



University of
Zurich^{UZH}

BIO392 Bioinformatics of Genome Variations

Survival | Classifications

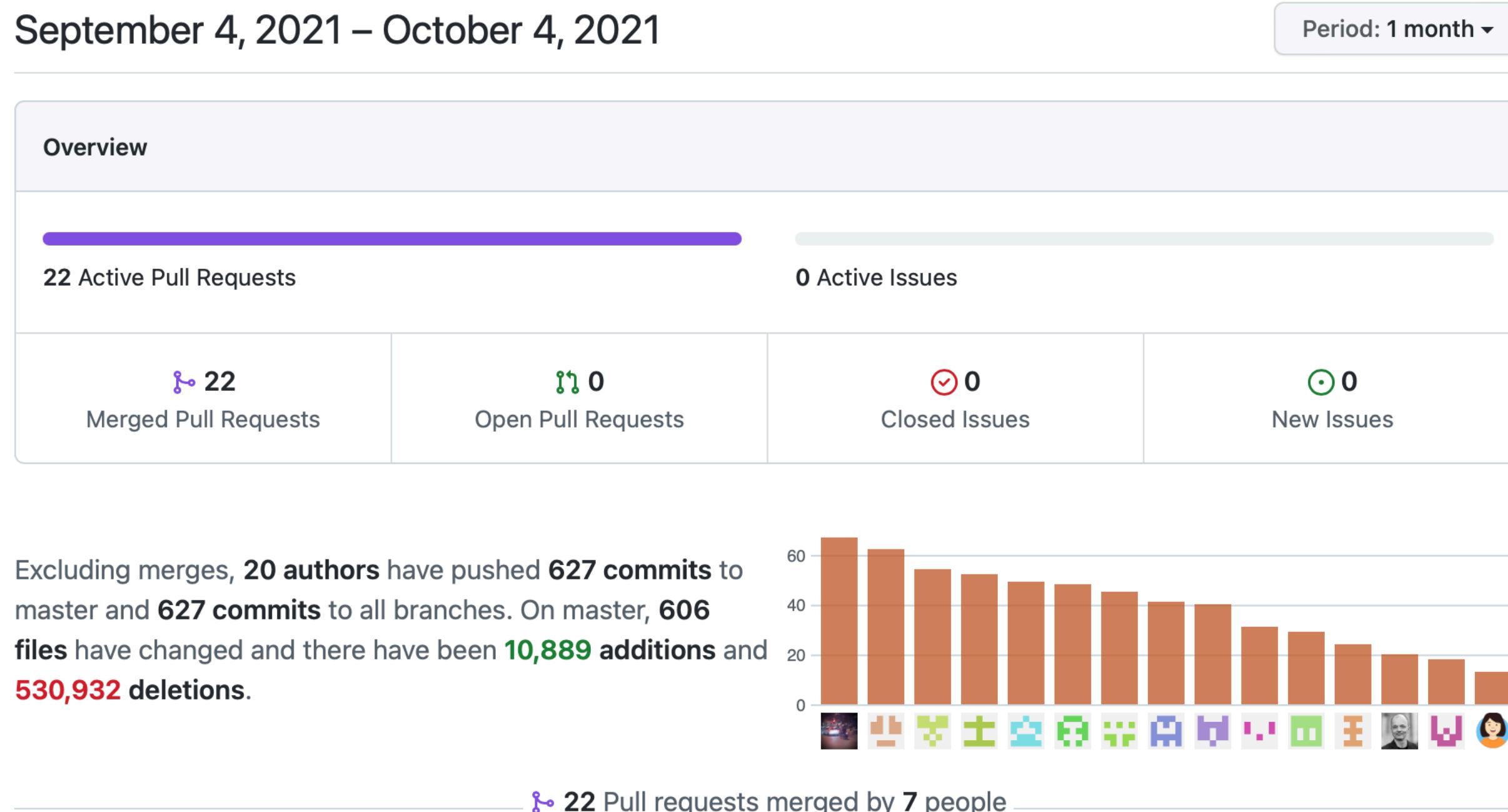
Michael Baudis **UZH SIB**
Computational Oncogenomics

Task: Exploration of different file formats

- Which genomic file formats exist & what are their use cases?
 - SAM
 - BAM
 - CRAM
 - VCF
 - FASTA
 - MPEG-G

BIO392 HS 2021

Github Activity



Survival

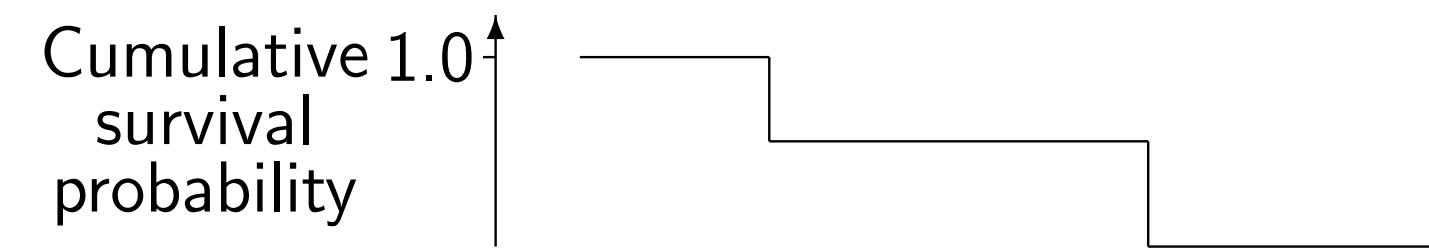
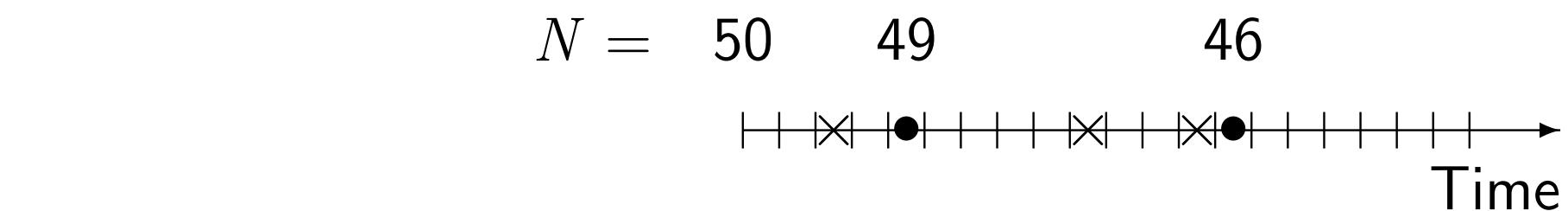
Kaplan-Meier Analysis of Survival Based on Conditional Probabilities

The Kaplan-Meier Method

- ▶ The most common method of estimating the survival function.
- ▶ A non-parametric method.
- ▶ Divides time into small intervals where the intervals are defined by the unique times of failure (death).
- ▶ Based on conditional probabilities as we are interested in the probability a subject surviving the next time interval given that they have survived so far.

Kaplan–Meier method illustrated

(● = failure and × = censored):



- ▶ Steps caused by multiplying by $(1 - 1/49)$ and $(1 - 1/46)$ respectively
- ▶ Late entry can also be dealt with

[Kaplan-Meier estimators \(km-na\)](#)

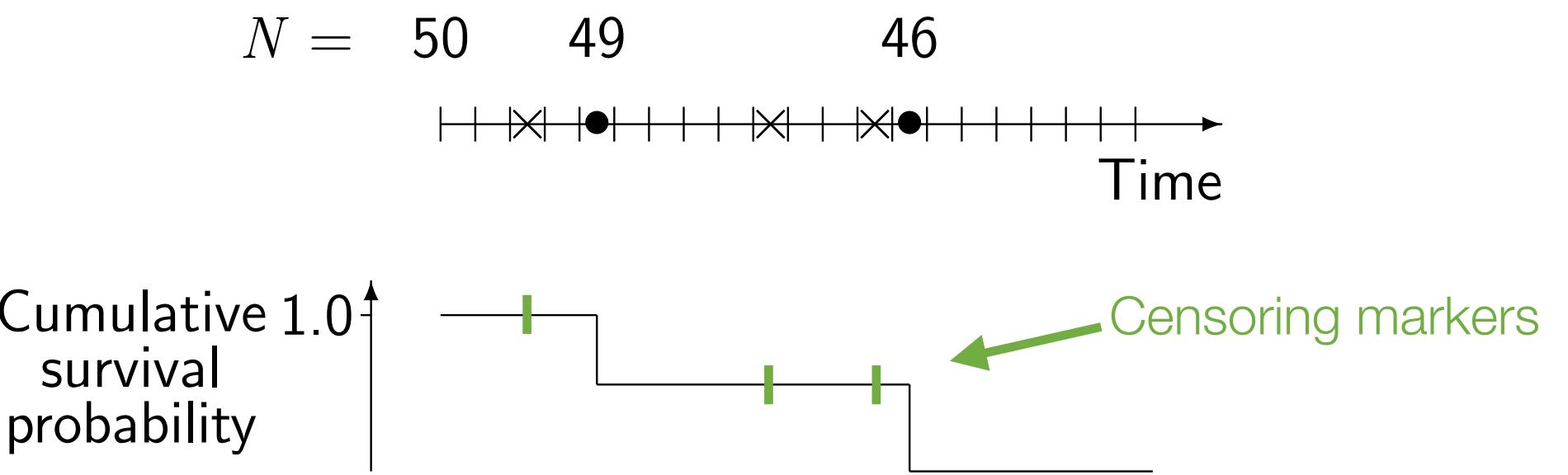
[Kaplan-Meier estimators \(km-na\)](#)

The Kaplan-Meier Method

- ▶ The most common method of estimating the survival function.
- ▶ A non-parametric method.
- ▶ Divides time into small intervals where the intervals are defined by the unique times of failure (death).
- ▶ Based on conditional probabilities as we are interested in the probability a subject surviving the next time interval given that they have survived so far.

Kaplan–Meier method illustrated

(● = failure and × = censored):



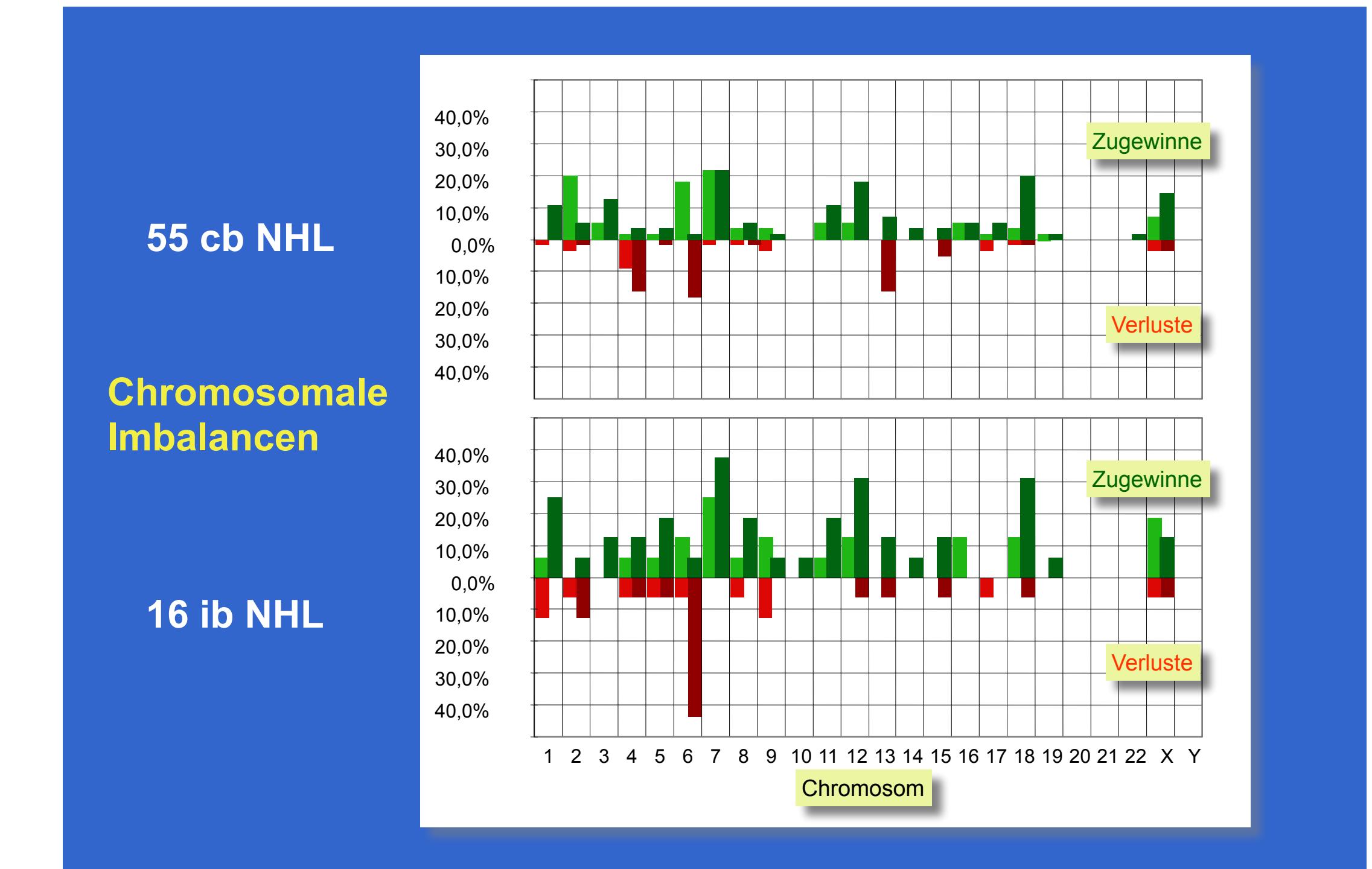
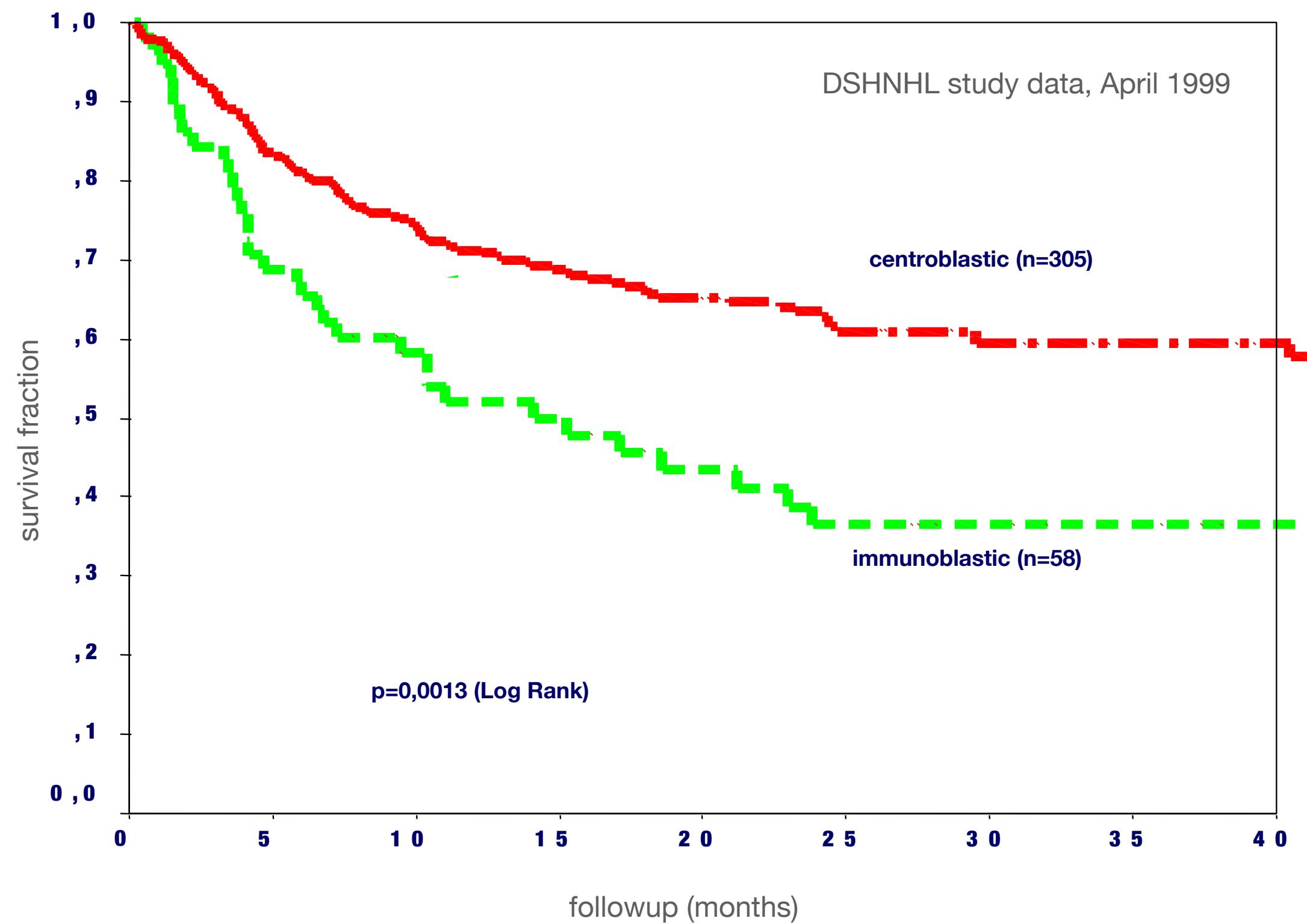
- ▶ Steps caused by multiplying by $(1 - 1/49)$ and $(1 - 1/46)$ respectively
- ▶ Late entry can also be dealt with

Kaplan-Meier estimators (km-na)

Kaplan-Meier estimators (km-na)

Cancer CNVs | Diagnostics | Prognosis

Single-study CNV frequencies correspond to diagnostic subsets



Kaplan-Meier Plots to Visualize Differential Risk

Multi-parametric "risk scores" in CLL Prognosis

Leukemia (2020) 34:1038–1051
<https://doi.org/10.1038/s41375-020-0727-y>

ARTICLE

Chronic lymphocytic leukemia

Prognostic model for newly diagnosed CLL patients in Binet stage A: results of the multicenter, prospective CLL1 trial of the German CLL study group

Manuela A. Hockstetter¹ · Raymonde Busch² · Barbara Eichhorst³ · Andreas Bühlert⁴ · Dirk Winkler⁴ · Jasmin Bahlo³ · Sandra Robrecht³ · Michael J. Eckart² · Ursula Vehling-Kaiser⁵ · Georg Jacobs² · Ulrich Jäger⁸ · Hans-Jürgen Hurtz² · Georg Hopfinger¹⁰ · Frank Hartmann¹¹ · Harald Fuss¹² · Wolfgang Abenhards¹³ · Ilona Blau¹⁴ · Werner Freier¹⁵ · Lothar Müller¹⁶ · Maria Goebeler¹⁷ · Clemens Wendtner^{1,3} · Kirsten Fischer³ · Carmen D. Herling³ · Michael Stärck¹ · Martin Bentz¹⁸ · Bertold Emmerich¹⁹ · Hartmut Döhner²⁰ · Stephan Stilgenbauer²⁰ · Michael Hallek³

Table 2a Results of the Cox's regression for OS and TTFT in CLL patients in whom all 30 baseline parameters were available.

Univariate comparison	Hazard ratio [HR]	95% Confidence Interval		<i>P</i> value
		Lower	Upper	
COX regression OS				
Cytogenetic Hierarchical Type				
del(17p) vs. not del(17p)/del(11q)	3.8	2.1	7.1	<0.001
del(11q) vs. not del(17p)/del(11q)	2.0	1.2	3.5	0.008
LDT				
<12 months vs. ≥12 months	1.9	1.3	2.8	0.001
Age, years				
>60 vs. ≤60	1.8	1.2	2.7	0.002
B2M, mg/dL				
>3.5 vs. ≤3.5	2.0	1.2	3.1	0.004
IGHV mutational status				
Unmutated vs. mutated	2.4	1.6	3.6	<0.001
COX regression TTFT				
Cytogenetic Hierarchical Type				
del(17p) vs. not del(17p)/del(11q)	2.2	1.2	4.1	0.009
del(11q) vs. not del(17p)/del(11q)	2.0	1.3	3.0	0.001
LDT				
vs. <12 months	2.3	1.7	3.1	<0.001
Age, years				
>60 vs. ≤60	1.3	1.0	1.7	0.037
B2M, mg/dL				
>3.5 vs. ≤3.5	1.5	1.0	2.3	0.049
IGHV mutational status				
Unmutated vs. mutated	4.4	3.2	5.9	<0.001

Table 2b Allocation of risk score points to the distinctive factors of the CLL1-PM.

	HR (95% CI)	<i>P</i>	Allocated risk score points
Characteristics			
Del(17p)	3.8 (2.1–7.1)	<0.001	3.5
Unmutated IGHV	2.4 (1.6–3.6)	<0.001	2.5
Del(11q)	2.0 (1.2–3.5)	0.008	2.5
Beta2-MG >3.5 mg/L	2.0 (1.2–3.1)	0.004	2.5
LDT<12 months	1.9 (1.3–2.8)	0.001	1.5
Age >60 years	1.8 (1.2–2.7)	0.002	1.5

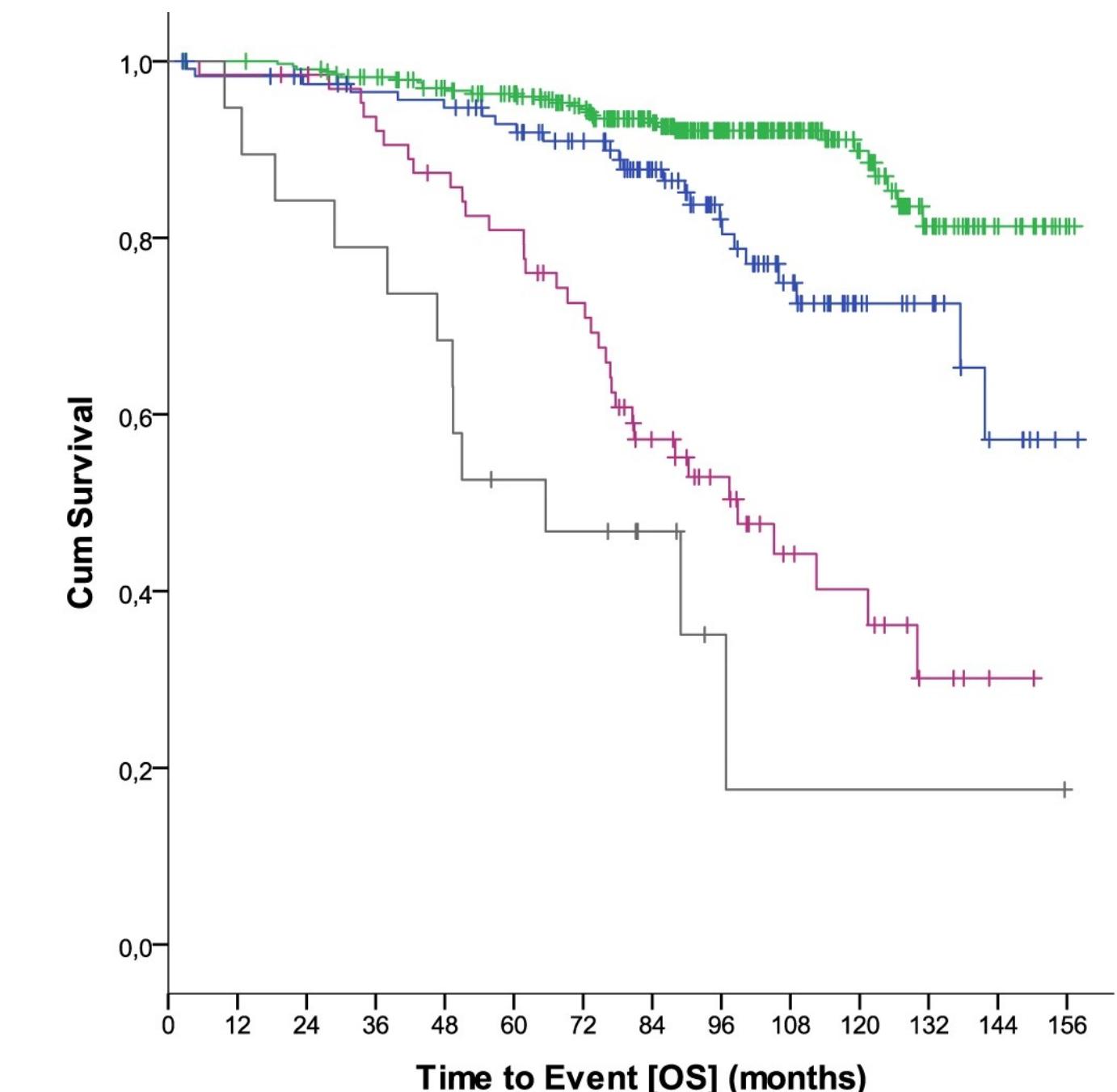
The assigned risk score points derived from the HR for OS of the individual factors.

Table 2c Patients and risk groups according to the CLL1 Prognostic Model (CLL1-PM). Patients and risk groups according to the CLL-IPI.

	Index score	Patients N (%)
Risk Groups according to the CLL1-PM		
Very low	0.0–1.5	336 (62.3)
Low	2.0–4.0	119 (22.1)
High	4.5–6.5	65 (12.1)
Very high	7.0–14.0	19 (3.5)
Risk Groups according to the CLL-IPI		
Low	0–1	360 (66.8)
Intermediate	2–3	141 (26.2)
High	4–6	33 (6.1)
Very high	7–10	5 (0.9)

OS overall survival, HR hazard ratio, Beta2-MG beta-2 microglobulin, IGHV immunoglobulin heavy-chain genes, LDT lymphocyte doubling time, TTFT time-to-first treatment.

- "a novel prognostic model (CLL1-PM) developed to identify risk groups, separating patients with favorable from others with dismal prognosis"
- " findings would be useful to effectively stratify Binet stage A patients, particularly within the scope of clinical trials evaluating novel agents"



***P* < 0.001**

Number at risk	0	12	24	36	48	60	72	84	96	108	120	132	144	156
Very low	336	335	331	322	306	294	262	215	160	113	68	33	15	2
Low	119	115	111	108	106	100	89	71	49	34	19	14	6	1
High	65	64	63	59	54	50	43	29	21	12	10	4	1	0
Very high	19	18	16	15	13	9	8	5	2	1	1	1	1	0

Discrimination: C-statistics, C = 0.739 (95% CI, 0.686–0.790)
AIC=445

Overall survival according to the CLL1-PM risk groups. The full analysis dataset is comprised of the dataset of 539 patients.

Cancer Classifications & Parameters

NCIt | ICD-O / WHO | TNM

ICD-O 3

WHO International Classification of Diseases for Oncology, 3rd Edition (ICD-O-3)

- used in cancer registries for coding the site (topography) and the histology (morphology) of neoplasms, usually obtained from a pathology report
- mix of "biology" (i.e. tumor morphology) and "clinical" (i.e. tumor site)

→ 2 codes per cancer

- "Adenocarcinoma" of the "Sigmoid colon"

8140/3

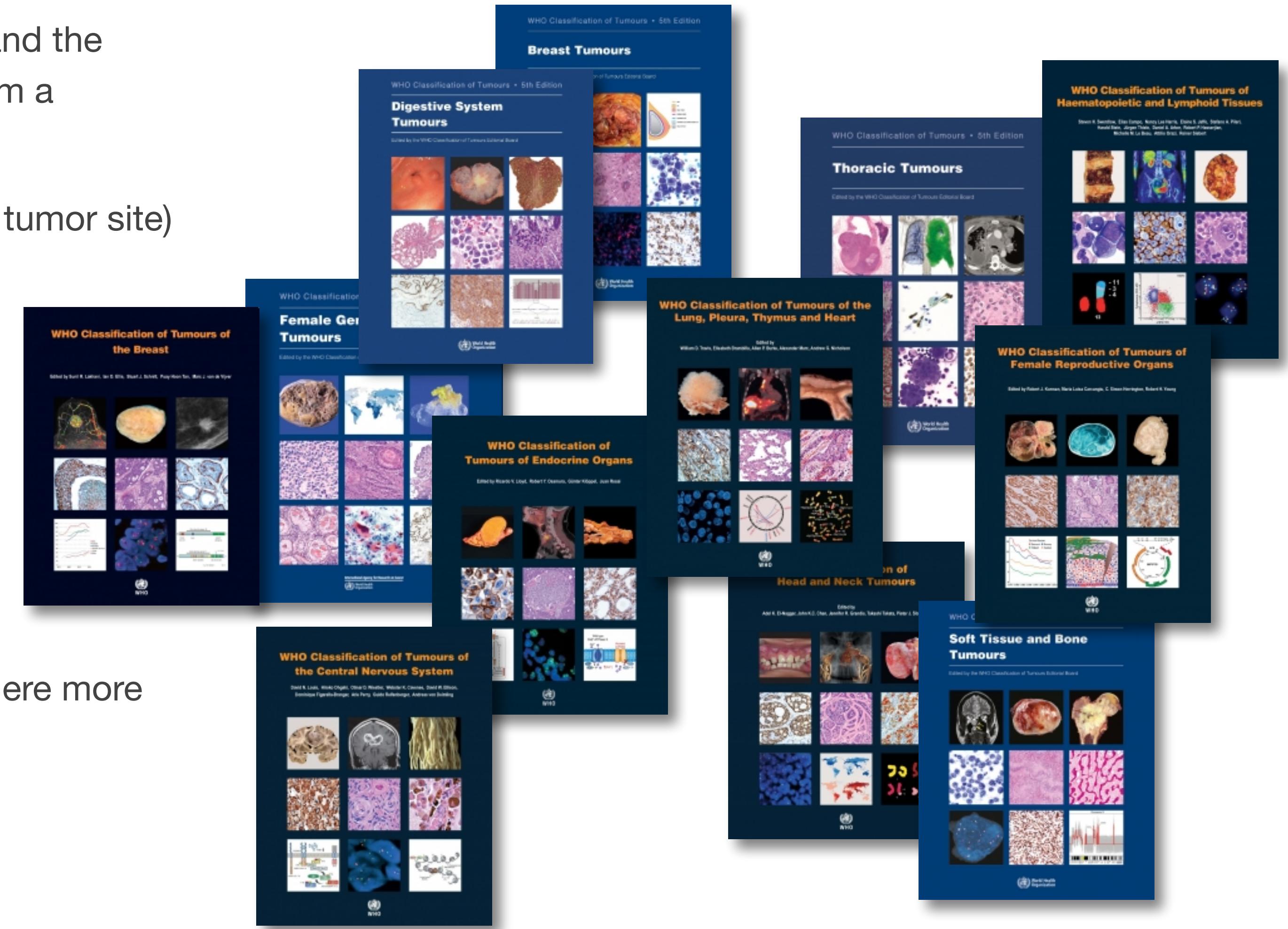
C18.7

- "Retinoblastoma" of the "Retina"

9510/3

C69.2

- widely accepted by pathologists but limited clinical use (there more ICD-10 or SNOMED)
- no ontology & not (truly) hierarchical
- many entities difficult to remap if using only single code



NCIt

Neoplasm Classifications in the NCI Thesaurus

- NCI's core reference terminology and biomedical ontology are collected in the NCI Thesaurus (NCIt)
- individual codes for site-specific occurrences of "biological" diagnoses

1 code per cancer

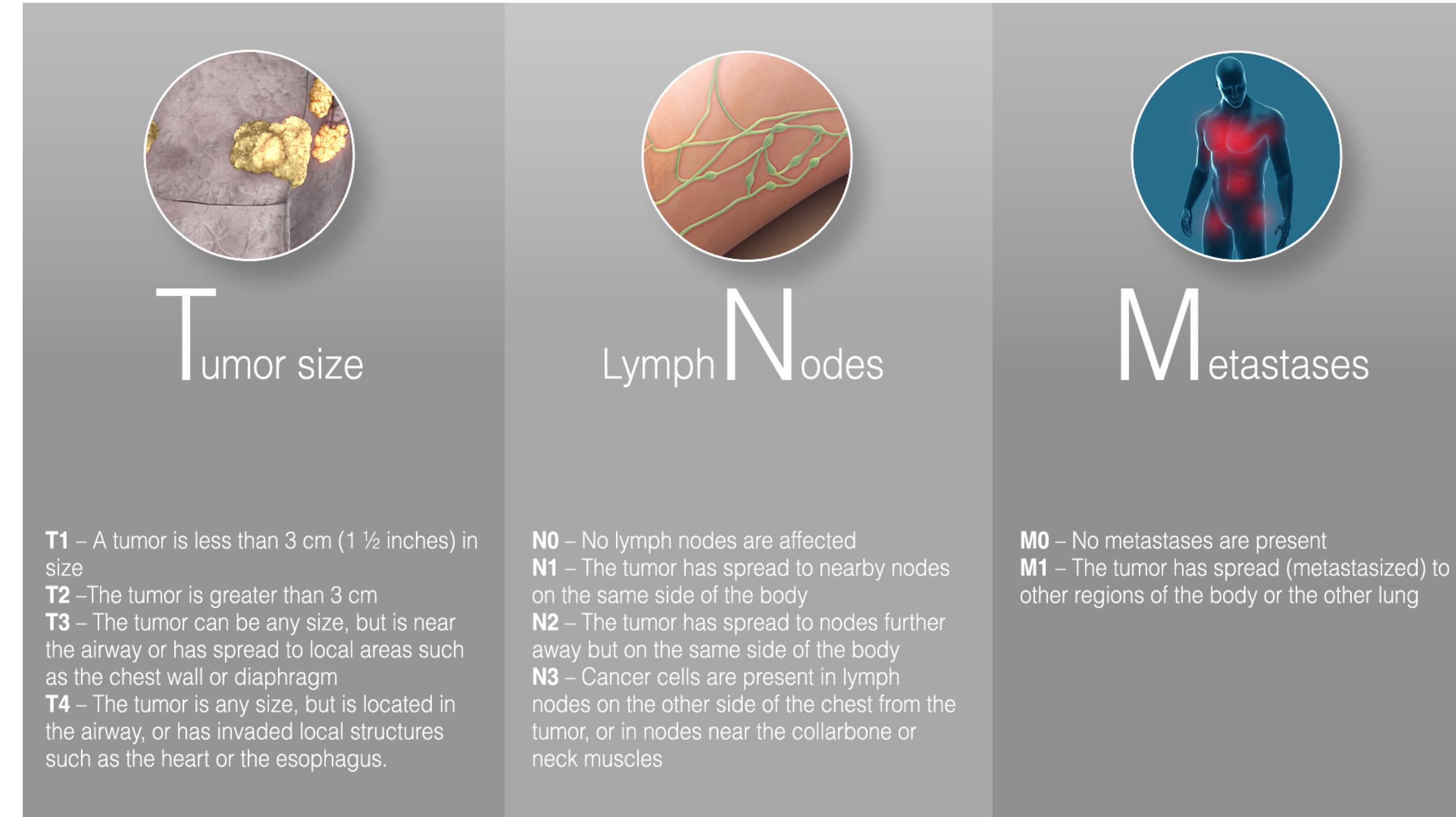
- ▶ **NCIT:C43584** - Rectosigmoid Adenocarcinoma
- ▶ **NCIT:C7541** - Retinoblastoma
- truly hierarchical ontology
- hierarchical system empowers "logical OR" queries
- terms can have multiple occurrences in diagnostic tree
- assignment of code to different groupings allows soft aggregation (e.g. a type of colorectal adenocarcinoma with all colon tumors or with all adenocarcinomas)

- ▼ NCIT:C3262: Neoplasm (116013 samples)
- ▼ NCIT:C3263: Neoplasm by Site (110893 samples)
 - NCIT:C156482: Genitourinary System Neoplasm (16534 samples)
 - NCIT:C2910: Breast Neoplasm (15957 samples)
 - NCIT:C3010: Endocrine Neoplasm (3521 samples)
 - NCIT:C3030: Eye Neoplasm (280 samples)
- ▼ NCIT:C3052: Digestive System Neoplasm (15289 samples)
 - NCIT:C172852: Digestive System Soft Tissue Neoplasm (99 samples)
 - NCIT:C27721: Digestive System Neuroendocrine Neoplasm (202 samples)
 - NCIT:C2877: Anal Neoplasm (61 samples)
 - NCIT:C3028: Esophageal Neoplasm (1865 samples)
- ▼ NCIT:C3141: Intestinal Neoplasm (5723 samples)
 - ▼ NCIT:C2956: Colorectal Neoplasm (5579 samples)
 - NCIT:C2953: Colon Neoplasm (4666 samples)
 - NCIT:C3350: Rectal Neoplasm (527 samples)
 - NCIT:C4610: Benign Colorectal Neoplasm (181 samples)
 - ▼ NCIT:C4877: Rectosigmoid Neoplasm (240 samples)
 - ▼ NCIT:C7420: Malignant Rectosigmoid Neoplasm (240 samples)
 - ▼ NCIT:C7421: Rectosigmoid Carcinoma (240 samples)
 - ▼ NCIT:C43584: Rectosigmoid Adenocarcinoma (240 samples)
 - NCIT:C43592: Rectosigmoid Mucinous Adenoca... (18 samples)
 - NCIT:C4978: Malignant Colorectal Neoplasm (5398 samples)
 - NCIT:C96152: Colorectal Neuroendocrine Neoplasm (11 samples)
 - NCIT:C4432: Small Intestinal Neoplasm (66 samples)

TNM

A Classification for Clinical Cancer Stage Parameters

- most widely used cancer staging system
- T** refers to the size and extent of the main tumor
- N** refers to the number / location of nearby lymph nodes that have cancer infiltration
- M** refers to whether the cancer has metastasized
- not used for leukemias / lymphomas
 - Binet and Rai in CLL
 - proportion of blasts in bone marrow or blood in leukemias
 - Lugano classification in lymphomas
- other disease specific staging systems may (co-) exist
 - e.g. a stage II breast cancer is determined by size & nodal involvement



Source: www.scientificanimations.com

TNM

A Classification for Clinical Cancer Stage Parameters

- most widely used cancer staging system
- T** refers to the size and extent of the main tumor
- N** refers to the the number / location of nearby lymph nodes that have cancer infiltration
- M** refers to whether the cancer has metastasized
- not used for leukemias / lymphomas
 - Binet and Rai in CLL
 - proportion of blasts in bone marrow or blood in leukemias
 - Lugano classification in lymphomas
- other disease specific staging systems may (co-) exist
 - e.g. a stage II lung cancer is determined by size & nodal involvement

TNM STAGING OF LUNG CANCER - 8th EDITION

The diagram illustrates the TNM staging of lung cancer, organized into three main sections:

- DISTANT METASTASIS (M)**: This section includes Stage IV B (Any T, Any N, M1c) and Stage IV A (Any T, Any N, M1a/b). Stage IV B is associated with multiple extrathoracic metastases (in one or more organs). Stage IV A includes single extrathoracic metastasis (including non-regional lymph nodes) and satellite (separate) tumor nodule(s) in contralateral lobe or Pleural or pericardial nodules or malignant effusion.
- No distant metastasis (M0)**: This section includes Stage III B, Stage III A, Stage II B, and Stage I A1.
- Distant metastasis (M1)**: This section includes Stage III C, Stage III B, Stage III A, Stage II B, Stage T1, Stage T2, Stage T3, and Stage T4.

Explanation of lymph node staging:

- For any N category, one or more of the groups marked by ● must be involved and the involvement of all groups marked by □ should be absent.
- The presence or absence of involvement in groups marked by □ does not alter N staging in the corresponding category.

PRIMARY TUMOR (T)

1- Size (greatest dimension)		2- Criteria of Extent		Local Invasion		Separate Tumor Nodule(s)	
Stage I A1 (T1 (mi) N0 M0)	T1 (mi): Minimally invasive adenocarcinoma (solitary adenocarcinoma, ≤ 3 cm with a lepidic growth and ≤ 5 mm invasion in any focus)	Endo-bronchial Location	No extension proximal to the lobar bronchus **	Main bronchus (regardless of the distance to the carina)** vs. Atelectasis or obstructive pneumonitis extending to the hilum (entire or part of the lung)	Chest wall (Including superior sulcus), phrenic nerve, parietal pleura and/or parietal pericardium	Diaphragm, Mediastinum, heart, great vessels, recurrent laryngeal nerve, esophagus and/or vertebral body	
Stage 0 (Tis N0 M0)	Tis: Carcinoma in situ	Local Invasion	None; the tumor is surrounded by lung or visceral pleura	Visceral pleura	Present in the same lobe of the primary tumor	Present in a different ipsilateral lobe	
Occult Carcinoma (Tx N0 M0)	Tx: Tumor is proven histopathologically (+ Cytology) but not detected by imaging or bronchoscopy	Separate Tumor Nodule(s)	Absent	Absent			

Lababede O, Meziane MA. The Eighth Edition of TNM Staging of Lung Cancer: Reference Chart and Diagrams. Oncologist. 2018;23(7):844-848.

Tasks

Survival analyses | Cancer classifications | Staging

- Familiarize yourself with the different concepts behind different disease classification systems - what are there use, advantages, problems? E.g. ICD-10, ICD-O, NCI
 - you can use Progenetix to explore e.g. ontology mapping
- Learn to "read" Kaplan-Meier plots (preparation for explorative analyses later this week).
- Achieve a principal understanding of TNM codes & write some "translations"
 - T1N1M0: small tumor with regional lymph node involvement and no detected distant metastases