

# CHALLENGES IN IMPLEMENTING AN AI-BASED AUTOCODING TOOL TO ENHANCE CODING EFFICIENCY

Dr. Karen Pien, MD – University Hospital Brussels

Dr. Maarten Lambrecht, MD – Solventum

Dr. Elric Verbruggen, MD - Solventum

## MEDICAL REGISTRATION TEAM - UZB

### Why?

- Reimbursement system
- Epidemiology
- Secondary data use

### What do we want?

- Excellent coding quality
- Coded as fast as possible

### Coding team:

- 721 beds Classical hosp – 100 daycare positions
- 7 coders – 6.6 FTE
- Highly skilled and trained coders

	2016	2019	2020	2021	2022	2023
Daycare	33.200	40.500	30.200	37.600	38.000	41.500
Classical hosp	32.100	32.700	27.500	29.500	30.000	32.000
Total	65.300	73.200	57.700	67.100	68.000	73.500

### Future of Medical Registration team:

Small(er) team of coding experts

Cost reduction

Shortage of nursing staff

How?

# 360 Encompass: Autocoding daycare admission

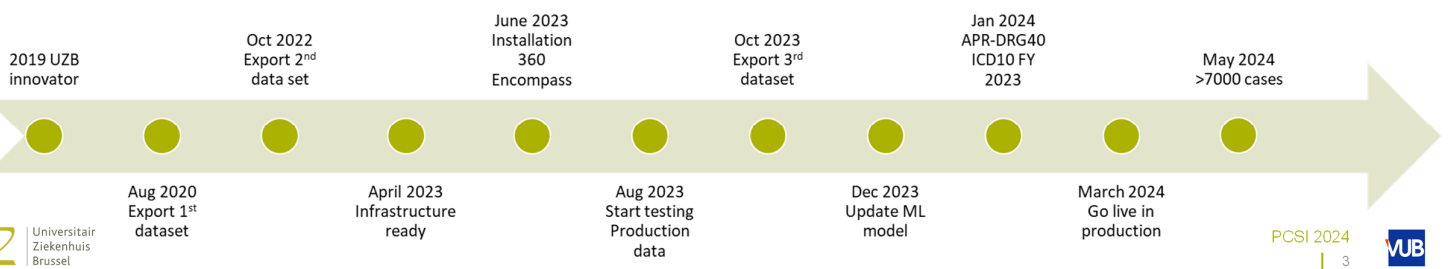
- Based on Machine Learning ([ECAN2 neural network](#))
- ICD-10-CM and PCS codes from free text (discharge letters, daily notes, ...)
- Dutch & French



≠ Computer assisted coding (CDC)

# UZB: Innovator Hospital

- Training data and pseudonymization
- Application testing
- Integration in PrimUZ



PrimUZ

Admin info

Discharge letter  
Daily notes  
...

ICD  
ICD-10 codes

REST API

**Case search** | All Cases | Saved for later

Search results (5/17)

Name	Gender	Age	Discharge 1	DRG	SDI	PCs	SDI no.	Speculation	UVC	Confidence	Manually Reviewed	Approved
[Redacted]	♂	5	21.11.2023	956	---	---	0	Pediatrie Gastro	No	Low	---	---
[Redacted]	♂	11	02.01.2024	261	2	K980	3	Pediatrie Gastro	No	Low	---	---
[Redacted]	♂	9	02.01.2024	862	---	---	0	Pediatrie Gastro	No	Low	---	---
[Redacted]	♂	14	02.01.2024	254	2	K803	6	Pediatrie Gastro	No	Low	---	---
[Redacted]	♀	17	02.01.2024	254	2	K580	4	Pediatrie Gastro	No	Low	---	---
[Redacted]	♀	15	02.01.2024	254	1	K900	2	Pediatrie Gastro	No	Medium	---	---
[Redacted]	♀	6	02.01.2024	254	1	K900	3	Pediatrie Gastro	No	Medium	---	---
[Redacted]	♀	6	02.01.2024	254	1	E739	2	Pediatrie Gastro	No	Medium	---	---
[Redacted]	♀	7	02.01.2024	254	1	E739	0	Pediatrie Gastro	No	Medium	---	---
[Redacted]	♀	2	02.01.2024	254	1	Z4659	3	Pediatrie Gastro	No	Medium	---	---

**Dietaadvies** | 11/2024 Uivoersier

Verantwoordelijke: [Redacted]

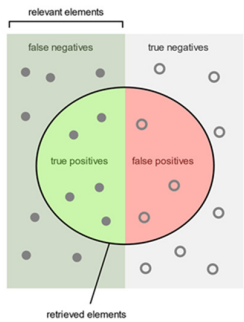
**Problems and Allergies** | Feb-2024, 18:17

**Codes (3)**

- E73.9 P [Redacted] Lactose intolerantie, unspecified
- Z91.0A8 Other immunological substance allergy status
- J45.20 Mild intermittent asthma, uncomplicated

# 000 F1 SCORE – MACHINE LEARNING ACCURACY

- Precision = how many of the “positive” predictions are correct /How many retrieved items are relevant?  
 $\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}}$
- Recall = how many of the positive class samples present in the dataset were correctly identified by the model / How many relevant items are retrieved?  
 $\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}$
- F1 score = harmonic mean of Precision & Recall



$$\text{Precision} = \frac{5}{8} = 62,5\%$$

$$\text{Recall} = \frac{5}{12} = 41,7\%$$

Test data performance	F1	Precision	Recall
Diagn&Proc	72.9%	71.2%	74.7%

Production data	F1	Precision	Recall
Diagnoses	55.9%	61%	51.5%
Procedures	58.8%	69.4%	51%

Production data + retraining (700 cases)	F1	Precision	Recall
Diagnoses	62.4%	64.8%	60.1%
Procedures	62.2%	68.3%	57.1%

# 000 Qualitative analysis – impact of data drift

## Test data-set

Case	DX Codes	PX Codes	PDX quality	Missing DX	Evidence quality	Bad evidences	Pdx	Pdx Prob	Drp Soi	Confidence
xxxxxxx	Poor	N/A	Poor	Some	Poor	0	G30.9	64	042.0	Medium
6248327	Good	Good	Good	0	Good	1	I25.10	95	191.1	Medium
6248298	Good	Good	Good	0	Good	0	I25.10	99	191.1	Medium
6248021	Good	Good	Good	0	Good	0	K01.1	95	114.1	Medium
6247878	Good	Good	Good	0	Good	0	I87.2	99	180.1	Medium
6247186	Good	Good	Good	1	Good	1	N40.1	99	501.1	Medium
6248482	Good	Good	Avg	0	Avg	0	C91.00	47	690.2	Low
6244004	Good	Good	Good		Good	0	G56.01	99	026.1	Low
6240069	Good	Good	Poor				Z5111	0	681.1	Low

## Production data

Case	DX Codes	PX Codes	PDX quality	Missing DX	Evidence quality	Bad evidences	Pdx	Pdx Prob	Drp Soi	Confidence
1	Good	Missing	Good	0	Poor	1	G89.29	96	861,1	High
2	Missing	Missing	Good	6	Good	0	M81.0	99	2,1	High
3	Good	Missing	Good		Good	0	G89.29	97	861,1	High
4	Poor	Good	Good		Poor	2	Z5111	99	696,2	Medium
5	Poor	Good	Good	4	Poor	2	Z5111	98	696,2	Medium
6	Good	Good	Good	4	Avg	0	Z5111	99	696,2	Medium
9	Good	Good	Good	0	Avg	1	Z0389	18	861,1	Medium
10	Good		Good	Some	Avg	0	G31.9	41	042,1	Medium
11	Poor	Poor	Poor		Poor	1	N20.1	88	465,1	Medium
12	Good	Avg	Good	Some	Good	0	B20	3	894,2	Low
13	Avg	Missing	Poor		Avg	1	Z5111	0	696,1	Low
14	-	Missing	Good		Good		K60.3	2	254,1	Low

## CONFIDENCE LEVELS

How to autocode with an F1 score of 62%?

Discharge	DRG	SOI	PDx	SDx no.	Specialism	LVC	Confidence ↓
21.07.2021	228	1	K4090	1	Chirurgie	No	● High
22.12.2023	263	1	K8020	1	Chirurgie	No	● High
19.03.2021	316	1	M720	2	Orthopedie	No	● High
15.05.2021	501	1	N471	0	Urologie	No	● Medium
19.08.2023	315	1	S52531A	1	Spoedgevallendienst	No	○ Low
19.07.2021	385	1	L723	0	Chirurgie	No	○ Low

Confidence ↓
● High
● High
● High
● Medium
○ Low
○ Low
○ Low

High confidence cases: F1 score of **88%** → “Autocodable”

precision: 91% | Recall 84%

Medium confidence cases: F1 score of **85%**

precision: 86% | Recall 84%

Low confidence cases: F1 score of **53%**

precision: 85% | Recall 39%

## UZB – 360 ENCOMPASS - > 5000 DAYCARE ADMISSIONS

1. Overall experience: 5/7 positive (71%) – 2/7 negative (29%)
2. Does 360 help you in the coding process: 5/7 positive – 1 neutral – 1 negative
3. What about the coding process/application:
  - Old coding process: “I want to adjust 360 to have the most “accurate” diagnoses”
  - Lay-out problems
  - Missing information

Change management, embedded in AI reality!

AI literacy

AI training



## > 5000 CODES CASES: 60% DID NOT REQUIRE ANY CODE REJECTION - 36% DID NOT REQUIRE ANY CODE ADDITION

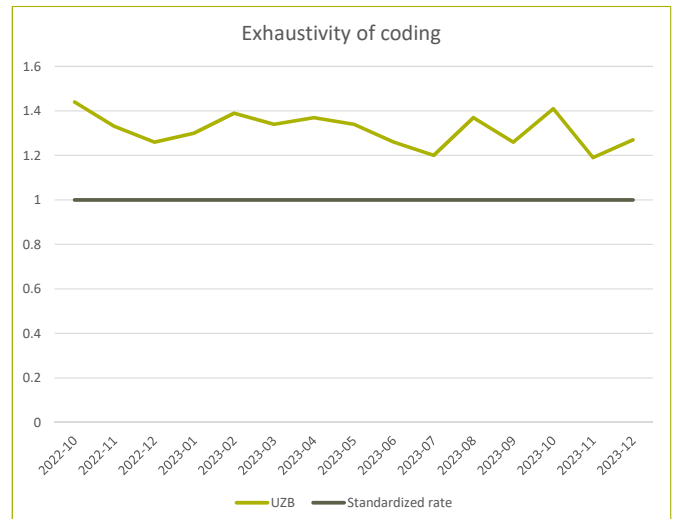
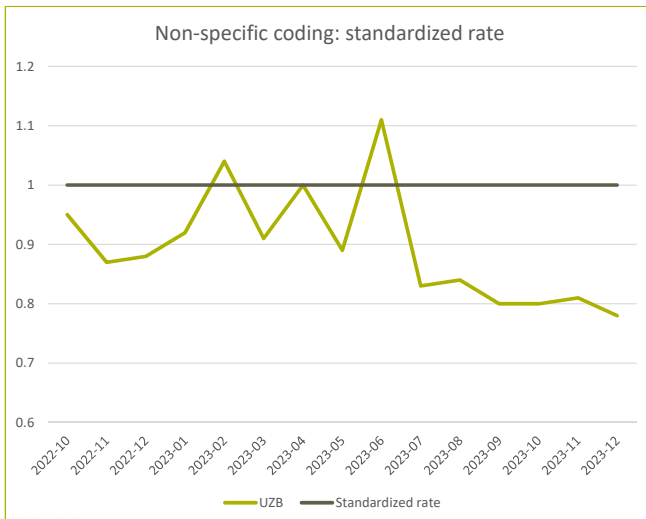
If there are few or no codes being adjusted, does this mean the system is performing well, or could it indicate that the coders were not sufficiently attentive?



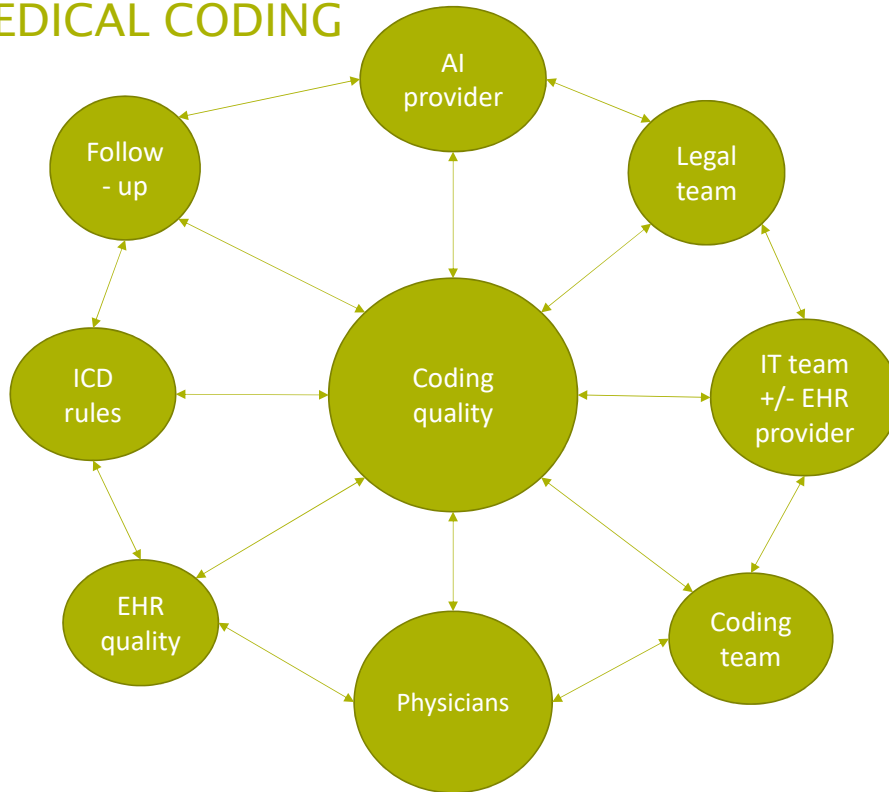
## CODING QUALITY – 3M/SOLVENTUM BENCHMARK

2023Q4: 11.878 daycare adm – 4900 coded 360 (41,25%)

➔ Gastroenterology: 60% coded

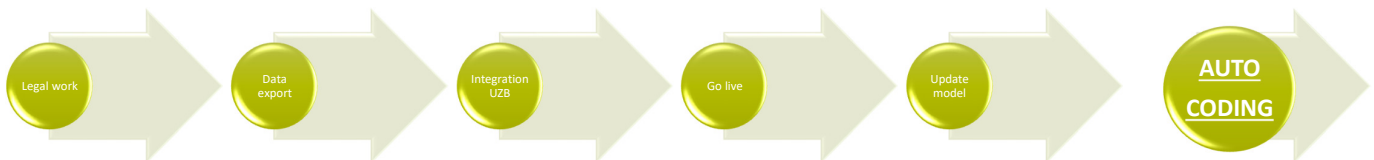


## AI AND MEDICAL CODING



## WHERE ARE WE NOW?

Geriatric daycare  
 Gastroenterology  
 Dental  
 ENT  
 Urology  
 Cardiology?  
 ...





# THANK YOU

Karen Pien: [karen.pien@uzbrussel.be](mailto:karen.pien@uzbrussel.be)

Maarten Lambrecht: [mlambrecht@solventum.com](mailto:mlambrecht@solventum.com)

Elric Verbruggen: [everbruggen@solventum.com](mailto:everbruggen@solventum.com)



Universitair  
Ziekenhuis  
Brussel

