Titolo rivista/libro: International Journal of Radiation Oncology Biology Physics

Titolo articolo/sezione: Predicting late faecal incontinence risk after radiotherapy for prostate cancer: New insights from external independent validation

Autore/i: A Cicchetti

ISSN: 0360-3016

DOI:

Anno: 2018

Volume:

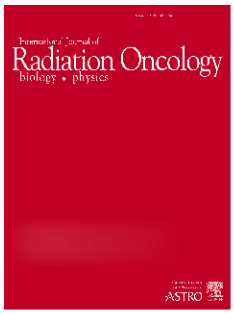
Fascicolo:

Editore:

Pag. iniziale:

Pag. finale:

Accepted Manuscript



Predicting late faecal incontinence risk after radiotherapy for prostate cancer: New insights from external independent validation

A. Cicchetti, B. Avuzzi, F. Palorini, F. Ballarini, C. Stucchi, G. Fellin, P. Gabriele, V. Vavassori, C. Degli Esposti, C. Cozzarini, C. Fiorino, T. Rancati, R. Valdagni

PII: S0360-3016(18)30806-X

DOI: 10.1016/j.ijrobp.2018.05.013

Reference: ROB 25000

To appear in: *International Journal of Radiation Oncology • Biology • Physics*

Received Date: 29 October 2017

Revised Date: 30 March 2018

Accepted Date: 2 May 2018

Please cite this article as: Cicchetti A, Avuzzi B, Palorini F, Ballarini F, Stucchi C, Fellin G, Gabriele P, Vavassori V, Esposti CD, Cozzarini C, Fiorino C, Rancati T, Valdagni R, Predicting late faecal incontinence risk after radiotherapy for prostate cancer: New insights from external independent validation, *International Journal of Radiation Oncology • Biology • Physics* (2018), doi: 10.1016/j.ijrobp.2018.05.013.

Predicting late faecal incontinence risk after radiotherapy for prostate cancer: New insights from external independent validation

A. Cicchetti[1,2], B. Avuzzi[3], F. Palorini[1], F. Ballarini[2,4], C. Stucchi[5], G. Fellin[6], P. Gabriele[7], V. Vavassori[8], C. Degli Esposti[9], C. Cozzarini[10], C. Fiorino[11], T. Rancati[1*] and R. Valdagni [1,3,12*]

*Tiziana Rancati and Riccardo Valdagni equally contributed to this work

[1]Fondazione IRCCS Istituto Nazionale dei Tumori, Prostate Cancer Program, Milan, Italy.

[2]Università degli Studi di Pavia, Physics Department, Pavia, Italy.

[3]Fondazione IRCCS Istituto Nazionale dei Tumori, Radiation Oncology 1, Milan, Italy.

[4] INFN-Section of Pavia, Pavia, Italy.

[5]Fondazione IRCCS Istituto Nazionale dei Tumori, Medical Physics, Milan, Italy.

[6]Ospedale Santa Chiara, Radiotherapy, Trento, Italy.

[7]Istituto di Candiolo- Fondazione del Piemonte per l'Oncologia IRCCS, Radiotherapy, Torino, Italy.

[8]Cliniche Humanitas-Gavazzeni, Radiotherapy, Bergamo, Italy.

[9]Ospedale Bellaria, Radiotherapy, Bologna, Italy.

[10]San Raffaele Scientific Institute, Radiation Oncology, Milan, Italy.

[11]San Raffaele Scientific Institute, Medical Physics, Milan, Italy.

[12]Università degli Studi di Milano, Department of Oncology and Hemato-oncology, Milan, Italy.

Running title: prediction of late faecal incontinence

Corresponding author:

Tiziana Rancati

Prostate Cancer Program

Fondazione IRCCS Istituto Nazionale dei Tumori

Via Venezian 1 – 20133 Milano, Italy

+39 02 23903009 tiziana.rancati@istitutotumori.mi.it

Conflict of interest statement: The authors have nothing to disclose

Summary: A model for prediction of late faecal incontinence was applied on a validation population finding substantial differences compared to the development cohort. Two out of three factors were confirmed as risk factors with similar Odds Ratios. Calibration plots showed a clear increasing probability of complication with the increase of dose. However, absolute toxicity rate was slightly underestimated, suggesting a possible role of hypofractionation beyond linear-quadratic model. Further possible explanations are discussed.

Predicting late faecal incontinence risk after radiotherapy for prostate cancer: New insights from external independent validation

Running title: prediction of late faecal incontinence

ABSTRACT

Aim: This study aimed to validate a previously published predictive model for late faecal incontinence (FI) in a contemporary population of prostate cancer patients treated with radical radiotherapy.

Materials and Methods: The validation included patients treated with IMRT (2010-2014). Prescribed dose range was 65-80Gy, including conventional and moderate hypo-fractionated treatments. Rectal toxicity was scored using LENT/SOMA, a minimum 2-year follow-up was considered. We chose to validate the model published by Rancati et al. for predicting chronic FI, developed on a 3DCRT population. It considered a longitudinal endpoint defined as the average toxicity grade during the follow-up. This continuous endpoint was dichotomized using a cut-off value of mean FI grade>1. The model included mean rectal dose (Dmean), previous diseases of the colon (COLO) and previous abdominal surgery (SURG). Doses were corrected to 2Gy/fraction using the linear-quadratic model and applying alpha/beta ratio=5Gy.

Results: 228 patients constituted the validation population. A mean FI grade>1 was scored in 25 patients (11%). Logistic regression confirmed risk factors reported in the literature, with similar Odds Ratios (ORs) for Dmean (1.05±0.03 vs 1.06±0.04) and SURG (1.90±1.70 vs 1.68±1.45); COLO was not confirmed. Consequently, the predictive models including Dmean/Dmean+SURG were evaluated using calibration plots. Both showed a clear discriminative trend, but the absolute observed toxicity rates were underestimated (i.e. absolute predicted rates were always lower than corresponding absolute observed rates). This result was consistent with an unexpected effect of hypofractionation (OR=2.20, conventional=8.1% vs hypofractionated=17.4%) beyond the standard correction using linear-quadratic model. Nevertheless, FI rate in the conventionally treated group was almost double than the one observed in the previously studied cohort (4.3% vs. 8.1%).

Conclusions: The study confirms previously published results indicating that abdominal surgery and rectal mean dose are risk factors for late FI. Calibration plots highlight a possible role of hypofractionation beyond linear-quadratic correction.

INTRODUCTION

Gastrointestinal toxicity has been the most recurring and disturbing toxicity after prostate cancer external beam radiotherapy up till the last decade [1]. Anorectal dysfunction occurred in a significant fraction of patients, with a non-negligible effect on their quality of life (QoL)[2]. In the conformal radiotherapy era, late rectal bleeding was the most frequent and investigated rectal toxicity symptom (results on a Pubmed search on number of papers about specific rectal symptoms in last 5/10 years are presented in Supplementary Material). With the coming of Intensity-Modulated (IMRT) and Image-Guided Radiotherapy (IGRT), and the simultaneous application of appropriate dose-volume constraints to the rectum [3,4] an important decrease in the insurgence of bleeding was observed (3-year incidence <15%, <10% and <5% for mild, moderate and severe events, respectively [3-5]).

Only recently have researchers begun to consider other symptoms, such as late faecal incontinence (FI) [3]. Even if FI occurs less frequently (≈5% of patients), it has an even stronger negative impact on QoL. Krol et al. [7] used the Expanded Prostate Index Composite Bowel Function (EPICB) to assess the impact of late anorectal dysfunction on QoL, and FI was the symptom with the largest impact (primarily related to embarrassment), while Bacon et al. showed that the level of worry caused by bowel morbidity was greater than that of sexual and urinary dysfunctions [8].

Trying to minimize late fecal incontinence is thus of high importance, in order to guarantee good durable QoL to the long-term prostate cancer survivors. Use of predictive models is an effective method to realize personalized treatment optimization. Some normal tissue complication probability (NTCP) models for late FI can be found in the literature [9-13], sometimes also including clinical modifying factors together with dosimetric features [9,10]. Their use in clinical practice is limited by the lack of external validation, which can establish their applicability in population other than the one used for model development, with particular interest in application in the IMRT domain. Development of validated models is essential to establish interventional studies aimed at changing patient management to reduce side-effects in cancer survivors.

In this study, we aimed at externally validating the NTCP model published by Rancati et al. [10], which considers a longitudinal definition of late FI. This definition is of clinical relevance, and also important for the social wellbeing of patients, as it can take both persistence and severity of incontinence symptoms into account. It was also shown it can better discriminate events clearly related to the radiotherapy with respect to events not directly due to radiotherapy [13,14]. The study was approved by Ethical Committee (INT 50/10)

METHODS AND MATERIALS

Validation population

The validation population consisted of patients treated in a multicentre setting between 2010 and 2014 in the frame of a prospective observational trial specifically designed to validate NTCP models for intestinal toxicity after exclusive external beam radiotherapy for prostate cancer. All patients were treated with radical intent by IMRT, with/without IGRT. The prescribed doses ranged between 65 and 80 Gy, including conventional (2 Gy/fr) and moderate hypofractionated (2.35-2.75 Gy/fr) schedules. Dose Constraints to the rectum used for treatment planning optimisation in the different centres are presented in Supplementary Materials. All doses were corrected to 2Gy-equivalent using the linear-quadratic model and applying an alpha/beta ratio of 4.8 Gy [15]. The linear-quadratic model in its formulation including treatment time correction was also considered, with gamma=0.7 Gy/day [15-16] to further diversify conventional and hypofractionated treatment involving the same 2Gy-equivalent doses.

Faecal incontinence scoring and endpoint definition

Toxicity was prospectively assessed before radiotherapy, at the end of treatment and every 6 months thereafter, till 2 year minimum follow-up. For this purpose, a self-reported questionnaire based on the LENT/SOMA system was used. The patient-reported questionnaire and the timing schedule were the same as used in the population considered for the NTCP model development [10].
The questionnaire scores incontinence as follows: Grade 1, unintentional stool discharge “sometimes” experienced; Grade 2, unintentional stool discharge “often” experienced or sporadically use of sanitary pads and Grade3, daily unintentional stool discharge or use of sanitary pads >2 times/week.

Endpoint definition followed the longitudinal characterization of late FI as presented by XXXX et al. [13]. Mean FI during follow-up was calculated as the average FI grade during the first 2 years after radiotherapy completion. Patients with at least three out four follow-up points were included in the analysis (the 2-year follow-up point was mandatory). The resulting synthetic score for the persistence and severity of incontinence symptoms is continuous, and can range from 0 to 3 (0 for patients registered with Grade 0 FI at each follow-up and 3 for patients with Grade 3 FI at each follow-up). Following the choice made for the original NTCP model development, a mean FI Grade>1 was considered as the toxicity endpoint

NTCP model for late faecal incontinence

The NTCP model for late FI proposed by Rancati et al. [10] is a logit model [17], with the toxicity probability given by:

$$Toxicity\ probability\ (Dmean) = \frac{1}{1+\left(\frac{Dmean_{50}}{Dmean}\right)^{k}} \quad [1]$$

where *Dmean* is the mean rectal dose, *k* is a parameter that determines the slope of the sigmoid dose-response curve and $Dmean_{50}$ is the mean rectal dose that results in 50% probability of experiencing late FI. In the model including only the mean dose, $Dmean_{50}$ has the same value for all patients, whereas in models including one or more clinical factors acting as dose-response modifiers, $Dmean_{50}$ takes on different values for patients with/without the clinical features. The dose modifying factor (dmf) is defined as the ratio of $Dmean_{50}$ for patients with/without the selected clinical feature.

Best fit parameters for the model including only mean rectal dose were: *k*=2 (68% confidence interval (CI): 0.7-2.8) and $Dmean_{50}$ 223.6 Gy (68% CI: 201.6-249.8 Gy). Parameters for the model with the inclusion of previous abdominal surgery were: *k*=2 (68% CI: 1.8-2.2), $Dmean_{50}$ 223.6 Gy (68% CI: 201.6-249.8 Gy) and dmf=0.73 (68% CI: 0.56-0.98). In the model with the inclusion of presence of colon disease *k*=2 (68% CI: 1.9-2.2), $Dmean_{50}$ is 223.6 (68% CI: 211.2-266.4) and dmf=0.64 (68% CI: 0.49-0.89).

Due to the presence of different radiotherapy schedules, mean doses were computed after linear-quadratic correction (without/with treatment time correction) of Dose-Volume Histograms (DVHs). A graphical interpretation of this procedure is reported in the Supplementary Material.

Comparison between development and validation population

External validation provides a measure of "generalizability" and "transportability" of the prediction model to cohorts that could be somewhat different from the one used for model development.

A summary of the characteristics of the here considered validation population and of the population (XXXXX trial [18]) used to fit the published NTCP models is shown in Table 1. The models were originally trained on a population of patients treated with three-dimensional conformal radiotherapy (3DCRT), whereas the external validation population consisted of patients irradiated with IMRT. Other differences were related to prescription doses (higher in the validation population) and to the fractionation scheme (30% patients in the validation population received moderate hypofractionation). Moreover, differences in geographical and temporal aspects were also encased in this validation study: patient were treated in different centres around Italy and in different time decades.

Figure 4, panel (a) shows the placement of this kind of study into the wide scenario of all possible validation analyses.

Statistical analysis

Validation of the effect size for the clinical and dosimetric features in the external validation population was performed by comparing the odds ratios (ORs, in the frame of univariate logistic analysis) for the development (XXXXXXX) and validation populations.
Replication of the NTCP models in the independent population was subsequently evaluated by calibration plot (calibration slope and R-squared), Brier score and Receiver Operating Characteristics Curve analyses.

RESULTS

In total 229 patients with 2-year follow up and dosimetric/clinical characteristics were available for the current analysis. The observed rate of mean FI Grade>1 in the first 2 years after radiotherapy completion was 10.9% (25/229 patients): 8.1% and 17.4% in the conventional fractionated and hypofractionated subpopulations, respectively (z-test for proportions, p=0.04, contingency table is shown in Supplementary Material). Details on the distribution of mean rectal dose (which is the relevant dosimetric feature in the considered model) are reported in figure 1.

The mean rectal dose and presence of previous abdominal surgery were confirmed as independent risk factors in the validation population, with ORs comparable to those previously published [10,13]. The OR for mean rectal dose was 1.06 (range 1.01-1.09) in the validation population vs 1.04 (range 1.01-1.07) in the developing set, while the OR for SURG was 1.6 (range 1-2.2) vs 1.9 (range 1.2-3.6).
The independent role of previous colonic disease was not confirmed in the validation population OR=0.3 (range 0.15-0.5) vs 2.7 (range 1.4-5.2): for this reason, the model including colon diseases was not considered for validation.

Figure 2 presents the calibration plots together with R-squared values for the considered models. The model exclusively including mean dose and that including abdominal surgery as dose-response modifying factor showed both a clear trend (i.e., increasing observed toxicity rates with increased predicted risk), but the absolute toxicity rates were underestimated (i.e. absolute predicted rates were always lower than corresponding absolute observed rates). Brier score values are shown in Supplementary Material together with its classical decomposition: reliability, resolution and uncertainty. The latter was the dominant coefficient in this validation study. Differences among models were exclusively affected by changes in the composition of the sample (conventional population only).
ROC curve analysis gave the same Area Under the Curve (AUC) of 0.64 in the development and validation population for the model including mean dose and previous abdominal surgery.

It may be hypothesized that this result could be related to a hidden effect of hypofractionation beyond the standard linear quadratic correction. As a matter of fact, hypofractionation results to be a risk factor in the validation population with OR=2.4 (range 1.6-3.7, FI rates 8.1% vs. 17.4%, in conventionally treated group vs hypofractionated patients).

Even the introduction of treatment time correction (Supplementary Material for details) did not completely explain the discrepancy between observed absolute toxicity rates and estimated rates (Figure 2c and 2d).

To try to understand the origin of this discrepancy, application of the NTCP models to the subset of conventionally treated patients (160 patients, FI rate 8.1%) was also tested. Calibration plots for this subgroup are reported in Figure 2e and 2f. Even in this case underestimation of observed toxicity rates is present.

Abdominal surgery– mean dose (surGy) conversion

The presence of previous abdominal surgery was converted into an equivalent effective dose, in order to have a better graphical view of the global underestimation. The conversion was computed starting from the dose-modifying factor for the logit model: a dose modifying factor for $Dmean_{50}$ of 0.73 corresponds to a modifying factor of 1.37 (=1/0.73) to be applied to the mean rectal doses of patients who underwent abdominal surgery before radiotherapy. In this way, we can consider patients with previous abdominal surgery has having an effective mean rectal dose which is 37% higher than the true physical mean dose (i.e. they receive dose X, but, due to the presence of the risk factor, this dose has the same effect of dose X·1.37 in a patient without the risk feature, Figure 3a).

When using the logit model, this shift in dose in not constant and increases with increasing physical mean doses. In the range of mean doses of the here considered validation population the shift is between 11 and 17 Gy. In this population, abdominal surgery is thus equivalent to have an “extra” mean dose of 11-17 Gy, which we chose to name 11-17 surGy to underline that this is not a physical dose, rather an effective dose describing a clinical risk factor. This effective dose was calculated for all patients who underwent abdominal surgery prior to radiotherapy. Figure 3b shows FI probability as a function of rectal effective mean dose together with observed toxicity rates in the validation population. This confirms the good ability of the model to describe increasing observed toxicity with increased effective mean dose. The offset between absolute observed and predicted late FI rates was also evident.

DISCUSSION

Currently, the validity and generalizability of toxicity prediction models is limited by a lack of external validation, which can ascertain their applicability in populations other than the one used in their training. Development of validated models is essential to establish interventional studies aimed at changing patient management to reduce side effects in cancer survivors.

In this work, we validated models for predicting late faecal incontinence by concentrating on a longitudinal definition that considers both the severity and duration of symptoms. Regardless the heterogeneity and differences between the two settings (training and validation population, Figure 4(a) and Table 1), several clinical implications can be drawn from the present analysis.

We investigated model validation in a particularly challenging situation, where generalization in treatment technique and fractionation schemes, together with geographic and time generalization, was required. Furthermore, the rate of FI in the validation cohort was more than twice as high as the rate in the development cohort (11% vs 4.3), the presence of prior bowel disease was twice as high in the validation cohort (12.7% vs 6.1%), and the rate of previous abdominal surgery was 4.6 times higher in the validation cohort (38.9% vs 8.4%). Nevertheless, we think that if we aim at gaining confidence that a previously developed model can somehow be used in the present clinical practice, we have to test it in new clinical settings, and determine which variables and relationships can be exported to new clinical settings and which features/relationships seem to be more related to peculiar settings. This process should be repeated across many different clinical settings, populations, and subgroups of interest. This procedure helps identify if and how updating or tailoring strategies can improve performance for particular settings, clusters or subgroups (rather than simply discarding the model). Of note, the predictive performance of a model tends to vary across settings, populations and periods. This implies that there is often heterogeneity in model performance, and that multiple external validation studies are needed to fully appreciate the generalizability of a prediction model.

A first important clinical observation is that the rate of late faecal incontinence did not decrease with the use of IMRT. In the XXXX training population, this rate was 4.3%, while in the recent IMRT population it was 8.1% in the conventionally fractionated subgroup and 17.4% in the hypofractionated patients. With respect to other rectal symptoms, we found a decrease in severe bleeding rate, an increase in pain and stool frequency and similar rates for mild/moderate bleeding and for acute toxicity. Detailed comparison of rectal symptom rates in the training cohort [3,4] and in the validation population [19] together with average values retrieved from literature [1, 20, 21, 22] is presented in Supplementary Material. This finding agrees with the studies of Wortel et al [23] and Al-Mamgani et al [24], who reported reduced rates for urinary toxicity and rectal bleeding using IMRT but no differences for faecal incontinence both as acute and late toxicity. This increase in FI rate is not explained by an increase in mean rectal dose (which is lower in

the IMRT population; Supplementary Material for details) or by known clinical risk factors, as demonstrated by the model calibration plots, with models correctly predicting increased toxicity rates with addition of risk factors, but still failing in estimation of absolute toxicity rates.

The published NTCP models proved to be robust with respect to confirmation of known risk factors such as rectal mean dose and presence of previous abdominal surgery, with concordance of the odds ratios in the two populations. These findings are also in agreement with several works published in the timeframe between the collection of data for model development and the validation study [12,14,21-24]. On the other hand, the presence of diseases of the colon was not confirmed as a risk factor.
The full NTCP models had a satisfactory calibration slope, indicating that it is a good tool for the selection of patients at higher risk of developing late FI, even if the absolute predictions were underestimated. Table 2 reports possible thresholds in mean rectal doses which could be suggested to discriminate patients at higher risk of late FI and to guide treatment optimization.

As already pointed out, calibration plots clearly showed a systematic underestimation of absolute observed FI rates. An initial hypothesis was that the presence of patients treated with hypofractionated regimens could partly explain this increased rate of late FI (hypofractionation was a risk factor in the validation population with OR=2.4), suggesting a role of larger daily doses beyond the one established by the linear quadratic model. Indeed, explicit introduction of hypofractionation as risk feature did not solve the discrepancy between absolute predicted and observed toxicity rates; moreover, this disagreement (even if smaller) was found also in the conventional subgroup.

Another possible explanation of underestimation could be the presence of a previously neglected risk factor. Irradiation of pelvic lymph node could play this role, due to the substantial difference in the fraction of patients with pelvic radiotherapy in the two populations (5% vs 22%, development vs validation population respectively). It is reasonable to think that dose to the bowel could play a role in increasing fecal control-like symptoms. Indeed, irradiation of lymph nodes resulted to be a risk factor in the IMRT group (OR=2.2, range 1.4-3.4, FI rate 8.9% vs 17.6%, patient without and with pelvic irradiation, respectively), but a model including this further risk factor (together with mean rectal dose and abdominal surgery) did not solved underestimation of toxicity rates (Supplementary Material for details).
A final check on application of the model to the conventional population without irradiation of lymph nodes also definitely revealed that the same underestimation was still present, detailed results in Supplementary Material.

A further investigated hypothesis was that the higher FI rates could be related to an effect of rectal volumes receiving high doses (above ≈78Gy), which could not be revealed by the older populations treated at lower prescription doses. However, these DVH regions were not significant risk factors (OR=1 for V75Gy and OR=0.96 for V80Gy).

Otherwise, in a complementary line of reasoning, the responsibility of the increased FI rates could be attributed to decreasing volumes of the spared structures into the pelvic region. Consistent with this hypothesis V5Gy was a risk factor with OR=1.05, but with p-value=0.15.

Another possibility could be an enhanced effect of IMRT spreading of low/medium doses (below ≈20-30Gy) into the pelvic floor, thus irradiating tissues/organs involved in the development of incontinence symptoms, such as small bowel, pelvic muscles, pudendal nerve or perirectal fat space [21,22,25-27]. Detailed analysis of doses in the peri-rectal space is needed to appreciate the validity of this hypothesis. A further work facing this aspect and the detailed analysis of dose-maps in the ano-rectal region is in progress.

In order to better understand if mis-calibration was due to differences between simulated rectal DVHs and accumulated rectal doses, sub-analysis was performed, stratifying patients with respect to presence of IGRT or not. The IGRT-population consisted of 128 pts (toxicity rate 1.7%), 63 were conventionally fractionated (toxicity rate 6.3%) and 65 had HF schemes (toxicity rate 17%). The no-IGRT population consisted mainly of conventionally fractionated patients (97/101, toxicity rate 9.3%). There was no difference in model calibration for the stratified populations, indicating no relevant effect of possible differences between planned and accumulated rectal doses. Details are reported in the Supplementary Material.

As a final consideration, Figure 4(b) depicts the possible course of validation studies as a function of complexity and robustness of the statistical approach. The present work could be placed in between the third and the fourth steps of this validation path, exhibiting comparable odds ratios and clear calibration trend, but a calibration slope>1 and failing calibration-in-the-large [28]. Additional investigation of other important factors is needed to re-gain calibration-in-the-large, together with further effort to model the possible effect of the hypofractionation beyond the linear-quadratic model, and possible role of organs/structures, other than the rectum and the anal canal, in insurgence of faecal incontinence symptoms.

Weakness of this study include the size of the validation population, which is only about one half of the population used for development, this in increasing uncertainties in model performance estimates. Another

possible limitation is related to differences in co-morbidity rates in the two populations, this heterogeneity could enhance the mis-calibration in absolute toxicity rates.

CONCLUSIONS

IMRT did not result in a decrease of incidence and severity of late faecal incontinence. The mean rectal dose resulted to be a validated dosimetric parameter associated with an increased risk of late FI, and the presence of previous abdominal surgery was confirmed to be a relevant dose-modifying factor.

The application of the models obtained by the 3DCRT era in modern RT practice shed light on possible effects of hypofractionation on radioinduced fecal incontinence, with fractionation correction following the linear-quadratic model apparently being insufficient to consider the effect of larger daily doses.

New scenarios were also opened, with possible need to consider doses outside the ano-rectal region in order to prevent late fecal incontinence.

Further investigation on larger prospectively followed populations is needed to confirm these results and to understand why late FI was not decreased in the recent IMRT population.

AKNOWLEDGEMENTS

XXXXXXXX.

REFERENCES

1. XXXXXXXXXXX
2. Stenmark MH, Conlon AS, Johnson S et al. Dose to the inferior rectum is strongly associated with patient reported bowel quality of life after radiation therapy for prostate cancer. Radiother Oncol. 2014 Feb;110(2):291-7. doi: 10.1016/j.radonc.2014.01.007.
3. XXXXXXXXXX
4. XXXXXXXXXX
5. Fonteyne V, Sadeghi S, Ost P, Vanpachtenbeke et al. Impact of changing rectal dose volume parameters over time on late rectal and urinary toxicity after high-dose intensity-modulated radiotherapy for prostate cancer: a 10-years single centre experience. Acta Oncol 2015;54: 854–61.
6. Gulliford SL, Foo K, Morgan RC et al. Dose–volume constraints to reduce rectal side effects from prostate radiotherapy: evidence from MRC RT01 Trial ISRCTN 47772397. Int J Radiat Oncol Biol Phys 2010;76:747–54.
7. Krol R, Smeenk RJ, van Lin EN, Hopman WP. Impact of late anorectal dysfunction on quality of life after pelvic radiotherapy. Int J Colorectal Dis. 2013 Apr;28(4):519-26.
8. Bacon CG, Giovannucci E, Testa M et al. The association of treatment-related symptoms with quality-of-life outcomes for localized prostate carcinoma patients. Cancer. 2002 Feb 1;94(3):862-71.
9. Defraene G, Van den Bergh L, Al-Mamgani A et al. The benefits of including clinical factors in rectal normal tissue complication probability modeling after radiotherapy for prostate cancer. Int J Radiat Oncol Biol Phys 2012;82:1233-1242
10. Rancati T, Fiorino C, Fellin G et al. Inclusion of clinical risk factors into NTCP modelling of late rectal toxicity after high dose radiotherapy for prostate cancer. Radiother Oncol 2011;100:124-130
11. Thor M, Olsson CE, Oh JH et al. Relationships between dose to the gastro-intestinal tract and patient-reported symptom domains after radiotherapy for localized prostate cancer. Acta Oncol 2015;54:1326-1334
12. Buettner F, Gulliford SL, Webb S et al. The dose–response of the anal sphincter region–an analysis of data from the MRC RT01 trial. Radiother Oncol 2012;103:347.
13. XXXXXXXXXXXX
14. Gulliford SL, Partridge M, Sydes MR et al. Parameters for the Lyman Kutcher Burman (LKB) model of Normal Tissue Complication Probability (NTCP) for specific rectal complications observed in clinical practise. Radiother Oncol 2012;102:347-351
15. Tucker SL, Thames HD, Michalski JM, et al. Estimation of α/β for late rectal toxicity based on RTOG 94-06. Int J Radiat Oncol Biol Phys. 2011 Oct 1;81(2):600-5.
16. Vogelius IR, Bentzen SM. Meta-analysis of the alpha/beta ratio for prostate cancer in the presence of an overall time factor: bad news, good news, or no news? Int J Radiat Oncol Biol Phys. 2013 Jan 1;85(1):89-94. doi: 10.1016/j.ijrobp.2012.03.004.
17. Fowler JF. 21 years of biologically effective dose. Br J Radiol. 2010 Jul;83(991):554-68. doi: 10.1259/bjr/31372149.
18. XXXXXXXXXXXXXX
19. XXXXXXXXXXXXXX
20. XXXXXXX

21. Di Franco R, Borzillo V, Ravo V et al. Rectal/urinary toxicity after hypofractionated vs conventional radiotherapy in low/intermediate risk localized prostate cancer: systematic review and meta analysis. Oncotarget. 2017 Mar 7;8(10):17383-17395. doi: 10.18632/oncotarget.14798.
22. Schaake W, van der Schaaf A, van Dijk LV et al. Normal tissue complication probability (NTCP) models for late rectal bleeding, stool frequency and fecal incontinence after radiotherapy in prostate cancer patients. Radiother Oncol. 2016 Jun;119(3):381-7. doi: 10.1016/j.radonc.2016.04.005.
23. Peeters ST, Hoogeman MS, Heemsbergen WD et al. Rectal bleeding, fecal incontinence, and high stool frequency after conformal radiotherapy for prostate cancer: normal tissue complication probability modeling. Int J Radiat Oncol Biol Phys. 2006 Sep 1;66(1):11-9.
24. Wortel RC, Incrocci L, Pos FJ et al. Late Side Effects After Image Guided Intensity Modulated Radiation Therapy Compared to 3D-Conformal Radiation Therapy for Prostate Cancer: Results From 2 Prospective Cohorts. Int J Radiat Oncol Biol Phys. 2016 Jun 1;95(2):680-9.
25. Al-Mamgani A, Heemsbergen WD, Peeters ST, Lebesque JV. Role of intensity-modulated radiotherapy in reducing toxicity in dose escalation for localized prostate cancer. Int J Radiat Oncol Biol Phys. 2009 Mar 1;73(3):685-91. doi: 10.1016/j.ijrobp.2008.04.063.
26. Loganathan A, Schloithe AC, Hutton J et al. Pudendal nerve injury in men with fecal incontinence after radiotherapy for prostate cancer. Acta Oncol. 2015 Jun;54(6):882-8. doi: 10.3109/0284186X.2015.1010693.
27. Gulliford SL, Ghose S, Ebert MA et al. Radiotherapy dose-distribution to the perirectal fat space (PRS) is related to gastrointestinal control-related complications. Clinical and Translational Radiation Oncology 2017;7:62-70
28. Moons KG, Altman DG, Reitsma JB et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015 Jan 6;162(1):W1-73. doi: 10.7326/M14-0698.

CAPTIONS TO FIGURES

Figure 1: Distribution (Box and whisker plot) of mean doses in the development (white) and validation (red) populations. All doses were corrected to 2Gy-equivalent using the linear-quadratic model and applying an alpha/beta ratio of 4.8 Gy [13] and including treatment time correction with gamma=0.7 Gy/day [13,14].

Figure 2: Calibration plots for the Normal Tissue Complication Probability models for the prediction of the endpoint "mean grade of late fecal incontinence ≥1". Calibration plots present rate of observed events in a group of patients (y-axis) vs mean predicted probability for the same group (x-axis). Groups of patients are ordered for increasing predicted probability. Error bars represent the confidence interval in observed frequencies as calculated from proportions in the study population and based on normal distribution of events. The left column of the figure reports calibration plots for models including only the mean rectal dose (panels a,c,e), while the right column presents calibration plots for models including mean rectal dose and presence of previous abdominal surgery as additional risk factors (panels b,d,f). The first row in the figure presents results for to models applied to the whole validation population, with correction of mean rectal doses using the linear-quadratic model and α/β=4.8 Gy, but no treatment time correction (panels a and b); the second row in the figure corresponds to models applied to the whole validation population, with correction of mean rectal doses using the linear-quadratic model and α/β=4.8 Gy, also including treatment time correction with gamma=0.7Gy/day (panels c and d). The third row presents results for the model applied to the subpopulation of conventionally treated patients (panels e and f).

Figure 3: (a) Graphical representation of the meaning of rectal effective mean dose for the patients with presence of previous abdominal surgery. A patient with abdominal surgery and mean dose of 33Gy has the same risk of late faecal incontinence as a patient without surgery and mean rectal dose of 45.2Gy (i.e. he has a 45.2Gy effective rectal mean dose, 45.2Gy=33Gy·1.37, using the modifying factor for presence of abdominal surgery). Examples for patients with surgery and mean rectal doses of 50Gy and 75Gy and for TD50 are also reported in the figure. (b) Probability of late longitudinal fecal incontinence, following the logit model reported by Rancati et al. [8] and as a function effective rectal mean dose (including correction for abdominal surgery). The continuous curve represents the predictive model, while symbols report observed toxicity rates in the validation population. Error bars represent the confidence interval in observed toxicity rates as calculated from proportions in the study population and based on normal distribution of events.

Figure 4: (a) Schematic description of possible validation trials and positioning of the present study into the wide panorama of the validation analyses (solid circular annuli), following TRIPOD Guidelines [22].

(b) graphical representation of possible course of validation studies as a function of complexity and robustness of the statistical approach. The present work could be placed in between the third and the

fourth steps of this validation path (black star in the figure), exhibiting comparable odds ratios and clear calibration trend, but a calibration slope >1 and failing calibration-in-the-large.

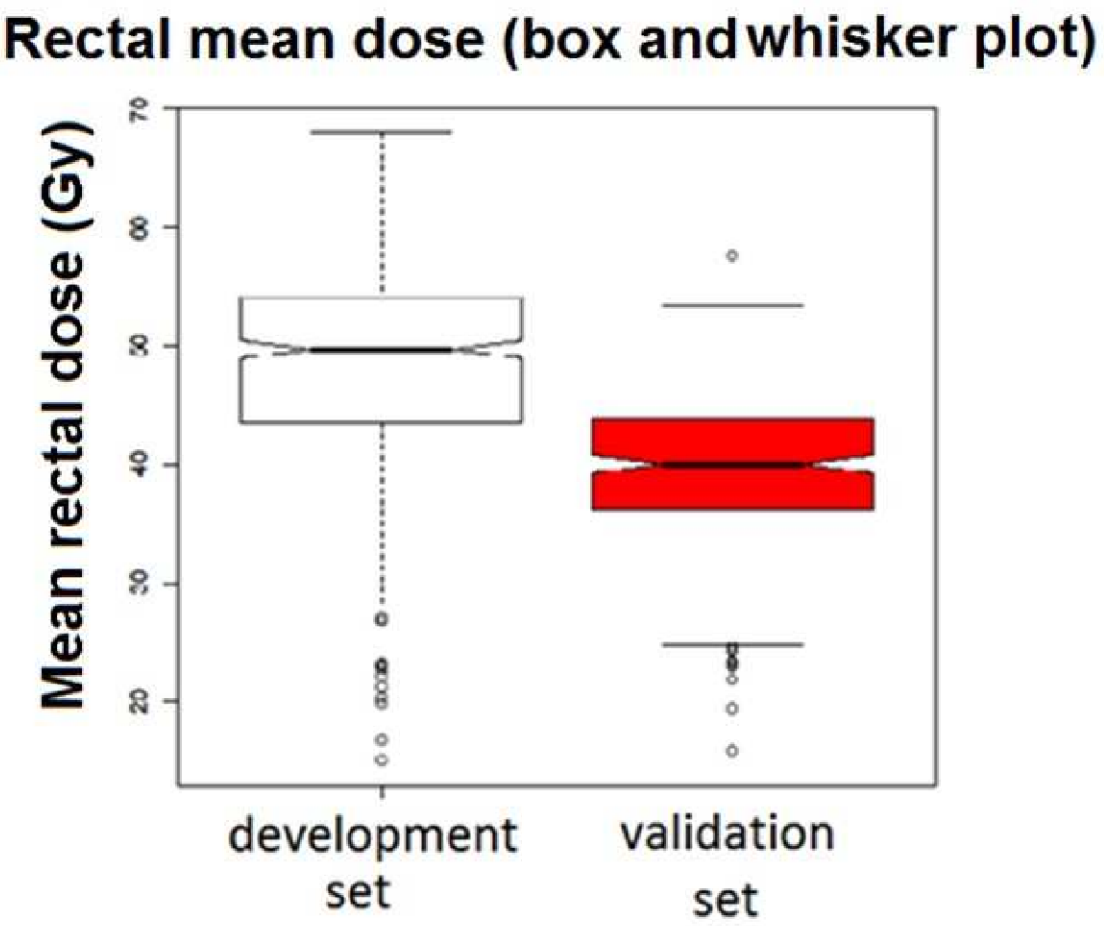
Tables

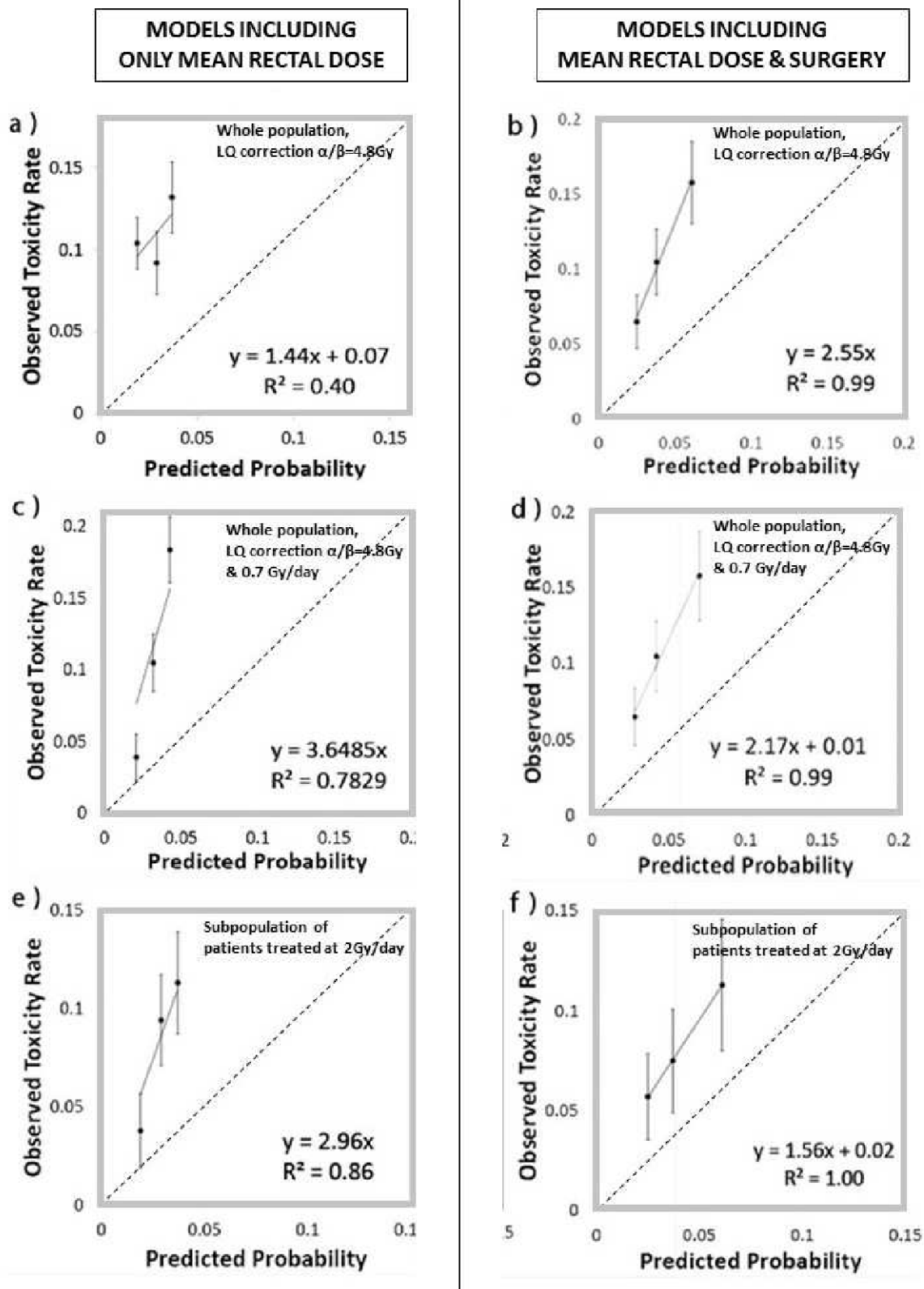
Table 1*:* Main characteristics of the development and validation populations.

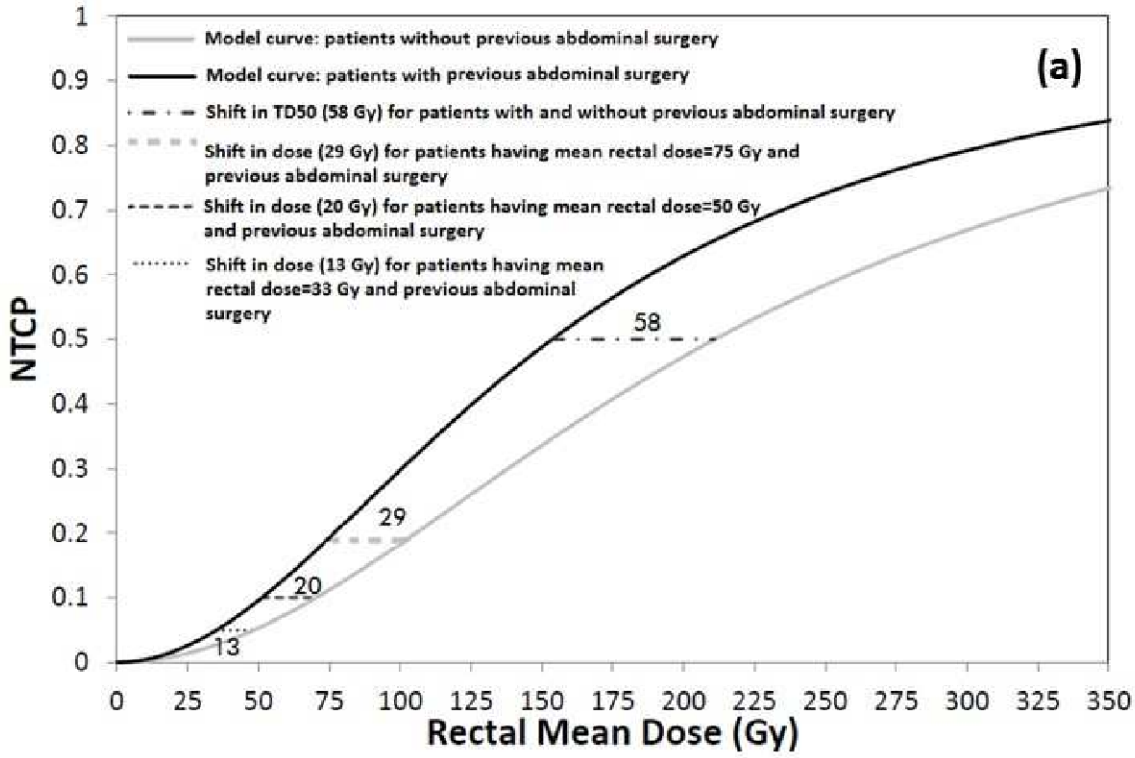
	Development population (XXXXXXXX)	Validation population (multi-centre setting)
Number of patients	506	228
Year of treatment	2002-2004	2010-2014
Location of Institutes	XXXXX	XXXXX
Average patient age (68% confidence interval)		71(65-77)
Radiotherapy technique	3DCRT	IMRT
Prescription dose range (Gy)	70-78	65-80
Prescription dose (corrected with α/β=3Gy) range (Gy)	70-78	72.8-83.9
Prescription dose (corrected with α/β=5Gy) range (Gy)	70-78	70.5-81.1
Dose/fraction (Gy/fr)	2 Gy	2.0 -2.75 Gy
Rate of longitudinal mean fecal incontinence grade >1 (absolute number of patients and %)	21 (4.3%)	25 (11%)
Average mean rectal dose corrected for linear-quadratic model without treatment time correction (Gy) (68% confidence interval)	44.0 (35-53)	37.7 (31.3-43.7)
Previous abdominal surgery rate (%)	8.4	38.9
Presence of bowel diseases rate (%)	6.1	12.7

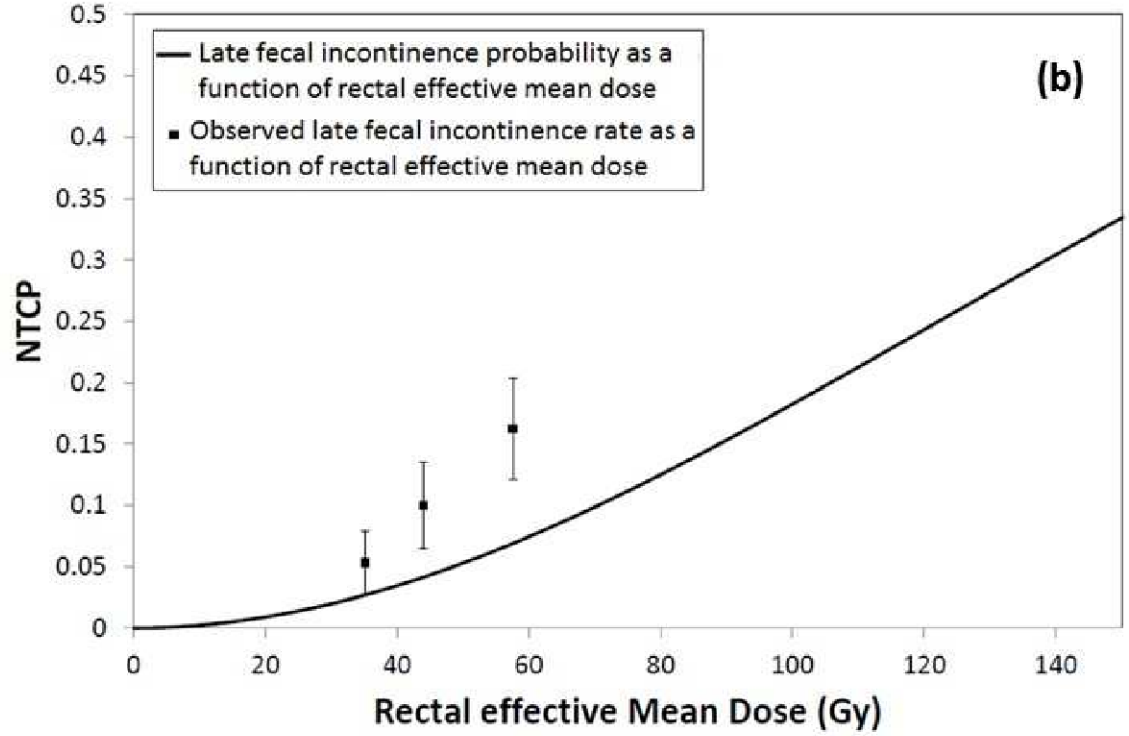
Table 2: possible thresholds in mean rectal doses which could be suggested to discriminate patients at higher risk of late faecal incontinence and to guide treatment optimization. Thresholds are reported for different populations. Faecal incontinence rates are evaluated in the validation population.

Group	**Overall faecal incontinence rate**	**Threshold in mean rectal dose**	**Faecal incontinence rate above dose threshold**	**Faecal incontinence rate below dose threshold**
All patients	11.0%	40 Gy	15.6%	5.6%
Patients without abdominal surgery	9.3%	40 Gy	12.2%	4.0%
Patients with abdominal surgery	13.5%	40 Gy	20.0%	6.8%
All patients (rectal effective mean dose)	11.0%	50 Gy	16.1%	8.3%


Rectal mean dose (box and whisker plot)
Mean rectal dose (Gy)
70
60
50
40
30
20
development set
validation set


MODELS INCLUDING ONLY MEAN RECTAL DOSE
MODELS INCLUDING MEAN RECTAL DOSE & SURGERY
a)
Whole population, LQ correction α/β=4.8Gy
y = 1.44x + 0.07
R² = 0.40
b)
Whole population, LQ correction α/β=4.8Gy
y = 2.55x
R² = 0.99
c)
Whole population, LQ correction α/β=4.8Gy & 0.7 Gy/day
y = 3.6485x
R² = 0.7829
d)
Whole population, LQ correction α/β=4.8Gy & 0.7 Gy/day
y = 2.17x + 0.01
R² = 0.99
e)
Subpopulation of patients treated at 2Gy/day
y = 2.96x
R² = 0.86
f)
Subpopulation of patients treated at 2Gy/day
y = 1.56x + 0.02
R² = 1.00
Observed Toxicity Rate
Predicted Probability


(a)
Model curve: patients without previous abdominal surgery
Model curve: patients with previous abdominal surgery
Shift in TD50 (58 Gy) for patients with and without previous abdominal surgery
Shift in dose (29 Gy) for patients having mean rectal dose=75 Gy and previous abdominal surgery
Shift in dose (20 Gy) for patients having mean rectal dose=50 Gy and previous abdominal surgery
Shift in dose (13 Gy) for patients having mean rectal dose=33 Gy and previous abdominal surgery
58
29
20
13
NTCP
Rectal Mean Dose (Gy)


(b)
Late fecal incontinence probability as a function of rectal effective mean dose
Observed late fecal incontinence rate as a function of rectal effective mean dose
NTCP
Rectal effective Mean Dose (Gy)