Automation in Planning

A. Scaggion, PhD

Recommended readings

Recent review articles with an extensive collection of literature

- M. Hussein et al, "Automation in intensity-modulated radiotherapy treatment planning - a review of recent innovations." *The British Journal of Radiology, 20180270.* (2018) https://doi.org/10.1259/bjr.20180270
- S. Breedveld et al, "Multi-criteria optimization and decision-making in radiotherapy", *European Journal of Operational Research,* **277**(1): 1-19 (2019) https://doi.org/10.1016/j.ejor.2018.08.019
- P. Meyer et al, "Survey on deep learning for radiotherapy", *Computers in Biology and Medicine,* **98**:126-146 (2018) https://doi.org/10.1016/j.compbiomed.2018.05.018
- A.M. Kalet et al, "Quality assurance tasks and tools: The many roles of machine learning." to be published in *Medical Physics*, https://doi.org/10.1002/mp.13445

Why are we fostering automation?

RT planning is a complex non convex problem with a non-unique solution which is always been tackled with a trial-and-error approach

Why are we fostering automation?

RT planning is a complex non convex problem with a non-unique solution which is always been tackled with a trial-and-error approach

- **EFFICIENCY**

  Save time and money

- **EFFICACY**

  Achieve better quality (within center and across centers)

- **RUBUSTNESS**

  Reduce variability (standardization)

- **EDUCATION**

  Share knowledge

Outline

- **Why we talk about quality and variability?**
- **Variability in Radiotherapy**
  - *does it really matter?* reported negative outcomes
  - *where does it come from?* main causes and actors
  - *how to tackle the problem?* some general examples
- **Variability in Treatment Planning**
  - *where does it come from?* a closer look
  - *automation can help with variability?* some reported examples
- **A clinical experience from Padova**
  - *Reduce inter- and intra-planner variability with a commercial knowledge-based planning solution*
- **Automation pitfalls and personal advices**
- **Conclusion**

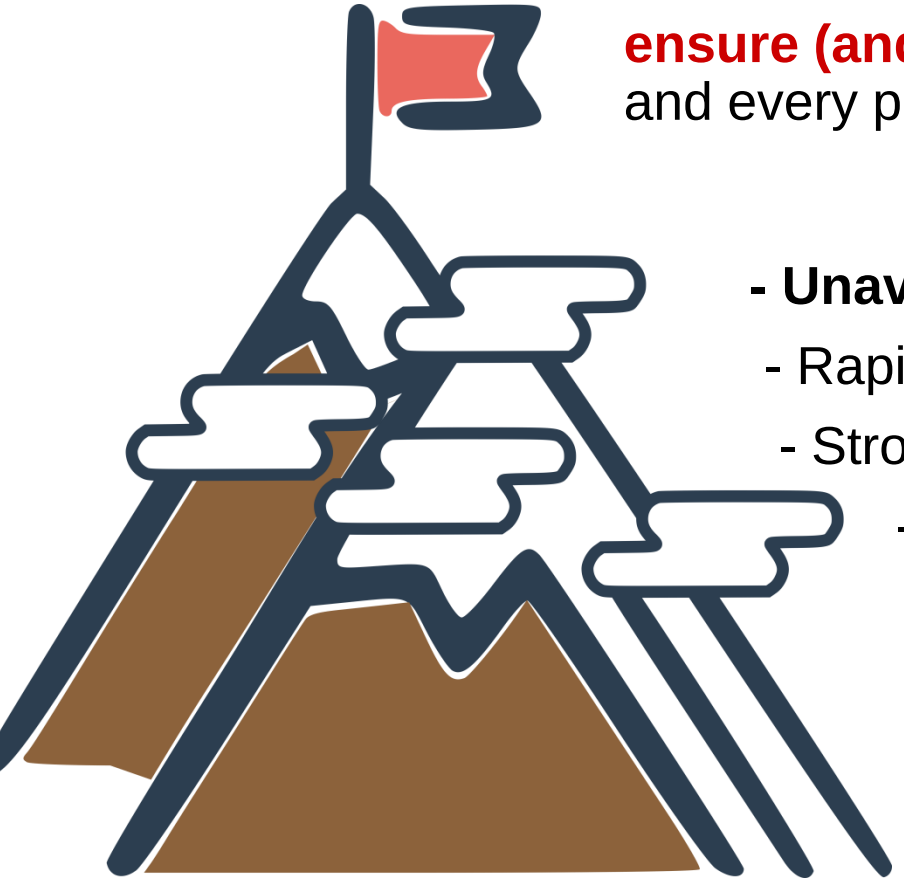
Why we talk about quality and variability?

«In radiotherapy, and medicine in general, there is not a "gold" standard for the best treatment.»

ensure (and assess) quality of cancer care to each and every patient without distinction.

Why we talk about quality and variability?

«In radiotherapy, and medicine in general, there is not a "gold" standard for the best treatment.»



ensure (and assess) quality of cancer care to each and every patient without distinction.

- **Unavoidable patient-specific variability**
- Rapid technology developments
- Strongly personalized treatments
- Spread of treatment protocols
- Increased life expectation and increased focus on QOL
- Demand for cost reduction (increase efficacy and efficiency)

Why we talk about quality and variability?

«In radiotherapy, and medicine in general, there is not a "gold" standard for the best treatment.»

ensure (and assess) quality of cancer care to each and every patient without distinction.

- **Unavoidable patient-speci**
- Rapid technology de
- Strongly perse
- Sp cols
- pectation and
- us on QOL
- nd for cost reduction (increase cacy and efficiency)

INCREASE VARIABILITY (MIGHT UNDERMINE QUALITY)

Negative outcomes of Variability in RT

(TROG 02.02) *«The impact of poor radiotherapy can greatly exceed the anticipated benefit of concurrent chemotherapy. […]»*

L.J. Peters, 2010 https://doi.org/10.1200/jco.2009.27.4498

(RTOG 9704) *«Failure to adhere to specified RT guidelines was associated with reduced survival [...]»*

R.A. Abrams, 2012 https://doi.org/10.1016/j.ijrobp.2010.11.039

«protocol-compliant RT may decrease failure rates and increase overall survival[…] »

A. Fairchild, 2013 https://doi.org/10.1016/j.ijrobp.2013.03.036

«In clinical trials, RT protocol deviations are associated with increased risks of treatment failure and overall mortality. »

N. Ohri, 2013 https://doi.org/10.1093/jnci/djt001

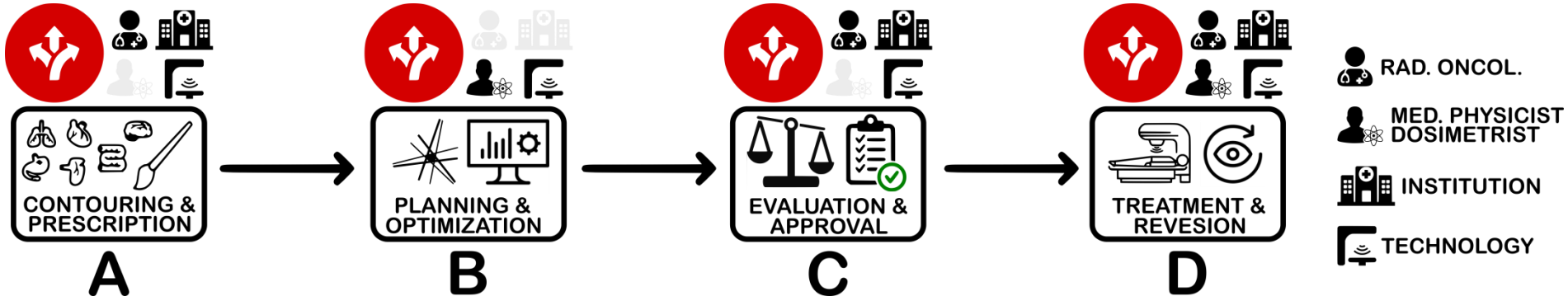
*«**Plan quality deficiencies** in* **RTOG 0126** *exposed patients to substantial excess risk for rectal complications.»*

K.L. Moore, 2015 http://dx.doi.org/10.1016/j.ijrobp.2015.01.046

«Individualized QA indicated that OAR sparing could frequently be improved in **EORTC-1219-DAHANCA-29** *study plans,* ***even though they met the trial's generic plan criteria.****»*

J.P. Tol, 2018 https://doi.org/10.1016/j.radonc.2018.10.005

Causes and actors of RT variability



A) target volume delineation, organ-at-risk delineation, prescription and fractionation, technology availability, combined, …

B) plan set-up, plan optimization, …

C) plan evaluation, protocol adherence, …

D) Imaging capabilities, IGRT/ART protocols, treatment evaluation and revision, …

Causes and actors of RT variability

A) target volume delineatio… …lineation, prescription and fra… …ogy availability, combined, …

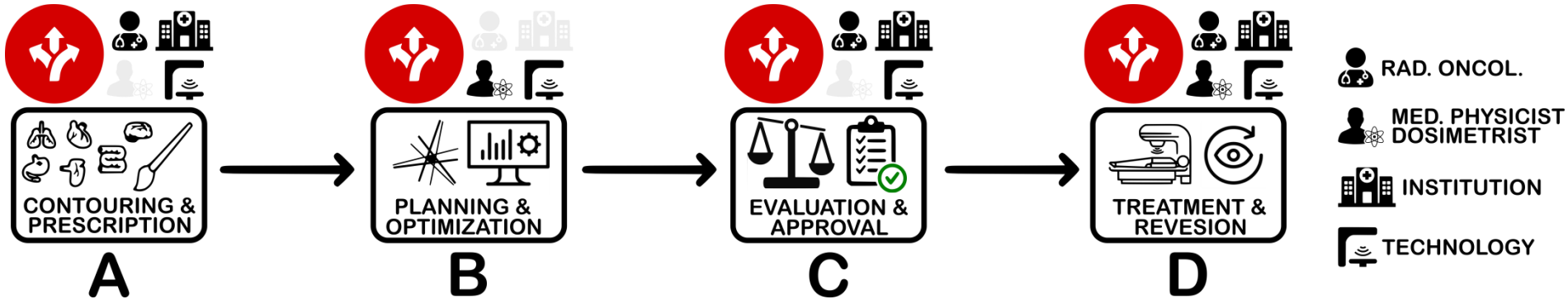
B) plan sc… …tion, …

C) …rotocol adherence, …

D) …apabilities, IGRT/ART protocols, treatment …ation and revision, …

THE MEDICAL PHYSICIST IS ONLY A PART OF A COMPLEX SYSTEM

Some examples of variability tackling



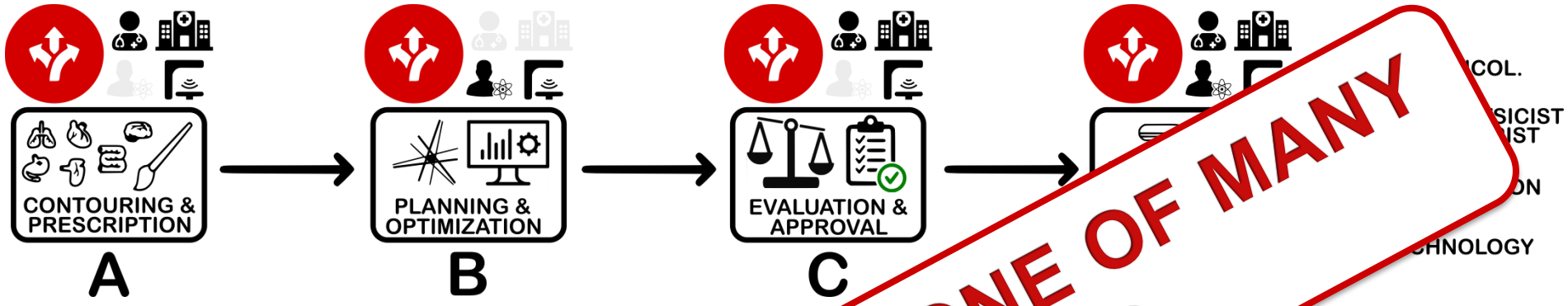
A) draw local protocols, adherence to national/international standards, participation in Quality Assurance programmes, **use of automatic or semi-automatic contouring solution**, technological update, …

B) continuous planners education, **use of automatic or semi-automatic planning solution**, …

C) adherence to national/international standards, continuous professional interaction and collaboration, draw local protocols, participation in Quality Assurance programmes, **use of automatic or semi-automatic decision-making approaches**, …

D) draw local protocols, adherence to national/international standards, **use of automatic or semi-automatic decision-making approaches,** …

Some examples of variability tackling



AUTOMATION IS ONLY ONE OF MANY POSSIBLE ACTIONS

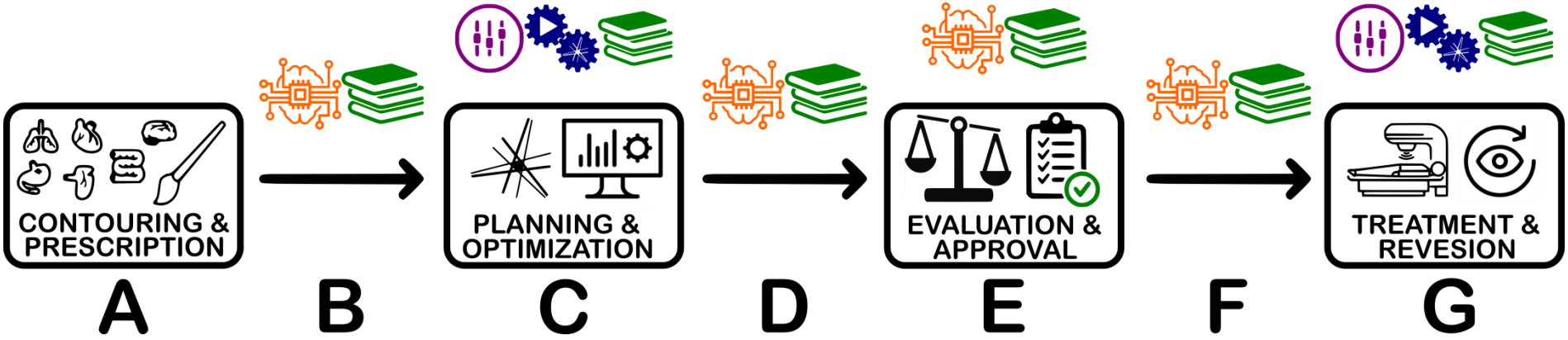
A) draw local protocols, adherence to n… …ards, participation in Quality Assuran… …**tomatic or semi-automatic contouring sol**… …te, …

B) continuous planner… …**matic or semi-automatic planning sol**…

C) adhere… …ial standards, continuous professional i… …n, draw local protocols, participation in Quality …es, **use of automatic or semi-automatic decision-…nes**, …

D) …l protocols, adherence to national/international standards, **use of …omatic or semi-automatic decision-making approaches,** …

Can automation tailor variability in RT?



A) OAR contouring (tumor contouring) and treatment definition

B) spot possible unavoidable tradeoffs and tailor prescription to dose constraints

C) drive optimization

D) comparison to historical standards (within center QA)

E) comparison to general standards (across center QA, automatic peer-revision)

F) evaluate need for/benefit of replanning

G) online/offline replanning

ARTIFICIAL INTELLIGENCE**

** Machine Learning, Deep Neural Network, ...

MCO

AUTOMATIC PLANNING

KBP

Main causes of treatment planning variability

- ACROSS CENTERS
- WITHIN CENTER

- Center experience and subspecializiation (yearly patients income)
- Available technology
- Planner's expertise and planning skills

Mainly related to

- Differences in treatment set-up (technique and geometry)
- Difficulty to a priori asses the attainable tradeoff between the PTV coverage and OAR sparing
- Differences in planning priorities during optimization (different choices for OAR sparing)
- Clinical workload (time for planning and pressure on planner)

Can automation solve variability in planning?

Whatever the strategy, planning automated solutions aim at:

- **Reply a predefined scheme of actions or a result similar to prior ones**

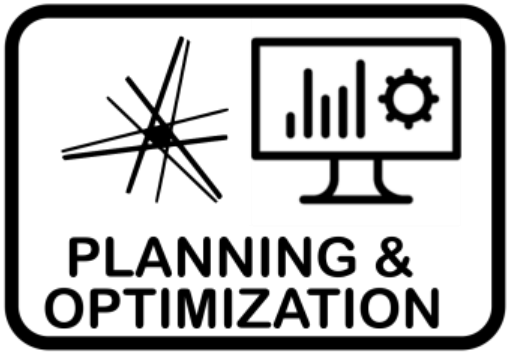
(might not be the best path but at least reduces the possible paths followed in search for a solution)

- **Tend to the better solution they know**

(might not be the best one but is at least the better)

- **Reduce the time spent tackling single challenges**

(system set-up might be time demanding, but reduces time when you might be in a hurry)



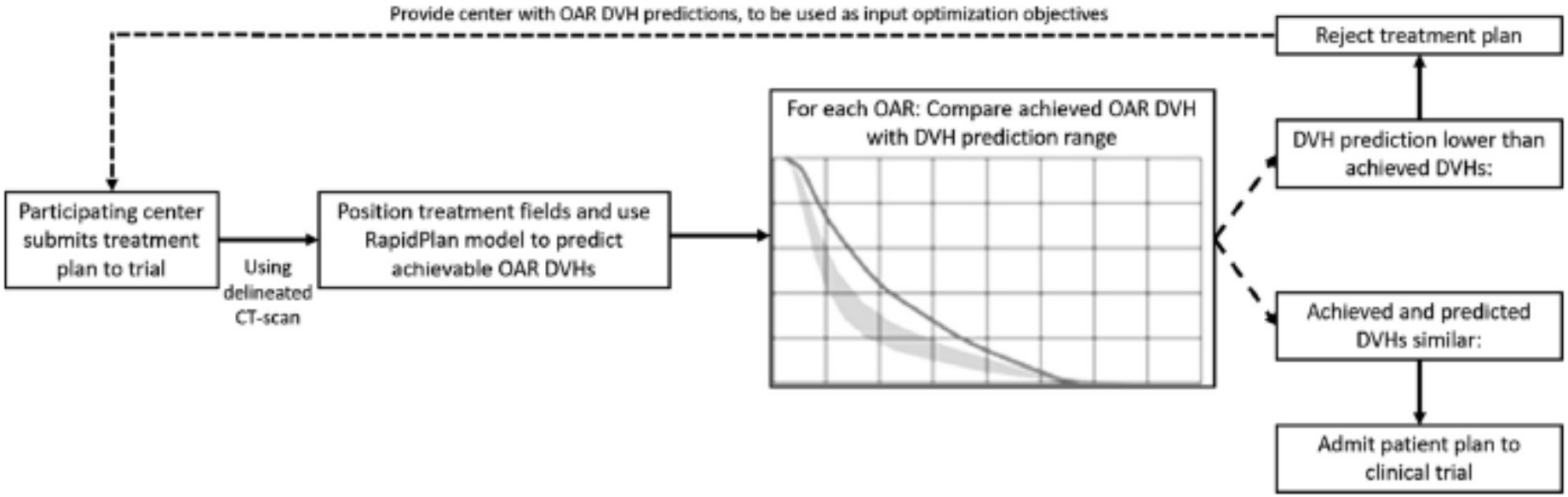
Single strategies might also:

- give you a bunch of possible good solutions (*a posteriori* MCO)
- be educational for the user (*a posteriori* MCO, KBP)
- be educational for others (KBP)
- be fairly automatic (PBAIO, *a priori* MCO),

EX 1. Automated patient-specific QA tool from KBP

JP Tol et al. *Radiotherapy and Oncology* **130** 75–81 (2019)

- KBP model generated from a single well-trained institution
- KBP model used to predict and plan treatment for patients included in a previous clinical trial (EORTC-1219-DAHANCA-29)
- Proposal of an automated patient-specific QA workflow

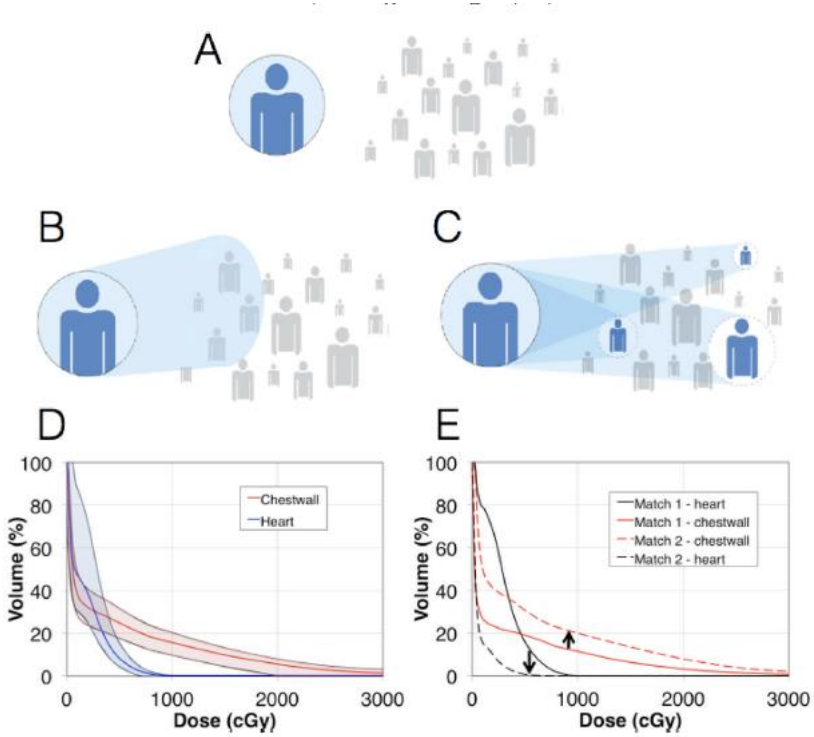


- **Such steps also have the potential to be largely automated.**

EX 2. Machine Learning for decision support

G.Valdes et al *Radiotherapy and Oncology* **125**(3), 392-397 (2017)

- The first artificial-intelligence based clinical decision support system (CDS) in radiation oncology
- CDS connects current assessments (patients) to past decisions (discrete treatment plans already treated)
- Past cases are collected and classified through: anatomical information, medical records, treatment intent, radiation transport
- A machine-learning algorithm is fed with these data
- For any new patient the “closer” and “more diverse” solution are proposed
- **Clinicians can be informed of dose tradeoffs between critical structures early in the treatment process**



A) Standard treatment planning

B) Knowledge-based planning

C) Treatment plan outcome decision support enabled using treatment plan classification.

D) DVH prediction provided by KBP

E) DVH illustrating distinct tradeoffs provided by a classification technique.

EX 3. QA models from Pareto-Optimal plans

Y. Wang et al *Phys. Med. Biol.* **61** 4268–4282 (2016) + **Y. Wang et al** *Med. Phys.* **46** 934-943 (2018)

- Database of fully automated generated Pareto-optimal treatment plans with consistent priorities (a single wishlist).
- Plans used to train existing OVH model.
- **OVH model applied to predict DVH as a QA tool.**

- Database of Pareto-optimal treatment plans generated with fully automatic multi-criterial treatment planning variyng the priority list
  (N patients x M priority lists → NxM treatment plans)
- This dataset contains intrinsically effect of inter-organ dependency and dataset inconsistency.
- **This database can be used to validate and characterize KBP prediction models** (available upon request).

PD experience – clinical implementation KBP (I)

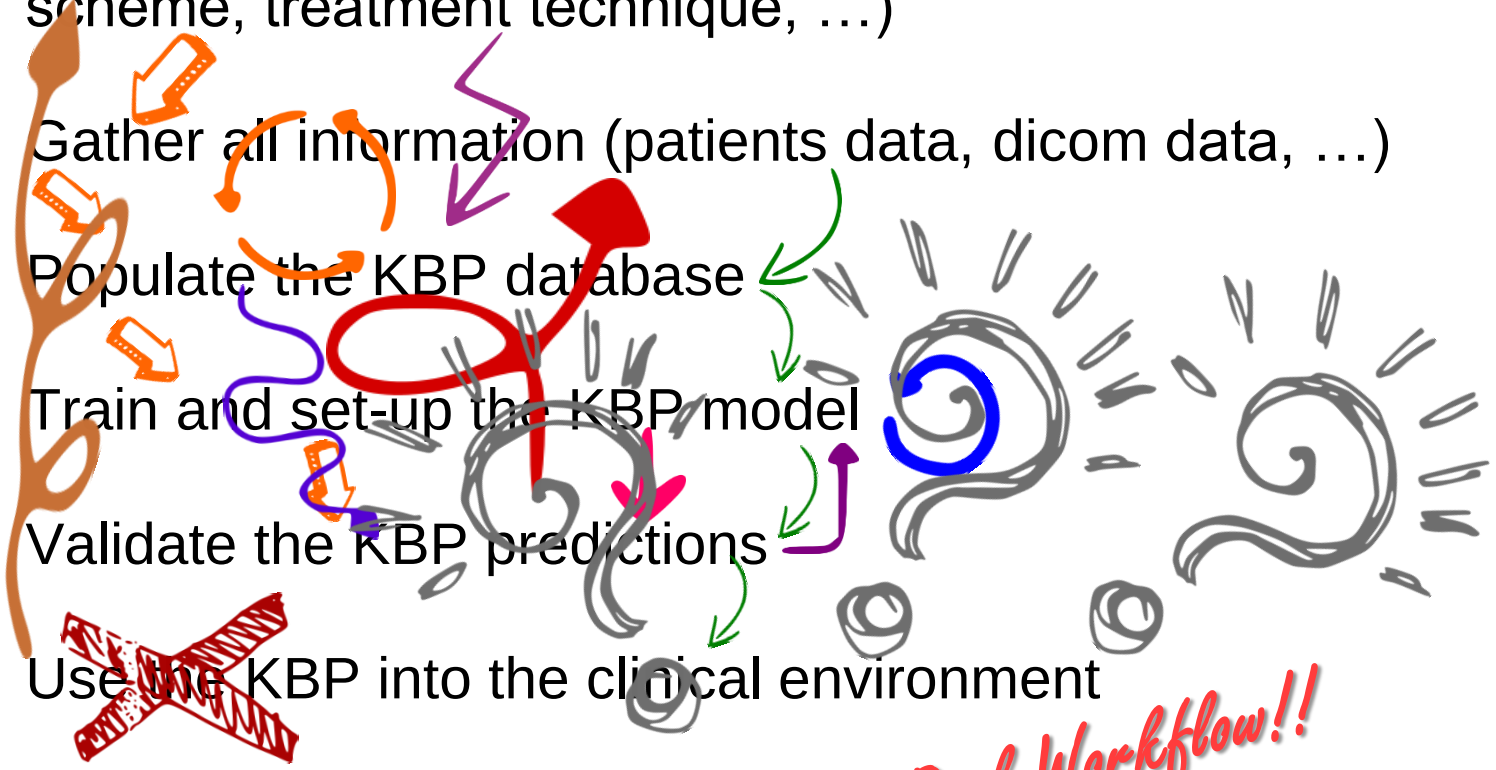
Ideal workflow:

1. Select and specify the model goal (site/disease, fractionation scheme, treatment technique, …)
2. Gather all information (patients data, dicom data, …)
3. Populate the KBP database
4. Train and set-up the KBP model
5. Validate the KBP predictions
6. Use the KBP into the clinical environment

PD experience – clinical implementation KBP (II)

Ideal workflow:

1. Select and specify the model goal (site/disease, fractionation scheme, treatment technique, …)
2. Gather all information (patients data, dicom data, …)
3. Populate the KBP database
4. Train and set-up the KBP model
5. Validate the KBP predictions
6. Use the KBP into the clinical environment



Real Workflow!!

PD experience – Goal, sample, population

Our experience with prostate:

SYSTEM: Varian RapidPlan v13.5, initially then upgraded to v15.5

GOAL: Radical prostate only – 78Gy/39fx or 70Gy/28fx, only VMAT

PATIENT SAMPLE: in total ~120 patients

- ~100 used for model training (initially 60 then further collection) treated between 2014 and 2016
- after database cleaning only 82 used (because of deviations)
- 20 patients reserved for closed-loop validation

MODEL POPULATION: CTV, PTV, rectum, bladder, femoral head left and right separately (after a major revision penile bulb was added and femurs were merged)

PD experience – Training and refinement (I)

MODEL TRAINING AND REFINEMENT:

- **1st APPROACH:** After every training we:
  1. Removed geometric outliers
  2. Visually checked every other outlier and every single OAR largely over prediction
  3. Re-planned under-optimized plans
  4. Re-train the model and begin again ,

Where/when should we stop?

How to cope with trade-offs? (KBP model is composed by as many models as OARs which do not take directly into account inter-organ dependencies) See **Y. Wang et al *Phys. Med. Biol.* 61 4268–4282 (2016)** and **Y. Wang et al *Med. Phys.* 46 934-943 (2018)** for a deeper perspective

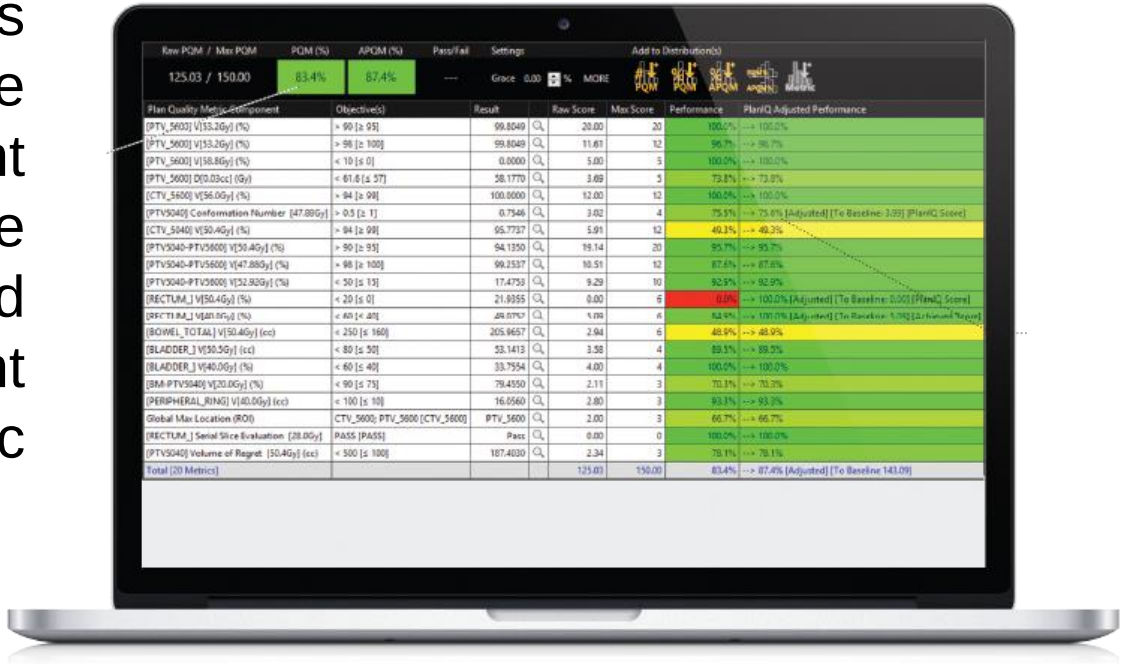
How to thoroughly compare to competitive plans?

PD experience – Quality Score

- **PQM%** (Plan Quality Metric) is a user-defined metric intended to quantify and compare plan quality
- PQM can be adjusted making use of a “feasibility” analysis built upon first principles to become **APQM%**

BE Nelms, *Pract. Radiat. Oncol.* **2**:296-305 (2012) S Ahmed, *Med. Phys*. **44**:5486-5497 (2017)

- It allows to rank plans (pertaining to same patient or different patients) following the user clinical practice and taking into account patient specific challenges



PD experience – Training and refinement (II)

MODEL TRAINING AND REFINEMENT:

- **2nd APPROACH: Firstly we ranked all plans through a quantitative quality score (APQM%).** After every training we:
  1. Removed geometric outliers (only if law-quality plans)
  2. Visually checked every other outlier and every single OAR largely over prediction
  3. Re-planned under-optimized plans (of lower-half of the rank) and re-ranked them
  4. Re-train the model and begin again

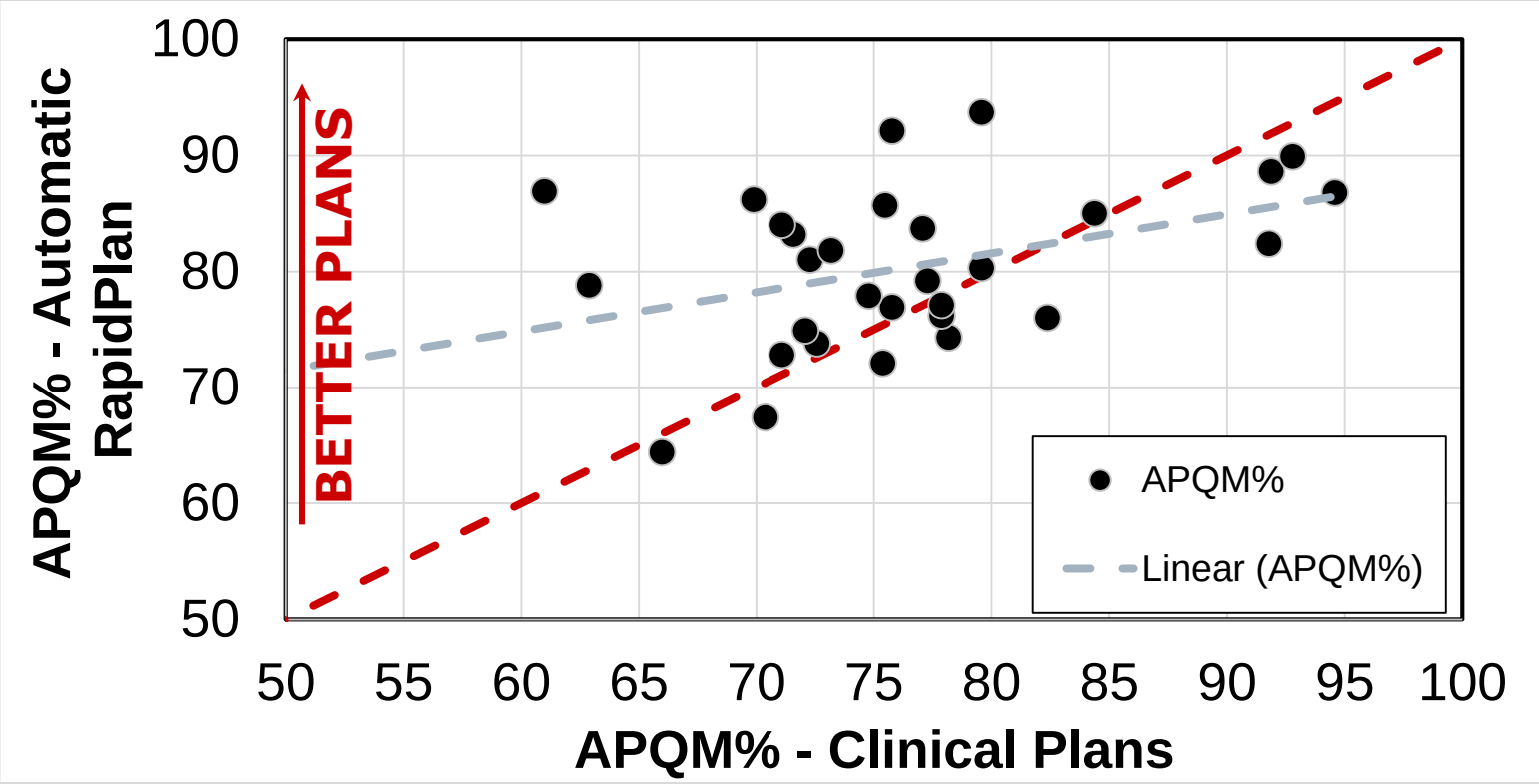
We stopped when there where no more plans lower than the initial first quartile

How to cope with trade-offs? PQM%

How to thoroughly compare to competitive plans? PQM%

PD experience – Validation

- To validate the model we automatically re-planned (without human interaction):
  - 20 randomly chosen patients within the model (closed-loop)
  - 20 patients outside the model (open-loop)

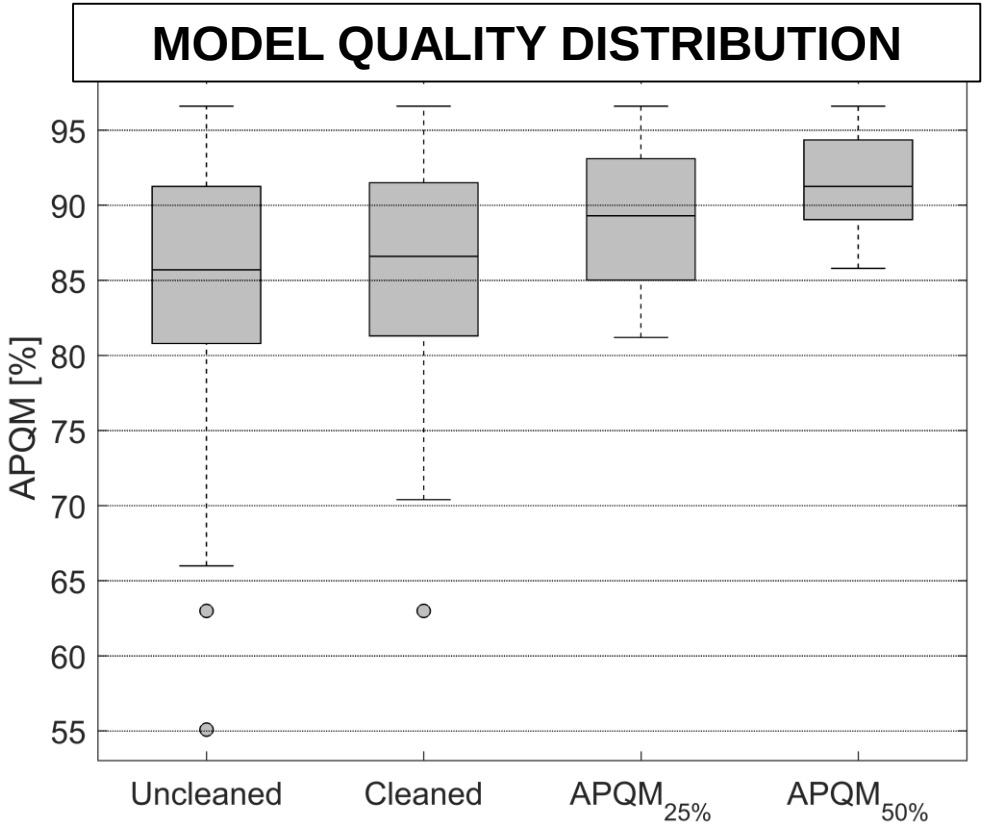


PD experience – Ancillary project (I)

M. Fusella et al *Med.Phys.* **45**(6):86-93 (2018)

- APQM% scoring have been proposed as a tool to help a feed forward population, train and validation of a KBP model

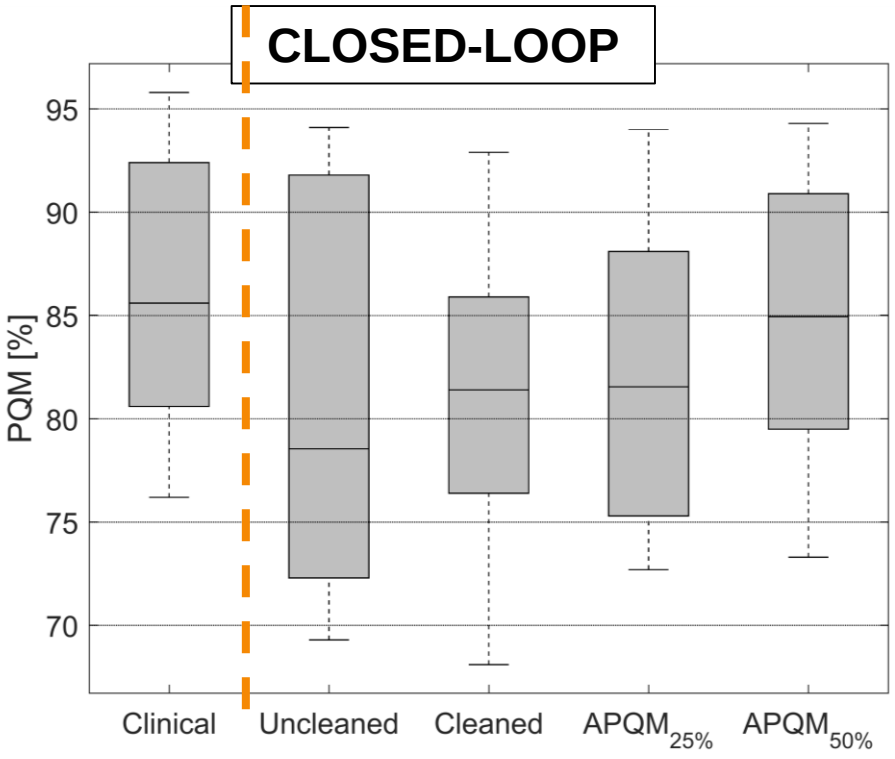
- 4 KBP models compared:
  - ***Uncleaned:*** 80 patients no other refinement
  - ***Cleaned:*** outliers removed to lead 69 patients
  - ***APQM25%:*** 60 patients removed the lower quartile of APQM% ranking
  - ***APQM50%:*** 40 patients removed the lower half of APQM% ranking

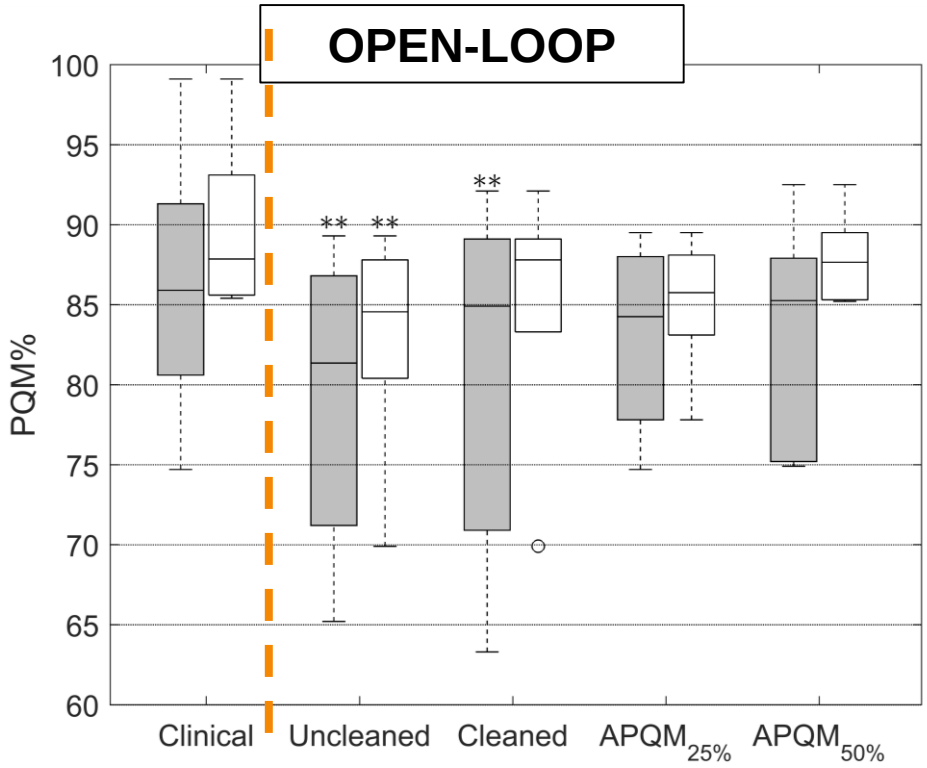


PD experience – Ancillary project (II)

M. Fusella et al *Med.Phys.* **45**(6):86-93 (2018)

- Open- and closed-loop validation of automatically optimized plans compared through PQM%
- **RESULT:** “Better plans in, better plans out”, but pay attention to the width of the population





PD experience - clinical impact of KBP (I)

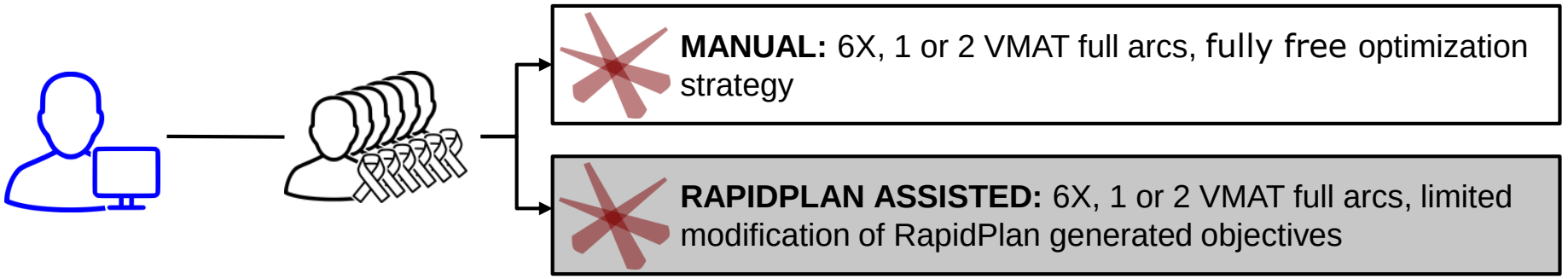
A. Scaggion et al. *Phys. Med.* **53**:86-93 (2018)

- ***Varian RapidPlan v13.5* prostate model**: 70 patients - prostate only - VMAT, validated through *open-loop* and *closed-loop* tests

15 patients used for prospective planning

7 planners: 6 resident operators + 1 internship student

- Each operator planned twice the same 15 patients with and without RapidPlan assistance

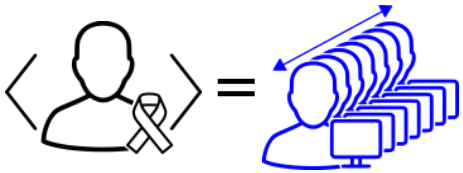


PD experience - clinical impact of KBP (II)

A. Scaggion et al. *Phys. Med.* **53**:86-93 (2018)

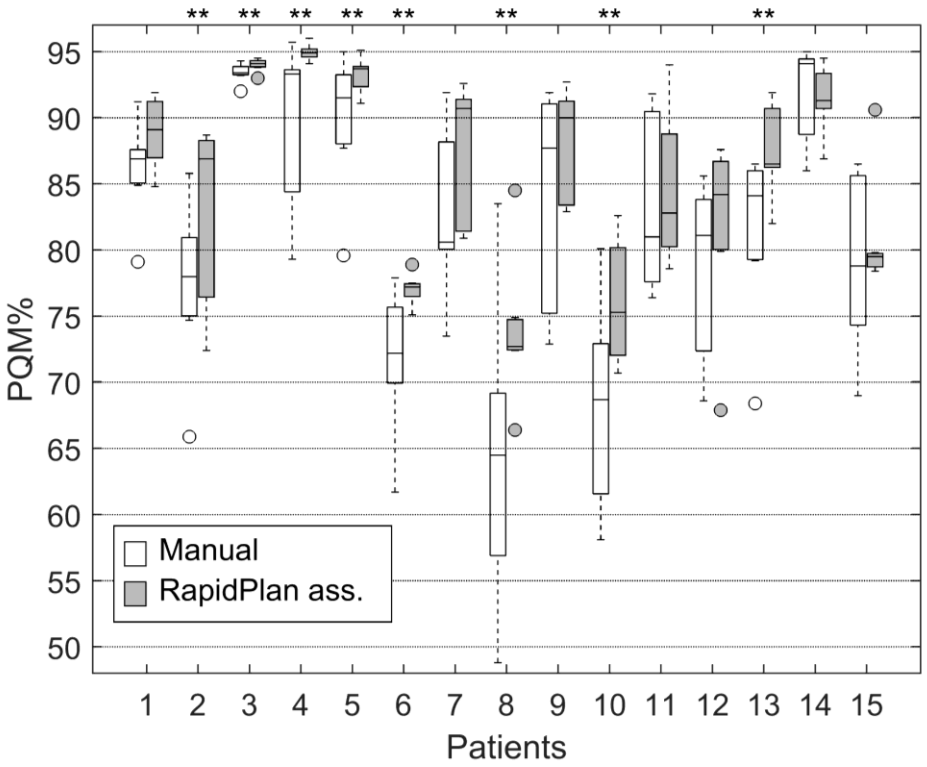
- The overall increase in plan quality is accompanied by a general reduction in its variability

- 11 out of 15 patients (73%) showed an increased mean quality (PQM%) **and** a reduced variability



INTER-PLANNER
(averaging over patients)

Metric		IQR		IQR reduction	
		Manual	RapidPlan	# of cases	p-value
Rectum	V_{40Gy} [%]	11.69±3.71	6.45±3.34	14 (93%)	<0.001*
	V_{60Gy} [%]	4.26±2.77	2.17±1.02	13 (86%)	0.002*
	V_{75Gy} [cc]	0.81±0.74	0.42±0.34	13 (86%)	0.001*
	<D> [Gy]	6.57±1.90	3.82±1.89	15 (100%)	<0.001*
Bladder	V_{40Gy} [%]	4.62±3.75	3.54±2.93	11 (73%)	0.107
	V_{65Gy} [%]	1.42±1.19	1.14±0.89	10 (66%)	0.389
	V_{75Gy} [cc]	1.26±0.69	1.12±0.67	9 (60%)	0.330
	<D> [Gy]	2.35±1.65	1.95±1.45	10 (66%)	0.277
Femur R	D_{1cc} [Gy]	4.77±2.65	2.49±1.04	12 (80%)	0.008*
	<D> [Gy]	2.09±0.87	1.45±0.78	10 (66%)	0.043*
Femur L	D_{1cc} [Gy]	5.66±2.33	2.76±1.57	13 (86%)	0.001*
	<D> [Gy]	3.02±1.05	1.56±0.97	14 (93%)	0.001*
PQM%		8.32±4.19	4.73±3.79	12 (80%)	0.046*

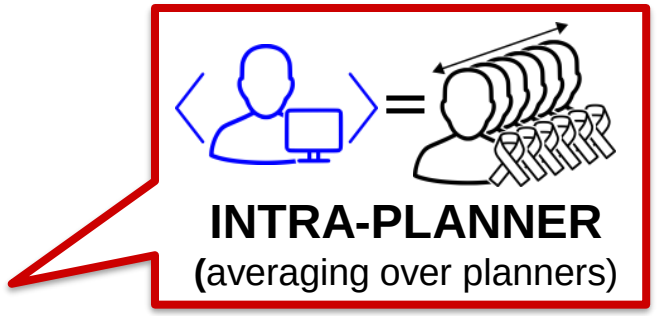


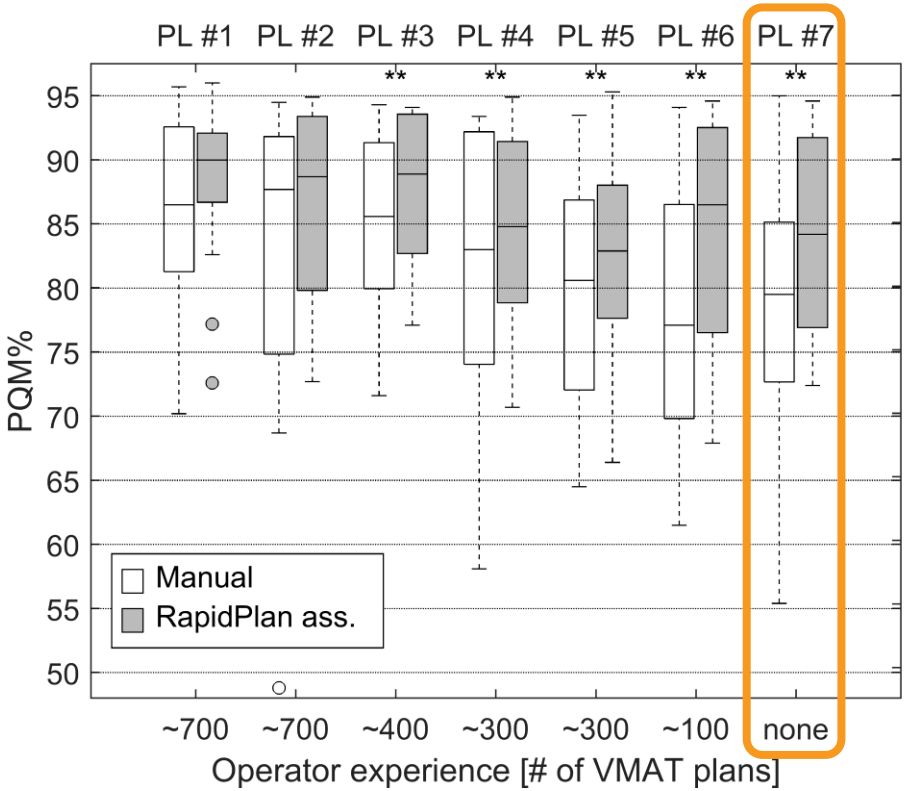
PD experience - clinical impact of KBP (III)

A. Scaggion et al. *Phys. Med.* **53**:86-93 (2018)

- Significant increase of overall quality for 5 out of 7 planners
- **Internship student raised to the level of a medium experienced planner**
- Intra-planner variability shows a reduction for all planners but #6
- The overall reduction is statistically significant (p-value=0.033).
- Similar results have been found by Wang in 2017 for left-side breast cancer

J Wang, 2017 https://doi.org/10.1186/s13014-017-0822-z

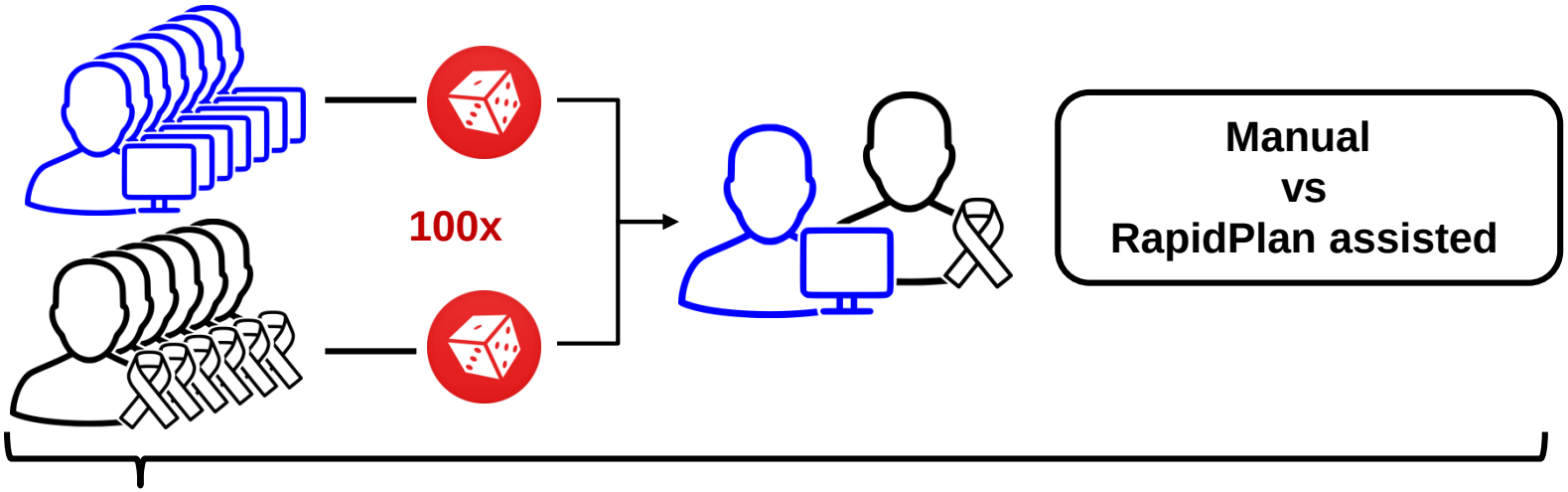




PD experience - clinical impact of KBP (IV)

A. Scaggion et al. *Phys. Med.* **53**:86-93 (2018)

- Clinical impact simulation based on bootstrap technique



- Average **+5.35%** increase of PQM% (C.I. = [4.78%,5.91%])
- **75.1% chance** better overall plan for every new patient (C.I. = [72.3%,77.8%]).

Take home messages

- **The automated system you will have tomorrow will be based on the knowledge you have today!** Start today to collect data as methodically and tidy as you can and to reduce variability as much as you can!
- Every automated process is the outcome of a lengthy and demanding effort to build it. **The effort you save tomorrow is worth the effort you spend today.**
- At the moment **most of the automated planning solution still require a certain degree of human interaction.** Some room for variability still remains.
- An automated planning solution leaves you more room to improve not more time to pursue your own business (disengaging is dangerous)
- Solutions based on prior knowledge require to gather large amount of cases. A lot of tasks that single centers can not undertake alone (pediatrics, rare disease, ultra-specific treatments, …) and, in my opinion, should not undertake alone.

Conclusion

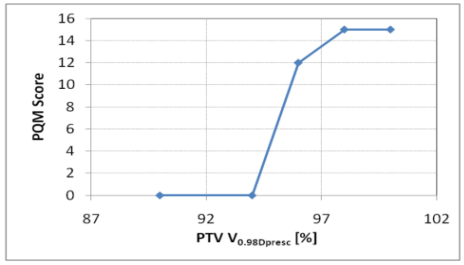
- Automation in planning is accomplished through different strategies but all of them has the **intrinsic capability to improve quality and limit variability.**
- **SMALL SCALE:** Automation narrows the space for human failure (planner interaction).
- **LARGE SCALE:** Automation opens the possibility to global improvement through large collaboration (multi-center shared libraries, patient-specific QA, …)
- In Padova we need approximately 18 months to set-up the first RapidPlan model. Most of them spent to deeply understand the tool.
  - hands-on experience is beneficial and crucial
  - set-up is a trial-error process based on human interaction (pay attention to your own decision)
  - KBP useful to increase quality and train new human resources

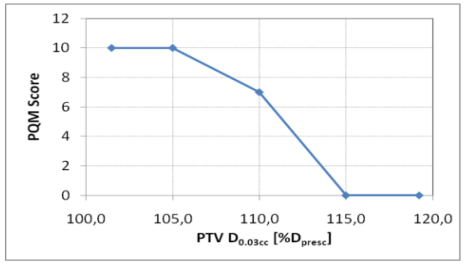

Thank You
Kiitos
Asante
Grazie
Merci
Dankon
Welalin
Terma Kasih
Salamat
Multumesc
Matondo
Nirringrazzjak
Mochchakkeram
Spasibo
Dank Je
Obrigado
Arigato
Kia Ora
Ua Tsaug Rau Koj
Raibh Maith Agat
Cam on ban
Matur Nuwun

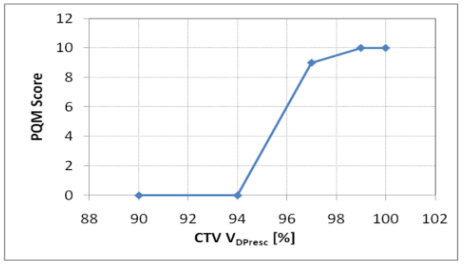
PD experience - Plan Quality Metric (I)

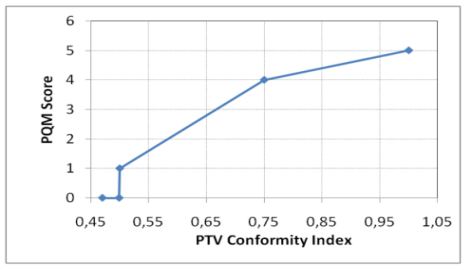
Structure	Metric	Definition	PQM value range Min	Max
PTV	$V_{0.98Dpresc}$ [%]	Percent of PTV volume ≥ 98% of the prescription dose	0	15
PTV	$D_{0.03\ cc}$ [Gy]	Dose [Gy] covering highest 0.03 cc of PTV	0	10
CTV	V_{Dpresc} [%]	Percent of CTV volume ≥ prescription dose	0	10
PTV	Conformity index	(PTV $V_{95\%}$ [cc])^2 / (PTV total volume [cc] * 0.98D_{presc} isosurface volume [cc])	0	5
Body - PTV	V_{Dpresc} [%]	Volume [cc] of tissue outside PTV ≥ D_{presc}	0	10
Rectum	V_{40Gy} [%]	Percent of rectum volume ≥ 40 Gy	0	10
Rectum	V_{65Gy} [%]	Percent of rectum volume ≥ 65 Gy	0	10
Rectum	V_{Dpresc} [cc]	Volume [cc] of rectum ≥ Dpresc	0	10
Rectum	Serial rectum	Number of axial planes with all rectum voxels exceeding 34 Gy	-10	0
Bladder	V_{40Gy} [%]	Percent of bladder volume ≥ 40 Gy	0	3
Bladder	V_{65Gy} [%]	Percent of bladder volume ≥ 65 Gy	0	7
Femur R	$D_{1\ cc}$ [Gy]	Dose [Gy] covering highest 1 cc of right femour	0	5
Femur L	$D_{1\ cc}$ [Gy]	Dose [Gy] covering highest 1 cc of left femour	0	5
	Global maximum location	Anatomic location of global maximum: CTV, PTV or elsewhere	0	5
Total			-10	105

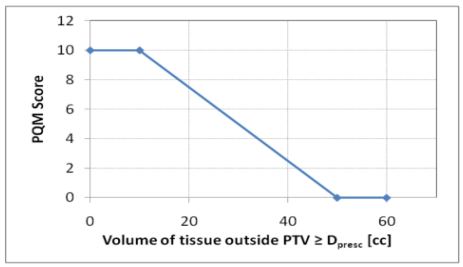
PD experience - Plan Quality Metric (II)

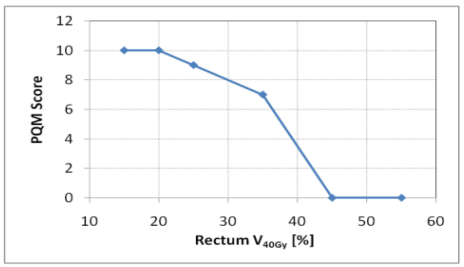


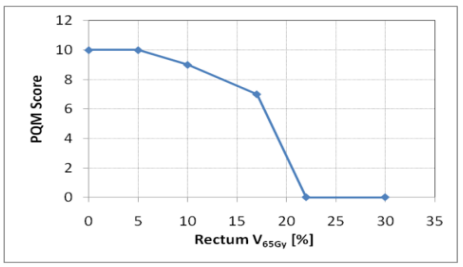


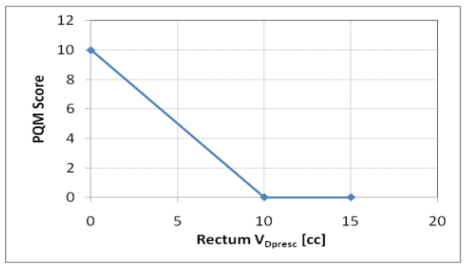


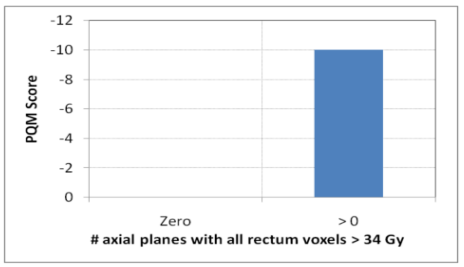


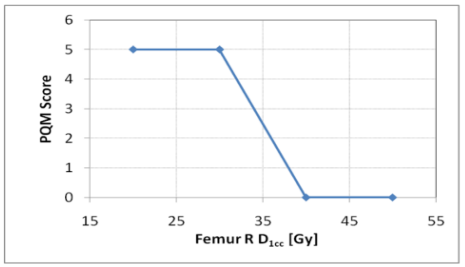


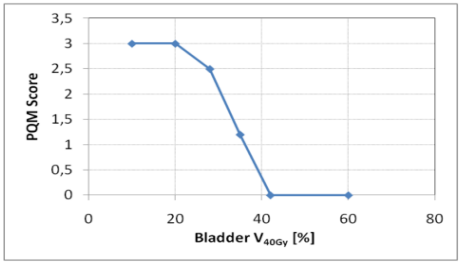


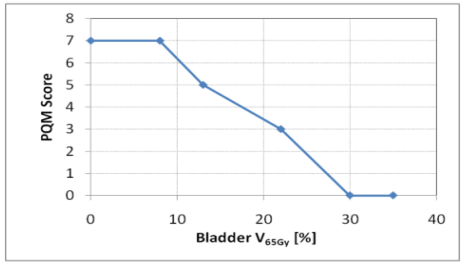


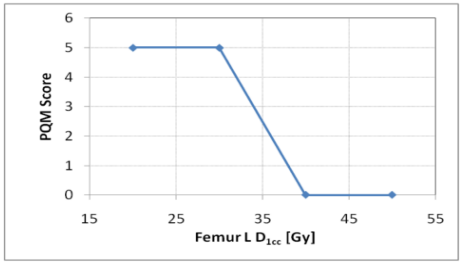


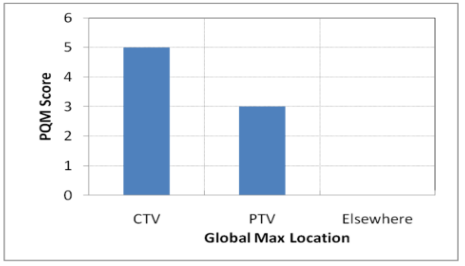












PD experience - plan quality DVH-based

- 46% plans unequivocal better sparing, 11% plans unequivocal worse sparing

Metric			Manual		RapidPlan assisted			
		Endpoint	mean ± std	[min;max]	mean ± std	[min;max]		p-value
PTV	$D_{98\%}$		98.5 ± 0.3	[97.5;98.9]	98.6 ± 0.3	[96.9;99.3]	😐	0.441
	$D_{2\%}$		105.6 ± 0.9	[103.7;109.4]	105.8 ± 1.0	[103.6;109.6]	😐	0.618
	HI		7.12 ± 1.16	[5.03;11.80]	7.21 ± 1.25	[4.46;12.68]	😐	0.837
	CI		1.00 ± 0.02	[0.97; 1.10]	0.99 ± 0.02	[0.97;1.06]	😐	0.089
Rectum	V_{40Gy} [%]	<= 45	30.43 ± 10.13	[11.39;58.07]	26.38 ± 8.60	[9.42;50.12]	😀	<0.001*
	V_{60Gy} [%]	<= 25	13.94 ± 6.33	[3.13;34.43]	11.31 ± 4.98	[3.42;25.69]	😀	<0.001*
	V_{75Gy} [cc]	<=10	2.38 ± 1.84	[0.03;10.09]	2.24 ± 1.62	[0.04;7.80]	😐	0.059
	<D> [Gy]		31.03±6.35	[18.90;46.07]	29.16±5.38	[19.18;41.29]	😀	<0.001*
Bladder	V_{40Gy} [%]	<=40	24.17 ± 10.38	[7.38;52.22]	23.41 ± 9.69	[7.49;44.23]	😐	0.091
	V_{65Gy} [%]	<=25	10.75 ± 5.19	[3.75;23.50]	10.37 ± 4.92	[3.65;22.19]	😐	0.118
	V_{75Gy} [cc]	<=10	13.19 ± 6.01	[4.55;32.99]	12.83 ± 5.80	[4.78;33.99]	😐	0.136
	<D> [Gy]		24.43±8.50	[8.35;41.59]	24.04±8.32	[8.99;39.24]	😐	0.149
Femur R	D_{1cc} [Gy]	<=45	29.67±8.50	[20.41;43.59]	26.68±3.35	[8.99;39.24]	😀	<0.001*
	<D> [Gy]		14.00±4.68	[7.28;21.12]	13.01±2.57	[8.17;18.62]	😀	<0.001*
Femur L	D_{1cc} [Gy]	<=45	30.29±5.23	[18.79;45.18]	27.95±3.68	[18.29;37.03]	😀	<0.001*
	<D> [Gy]		14.58±3.14	[6.70;23.24]	13.68±2.68	[7.06;21.36]	😀	<0.001*