**Profil pacjenta do analizy przeżycia**

**1. Przegląd struktury danych**

**Plik:** clinical.tsv  
**Wielkość:** 5 546 wierszy × 210 kolumn  
**Zakres nowotworu:** TCGA‐BRCA (rak piersi)

|  |  |  |  |
| --- | --- | --- | --- |
| Blok kolumn (prefiks) | Liczba pól | Przykłady pól | Komentarz |
| project. | 1 | project\_id | metadane projektu; stała wartość TCGA-BRCA. |
| cases. | 9 | case\_id, days\_to\_lost\_to\_followup | unikatowy ID + info o cenzorowaniu. |
| demographic. | 22 | age\_at\_index, gender, race, ethnicity, vital\_status, days\_to\_death | podstawowe zmienne prognostyczne i zmienna zdarzenia. |
| diagnoses. | 132 | age\_at\_diagnosis, ajcc\_pathologic\_stage, tumor\_grade, primary\_diagnosis, TNM | najbogatszy blok – opis choroby w momencie rozpoznania. |
| treatments. | 46 | treatment\_type, treatment\_or\_therapy, treatment\_outcome, days\_to\_treatment\_start/end | informacje o leczeniu; część może być modelowana jako współzmienne czasowe. |

**2. Kluczowe kategorie zmiennych dla „profilu pacjenta”**

**2.1 Czas i zdarzenie**

* demographic.vital\_status (Alive/Dead)
* demographic.days\_to\_death
* diagnoses.days\_to\_last\_follow\_up

**2.2 Demografia**

* diagnoses.age\_at\_diagnosis *(preferowane)* lub demographic.age\_at\_index
* demographic.race, demographic.ethnicity
* demographic.gender *(jednorodne w BRCA, ale przydatne w analizach wielokohortowych)*

**2.3 Stadium i cechy guza**

* diagnoses.ajcc\_pathologic\_stage
* TNM: diagnoses.ajcc\_pathologic\_t, \_n, \_m
* diagnoses.tumor\_grade
* diagnoses.primary\_diagnosis, diagnoses.morphology
* Dodatkowe: diagnoses.tumor\_size, diagnoses.lymphatic\_invasion\_present, diagnoses.metastasis\_at\_diagnosis

**2.4 Podtyp molekularny**

Dołączenie ramek z pakietu **TCGAbiolinks**:

library(TCGAbiolinks)

sub <- TCGAquery\_subtype("BRCA") # kolumna BRCA\_Subtype\_PAM50

Łączenie po pierwszych 12 znakach cases.case\_id ↔️ sub$patient.

**2.5 Leczenie (baseline)**

* treatments.treatment\_or\_therapy *(yes/no)*
* treatments.treatment\_type *(kategoryczna: Surgery, Radiation, Hormone, Chemo…)*
* treatments.regimen\_or\_line\_of\_therapy

**2.6 Stan ogólny i historia choroby**

* diagnoses.ecog\_performance\_status, diagnoses.karnofsky\_performance\_status
* diagnoses.prior\_malignancy, diagnoses.prior\_treatment

**2.7  Kolumny szczególnie przydatne w analizie przeżycia**

|  |  |  |
| --- | --- | --- |
| Kategoria kliniczna | Kluczowe pola (TSV) | Uzasadnienie |
| Czas & zdarzenie | vital\_status, days\_to\_death, days\_to\_last\_follow\_up, days\_to\_last\_known\_disease\_status | Definiują **czas obserwacji** i **event** (zgon, cenzorowanie). |
| Wiek | age\_at\_diagnosis / age\_at\_index | Najsilniejszy pojedynczy czynnik prognostyczny. |
| Płeć | gender | Potencjalny czynnik ryzyka w analizach wielokohortowych. |
| Rasa / etniczność | race, ethnicity | Korekta dla czynników społeczno‑genetycznych. |
| Stadium choroby | ajcc\_pathologic\_stage, TNM (ajcc\_pathologic\_t, \_n, \_m) | Zaawansowanie choroby w chwili rozpoznania. |
| Stopień histologiczny | tumor\_grade, primary\_gleason\_grade, secondary\_gleason\_grade | Agresywność guza. |
| Podtyp molekularny | BRCA\_Subtype\_PAM50, IC10 (z pakietu *TCGAbiolinks*) | Różnice biologiczne wpływające na rokowanie. |
| Cechy inwazji / przerzutów | lymphatic\_invasion\_present, perineural\_invasion\_present, metastasis\_at\_diagnosis | Wskaźniki agresywności i rozsiewu. |
| Stan ogólny | ecog\_performance\_status, karnofsky\_performance\_status | Silny predyktor przeżycia niezależny od samej choroby. |
| Terapie bazowe | treatment\_or\_therapy, treatment\_type, regimen\_or\_line\_of\_therapy | Pozwala modelować wpływ leczenia (tak/nie lub kategoriami). |
| Historia onkologiczna | prior\_malignancy, prior\_treatment | Selekcja pacjentów naïve vs wcześniej leczonych. |

**3. Minimalny zestaw do modelu Cox (OS)**

|  |  |  |
| --- | --- | --- |
| Zmienna | Źródło | Typ |
| Czas obserwacji | ifelse(is.na(days\_to\_death), days\_to\_last\_follow\_up, days\_to\_death) | liczba dni |
| Status zdarzenia | vital\_status == "Dead" | binarna |
| Wiek | age\_at\_diagnosis | liczbowy |
| Stadium | ajcc\_pathologic\_stage | kategoryczny |
| Grade | tumor\_grade | kategoryczny |
| Podtyp PAM50 | z TCGAquery\_subtype | kategoryczny |

