provided to the ANRDR from **dosimetry service providers** and **uranium operators** according to the ANRDR format as a .csv file
- If your system uses different formats, the first step is to convert all files into the same format

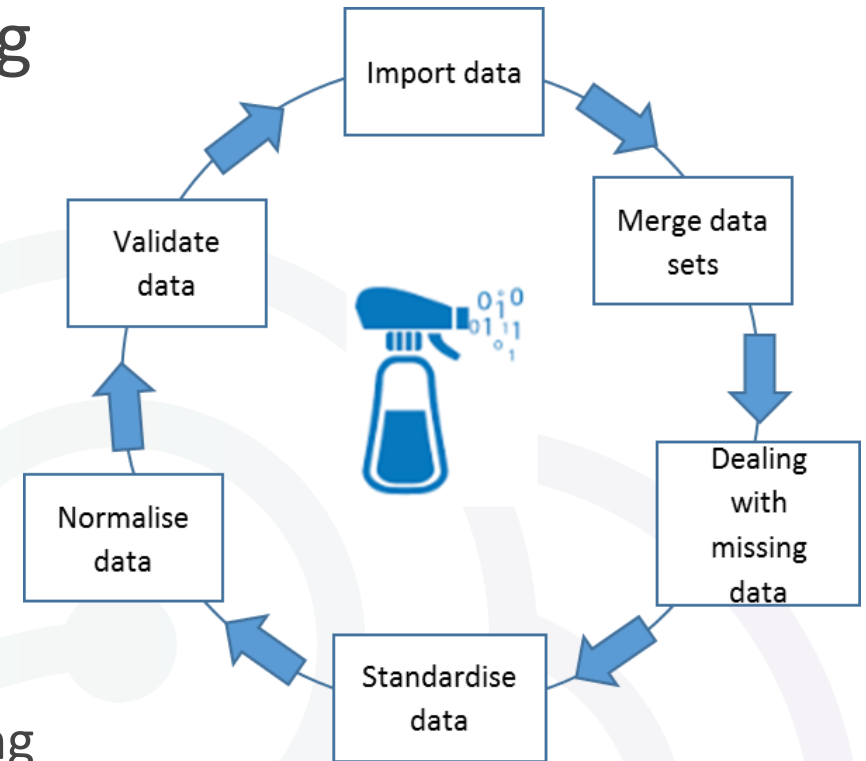


Merge datasets

- If data provided in different formats, merging files into a common database using any statistical software like SPSS, Stata, Excel
- Better method is to mandate standardised format to minimise errors and missing data

Missing data

- To reduce missing data:
 - Mandatory fields (e.g. ANRDR no longer accepts missing 'sex' field)
 - Included in the data validation process

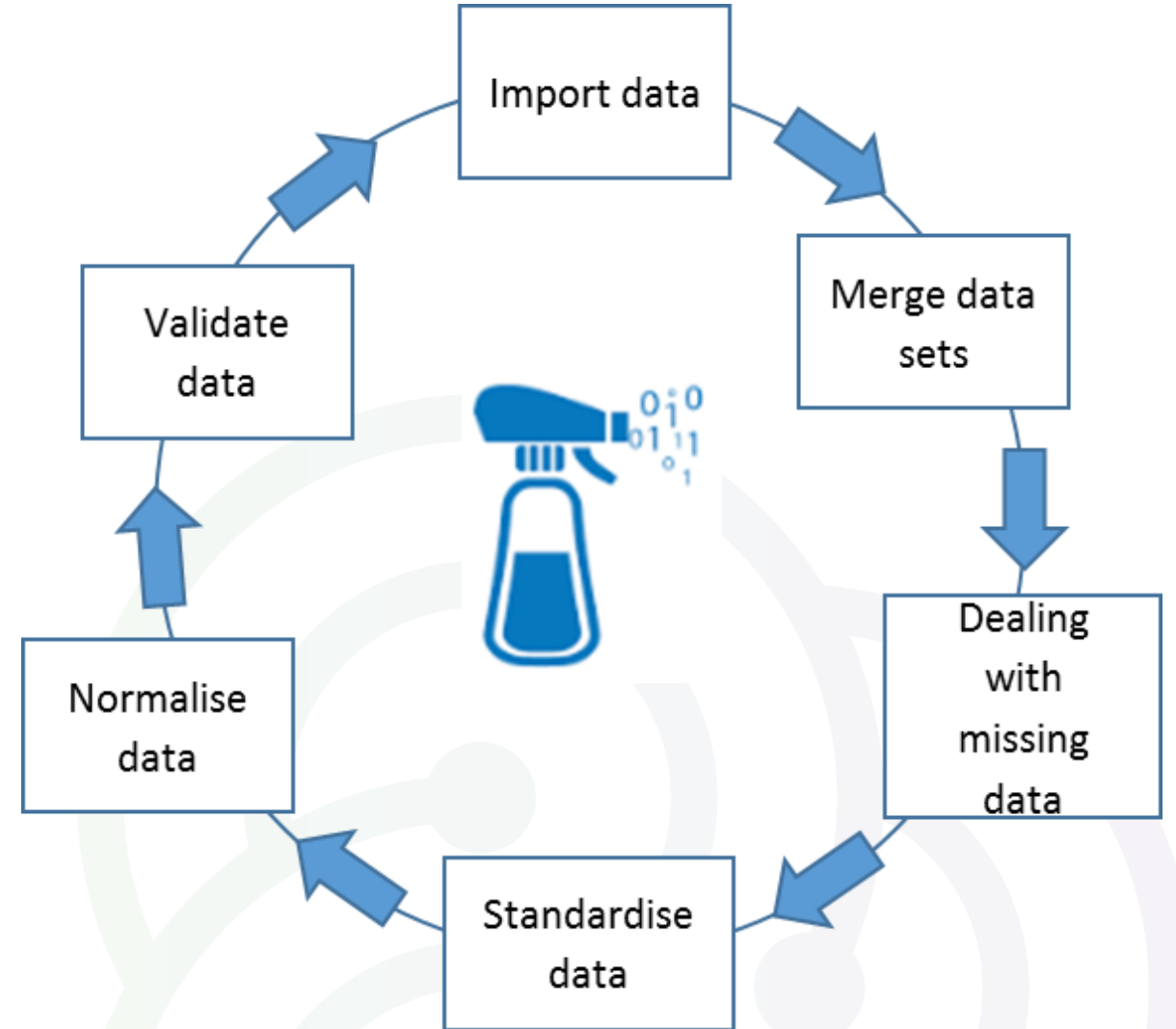


Standardised data

- Check data to ensure that the required and same type of data exists in each column (validation)
 - E.g. in the 'NDN' column you should not have 'date'

Normalised data

- Each column must have the same format
 - E.g. same scale for doses (mSv)
 - E.g. same format for sex column (code 1,2,3, not F/M/X)



Validating data

- Essential part of any data handling task is to check the accuracy and quality of data (correct and useful for the context)
- Happens at the point of data entry (before it's saved into the database)
- Improves the quality of data entered into a database
- Error report automatically generated if any validation fails
- Different methods can be used to validate data:
 - Data presence: mandatory fields cannot be left empty
 - Data type: only numbers accepted in certain fields (not letters or other characters)
 - Data format: dates of birth in YYYYMMDD format (not DDMMYYYY, MMDDYYYY or DMYY)
 - Data range: incorrectly recorded dates of birth (<16 or >80 years old at time of registration)
 - Data length: Australian Business Number field must be exactly 11 digits
 - Data lookup: Australian Business Register

Data cleansing

- Can occur any time after the data is entered into the database
- Data cleansing is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate or incomplete data within the dataset
- Validation and cleansing both deal with detection and removal of incorrect records *before* data analysis occurs
- Duplicate entries were created for workers in the ANRDR due to slight variation in personal information (if not 100% match)
- Can occur if a worker has multiple employers or changes employers; manual data entry can include errors; name change due to marriage
- Caused by lack of available national identification number
 - no social security number
 - not allowed to use Tax File Number or Medicare Number

De-duplication

- ARPANSA purchased record-matching software. The software works by identifying common features in personal details across all individuals in the database and making suggestions for 'possible matches' by providing a percentage match to assist with identifying separate profiles for the same individual
- Unique identification number eliminates duplication issues

Record-matching software

- Enables us to identify duplicate data records or database entries in order to merge the data, and improve data quality and data accuracy
- First run removed 27,000 duplicate workers!

[illegible]

Statistical analysis



Average & max effective doses with workforce numbers for ARPANSA

Average ARPANSA staff doses are well below 1 mSv/y

An average of less than 0.2 mSv/y over the 30 year period

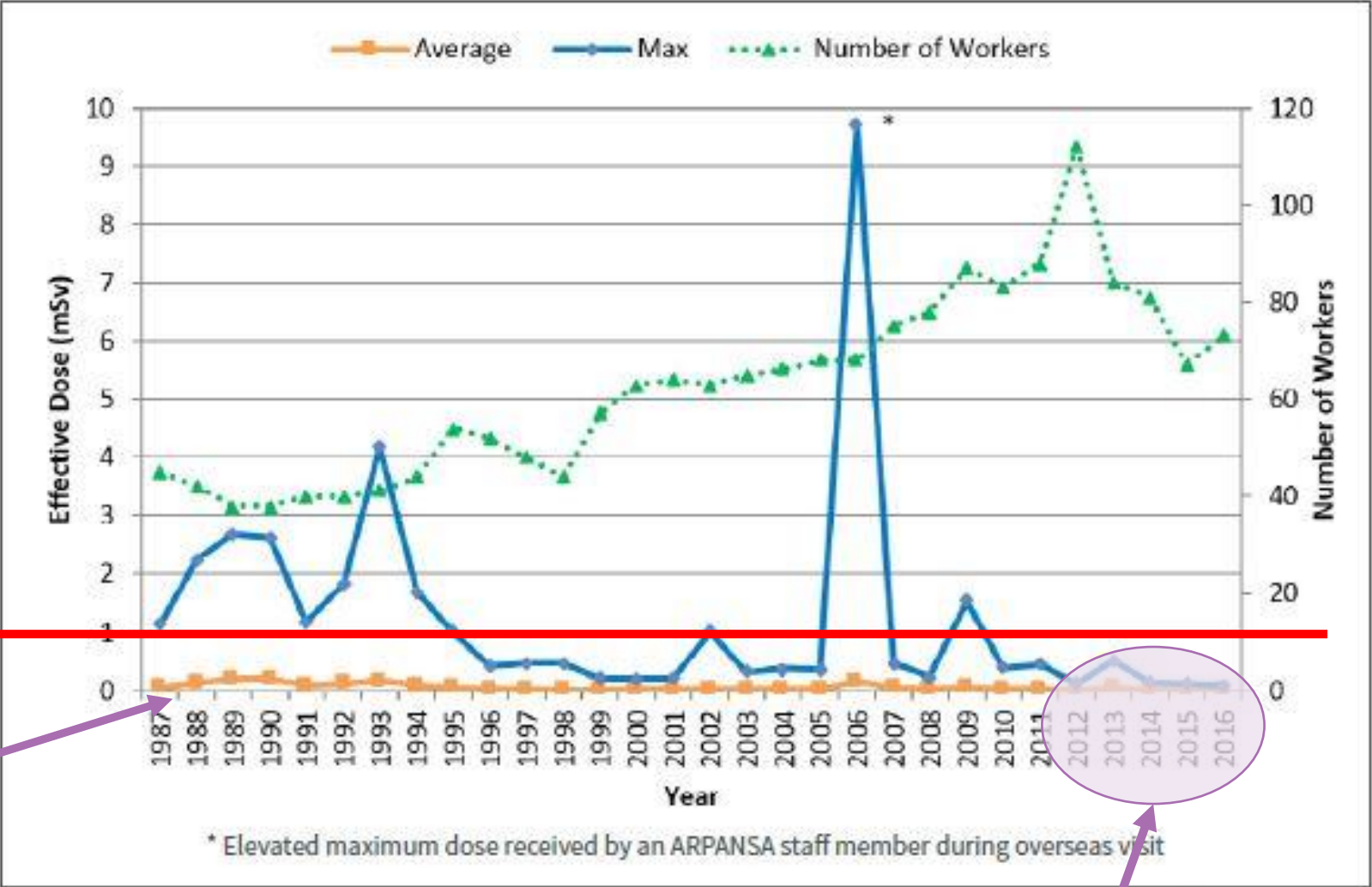


Figure 1: Average and maximum effective doses with staff numbers for ARPANSA staff (1987–2016)

Average exposures over the last 5 years are less than 0.1 mSv/y.

Uranium industry average & max doses with workforce numbers

The number of workers in the industry has also been declining since 2014 as a result of reduced market conditions.

Demonstrates that **average** and **maximum** effective doses for uranium industry workers continue to proceed in a **downward trend from 2014** with average and maximum effective doses in 2016 of 0.75 mSv and 5.6 mSv, respectively.

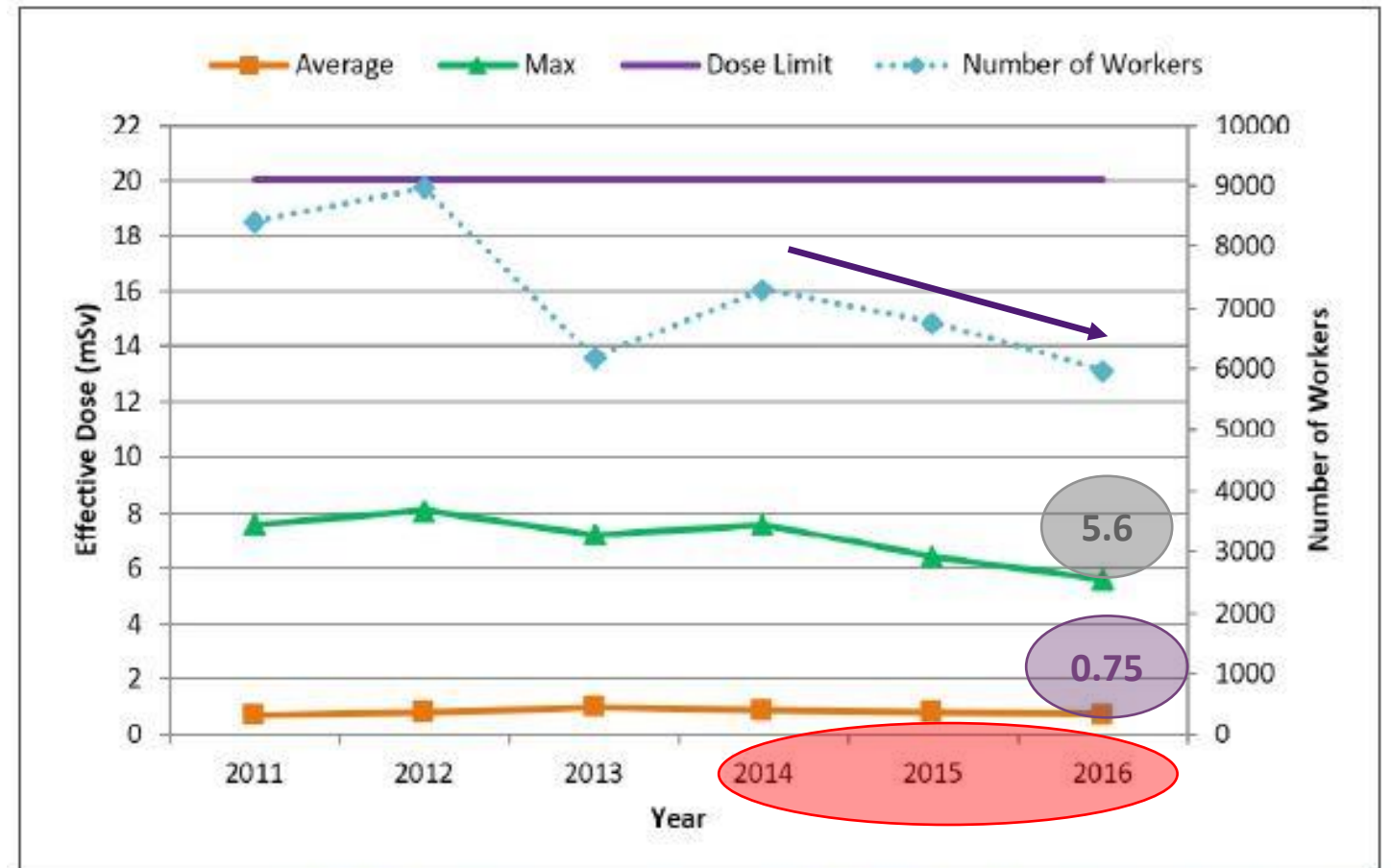


Figure 3: Uranium industry average and maximum effective doses with workforce numbers (2011–2016)

Uranium industry average effective doses by category

The mining category continues to be the highest exposed worker group in the Australian uranium industry.

This can be attributed to the fact that due to its large workforce, the mining category is heavily skewed towards underground operations.

A continued *downward trend* in occupational exposures for this work category with an average effective dose in 2016 of 1.24 mSv.

In the same year, average exposures to processing and other workers remain low at 0.44 mSv and 0.25 mSv, respectively.

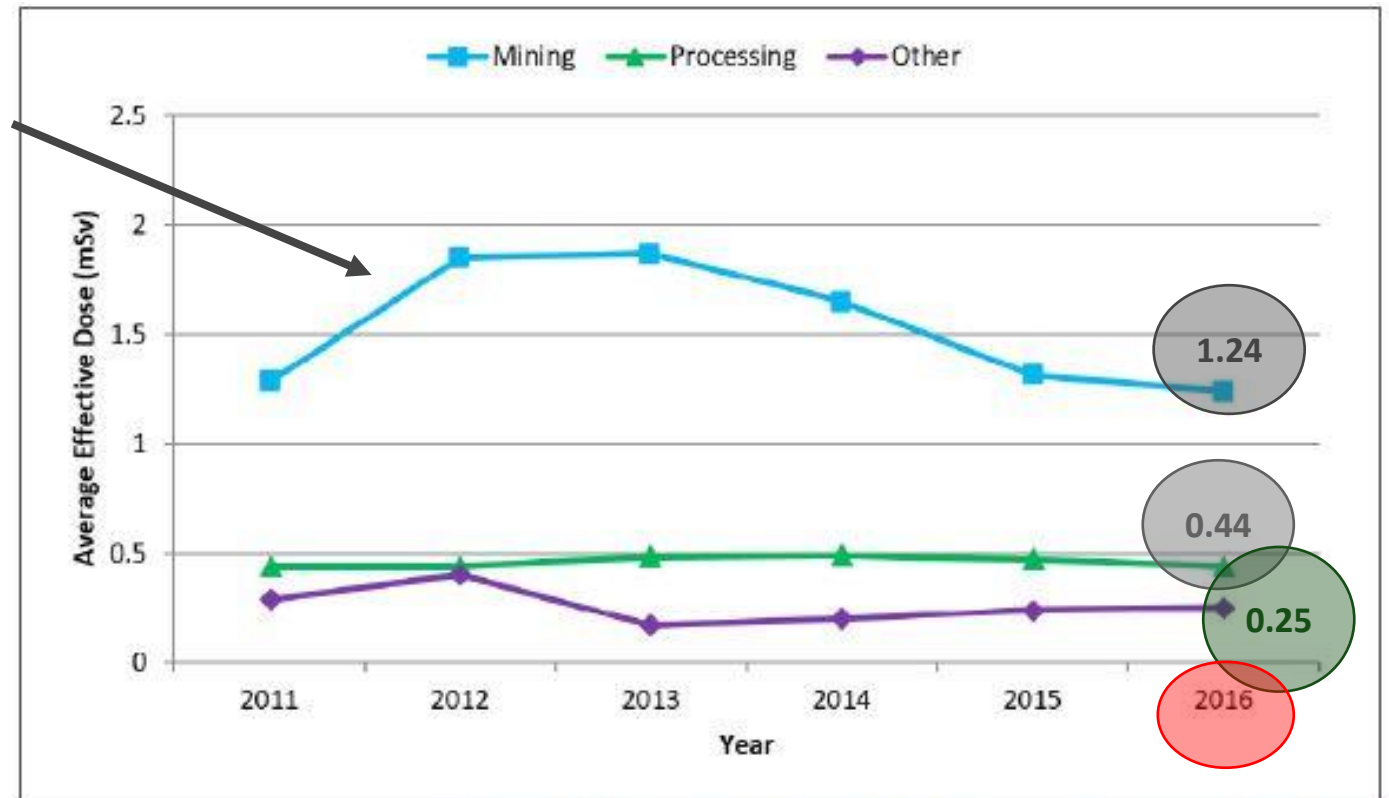


Figure 4: Uranium industry average effective doses by worker categories (2011–2016)

Collective effective dose with uranium production

Demonstrates that over the period 2014-2016 the **collective effective doses have decreased** while **uranium production has increased**

The decrease in the collective effective dose is due to a decrease in uranium industry worker numbers and a decrease in average effective doses over that period

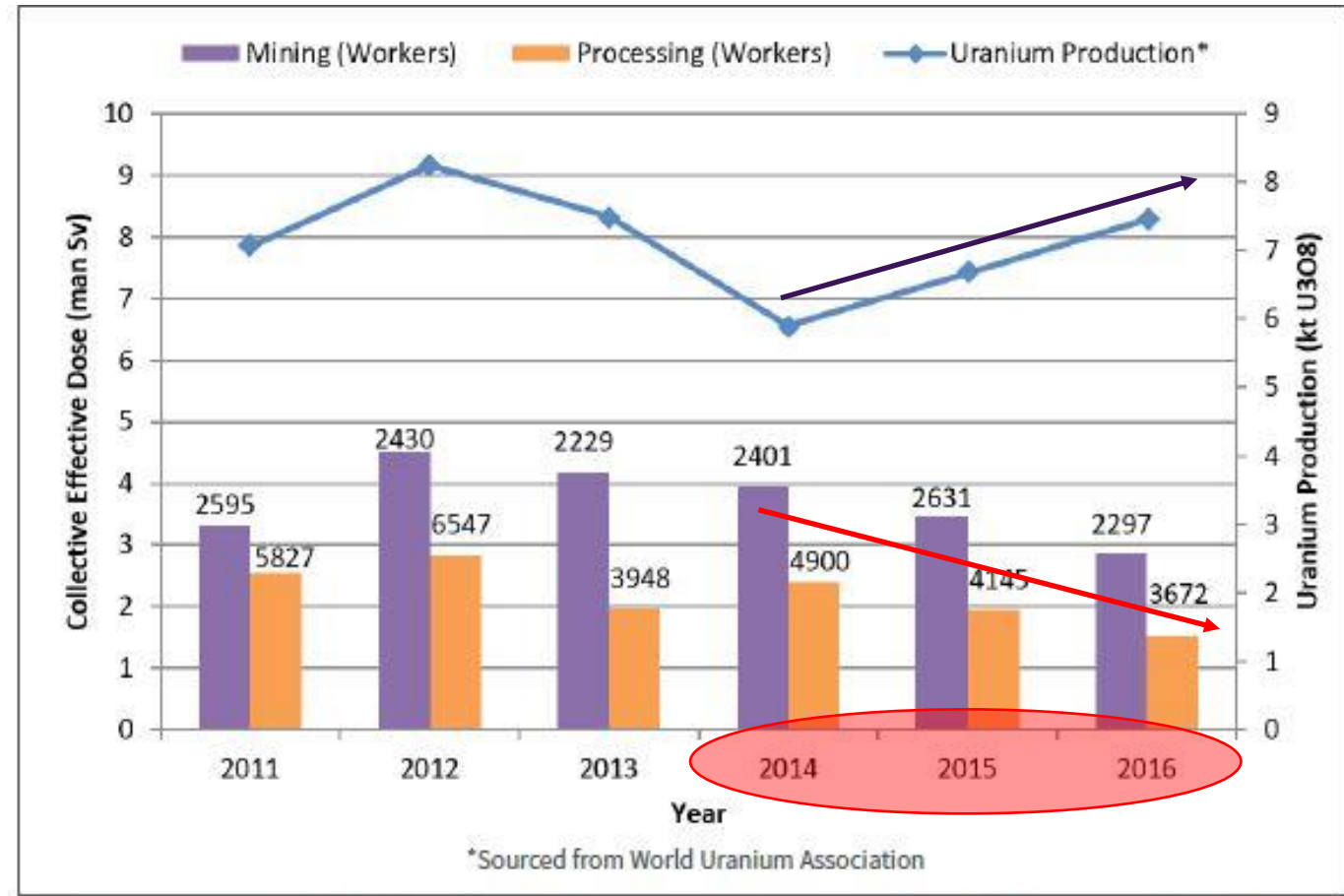
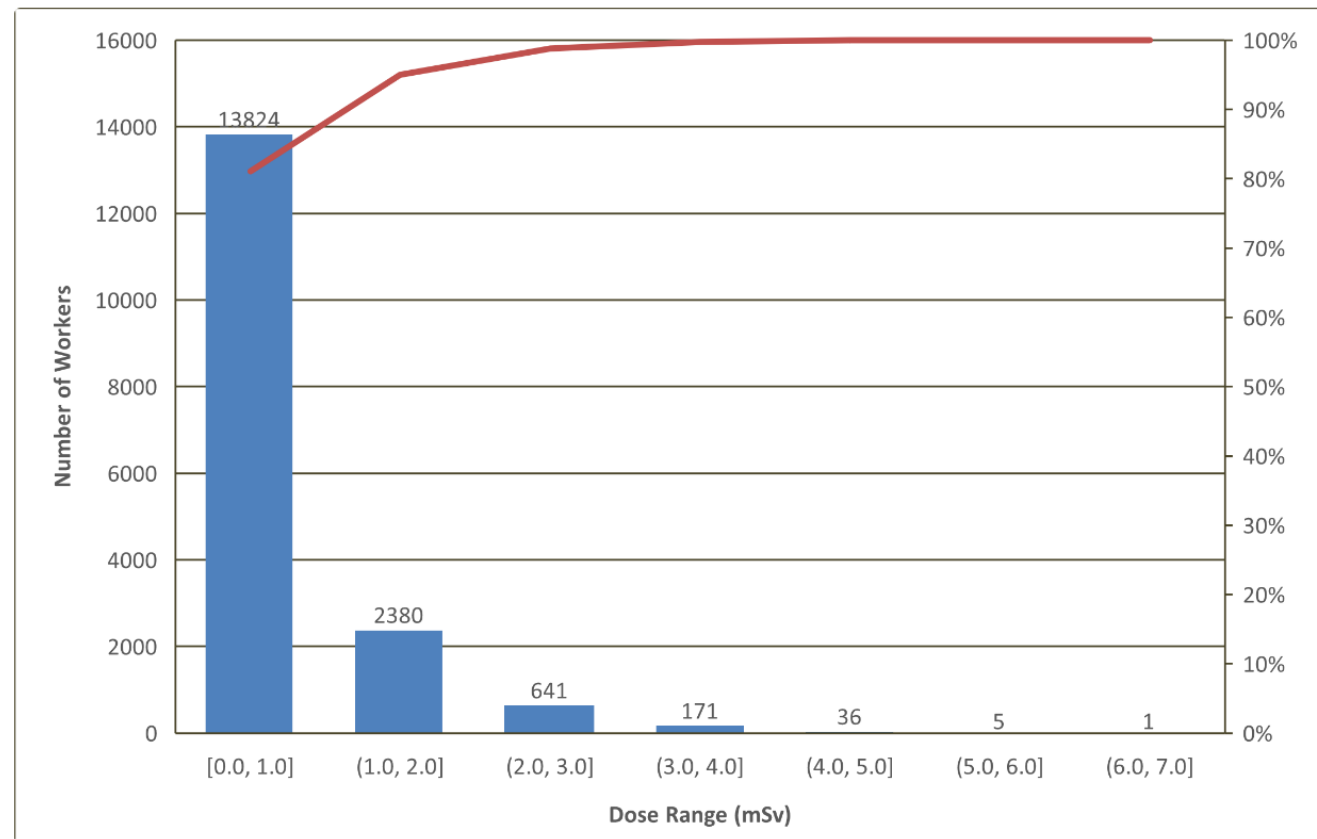


Figure 2: Australian uranium industry collective effective doses and uranium production with worker numbers (2011–2016)

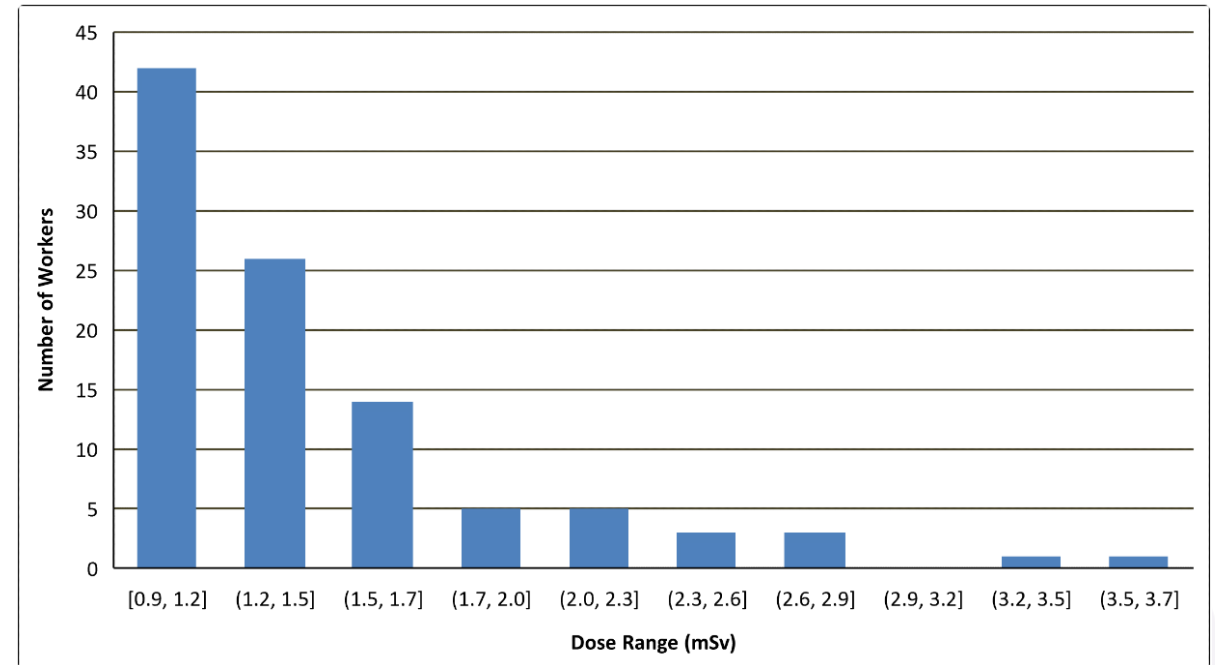
Dose distribution for all ANRDR data 2020

- 81% of workers received a dose <1 mSv
- 99% of workers received a dose <3 mSv
- Effective way to show the distribution of doses and can minimise the impact of data skewing for doses reported as <MRD



Top 100 doses for Commonwealth organisations

- Top 100 doses range from 0.89 – 3.74 mSv with an average of 1.42 mSv in 2020
- Can help regulators tailor their inspection regime and facilitate discussions with licence holders



**What we could
do in the future
with available
data**



Table 5. Collective dose (man mSv) in Ireland by dose category (1996–2005).

Sector/dose interval	0–1 mSv	1–5 mSv	5–10 mSv	10–20 mSv	> 20 mSv
Medical	625.7	181.5	20.3	15.9	—
Industrial	262.1	349.7	92.4	17.3	37.7
Education/research	30	6.3	—	—	—

Breakdown of annual doses by job category for all of Canada

Job Category	Distribution of workers over dose intervals							Number of Workers	Avg. Dose (mSv)	Avg. of Positive Doses
	0 mSv	>0-1 mSv	>1-2 mSv	>2-5 mSv	>5-20 mSv	>20-50 mSv	>50 mSv			
Shared among all sectors										
Administrator	473	195	3	1	0	0	0	672	0.11	0.36
Office staff	2942	375	17	8	0	0	0	3342	0.04	0.36
Safety officer	303	73	2	4	1	0	0	383	0.10	0.48
Student	14430	842	55	38	8	1	0	15374	0.03	0.49
Industry and Research										
Aircrew	11	2	3	0	0	0	0	16	0.26	0.85
Ground transportation	41	37	6	10	3	0	0	97	0.74	1.29
Industrial radiographer	1207	517	250	446	422	6	0	2848	2.06	3.58
Instructor (non-medical)	306	35	0	1	0	0	0	342	0.04	0.38
Instrument technician	1490	543	34	29	8	0	0	2104	0.17	0.58
Laboratory technician (industrial)	1968	649	69	59	12	0	0	2757	0.21	0.73
Nuclear fuel processor	229	343	86	71	27	0	0	756	0.86	1.24
Scientist/Engineer (field)	785	668	80	35	11	0	0	1579	0.34	0.68

Table 4. Summary dose data for occupational radiation exposure in Ireland (1996-2000) and (2001–05).

Sector/year	1996–2000			2001–05		
	Collective dose (man mSv)	Workers receiving doses	Average dose ^a (mSv)	Collective dose (man mSv)	Workers receiving doses	Average dose ^a (mSv)
Medical	544.7	1915	0.28	298.7	924	0.32
Industrial	404.9	575	0.70	354.3	448	0.79
Education/research	26.1	135	0.20	10.2	43	0.24
All Sectors	975.7	2625	0.37	663.2	1,415	0.47

^aPer worker receiving a measurable dose.

Number of workers (top) and average whole body dose in mSv (bottom) by job category and province/territory

Job Sector and Category	Nfld.	P.E.I.	N.S.	N.B.	Que.	Ont.	Man.	Sask.	Alta.	B.C.	N.W.T.	Yukon	Canada
Tradesmen	0	0	19	1	18	81	0	9	61	4	0	0	193
	-	-	0.00	9.19	0.00	0.22	-	0.00	0.19	0.00	-	-	0.20
Well logger	5	0	0	10	3	68	0	13	2584	5	0	0	2688
	0.05	-	-	0.02	0.86	0.39	-	0.34	0.40	0.19	-	-	0.40
Sector totals	247	14	342	379	3133	6687	319	690	5861	913	7	2	18480
	0.17	0.68	0.16	0.44	0.13	0.40	0.06	0.36	1.27	0.62	0.29	0.00	0.63

Medicine

Chiropractor	1	0	3	3	651	237	76	3	127	16	0	0	1117
	0.00	-	0.00	0.25	0.02	0.03	0.01	0.00	0.04	0.24	-	-	0.02
Dental assistant	183	61	315	231	3133	5045	868	324	3346	594	18	3	14083
	0.04	0.00	0.02	0.01	0.00	0.01	0.00	0.01	0.01	0.01	0.00	0.00	0.01

Table 6. Average annual collective dose (man mSv) in Ireland by work practice (2001–05).

Work practice	Annual collective dose (man mSv)
Veterinary radiology	4.9
Dental radiology	3.1
Diagnostic radiology	38.2
Radiotherapy	2.5
Nuclear medicine	9.3
Other medical	1.8
Industrial radiography—fixed	0.5
Industrial radiography—site	44.6
Transport	5.7
Other industrial	20.1
Education/research	2.0
Total	132.7

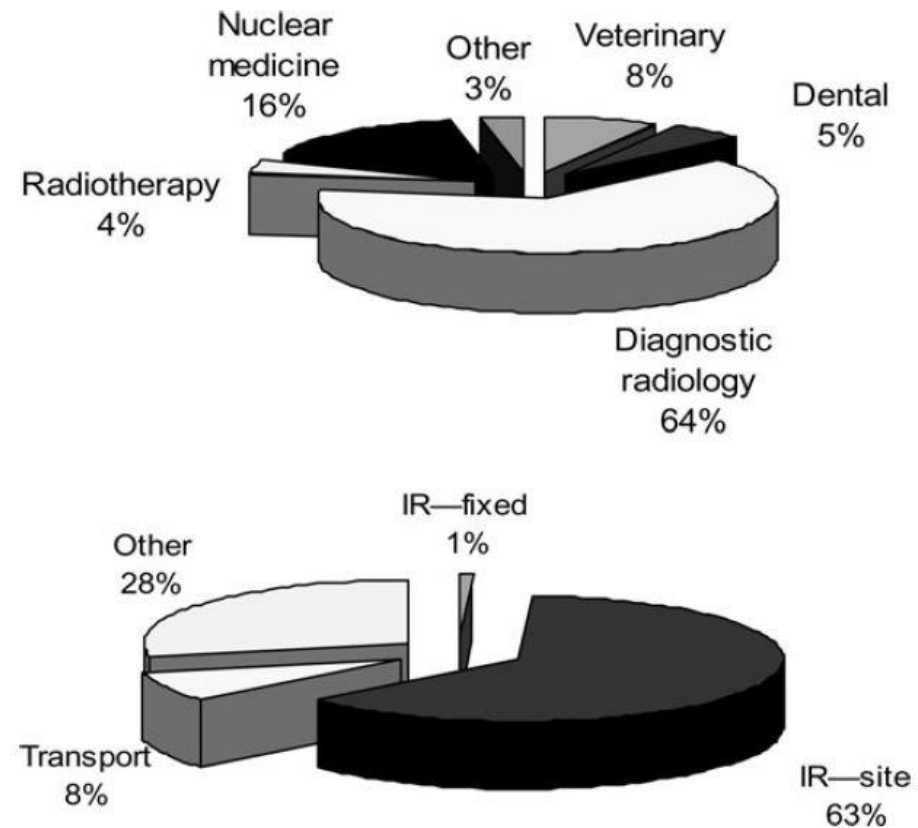


Figure 7. Distribution of annual collective doses in Ireland for the medical and industrial sectors (2001–05).

Dose distribution broken down by job sector, age and sex

Job Sector	Age	Statistic	Sex		Overall
			Male	Female	
Industry and Research	45-54	Number of Workers	3487	825	4312
		Average dose (mSv)	0.52	0.13	0.45
		% tritium	0.31	1.74	0.39
		% radon progeny	0.00	0.00	0.00
	55 and up	Number of Workers	1730	255	1985
		Average dose (mSv)	0.33	0.21	0.32
		% tritium	0.35	0.00	0.32
		% radon progeny	0.00	0.00	0.00
	Total	Number of Workers	14539	3941	18480
		Average dose (mSv)	0.75	0.18	0.63
		% tritium	0.19	1.28	0.25
		% radon progeny	0.00	0.00	0.00
Medicine	Below 25	Number of Workers	392	5512	5904
		Average dose (mSv)	0.22	0.06	0.07
		% tritium	0.00	0.00	0.00
		% radon progeny	0.00	0.00	0.00
	25-34	Number of Workers	3066	16652	19718
		Average dose (mSv)	0.18	0.07	0.09
		% tritium	0.43	0.00	0.13

Mann-Kendall Test for Trend

The purpose of the Mann-Kendall (MK) test is to statistically assess if there is an upward or downward trend of the variable of interest over time and whether the trend is statistically significant. Does not test for magnitude.

An upward (downward) trend means that the variable consistently increases (decreases) through time.

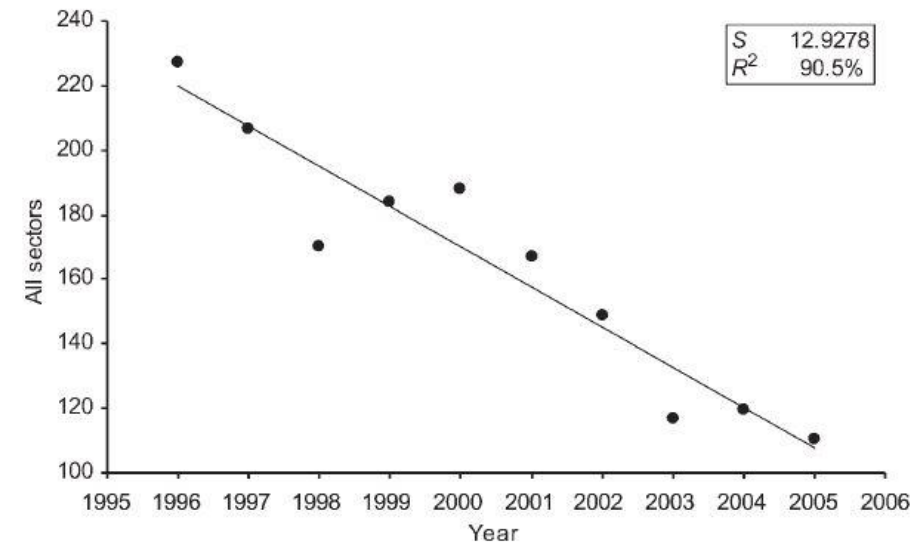


Figure 4. Mann–Kendall test of statistical significance of downward trend in collective doses for all sectors (1996–2005).

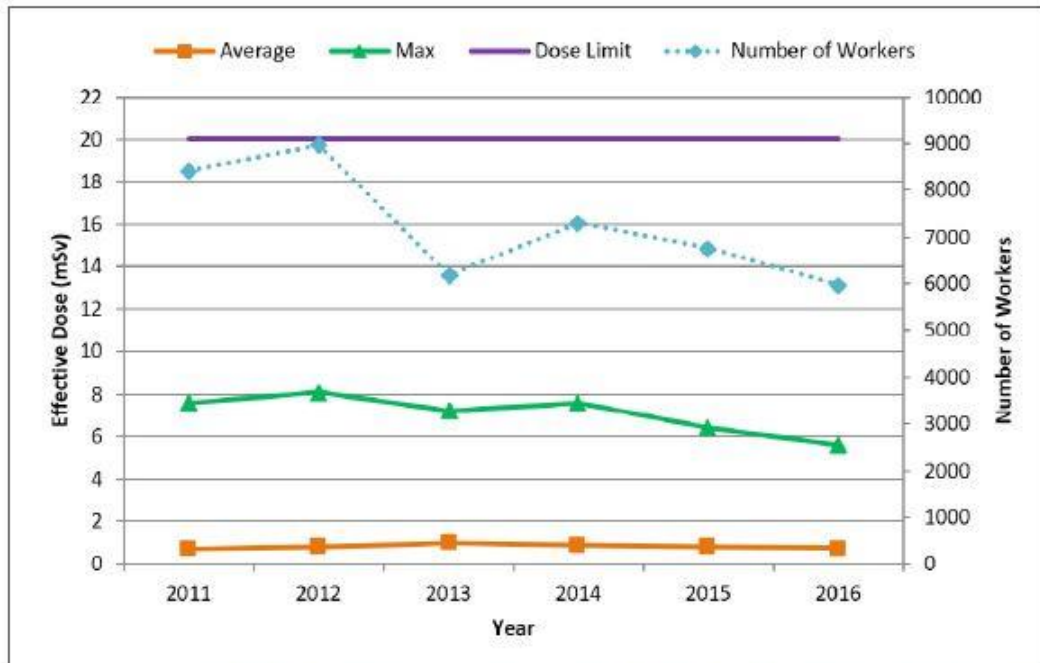


Figure 3: Uranium industry average and maximum effective doses with workforce numbers (2011–2016)

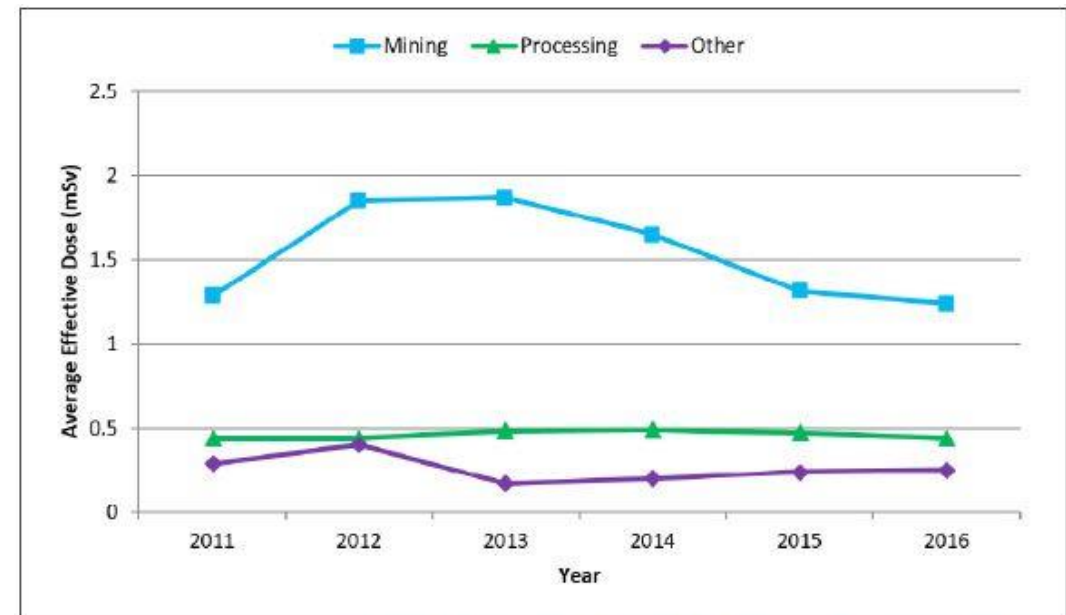


Figure 4: Uranium industry average effective doses by worker categories (2011–2016)

Table 2 Standardised mortality ratios (SMRs)^a by broad cause, gender and industrial classification

Gender	Industrial classification	Number of deaths		Unadjusted		Social-class adjusted	
		Observed	Expected	SMR	95% CI ^b	SMR	95% CI
All causes							
Both	All	26731	33014.00	81	80–82	84	83–85
	Industrial	18285	19660.37	93	92–94	82	80–83
	Non-industrial	8146	12950.43	63	62–64	90	88–92
Men	All	25841	31852.94	81	80–82	84	83–85
Women	All	890	1161.06	77	72–82	84	79–90
χ^2 for heterogeneity in SMR between men and women				2.77		0.01	
All malignant neoplasms							
Both	All	8107	9666.63	84	82–86	82	81–84
	Industrial	5394	5640.72	96	93–98	80	78–82
	Non-industrial	2622	3905.75	67	65–70	88	85–91
Men	All	7752	9229.92	84	82–86	82	80–84
Women	All	355	436.71	81	73–90	84	76–93
χ^2 for heterogeneity in SMR between men and women				0.36		0.17	

^aBased on the general population of England and Wales. ^bConfidence interval.

Standardised Mortality Ratio (SMR)

SMR quantifies the increase or decrease in mortality of a study cohort with respect to the general population

SMR is a ratio between the observed number of deaths in a study population and the number of deaths that would be expected in the general population

A ratio >1 indicates that more mortality has occurred than expected, a ratio <1 indicates less than expected

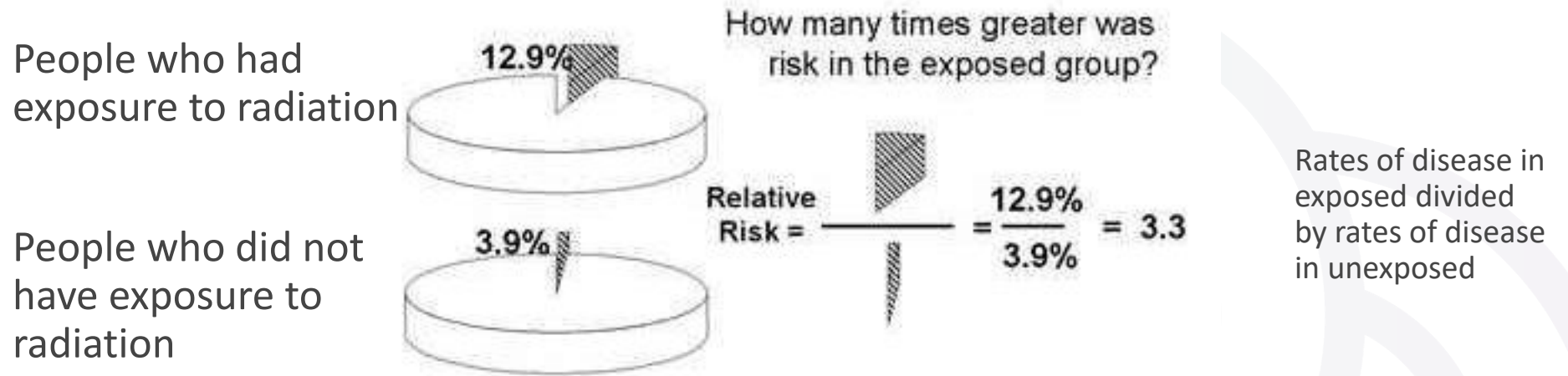
Example of SMR interpretation

$$26731/33014 = 0.81$$

0.81<1 therefore there were 19% less deaths than expected

Relative Risk (RR)

RR is often used when the study involves comparing the chance of an event occurring between two groups



The risk of the disease was **3.3 times greater** in the people who had exposure to radiation compared to those who did not have exposure to radiation

RR=1 means exposure has no effect on risk

Excess Relative Risk (ERR)

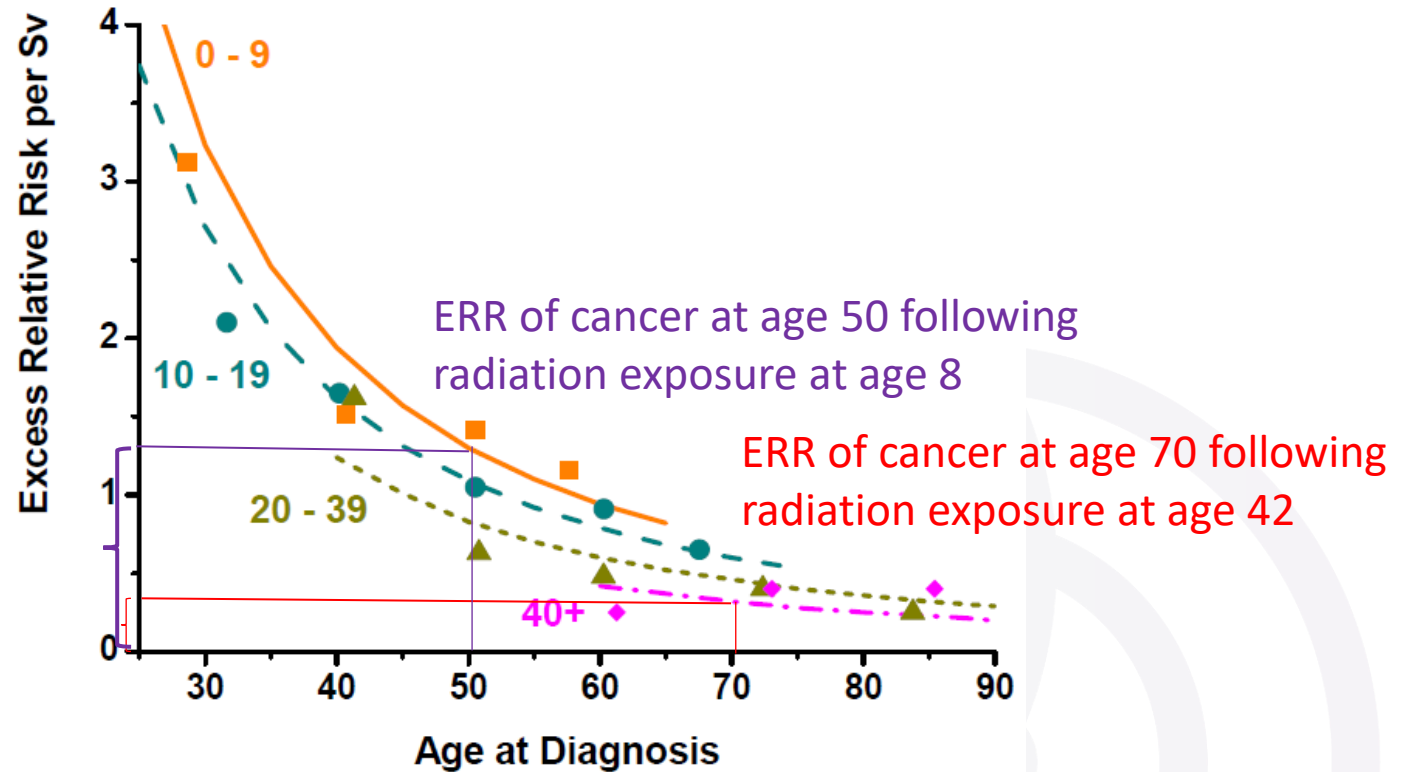
ERR is the **proportional increase in risk** following exposure.

$$\text{ERR} = \text{RR} - 1$$

ERR assumes that if a dose of radiation is administered to a population, then after some latent period there is an increase in cancer rate.

The additional risk of a disease that can be associated with an exposure.

The fraction by which the risk for an exposed person exceeds that of a person who was not exposed.



In summary, in the future we could

- Report information by effective dose distribution, year, age categories
- Use pie charts and frequency tables (sometimes tables are better than graphs)
- Assess trends using statistical tests (e.g. Mann-Kendall)
- See the effect of sex, age, industry type, etc. (one at a time or all together) on average effective doses
- Link this data with other national datasets, e.g. Cancer Council information and perform advanced statistical tests

Thank you

ANRDR@arpansa.gov.au

arpansa.gov.au



ARPANSAGovernment



ARPANSA



ARPANSANews



ARPANSA