SUPPLEMENTAL MATERIAL

Data curation

First, we corrected issues like mislabeled task numbers and patient IDs. We removed duplicate submissions where every field of the spreadsheet was the same. To avoid biasing participants between Task 1 and planar dosimetry of Task 2, we provided a sensitivity value for the camera that was double of the true one (participants were aware of a scaling factor but did not know the exact value). Affected values (e.g., dose, but not volume) for Task 2 were divided by 2 to undo this scaling. The different data columns were converted to their corresponding element type (e.g. float or string). Errors and inconsistencies in the conversion were monitored and acted as quality control for the curation of the data. Differences in regional formatting (e.g., use of commas or periods to represent the decimal) were harmonized.

Since the focus was variation in absorbed dose, we first focused on AD values with large deviations from the mean. This led to the discovery of likely errors such as reporting AD in units different than Gy. Volumes of kidneys and lesions were assessed and indicated that some values had been reported in the incorrect cells (e.g. right and left kidneys). This was straightforward for tasks 4 and 5 where VOIs were provided, but was also possible for other tasks given the substantially different relative sizes of these objects. These issues were corrected. By analyzing the volumes of VOIs, we also identified submissions for tasks 4 and 5 that did not use the provided VOIs; these were therefore excluded.

When absorbed dose was reported for the combined kidneys, we verified that the volume of the VOI of the total kidneys corresponded to the sum of volume of the individual left and right kidneys. Similarly, we checked if the total kidney AD was mistakenly reported as the sum of the individual kidney ADs. Data was excluded in these cases as it represents a fundamental error and not an error in method or judgement.

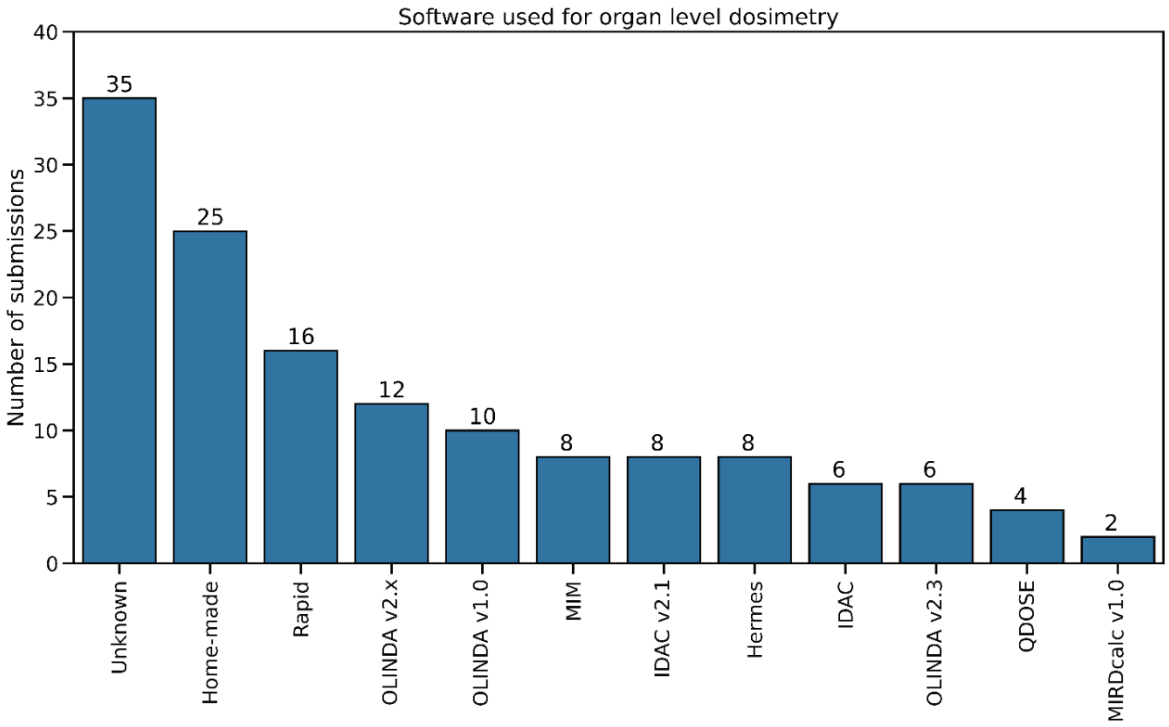
To identify outliers, we created box plots. This identified some values that seemed very likely to be errors, e.g., liver ADs above 100 Gy. We carefully scrutinized submissions containing outliers, including reading the methodology summary of dosimetry methods. When the outliers could not be traced to an obvious mistake, the submitters were contacted. In many cases, an error in reporting (e.g., units, etc.), was identified and corrected. Finally, data points more than 2 standard deviations from the mean were examined and submitters contacted to correct for possible mistakes or systematic deviations. We did not include submissions for which large outliers could not be resolved in subsequent analysis because we believe they represent errors in conception or implementation of the calculations and not differences in judgement, methods or tools.

As an example of the process of scrutinizing outliers, we identified a systematic deviation in results calculated with one software package (Olinda/EXM, Hermes Medical Solutions, Sweden). We traced this to an error in mass correction using some versions of the software (versions before 2.2.3 as verified by the vendor) and when using the ICRP-89 based phantoms. We excluded all submissions (n=12) that were affected by that error.

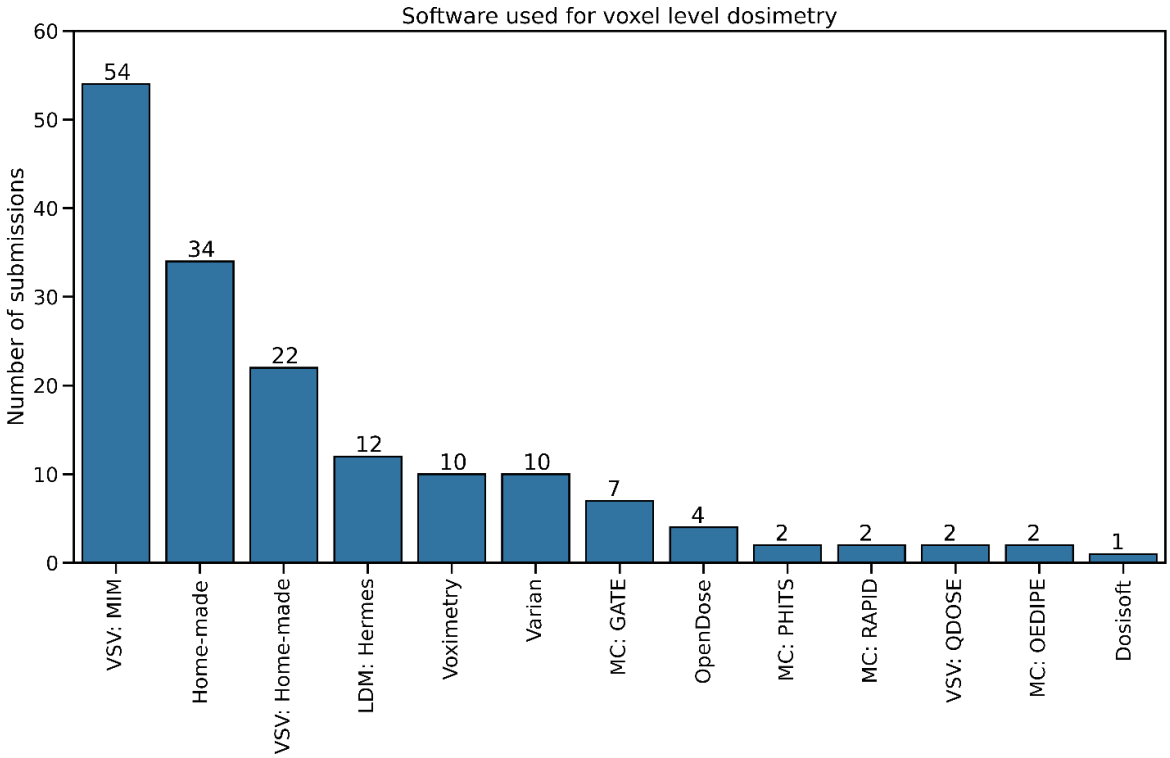
Some kinds of variations were deemed not relevant to the goals of the challenge. For example, errors in completing the spreadsheet (e.g., incorrect units, switching left and right kidneys, errors in specific versions of a software package) were considered not to represent fundamental variability in method, tools, or practice and were thus corrected, if possible, or excluded from the variability analysis.

General observations

Supplemental Figures 1 and 2 provide an overview of the software used for organ-level and voxel-level dosimetry.



Supplemental Figure 1: Organ level dosimetry software used by the challenge participants.



Supplemental Figure 2: Voxel level dosimetry software used by the challenge participants. VSV refers to voxel S value methods, while MC stands for Monte Carlo simulation based methods.

Supplemental Table 1 provides the mean and standard deviation of healthy organ and lesion AD values per task.

Supplemental Tables 1: Descriptive statistics of organ and lesion absorbed dose results across tasks and patient.

Task	Patient	Statistics	Liver	Spleen	R Kidney	L Kidney	Total Kidney
2	A	Mean ± SD (Gy)	0.75 ± 0.74	1.60 ± 1.45	0.96 ± 0.91	1.08 ± 1.20	1.73 ± 1.52
		Min (Gy)	0.09	0.17	0.20	0.21	0.24
		25% (Gy)	0.17	0.37	0.30	0.31	0.45
		50% (Gy)	0.42	1.05	0.63	0.56	1.52
		75% (Gy)	1.08	2.82	1.34	1.29	2.36
		Max (Gy)	2.15	4.58	2.88	3.59	4.25
		COV (%)	98.42	90.57	95.25	110.99	87.70
		QCD (%)	72.62	76.96	63.73	61.00	68.28
	B	Mean ± SD (Gy)	0.87 ± 0.83		2.68 ± 1.50	2.69 ± 1.59	2.63 ± 1.35
		Min (Gy)	0.15		1.37	0.46	0.90
		25% (Gy)	0.23		1.56	2.04	1.65
		50% (Gy)	0.63		1.92	2.55	2.68
		75% (Gy)	1.22		3.84	3.97	3.80
		Max (Gy)	2.41		4.72	4.43	4.04
		COV (%)	94.46		56.07	59.19	51.18
		QCD (%)	68.56		42.20	32.27	39.52

Task	Patient	Statistics	Liver	Spleen	R Kidney	L Kidney	Total Kidney	Lesion 1	Lesion 2	Lesion 3	Lesion 4
3	A	Mean ± SD (Gy)	1.95 ± 1.25	4.09 ± 3.66	2.87 ± 2.37	3.04 ± 1.56	2.56 ± 0.61	25.72 ± 14.40	34.43 ± 30.64		
		Min (Gy)	0.16	0.34	0.41	0.41	1.62	5.60	0.26		
		25% (Gy)	1.24	2.61	2.06	2.60	2.18	17.45	13.52		
		50% (Gy)	1.67	3.04	2.30	2.95	2.79	24.46	25.42		
		75% (Gy)	2.48	4.03	2.86	3.42	3.02	26.99	44.68		
		Max (Gy)	4.71	15.70	9.87	5.81	3.10	64.06	109.60		
		COV (%)	64.09	89.57	82.50	51.43	23.94	55.99	88.97		
		QCD (%)	33.10	21.35	16.33	13.62	16.16	21.45	53.54		
	B	Mean ± SD (Gy)	2.01 ± 1.96		4.65 ± 2.18	3.41 ± 1.13	4.17 ± 1.20	7.26 ± 12.90	26.08 ± 21.42	12.29 ± 22.58	6.09 ± 6.42
		Min (Gy)	0.20		0.53	1.43	2.95	0.99	5.43	1.56	0.02
		25% (Gy)	1.27		3.78	2.85	3.23	2.10	11.70	2.91	2.28
		50% (Gy)	1.46		4.50	3.20	3.84	3.49	24.05	3.79	5.14
		75% (Gy)	1.68		5.24	3.99	5.19	4.71	26.94	5.15	6.75
		Max (Gy)	8.19		9.59	5.48	5.67	45.75	81.82	83.92	24.31
		COV (%)	97.60		46.91	33.14	28.82	177.60	82.15	183.72	105.40
		QCD (%)	13.90		16.13	16.61	23.35	38.46	38.45	27.81	49.52

Task	Patient	Statistics	Liver	Spleen	R Kidney	L Kidney	Total Kidney	Lesion 1	Lesion 2	Lesion 3	Lesion 4
1	A	Mean ± SD (Gy)	1.77 ± 0.70	4.09 ± 1.64	3.11 ± 1.03	3.25 ± 1.00	3.19 ± 1.21	31.81 ± 16.44	38.31 ± 29.63		
		Min (Gy)	0.44	0.72	0.59	0.57	1.78	5.29	3.40		
		25% (Gy)	1.37	3.08	2.45	2.56	2.52	23.38	25.14		
		50% (Gy)	1.63	3.59	2.76	3.00	2.68	27.11	33.21		
		75% (Gy)	2.28	4.93	3.71	3.76	3.36	36.87	46.58		
		Max (Gy)	3.39	11.70	7.78	5.84	8.29	102.00	218.00		
		COV (%)	39.78	40.11	32.98	30.91	37.81	51.68	77.34		
		QCD (%)	25.00	23.14	20.53	19.02	14.29	22.40	29.90		
	B	Mean ± SD (Gy)	1.62 ± 0.48		5.21 ± 1.49	3.86 ± 1.31	4.61 ± 1.33	3.38 ± 0.99	24.59 ± 13.65	3.44 ± 1.14	4.85 ± 1.80
		Min (Gy)	0.39		1.02	0.74	1.95	0.72	1.76	0.77	0.87
		25% (Gy)	1.39		4.61	3.26	4.00	3.09	14.28	2.97	3.60
		50% (Gy)	1.56		5.04	3.70	4.44	3.30	25.72	3.17	4.84
		75% (Gy)	1.79		5.96	4.30	5.43	3.72	32.85	3.85	5.78
		Max (Gy)	3.43		8.47	8.11	7.54	6.60	81.10	7.00	8.47
		COV (%)	29.64		28.66	33.79	28.94	29.35	55.51	33.07	37.02
		QCD (%)	12.54		12.77	13.82	15.11	9.27	39.41	12.83	23.22

Task	Patient	Statistics	Liver	Spleen	R Kidney	L Kidney	Total Kidney	Lesion 1	Lesion 2	Lesion 3	Lesion 4
4	A	Mean ± SD (Gy)	2.12 ± 0.88	3.94 ± 1.23	2.88 ± 0.76	3.10 ± 0.88	3.04 ± 0.69	23.95 ± 12.60	45.31 ± 38.09		
		Min (Gy)	0.33	0.85	0.73	0.75	2.31	5.17	11.34		
		25% (Gy)	1.42	3.14	2.56	2.60	2.61	19.12	29.95		
		50% (Gy)	2.10	3.50	2.67	2.80	2.75	21.00	40.09		
		75% (Gy)	2.61	5.21	3.45	3.70	3.27	25.06	47.06		
		Max (Gy)	4.51	6.50	4.40	4.87	4.60	72.90	226.00		
		COV (%)	41.36	31.14	26.43	28.57	22.82	52.61	84.06		
		QCD (%)	29.53	24.82	14.82	17.51	11.22	13.46	22.22		
	B	Mean ± SD (Gy)	1.86 ± 0.84		5.08 ± 1.20	3.67 ± 1.02	4.33 ± 1.03	3.36 ± 1.01	36.60 ± 19.99	3.12 ± 0.74	4.31 ± 1.83
		Min (Gy)	0.42		1.32	1.09	1.82	0.89	5.41	0.84	1.14
		25% (Gy)	1.58		4.49	3.35	3.78	2.80	25.13	2.86	3.48
		50% (Gy)	1.65		5.09	3.60	4.60	3.27	35.93	3.17	4.14
		75% (Gy)	1.87		5.59	4.26	4.98	3.60	42.09	3.44	4.50
		Max (Gy)	5.02		7.90	5.97	5.67	6.73	95.08	5.60	12.53
		COV (%)	45.04		23.53	27.80	23.73	29.89	54.60	23.67	42.57
		QCD (%)	8.57		10.94	12.03	13.70	12.50	25.23	9.15	12.81

Task	Patient	Statistics	Liver	Spleen	R Kidney	L Kidney	Total Kidne	Lesion 1	Lesion 2	Lesion 3	Lesion 4
5	A	Mean ± SD (Gy)	2.03 ± 0.86	3.43 ± 1.03	2.65 ± 0.83	2.82 ± 0.89	2.64 ± 0.24	21.43 ± 5.56	31.62 ± 11.57		
		Min (Gy)	0.97	2.80	1.59	1.54	2.29	17.93	17.42		
		25% (Gy)	1.27	2.97	2.31	2.48	2.51	19.50	28.06		
		50% (Gy)	2.25	3.16	2.50	2.68	2.62	20.38	29.93		
		75% (Gy)	2.45	3.30	2.57	2.75	2.69	21.21	30.48		
		Max (Gy)	4.78	7.59	5.83	6.18	3.14	47.21	83.20		
		COV (%)	42.48	29.88	31.38	31.54	9.15	25.96	36.60		
		QCD (%)	31.66	5.33	5.33	5.08	3.37	4.19	4.13		
	B	Mean ± SD (Gy)	1.55 ± 0.33		4.57 ± 0.96	3.50 ± 0.59	4.30 ± 0.23	3.13 ± 0.28	28.42 ± 7.47	2.98 ± 0.28	3.98 ± 0.79
		Min (Gy)	1.12		0.81	2.00	3.91	2.82	2.77	2.65	0.59
		25% (Gy)	1.38		4.40	3.43	4.11	2.92	27.90	2.76	3.88
		50% (Gy)	1.51		4.78	3.67	4.34	3.12	29.92	2.96	4.08
		75% (Gy)	1.60		4.90	3.79	4.48	3.20	31.20	3.06	4.28
		Max (Gy)	2.84		5.97	4.68	4.60	4.13	40.30	3.91	5.36
		COV (%)	21.43		20.96	16.74	5.40	8.87	26.28	9.38	19.73
		QCD (%)	7.32		5.33	5.03	4.28	4.57	5.58	5.09	4.93