University of Zurich^UZH

BIO392
Bioinformatics of Genome Variations

Survival | Classifications

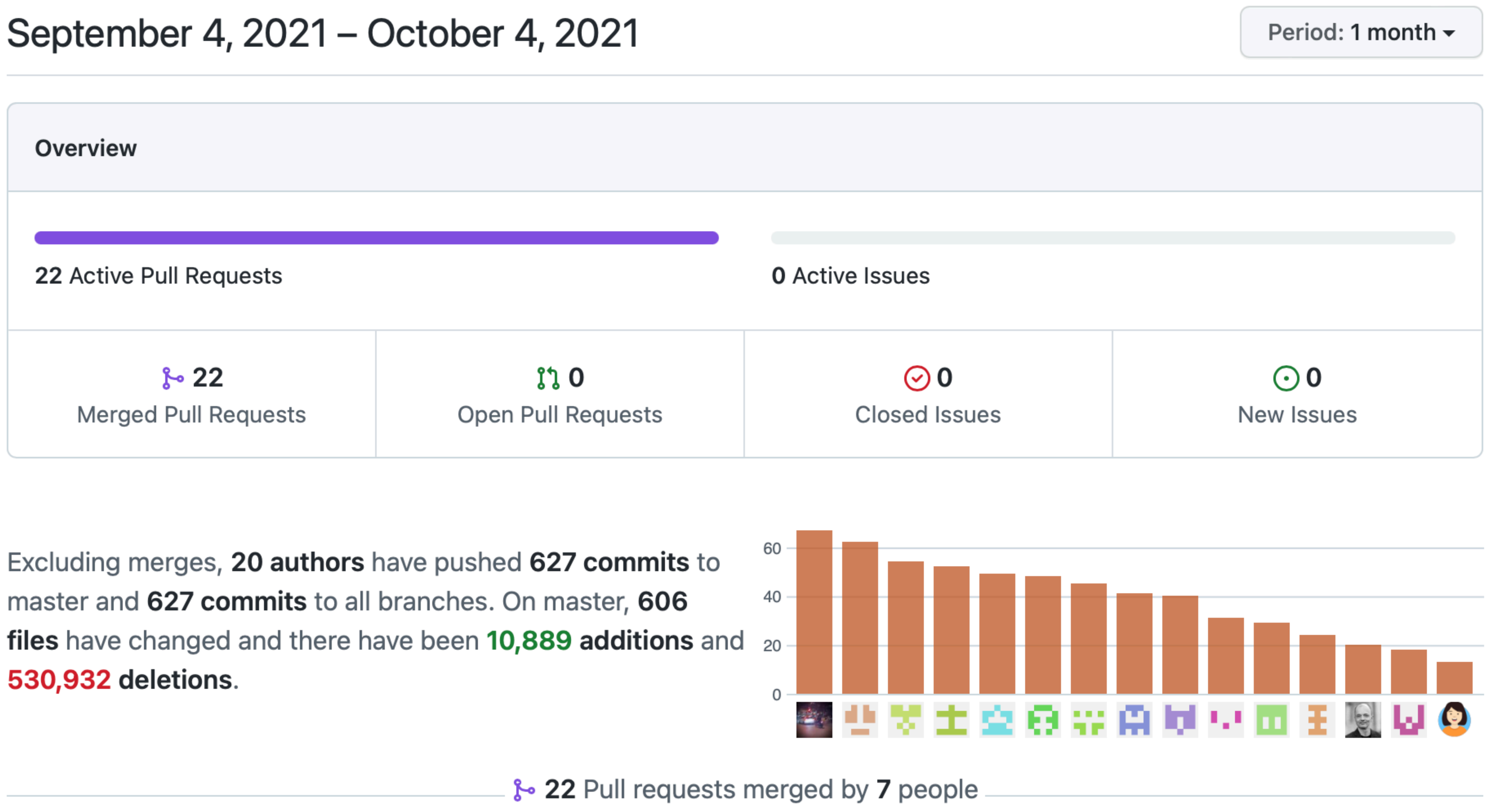
Michael Baudis **UZH** **SIB**
Computational Oncogenomics

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



September 4, 2021 – October 4, 2021

Period: **1 month** ▾

Overview

22 Active Pull Requests

0 Active Issues

22	0	0	0
Merged Pull Requests	Open Pull Requests	Closed Issues	New Issues

Excluding merges, **20 authors** have pushed **627 commits** to master and **627 commits** to all branches. On master, **606 files** have changed and there have been **10,889 additions** and **530,932 deletions**.

22 Pull requests merged by **7** people

Survival

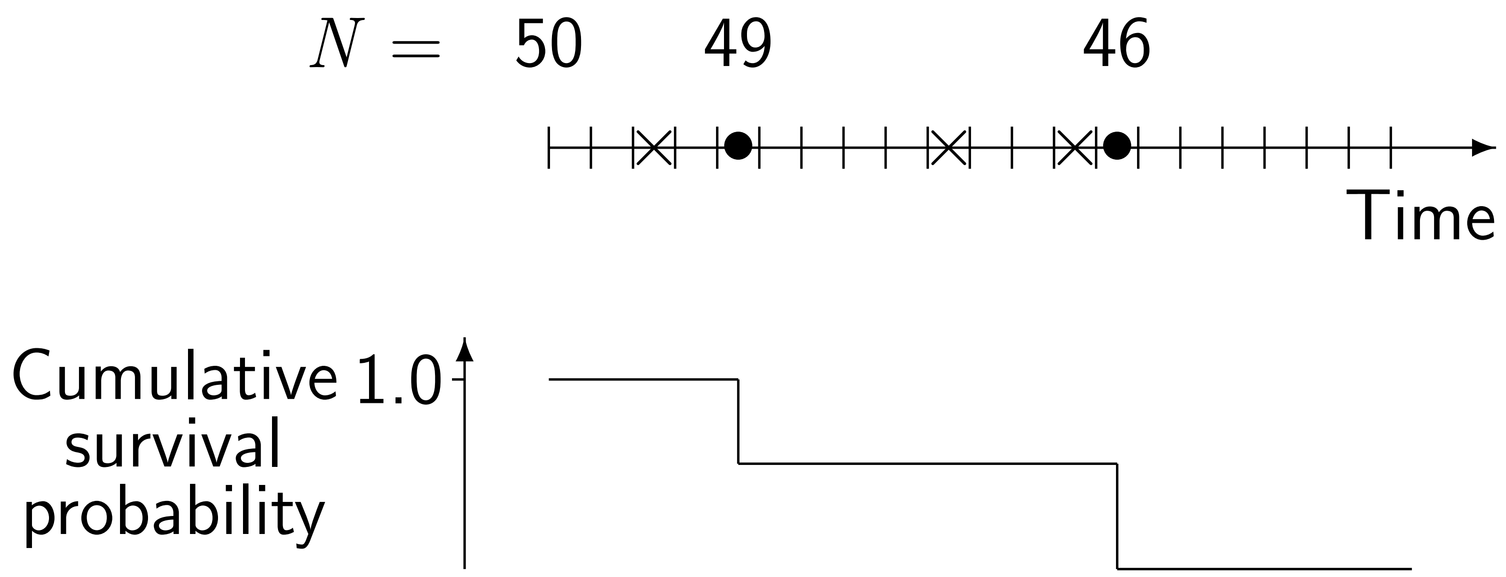
Kaplan-Meier Analysis of Survival Based on Conditional Probabilities

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 $\times$ $=$ censored):



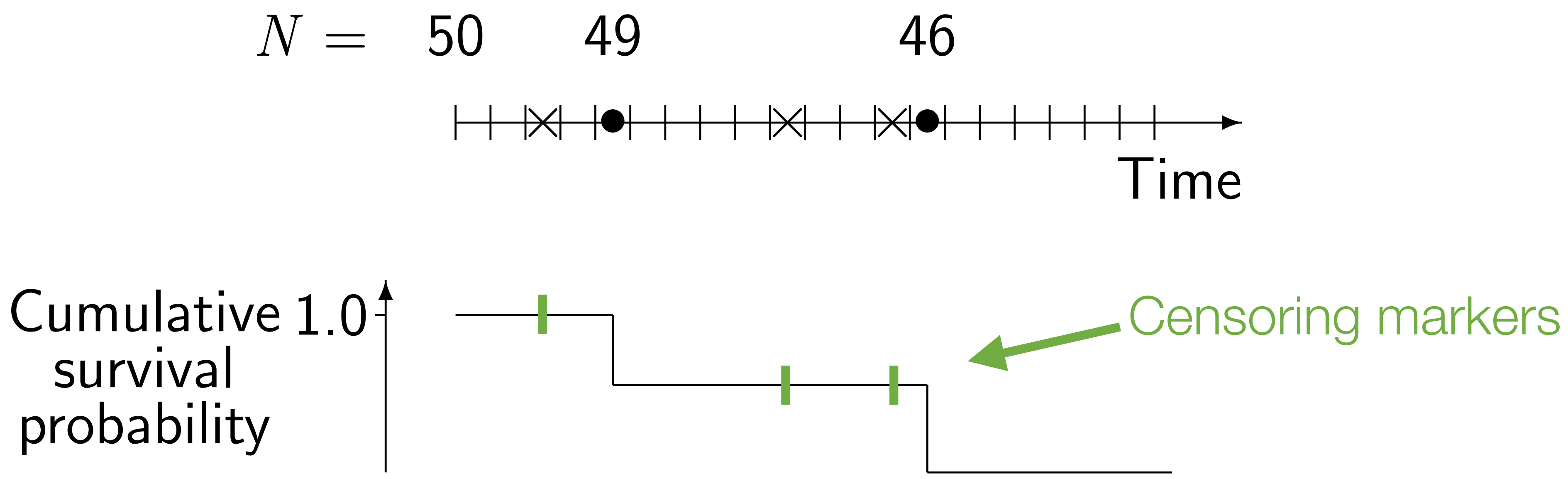
- Steps caused by multiplying by $(1-1/49)$ and $(1-1/46)$ respectively
- Late entry can also be dealt with

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

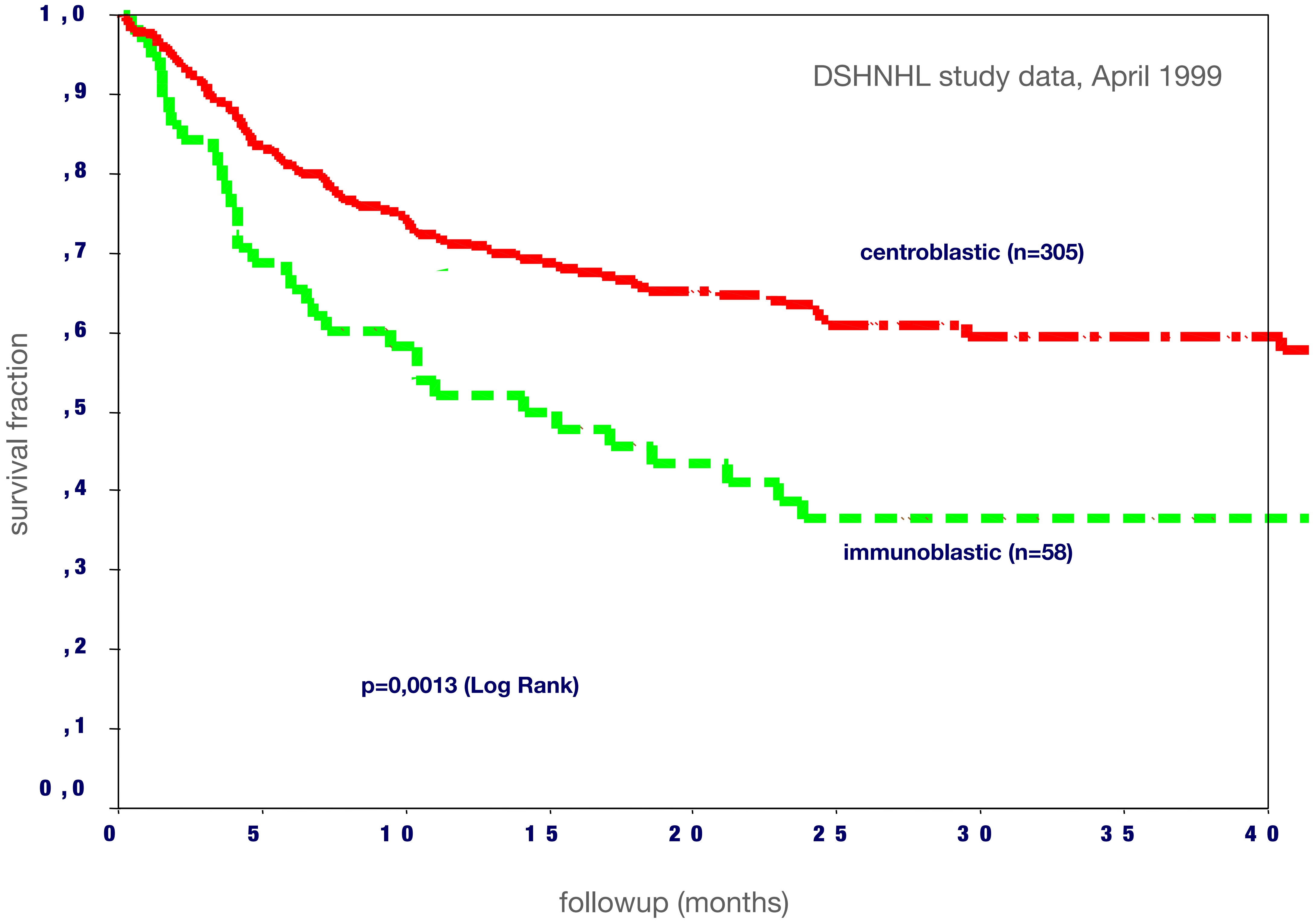
(● = failure and × = censored):

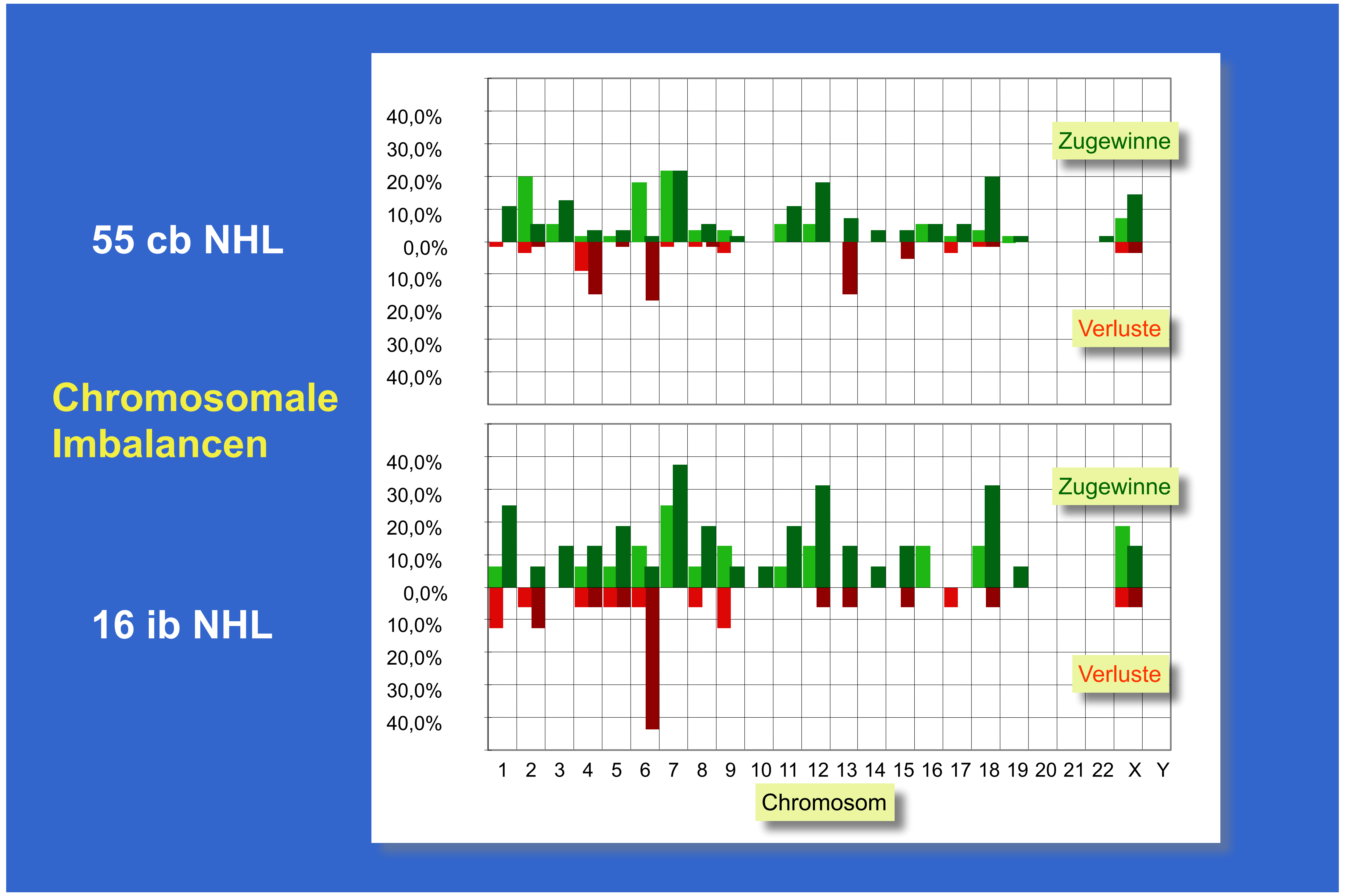


- Steps caused by multiplying by $(1 - 1/49)$ and $(1 - 1/46)$ respectively
- Late entry can also be dealt with

Cancer CNVs | Diagnostics | Prognosis

Single-study CNV frequencies correspond to diagnostic subsets





Michael Baudis | Presentation at DGHO annual Meeting | Graz 2000-10-21

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. Hoechstetter[1] · Raymonde Busch[2] · Barbara Eichhorst[3] · Andreas Bühler[4] · Dirk Winkler[4] · Jasmin Bahlo[3] · Sandra Robrecht[3] · Michael J. Eckart[5] · Ursula Vehling-Kaiser[6] · Georg Jacobs[7] · Ulrich Jäger[8] · Hans Jürgen Hurtz[9] · Georg Hopfinger[10] · Frank Hartmann[11] · Harald Fuss[12] · Wolfgang Abenhardt[13] · Ilona Blau[14] · Werner Freier[15] · Lothar Müller[16] · Maria Goebeler[17] · Clemens Wendtner[1,3] · Kirsten Fischer[3] · Carmen D. Herling[3] · Michael Starck[1] · Martin Bentz[18] · Bertold Emmerich[19] · Hartmut Döhner[20] · Stephan Stilgenbauer[20] · Michael Hallek[3]

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		*P* 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.**	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)	*P*	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

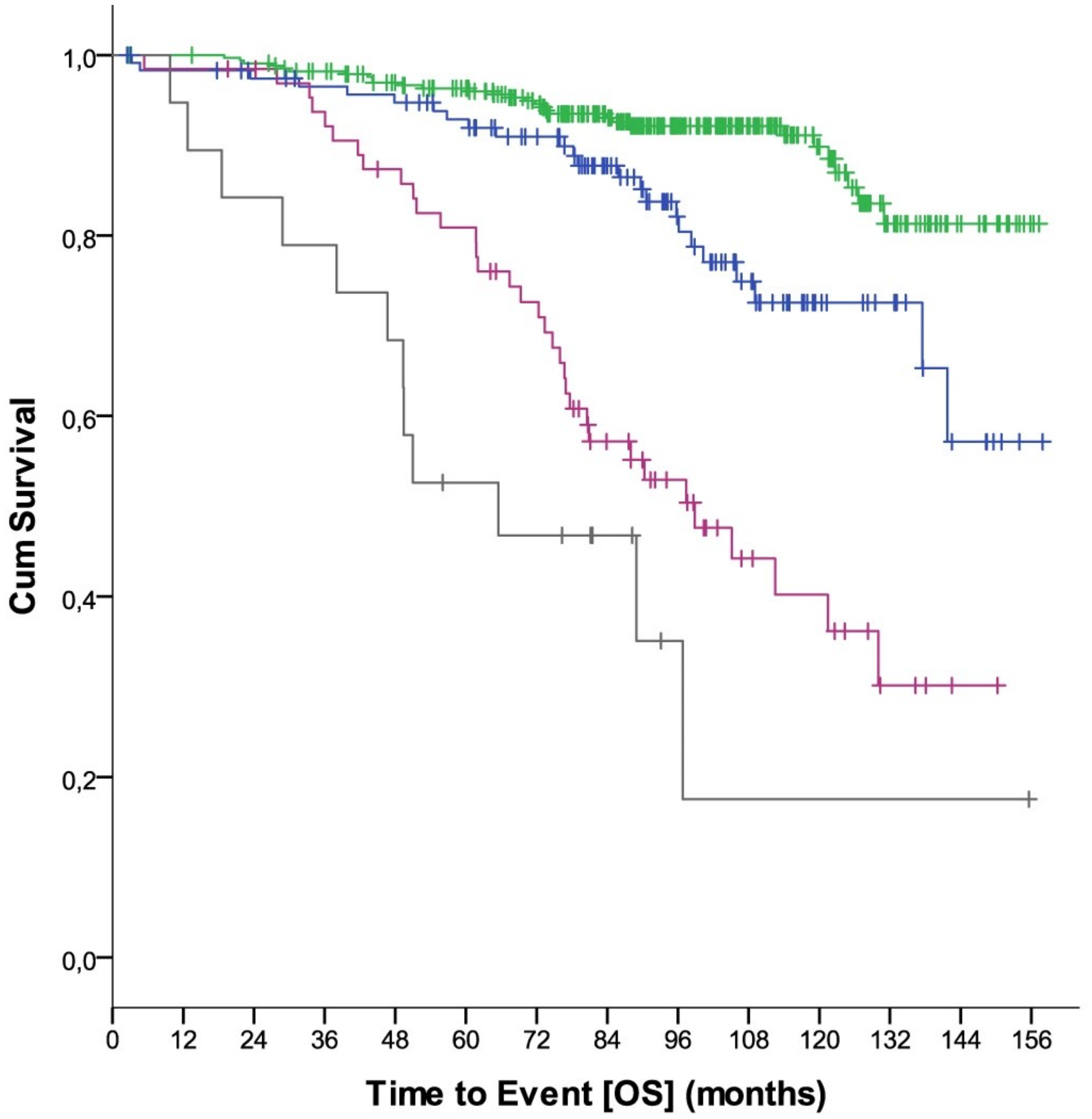
The assigned risk score points derived from the HR for OS of the individual factors.

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 accroding to the CLL1-PM		539
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		539
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.

2018-12-28___Yosifov,-Stilgenbauer-et-al.__The-CLL-Success-Story__HemaSphere-Review.pdf

Cancer Classifications & Parameters

NCIt | ICD-O / WHO | TNM

International Agency for Research on Cancer
World Health Organization

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**

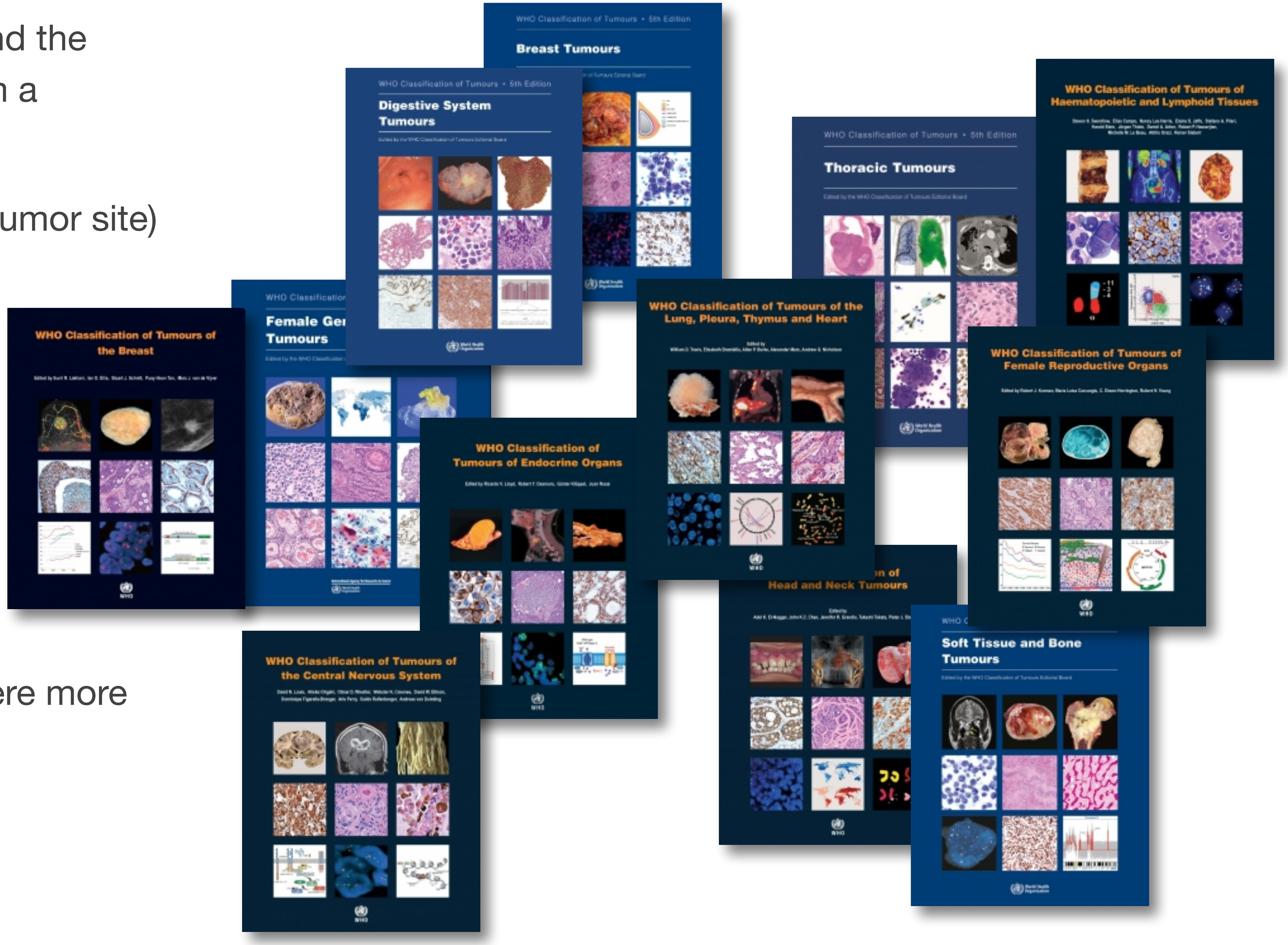
▸ "Adenocarcinoma" of the "Sigmoid colon"

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


WHO Classification of Tumours • 5th Edition
Breast Tumours
Digestive System Tumours
Thoracic Tumours
WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues
WHO Classification of Tumours of the Breast
Female Genital Tumours
WHO Classification of Tumours of the Lung, Pleura, Thymus and Heart
WHO Classification of Tumours of Female Reproductive Organs
WHO Classification of Tumours of Endocrine Organs
Head and Neck Tumours
Soft Tissue and Bone Tumours
WHO Classification of Tumours of the Central Nervous System

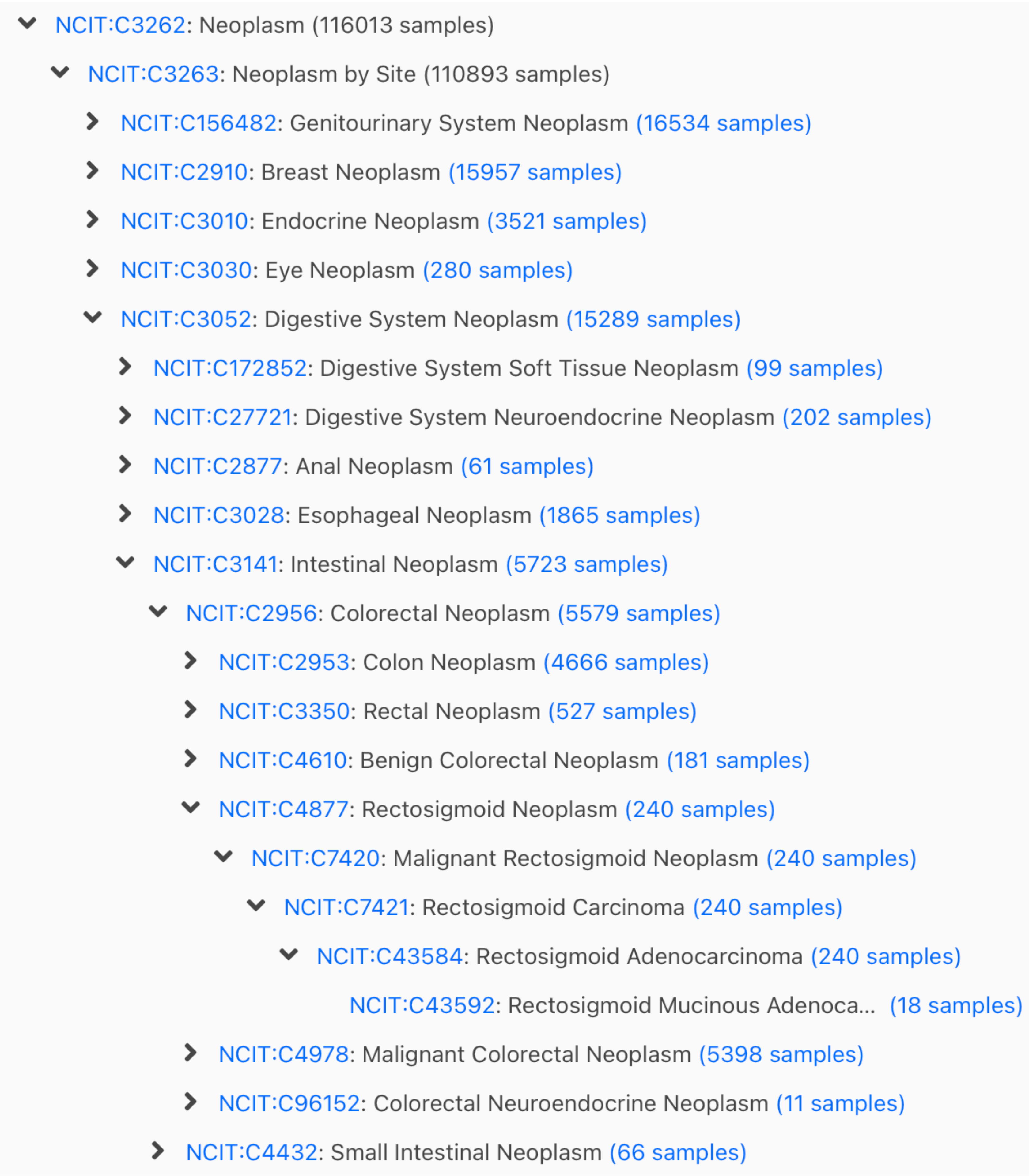
NCIt

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)



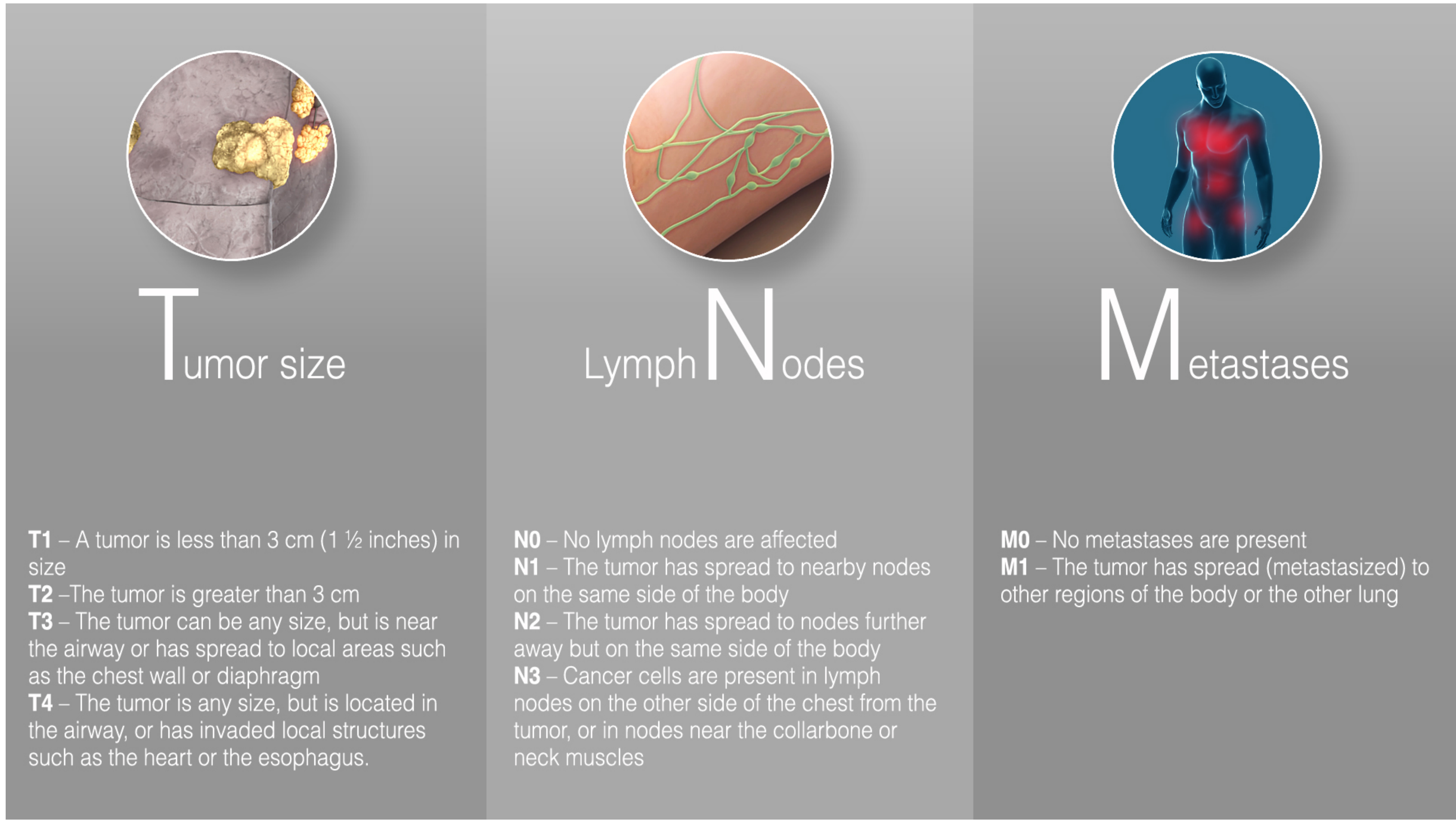
- ⌄ NCIT:C3262: Neoplasm (116013 samples)
  - ⌄ NCIT:C3263: Neoplasm by Site (110893 samples)
    - › NCIT:C156482: Genitourinary System Neoplasm (16534 samples)
    - › NCIT:C2910: Breast Neoplasm (15957 samples)
    - › NCIT:C3010: Endocrine Neoplasm (3521 samples)
    - › NCIT:C3030: Eye Neoplasm (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 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: https://www.cancer.gov/about-cancer/diagnosis-staging/staging



Tumor size

T1 – A tumor is less than 3 cm (1 ½ inches) in size
T2 –The tumor is greater than 3 cm
T3 – The tumor can be any size, but is near the airway or has spread to local areas such as the chest wall or diaphragm
T4 – The tumor is any size, but is located in the airway, or has invaded local structures such as the heart or the esophagus.

Lymph Nodes

N0 – No lymph nodes are affected
N1 – The tumor has spread to nearby nodes on the same side of the body
N2 – The tumor has spread to nodes further away but on the same side of the body
N3 – Cancer cells are present in lymph nodes on the other side of the chest from the tumor, or in nodes near the collarbone or neck muscles

Metastases

M0 – No metastases are present
M1 – The tumor has spread (metastasized) to other regions of the body or the other lung

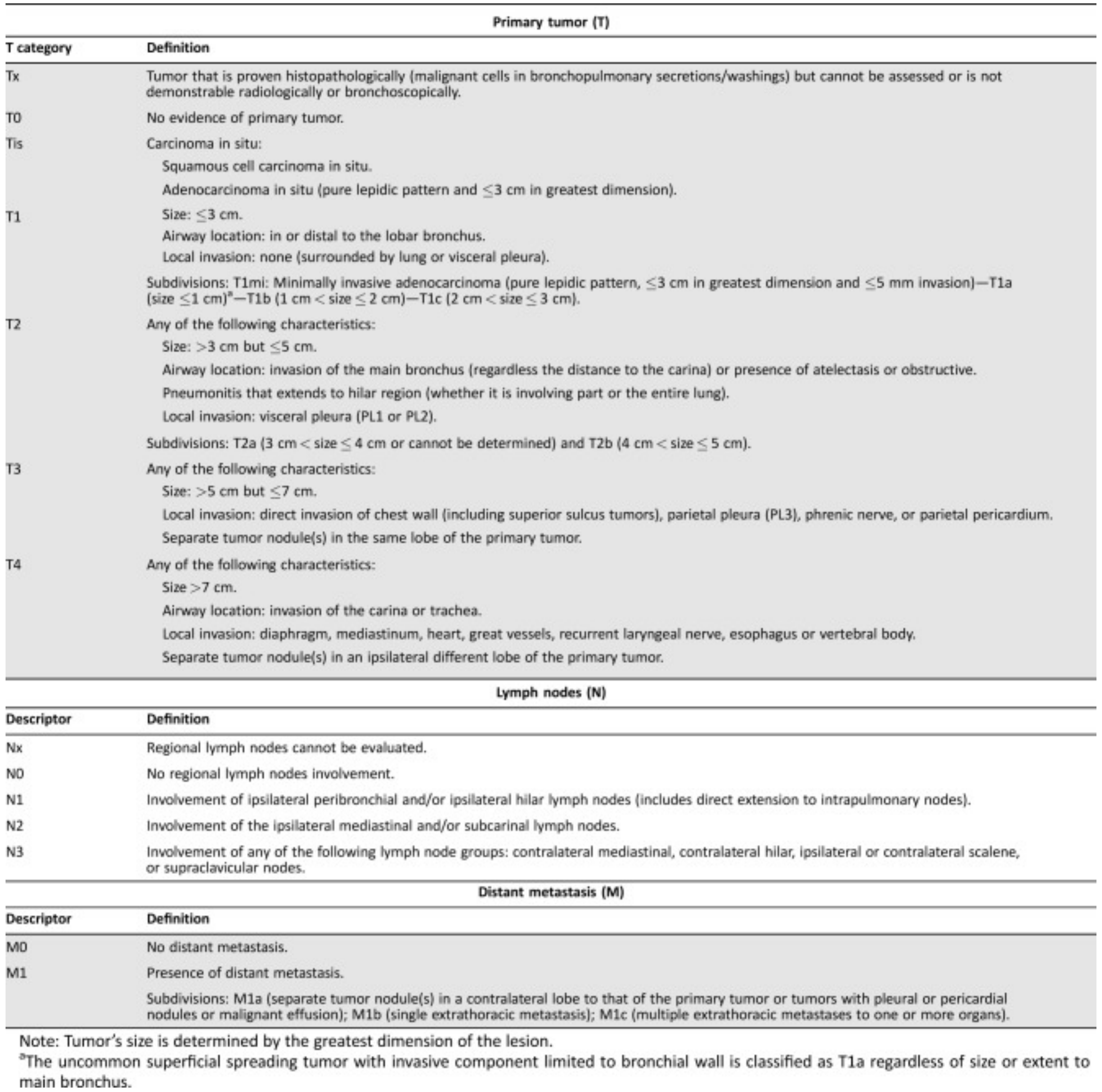
Source: www.scientificanimations.com

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 breast cancer is determined by size & nodal involvement

Source: https://www.cancer.gov/about-cancer/diagnosis-staging/staging



Primary tumor (T)

T category	Definition
Tx	Tumor that is proven histopathologically (malignant cells in bronchopulmonary secretions/washings) but cannot be assessed or is not demonstrable radiologically or bronchoscopically.
T0	No evidence of primary tumor.
Tis	Carcinoma in situ: Squamous cell carcinoma in situ. Adenocarcinoma in situ (pure lepidic pattern and ≤3 cm in greatest dimension).
T1	Size: ≤3 cm. Airway location: in or distal to the lobar bronchus. Local invasion: none (surrounded by lung or visceral pleura). Subdivisions: T1mi: Minimally invasive adenocarcinoma (pure lepidic pattern, ≤3 cm in greatest dimension and ≤5 mm invasion)—T1a (size ≤1 cm)[a]—T1b (1 cm < size ≤ 2 cm)—T1c (2 cm < size ≤ 3 cm).
T2	Any of the following characteristics: Size: >3 cm but ≤5 cm. Airway location: invasion of the main bronchus (regardless the distance to the carina) or presence of atelectasis or obstructive. Pneumonitis that extends to hilar region (whether it is involving part or the entire lung). Local invasion: visceral pleura (PL1 or PL2). Subdivisions: T2a (3 cm < size ≤ 4 cm or cannot be determined) and T2b (4 cm < size ≤ 5 cm).
T3	Any of the following characteristics: Size: >5 cm but ≤7 cm. Local invasion: direct invasion of chest wall (including superior sulcus tumors), parietal pleura (PL3), phrenic nerve, or parietal pericardium. Separate tumor nodule(s) in the same lobe of the primary tumor.
T4	Any of the following characteristics: Size >7 cm. Airway location: invasion of the carina or trachea. Local invasion: diaphragm, mediastinum, heart, great vessels, recurrent laryngeal nerve, esophagus or vertebral body. Separate tumor nodule(s) in an ipsilateral different lobe of the primary tumor.

Lymph nodes (N)

Descriptor	Definition
Nx	Regional lymph nodes cannot be evaluated.
N0	No regional lymph nodes involvement.
N1	Involvement of ipsilateral peribronchial and/or ipsilateral hilar lymph nodes (includes direct extension to intrapulmonary nodes).
N2	Involvement of the ipsilateral mediastinal and/or subcarinal lymph nodes.
N3	Involvement of any of the following lymph node groups: contralateral mediastinal, contralateral hilar, ipsilateral or contralateral scalene, or supraclavicular nodes.

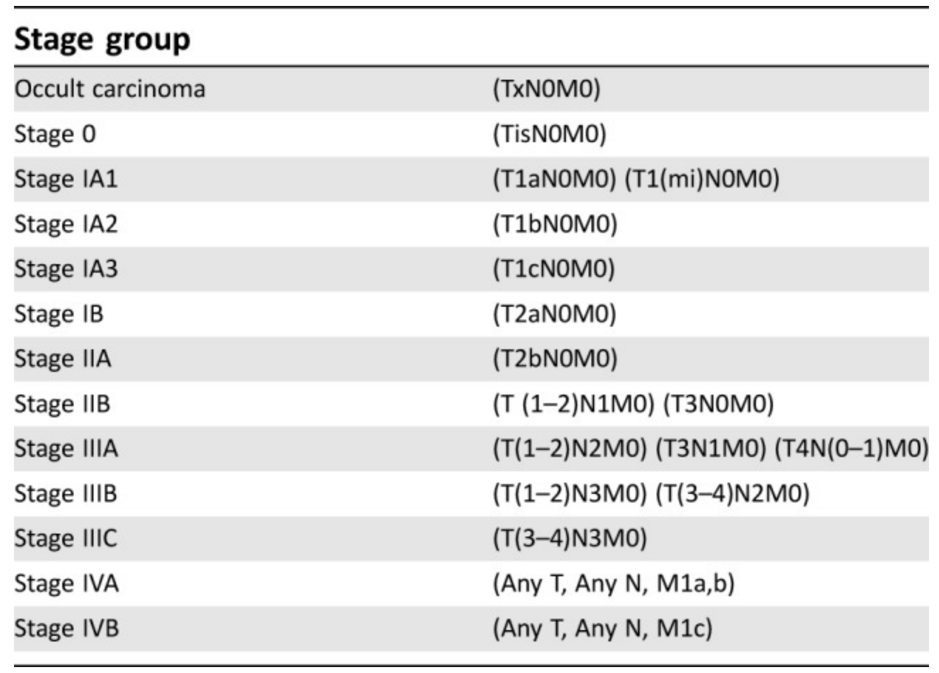
Distant metastasis (M)

Descriptor	Definition
M0	No distant metastasis.
M1	Presence of distant metastasis. Subdivisions: M1a (separate tumor nodule(s) in a contralateral lobe to that of the primary tumor or tumors with pleural or pericardial nodules or malignant effusion); M1b (single extrathoracic metastasis); M1c (multiple extrathoracic metastases to one or more organs).

Note: Tumor's size is determined by the greatest dimension of the lesion.

[a] The uncommon superficial spreading tumor with invasive component limited to bronchial wall is classified as T1a regardless of size or extent to main bronchus.

Lababede O, Meziane MA. The Eighth Edition of TNM Staging of Lung Cancer: Reference Chart and Diagrams. Oncologist. 2018;23(7):844-848. doi:10.1634/theoncologist.2017-0659



Stage group	
Occult carcinoma	(TxN0M0)
Stage 0	(TisN0M0)
Stage IA1	(T1aN0M0) (T1(mi)N0M0)
Stage IA2	(T1bN0M0)
Stage IA3	(T1cN0M0)
Stage IB	(T2aN0M0)
Stage IIA	(T2bN0M0)
Stage IIB	(T (1–2)N1M0) (T3N0M0)
Stage IIIA	(T(1–2)N2M0) (T3N1M0) (T4N(0–1)M0)
Stage IIIB	(T(1–2)N3M0) (T(3–4)N2M0)
Stage IIIC	(T(3–4)N3M0)
Stage IVA	(Any T, Any N, M1a,b)
Stage IVB	(Any T, Any N, M1c)

Tasks

Survival analyses | Cancer classifications | Staging

- Familiarize yourself with the different concepts behind different disease clasification systems - what are there use, advantages, problems? E.g. ICD-10, ICD-O, NCIt
  - 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