The relationship between waiting time for radiotherapy and clinical outcomes: A systematic review of the literature
Zheng Chena, Will Kinga, Robert Pearceyb, Marc Kerbaa, William J. MacKillopab, *
aQueen’s Cancer Research Institute, Queen’s University, Kingston, Ont., Canada, bCross Cancer Institute, Edmonton, Alta., Canada

Abstract

Purpose: To synthesize the direct clinical evidence relating waiting times (WTs) for radiotherapy (RT) to the outcomes of RT.

Methods and materials: We did a systematic review of the literature between 1975 and 2005 to identify clinical studies describing the relationship between WTs and outcomes of RT. Only high quality (HQ) studies that had adequately controlled for confounding factors were included in the primary analysis. WTs that had originally been reported as a categorical variable were converted to a continuous variable based on the distribution of WTs in each category. Meta-analyses were done using a fixed-effect model.

Results: The systematic review identified 44 relevant studies. Meta-analyses of 20 HQ studies of local control demonstrated a significant increase in the risk of local failure with increasing WT, RRlocal recurrence/month = 1.14, 95% Confidence Intervals (CI): 1.09–1.21. For post-operative RT for breast cancer; RRlocal recurrence/month = 1.11, 95%CI: 1.04–1.19. For post-operative RT for head and neck cancer, RRlocal recurrence/month = 1.28, 95%CI: 1.08–1.52. For definitive RT for head and neck cancer, RRlocal recurrence/month = 1.15, 95%CI: 1.02–1.29.

There was little evidence of any association between WTs and the risk of distant metastasis. Meta-analyses of the 6 HQ studies of breast cancer showed RRmetastasis/month = 1.04, 95%CI: 0.98–1.09. Meta-analyses of 4 HQ studies of breast cancer showed no significant decrease in survival with increasing WT, RRdeath/month = 1.06, 95%CI: 0.97–1.16, but there was a marginally significant decrease in survival in 4 HQ studies of head and neck cancer, RRdeath/month = 1.16, 95%CI: 1.02–1.32.

Conclusions: The risk of local recurrence increases with increasing WTs for RT. The increase in local recurrence rate may translate into decreased survival in some clinical situations. WTs for RT should be as short as reasonably achievable.

Keywords: Systematic review; Meta-analyses; Radiotherapy; Waiting time; Local recurrence; Metastasis; Survival

Waiting lists for radiotherapy (RT) were first identified as a cause for concern in the medical literature in a report from Norway almost 20 years ago [1], and have since been reported in many other countries including the UK [2], Australia [3], New Zealand [4], Denmark [5], Germany [6], Spain [7], Italy [8], and Canada [9]. The underlying cause of the problem is a continuing increase in the demand for RT due to the increasing incidence of cancer attributable to population aging, and to the discovery and adoption of new indications for RT [10]. When the increasing demand is not matched by a commensurate increase in treatment capacity, waiting lists for RT become inevitable.

Many different types of evidence suggest that it would be prudent to avoid unnecessary delays in initiating RT [11]. There is a vast amount of indirect evidence that any delay in starting RT would be expected to have some adverse effect on local control. There is strong experimental evidence that the probability of eradicating a tumor by RT is inversely related to the number of clonogenic cells it contains [12]. The quantitative relationship between probability of cure and the burden of clonogenic cells is well explained by robust radiobiological theory [13]. There is also abundant clinical evidence that the probability of local control of many different types of human cancer is inversely related to the volume of the tumor [14–18]. There is overwhelming evidence that most human cancers grow inexorably, although their growth rates vary widely [19–21]. Two recent studies have shown clear evidence of tumor progression in a high proportion of patients waiting to start radiotherapy for head and neck cancer [5,22]. Radiobiological models have been used to estimate the impact of delay in RT on local control from knowledge of human tumor growth rates and the observed relationship between tumor volume and the probability of local control [23]. However, in the
era of "evidence-based medicine" [24], there is skepticism about any recommendations that are based only on indirect evidence [11]. Estimates of the adverse consequences of delay based on radiobiological models have therefore not persuaded policy makers that waiting for RT poses a danger to patients [11]. The purpose of this paper was to synthesize the direct evidence that waiting for RT may affect clinical outcomes.

The project was undertaken in response to an unusual request from the Canadian federal and provincial governments for assistance from the research community in establishing "evidence-based benchmarks" for waiting times for selected procedures. Canada has experienced increasing problems with waiting lists for many different medical services [25–28]. These problems have been widely reported in the media and have gradually eroded public confidence in the Canadian health system. Public opinion polls have shown that "wait times" for medical care have become the greatest concern of many Canadian voters [29].

In November 2004, in response to increasing political pressure for action to reduce WTs, the federal and provincial governments of Canada made a joint announcement that they would establish "evidence-based benchmarks" for WTs for cancer treatment, cardiac surgery, joint replacements, cataract surgery, and diagnostic imaging [30]. The governments also announced that they would develop a national process for monitoring and reporting compliance with these benchmarks, although they stopped short of announcing the "wait time guarantees" that some advocacy groups were looking for [31]. In order to assist the government in establishing its waiting time "benchmarks", the Institute of Health Services Research and Policy of the Canadian Institutes of Health Research (CIHR) announced a fast-tracked Request for Applications to synthesize the relevant clinical evidence. This unique funding competition offered health services researchers the unusual opportunity to play a direct role in policy development. Our group was funded through this program to synthesize the direct clinical evidence relevant to the establishment of benchmarks for WTs for RT.

A randomized clinical trial (RCT) is the best way to detect and measure differences in the effectiveness of alternative treatment strategies [24]. However, no RCTs have been done to explore the effects of treatment delay on the outcomes of RT, and none are likely to be done in future, because most observers agree that this approach would be unethical in this context [32]. The best available evidence about the relationship between WTs for RT and outcomes therefore comes from retrospective observational studies. There have now been many clinical reports in which the outcomes of RT have been compared between patients who have waited longer or shorter periods for RT. Most have included quite small numbers of cases and have lacked the statistical power to detect the relatively small effects that have been predicted theoretically [32]. The purpose of the present study was to review and synthesize the clinical literature pertaining to the relationship between WTs and the outcomes of RT, in order to address the following specific questions: Does delay adversely affect the outcome(s) of RT, and if so, what is the quantitative relationship between the duration of delay and the risk of an adverse outcome?

Several previous systematic reviews of the field, including one from our own group, have reached somewhat different conclusions based on the available information [32–36]. Most have focused on the relationship between WTs and local control [32,33,35,37], and three have dealt exclusively with breast cancer [33,35,37]. In the present review, we have taken a more comprehensive approach; we have included all types of cancer, and we have described and synthesized what is known about the relationship between WTs for RT and the probability of distant metastasis and survival, as well as the probability of local control. The present review also attempts to correct methodological weaknesses in previous studies, as discussed in detail in the methods section, and includes the results of a number of important new studies published in the four years since our last review.

Methods and material

Search strategy

We first searched the indexed databases including PubMed, HealthSTAR, CancerLit, Cochrane Library from 1975 to July 2005, using the keywords or Medical Subject Headings (MeSH) limited to humans: "waiting lists", "wait times", "delay", "interval", "timing", "radiotherapy", "radiation", "irradiation", "outcome", "local recurrence", "metastasis", "survival" in all the languages. Abstracts of these reports were scanned to exclude those that were clearly irrelevant. We next conducted manual searches of the reference lists of the key articles to identify relevant studies we might have missed on the primary search. We also reviewed abstracts from the annual meetings of the American Society for Therapeutic Radiology and Oncology (ASTRO) and the Canadian Association for Radiation Oncology (CARO). Where the same patient population has been studied in different reports, the most recent one was used. If the reports provided incomplete information or ambiguous description, we attempted to contact the authors for additional information.

Inclusion criteria

Only clinical studies that met the following criteria were identified as relevant to our analysis: (a) all patients were treated with RT; (b) WT for RT was defined, measured, and reported; (c) one or more of the following outcomes was measured and reported: local control/failure, metastasis, survival, quality of life. Studies which measured tumor growth while waiting for RT, but did not measure outcomes were not included. Studies which modeled the impact of WTs on outcomes based on observed or expected tumor growth rates, but did not measure outcomes directly were also not included.

Quality criteria

We identified no relevant randomized trials. All the relevant studies which were identified were observational. The fundamental problem with observational studies is that the comparison groups may be imbalanced with respect to patient-related, disease-related, or treatment-related...
factors, which might influence the outcome(s) of interest. We therefore made the a priori decision to confine our primary analyses to the studies in which the comparison groups (i.e. those who had waited for longer or shorter periods) were reasonably similar with respect to other prognostic factors associated with the outcome of interest, or which reported associations between WTs and outcomes after appropriate adjustment for differences in the relevant prognostic factors. We have used the term “High quality” (HQ) to describe the studies that met those criteria. We operationalized this approach by addressing the following questions: (a) Was the distribution of the relevant prognostic factors adequately described in the groups of patients which were compared? If NO, then classify the study as not of high quality (NHQ); if YES, proceed to question (b), Were the comparison groups balanced with respect to the relevant prognostic factors? If YES, classify the study as HQ; if NO, proceed to question (c). Were the reported results appropriately adjusted for any differences in the relevant prognostic factors? If YES, then classify as HQ; if NO, then classify as NHQ. Stage, grade, RT dose/fractionation, volumes and technique, and type of systemic treatment were deemed to be relevant in all studies. The status of the resection margins was deemed relevant for all studies of post-operative RT. In our previous review [32], we used a nine-point scale to classify the quality of studies, but we abandoned this approach here in favor of the binary HQ/NHQ approach, because we believed that high scores in some aspects of study design should not be allowed to compensate for low scores in other areas of design that might independently bias results [38].

The relevant studies were reviewed by four independent reviewers. Initially, the four reviewers unanimously agreed on the quality of 44 papers. Initial disagreements regarding the status of the remaining seven papers were usually attributable to one or more of the reviewers having missed some key points of information in the text, tables or figures. All disagreements were resolved by consensus.

Definition of delay

WT was defined as the interval from the date of diagnosis to start of RT for the primary RT, or from the date of surgery to the start of RT for the post-operative RT.

In many of the original reports, individual patient WTs were not reported. Patients were instead categorized into two or three groups, that had waited longer or shorter periods, and outcomes were compared between these groups. The cutoff points that separated the groups varied from study to study. In a few more-recent studies, individual WT was described and its association with outcomes was explored in a multivariate analysis which included individual WT as an independent continuous variable.

Combining the results of these disparate studies presents challenges. In our previous meta-analyses we were only able to synthesize the results of studies that compared outcomes of groups that had waited longer and shorter periods when they had used a common cutoff point to separate longer and shorter WTs. Studies that had not used the consensus cutoff point had to be left out of the meta-analyses. More importantly, this approach only permitted us to compare outcomes between patients who had experienced “longer” and “shorter” WTs, and it did not permit us the quantify the impact of delay per unit time, which is the quantity of interest from the clinical and policy perspectives.

Data extraction

For the purposes of this study, Hazard Ratios (HRs) and Relative Risks (RRs) were treated as equivalent estimates of effect size [38]. In order to synthesize the results of studies with disparate WT variables (e.g. continuous and various categorical representations), we converted the WT effect size reported in each study to a regression coefficient (β) and standard error (SE) corresponding to a continuous representation per month of WT. For studies with categorical WT representations, the median was assigned to reflect the central value of WT for each category. If the original study did not provide the median WT in each category, it was estimated from the boundaries of the category, assuming that the shape of the WT distribution was similar to that observed in Ontario, for which complete data were available to us.

Regression coefficients and standard errors for continuous, dichotomous, and ordinal representations were converted as follows: (a) For the studies with continuous measurement of WTs, the β was calculated as log(RR), and the corresponding SE was calculated as (log(upper CI)−log(lower CI))/3.92. The unit of HR was converted into months before calculation. (b) For studies with only two WT groups, for which only crude rates of outcomes were provided, RR and CI were computed from 2 by 2 contingency table. The β was calculated as log(RR)/((x_n−x_0)*3.92), and the corresponding SE was calculated as (log(upper CI)−log(lower CI))/((x_n−x_0)*3.92), where x_n denotes the exposure at group n level and x_0 denotes the exposure at reference group. If only a p-value was provided, SE was calculated as the “test-based” method, SE = (log(RR))/Z_p, where Z_p is the value of a unit-normal test (e.g. Z_p = 1.96 if p = 0.05, two-tail test). (c) For the studies with more than two groups, weighted regression was used to estimate the β [38]. The SE was then computed employing the approach described in Greenland and Lognecker [39].

Meta-analyses

The adjusted regression coefficients from individual studies were combined using a fixed-effect model. The inverse-variance (1/SE^2) was used to weigh individual studies. The primary analysis was confined to HQ studies. A secondary analysis, including all relevant studies, was done to explore the impact of exclusion of studies which did not meet our criterion of adequately controlling for potential confounding factors. In our previous review, we did an a priori stratification by primary cancer site on the effect of delay on local control in the analysis. On reflection, we considered this to be inappropriate since the fundamental mechanism by which delay might affect local control was common to all types of cancer. We therefore analyzed the results of all studies of local control together, although we also did stratified analyses of specific types of cancer. When examining the relationship between WTs and overall survival, we did not combine the results of studies of different types of cancer.
of cancer, because it was self-evident that the impact of local failure on survival would mostly depend on the potential for surgical salvage which is dependant on primary site.

The homogeneity assumption in the meta-analyses was first assessed by the chi-square statistics:

\[ X^2_h = \sum \frac{w_i(b_i - E(b))}{E(b)} \]

(\(X^2_h\) is also referred to as Cochran's \(v^2\) [38]. This statistic is known as having high specificity but with low sensitivity [38]. The quantity \(I^2\) was then used to evaluate the extent of heterogeneity. This statistic has routinely been reported in Cochrane Reviews to help assess the consistency of meta-analyses since 2003 [40]. \(I^2\) is calculated as

\[ I^2 = 100\% \times \frac{(Q - df)}{Q} \]

where \(Q\) is the Cochran's \(v^2\) [40]. \(I^2\) can be interpreted as the proportion of total variation due to heterogeneity [41]. A value greater than 50% indicates there may be substantial heterogeneity [42]. The meta-regression approach was used to further explore the sources of heterogeneity [38].

To detect publication bias, we examined the asymmetry of standard error-based funnel plots using the linear regression method suggested by Egger et al. [43]. Briefly, the standard normal deviate (SND) is regressed against its precision. SND is defined as the effect size divided by its standard error, and precision is defined as the inverse of the standard error (regression equation: \(\text{SND} = a + b \times \text{precision}\)). The intercept \(a\) provides the quantity testing whether there is a statistically significant asymmetry. If \(a\) is different from 0, there is statistical significant asymmetry, otherwise there is no evidence of asymmetry. In addition, the ‘trim and fill’ approach was used to adjust the overall effect size to take account of publication bias [44,45]. Briefly, by assuming the symmetric distribution of studies in the funnel plot around the overall effect, the asymmetric studies indicated by L\(^*\) statistic were ‘trimmed’ off. The overall effect was re-estimated based on the remaining studies, and then the asymmetric ‘trimmed’ studies were used to generate their mirror image studies around the modified overall effect to make the funnel plot symmetrical. The overall effect was finalized by accounting for all the observed and the artificial studies.

**Results**

We identified 72 potentially relevant original reports, all of which were reviewed in detail. Forty-four studies, involving a total of 26,231 patients, reported on the relationship between WTs for RT, and one or more of the outcomes of interest. HQ studies involving 12,463 patients were identified and form the basis for the primary analysis [46–88]. Table 1 shows that the majority of these studies were conducted either in North America or in Europe, and most were published after 1995. The majority focused on breast

<table>
<thead>
<tr>
<th>Geographic area</th>
<th>Canada (8 (18.2%))</th>
<th>5883</th>
<th>8 (32.0%)</th>
<th>5883</th>
</tr>
</thead>
<tbody>
<tr>
<td>Europe</td>
<td>12 (27.3%)</td>
<td>14029</td>
<td>9 (36.0%)</td>
<td>4203</td>
</tr>
<tr>
<td>United States</td>
<td>18 (40.9%)</td>
<td>4538</td>
<td>5 (20.0%)</td>
<td>1861</td>
</tr>
<tr>
<td>Other</td>
<td>6 (13.6%)</td>
<td>1781</td>
<td>3 (12.0%)</td>
<td>516</td>
</tr>
<tr>
<td>Primary sites</td>
<td>Breast</td>
<td>24 (54.5%)</td>
<td>19469</td>
<td>11 (44.0%)</td>
</tr>
<tr>
<td></td>
<td>Head and Neck</td>
<td>14 (31.8%)</td>
<td>5091</td>
<td>9 (36.0%)</td>
</tr>
<tr>
<td></td>
<td>Sarcoma</td>
<td>2 (4.5%)</td>
<td>857</td>
<td>1 (4.00%)</td>
</tr>
<tr>
<td></td>
<td>Cervix</td>
<td>1 (2.27%)</td>
<td>200</td>
<td>1 (4.00%)</td>
</tr>
<tr>
<td></td>
<td>Lung</td>
<td>1 (2.27%)</td>
<td>340</td>
<td>1 (4.00%)</td>
</tr>
<tr>
<td></td>
<td>Brain</td>
<td>1 (2.27%)</td>
<td>182</td>
<td>1 (4.00%)</td>
</tr>
<tr>
<td></td>
<td>Various</td>
<td>1 (2.27%)</td>
<td>92</td>
<td>1 (4.00%)</td>
</tr>
<tr>
<td>Year</td>
<td>1975–1989</td>
<td>1 (2.27%)</td>
<td>436</td>
<td>1 (4.00%)</td>
</tr>
<tr>
<td></td>
<td>1990–1994</td>
<td>10 (22.7%)</td>
<td>2294</td>
<td>3 (12.0%)</td>
</tr>
<tr>
<td></td>
<td>1995–2000</td>
<td>21 (47.7%)</td>
<td>9989</td>
<td>11 (44.0%)</td>
</tr>
<tr>
<td></td>
<td>2001–2005</td>
<td>12 (27.3%)</td>
<td>13512</td>
<td>10 (40.0%)</td>
</tr>
<tr>
<td>Sample size</td>
<td>&lt;100</td>
<td>9 (20.5%)</td>
<td>528</td>
<td>3 (12.0%)</td>
</tr>
<tr>
<td></td>
<td>100–299</td>
<td>12 (27.3%)</td>
<td>2662</td>
<td>6 (24.0%)</td>
</tr>
<tr>
<td></td>
<td>300–499</td>
<td>8 (18.2%)</td>
<td>3234</td>
<td>6 (24.0%)</td>
</tr>
<tr>
<td></td>
<td>500–999</td>
<td>11 (25.0%)</td>
<td>7202</td>
<td>8 (32.0%)</td>
</tr>
<tr>
<td></td>
<td>≥1000</td>
<td>4 (9.09%)</td>
<td>12605</td>
<td>2 (8.00%)</td>
</tr>
<tr>
<td>Status of publication</td>
<td>Abstract</td>
<td>3 (6.82%)</td>
<td>2883</td>
<td>2 (8.00%)</td>
</tr>
<tr>
<td></td>
<td>Full-text</td>
<td>41 (93.2%)</td>
<td>23348</td>
<td>23 (92.0%)</td>
</tr>
<tr>
<td>Outcomes</td>
<td>LC/LF(^a)</td>
<td>38 (86.4%)</td>
<td>17073</td>
<td>20 (80.0%)</td>
</tr>
<tr>
<td></td>
<td>Metastasis</td>
<td>10 (22.7%)</td>
<td>6261</td>
<td>8 (32.0%)</td>
</tr>
<tr>
<td></td>
<td>Overall Survival</td>
<td>18 (40.9%)</td>
<td>13903</td>
<td>14 (56.0%)</td>
</tr>
</tbody>
</table>

\(^a\) LC/LF: Local Control/Local Failure.
Table 2

Characteristics of all the studies describing local recurrence

<table>
<thead>
<tr>
<th>Study</th>
<th>Cancer sites</th>
<th>Adjuvant chemo</th>
<th>Sequence of chemo and RT&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Roles of RT</th>
<th>Sample size</th>
<th>Measures of WT</th>
<th>HQ study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Froud, 2000 [50]</td>
<td>Breast</td>
<td>NO</td>
<td>Post-lumpectomy</td>
<td>Categorical</td>
<td>1962</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>Vujovic, 1998 [49]</td>
<td>Breast</td>
<td>NO</td>
<td>Post-lumpectomy</td>
<td>Continuous</td>
<td>568</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>Whelan, 1996 [47]</td>
<td>Breast</td>
<td>NO</td>
<td>Post-lumpectomy</td>
<td>Categorical</td>
<td>400</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>Fourquet, 1995 [46]</td>
<td>Breast</td>
<td>NO</td>
<td>Post-lumpectomy</td>
<td>Categorical</td>
<td>1839</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>Nixon, 1994 [53]</td>
<td>Breast</td>
<td>NO</td>
<td>Post-lumpectomy</td>
<td>Continuous</td>
<td>644</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>Bahena, 1998 [52]</td>
<td>Breast</td>
<td>NO</td>
<td>Post-operative</td>
<td>Categorical</td>
<td>623</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>Clarke, 1985 [48]</td>
<td>Breast</td>
<td>NO</td>
<td>Post-operative</td>
<td>Categorical</td>
<td>436</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>Benk, 2004 [76]</td>
<td>Breast</td>
<td>YES</td>
<td>CONC/C → RT</td>
<td>Continuous</td>
<td>482</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>Hebert-Croteau, 2004 [77]</td>
<td>Breast</td>
<td>YES</td>
<td>Not specified</td>
<td>Continuous</td>
<td>1004</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>Ampil, 1999 [51]</td>
<td>Breast</td>
<td>YES</td>
<td>C → RT</td>
<td>Continuous</td>
<td>26</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>Meek, 1996 [59]</td>
<td>Breast</td>
<td>YES</td>
<td>C → RT/SAND/CONC</td>
<td>Continuous</td>
<td>310</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>Wallgren, 1996 [74]</td>
<td>Breast</td>
<td>YES</td>
<td>C → RT</td>
<td>Continuous</td>
<td>433</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>Wallgren, 1996 [74]&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Breast</td>
<td>YES</td>
<td>C → RT</td>
<td>Continuous</td>
<td>285</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>Hartsell, 1995 [55]</td>
<td>Breast</td>
<td>YES</td>
<td>C → RT/SAND/CONC</td>
<td>Continuous</td>
<td>84</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>Leonard, 1995 [58]</td>
<td>Breast</td>
<td>YES</td>
<td>C → RT</td>
<td>Continuous</td>
<td>261</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>Slotman, 1994 [54]</td>
<td>Breast</td>
<td>YES</td>
<td>RT → C</td>
<td>Continuous</td>
<td>508</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>Metz, 2000 [87]</td>
<td>Breast</td>
<td>YES</td>
<td>C → RT</td>
<td>Continuous</td>
<td>221</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>Yock, 2004 [79]</td>
<td>Breast</td>
<td>YES</td>
<td>C → RT</td>
<td>Continuous</td>
<td>290</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>McCormick, 1996 [60]</td>
<td>Breast</td>
<td>YES</td>
<td>C → RT vs RT → C</td>
<td>Continuous</td>
<td>471</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>Recht, 1996 [61]</td>
<td>Breast</td>
<td>YES</td>
<td>C → RT vs RT → C</td>
<td>Continuous</td>
<td>244</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>Buchholz, 1993 [56]</td>
<td>Breast</td>
<td>YES</td>
<td>C → RT</td>
<td>Continuous</td>
<td>105</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>Nguyen, 1993 [86]</td>
<td>Breast</td>
<td>YES</td>
<td>C → RT</td>
<td>Continuous</td>
<td>187</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>Recht, 1991 [57]</td>
<td>Breast</td>
<td>YES</td>
<td>C → RT/SAND/CONC</td>
<td>Continuous</td>
<td>286</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>Leon, 2003 [82]</td>
<td>Head and Neck</td>
<td>NO</td>
<td>Definitive</td>
<td>Continuous</td>
<td>800</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>Fortin, 2002 [63]</td>
<td>Head and Neck</td>
<td>NO</td>
<td>Definitive</td>
<td>Continuous</td>
<td>623</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>Brouha, 2000 [65]</td>
<td>Head and Neck</td>
<td>NO</td>
<td>Definitive</td>
<td>Continuous</td>
<td>362</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>O'Sullivan, 1998 [62]</td>
<td>Head and Neck</td>
<td>NO</td>
<td>Definitive</td>
<td>Continuous</td>
<td>644</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>Barton, 1997 [66]</td>
<td>Head and Neck</td>
<td>NO</td>
<td>Definitive</td>
<td>Continuous</td>
<td>581</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>Lee, 1994 [64]</td>
<td>Head and Neck</td>
<td>NO</td>
<td>Definitive</td>
<td>Continuous</td>
<td>290</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>Marshak, 2004 [80]</td>
<td>Head and Neck</td>
<td>NO</td>
<td>Post-operative</td>
<td>Continuous</td>
<td>44</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>Suwinski, 2003 [81]</td>
<td>Head and Neck</td>
<td>NO</td>
<td>Post-operative</td>
<td>Continuous</td>
<td>868</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>Dixit, 1998 [67]</td>
<td>Head and Neck</td>
<td>NO</td>
<td>Post-operative</td>
<td>Continuous</td>
<td>61</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>Trotti, 1993 [69]</td>
<td>Head and Neck</td>
<td>NO</td>
<td>Post-operative</td>
<td>Continuous</td>
<td>30</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>Ampil, 1993 [70]</td>
<td>Head and Neck</td>
<td>NO</td>
<td>Post-operative</td>
<td>Continuous</td>
<td>70</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>Kajanti, 1991 [68]</td>
<td>Head and Neck</td>
<td>NO</td>
<td>Post-operative</td>
<td>Continuous</td>
<td>63</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>Schiff, 1990 [75]</td>
<td>Head and Neck</td>
<td>NO</td>
<td>Post-operative</td>
<td>Continuous</td>
<td>111</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>Balo, 2004 [84]</td>
<td>Sarcoma</td>
<td>YES</td>
<td>C → RT</td>
<td>Continuous</td>
<td>799</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>Schwartz, 2002 [85]</td>
<td>Sarcoma</td>
<td>YES</td>
<td>C → RT</td>
<td>Continuous</td>
<td>58</td>
<td>NO</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Abbreviation: C → RT: Chemo prior to RT; SAND: "Sandwich"; CONC: Concurrent; RT → C: RT prior to chemo.

<sup>b</sup> The Wallgren study was separated into two independent studies because two independent populations were examined.
cancer or head and neck cancer. Three reports were available in abstract form only. All the others were available as full-text manuscripts. Most studies reported on local control rates, many also reported on survival and some also reported rates of metastasis.

Table 2 describes the 38 studies which examined the relationship between WTs and the risk of local recurrence. We were able to calculate RR of local recurrence in 36 studies. This was not possible for the two remaining studies because no local failures were found in either one or other arm of these studies. Figs. 1 and 2 illustrate how the raw data from the original studies were converted into RR’s per month of delay. Fig. 1 shows the relationship between WTs and the rate of local recurrence in the eight HQ studies of breast cancer in which outcomes had been described in subgroups of patients defined by WT. This type of graph was first used for this purpose by Heinmüller [36]. The points show the local recurrence rates observed in subgroups of patients who waited for longer or shorter periods. The slopes of the line illustrates the strength of the association between WTs and the risk of local recurrence. In studies that reported the rate of local recurrence in only two groups of cases, the relationship between WTs and risk of local recurrence is represented by a straight line. In studies that had more than two groups, the slope is free to vary if the trend in risk is not constant over time. Most studies show an up-going trend, indicating an increase in the risk of recurrence with increasing WT. The baseline WT in the groups of cases that had waited the shortest period varied from study to study. The slopes of the lines which indicates the strength of the relationship between WTs and the risk of recurrence also varied, but there is no evidence of any systematic relationship between the baseline WT and the strength of the association between WTs and risk of recurrence. Thus the collective results of these studies do not suggest that there is a threshold below which delay has a lesser impact on the risk of local recurrence, although one study showed an initial downward trend followed by an upward trend. This question was addressed more quantitatively in the meta-regression (see below).

For the purposes of the meta-analyses it was assumed that the relationship between WTs and risk of local recurrence was constant over time. For studies with >2 subgroups, the overall RR was estimated by the weighted mean RR/month of each sequential pair of observations. Fig. 2 illustrates the RRs derived in this way from the original breast cancer studies described in Fig. 1. In Fig. 2, we have also shown the results of three additional breast cancer studies in which WTs had been treated as a continuous variable and the RR had been derived by regression analysis. The slopes of the lines in Fig. 2 represent the RR’s that were incorporated into the meta-analyses described below. The figure also shows the weighted average of the results of the 11 breast cancer studies along with the 95% confidence limits on this estimate derived from the meta-analyses.

Fig. 3 shows the RRs of local recurrence and 95%CI derived from all 20 HQ primary studies (all cancer sites), in the form of a conventional forest plot. The overall RR of local recurrence = 1.14 (CI: 1.09–1.21) per month of delay. The Cochran $\chi^2$ test showed significant heterogeneity ($\chi^2 = 34.6, \ p = 0.016$). We calculated $I^2$ as a measure of the heterogeneity among these 20 studies, and found a moderate heterogeneity ($I^2 = 45\%$) in comparison to 509 other meta-analyses in the Cochrane Database of Systematic Reviews [40]. Meta-regression showed no significant association between the primary cancer site ($p = 0.44$) or the baseline WT ($p = 0.32$), and the strength of the association between WTs and the risk of local recurrence.

Fig. 3 also shows the results of stratified meta-analyses in which studies were grouped based on the primary site of the cancer and the context in which RT was used (post-opera-
tive versus definitive RT). The RR\textsubscript{local recurrence} following post-operative RT for breast cancer was 1.11 per month of delay (95\%CI: 1.04–1.19). The value of RR\textsubscript{local recurrence} estimated from studies which included only patients who had not received chemotherapy (1.11, 95\%CI: 0.94–1.33) was the same as the value derived from studies that included patients who had received chemotherapy (1.11, 95\%CI: 1.03–1.19). The RR\textsubscript{local recurrence} following post-operative RT for head and neck cancer was 1.28 per month of delay (95\%CI: 1.08–1.52). The RR\textsubscript{local recurrence} following definitive RT for head and neck cancer was 1.15 per month of delay (95\%CI: 1.02–1.29).

Fig. 4 shows a funnel plot that illustrates that the degree of asymmetry of the individual study results around the pooled log(RR) of 0.13 for all 20 HQ studies combined. There appears to be some asymmetry. The degree of asymmetry was not statistically significant by Egger’s method [43] (p = 0.12), but this approach is known to be fairly insensitive [44,45]. We therefore used the ‘trim and fill’ approach to adjust our estimate of effect size for the observed asymmetry, on the assumption that it reflected publication bias. Four studies on the right side of the graph which were not matched by similar studies on the left were first identified as described by Duval and Tweedie [44,45]. These studies were then “trimmed”, i.e. excluded from the pool, and the RR was recalculated based on the residual studies. In the final step, these four studies were replaced in the pool along with four “mirror image” studies created to compensate for unreported studies. The meta-analysis was repeated to derive confidence limits on the adjusted RR. The corrected estimate of RR\textsubscript{local recurrence} was 1.12, 95\%CI: 1.06–1.18, indicating that the positive association between WTs and the risk of local recurrence was unlikely to be explained by publication bias.

Fig. 5 shows the results of secondary analysis in which we included the 16 additional studies that had been excluded from the primary analysis because they did not meet our HQ criteria for adequate control of possible confounding.

Please cite this article in press as: Chen Z et al., The relationship between waiting time for radiotherapy ...., Radiother Oncol (2007), doi:10.1016/j.radonc.2007.11.016
factor. The results of the meta-analyses of all 36 studies were similar to those of the primary analysis confined to the 20 HQ studies. For all 36 studies, the RRlocal recurrence was 1.13 (95%CI: 1.08–1.18), compared to 1.14 (95%CI: 1.09–1.21) for the 20 HQ studies. Thus there was no indication that the results of our primary analysis were affected by exclusion of the lower quality studies.

### Relationship between waiting time and risk of distant metastasis

Table 3 describes the 10 individual studies that explored the relationship between WTs and the risk of distant metastasis. Fig. 6 shows the RR per month of delay in RT in the eight HQ studies that met our criteria for adequate control for potential confounding factors in the form of a forest plot. Only in breast cancer were there sufficient studies to justify a meta-analysis. The $RR_{\text{metastasis/month}} = 1.04$, 95%CI: 0.98–1.09 for post-operative RT for breast cancer.

### Relationship between Waiting Time and Overall Survival

Table 4 describes the 18 studies that explored the relationship between WTs for RT and overall survival. Fig. 7 shows the $RR_{\text{death/month}}$ in the 14 HQ studies that met our criteria for adequate control for potential confounding factors. There was no significant association between WTs for RT and survival in the 4 HQ studies of breast cancer, $RR_{\text{death/month}} = 1.06$, 95%CI: 0.97–1.16. There was a marginally significant decrease in survival with increasing WTs in the 5 HQ studies of head and neck cancer $RR_{\text{death/month}} = 1.16$, 95%CI: 1.02–1.32. The single study that explored the relationship between delay and survival in cervical cancer showed a trend toward decreased survival with longer WTs, but this was not statistically significant.

### Relationship between waiting time and quality of life

We identified no studies that quantitatively examined the relationship between WTs for RT and quality of life.

### Discussion

The main finding of this study is that delay in starting RT is associated with an increase in the risk of local recurrence. This association has been shown to be statistically significant in breast cancer and in head and neck cancer. As might be predicted based on their respective growth rates, the association between WTs and the risk of local recurrence is stronger in head and neck cancer than in breast cancer, although that difference is not statistically significant. Consistent with the different clinical consequences of local failure in these two sites, there was also some evidence of a decrease in survival in head and neck cancer with increasing WTs, but less evidence of any similar association in breast cancer. There is no evidence to suggest that the relationship between delay in RT and local recurrence is unique to these two cancer sites. To date, there have been no studies with sufficient power to rule out the presence of an association of similar magnitude in other cancer sites. In contrast, we found no significant association between delay in RT and the risk of distant metastasis in any site. In the context of breast cancer, there was sufficient information available to suggest that any increase in the risk of metastasis must be relatively small compared to the increase in the risk of local recurrence.

We considered the possibility that the observed association between WTs and the risk of local recurrence might be due to selection bias. Although none of the original studies compared outcomes in groups of patients who had been randomized to longer or shorter delays, we did take steps to ensure that the comparison groups were either well balanced with respect to relevant prognostic factors, or that those factors had been adequately controlled for in the analysis. Moreover, there is little reason to suspect an imbalance between the groups that would bias the results in favor of those who had been treated more promptly. In general, doctors would be expected to select worse rather better cases for earlier treatment and this would lead to underestimation rather than overestimation of the magnitude of the association between delay and local recurrence. The absence of any significant association between delay and the risk of metastasis in breast cancer makes it particularly unlikely that the observed association between delay and local control is due to longer delays in the treatment of patients with more aggressive or advanced cancers. It remains possible that patients with head and neck cancer who have had more extensive surgery for more advanced cancers may have to wait longer for RT because of delayed wound healing and this might indeed create an artificial association between delay and local control. This effect could not, of course, explain the significant association between WTs and the risk of local recurrence in patients who receive definitive RT of head and neck cancer.
The possibility of publication bias was also considered. It seems likely that investigators would be more committed to publishing positive findings and that editors would be more likely to consider papers that report positive results. However, the funnel plot analysis showed only a modest degree of asymmetry providing little evidence of publication bias.

Fig. 5. Meta-analyses of all 36 studies of the relationship between WTs and local control. The forest plot shows the results of the meta-analyses of local recurrence including HQ and NHQ studies.
Moreover, adjusting for the observed asymmetry had only a small impact on the strength of the observed association between WTs and local control.

Although we tried to minimize the impact of bias on our results, we acknowledge that all the observational studies that were included in our meta-analyses would be classified as no better than level 3 or 4 evidence in Sackett’s typology of levels of evidence. Thus, when viewed through the lens of “evidence-based medicine”, our results might seem inconclusive. However, we believe that this is the wrong paradigm to apply in this context. Sackett, in fact, recommends a different approach to the evaluation of the potential harms of treatment which is much closer to the approach generally used in evaluating environmental hazards. Recognizing that it is usually impossible to do randomized clinical trials purely to evaluate potential harm to patients, he recommends that in appraising evidence in this context, we begin by asking the question, ’Was the type of study done the strongest that could have been performed under the circumstances?’ He then recommends that we apply Hill’s criteria for cause and effect relationships to the entire body of evidence that bares on the question. For the reasons outlined in the introduction, randomized trials cannot be done to evaluate the adverse effects of treatment delay and observational cohort studies represent the best source of direct evidence that delay is harmful. However, they are not the only source of relevant evidence. There is also strong indirect evidence that delay increases the risk of local failure, based on good clinical evidence that: (a) tumors progress while patients are waiting for RT; and (b) the probability of local tumor control by RT is dependant on the extent of the disease. Moreover, the relationship between tumor volume and local control is well explained by robust radiobiological theory, which has been validated experimentally. When Hill’s criteria are applied to the entire literature that links delay in RT to increased rates of local recurrence, the evidence that delay causes an increase in the risk of local recurrence appears to be essentially unassailable, at least in the context of breast and head and neck cancer.

Additional studies are still required to quantify the risks of

<table>
<thead>
<tr>
<th>Study</th>
<th>Cancer Site</th>
<th>Role of RT</th>
<th>Sample size</th>
<th>Measurement of WT</th>
<th>HQ study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hebert-Croteau, 2004</td>
<td>Breast</td>
<td>Post-lumpectomy</td>
<td>1004</td>
<td>Categorical</td>
<td>YES</td>
</tr>
<tr>
<td>Froud, 2000</td>
<td>Breast</td>
<td>Post-lumpectomy</td>
<td>1962</td>
<td>Categorical</td>
<td>YES</td>
</tr>
<tr>
<td>Wallgren, 1996</td>
<td>Breast</td>
<td>Post-lumpectomy</td>
<td>285</td>
<td>Categorical</td>
<td>YES</td>
</tr>
<tr>
<td>Wallgren, 1996</td>
<td>Breast</td>
<td>Post-lumpectomy</td>
<td>433</td>
<td>Categorical</td>
<td>YES</td>
</tr>
<tr>
<td>Meek, 1996</td>
<td>Breast</td>
<td>Post-lumpectomy</td>
<td>310</td>
<td>Categorical</td>
<td>NO</td>
</tr>
<tr>
<td>Nixon, 1994</td>
<td>Breast</td>
<td>Post-lumpectomy</td>
<td>644</td>
<td>Categorical</td>
<td>NO</td>
</tr>
<tr>
<td>Yock, 2004</td>
<td>Breast</td>
<td>Post-operative</td>
<td>290</td>
<td>Categorical</td>
<td>YES</td>
</tr>
<tr>
<td>Recht, 1996</td>
<td>Breast</td>
<td>Post-operative</td>
<td>244</td>
<td>Categorical</td>
<td>YES</td>
</tr>
<tr>
<td>Lee, 1994</td>
<td>Head and Neck</td>
<td>Definitive</td>
<td>290</td>
<td>Continuous</td>
<td>YES</td>
</tr>
<tr>
<td>Ballo, 2004</td>
<td>Sarcoma</td>
<td>Post-operative</td>
<td>799</td>
<td>Categorical</td>
<td>YES</td>
</tr>
</tbody>
</table>

Table 3

Characteristics of all the studies describing metastasis

- The Wallgren study was separated into two independent studies because two independent populations were examined.

Fig. 6. Meta-analyses of eight HQ studies describing the relationship between WTs and metastasis.

Moreover, adjusting for the observed asymmetry had only a small impact on the strength of the observed association between WTs and local control.

Although we tried to minimize the impact of bias on our results, we acknowledge that all the observational studies that were included in our meta-analyses would be classified as no better than level 3 or 4 evidence in Sackett’s typology of levels of evidence. Thus, when viewed through the lens of “evidence-based medicine”, our results might seem inconclusive. However, we believe that this is the wrong paradigm to apply in this context. Sackett, in fact, recommends a different approach to the evaluation of the potential harms of treatment which is much closer to the approach generally used in evaluating environmental hazards. Recognizing that it is usually impossible to do randomized clinical trials purely to evaluate potential harm to patients, he recommends that in appraising evidence in this context, we begin by asking the question, ”Was the type of study done the strongest
delay more thoroughly and to identify potential differences in risk among different cancer sites, but at present there is no evidence that delay is safe in any context.

For the purposes of policy development, absolute risks are more relevant than relative risks. We therefore converted the RRs derived from the meta-analyses into estimates of the...
increment in risk attributable to an increment of one month in the WT for RT. In these calculations, the mean of the local recurrence rates reported in the groups with the shortest WT in each study was used to estimate the baseline rate of local recurrence. The mean baseline rate of local recurrence following post-operative RT for breast cancer was 8.5%, with a range of 2.0–13.0%. The observed RR\text{local recurrence/month} = 1.11 for post-operative RT for breast cancer therefore translates into an absolute increase in the risk of recurrence of 1.0% per month of delay in starting RT. The mean baseline rate of local recurrence following post-operative RT for head and neck cancer was 22.7%, with a range of 9.9–25.5%. The observed RR\text{local recurrence/month} = 1.28 for post-operative RT for head and neck cancer therefore translates into an absolute increase in the risk of recurrence of 6.3% per month of delay. The mean baseline rate of local recurrence following definitive RT for head and neck cancer was 24.7%, with a range of 9.0–27.0%. The observed RR\text{local recurrence/month} = 1.15 for definitive RT for head and neck cancer therefore translates into an absolute increase in the risk of recurrence of 3.7% per month of delay.

Although the average increase in risk per month of delay in the individual patient is not large, it may have a very important detrimental effect on the overall value of an RT program because it potentially affects every patient who needs RT. As pointed out recently by Ravnsbaek et al. [5], one of the sad ironies of modern radiotherapy in some publicly funded health systems is that the negative effects of the prevailing delays in RT are probably sufficient to cancel out the positive effects of the many advances in radiotherapeutics of the last twenty years. The reverse of this is that shortening WTs represents a straightforward opportunity to improve local control rates across the board. An RT program with chronic waiting list could expect to achieve an absolute increase in local control of 5% and 10% in head and neck cancer simply by reducing its WT by 6 weeks.

How long, then, is it reasonable for patients to wait for RT? Given that there is no theoretical reason to believe that there is a threshold below which delay is safe, we believe that it is prudent to apply the principle that delays in RT should be As Short As Reasonably Achievable (ASARA), modeled on the ALARA principle which guides risk management in the field of radiation protection [23]. The involvement of a broad range of stakeholders is now required to translate that principle into context specific, operational guidelines.

Acknowledgements

This work was supported by a grant from Canadian Institutes of Health Research. The authors thank Drs. Vujovic, Froud, Olivotto and Kajanti, for their support and for providing additional information about their previously published case series.

* Corresponding author. William J. Mackillop, Division of Cancer Care and Epidemiology, QCRI Level 2, Queen’s University, Kingston, Ont., Canada. E-mail address: William.Mackillop@krcc.on.ca

Received 12 July 2007; received in revised form 13 November 2007; accepted 14 November 2007

References

[21] Shackney SE, McCormack GW, Cuchural Jr GJ. Growth rate of local recurrence. The mean baseline rate of local recurrence following post-operative RT for head and neck cancer therefore translates into an absolute increase in the risk of recurrence of 6.3% per month of delay. The mean baseline rate of local recurrence following definitive RT for head and neck cancer was 24.7%, with a range of 9.0–27.0%. The observed RR\text{local recurrence/month} = 1.15 for definitive RT for head and neck cancer therefore translates into an absolute increase in the risk of recurrence of 3.7% per month of delay.

Please cite this article in press as: Chen Z et al., The relationship between waiting time for radiotherapy ..., Radiother Oncol (2007), doi:10.1016/ j.radonc.2007.11.016


[29] Ipsos News Centre. Canadians agree that a patient wait time guarantee is the most important priority for Canada’s New Government. 11.29.2006.


