Recommendation for the Frequency of Calibration Audits by a Quality Assurance Center Supporting NCI sponsored Clinical Trials.

This document represents work in progress within the Advanced Technology QA Consortium (ATC). Comments are welcome and should be directed to Marcia Urie, PhD or David Followill, PhD, co-chairs of the ATC Credentialing and QA Committee.

Executive Summary

The focus of this report is to provide data from the Radiological Physics Center (RPC) and the literature from which a consensus opinion could be developed regarding annual reference dosimetry TLD audits. Such audits would be conducted by a recognized quality assurance organization and would be performed by NCI cooperative group members, or their affiliates and collaborators. These institutions are primarily located in the USA and Canada.

After an extensive literature search, this committee was unable to find any peer-reviewed published data assessing the possible resulting impact on clinical trial outcome from a change in frequency (increase or decrease) of reference dosimetry audits. The following report analyzing data from the RPC, however, does provide data that indicate that a reduction in the frequency of TLD audits would likely lead to an increase in the number of institutions with an undetected calibration error. The data provided also indicate that in any given year, photon beam audits failed to meet the 5% agreement criteria at 7-14% of the institutions (3-5% of the photon beams) that enroll the majority of clinical trial patients. Including the impact of the number of patients enrolled onto clinical trials from institutions with TLD discrepancies, the percent of total patients per year potentially affected by photon beams with a TLD audit outside of the 5% criterion range from 5 to 12%.

Published data indicate that dose errors as small as 3% can affect the quality and outcome of a trial unless this is anticipated and the trial is powered appropriately. [3-5] As indicated in the World Health Organization report[1] the majority of radiotherapy errors originate from human error which is the probable cause for the majority of the beam calibration errors since it is the one dosimetry parameter most likely to change with time and subject to human interpretation of the calibration protocols. A reduction in the audit frequency and a corresponding increase in the number of undetected calibration errors may increase the uncertainty in tumor dose of patients entered onto trials. It is the consensus of this committee that uncertainties in the doses delivered to clinical trial patients should be kept small to avoid increasing patient accrual goals. The following recommendations are made:

1. The frequency of the reference dosimetry TLD audit should remain annual for participating instituions within the USA and Canada as is currently the parctice.

2. International institutions (outside the USA and Canada) wishing to participate in NCI sponsored clinical trials will have all of their photon beam calibrations verified using a TLD audit prior to enrolling patients onto the trial. Any new machine put into clinical service at the facility will

also have all of its photon beams audited as well. These audits should be performed at a minimal interval of every two years.

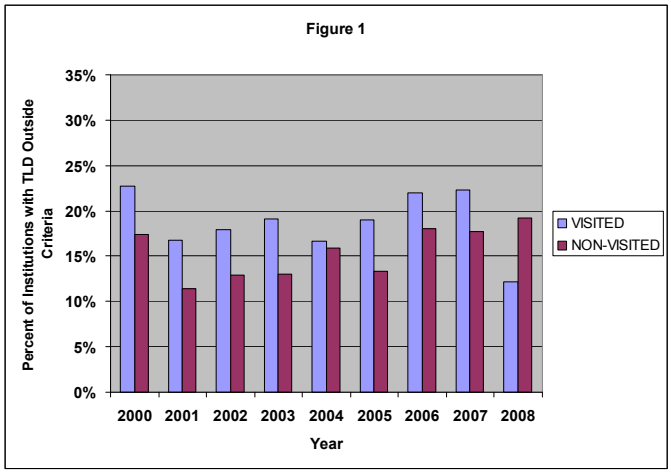
3. The RPC and EORTC will jointly conduct a study to determine the consequences and impact on clinical trial dosimetry of varying the frequency of the TLD audit.

Report

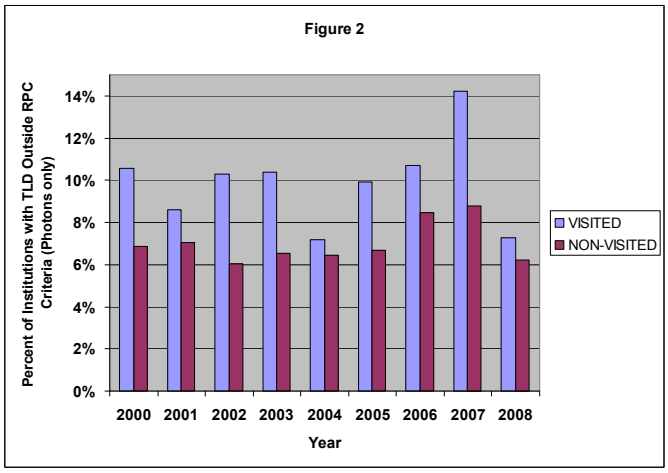
Currently, there are four quality assurance centers throughout the world that have large TLD audit programs to monitor the machine output of megavoltage radiotherapy machines. These groups include the Radiological Physics Center (RPC), International Atomic Energy Agency (IAEA), Radiation Dosimetry Services (RDS) and European Quality Assurance Laboratory/European Society for Therapeutic Radiology and Oncology (EQUAL-ESTRO). Of these four centers, the RPC has the largest TLD program and monitors all of the institutions (1669 institutions) that participate in NCI sponsored clinical trials, both within the USA and internationally. The RPC initiated its TLD program for photon beams in 1977. In 1982 electron beams were included, and in 2007, measurements of proton beams were initiated. (It is noted that the IAEA monitors about the same number of institutions as the RPC, about 1600. However, the IAEA measures far fewer beams per year, and reported measuring only 828 beams in 2006-2007 whereas the RPC measures approximately 13,000 beams annually.)

The aim of the RPC, as funded by the NCI, is to assure NCI of the correctness and consistency of the physical data for radiotherapy patients entered onto clinical trials. A secondary aim is to provide feedback to the participating institutions to correct any errors discovered by the RPC. These aims serve to help minimize the uncertainty in the radiotherapy doses delivered to clinical trial patients. The TLD audit of the machine output has been and continues to be an integral part of the RPC QA program. It is the only QA process that reaches every participating institution on an annual basis and as such is the one audit tool that provides continued assurance that the basic output of each machine used to calculate the tumor dose for each patient is accurate and consistent. In addition to the machine output audit, the TLD audit also serves as a mechanism to gather demographic data from each of the participating institutions such as personnel, therapy machines, treatment planning systems, etc. A third benefit of the TLD audit is that it raises the awareness of the institution to the need for accuracy in their machine calibration and extra attention is given to this need when the TLD are sent. New megavoltage machines of the same make, model, and energy used to treat patients these days are fairly similar in terms of their dosimetry properties; however, the one dosimetric quantity that is unique to each machine and highly dependent on personnel and human intervention is the machine output calibration. Numerous explanations for the initial TLD discrepancies in beam output calibrations have been identified, including errors made during calibration of machines by young or inexperienced personnel, errors in spreadsheets used to calculate the output, allowing the beam output to drift beyond the limits set forth by the institution, setup errors with the TLD, incorrect calibration parameters used, incorrect transfer of the reference calibration to the monthly output check system, problems with machine function, etc. A recent publication[1] by the World Health Organization titled “Radiotherapy Risk Profile” states that based on data from the Nuclear Regulatory Commission, 60% or more of radiotherapy incidents are due to human error. Whatever the reason for the discrepancy, the TLD audit provides a mechanism to identify potential calibration problems, enabling the institution to resolve any discrepancies and ensure that their output is correct and consistent with the other clinical trial participating institutions.

Over the past 8 years approximately 5% of all of the megavoltage beams audited with TLD have fallen outside of the RPC's ±5% dose criteria requiring some action and followup by the RPC staff. Today this would represent approximately ~140 photon and ~550 electron beams from nearly 3200 machines used to treat clinical trial patients. Of the approximately 770 institutions the RPC physicists have visited since its inception to conduct an on-site dosimetry review visit and who contribute ~85% of all clinical trial patients that receive radiotherapy, approximately 15 - 20% of these institutions per year (~150 institutions) (figure 1) have one or more photon or electron beams outside of the RPC's criteria requiring an investigation by the RPC. Of the remaining smaller institutions which contribute very few patients ~15% have a beam outside of the RPC's criteria each year. Very few institutions (≤ 50 out of 150 per year) have unacceptable TLD results in two consecutive years. This is because the RPC investigates the discrepancies and follows up with the institution to make sure the errors have been resolved. Performing less frequent audits will result in more institutions with undiscovered calibration errors for longer periods of time.



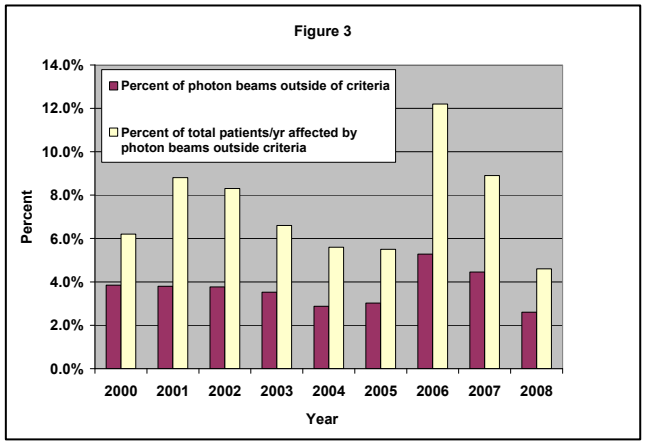
Because the majority of the NCI sponsored clinical trials employ only photon beams, figure 2 illustrates the percent of institutions with TLD results, over the past 8 years, from photons beams only, whereas figure 1 showed the percent of institutions for TLD results from both photon and electron beams. Figure 2 indicates that when only photon beams are considered, the percentage of institutions having a beam outside of the RPC's criteria is reduced from 15–20% to 6–14 %. A further breakdown of these data in terms of photon beams only, and not institutions, show that the percent of photon beams that have the initial TLD outside of the RPC's 5% criterion range from 3 to 5% as seen in figure 3. The larger trial contributors, i.e., the visited institutions that contribute the majority of clinical trial patients, are more likely to have one or more TLD problems. A further breakdown of the TLD results shown in Figures 1 and 2, isolating USA institutions, Canadian institutions and the rest of the international facilities can be seen in the Appendix A of this report in Figures A1-A6. One should note that when considering the percentages for the non-USA institutions in all figures, these values represent small numbers of institutions, but we believe that they are a representative sampling of the international radiation oncology sites that contribute patients to clinical trials.



It is difficult to discern the actual dosimetry errors from the simple mistakes such as setup errors. To the best of our knowledge, the data presented in this report do not include the results from common known irradiation mistakes based on information provided by the institutions. The RPC TLD results

represent potential calibration errors at participating institutions as supported by the fact that the frequency of reference calibration errors detected during the RPC's on-site dosimetry reviews using an ion chamber is 3% for photon beams only as compared to an average of 4% from the TLD audits. The percent of institutions with confirmed machine output calibration errors using an ion chamber is 13% which is close to the percentages (7-11%) from the TLD results (figure 2) over the same time period (2000-2008).

If one takes a closer look at the TLD results for institutions in terms of their patient contributions and limiting the data to photon beams, the percent of total patients per year affected, on average, by the photon beams outside the RPC's TLD criterion ranges from 5 to 12% as seen in Figure 3. These data were derived from knowing the numbers of patients put onto clinical trials by each institution, the number of photon beams at each institution and the fraction of the photon beams outside of the RPC's 5% criterion for the TLD audit. A more detailed description of this analysis can be found in Appendix B. A statistical analysis would be needed to determine whether 5-12% of the patients having dosimetry errors of ≥5% would have an impact on a trial outcome. In addition it is not known how reducing or increasing the frequency of the audit would impact on the outcome of a trial due to the uncertainty in patient dosimetry. Tom Pajak, an RTOG statistician, has stated that if a study involving RT is ever questioned as to the outcome, without proper radiation therapy QA, the results of the study would be less likely to be accepted, or additional patients might be required to improve the power of the study. Studies that include radiation therapy have the luxury of being able to quantify the delivered doses and have a high degree of certainty that the dose delivered is correct because of the QA programs currently in place. The purpose of these QA programs (including the annual TLD audit) is to provide the NCI with the highest quality data for its clinical trial research programs.



Is a TLD audit of the machine output necessary? It is the consensus of this committee that the TLD audit serves to provide assurance that the basic radiotherapy dosimetry at participating institutions is correct, helping to reduce the uncertainty in the delivered tumor doses to patients entered onto clinical trials and allowing the QA centers to focus their efforts on more specific QA or protocol issues. As such, the committee recommends that the annual TLD audit be continued for all clinical trial participants in the USA and Canada as is currently the practice. Numerous errors have been detected at the institutions monitored by the RPC, demonstrating the risk that such errors could go undetected without the regular oversight of the TLD program. An example of this occurred at the Moffit Cancer Center where 77 patients were overdosed by 50% with photons, due to an error in the beam calibration, over a 10 month period.[2] RPC data show that, despite the emphasis on board certification and licensure of medical physicists (in some states), a number of calibration errors continue to occur at institutions participating in clinical trials.

Should the TLD audit be administered annually, biennally or even less frequent? There exists no data to suggest the optimum frequency of the TLD audit in terms of its impact on clinical trial dosimetry. The precedent for performing the TLD audit annually was established by the RPC many years ago and all

current trial data and results are based on having this level of QA. Published data by Bentzen et al[3] reviewing EORTC trial results indicate that, based on their TLD audit program, in cases where the beam calibration was low or high there were decreases in tumor control probability or increases in normal tissue morbidity, respectively, when looking at the clinical dose response data. Bentzen et al[3] also indicated that sequential TLD audits improved the uniformity of the clinical outcome and that small deviations in beam output might lead to clinically important variations in outcome. These same conclusions were reached by Pettersen et al[4] when they discussed the impact of dosimetry quality assurance and its impact on sample size in randomized clinical trials as well as by Boyer and Schultheiss[5] who looked at the effect of dosimetry uncertainty on complication-free local tumor control. It is further recommended by this committee that a study to determine an appropriate frequency of the TLD audit should be conducted. The RPC and EORTC QA office have consulted statisticians and have designed a study to examine the frequency and errors detected by the TLD audit. The proposed study can be found in Appendix C.

It is the consensus of this committee that, until the study to determine the optimum frequency of the TLD audit is completed, international participants such as those in the EORTC, shall perform an initial independent beam calibration TLD audit of all photon energies at the institution prior to enrolling patients onto NCI sponsored clinical trials. In addition any new machine installed at the international participating institution will also have all of its photon beams similarly audited with TLD or similar dosimeter. This audit for international institutions should be conducted every two years as a minimum.

Mailed TLD audits have been and continue to be an integral part of quality assurance for clinical trials. The RPC's annual TLD audit of the machine calibration gives the NCI assurance that clinical trial patient radiation doses are accurate and consistent, reducing the risk that the outcome of a study was influenced by uncertainty in radiation doses as noted by Bentzen et al[3]. When conducting clinical trials it is important to have periodic documentation showing that the participants are qualified to perform their assigned tasks such as delivering the therapeutic dose to the patient in an accurate manner according to the specifications of the protocol.

References

[1] "Radiotherapy Risk Profile", World Health Organization, Geneva, Switzerland, 2008.

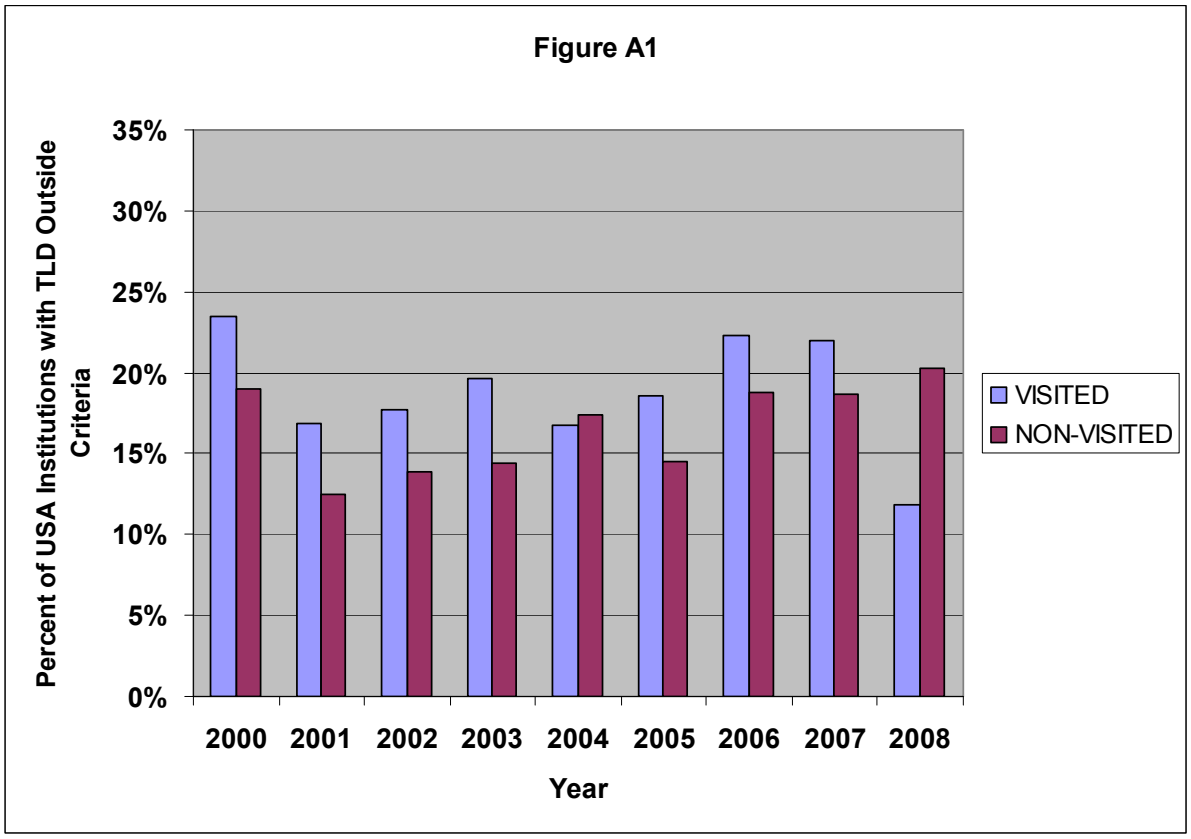
[2] http://www.usatoday.com/news/health/2005-04-02-radiation-overdoses_x.htm, accessed March 18, 2009.

[3] Bentzen et al, Clinical impact of dosimetry quality assurance programmes assessed by radiobiological modeling of data from the thermoluminescent dosimetry study of the European Organization for Research and Treatment of Cancer., European J. Cancer (36), 615-620, 2000.

[4] Pettersen et al, Quality assurance of dosimetry and the impact on sample size in randomized clinical trials., Radiotherapy and Oncology (86), 195-199, 2008.

[5] Boyer, A. and Schultheiss, T., Effects of dosimetric and clinical uncertainty on complication-free local tumor control., Radiotherapy and Oncology (11), 65-71, 1988.

Appendix A

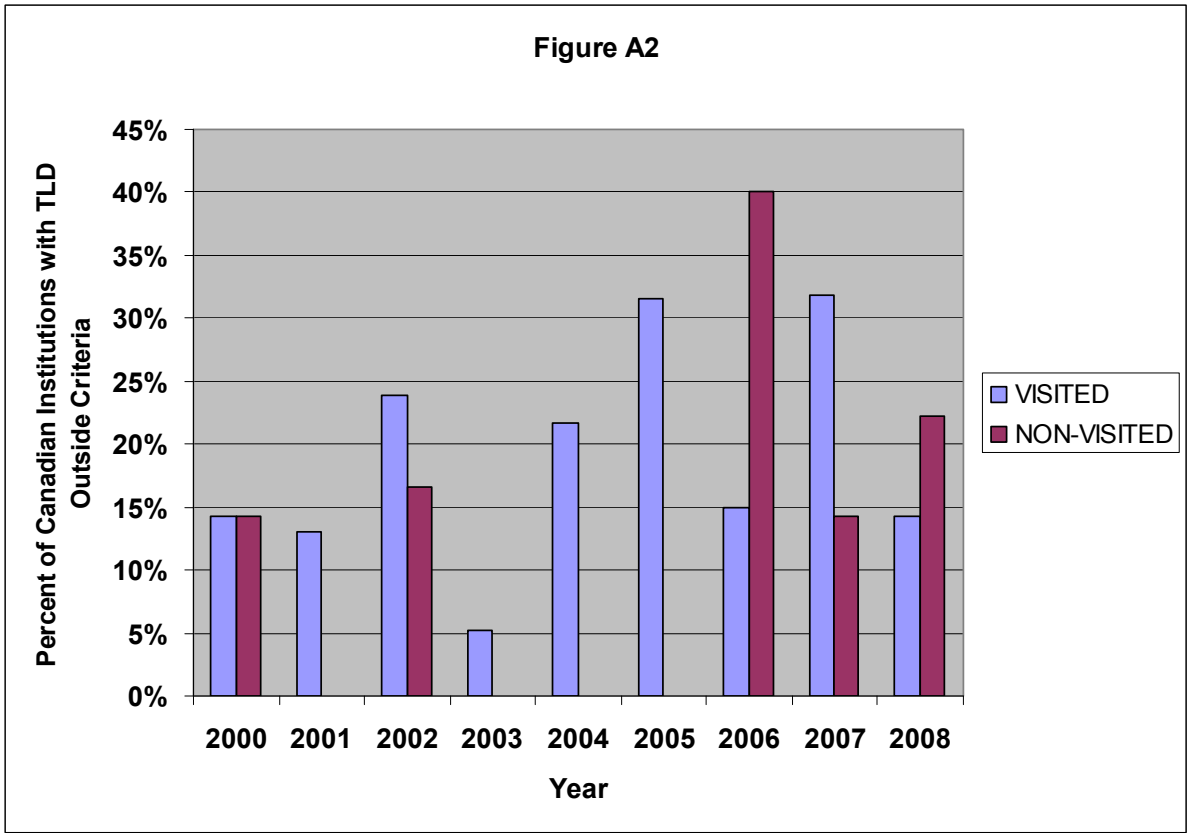


CNT	TOTAL	%	YEAR	STATUS
78	411	19%	2000	NON-VISITED
55	439	13%	2001	NON-VISITED
64	462	14%	2002	NON-VISITED
69	480	14%	2003	NON-VISITED
92	528	17%	2004	NON-VISITED
91	626	15%	2005	NON-VISITED
127	675	19%	2006	NON-VISITED
140	748	19%	2007	NON-VISITED
98	484	20%	2008	NON-VISITED

CNT	TOTAL	%	YEAR	STATUS
146	621	24%		VISITED
106	630	17%		VISITED
115	648	18%		VISITED
130	661	20%		VISITED
113	676	17%		VISITED
128	688	19%		VISITED
156	699	22%		VISITED
158	718	22%		VISITED
81	682	12%		VISITED

YEAR	ACTIVE
2000	1076
2001	1111
2002	1152
2003	1184
2004	1246
2005	1357
2006	1421
2007	1514
2008	1229

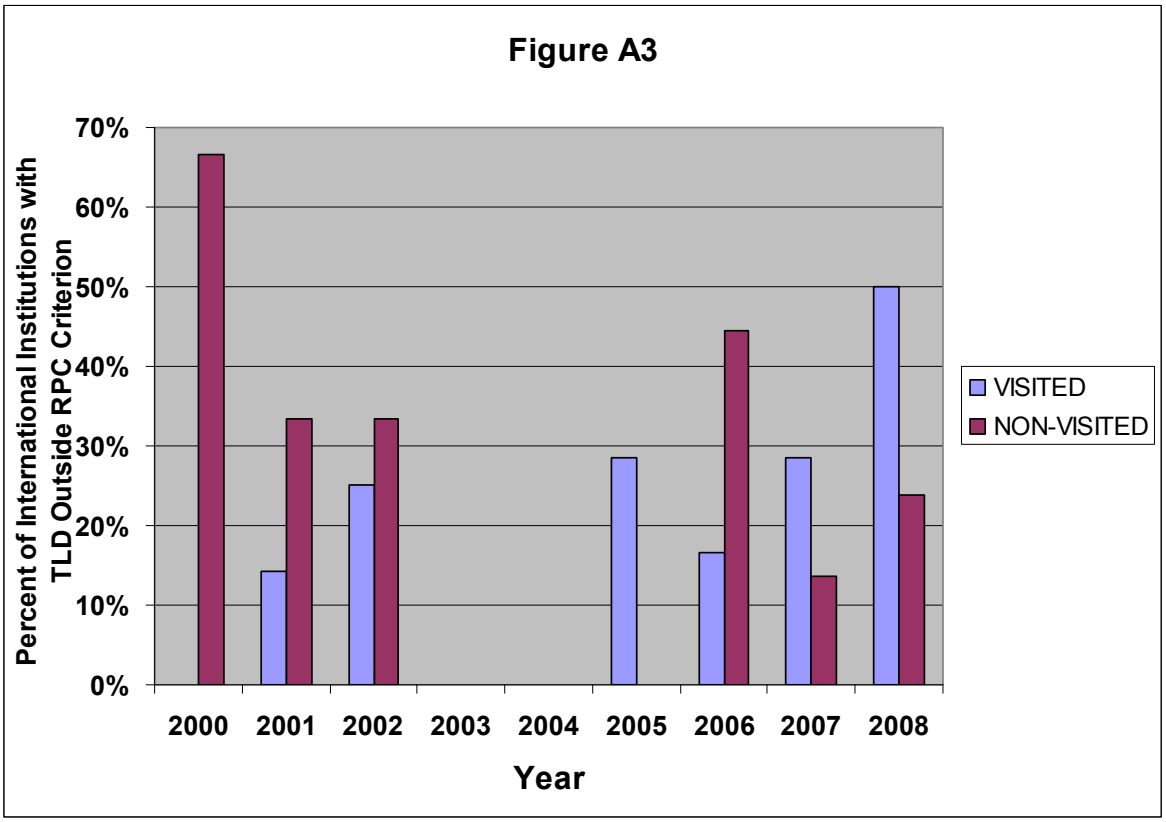
Figure A2



CNT	TOTAL	%	YEAR	STATUS
1	7	14%	2000	NON-VISITED
0	7	0%	2001	NON-VISITED
1	6	17%	2002	NON-VISITED
0	5	0%	2003	NON-VISITED
0	6	0%	2004	NON-VISITED
0	5	0%	2005	NON-VISITED
2	5	40%	2006	NON-VISITED
1	7	14%	2007	NON-VISITED
2	9	22%	2008	NON-VISITED

CNT	TOTAL	%	YEAR	STATUS
3	21	14%		VISITED
3	23	13%		VISITED
5	21	24%		VISITED
1	19	5%		VISITED
5	23	22%		VISITED
6	19	32%		VISITED
3	20	15%		VISITED
7	22	32%		VISITED
3	21	14%		VISITED

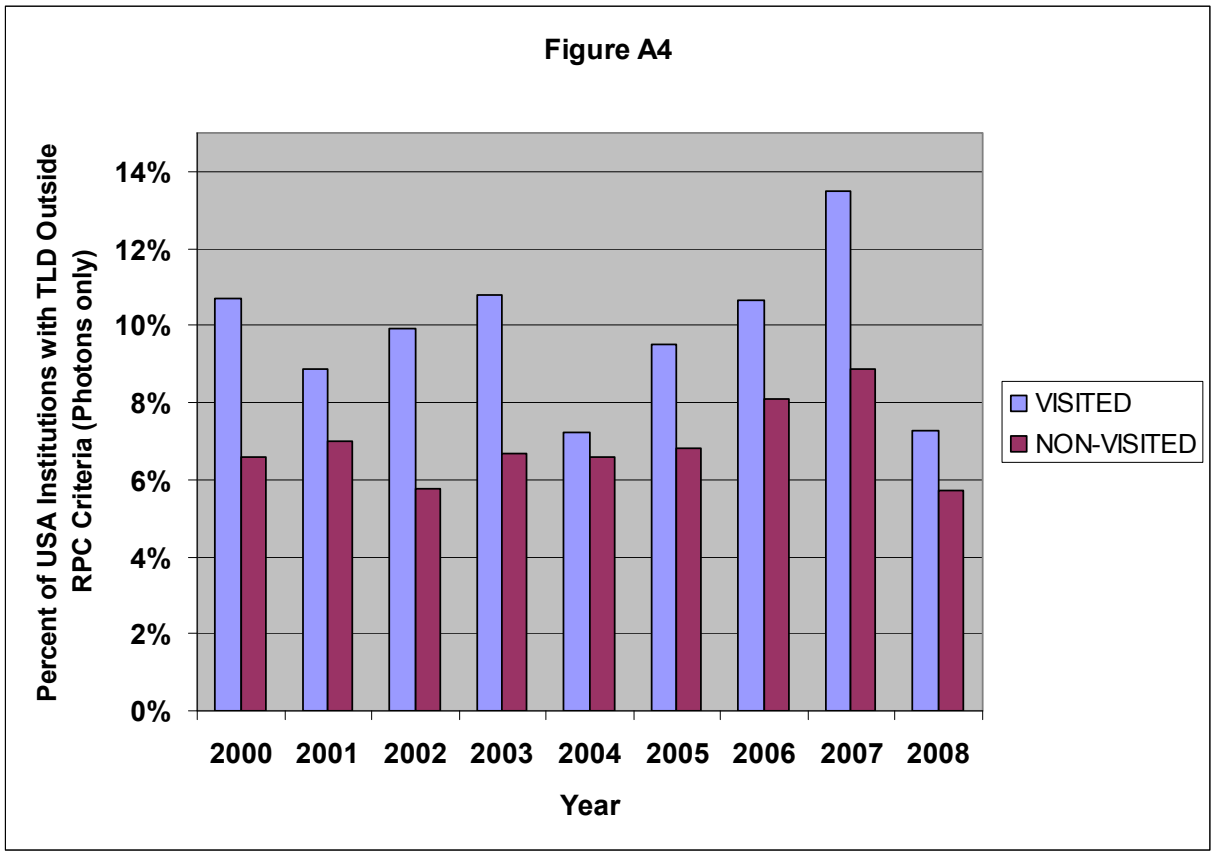
YEAR	ACTIVE
2000	28
2001	30
2002	27
2003	24
2004	29
2005	24
2006	25
2007	29
2008	30



CNT	TOTAL	%	YEAR	STATUS
2	3	67%	2000	NON-VISITED
1	3	33%	2001	NON-VISITED
1	3	33%	2002	NON-VISITED
0	4	0%	2003	NON-VISITED
0	3	0%	2004	NON-VISITED
0	4	0%	2005	NON-VISITED
4	9	44%	2006	NON-VISITED
3	22	14%	2007	NON-VISITED
11	46	24%	2008	NON-VISITED

CNT	TOTAL	%	YEAR	STATUS
0	7	0%		VISITED
1	7	14%		VISITED
2	8	25%		VISITED
0	6	0%		VISITED
0	8	0%		VISITED
2	7	29%		VISITED
1	6	17%		VISITED
2	7	29%		VISITED
3	6	50%		VISITED

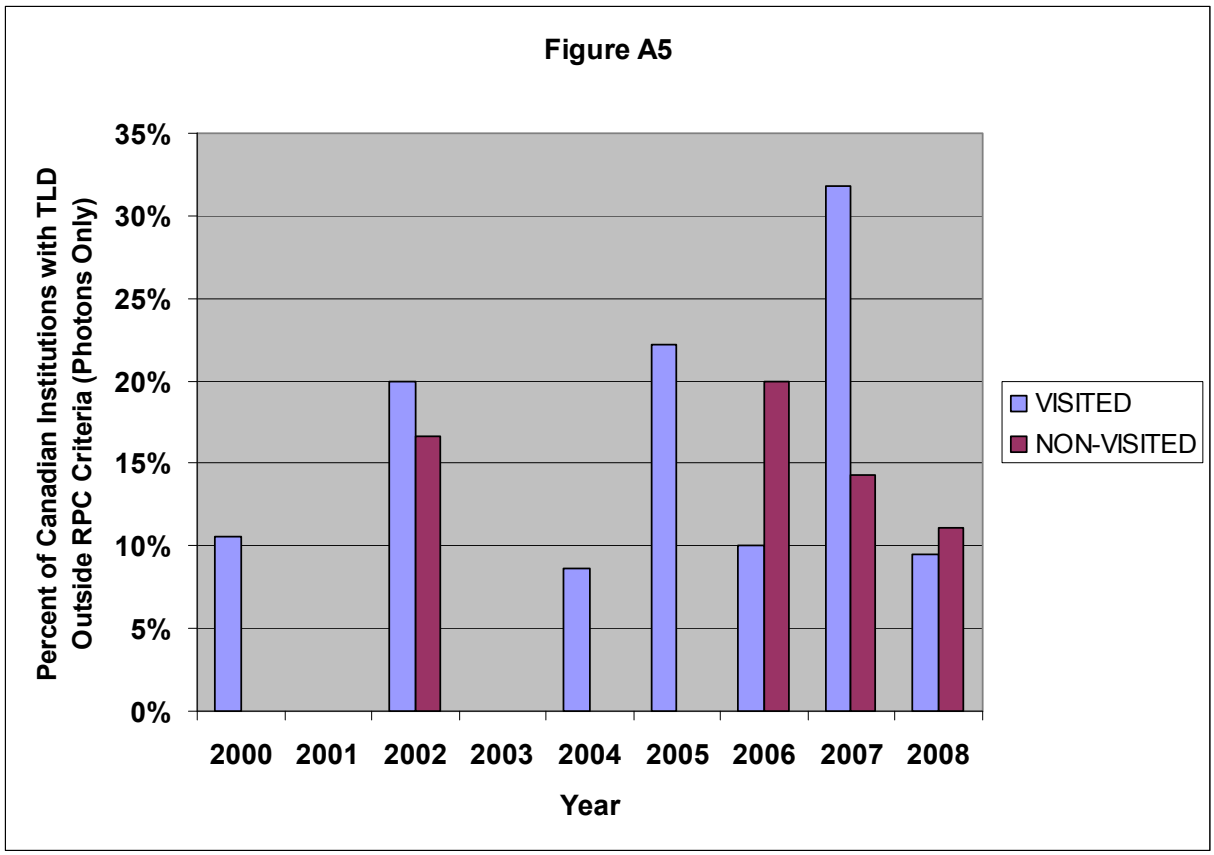
YEAR	ACTIVE
2000	10
2001	10
2002	11
2003	10
2004	11
2005	11
2006	15
2007	29
2008	52



CNT	TOTAL	%	YEAR	STATUS
27	411	7%	2000	NON-VISITED
29	415	7%	2001	NON-VISITED
27	468	6%	2002	NON-VISITED
33	495	7%	2003	NON-VISITED
37	560	7%	2004	NON-VISITED
39	574	7%	2005	NON-VISITED
57	705	8%	2006	NON-VISITED
69	777	9%	2007	NON-VISITED
41	716	6%	2008	NON-VISITED

CNT	TOTAL	%	YEAR	STATUS
65	607	11%		VISITED
48	540	9%		VISITED
61	614	10%		VISITED
67	621	11%		VISITED
46	638	7%		VISITED
56	589	10%		VISITED
69	647	11%		VISITED
90	668	13%		VISITED
42	578	7%		VISITED

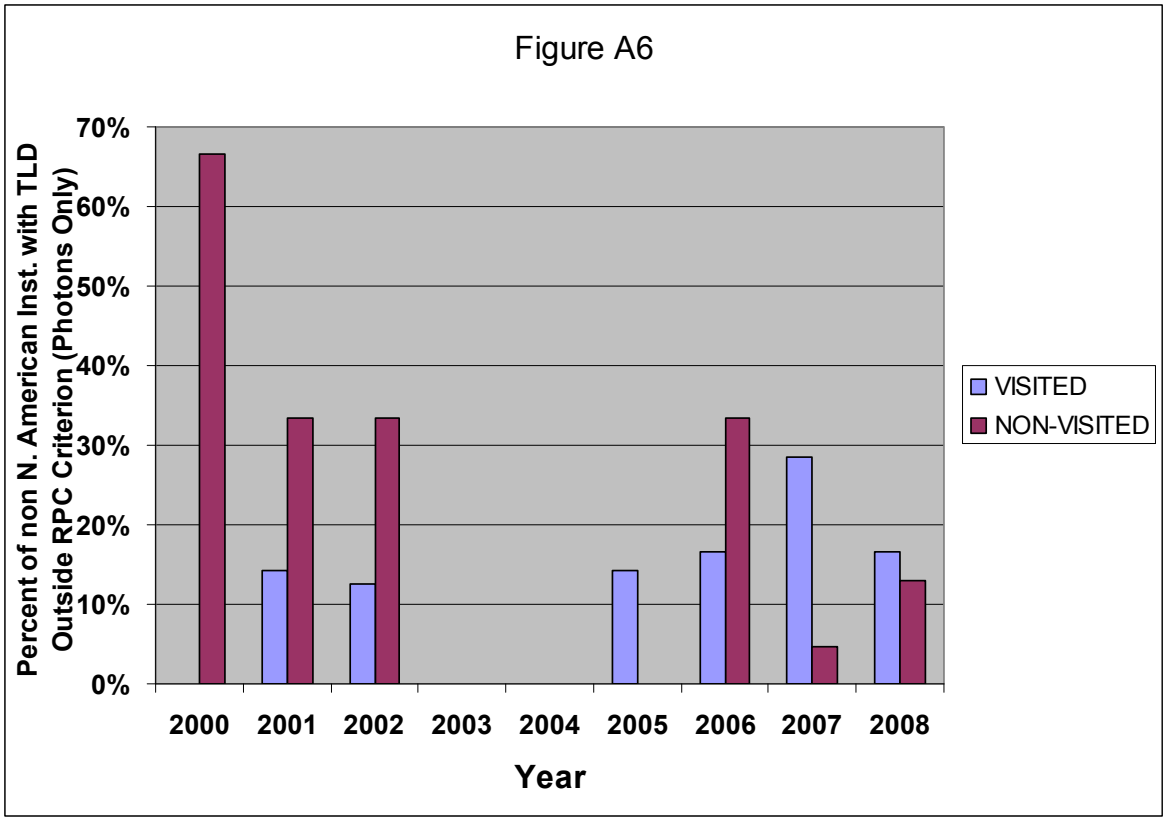
YEAR	ACTIVE
2000	1018
2001	955
2002	1082
2003	1116
2004	1198
2005	1163
2006	1352
2007	1445
2008	1292



CNT	TOTAL	%	YEAR	STATUS
0	7	0%	2000	NON-VISITED
0	7	0%	2001	NON-VISITED
1	6	17%	2002	NON-VISITED
0	5	0%	2003	NON-VISITED
0	6	0%	2004	NON-VISITED
0	5	0%	2005	NON-VISITED
1	5	20%	2006	NON-VISITED
1	7	14%	2007	NON-VISITED
1	9	11%	2008	NON-VISITED

CNT	TOTAL	%	YEAR	STATUS
2	19	11%		VISITED
0	23	0%		VISITED
4	20	20%		VISITED
0	19	0%		VISITED
2	23	9%		VISITED
4	18	22%		VISITED
2	20	10%		VISITED
7	22	32%		VISITED
2	21	10%		VISITED

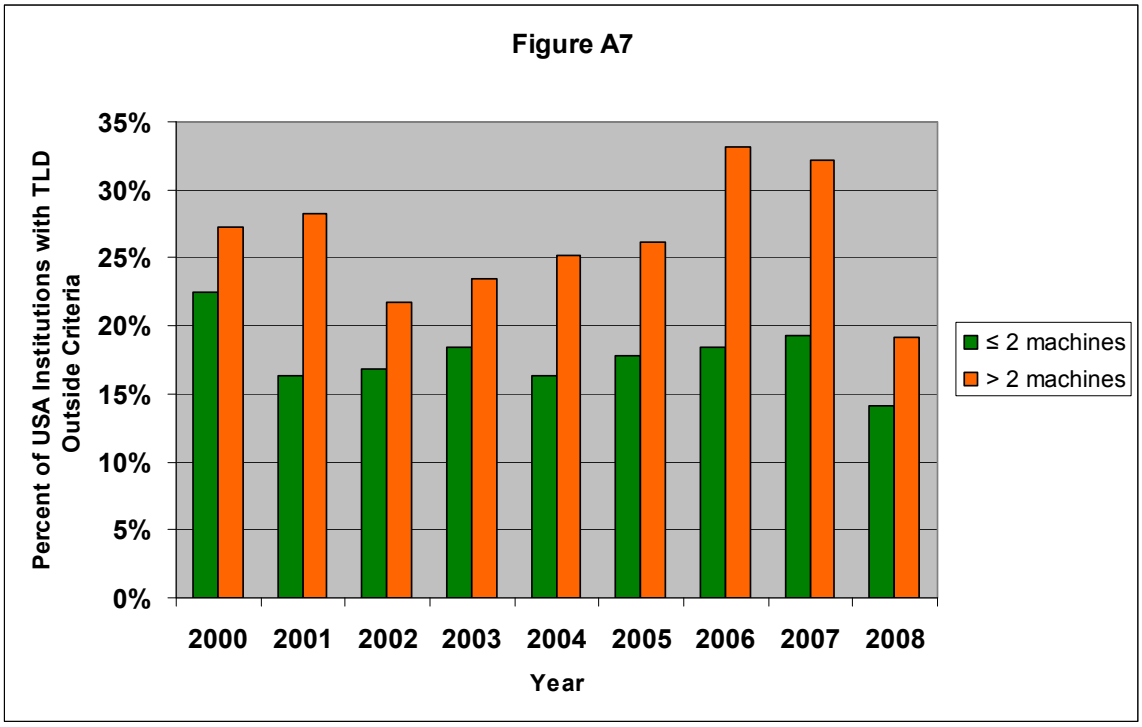
YEAR	ACTIVE
2000	26
2001	30
2002	26
2003	24
2004	29
2005	23
2006	25
2007	29
2008	30



CNT	TOTAL	%	YEAR	STATUS
2	3	67%	2000	NON-VISITED
1	3	33%	2001	NON-VISITED
1	3	33%	2002	NON-VISITED
0	4	0%	2003	NON-VISITED
0	3	0%	2004	NON-VISITED
0	4	0%	2005	NON-VISITED
3	9	33%	2006	NON-VISITED
1	21	5%	2007	NON-VISITED
6	46	13%	2008	NON-VISITED

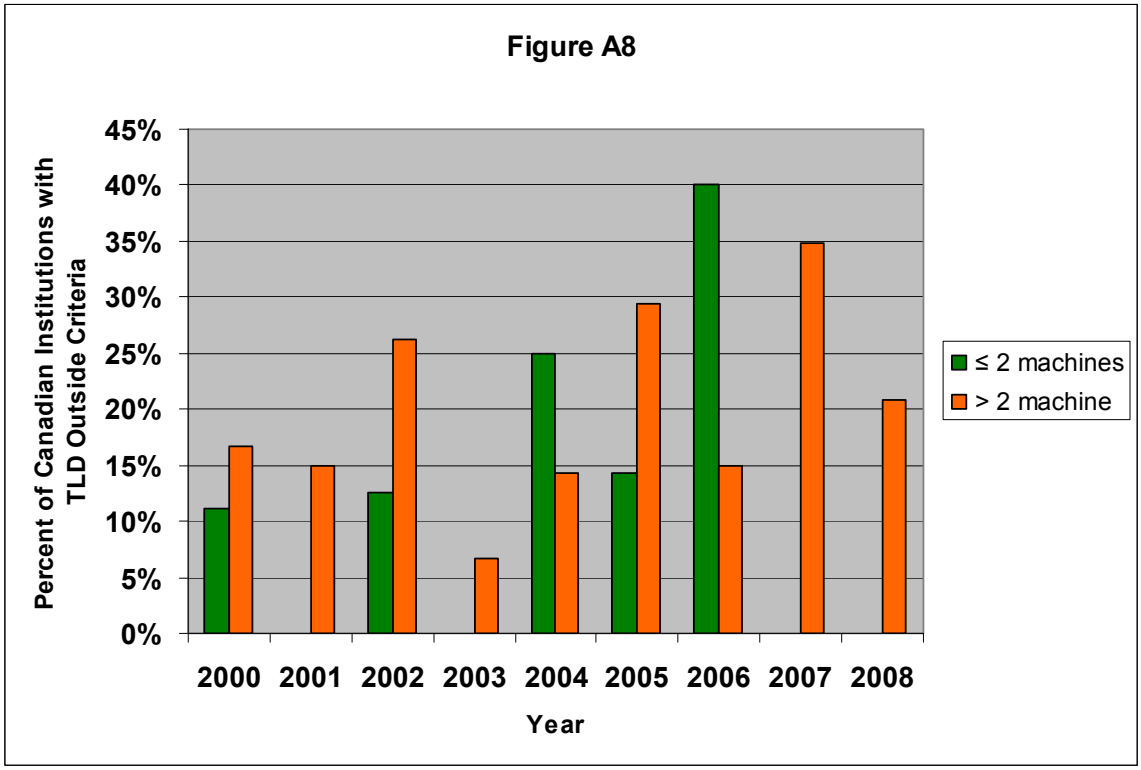
CNT	TOTAL	%	YEAR	STATUS
0	7	0%		VISITED
1	7	14%		VISITED
1	8	13%		VISITED
0	6	0%		VISITED
0	8	0%		VISITED
1	7	14%		VISITED
1	6	17%		VISITED
2	7	29%		VISITED
1	6	17%		VISITED

YEAR	ACTIVE
2000	10
2001	10
2002	11
2003	10
2004	11
2005	11
2006	15
2007	28
2008	52



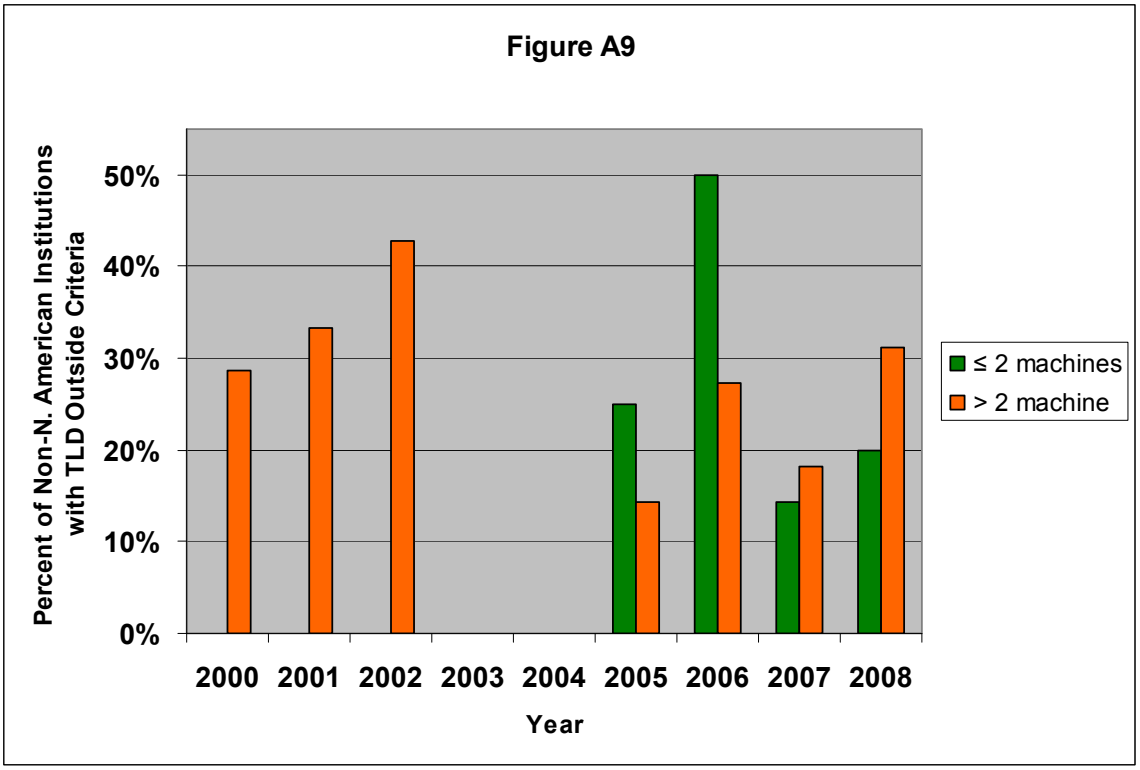
CNT	TOTAL	%	YEAR	Machines
164	730	22%	2000	≤ 2 machines
115	704	16%	2001	
134	796	17%	2002	
157	851	18%	2003	
151	926	16%	2004	
166	934	18%	2005	
202	1099	18%	2006	
226	1174	19%	2007	
151	1070	14%	2008	

CNT	TOTAL	%	YEAR	Machines
27	99	27%		> 2 machines
28	99	28%		
27	124	22%		
31	132	23%		
38	151	25%		
41	157	26%		
55	166	33%		
57	177	32%		
34	177	19%		

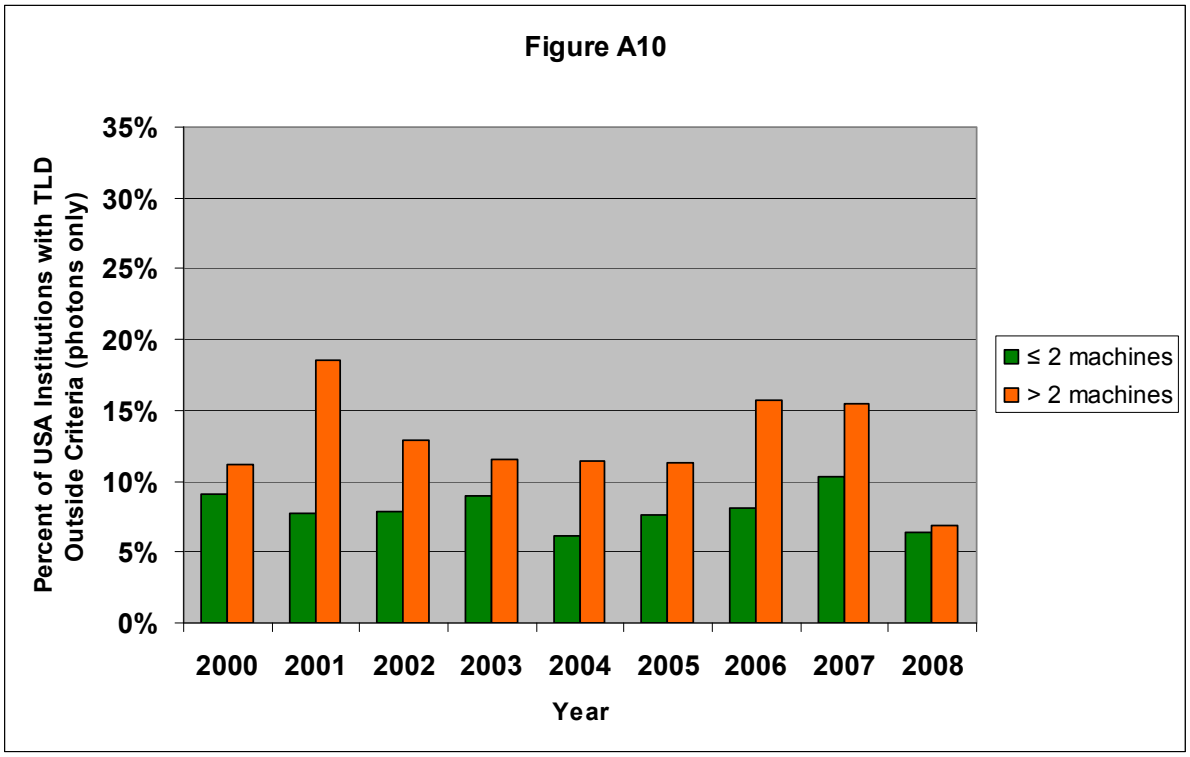


CNT	TOTAL	%	YEAR	Machines
1	9	11%	2000	≤ 2 machines
0	10	0%	2001	
1	8	13%	2002	
0	9	0%	2003	
2	8	25%	2004	
1	7	14%	2005	
2	5	40%	2006	
0	6	0%	2007	
0	6	0%	2008	

CNT	TOTAL	%	YEAR	Machines
3	18	17%		> 2 machine
3	20	15%		
5	19	26%		
1	15	7%		
3	21	14%		
5	17	29%		
3	20	15%		
8	23	35%		
5	24	21%		

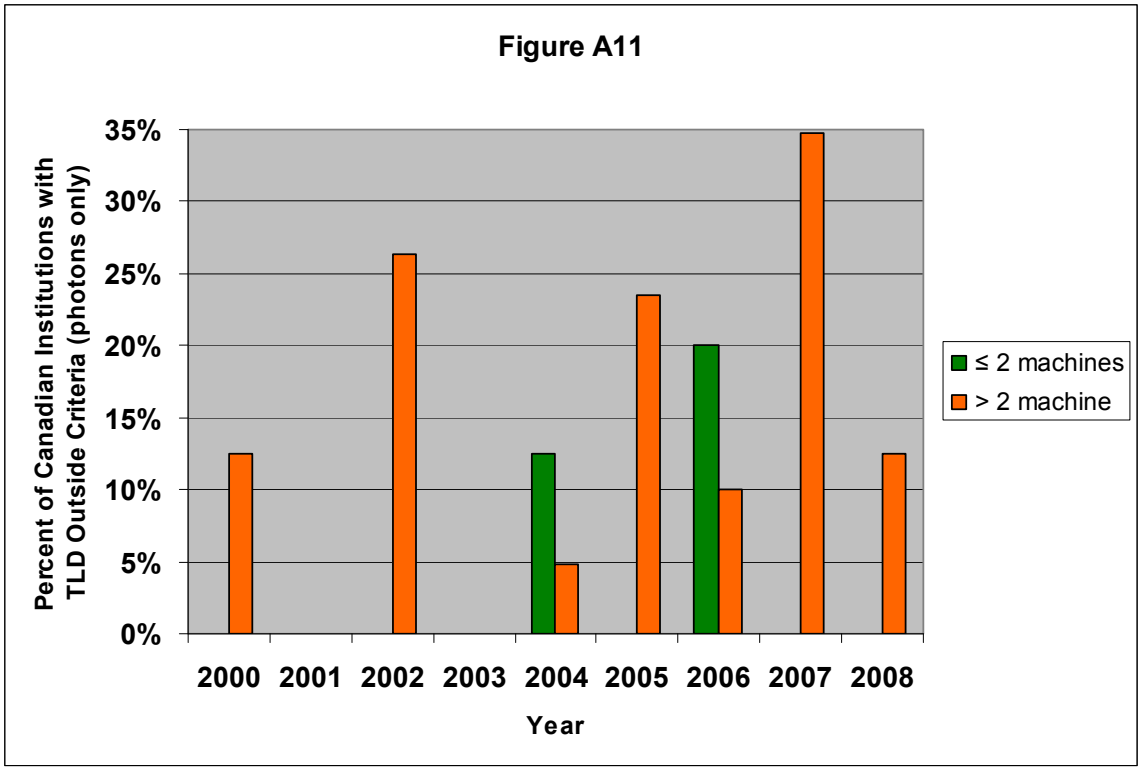


CNT	TOTAL	%	YEAR	Machines	CNT	TOTAL	%	YEAR	Machines
0	2	0%	2000	≤ 2 machines	2	7	29%		> 2 machin
0	2	0%	2001		2	6	33%		
0	3	0%	2002		3	7	43%		
0	3	0%	2003		0	6	0%		
1	3	33%	2004		0	8	0%		
1	4	25%	2005		1	7	14%		
2	4	50%	2006		3	11	27%		
1	7	14%	2007		4	22	18%		
4	20	20%	2008		10	32	31%		



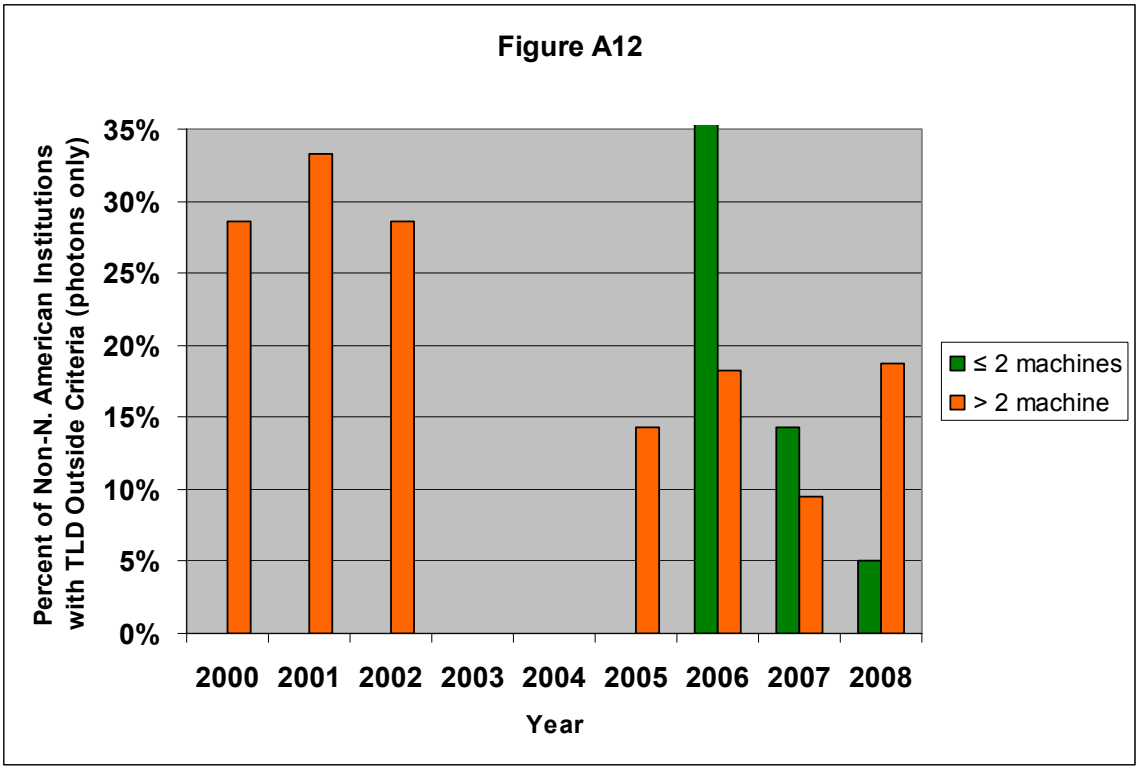
CNT	TOTAL	%	YEAR	Machines
65	713	9%	2000	≤ 2 machines
53	689	8%	2001	
62	790	8%	2002	
76	846	9%	2003	
56	919	6%	2004	
70	914	8%	2005	
88	1082	8%	2006	
120	1169	10%	2007	
68	1056	6%	2008	

CNT	TOTAL	%	YEAR	Machines
11	98	11%		> 2 machines
18	97	19%		
16	124	13%		
15	130	12%		
17	149	11%		
17	150	11%		
26	165	16%		
27	175	15%		
12	175	7%		

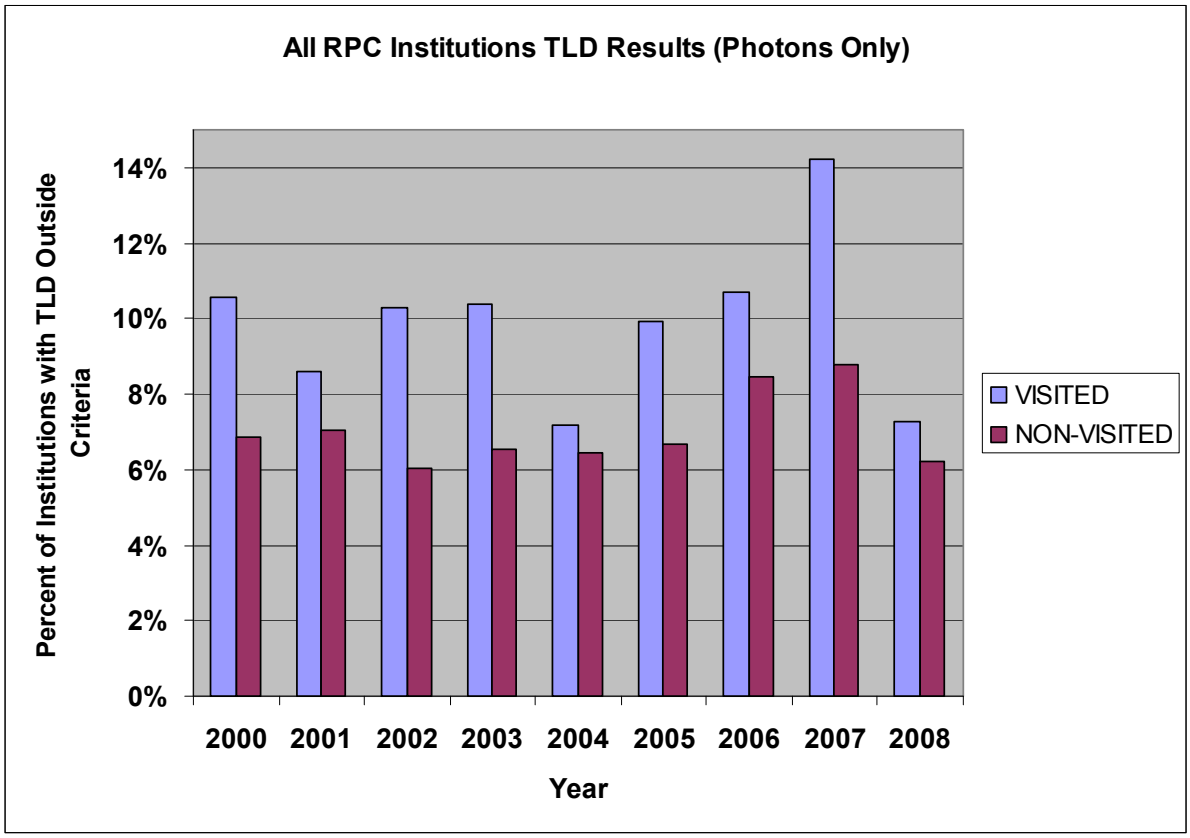


CNT	TOTAL	%	YEAR	Machines
0	9	0%	2000	≤ 2 machines
0	10	0%	2001	
0	7	0%	2002	
0	9	0%	2003	
1	8	13%	2004	
0	6	0%	2005	
1	5	20%	2006	
0	6	0%	2007	
0	6	0%	2008	

CNT	TOTAL	%	YEAR	Machines
2	16	13%		> 2 machine
0	20	0%		
5	19	26%		
0	15	0%		
1	21	5%		
4	17	24%		
2	20	10%		
8	23	35%		
3	24	13%		



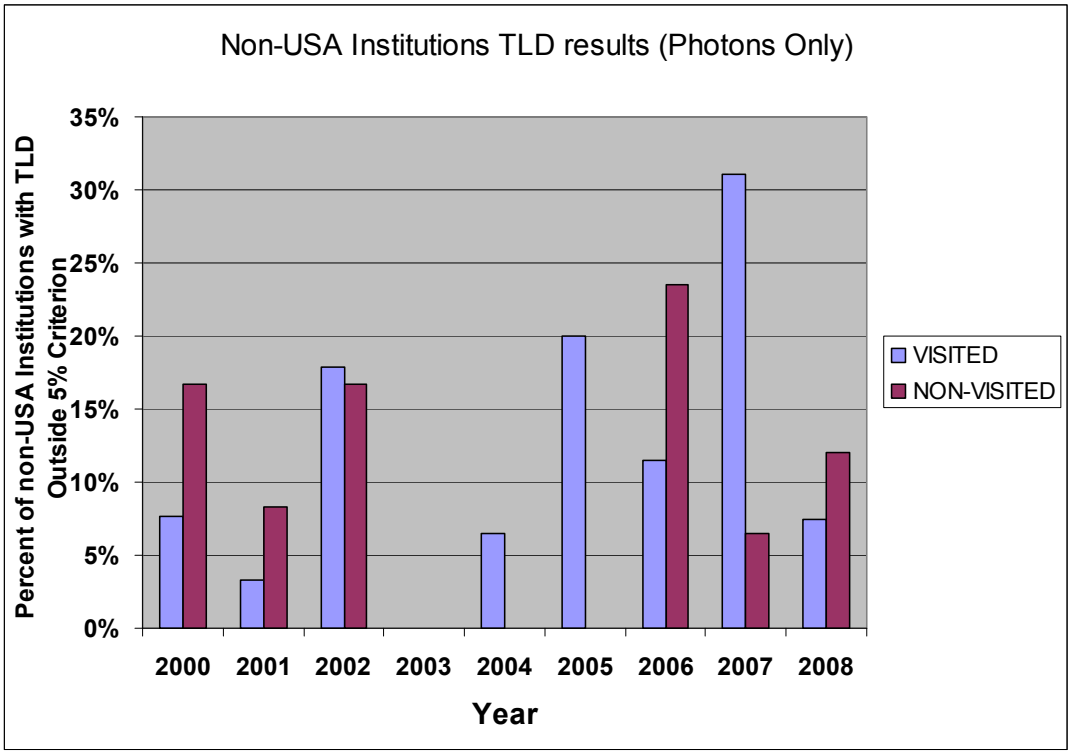
CNT	TOTAL	%	YEAR	Machines	CNT	TOTAL	%	YEAR	Machines
0	2	0%	2000	≤ 2 machines	2	7	29%		> 2 machin
0	2	0%	2001		2	6	33%		
0	3	0%	2002		2	7	29%		
0	3	0%	2003		0	6	0%		
0	3	0%	2004		0	8	0%		
0	4	0%	2005		1	7	14%		
2	4	50%	2006		2	11	18%		
1	7	14%	2007		2	21	10%		
1	20	5%	2008		6	32	19%		



CNT	TOTAL	%	YEAR	STATUS
29	423	7%	2000	NON-VISITED
30	427	7%	2001	NON-VISITED
29	480	6%	2002	NON-VISITED
33	504	7%	2003	NON-VISITED
37	572	6%	2004	NON-VISITED
39	586	7%	2005	NON-VISITED
61	722	8%	2006	NON-VISITED
71	808	9%	2007	NON-VISITED
48	772	6%	2008	NON-VISITED

CNT	TOTAL	%	YEAR	STATUS
67	633	11%		VISITED
49	570	9%		VISITED
66	642	10%		VISITED
67	646	10%		VISITED
48	669	7%		VISITED
61	614	10%		VISITED
72	673	11%		VISITED
99	697	14%		VISITED
44	605	7%		VISITED

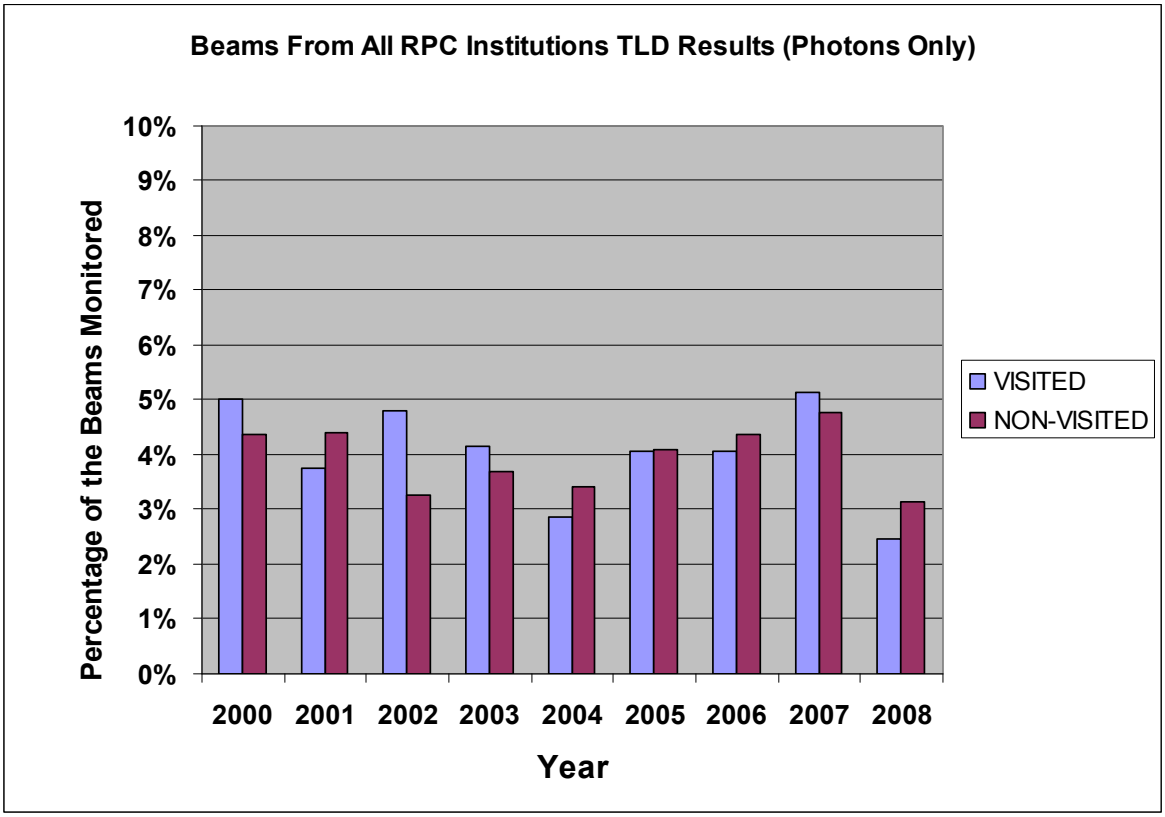
YEAR	ACTIVE
2000	1056
2001	997
2002	1122
2003	1150
2004	1241
2005	1200
2006	1395
2007	1505
2008	1377



CNT	TOTAL	%	YEAR	STATUS
2	12	17%	2000	NON-VISITED
1	12	8%	2001	NON-VISITED
2	12	17%	2002	NON-VISITED
0	9	0%	2003	NON-VISITED
0	12	0%	2004	NON-VISITED
0	12	0%	2005	NON-VISITED
4	17	24%	2006	NON-VISITED
2	31	6%	2007	NON-VISITED
7	58	12%	2008	NON-VISITED

CNT	TOTAL	%	YEAR	STATUS
2	26	8%		VISITED
1	30	3%		VISITED
5	28	18%		VISITED
0	25	0%		VISITED
2	31	6%		VISITED
5	25	20%		VISITED
3	26	12%		VISITED
9	29	31%		VISITED
2	27	7%		VISITED

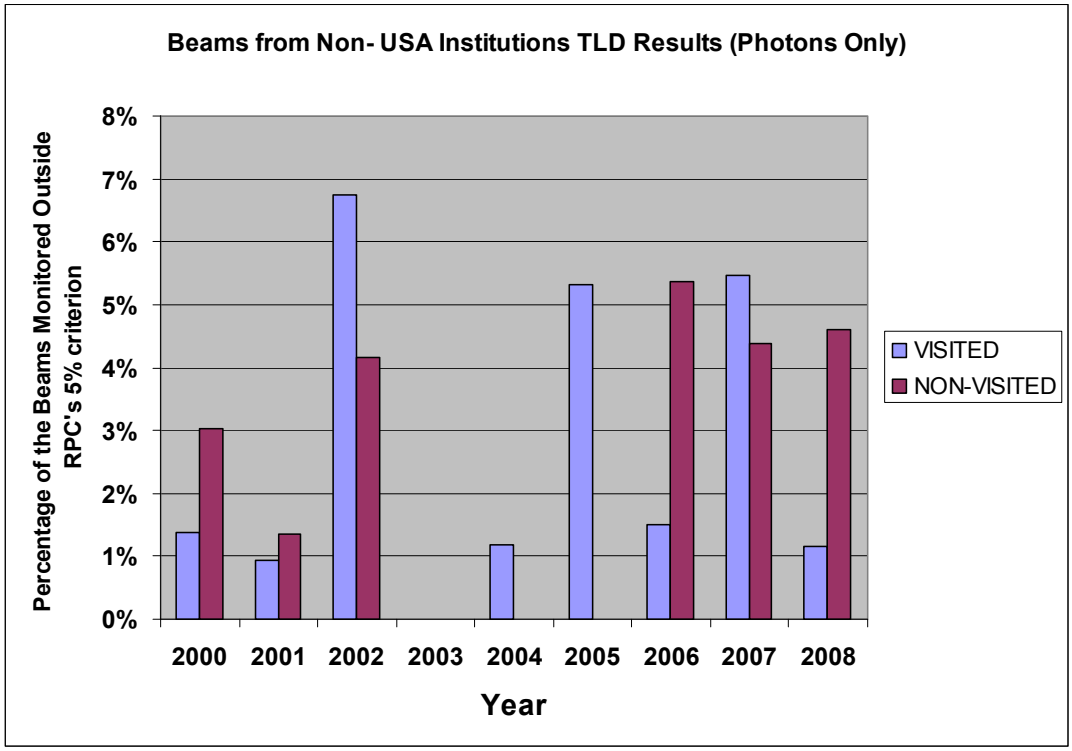
YEAR	ACTIVE
2000	38
2001	42
2002	40
2003	34
2004	43
2005	37
2006	43
2007	60
2008	85



CNT	TOTAL	%	YEAR	STATUS
43	981	4%	2000	NON-VISITED
45	1021	4%	2001	NON-VISITED
38	1168	3%	2002	NON-VISITED
46	1246	4%	2003	NON-VISITED
50	1463	3%	2004	NON-VISITED
61	1491	4%	2005	NON-VISITED
83	1903	4%	2006	NON-VISITED
103	2158	5%	2007	NON-VISITED
67	2126	3%	2008	NON-VISITED

CNT	TOTAL	%	YEAR	STATUS
105	2091	5%	2000	VISITED
73	1948	4%	2001	VISITED
111	2318	5%	2002	VISITED
98	2367	4%	2003	VISITED
74	2574	3%	2004	VISITED
94	2313	4%	2005	VISITED
112	2753	4%	2006	VISITED
154	2994	5%	2007	VISITED
62	2529	2%	2008	VISITED

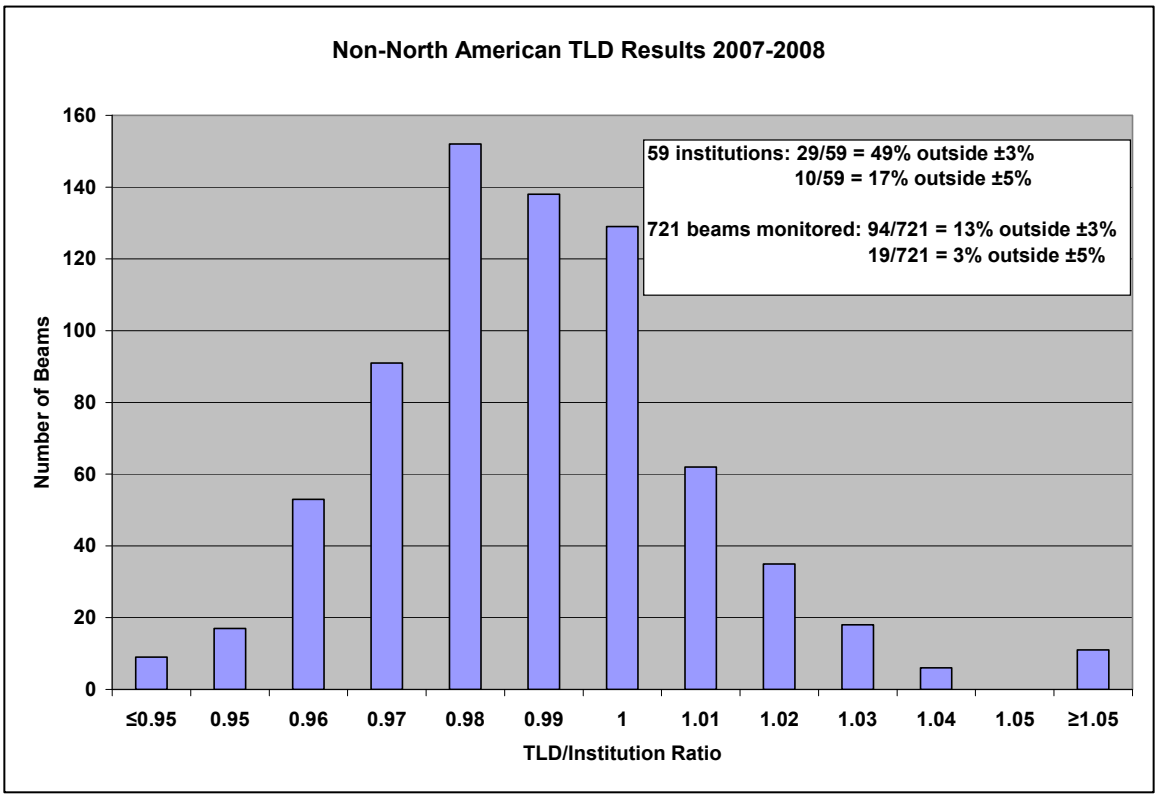
YEAR	# SENT TL
2000	3072
2001	2969
2002	3486
2003	3613
2004	4037
2005	3804
2006	4656
2007	5152
2008	4655

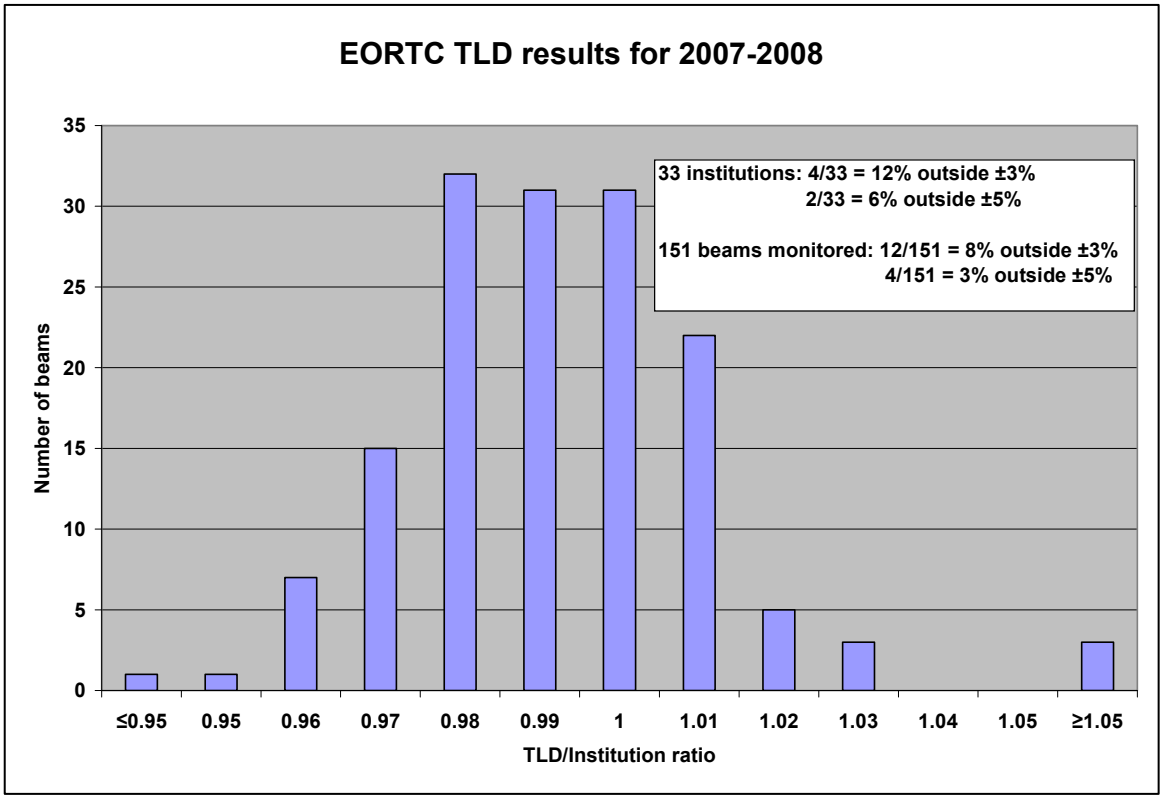


CNT	TOTAL	%	YEAR	STATUS
2	66	3%	2000	NON-VISITED
1	74	1%	2001	NON-VISITED
2	48	4%	2002	NON-VISITED
0	47	0%	2003	NON-VISITED
0	65	0%	2004	NON-VISITED
0	56	0%	2005	NON-VISITED
5	93	5%	2006	NON-VISITED
6	137	4%	2007	NON-VISITED
11	239	5%	2008	NON-VISITED

CNT	TOTAL	%	YEAR	STATUS
2	145	1%	2000	VISITED
2	216	1%	2001	VISITED
13	193	7%	2002	VISITED
0	207	0%	2003	VISITED
3	254	1%	2004	VISITED
10	188	5%	2005	VISITED
4	265	2%	2006	VISITED
17	311	5%	2007	VISITED
3	260	1%	2008	VISITED

YEAR	# SENT TL
2000	211
2001	290
2002	241
2003	254
2004	319
2005	244
2006	358
2007	448
2008	499


Non-North American TLD Results 2007-2008
59 institutions: 29/59 = 49% outside ±3%
10/59 = 17% outside ±5%
721 beams monitored: 94/721 = 13% outside ±3%
19/721 = 3% outside ±5%
Number of Beams
0
20
40
60
80
100
120
140
160
≤0.95
0.95
0.96
0.97
0.98
0.99
1
1.01
1.02
1.03
1.04
1.05
≥1.05
TLD/Institution Ratio


EORTC TLD results for 2007-2008
33 institutions: 4/33 = 12% outside ±3%
2/33 = 6% outside ±5%
151 beams monitored: 12/151 = 8% outside ±3%
4/151 = 3% outside ±5%
Number of beams
0
5
10
15
20
25
30
35
≤0.95
0.95
0.96
0.97
0.98
0.99
1
1.01
1.02
1.03
1.04
1.05
≥1.05
TLD/Institution ratio

Appendix B

Description of the TLD Results for Institutions in Terms of Their Patient Contributions

The data presented in figure 3 was derived from two main sources. The first being from information recieved by the RPC from the other Quality Assurance Offices and study groups regarding the number of patients each institution had placed onto trials with a radiotherapy component since 2000. The second source came from the TLD data and radiotherapy machine demographics for each participating institution that is kept in the RPC database.

Since the data in figure 3 is expressed as a percent of total patients per year affected, on average, by the photon beams outside the RPC's TLD criterion, the first value needed from the the two sources above was the total number of patients placed on trials per year to be used as the denominator to determine the percentage. Since 2000, the RPC has received data showing that 9940 patients treated with radiotherapy have been placed on NCI sponsored trials. Assuming an equal submission of patients each year, 1104 patients per year treated with radiotherapy have been placed on NCI sponsored trials.

Queries to the RPC database for each calendar year yields three quantities for each institution:

1. the number of patients (P) placed on trials (the total number of patients per institution since 2000 is known so the number of patients per year was calculated as the total number of patients divided by 9)
2. the total number of photon beams (B)
3. the number of photon beams outside the RPC's criterion (T) determined from the TLD audits but excluding those TLD results that were miss-irradiations

The number of patients per institution per year was then multiplied by the fraction of photon beams from that institution that had a TLD result outside the RPC's criterion. This product yielded the number of patients per institution that might have been affected by the photon beams with TLD results outside of the RPC's criterion. The total number of patients per institution that might have been affected by the photon beams with TLD results outside of the RPC's criterion were then summed for the specific calendar year. This sum for the specific year was then divided by the total number patients per year (1104 patients) treated with radiotherapy placed on NCI trials to yield a percentage.

$$\frac{(P)\bullet(T/B)\bullet 100}{1104}$$

The final percentage represents the potential percentage of patients that might be affected by the photon beams found to be outside of the RPC's criterion as shown in figure 3 for each calendar year.

Appendix C

Statistical Analysis Plan
QA-RT project: Comparison of ERDA policy in US (RTOG) and Europe (EORTC)

Study Number	**40084-22085 (with possibly some extra information from Glioma trial (NCCTG)**
Statistician responsible for this SAP at EORTC	**Laurence Collette**
Coordinators of the project within EORTC:	**EORTC ROG QART: Coen Hurkmans and Akos Gulyban**
Date:	January 4. 2010
Version	Version 4.1

Aims of the project

This is a plan for a prospective analysis of the optimal frequency and extension of External Reference Dosimetry Audit (ERDA) for clinical trials in a sub-study to the coming RTOG 0848 / EORTC 40084 adjuvant pancreas intergroup phase III trial.

External Reference Dosimetry Audit (ERDA) is implemented in EORTC and US clinical trials, but differs both in content and evaluation criteria:

- **EU-QART**: ERDA is performed every 2 years with checks of two selected beams (highest and lowest photon energy beams, not necessarily from the same machine, selected by the site) in the usual EORTC-ROG practice
- **US-QART**: ERDA performed every year and all photon beam energies of all machines are tested (RPC/US)).

This affects inter-group trial collaboration, therefore EORTC-ROG/QART and ATC together with RTOG decided to launch a specific sub-project to assess the optimal frequency and extension of the External Reference Dosimetry Audit program, using analysis of the RPC TLD audits and an upfront defined and agreed methodology confirmed by EORTC and ATC.

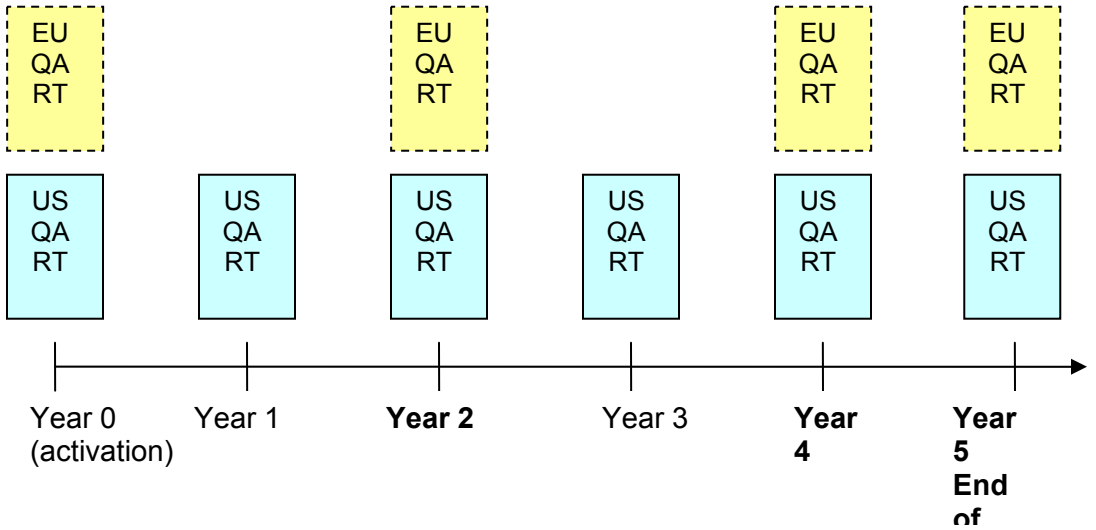
For this purpose, the EORTC agreed to use the US ERDA schedule and extension for all EORTC sites that participate in the EORTC 40084-22084 in collaboration with RTOG/RPC. Both groups also agreed that a statistical analysis of the resulting data will be performed in order to determine if the EORTC approach to ERDA schedule and extension would result in similar outcome.

The exact number of EORTC sites that will contribute to this study is as yet unknown, but it is anticipated to be around 40 for the EORTC 40084-22084 trial, as the EORTC committed to enter about 400 patients. Assuming there is on average 3.9 machines by institution (Budiharto T. et al. Radiation Oncology 2008), and 2 beam energies will be tested for each machine, it results that on average 7.8 measurements will be taken for each institution at each US-QART yearly assessment. Therefore a total of 312 beams are expected to be measured in EORTC sites at each yearly measurement. Thus, if all sites contribute to the whole of the study (i.e. 6 yearly assessments, see diagram), this would induce 1872 measurements for EORTC beams in total.

The trial recruitment period is foreseen to last 5 years, and the final analysis of the main clinical trial endpoint will take place three years later. The analysis of the ERDA-side study will be performed as soon as the data from year 5 are available.

EORTC centers will undergo the US-type of QART procedure prior to activation (time 0) and then yearly for the duration of the study recruitment (i.e. maximum 5 years) as indicated by the blue solid boxes in the scheme below.

The objective of the project is to investigate if the EU scheme (indicated by the yellow dashed boxes in the scheme, would have produced qualitatively as good results.



In this figure, the blue blocks represent the measures that are effectively made in this study. The yellow dashed boxes represent the results that could be recalculated from the database.

To prevent that knowledge of the US-QART results would influence the further EU-QART results in any given institution, the following measures will be taken;

- The US-QART is performed for all photon energies at participating institutions.
- The US-QART results are not shared with the institutions during the study as this information will affect how the institution responds to any problems found. The EU sites will only be informed of the results for the selected beams in years 0, 2, 4 and 5 during the study. After the last measurements in the last year of this study are performed, all measured data will be released to e.g., the EORTC ROG.

This design would allow for comparisons to be made at year 0 (EU vs US, baseline), year 2 (EU every two years vs US every year), year 4 (EU every two years vs US every year, plus informing EU of US results in year 2), and year 5 (EU every year vs US every year, plus informing EU of US results in year 4).

Primary hypothesis:

The primary objective is to demonstrate that, if the EU-QART schedule had been applied, the percentage of beams with dosimetry errors that would have remained undetected but would have been detected using the US-ERDA approach is acceptable.

For the purpose of this demonstration and to fully determine the "EU-QART" policy, strategies must be defined that result in the selection of two beams in each center.
The following four strategies will be assessed, the first one represents the EU-QART strategy currently used by EORTC and its comparison with the US-QART is the primary goal of this study. However three further experimental strategies will also be explored, to assess if they result in better detection rates. This will be used to optimize EU-QART strategy in the future. The results of the experimental strategies will be recalculated on the basis of the measurements available in the database.

EU-QART strategy	Allow the institution to select the machine (s) and energie(s) (2/site)
Experimental Strategy 1	Select most recent machine, test highest and lowest beam energies
Experimental Strategy 2	Select oldest machine, test highest and lowest beam energies
Experimental Strategy 3	Select machine with highest energy, test highest and lowest beam energies

The following hypotheses will be tested for the EU-ERDA strategy. Similar tests may be performed to study the experimental strategies, in an exploratory manner.

Hypothesis 1 (2 vs all beams tested in an institution):
Is it as good to test two selected beam versus testing all beams of all machines in a given institution in the EORTC framework?

Table 1: Variables to investigate performance of institutions.

Time t=0,1,2,4

$N_{t,i}=A_{t,i}+B_{t,i}+D_{t,i}$

		US-QART result	
		Pass	Fails
EU-QART	Pass	$A_{t,i}$	$B_{t,i}$
result	Fails	0 (by construction)	$D_{t,i}$

Test plan for Hypothesis 1
For assessment times t=0,2,4,5
Show that P[disagreement between US- and EU- QART]<5%, which is testing on the agreement between the two tests (one being nested in the other – i.e.: one empty cell by construction). This is equivalent to P[disagreement between EU & US]=$B_{t,i}/N_{t,i}$ must be <5% for t=0,2,4,5 (See Table 1).

Simulations:

- Simulations of 1000 repeats of 40 binary results B(0.05) that represent whether the results for a given site at a given time t are discordant (B=1) or concordant (B=0) were obtained assuming an AR(1) correlation matrix between repeated observations over time in the same institution, with parameter ρ=0.5 (autocorrelation) using the R package "binarySimCLF"
- The 95% bootstrap confidence interval around the probability of discordance for the 160 observations was (0.012 – 0.094)
- Simulations of the same set up with autocorrelation parameter ρ=0.7 gives a confidence interval (0.012 – 0.105)
- Simulations of 1000 repeats of 40 binary results Ranging B(0.10) for T=0 to B(0.05) for T=4 with autocorrelation parameter ρ=0.7 Lead to a 95% bootstrap confidence interval around the probability of discordance (average simulated p being 0.075 over the four time points) The confidence interval was (0.025 – 0.143)

For each of the four assessment times considered, each institution will be cross tabulated according to its results by EU-QART vs US-ERDA results

Hypothesis 2 (skipping every other year):
If the EU-QART schedule was used, the assessments at times 1 and 3 would be skipped. We must therefore show that the overall beam failure rate of EU sites at times 1 and 3 is low enough (≤5%). Indeed, these are the failures that would have remained undetected if one tests (all beams) only every other year

One can also assess, if the year 5 failure rate with the EU-QART assessment, that takes places only one year after the former assessment, is lower than the year 4 or year 2 EU-QART failure rates, that each take place 2 years after the former assessment.

Test plan for Hypothesis 2:

a) Let B1 be the number of **beams** assessed at time 1 and B3 be the number of beams assessed at time 3. Let F1 be the number of beams that failed by US-ERDA assessment at time 1 and F3 be the number of beams that failed by US-ERDA assessment at time 3.
One should show that p1=(F1/B1)≤5% and p3=(F3/B3) ≤5%. We will estimate the rates with their exact 2-sided 90% and 95% confidence intervals. The hypothesis will be demonstrated if the upper bound of the 90% CI is ≤5%.

b) Let S1 be the number of **sites** assessed at time 1 and S3 be the number of sites assessed at time 3. Let N1 be the number of sites that failed by US-ERDA assessment by strategy k at time 1 and N3 be the number of sites that failed by US-ERDA assessment at time 3.
One should show that p=(N1+N2)/(S1+S2))<12%. We will estimate the rates with their exact binomial 90% and 95% confidence intervals. The hypothesis will be demonstrated if the upper

bound of the 90% confidence interval is <12%. For this estimation we will assume that the outcomes of a given institution at years 1 and 3 are independent.

c) Compare the site failure rate with EU-QART at time 5 with the site failure rate by EU-QART at time 2 and time 4.

Power calculations:

a. For the test a), one will assume as before that with 40 sites with 2 beam energies tested on all machines and an average of 3.9 machines per site, we will have roughly 312 values available at each time point. With this number, one will have 90% power of rejecting the hypothesis H0: pi≥0.05 (I;e. 5%) under H1: pi=0.02 (i.e. 2%) at the 1-sided 5% significance level, at each timepoint i=1,3. This estimation does not account for possible within-institution correlation or within-machine correlation between beam energy measures.

b. For the test b) there will be a maximum of 40 observations at each timepoint, therefore the two timepoints will be pooled. Under the assumption that there is no within-institution correlation over time. With N=63 site measurements, we have 70% power of rejecting H0: p≥0.12 (i.e. 12%) under H1: p=0.04 (i.e. 4%)5 at the 1-sided 5% significance level (using binomial distribution). To reach 80% power, one would need 74 observations. .

c. Test c) severely lacks statistical power. Assuming that observations are independent over time, with N=40 sites at every timepoint, 80 observations (2x40) would provide only 60% power for a two-sided test of H0: both failure rates = 12% vs H1, one of the failure rates = 12% and the other =32 %. A very unlikely difference of 20%!

Pooling the timepoints 2 and 4, to use N=120 observations would provide 78% power under H1: failure rate at year 5=12% and overall failure rate for years 2+4=32

Exploratory hypotheses (these have low statistical power and will only be discussed (without thorough statistical analysis) in the results that will come out of this study. Only if the data show that EU results deviate significantly from US results, these tests might become statistically significant.

Hypothesis 3:

It is claimed that since in Europe, several physicists cross- validate the beam calibration, there is a high correlation between the performance of all machines within an institution

Estimate the within-institution correlation coefficient between all beam energies tested within the institutions (show ρ >0.90, for continuous measure), estimate agreement between the EU-QART results (2 beams) and outcome of the non selected the other institution's machines that are not selected by the EU-QART strategy

Statistical inference for hypothesis 3

a) Estimate overall agreement coefficient between beams tested within institutions (as a binary result pass/fails), with 95% confidence interval
b) Estimate the overall correlation coefficient between beam results tested within institutions (as a continuous measure of the % of relative deviation from the target dose) with 95% confidence interval

Hypothesis 4:
It is possible that, possibly because the institution/equipment conditions are not the same in Europe as in US, the overall failure rate of EORTC sites/beams, even using the US-QART approach is lower than that reported in US (12% of institutions fail, 5% of all beam tested fail).

We will estimate the overall probability of failure of EORTC sites and beams and calculate the 95% confidence interval
One must show

a) the overall rate of institution failure is <12-13%%
b) the overall rate of beam failure is <5%

In order to account for the autocorrelation between measurements on beams within a **single** institution and within a beam over time, a generalized logistic mixed model with only the intercept term will also be fit, with either AR(1) or CS correlation structure blocked by center.

Hypothesis 5:
We wish to estimate the impact of institution's equipment, staffing and other characteristics on the risk of failure of EU sites/Beams when using the selected EU-QART or the US-QART approach

Testing this hypothesis will make use of the extra-information from institutions that are collected through the Facility Questionnaire, for all EORTC institutions.

Test plan for Hypothesis 5

Influence of institution specific factors related to staffing, equipment and institution on ERDA test results

a) Using generalized linear models for repeated binary observations (with either AR(1) or CS correlation structure blocked by center), with the outcome variable being the ERDA result of institution j at time i the US-QART policy (with i= 0, 1,2,3,4), we will assess the influence of the following factors. Univariate tests at the 5% significance level will first be conducted. If several factors are significant in the univariate analysis, they will then be entered in a multivariate model to assess their relative influence.

- number of physicists/linac
- number physicists/site
- number physicists/patient
- number of linacs/site
- number of patients/linac
- country (probably not enough power)

b) A similar analysis will be conducted using the numeric outcome of the test (ie: % deviation from the target dose) using generalized linear mixed models.

General comments

This analysis will primarily aim at estimating the various probabilities listed in the analysis plan. Confidence intervals will be estimated at the 2-sided 95% confidence level.

The analysis should account for the following features, wherever required and feasible (i.e.: provided models can converge)

- Correlation between measurements of a same beam at different times
- Correlation between measurements of beams within a same institutions
- Drop out of institutions (if not all institutions are initiated at the same time, some institutions may have only 3 or 4 measurements available instead of 5
- In general, we will assume non informative drop out.

Appendix: Statistical considerations and (rejected) alternatives, for informative purpose only to draft the final ERDA SAP.

Hypothesis 1 (2 vs all beams tested in an institution):
Is it as good to test two selected beam versus testing all beams of all machines in a given institution in the EORTC framework?

Test plan for Hypothesis 1

For assessment times i=0,2,4,5 ,
A) show P[site fails by US approach at this assessment | site passes by EU approach for this same assessment] < 5% , which is showing that PPV of EU-QART (2 beams) to predict outcome of US-QART (all beams) is >95%OR
B) Show that P[disagreement between US- and EU- QART]<5%, which is testing on the agreement between the two tests (one being nested in the other – ie: one empty cell by construction)

Power calculations for B are easier because the statistic is Bi/Ni and estimates of Ni are known whereas with approach A, it is Bi/(Ai+Bi) which itself will need to be based on the expected failure rate with EU-QART approach (which we will assume to be 15% approximately)..
However, with N=40 per timepoint we will have no power to discard an overall disagreement rate <5%.

Simulations of 1000 repeats of 40 binary results B(0.05) that represent whether the results for a given site at a given time t are discordant (B=1) or concordant (B=0) were obtained assuming an AR(1) correlation matrix between repeated observations over time in the same institution, with parameter ρ=0.5 (autocorrelation) using the R package "binarySimCLF"

The 95% bootstrap confidence interval around the probability of discordance for the 160 observations was (0.012 – 0.094)

Simulations of the same set up with autocorrelation parameter ρ=0.7 gives a confidence interval (0.012 – 0.105)

Simulations of 1000 repeats of 40 binary results Ranging B(0.10) for T=0 to B(0.05) for T=4 with autocorrelation parameter ρ=0.7
Lead to a 95% boostrap confidence interval around the probability of discordance (average simulated p being 0.075 over the four time points)
The confidence interval was (0.025 – 0.143)

Or
C) look into Kappa statistic and demonstrate that Kappa is high enough (to be defined) However Kappa is sensitive to skewed margins..
D) Consider Mc Nemar's test of equal marginal distributions

For each of the four assessment times considered, each institution will be cross tabulated according to its results by EU-QART vs US-ERDA results

Time i=0,1,2,4

$N_i=A_i+B_i+D_i$
centers tested

		US-QART result	
		Pass	Fails
EU-QART	Pass	A_i	B_i
result	Fails	0 (by construction)	D_i

P[disagreement between EU & US]=B_i/N_{ii} must be <5% for i=0,2,4,5

The following theoretical set up gives a Kappa value of 0.77.

Ni=40,
Ai+Bi=0.85*40=34
Di=6

EU\US	Pass	Fail	
Pass	85%/34	5% / 2	90% /36
Fail	0	10%/4	10% / 4
	85%/34	15%/ 6	100% / 40

Sample size calculations for approach C:

Assuming marginal probability distribution as above (ie: 10% of sites fail the EU-QART, whereas 15% fail the US-QART at each time), to test the hypothesis
H0: Kappa≤0.80 vs H1: Kappa>0.80 at 1-sided alpha 5% requires
330 cases for 80% power if the true kappa value is 0.90
102 cases for 80% power if the true Kappa value is 0.95
47 cases for 80% power if the true kappa value is 0.98

If the marginal distributions are more skewed (ie: lower rate of site failures) then the numbers required for each test further increase
Eg, if 5% sites fail the EU-QART and 10% sites fail the US-QART then the numbers become
H0: Kappa≤0.80 vs H1: Kappa>0.80 at 1-sided alpha 5% requires
508 cases for 80% power if the true kappa value is 0.90
157 cases for 80% power if the true Kappa value is 0.95
72 cases for 80% power if the true kappa value is 0.98

(numbers were computed with N.cohen.kappa from R-package “concord” with code:
N.cohen.kappa(0.10,0.15,0.95,0.80,0.05,0.80,F))

At most, we will have 4x40 site with paired results (EU,US), however, these will be correlated (repeated observations inside a center) therefore, even pooling the information over time points, we will have the equivalent of <160 independent observations.

Sample size for D (Mc Nemar):

For a table similar to that above, a Mc-Nemar tests with 1-sided alpha=0.05 and Power 0.80 under the assumption that the % discordant pairs is 5% and the expected difference in marginal probabilities of site failure (EU-US) is 5%, would require 120 observations..

(Caculations using EAST 5.3)

Hypothesis 2(skipping every other year):

If the EU-QART schedule was used, the assessments at times 1 and 3 would be skipped. We must therefore show that the overall beam failure rate of EU sites at times 1 and 3 is low enough (<5%). Indeed, these are the failures that would have remained undetected if one tests (all beams) only every other year

One can also assess,if the year 5 failure rate with the EU-QART assessment, that takes places only one year after the former assessment, is lower than the year 4 or year 2 EU-QART failure rates, that each take place 2 years after the former assessment.

Test plan for Hypothesis 2:

d) Let B1 be the number of **beams** assessed at time 1 and B3 be the number of beams assessed at time 3. Let F1 be the number of beams that failed by US-ERDA assessment at time 1 and F3 be the number of beams that failed by US-ERDA assessment at time 3.
One should show that (F1/B1)<5% And (F3/B3)<5%. We will estimate the rates with their exact 2-sided 90% and 95% confidence intervals. The hypothesis will be demonstrated if the upper bound of the 90% CI is <5%.

e) Let S1 be the number of **sites** assessed at time 1 and S3 be the number of sites assessed at time 3. Let N1 be the number of sites that failed by US-ERDA assessment by strategy k at time 1 and N3 be the number of sites that failed by US-ERDA assessment at time 3.
One should show that (N1+N2)/(S1+S2))<12%. We will estimate the rates with their exact binomial 90% and 95% confidence intervals . The hypothesis will be demonstrated if the upper bound of the 90% confidence interval is <5%. For this estimation we will assume that the outcomes of a given institution at years 1 and 3 are independent.

f) Compare the site failure rate with EU-QART at time 5 with the site failure rate by EU-QART at time 2 and time 4.

Power calculations:

For the test a), one will assume as before that with 40 sites with 2 beam energies tested on all machines and an average of 3.9 machines per site, we will have roughly 312 values available at each time point.
With this number, one will have 90% of rejecting the hypothesis H0: p≥0.05 under H1: p=0.02 at the 1-sided 5% significance level, at each timepoint i=1,3. This estimation does not account for possible within-institution correlation or within-machine correlation between beam energy measures.

For the test b) there will be a maximum of 40 observations at each timepoint, therefore the two timepoints will be pooled. Under the assumption that there is no within-institution correlation over time.
With N=63 site measurements, we have 70% power of rejecting H0: p≥0.05 under H1: p=0.005 at the 1-sided 5% significance level (using binomial distribution). To reach 75% power, one would need 93 observations to have 75% power
Under p=0.003, we would have 75% power with 59 observations.

Test c) severely lacks statistical power. Even assuming that observations are independent over time, with N=40 sites at every timepoint, 80 observations (2x40) would provide 75% power for a two-sided test of H0: failure rates = 5% vs H1, one of the failure rates = 5% the other =30%. Pooling the timepoints 2 and 4, to use N=120 observations would provide 75% power under H1: failure rate at year 5=5% and overall failure rate for years 2+4=25%. A very unlikely difference of 20%!
Should the failure rate be higher than 5% under H0, the picture is worse..

Exploratory hypotheses

Hypothesis 3:
It is claimed that since in Europe, several physicists cross- validate the beam calibration, there is a high correlation between the performance of all machines within an institution

Estimate the within-institution correlation coefficient between all beam energies tested within the institutions (show ρ >0.90, for continuous measure), estimate agreement between the EU-QART results (2 beams) and outcome of the non selected the other institution's machines that are not selected by the EU-QART strategy

Statistical inference for hypothesis 3

c) Estimate overall agreement coefficient between beams tested within institutions (as a binary result pass/fails), with 95% confidence interval
d) Estimate the overall correlation coefficient between beam results tested within institutions (as a continuous measure of the % of relative deviation from the target dose) with 95% confidence interval

Hypothesis 4:
It is possible that, possibly because the institution/equipment conditions are not the same in Europe as in US, the overall failure rate of EORTC sites/beams, even using the US-QART approach is lower than that reported in US (12% of institutions fail, 5% of all beam tested fail). (This would take some thought as to what conditions one would consider and then there still would not be any direct connection between the conditions and ERDA failure.)

We will estimate the overall probability of failure of EORTC sites and beams and calculate the 95% confidence interval
One must show

c) the overall rate of institution failure is <12-13%%
d) the overall rate of beam failure is <5%

In order to account for the autocorrelation between measurements on beams within a **single** institution and within a beam over time, a generalized logistic mixed model with only the intercept term will also be fit, with either AR(1) or CS correlation structure blocked by center.

Hypothesis 5:
We wish to estimate the impact of institution's equipment, staffing and other characteristics on the risk of failure of EU sites/Beams when using the selected EU-QART or the US-QART approach

Testing this hypothesis will make use of the extra-information from institutions that are collected through the Facility Questionnaire, for all EORTC institutions.

Test plan for Hypothesis 5

Influence of institution specific factors related to staffing, equipment and institution on ERDA test results

c) Using generalized linear models for repeated binary observations (with either AR(1) or CS correlation structure blocked by center), with the outcome variable being the ERDA result of institution j at time i the US-QART policy (with i= 0, 1,2,3,4), we will assess the influence of the following factors. Univariate tests at the 5% significance level will first be conducted. If several factors are significant in the univariate analysis, they will then be entered in a multivariate model to assess their relative influence

- number of physicists/linac
- number physicists/site
- number physicists/patient
- number of linacs/site
- number of patients/linac
- country (probably not enough power)

d) A similar analysis will be conducted using the numeric outcome of the test (ie: % deviation from the target dose) using generalized linear mixed models.