Evidence-based accuracy requirements in radiation oncology – tumor focus

Søren M Bentzen, Ph.D., D.Sc.
Departments of Human Oncology; Medical Physics; Biostatistics and Medical Informatics, University of Wisconsin Carbone Cancer Center, Madison, Wisconsin, USA

bentzen@humonc.wisc.edu
Accuracy is a quality, not a quantity

... so how DO we know when we are accurate enough?
Evolution of RT Data Collection Required for Advanced Technology Clinical Trials

Prior to 3DOG/RTOG 9406
- Film, paper forms (CRFs).
- Postal

3DOG/RTOG 9406 to present
- Data objects (typical patient accrual digital data set ~ 100 MB)
  - Volumetric, digital images (planning CT)
  - Contours (TV, OAR)
  - 3D dose distributions (fractionation)
- Participants mostly submit data digitally

2010+ Protocols
- Additional data objects needed for 2010 trials
  - Diagnostic imaging studies (pre- & post, MRI, MRS, PET/CT,...)
  - Treatment verif. images (planar, kV CBCT, MV CT, Calypso,...)
  - Treat. plans (parameters for dose recalc.)

Thanks to Jim Purdy
Welcome!

Who Should Attend?

This topic is extremely important for all those clinical physicists, radiation oncologists, and radiotherapy equipment manufacturers who are striving to improve the precision and accuracy of modern day radiotherapy.

Why Attend?

1. Radiotherapy is under scrutiny for mistreatments. What are you doing to better expend your quality management resources so that it doesn’t happen at your facility? The SS is a cost-effective choice.

2. What has the IAEA on IMRT said about how dose prescriptions should be reported? If you don’t know, you should attend the Summer School.

3. Rank the following in terms of likely impact to patient outcome. Correct answers in parentheses:

   - dose calculation accuracy (3)
   - calibration accuracy (4)
   - setup uncertainty (2)
   - uncertainty in contouring the gross tumor volume (1)

If you hesitated in your answer, you should attend the Summer School.
3. Rank the following in terms of likely impact to patient outcome.

- dose calculation accuracy (3)
- calibration accuracy (4)
- setup uncertainty (2)
- uncertainty in contouring the gross tumor volume (1)
Precision and accuracy: the target analogy

Not precise, NOT ACCURATE

[in metrology the blue darts would be seen as “accurate” = unbiased]

Precise (i.e. reproducible), but biased = NOT ACCURATE

Precise AND unbiased = ACCURATE
Statistics of radiation delivery

Mean Squared Error:

\[ MSE = \sigma^2 + bias^2 \]

Root Mean Squared Error:

\[ RMSE = \sqrt{MSE} \]
Comparing radiation dose deliveries

Deviation from planned dose [%]

Root Mean Square Errors

Probability density

0.32%

1.50%

2.04%
Dosimetric precision *in vivo*

Entrance and exit Si diode measurements in 11 patients with HNSCC

\*Leunens, ..., van der Schueren, *R&O* 25: 242 (1992)
"I guess I should warn you that if I turn out to be particularly clear, you've probably misunderstood me."

*Alan Greenspan at his 1988 confirmation hearings*
The normalized dose-response gradient

\[ \gamma_n = D \cdot \frac{dP}{dD} \approx D \cdot \frac{\Delta P}{\Delta D} = \frac{\Delta P}{\Delta D / D} \]

Response probability

Dose (Gy)

\( \Delta P \)

\( \Delta D \)

\( \text{P(D)} \)

\( D_{50} \)
Steepness of DR curves for HNSCC

[Graph showing steepness of DR curves for different anatomical regions.]
Steepness of normal-tissue dose-response curves

- Laryngeal edema
- Frozen shoulder
- Subcutaneous fibrosis
- Telangiectasia
- Telangiectasia
- Lung, early
- Lung, late
- Recto-sigmoid

\[ \gamma_{50} \]

HNSCC

Influenced by dose inhomogeneity (?)

Bentzen *R&O* 32: 1 (1994)
Sensori-neural hearing loss

\[ \gamma_{50} = 0.70 \text{ with } 95\% \text{ CI (0.22; 1.18)} \]

Sensori-neural hearing loss

\[ \gamma_{50} = 3.4 \text{ with } 95\% \text{ CI (0.3; 6.5)} \]

Adjusting for patient’s age and pre-RT hearing level

49 y.o.
43 dB pre-RT hearing level

$\gamma_{50}$ vs. local $\gamma$-value

- $\gamma$ varies with position on the dose-response curve, i.e. with the response level.
- The curve is still parameterized in terms of $D_{50}$ and $\gamma_{50}$
- However, in most situations the local $\gamma$-value should be applied
γ-value for dose-per-fraction escalation

- At a reference dose per fraction of $d_r$ the relationship between $\gamma_N$ and $\gamma_d$ is

$$\gamma_N = \gamma_d \cdot \frac{\alpha/\beta + 2 \cdot d_r}{\alpha/\beta + d_r}$$

- Two asymptotic results are

$$\lim_{d_r \to \infty} \gamma_N = 2 \cdot \gamma_d$$

$$\lim_{d_r \to 0} \gamma_N = \gamma_d$$

**EXAMPLE:**
Assume $\alpha/\beta = 2$ Gy

- $d_r = 2$ Gy \quad $\gamma_N = 1.5 \cdot \gamma_d$
- $d_r = 6$ Gy \quad $\gamma_N = 1.75 \cdot \gamma_d$

[Bentzen Acta Oncol 44: 825 (2005)]
Uncertainty components

The delivered dose can be decomposed as:

\[ \hat{D} = D_p + b + \varepsilon \]

where \( D_p \) is the planned (intended or acceptable to the physician) dose, \( b \) is the bias, and \( \varepsilon \) is a random error.
Estimating the clinical effect of imprecision

The expectation value of \( P(D) \) in the presence of variability in dose is
\[
( \_ \_ ) ( \_ \_ )
\]

The change in \( P(D) \) due to imprecision in dose delivery is
\[
( \_ \_ ) ( \_ \_ )
\]

The Taylor expansion of \( P(D) \) is
\[
( \_ \_ ) ( \_ \_ ) ( \_ \_ ) ( \_ \_ \_ ) - ( \_ \_ ) ( \_ \_ )
\]

The first order term cancels out in the convolution integral due to the symmetry of the p.d.f.
Second derivative of dose-response function

- Logistic dose-response curve: $D_{50}=60$ Gy, $\gamma_{50}=1.8$

\[ P''(D) \approx 21\% = \text{MAXIMUM INCREASE IN NTCP} \]

\[ P(D) \approx 79\% = \text{MAXIMUM LOSS OF TCP} \]
Loss of TCP with reduced precision

Around TCP=79%
Increase in NTCP with reduced precision

\[ \text{Precision, } \sigma_x \]

Around NTCP=21%
So, what’s the accuracy target then?

**Tumor control**

- **GOAL:** lack of accuracy results in <5% loss of tumor control probability
- **ASSUMPTION:** $\gamma_{50,N} = 3$
- What is $\gamma$ at the 80% level where the precision requirement is tightest?

![Graph showing local steepness vs response level with marked points $\gamma_{50} = 3$ and $\gamma_{80} = 2.1$.]
So, what’s the accuracy target then?

**Tumor control**

- **GOAL:** lack of accuracy results in <5% loss of tumor control probability
- **ASSUMPTION:** $\gamma_{50,N} = 3$
- What is $\gamma$ at the 80% level where the precision requirement is tightest?

**Answer:** $\gamma_{80,N} \approx 2.1$

- Accepting a maximum loss due to bias of 3%  
  bias 1.5%
- This caps the loss due to imprecision at 2%
Loss of TCP with reduced precision

NOTE: These curves refer to $\gamma_{50}$ NOT the local $\gamma$. 
So, what’s the accuracy target then?

**Tumor control**

- **GOAL:** lack of accuracy results in <5% loss of tumor control probability
- **ASSUMPTION:** $\gamma_{50,N} = 3$
- What is $\gamma$ at the 80% level where the precision requirement is tightest?
- Answer: $\gamma_{80,N} \approx 2.1$
- Accepting a maximum loss due to bias of 3%  
  ![Bias](bias.png)
- This caps the loss due to imprecision at 2%  
  ![Imprecision](imprecision.png)
- If treatment is delivered in, say, 30F this could be split as
  - $\sigma_{\text{course}}$ 4.5%
  - $\sigma_{\text{Fx}}$ 2.2%  
  - $\sigma_{\text{fx}}$ 12%
So, what if we gave 15 Gy x 3 instead?

**Tumor control**

- **GOAL:** lack of accuracy results in <5% loss of tumor control probability
- **ASSUMPTION:** tumor $\alpha/\beta = 8$ Gy, $\gamma_{50,d} = 2.5$ $\gamma_{50,N=3}=4.1$
- At the 80% level where the precision requirement is tightest $\gamma_{80,N}\approx2.9$
- Accepting a maximum loss due to bias of 3% $bias 1\%$
- This caps the loss due to imprecision at 2% $\sigma_c 3.9\%$
- As treatment is delivered in 3F this now yields
  
  $\sigma_{course} 3.5\%$
  
  $\sigma_{Fx} 1.7\%$ $\sigma_{fx} 3.0\%$
So, what about the NT accuracy target?

NTCP

- **GOAL:** lack of accuracy results in <5% increase in NTCP
- **ASSUMPTION:** $\gamma_{50,N} = 4$
- At the 20% level where the precision requirement is tightest $\gamma_{20,N} \approx 2.3$
- Accepting a maximum loss due to bias of 3% \( bias \ 1.3\% \)
- This caps the loss due to imprecision at 2% \( \sigma_c \ 4\% \)
- If treatment is delivered in, say, 30F this could be split as
  
  \[
  \sigma_{\text{course}} \ 3.5\% \\
  \sigma_{\text{Fx}} \ 1.9\% \quad \sigma_{\text{fx}} \ 10.6\%
  \]
Clinical outcome data from 1487 patients in the two 13F test arms of the START A trial


\[ \gamma_{eff} = 0.2 \]

\[ \gamma_{30} = 1.7 \]

...without adjusting for age, chemotherapy, tamoxifen, breast size, and surgical deficit

Yarnold, Bentzen et al. *IROBP* 79:1 (2011)
Missing a fraction of $D_p$ to a fractional volume

Target moved (or dose distribution moved)
TAKEHOME MESSAGE:
IF YOU MAKE MISTAKES,
MAKE SURE TO MAKE THEM RANDOMLY!

3. Rank the following in terms of likely impact to patient outcome.

- dose calculation accuracy (3)
- calibration accuracy (4)
- setup uncertainty (2)
- uncertainty in contouring the gross tumor volume (1)
EPILOGUE: Accuracy in a (humbling) perspective

% of patients who lost >3 Gy due to poor compliance

OVERALL: $\sigma_{\text{dose}}=9\%$

AT THE PATIENT LEVEL!!!
Co-factors in the expression of RT effects

<table>
<thead>
<tr>
<th>Endpoint</th>
<th>Risk factor</th>
<th>OR</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Various</td>
<td>Connective tissue disease</td>
<td>2.0</td>
<td>Hoelscher</td>
</tr>
<tr>
<td>Telangiectasia of the skin</td>
<td>Moist desquamation during RT</td>
<td>3.0</td>
<td>Bentzen</td>
</tr>
<tr>
<td>Radiological lung fibrosis</td>
<td>Tamoxifen during RT</td>
<td>3.2</td>
<td>Bentzen</td>
</tr>
<tr>
<td>Pneumonitis</td>
<td>Non-smokers/smokers</td>
<td>4.2</td>
<td>Liao</td>
</tr>
<tr>
<td>Sensori-neural hearing loss</td>
<td>5 years increment in patient’s age</td>
<td>10</td>
<td>Honore</td>
</tr>
<tr>
<td>Gastrointestinal injury RT for Hodgkin’s Disease</td>
<td>Previous laparotomy</td>
<td>12</td>
<td>Cosset</td>
</tr>
</tbody>
</table>

*$\gamma_{50}=3$, $D=70$ Gy, $p_0=10\%$
EPILOGUE: Accuracy in a (humbling) perspective

Variation in Mean Equivalent Lung Dose

18 patients with NSCLC receiving chemo-RT
Average MELD=10.2 Gy
$C_{\text{MELD}}=42\%$
Thanks to Fridtjof Nüsslin, Jan-Willem Leer, Jacques Bernier & Jake van Dyk for inspiring discussions on aspects of this topic.
Influence of dosimetric variation on $\gamma_{50}$

Inflation of sample size needed in RCT

\[ \gamma_{50,\text{biol}} \]

\[ \sigma_{\text{dose}} \]

1.05  1.10  1.15  1.20  1.30  1.50  1.80  2.0  3.0
# Intermediate/high risk prostate cancer

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Weight</th>
<th>Gamma (IV, Fixed, 95% CI)</th>
<th>Gamma (IV, Fixed, 95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-randomized studies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zelefsky 1998</td>
<td>4.2%</td>
<td>2.50 [1.09, 3.91]</td>
<td></td>
</tr>
<tr>
<td>Lyons 2000</td>
<td>3.2%</td>
<td>3.10 [1.47, 4.73]</td>
<td></td>
</tr>
<tr>
<td>Pollack 2000</td>
<td>2.5%</td>
<td>3.10 [1.28, 4.92]</td>
<td></td>
</tr>
<tr>
<td>Hanks 2000</td>
<td>10.8%</td>
<td>1.30 [0.42, 2.18]</td>
<td></td>
</tr>
<tr>
<td>Kupelian 2005</td>
<td>12.4%</td>
<td></td>
<td>0.89 [0.15, 0.63]</td>
</tr>
<tr>
<td><strong>Subtotal (95% CI)</strong></td>
<td>33.1%</td>
<td><strong>1.63 [1.13, 2.14]</strong></td>
<td></td>
</tr>
<tr>
<td>Randomized studies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zietman 2005</td>
<td>4.3%</td>
<td>1.60 [0.21, 2.99]</td>
<td></td>
</tr>
<tr>
<td>Peeters 2006</td>
<td>27.8%</td>
<td>0.65 [0.10, 1.20]</td>
<td></td>
</tr>
<tr>
<td>Dearnaley 2007</td>
<td>32.3%</td>
<td>0.90 [0.39, 1.41]</td>
<td></td>
</tr>
<tr>
<td>Kupan 2008</td>
<td>2.5%</td>
<td>0.40 [1.56, 5.24]</td>
<td></td>
</tr>
<tr>
<td><strong>Subtotal (95% CI)</strong></td>
<td>66.9%</td>
<td><strong>0.93 [0.58, 1.29]</strong></td>
<td></td>
</tr>
</tbody>
</table>

Test for subgroup differences: Chi² = 4.94, df = 1 (P = 0.03), I² = 79.8%

---

Diez, Vogelius, Bentzen *IJROBP 77*: 1066 (2010)
Dose-response stratified for *in vitro* radiosensitivity

**Graph:**
- **Y-axis:** Local control (%)
- **X-axis:** Biologic dose (Gy)
- Curves for 10%, 20%, 50%, 80%, and 90% local control

*References:*
- Bentzen *IJRB* 61: 417 (1991)
When is dose delivery precise enough?

Distribution of delivered dose

<table>
<thead>
<tr>
<th>CV(%)</th>
<th>D_5 (Gy)</th>
<th>TCP</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>68.73</td>
<td>50.0</td>
</tr>
<tr>
<td>5</td>
<td>71.76</td>
<td>62.9</td>
</tr>
<tr>
<td>2</td>
<td>72.85</td>
<td>67.2</td>
</tr>
<tr>
<td>1</td>
<td>73.01</td>
<td>67.9</td>
</tr>
<tr>
<td>0.1</td>
<td>73.07</td>
<td>68.1</td>
</tr>
</tbody>
</table>