AUTO-PLANNING
Promises, Pitfalls, and Preparations

Ben Nelms, Ph.D.
# LET’S GO FOR A DRIVE

<table>
<thead>
<tr>
<th><strong>1</strong></th>
<th><strong>PURPOSE</strong></th>
<th>The reason for this Sunday drive.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2</strong></td>
<td><strong>BACKGROUND</strong></td>
<td>The road behind us, and the road ahead.</td>
</tr>
<tr>
<td><strong>3</strong></td>
<td><strong>CHALLENGES</strong></td>
<td>Planning for potholes, traffic delays, and forks in the road.</td>
</tr>
<tr>
<td><strong>4</strong></td>
<td><strong>VARIOUS METHODS</strong></td>
<td>Choosing a vehicle.</td>
</tr>
<tr>
<td><strong>5</strong></td>
<td><strong>BIG(GER) PICTURE</strong></td>
<td>We have reached our destination: <em>a new starting line</em>.</td>
</tr>
</tbody>
</table>
PURPOSE

The reason for this Sunday drive.
The purpose of this talk:

- **IS NOT** to tell you that auto-planning is happening whether you like it or not.
- **IS** to envision a “best future” where auto-planning is
  - designed well and
  - implemented appropriately,
  - to the benefit of all stakeholders.
The purpose of this talk:

- **IS NOT** to tell you how auto-planning works and how it will be implemented.
- **IS** to help you start thinking about how you can play a role to help reach the best future that you and your team can imagine (and how to avoid bad futures).
The purpose of this talk:

- **IS NOT** to tell you how auto-planning will end your employment.
- **IS** to get excited about how auto-planning might transform your job.
The purpose of this talk:

- **IS NOT** to pretend to give all the answers.
- **IS** to get you asking the critical questions.
BACKGROUND

The road behind us, and the road ahead.
AN APT EXAMPLE (FROM OUTSIDE RAD ONC)

- Let’s talk about: the spreadsheet!
- The story of “VisiCalc”
  - “Spreadsheets!” Planet Money podcast (Feb 2015).
Accountants in the 1970s did all their calculations on paper.
One error would propagate and waste all the downstream work.
Laborious work, and expensive! (Required FTE accountants, or contracted work at a premium rate per hour.)
Then came Dan Bricklin’s daydream...
**An Apt Example (From Outside RAD Onc)**

Apple II’s video game paddle was used to move from “cell” to “cell.” (Later upgraded to using arrow keys.)

Cell value edited → effects on dependent cells were updated in real time.
AN APT EXAMPLE (FROM OUTSIDE RAD ONC)

- BIG Impact
  - *Increased accuracy, lower costs.* Better, cheaper.
  - *Boon for businesses* of all kinds.
  - *Did it eliminate jobs?* YES, some. The number of bookkeepers and accounting clerks fell; they were replaced by spreadsheets.
  - *Did it make jobs?* YES, better ones. The volume of accounting work actually increased, i.e. clients contracted out more accounting projects. They got hooked on the power of data and running scenarios.
The Atlantic
THE ROBOT
Will See You Now
Is Your Doctor Becoming Obsolete?
By Jonathan Cohn
March 2013
Is Auto-Planning a New Thing?

- Catalyst: the ability to deliver intensity modulated radiation.
- Computers are much better at optimizing complex beam patterns than humans.
- ➔ Inverse planning, i.e. asking computer to do the complex problem that is impractical to do manually.
Is Auto-Planning a New Thing?

- Beam directions selected by planner
- Beam shape based on BEV anatomy
- Beams added (and beam angles tweaked) until cumulative dose is deemed satisfactory

- Beam directions based on proven template or optimized as parts of a rotational arc
- Beam’s intensity is modulated based on computerized optimization
- Optimization continues until acceptance
Is Auto-Planning a New Thing?

- Computer-aided planning is not a new thing.
- Computer-optimized plans are not a new thing.
- In this regard, **auto-planning is not a new thing.**
- But... **Auto-planning today is not automated.**
Inverse Planning: Not Automated

- IMRT and VMAT planning may take a long time, more than conventional 3D planning.
- Lots of user interaction required to
  - Steer the optimizer
  - Iterate-and-tweak
  - Manage objectives & assess results
- High variability in the quality of output plans
A FEW RELEVANT CONCLUSIONS

- There is large variability in plan quality:
  - For all modalities.
  - For all TPS models.
  - For all educational degrees and work experience.
- There are very few instances of statistical difference between sub-populations.
- Therefore, as it stands today:
  - Plan quality is mostly determined by planner skill.
  - (And remember, this is for inverse planning.)
**Lowest Observed Score**
- **Modality:** VMAT
- **TPS:** Eclipse
- **MLC:** 120-leaf

**Highest Observed Score**
- **Modality:** VMAT
- **TPS:** Eclipse
- **MLC:** 120-leaf
GOALS #1-2 OF AUTO-PLANNING

**CURRENT MODE**
- High variation
- Average quality is low
- Lots of low quality items
- Few high quality items

**NEW MODE**
- (1) Lower variation
- (2) Higher quality
- Fewer low quality items
- More high quality items
YOU HELPED! 2011 AAMD PLAN CHALLENGE

Variation in external beam treatment plan quality: An inter-institutional study of planners and planning systems
Benjamin E. Nelms PhD,a,b,* Greg Robinson CMD,c Jay Markham CMD,c Kyle Velasco CMD,c Steve Boyd CMD,c Sharath Narayan CMD,c James Wheeler MD, PhD,d Mark L. Sobczak MD,e

Alredy cited by 57 papers in the last five years.
33 of those papers are about auto-planning.
GOALS #3-4 OF AUTO-PLANNING

- **$ Reduce costs**
  - Assumption is that virtual / automated planning will cost less over time than human labor.

- **Increase throughput**
  - A computerized/virtual workforce can work around the clock (i.e. duty cycle approaches 100%)
  - Higher workload accommodated by simply adding more processing power.
## Goals vs. Stakeholders

<table>
<thead>
<tr>
<th>Goals</th>
<th>Investors &amp; Executives</th>
<th>Payors</th>
<th>Physicians</th>
<th>Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher Quality</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Lower Variability</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Lower Cost</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Higher Throughput</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
Within the next ten years treatment planning will become fully automated without the need for human intervention

Michael B. Sharpe Ph.D., Kevin L. Moore Ph.D., Colin G. Orton Ph.D.

First published: 10 November 2014  Full publication history
DOI: 10.1118/1.4894496  View/save citation
Salient Points For/Against the Proposition

- Arguments **FOR** (the prediction of full automation within 10 years), MB Sharpe
  - “Within a decade, radiation treatment planning will become fully automated without the need for human intervention because (i) we will exploit pertinent trends in the manufacturing and informatics industries, (ii) the precedent is already established, and (iii) it is imperative to improving quality and continuing advancements in care.”
  - “Today...advanced functions are automated, such as image registration, organ delineation, and dose optimization. Using commercial tools, it is now possible to control workflow so as to fully create, evaluate and document a plan with minimal intervention.”
  - Generate automated plan with “robust quality” designed in; and then tailor, or personalize, per patient via adaptive methods.
Arguments **AGAINST** (the prediction of full automation within 10 years), KL Moore

- “I ... contend that the odds of fully automated treatment planning being the norm in ten years’ time must be rated as extremely unlikely.”
- (On difficulties posed by contouring) “Impressive though the last decade has been for the field of autosegmentation, it strains credulity that a decade's time would be enough to herald a universal autosegmentation platform that not only identifies all normal anatomical structures across all imaging modalities but also flawlessly incorporates every patient's unique clinical circumstances into fully automated tumor volume contouring.”
- (A revised prediction) “Semiautomated (i.e., computer-assisted) treatment planning will be used in the large majority cases, with some form of knowledge-based and/or computer-aided multicriterial optimization removing most of the present-day human variability from the optimization process.”
Planning for potholes, traffic delays, and forks in the road.
CHALLENGE [1]: CONTOURING

Inter-observer variability in rectum contouring
CHALLENGE [1]: CONTOURING

Inter-observer variability in rectum contouring (total volume)
Challenge [1]: Contouring

Inter-observer variability in rectum contouring (overlap)
CHALLENGE [1]: CONTOURING

Inter-observer variability in hippocampus
CHALLENGE [1]: CONTOURING

Inter-physician variability in left hippocampus contouring (total volume)
CHALLENGE [1]: CONTOURING

Inter-physician variability in left hippocampus contouring (overlap)
QUESTION: How can you automate contouring when there is no standard of accuracy?

QUESTION: What is the relative gain from ever-improving technology (plan optimization, auto-planning, treatment delivery, IGRT, QA, etc.) if there is rampantly inconsistent anatomy modeling?
CHALLENGE [2]: PLAN QUALITY MEASURES

“We have a comprehensive list of plan metrics and objectives for this patient. Each is prioritized and weighted so that we can render an objective plan score.

Okay, now create for me the best plan!”
**Challenge [2]: Plan Quality Measures**

**Question:** Does the auto-planning system incorporate (for inputs and outputs) an objective plan scoring system so that the output quality is inherently verified?

**Question:** Can you get a physician team to agree on plan quality metrics and objectives? That is, can you drive consensus on: “What is a good plan?”
Challenge [3]: Validation

“We have a vast array of test data, covering all the types of cases we intend to plan with the auto-planning engine.

We want to learn how well it works (or doesn’t work) for each case type, so we can implement safely and with confidence.”

Input Data
Patient data and planning goals

Output Plans
How did the system perform (objective comparisons of new vs. current method)?
Where did it fail?
Challenge [3]: Validation

- This is a good model to follow.

**Journal of Applied Clinical Medical Physics**

*Initial evaluation of automated treatment planning software*

Dawn Gintz, Kujtim Latifi, Jimmy Caudell, Benjamin Nelms, Geoffrey Zhang, Eduardo Moros, Vladimir Feygelman

*First published: 8 May 2016*
**Challenge [3]: Validation**

**Question:** What is your plan to validate the auto-planning method to measure how well it performs? And for which subsets of patient plan types?

**Question:** Do you have a robust enough test suite? (Patient datasets and plan scoring strategies)

**Question:** Is it a cross-disciplinary team effort, or is the validation team in a silo? If in a silo, is it the right silo?
**Challenge [4]: Personalization**

“Okay team, we have a unique set of goals for this particular patient.

Start with protocol XYZ but change metrics/goals 1, 2, and 3…”

**Question:** How well does the auto-planning solution work when you don’t use “one size fits all” treatment plan goals? Does it perform just as well for goals personalized for any given patient?
"Hey, the patient anatomy has changed appreciably from the initial fraction.

Re-plan it based on the changes. And by the way, tweak A, B, and C, too. And deform the new volumes to re-estimate cumulative DVH.

Plan "i" (replaces Plan$_0$ or Plan$_{i-1}$)

**QUESTION:** Where do the new (adjusted) anatomy contours come from? What about the 3D density matrix?

**QUESTION:** Is it fast enough for "online" adaptive? Or do you plan "offline" adaptive only?
As your department drives towards auto-planning, do so with open eyes and following a clear path.
VARIOUS METHODS

Choosing a vehicle.
Specific methods are interesting, keep your eye on the prize.

- Measured Results vs. Agreed-Upon Goals.

"The proof of the pudding, is in the eating."
If the high-level goals are:
• Higher quality, lower variability
• Higher efficiency / throughput
• Lower (and predictable) costs

Then does it matter how you achieve them?
A. Virtual treatment planner (e.g. software), or
B. Improved system still powered by humans, or
C. Combination of both?
**KEEP YOUR EYE ON THE PRIZE**

- Analogy: Commissioning the accuracy of a TPS dose calculation algorithm
  - Judge not by impressive terminology or fancy brand names; **judge by performance**
    - Result vs. Reliable Standard
    - Calculated dose vs. measured dose
  - Analyzed **across the spectrum of situations** (patient shape, size, densities; delivery hardware; couch, immobilization etc.)
[1] “Knowledge-Based”

- Collect a history of “good plans” for past patients
- History of plans categorized into bins (machine, energy, modality, etc.)

Auto-Planning engine uses achievements of past plans to predict what can be done for new patients, then optimizes to that prediction.
[2] SMARTER, PATIENT-SPECIFIC INPUTS

Use physics-based algorithms to predict, based only on anatomy, the feasibility limits for organs-at-risk (given specific target Rx).

TPS Optimizer is now challenged to improve beyond where you might have stopped otherwise (or stop you from trying to do the impossible).
A novel method for a priori estimation of best feasible DVH for organs at risk: Validation for Head and Neck VMAT planning

(Conditionally accepted to Medical Physics, upon final submission)
[3] The Online Plan Factory (cont.)

- What does a “plan factory” look like?
  - A virtual group; **online community**
  - **Proven skills**, “certified” per plan type
  - **Cloud-based system** for (1) posting jobs (patient data, plan requirements, constraints) and (2) accepting jobs and uploading plan outputs.
  - Integrated (and mutual) **ranking system** to help other users

- In other words, what does it look like?
  
  *Like everything else does these days.*
Imaging
Anatomy modeling

Define plan objectives/scoring
Submit data and requirements
Receive plan data (plan quality documented)

Verify dose calc and delivery
Approve treatment

- Plan scoring uses the same (online) method as you would locally.
- Plan scoring explicitly based on your defined plan objectives.
## Strengths / Weaknesses

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Knowledge Based</th>
<th>Smarter Inputs</th>
<th>Online Plan Factory</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1] Contouring Accuracy</td>
<td>WEAKNESS</td>
<td>WEAKNESS</td>
<td>WEAKNESS</td>
</tr>
<tr>
<td>Consensus on what are correct contours &amp; automating contouring to high accuracy for all patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[2] Objective Measures of Plan Quality</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agreed-upon plan objectives &amp; plan scoring across all physicians for all patients</td>
<td>??</td>
<td>??</td>
<td>??</td>
</tr>
<tr>
<td>[3] Validation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proving when, where, and how well it works</td>
<td>Dependent on above</td>
<td>Dependent on above</td>
<td>Dependent on above</td>
</tr>
<tr>
<td>[4] Personalization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Can you feed unique, patient-specific goals into the method?</td>
<td>???</td>
<td>STRENGTH</td>
<td>STRENGTH</td>
</tr>
<tr>
<td>Can the method be used to quickly re-design plan based on changes (e.g. anatomy)</td>
<td>???</td>
<td>???</td>
<td>WEAKNESS</td>
</tr>
</tbody>
</table>
## Without Objective Measures of Plan Quality

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Knowledge Based</th>
<th>Smarter Inputs</th>
<th>Online Plan Factory</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Contouring Accuracy</strong></td>
<td>WEAKNESS</td>
<td>WEAKNESS</td>
<td>WEAKNESS</td>
</tr>
<tr>
<td>Consensus on what are correct contours &amp; automating contouring to high accuracy for all patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Consensus Measures of Plan Quality</strong></td>
<td>WEAKNESS</td>
<td>WEAKNESS</td>
<td>WEAKNESS</td>
</tr>
<tr>
<td>Agreed-upon plan objectives &amp; plan scoring across all physicians for all patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Validation</strong></td>
<td>WEAKNESS</td>
<td>WEAKNESS</td>
<td>WEAKNESS</td>
</tr>
<tr>
<td>Proving when, where, and how well it works</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Personalization</strong></td>
<td>???</td>
<td>STRENGTH</td>
<td>STRENGTH</td>
</tr>
<tr>
<td>Can you feed unique, patient-specific goals into the method?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Application to Adaptive RT</strong></td>
<td>???</td>
<td>???</td>
<td>WEAKNESS</td>
</tr>
<tr>
<td>Can the method be used to quickly re-design plan based on changes (e.g. anatomy)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**WITH OBJECTIVE MEASURES OF PLAN QUALITY**

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Knowledge Based</th>
<th>Smarter Inputs</th>
<th>Online Plan Factory</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Contouring Accuracy</strong></td>
<td>WEAKNESS</td>
<td>WEAKNESS</td>
<td>WEAKNESS</td>
</tr>
<tr>
<td>Consensus on what are correct contours &amp; automating contouring to high accuracy for all patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Consensus Measures of Plan Quality</strong></td>
<td>STRENGTH</td>
<td>STRENGTH</td>
<td>STRENGTH</td>
</tr>
<tr>
<td>Agreed-upon plan objectives &amp; plan scoring across all physicians for all patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Validation</strong></td>
<td>STRENGTH</td>
<td>STRENGTH</td>
<td>STRENGTH</td>
</tr>
<tr>
<td>Proving when, where, and how well it works</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Personalization</strong></td>
<td></td>
<td>STRENGTH</td>
<td>STRENGTH</td>
</tr>
<tr>
<td>Can you feed unique, patient-specific goals into the method?</td>
<td>???</td>
<td>STRENGTH</td>
<td></td>
</tr>
<tr>
<td><strong>Application to Adaptive RT</strong></td>
<td></td>
<td>???</td>
<td>WEAKNESS</td>
</tr>
<tr>
<td>Can the method be used to quickly re-design plan based on changes (e.g. anatomy)</td>
<td>???</td>
<td>???</td>
<td></td>
</tr>
</tbody>
</table>
THE BIG(GER) PICTURE

We have reached our destination: A New Starting Line.
UNDENIABLE TRENDS

1. **Economics.** The cost of healthcare (at least in the U.S.A.) is unsustainable.

2. **Demographics.** Increase in # cancer patients. Changing workforce demographics.

3. **Technology.** If it can be automated, it will be automated.

4. **Patient Expectations.** The promise of “personalized” medicine.
Organization for Economic Cooperation and Development (OECD), an international economic group comprised of 34 member nations

USA vs. The Rest, noteworthy stats from the 2012 Report (analyzing the year 2010):
- 17.6% of GDP (OECD average: 9.5%, next highest is The Netherlands at 12%)
- 2.4 physicians per 1000 (OECD average: 3.1)
- 2.6 hospital beds per 1000 (OECD average: 3.4)
- 78.7 life expectancy (OECD average: 79.8)
At 17.6% of GDP in 2010, US health spending is one and a half as much as any other country, and nearly twice the OECD average.

Total health expenditure as a share of GDP, 2010 (or nearest year)

1. In the Netherlands, it is not possible to clearly distinguish the public and private share related to investments.
2. Total expenditure excluding investments.
Information on data for Israel: [http://dx.doi.org/10.1787/888932315602](http://dx.doi.org/10.1787/888932315602).

Source: OECD Health Data 2012.
US spends two-and-a-half times the OECD average

Total health expenditure per capita, public and private, 2010 (or nearest year)

1. In the Netherlands, it is not possible to clearly distinguish the public and private share related to investments.
2. Total expenditure excluding investments.
Information on data for Israel: [http://dx.doi.org/10.1787/888932315602](http://dx.doi.org/10.1787/888932315602).

Source: OECD Health Data 2012.
“Medical costs are the tapeworm of American economic competitiveness.”

– Warren Buffett (May 2017)
Cancer Research

Projecting Cancer Incidence and Deaths to 2030: The Unexpected Burden of Thyroid, Liver, and Pancreas Cancers in the United States

Lola Rahib, Benjamin D. Smith, Rhonda Aizenberg, Allison B. Rosenzweig, Julie M. Fleshman, and Lynn M. Matrisian
The Global Cancer Observatory (GCO) is an interactive web-based platform presenting global cancer statistics to inform cancer control and research.

- **Cancer Today**: Provides data visualization tools that present current national estimates of the incidence, mortality, and prevalence of 27 cancer types in 184 countries, by sex and age group. Go to website.

- **Cancer Over Time**: Undergoing further development; will provide data visualization tools that document the changing rates of cancer incidence and mortality over the course of half a century, based on high-quality data from 40 countries. Go to website.

- **Cancer Tomorrow**: Undergoing further development; will provide data visualization tools that predict the future cancer burden worldwide to 2035. Go to website.

- **Cancer Causes**: An expanding set of visualization tools that links the cancer burden to underlying causes, quantifying the extent to which different cancers are attributable to key lifestyle and environmental risk factors worldwide. Go to website.

http://gco.iarc.fr/
Participation in Plan Studies: Ratio of #Physicists / #Dosimetrists
DEMOGRAPHICS OF ANOTHER VARIETY

- Who does what?
  - Roles are changing.
  - Lines are blurring.
  - Boxes on flow charts are moving around.

- The ecosystem of tasks
  - Some conventional tasks are being obviated, others automated or outsourced.
  - New tasks are being introduced.
  - In general, tasks are migrating towards those with proven skillsets and interests.
#3 Automation Specifically Auto-Planning

- **DO**
  - Accept it in theory, because it’s happening.
  - Be critical and test it (or even help design it!), because we need it to work.
  - Embrace it when it does work, then move on to the new and remaining challenges.

- **DON’T**
  - Reject it in theory, because that would be short-sighted.
  - Accept it on faith, promises, or marketing, because it might not work very well and it could do more harm than good.
  - Rebel against it, because you will be left behind.
#4: **Patient Expectations**

- Patients expect personalized care.
- Your institution is probably marketing it already, anyway.

**Paving the Way for Personalized Medicine**

*FDA’s Role in a New Era of Medical Product Development*
IN OTHER WORDS...

- This is our challenge:
  - Use much less money
  - To treat far more patients
  - With personalized medicine
  - For better outcomes.

*Easy, right?*
The FINISH of this talk...

The START of something else.
Beyond Tomorrow & Outside the Box

Clinical Team

- Radiation Biologist, M.D.
- Personalized Treatment Manager

- Process patient history & genomics
- Design optimal, customized dose Rx
- Monitor progress & track outcomes

- Manage patient imaging
- Manage and verify the anatomy modeling
- Manage and verify the treatment plan
- Be with patient, per fraction, to manage setup, IGRT, and adaptive

- Machine install & acceptance testing
- Software maintenance & system tests
- Data management & communications
- Test and maintain the Plan Engine

Physics & Eng.

Outsourced or Automated

- Anatomy Modeling System
- Plan Engine
- Radiation System Engineer