ANALÝZA HISTORICKÝCH DAT NÁRODNÍHO ONKOLOGICKÉHO REGISTRU ČR: INFORMAČNÍ POTENCIÁL A RIZIKA ZKRESLENÍ

HISTORICAL DATA OF CZECH NATIONAL CANCER REGISTRY: INFORMATION VALUE AND RISK OF BIAS

MUŽÍK J.[1], KOPTÍKOVÁ J.[1], DUŠEK L.[1], ŽALOUDÍK J.[2], VYZULA R.[2], ABRAHÁMOVÁ J.[3]

[1] LÉKAŘSKÁ FAKULTA MASARYKOVY UNIVERZITY, BRNO
[2] MASARYKŮV ONKOLOGICKÝ ÚSTAV, BRNO
[3] FAKULTNÍ THOMAYEROVA NEMOCNICE, PRAHA

Souhrn
Článek se zabývá otázkou kvality dlouhodobě sbíraných epidemiologických dat v populačních registrech. Jako model je využita databáze Národního onkologického registru ČR (NOR), která nabízí kompletní epidemiologická data všech diagnostických skupin zhoubných nádorů od roku 1977. Analýza kvality dat NOR je nutným předpokladem jejich využití v hodnocení výsledků léčebné péče a je také základem odpovědné prezentace českých dat v zahraničí. Výsledkem auditu dat NOR je soubor validních záznamů s řádně zaznamenanou diagnostikou a soubor neúplných záznamů s identifikovanou objektivní příčinou problémů (DCO, diagnóza při pitvě, velmi časná úmrtí, apod.). Nad rámec těchto dvou kategorií bylo zachyceno 5,1 % záznamů s chybně neuvedeným klinickým stadiem i TNM klasifikací. Chybné záznamy typologicky odpovídají celkové databázi NOR a lze je tedy v závažných analýzách vypustit, aniž bychom se dopustili systémového zkreslení výstupů. Možnost zkreslení existuje pouze při některých regionálních srovnáních.

Klíčová slova: epidemiologie zhoubných nádorů, populační registry, kvalita dat.

Summary
The paper is focused on quality assessment of cancer population-based registries. The analyses are methodically based on the database of Czech National Cancer Registry (NCR) that offers valuable model with epidemiological data of all diagnostic groups of cancer, all collected continuously since 1977. Quality control is necessary for applications of NCR in health care evaluation and for international presentation of Czech cancer epidemiology. The audit resulted in a set of diagnostically completed and verified records and in a set of records with incomplete diagnostics (namely TNM) due to well distinguished objective reasons (DCO cases, cases diagnosed at autopsy, early death after diagnosis, etc.). All the objective reasons were recognized and the database was enriched by codes that identify the status of the record and its reason. In addition to it, we found 5.1 % of cases with wrong and/or incomplete diagnostics and without any acceptable explanation. These cases typologically correspond to the whole NCR database and so they can be excluded from serious analyses with a minimum risk of bias. Increased probability of bias due to erroneously incomplete records exists only in regionally specific comparisons.

Key words: cancer epidemiology, population registries, quality of data.

Problémy s retrospektivní analýzou dat populačních onkologických registrů

Plošné populační registry mají smysl pouze tehdy, jsou-li skutečně plošně reprezentativní, dlouhodobě standardně vedené a pravidelně aktualizované. Při dobrém organizačním zázemí, adekvátním rozpočtu a únosném rozsahu sbíraných dat lze toto vše zajistit i pro zhoubné nádory, i když v univerzální onkologické databázi nezůstane prostor pro specifické záznamy jednotlivých diagnostických skupin. Daleko větším problémem je vývoj onkologie samotné, změna klasifikačních systémů, rozvoj diagnostických a terapeutických metod – těmto změnám samozřejmě nelze bránit, ale rovněž na ně nelze běžící registr „za pochodu" adaptovat a neustále jej tak měnit. Plošně sbíraná data tak kumulují množství vlivů, které mohou vážně zkreslovat výstupy, zvláště dlouhodobé trendy. Vzhledem k rozmanitosti takto vznikajících problémů neexistuje ani žádné přímočaré a jednoduché řešení, jak ostatně ukazuje i následující výčet:

- Problémy s úplností registru a jeho záznamů

■ U plošných epidemiologických registrů je úplnost záznamů zcela zásadní požadavek podmiňující smysl registrace. Nutno dodat, že u onkologických registrů požadavek nelehký vzhledem k diagnostické rozmanitosti a k velkému počtu hlásících míst. Úplností zde rozumíme jednak reprezentativní počet záznamů, jednak vnitřní úplnost jednotlivých záznamů, kde nesmí chybět zásadní datové komponenty, jako je diagnostická identifikace nádoru, pokročilost onemocnění a aktualizovaný údaj o stavu pacienta. I renomované registry se potýkají se zpožděním v hlášení incidence a s dohledáváním nálezů diagnostikovaných až při pitvě nebo při prohlídce mrtvého. Z tohoto důvodu je velmi cenné mít srovnávací nezávislou databázi zemřelých, proti které se mohou epidemiologická data zpětně kontrolovat.

- Problémy s věrohodností registrovaných dat

■ Věrohodnost úzce souvisí s úplností, neúplné záznamy

jsou samozřejmě i nevěrohodné, a to samé platí i o analytických výstupech na nich postavených. Řešením je nastavení registru tak, aby obsahoval pouze nutné minimum parametrů, které lze věrohodně sbírat v populačním měřítku. Jakékoli detailní parametry by měly být předmětem specializovaných klinických projektů, u kterých se mnohem lépe zajišťuje kontrola dat.

- Změny regionálního uspořádání v oblasti pokryté registrem

■ Velmi nepříjemný problém komplikující zpětné transformace dat. Všechny záznamy by měly být jednoznačně přiřazeny k regionu a místu a tyto kódy by měly umožňovat jasný převod při změnách.

- Změny počtu a typu hlásících míst

■ Hlásícími místy zde myslíme především zdravotnická zařízení. Jednoznačné přiřazení diagnostikovaného a léčeného pacienta k danému zdravotnickému zařízení je velmi podstatnou podmínkou pro budoucí využitelnost záznamů. Každý záznam tak může být přiřazen určitému typu zdravotnického zařízení. Tímto přiřazením je mimo jiné vyjádřena odpovědnost daného zařízení za pacienta a za záznam. V případě, že péče o pacienta přechází v čase na více zdravotnických zařízení, mělo by ze záznamu být patrné, které pracoviště bylo odpovědné za diagnostiku a primární léčbu a které za dispenzární péči, popřípadě léčbu relapsů a progresí.

- Změny terapeutických postupů

■ Objektivní faktor, proti kterému se nelze bránit, ale který nicméně velmi komplikuje retrospektivní hodnocení záznamů v onkologických registrech. Zlepšování terapeutických možností jistě ovlivňuje vývoj mortality, ale tato skutečnost nemůže být důvodem pro zavádění podrobných klinických záznamů do epidemiologických registrů. V plošných registrech jsou únosné obecné záznamy o aplikovaných léčebných modalitách, jakékoli další sledování patří do klinických specializovaných databází.

- Změny diagnostických technik a diagnostického přístupu

■ Vývoj diagnostiky mění především relativní podíly klinických stadií zachycených nádorových onemocnění. U kvalitního plošného registru by vývoj diagnostiky neměl dlouhodobě zvyšovat hlášení o incidenci, mění ale vnitřní strukturu záznamů. Proto je tak významná diagnostická identifikace záznamů, neboť znalost o klinickém stadiu je podmínkou správné interpretace trendů.

- Změny klasifikačního systému nádorů

■ Zdokonalování systému klasifikace nádorů představuje problém u kategorií, kde se mění způsob určení klinického stadia z TNM komponent. Jsou-li v registru vedeny parametry TNM i klinické stadium a jsou-li tyto záznamy úplné, nepředstavuje změna v klasifikaci problém při zpětném hodnocení dat, neboť data lze zpětně transformovat.

Jak je patrné, většině problémů je možné předcházet již při plánování plošného registru (především nastavením rozumného rozsahu sbíraných dat), na některé problémy se musí registr adaptovat za chodu, zatímco některé problémy se vyřešit objektivně nedají a výstupy z registru s nimi musí analyticky počítat. Obecně lze doporučit následující strategická opatření, která minimalizují budoucí problémy:

A. Orientovat plošný registr především na epidemiologická data, bez sběru specifických diagnostických a terapeutických záznamů, které jsou poplatné vývoji v rámci jednotlivých diagnóz.
B. Kvalitně a robustně kódovat místní příslušnost každého nového záznamu a rovněž záznam jednoznačně přiřazovat k zdravotnickému zařízení, které je odpovědné za léčbu a později i k zařízení odpovědnému za dispenzarizaci pacienta. Tato opatření připravují registr na zpětné převody při změně regionálního uspořádání a usnadní i flexibilní třídění podle typu zdravotnického zařízení.
C. Dbát na korektnost a úplnosti diagnostických záznamů. U většiny solidních nádorů jde především o identifikaci diagnózy a lokalizace primárního tumoru, dále o určení TNM a klinického stadia.

Především poslední z výše uvedených opatření je třeba považovat za zásadní pravidlo určující kvalitu celého registru. Ne u všech záznamů mohou být ovšem diagnostická data kompletně k dispozici (např. je-li diagnóza provedena při pitvě), ale i bez ohledu na tuto skutečnost je diagnostická identifikace předpokladem využitelnosti dat plošných registrů. Jistě není třeba vysvětlovat, že neznáme-li diagnózu, nemůžeme pracovat ani s tak základními údaji jako je incidence a mortalita, a celý registr ztrácí smysl. U zhoubných nádorů je nadto i TNM klasifikace a klinické stadium na stejné úrovni jako diagnóza, neboť u různě pokročilých onemocnění se liší terapeutické možnosti a postupy, a v konečném důsledku i dosažitelné přežití pacientů.
V tomto příspěvku se dále zabýváme hodnocením kvality dlouhodobě sbíraných epidemiologických dat. Pro metodické rozbory a úvahy máme v ČR ideální model v databázi Národního onkologického registru (NOR), která nabízí rozsáhlá epidemiologická data všech diagnostických skupin zhoubných nádorů od roku 1977. V dalším rozboru se tedy zaměříme i na kvalitu epidemiologických dat NOR, jejich konzistenci a možná rizika zkreslení. Tato analýza je nutná nejen z pohledu využitelnosti NOR pro hodnocení výsledků léčebné péče, ale je také základem odpovědné prezentace českých epidemiologických dat v zahraničí.

Diagnostická identifikace zhoubných nádorů jako základ kvality plošných registrů

Chybějící nebo nevěrohodný záznam o diagnóze diskvalifikuje všechna další pořízená data a znemožňuje exaktní zařazení daného záznamu do epidemiologických analýz. Čtenář se nyní jistě diví, proč zdůrazňujeme tak samozřejmou věc, jako je uvedení diagnózy do epidemiologického registru. Řečeno odborně, u onkologických registrů je identifikace diagnózy podle MKN-O systému, kódem Cxx, podmínkou nutnou, nikoli však postačující. Skutečně věrohodný a využitelný záznam musí totiž zahrnovat i údaje o pokročilosti onemocnění, což je u většiny tzv. solidních nádorů TNM systém a z něj vyplývající klinické stadium, u některých dalších diagnóz (např. hematoonkologických) potom i jiné ukazatele pokročilosti (rizikový typ onemocnění, grade). Tyto záznamy jednak potvrzují samotnou diagnózu, jednak určují pokročilost nemoci v době diagnózy. Klinické stadium a hodnoty TNM mají jednoznačný vliv na variabilitu výsledného přežití a vysvětlují i řadu regionálních rozdílů a časových trendů (Sant a kol., 2003). I u vel-

Tabulka 1: Přehled diagnóz zhoubných novotvarů

C00	ZN rtu
C01	ZN kořene jazyka
C02	ZN jiných a neurčených částí jazyka
C03	ZN dásně - gingivy
C04	ZN ústní spodiny
C05	ZN patra
C06	ZN jiných a neurčených částí úst
C07	ZN příušní (parotické) žlázy
C08	ZN jiných a neurčených slinných žláz
C09	ZN mandle - tonzily
C10	ZN ústní části hltanu - orofaryngu
C11	ZN nosohltanu - nazofaryngu
C12	ZN pyriformního sinu
C13	ZN hypofaryngu
C14	ZN jiných a nepřesně určených lokalizací rtu, ústní dutiny a hltanu
C15	ZN jícnu
C16	ZN žaludku
C17	ZN tenkého střeva
C18	ZN tlustého střeva
C19	ZN rektosigmoideálního spojení
C20	ZN konečníku - recta
C21	ZN řiti a řitního kanálu
C22	ZN jater a intrahepatálních žlučových cest
C23	ZN žlučníku
C24	ZN jiných a neurčených částí žlučových cest
C25	ZN slinivky břišní
C26	ZN jiných a nepřesně určených trávicích orgánů
C30	ZN nosní dutiny a středního ucha
C31	ZN vedlejších dutin
C32	ZN hrtanu
C33	ZN průdušnice - trachey
C34	ZN průdušky - bronchu a plíce
C37	ZN brzlíku - thymu
C38	ZN srdce, mezihrudí - mediastina a pohrudnice - pleury
C39	ZN jiných a nepřesně určených lokalizací v dých. soustavě a nitrohrud.orgánech
C40	ZN kosti a kloubní chrupavky končetin
C41	ZN kosti a kloubní chrupavky jiných a neurčených lokalizací
C43	Zhoubný melanom kůže
C44	Jiný ZN kůže
C45	Mezoteliom - mesothelioma
C46	Kaposiho sarkom
C47	ZN periferních nervů a autonomní nervové soustavy
C48	ZN retroperitonea a peritonea
C49	ZN jiné pojivové a měkké tkáně
C50	ZN prsu
C51	ZN vulvy
C52	ZN pochvy - vaginy
C53	ZN hrdla děložního - cervicis uteri
C54	ZN těla děložního
C55	ZN dělohy, část NS
C56	ZN vaječníku
C57	ZN jiných a neurčených ženských pohlavních orgánů
C58	ZN placenty
C60	ZN pyje
C61	ZN předstojné žlázy - prostaty
C62	ZN varlete
C63	ZN jiných a neurčených mužských pohlavních orgánů
C64	ZN ledviny mimo pánvičku
C65	ZN ledvinné pánvičky
C66	ZN močovodu - ureteru
C67	ZN močového měchýře - vesicae urinariae
C68	ZN jiných a neurčených močových orgánů
C69	ZN oka a očních adnex
C70	ZN mozkomíšních plen
C71	ZN mozku
C72	ZN míchy, mozkových nervů a jiných částí centrální nervové soustavy
C73	ZN štítné žlázy
C74	ZN nadledviny
C75	ZN jiných žláz s vnitřní sekrecí a příbuzných struktur
C76	ZN jiných a nepřesně určených lokalizací
C77	Sekundární a neurčený ZN mízních uzlin
C78	Sekundární ZN dýchací a trávicí soustavy
C79	Sekundární ZN jiných lokalizací
C80	ZN bez určení lokalizace
C81	Hodgkinova nemoc
C82	Ne-Hodgkinův folikulární (nodulární) lymfom
C83	Ne-Hodgkinův (difúzní) lymfom
C84	Periferní a kožní T-buněčné lymfomy
C85	Ne-Hodgkinův lymfom, jiných a neurčených typů
C88	Zhoubné imunoproliferativní nemoci
C90	Mnohočetný myelom a plasmocytární novotvary
C91	Lymfoidní leukémie
C92	Myeloidní leukémie
C93	Monocytární leukémie
C94	Jiné leukémie určených buněčných typů
C95	Leukémie neurčeného buněčného typu
C96	Jiné zhoubné novotvary mízní, krvetvorné a příbuzné tkáně
C97	Zhoubné novotvary mnohočetných samostatných (primárních) lokalizací

mi zjednodušených registrů je nutné trvat na zaznamenávání všech tří komponent TNM, tento systém je nenahraditelný uvedením klinického stadia, neboť může mít sám o sobě větší informační hodnotu než integrující staging (Sant a kol., 2001). Bylo například prokázáno, že variabilita ve velikosti tumoru (T) v rámci kategorie N0/M0 může vysvětlovat výrazné regionální rozdíly v přežití žen s karcinomem prsu (Shetty a Reinman, 1997). Složky TNM systému vystupují často v populačních i klinických prognostických modelech jako nezávislé prediktory (Sant a kol., 2003). Dílčí komponenty TNM systému mají navíc významnou retrospektivní korekční úlohu, neboť podle změn v TNM klasifikaci můžeme zpětně korigovat i interpretace výsledků pro různá klinická stadia.

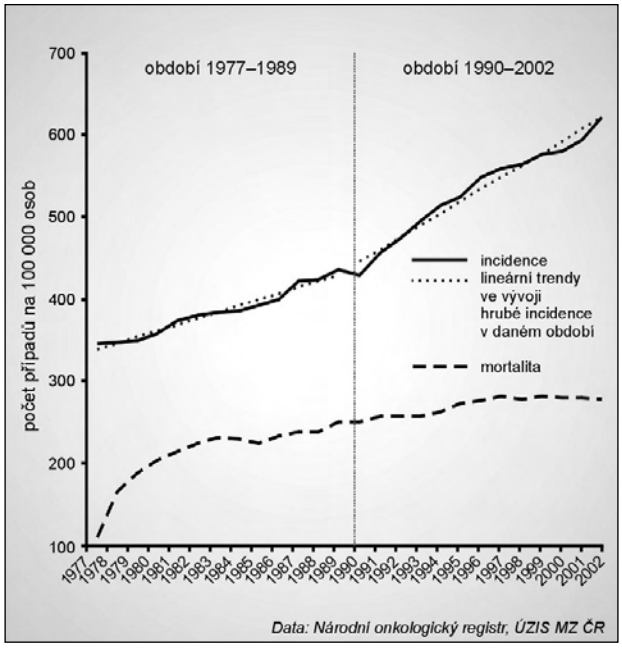
Tabulka 1 znázorňuje přehled onkologických diagnóz, z nichž většina je klasifikovatelná TNM systémem. Z přehledu je patrná rozmanitost systému, který se nadto metodicky vyvíjí v čase a pokrývá stále větší spektrum diagnóz a stadií (Tabulka 2). Při zpětné analýze dat dlouhodobého registru typu NOR si však musíme být těchto změn vědomi, neboť ovlivňují interpretaci trendů i u těch nejzákladnějších epidemiologických dat, jako je incidence a mortalita. Lze prohlásit, že bez znalosti klinického stadia nejsme v onkologii schopni interpretovat téměř žádnou změnu v epidemiologických parametrech. Jako téměř učebnicový příklad lze uvést často medializovaný nárůst hrubé incidence zhoubných nádorů v ČR (Obrázek 1), který ale při detailním rozboru jednotlivých diagnóz nemusí zdaleka odrážet jen rizikový vývoj. Například v případě dramaticky rostoucí incidence karcinomu prsu nám znalost klinického stadia a TNM umožní zjištění, že jde o trend dikto-

Tabulka 2: Stručný přehled vývoje TNM klasifikace zhoubných novotvarů

Verze TNM klasifikace	Platnost v NOR	Hlavní změny
II. vydání TNM, 1974 (1)	1976-1981	Provedeny změny a doplňky klasifikací 23 diagnóz vydaných v I. vydání TNM z roku 1968, doplnění o další diagnózy. II. vydání obsahuje klasifikace diagnóz C00*, C01-C06, C09*, C10*, C11*, C13*, C32, C15, C16*, C34, C50, C51*, C52*, C53, C54, C55*, C56, C57*, C64*, C67*, C60*, C61*, C62*, C73*, C44*
III. vydání TNM, 1982 (2)	1982-1994	Změny a doplňky klasifikací diagnóz vydaných ve II. vydání, zavedení klinických stadií pro diagnózy C00, C09-C13, C16, C51, C52. Nově zavedené klasifikace pro další diagnózy: C18-C20, C21*, C43, C47, C49.
IV. vydání TNM, 2. revize, 1992 (3)	1995-2000	Změny a doplňky klasifikací diagnóz vydaných ve III. vydání (především u C18-20), zavedení klinických stadií pro diagnózy C21, C60-C63, C64, C67, C44. Nově zavedené klasifikace pro další diagnózy: C07, C08, C17, C31, C22-C25, C40, C41, C45, C65, C66, C69, C71.
V. vydání TNM, 1997 (4)	2001-2004	Většina nádorových lokalizací zůstává beze změn nebo obsahuje menší úpravy s ohledem na nové údaje týkající se prognózy a nových metod ke stanovení prognózy (změny u C11, C58, C51, C53, C57, C61, C62, C64, C67, zrušena klasifikace u C71). Nově zavedené klasifikace pro další diagnózy: C58, C57.0
VI. vydání TNM, 2002 (5)	od 2005	Většina nádorových lokalizací zůstává beze změn nebo zaznamenává menší úpravy, které odráží nové informace týkající se prognózy onemocnění (změny u C22-C25, C38.4, C40, C41, C50, C69.3, C69.4)

* je zavedena jen TNM klasifikace bez dělení do klinických stadií

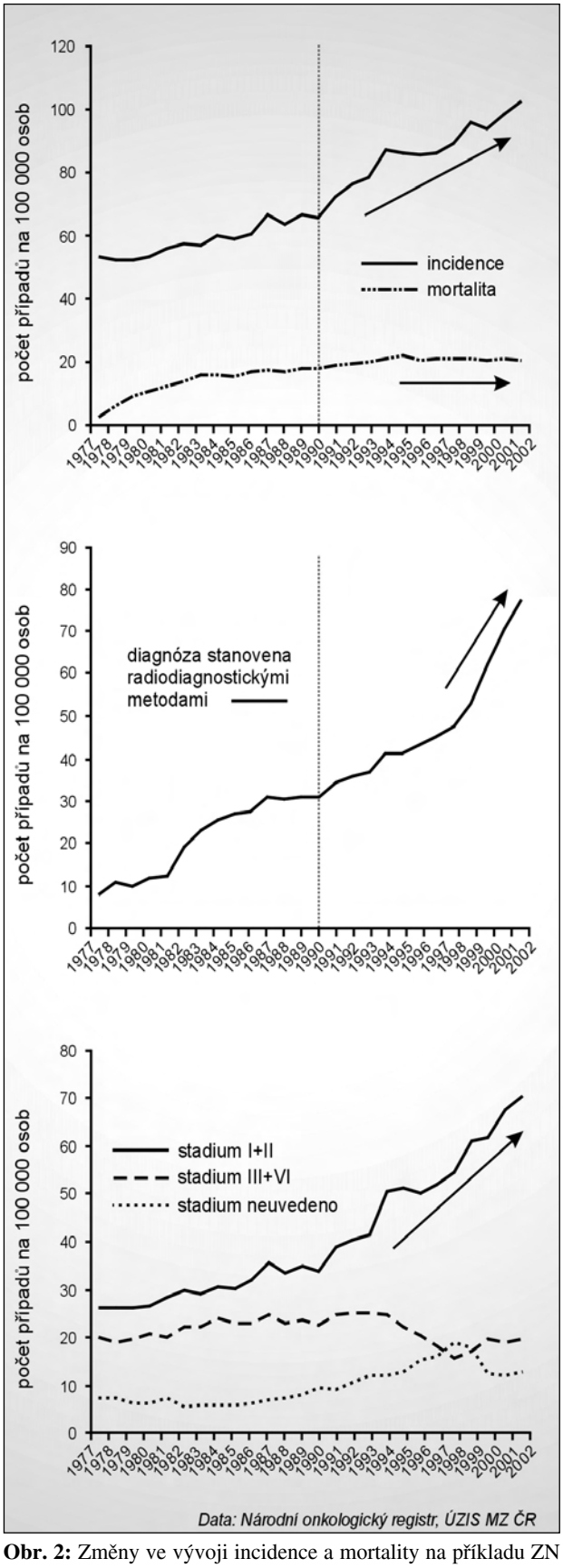
1. International Union Against Cancer (UICC): TNM Classification of Malignant Tumours. 2nd ed. Geneva, 1974.
2. International Union Against Cancer (UICC): TNM Classification of Malignant Tumours.. 3rd ed. M.H. Harmer, ed. Geneva, 1978. Enlarged and revised 1982.
3. International Union Against Cancer (UICC): TNM Classification of Malignant Tumours.. 4th ed. P. Hermanek, L.H. Sobin, eds. Berlin, Heidelberg, New York: Springer Verlag, 1987. Revised 1992.
4. International Union Against Cancer (UICC): TNM Classification of Malignant Tumours.. 5th ed. Sobin LH, Wittekind Ch (ed.), New York, Wiley-Liss. 1997
5. International Union Against Cancer (UICC): TNM Classification of Malignant Tumours.. 6th ed. Sobin LH, Wittekind Ch (ed.), New York, Wiley-Liss. 2002



Obr. 1.: Vývoj incidence a mortality všech nádorů (C00–C96, D03, D05, D06)

vaný rostoucím počtem nově diagnostikovaných časných klinických stadií a z větší části tedy o důsledek zlepšující se diagnostiky. Svou roli zde hraje nástup mamografie a v posledních letech již i projevující se vliv screeningu karcinomu prsu (dokumentující příklad v datech – viz Obrázek 2).

Metodický vývoj TNM klasifikace zhoubných nádorů je samozřejmě nepříjemnou komplikací pro analýzy dlouhodobých trendů. Z pohledu klinicky využitelných informací to ale tak zásadní není, neboť vzhledem k vývoji diagnostických technik a léčebných možností je u většiny diagnóz nutné zaměřit analýzy na záznamy pořízené v posledních 5–15 letech. To sice nijak nesnižuje složitost problému například pro hodnocení dlouhodobého přežití, pozitivum je nicméně v tom, že takto lze smysluplně pracovat pouze s daty zadávanými podle novějších a úplnějších pravidel TNM klasifikace (Tabulka 2). Nadto se postupem času u starších registrů zlepšuje i organizační zázemí a registry produkují data s rostoucí kvalitou. K významnému zkvalitňování dochází jistě i u Národního onkologického registru ČR, jak dokumentuje analýza mapující náhodnou složku variability incidenčních dat (Obrázek 3). Regionální i časové (meziroční) fluktuace jsou u záznamů NOR zadaných po roce 1990 významně nižší než u historických dat, a nově dosahují akceptovatelné úrov-



Obr. 2: Změny ve vývoji incidence a mortality na příkladu ZN prsu (C50)

ně do 5–10 % celkové variability. Zlepšení je patrné téměř u všech diagnostických skupin s výjimkou některých málo četných diagnóz, kde je vyšší variabilita vysvětlitelná relativně malým vzorkem a někdy i obtížnou diagnostikou v dané lokalizaci (např. ZN nádor jícnu nebo ZN nosohltanu, Obrázek 3).

Metodický rozbor kvality epidemiologických dat

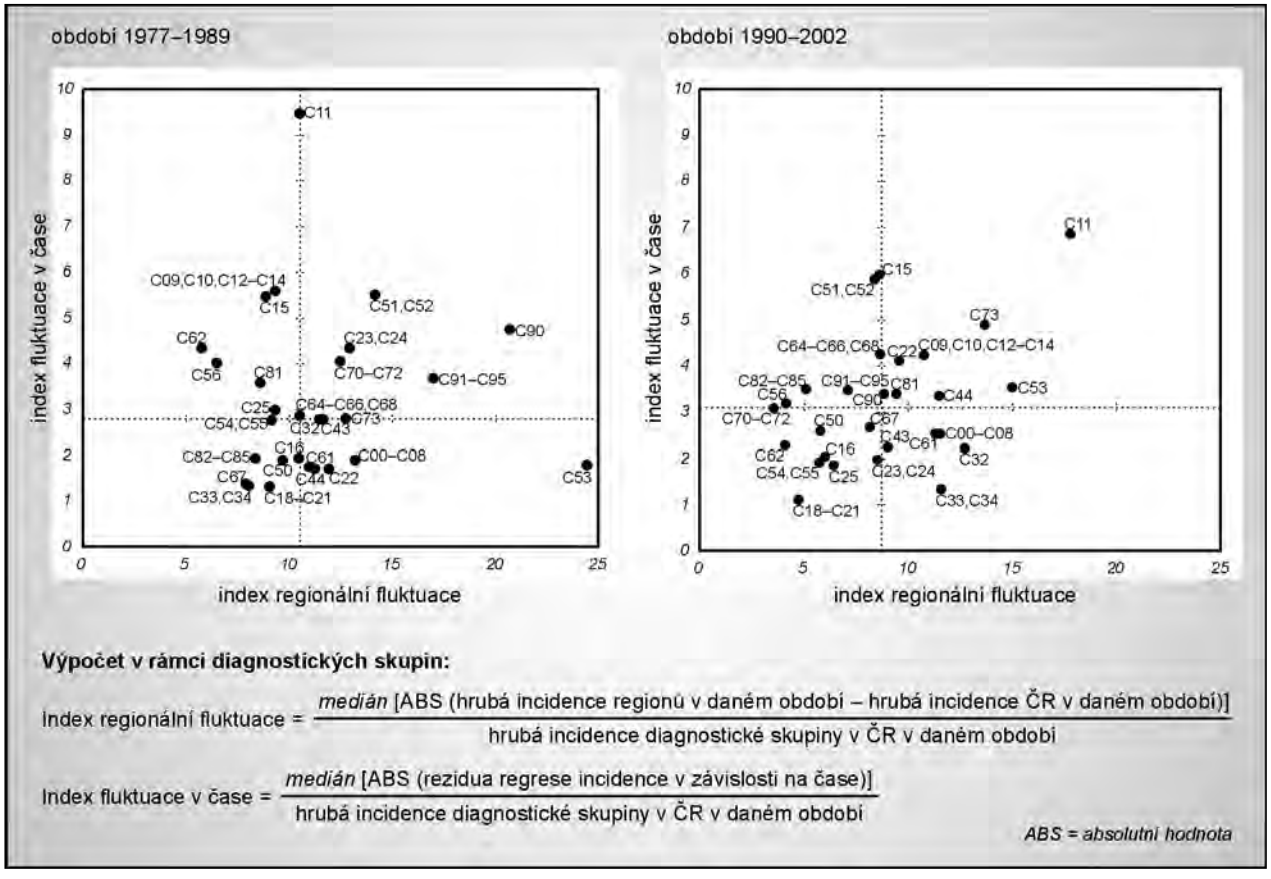
Nikdo samozřejmě nemůže očekávat, že analýza historických dat v rozsáhlém populačním registru bude probíhat bezproblémově. Již samotný vývoj klasifikačních systémů představuje prvek variability, který mnohdy nelze zpětně korigovat. Seriózní publikace dat populačních registrů přinášejí i informace o podílu různě problematických záznamů. Často bývá citována cca 90–95% věrohodnost anebo naopak 5–10% chyba v diagnostických záznamech populačních onkologických dat (např. Eaker a kol., 2006). Tedy i velmi renomované populační studie se potýkají s určitým procentem problematických záznamů. Například aktuální analýza agregovaných evropských registrů v projektu EUROCARE-3 vykázala na více než 6,5 milionech záznamů 1,1 % ztráty během období sledování (follow-up), dále 4,2 % záznamů typu DCO a 1,2 % ZN diagnostikovaných při pitvě (Capocaccia et al., 2003).

Za minimální atributy konzistence dat lze považovat logickou posloupnost věku pacienta a věku při diagnóze, a dále správné vyplnění diagnostických záznamů. V každém případě musí být neúplné záznamy nebo zcela chybné záznamy v rámci vstupního auditu vypuštěny nebo označeny jako nerelevantní pro další zpracování a musí být posouzen jejich vliv na reprezentativnost zbývajících záznamů. Důvodů ke snížení informační hodnoty jednotlivých záznamů přitom může být více a mohou mít různou váhu v různých situacích. Kontrola kvality dat registru by tedy měla vést k označení primárních záznamů kódy, které určí původ, úplnost a věrohodnost záznamu:

- sumární kód definující celkovou věrohodnost a využitelnost záznamu, například s následujícími položkami v číselníku:
  - o záznam úplný, plně využitelný
  - o záznam neúplný z objektivních důvodů, jinak konzistentní
  - o záznam se závažnými problémy a sníženou kvalitou
  - o záznam chybný a nevyužitelný
- kód označující chybějící hodnoty v primárním záznamu
- kód označující záznam s hodnotami vypuštěnými z důvodu nekonzistence
- kód označující hlavní příčiny neúplnosti záznamu (objektivní příčiny / chyba)

Provedení takového auditu musí být komplexní, neboť výsledek omezí dostupnost některých záznamů pro závažné analýzy. Audit tedy musí monitorovat všechny situace a příčiny, které vedou k neúplné diagnostické identifikaci záznamů. Jak vyplývá z následujícího přehledu, mnoho důvodů je objektivních a mnoho má i svou vlastní informační hodnotu již tím, že byly rozpoznány (viz např. Jemal a kol., 2004):

1) Vyčlenění diagnostických skupin, pro které TNM klasifikace nemá smysl a tudíž je u nich neúplnost záznamů v podstatě správná (v ČR je tato situace je historickým problémem univerzální hlášenky NOR, která takto vyřadila diagnostickou identifikaci např. leukémií).



Obr. 3: Hodnocení regionální a časové fluktuace incidence zhoubných nádorů v NOR

2) Vyčlenění záznamů těch diagnostických skupin, pro které systém TNM v době diagnózy nebyl zaveden a definován.
3) Neúplná diagnostika z důvodu diagnózy při prohlídce mrtvého (záznamy DCO) nebo z důvodu diagnózy při pitvě.
4) Neúplná diagnostika z důvodu velmi časného úmrtí pacienta po diagnóze, ještě před zahájením protinádorové terapie.
5) Neúplná diagnostika z důvodu neočekávaného vývoje situace, přestěhování pacienta, odmítnutí léčby, apod.
6) Skutečně chybně neúplné nebo jinak problematické záznamy, které nespadají do žádné z výše uvedených kategorií.

Je zřejmé, že neúplné záznamy existují a u historicky starších dat jejich podíl dosahuje poměrně vysokých hodnot. Absolutní vymazání již provedeného záznamu z databáze je ovšem extrémní zásah, ke kterému bychom měli přistoupit pouze v případě, kdy není ani jisté, zda šlo o zhoubný nádor. Výše navržený systém kódů dává správcům dat možnost záznamy třídit podle různého stupně neúplnosti, aniž by je museli přímo mazat. Otvírá se tím prostor pro následující korekční řešení:

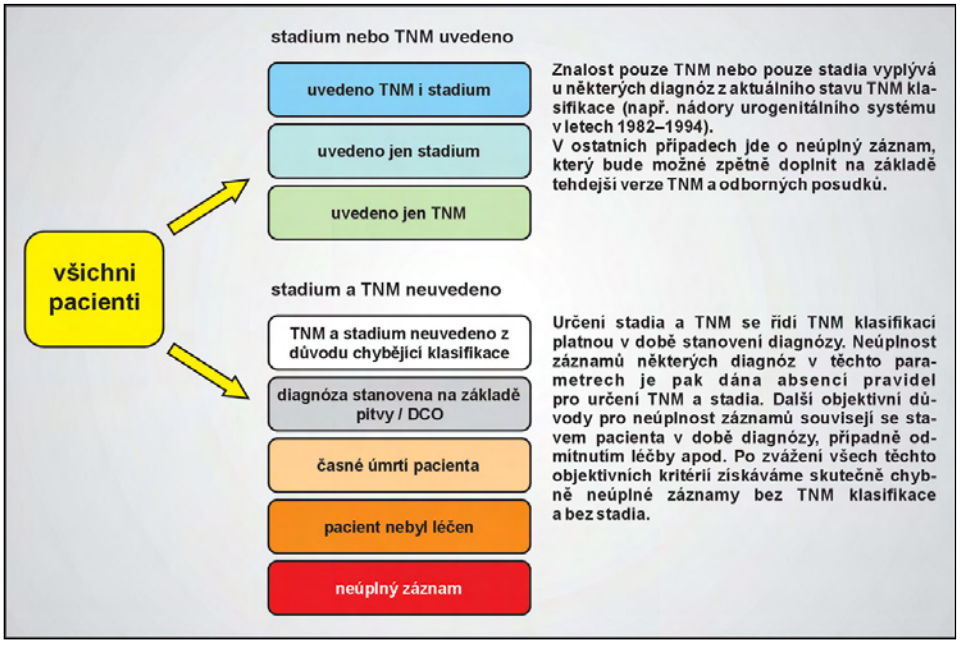
- Záznamy, které mají vedenu alespoň diagnózu ZN, mohou být zahrnuty do základních epidemiologických sumarizací.
- U záznamů, kde je uvedeno buď jen klinické stadium nebo jen TNM, je možné zpětné doplnění druhé položky podle pravidel platných v době diagnózy a na základě expertních posudků.
- Záznamy bez uvedení klinického stadia a bez uvedení TNM, ať již chybou nebo z objektivních důvodů, lze zpětně pravděpodobnostně klasifikovat. Jedním ze standardních řešení tohoto problému je využití tzv. imputačních statistických metod, které pomocí regresních modelů nebo modelových rozložení doplní chybějící údaje nejvíce pravděpodobnou hodnotou. Je-li výsledný soubor konzistentní a projde-li logickými kontrolami, lze s ním následně pracovat jako s úplným (Schenker a Taylor, 1996; Horton a Lipsitz, 2001; Bernaards a kol., 2003). K tomuto účelu však musí být použity pouze techniky, jejichž výsledek lze zpětně kontrolovat, a takto doplněné záznamy musí být jasně označeny.

Všechny výše uvedené rozbory lze ovšem provádět pouze u záznamů, které se vůbec do databáze registru dostaly, audit existujících záznamů neřeší případnou nereprezentativnost registru jako takového. Zde může pomoci pouze práce s externími zdroji dat, demografickými registry nebo mezinárodními databázemi. V ČR tuto roli plní registr zemřelých, z jehož záznamů lze zpětně korigovat odhady mortality. Systémové zkreslení populačních epidemiologických dat lze prokázat následujícími nepřímými metodami (řazeno od nespecifických analýz až po analýzy vyžadující údaje od konkrétních diagnóz):

1) Sledování regionálních a časových fluktuací. Zvýšený podíl této složky variability indikuje zhoršenou kvalitu dat o incidenci a mortalitě a zřejmě i organizační problémy registru.
2) Demografická typologie pacientů, výskyt pohlaví nebo

věkových kategorií podle publikovaných dat nebo ve srovnání s jinými referenčními databázemi.

3) Sledování zkreslení údajů o incidenci zpožděnou registrací. Velmi závažný fakt, který může vést k podhodnocení incidence u řady diagnóz. V studii provedené na databázi SEER (NCI, 2001) zjistili Clegg a kol. u některých diagnóz až 12% podhodnocení v důsledku zpožděného hlášení. Toto zjištění se přitom týká funkčního registru, jehož standardní doba zpoždění hlášení od data diagnózy je 1–2 roky (Clegg a kol., 2002).



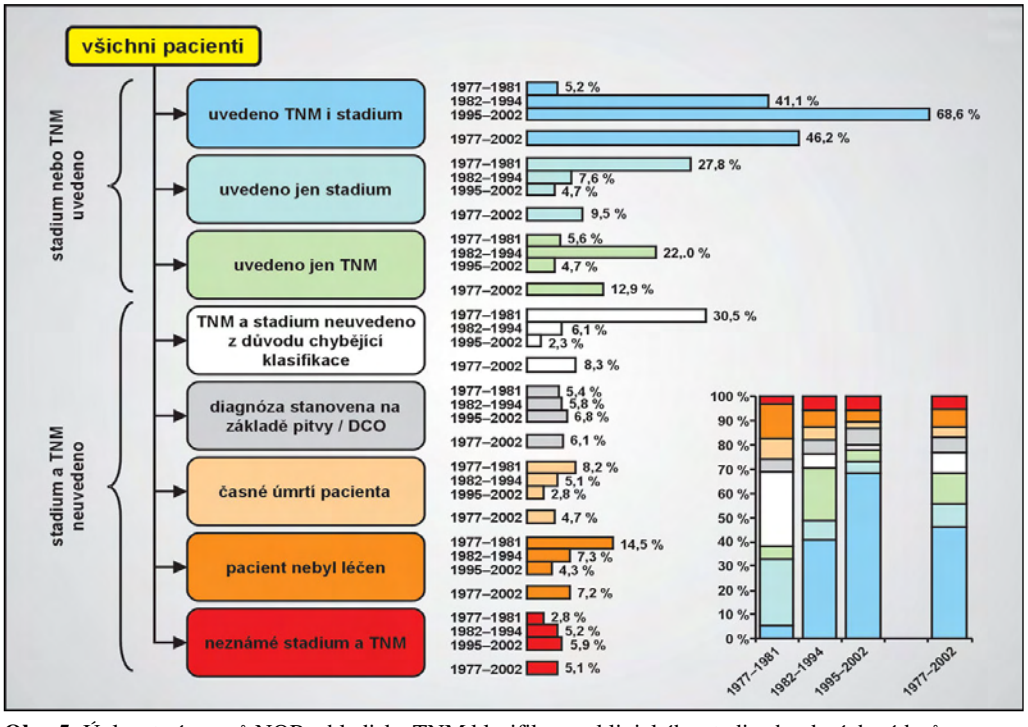
Obr. 4: Základní schéma dělící záznamy NOR podle úplnosti TNM klasifikace

4) Záchyt diagnózy při pitvě a záchyt diagnózy u mrtvého (DCO). Zde opět platí, že zvýšený podíl těchto záchytů může indikovat nereprezentativnost populačního registru. Musíme ovšem brát v úvahu specifika jednotlivých diagnóz, států a také historické souvislosti. Např. v ČR bylo množství diagnóz při pitvě významně zvýšeno před rokem 1989, kdy tehdejší systém řízení zdravotnictví vedl k vysokému počtu prováděných pitev.

5) Výskyt opakovaných malignit. Tento faktor mezi populačními registry značně kolísá, nicméně platí pravidlo, že s rostoucí dobou sledování a velikostí registru by počet opakovaných malignit měl narůstat. Nízký výskyt opakovaných malignit u dlouhodobých registrů (< 3 %) lze považovat za ukazatel zhoršené reprezentativnosti registru (Capocaccia a kol., 2003).

6) Hodnocení rozdílů mezi diagnózami v úplnosti záznamů. Např. Capocaccia a kol. (2003) prokázali na analýze databáze EUROCARE-3, že diagnosticky obtížné a těžko dostupné diagnózy (plíce, trávicí trakt) měly nejmenší podíl ověřených diagnostických nálezů, zatímco dobře diagnostikovatelné nálezy (prsa, melanom) vykazovaly záznamy v lepším stavu.

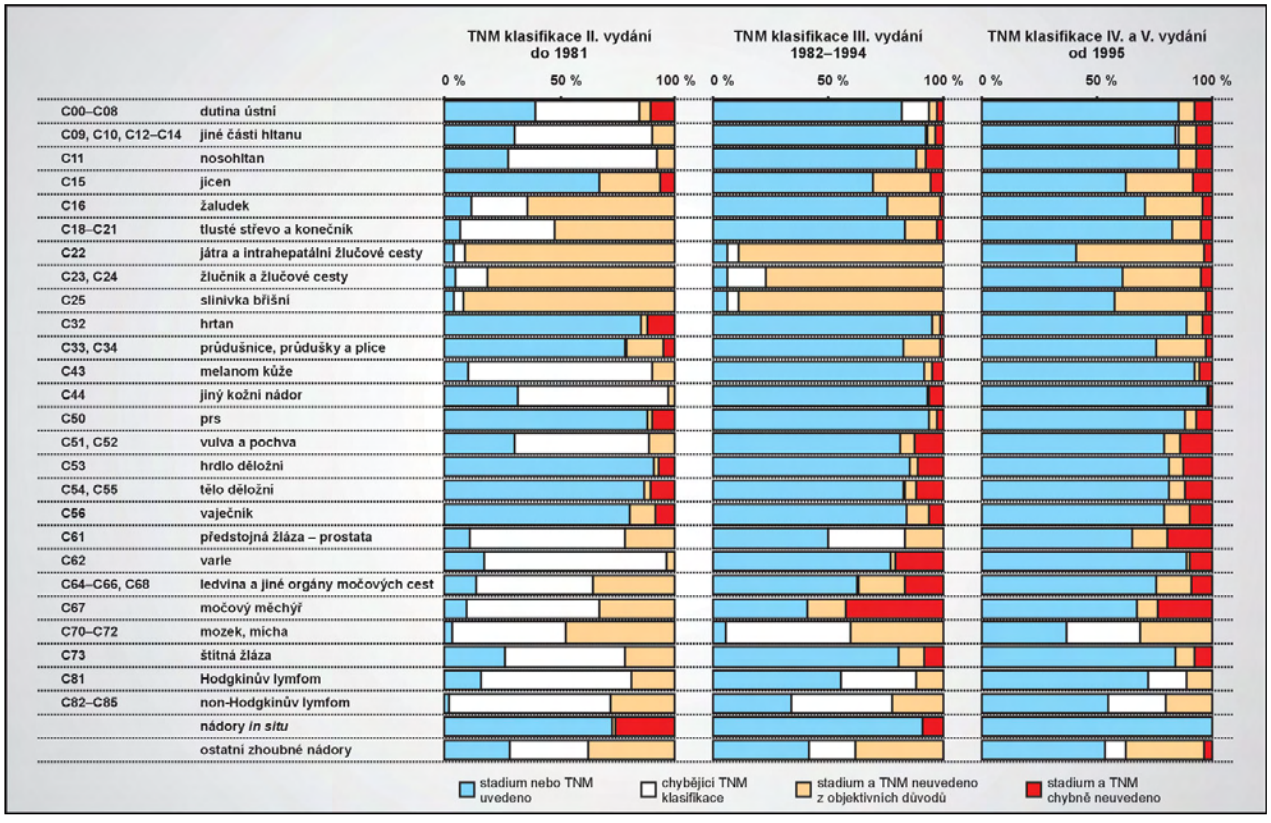
7) Analýza přežití, která by při dodržování standardních léčebných postupů měla u reprezentativního registru vykazovat výsledky srovnatelné s mezinárodními daty. K tomuto účelu se dají využít diagnózy se spíše špatnou prognózou, které nabízejí úplnější data o přežití za kratší časové období, např. akutní leukémie (Capocaccia a kol., 2003).



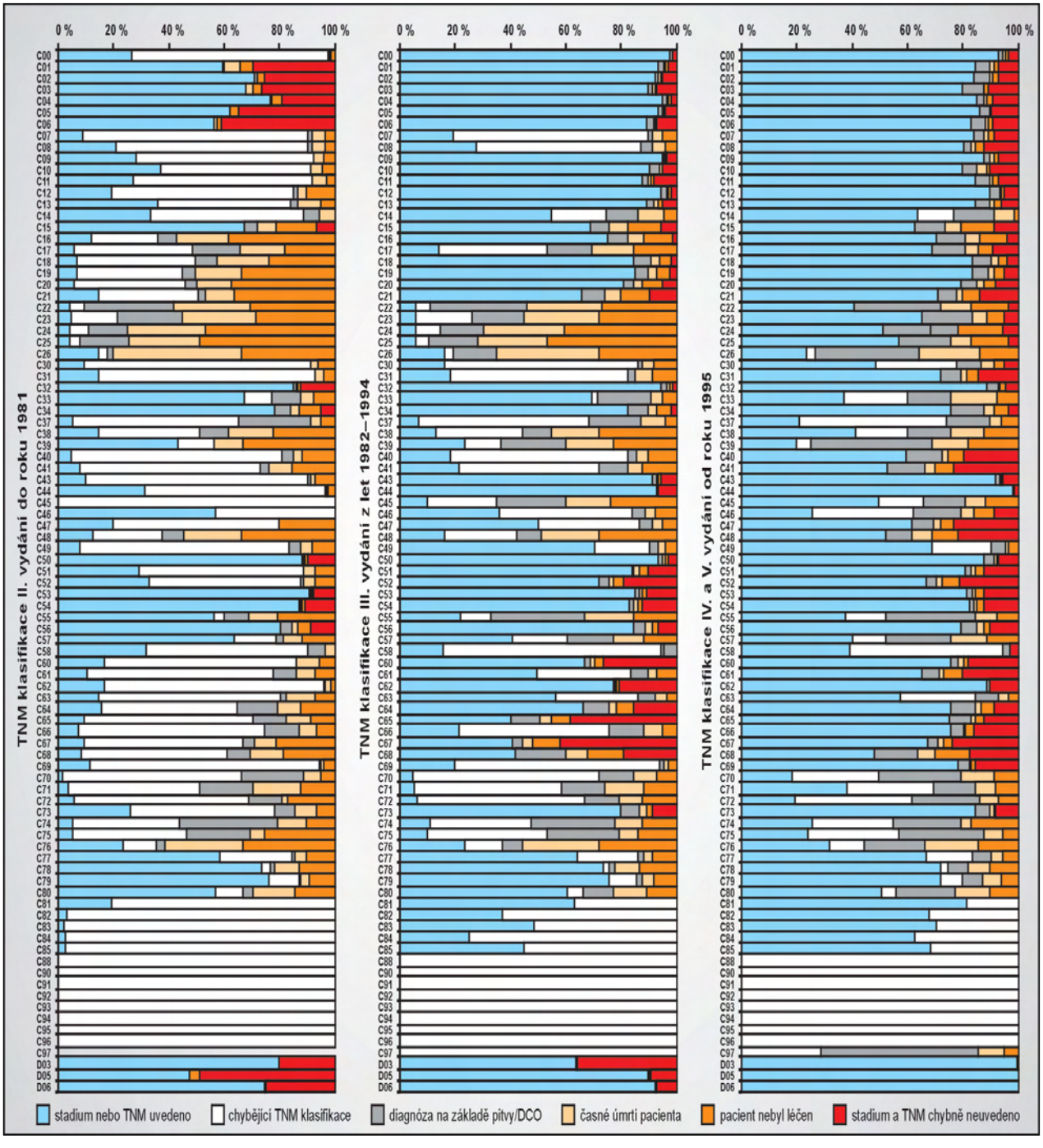
Obr. 5: Úplnost záznamů NOR z hlediska TNM klasifikace a klinického stadia zhoubných nádorů

Tabulka 3: Skupiny diagnóz NOR z pohledu určení klinického stadia a TNM. Období 1995-2002 (TNM klasifikace IV. a V. vydání)

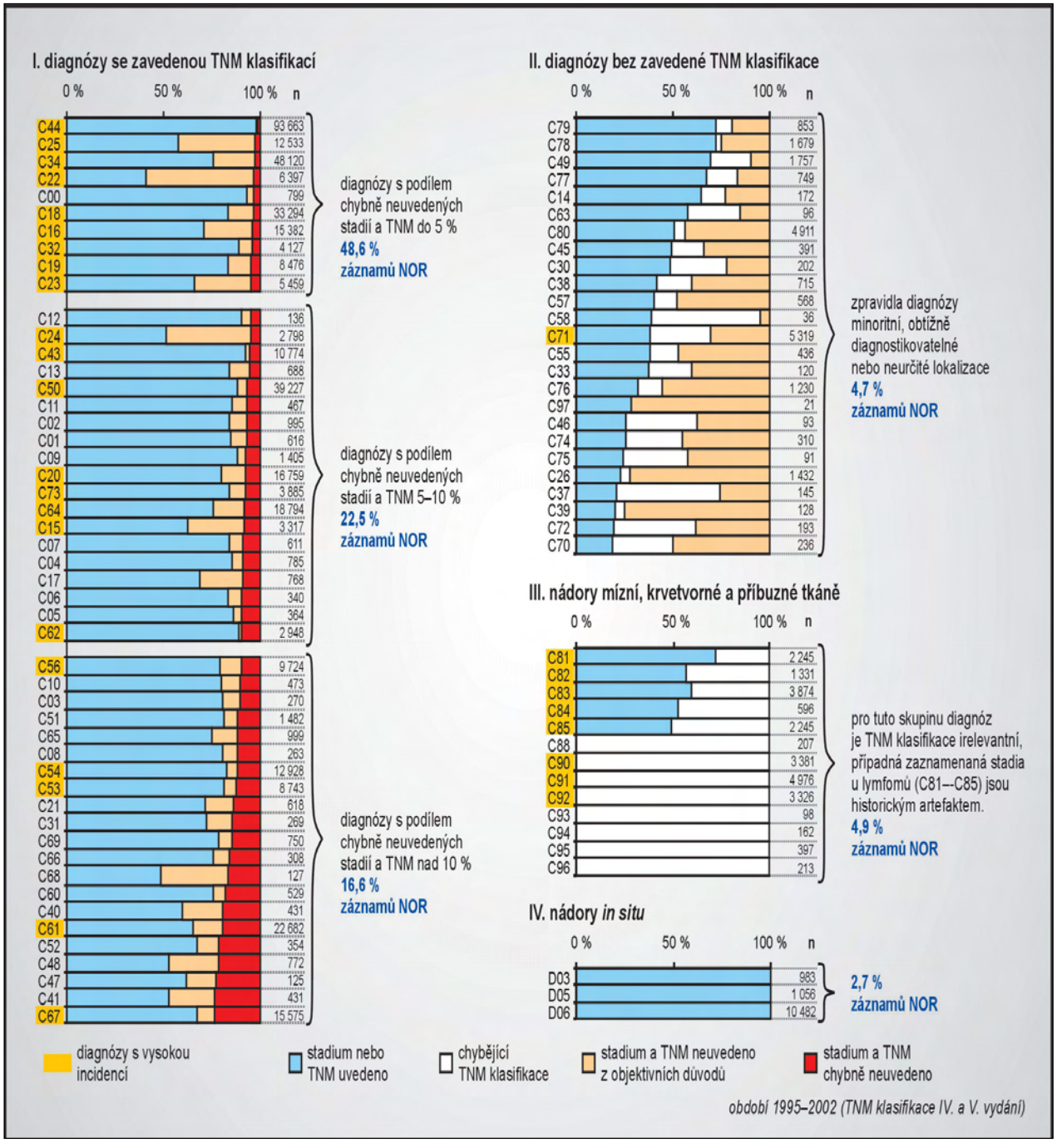
			Stadium nebo TNM uvedeno			Stadium a TNM neuvedeno				
		N	Stadium a TNM uvedeno	Uvedeno stadium	Uvedeno TNM	Chybějící TNM klasifikace	Diagnóza na základě pitvy/DCO	Časné úmrtí pacienta	Pacient nebyl léčen	Stadium a TNM neuvedeno
C00-C08	Dutina ústní	5043	79,3%	2,6%	3,3%		4,1%	1,1%	1,7%	7,8%
C09-C10,C12-C14	Jiné části hltanu	2874	78,3%	3,2%	2,7%	0,8%	4,3%	2,2%	1,5%	7,0%
C11	Nosohltan	467	79,0%	3,9%	1,9%		5,1%	0,6%	2,6%	6,9%
C15	Jícen	3317	54,7%	3,0%	5,1%		10,8%	5,8%	11,8%	8,8%
C16	Žaludek	15382	62,9%	2,3%	5,4%		10,1%	5,8%	9,6%	4,0%
C18-C21	Tlusté střevo a konečník	59147	75,8%	2,0%	4,4%		6,4%	2,3%	3,7%	5,3%
C22	Játra a intrahepatální žlučové cesty	6397	28,7%	3,2%	8,8%		31,0%	11,3%	13,4%	3,6%
C23,C24	Žlučník a žlučové cesty	8257	47,4%	3,1%	10,1%		17,6%	7,1%	9,6%	5,1%
C25	Slinivka břišní	12533	42,5%	2,3%	12,4%		18,5%	7,7%	13,2%	3,4%
C32	Hrtan	4127	84,4%	2,3%	2,2%		3,8%	1,1%	1,7%	4,5%
C33,C34	Průdušnice, průdušky a plíce	48240	66,2%	2,2%	6,9%	0,1%	12,0%	4,0%	5,1%	3,5%
C43	Melanom kůže	10774	87,2%	1,6%	3,2%		1,6%	0,3%	0,5%	5,6%
C44	Jiný kožní nádor	93663	93,7%	0,6%	3,7%		0,1%	0,0%	0,1%	1,8%
C50	Prs	39227	81,9%	2,2%	3,7%		3,5%	1,0%	1,0%	6,7%
C51,C52	Vulva a pochva	1836	69,4%	5,4%	3,6%		2,2%	1,8%	3,8%	13,7%
C53	Hrdlo děložní - cervicis uteri	8743	70,0%	8,0%	3,4%		2,0%	1,3%	2,6%	12,6%
C54,C55	Děloha	13364	72,2%	5,1%	3,5%	0,5%	2,8%	1,1%	2,7%	12,1%
C56	Vaječník	9724	69,6%	4,9%	4,9%		5,6%	2,4%	2,3%	10,2%
C61	Předstojná žláza – prostata	22682	55,2%	5,2%	4,7%		6,4%	1,5%	7,1%	19,9%
C62	Varle	2948	83,0%	2,4%	3,4%		0,9%	0,2%	0,2%	9,9%
C64-C66,C68	Ledvina a jiné orgány močových cest	20228	64,9%	2,8%	7,6%		9,1%	2,1%	4,5%	9,0%
C67	Močový měchýř	15575	59,2%	5,2%	3,2%		3,6%	1,8%	3,2%	23,7%
C72	Mozek a mícha	5748	8,1%	23,9%	4,6%	31,6%	15,9%	7,9%	7,9%	
C73	Štítná žláza	3885	77,9%	3,1%	3,2%		5,8%	1,1%	1,0%	7,9%
C81	Hodgkinův lymfom	2245	1,7%	70,5%	0,1%	27,6%				
C85	Non-Hodgkinův lymfom	8259	3,6%	50,2%	0,9%	45,3%				
C90	Mnohočetný myelom	3381				100%				
C91-C95	Leukémie	8959				100%				
	Nádory in situ	12521	98,5%	0,6%	0,9%					
	Ostatní zhoubné nádory	19689	29,0%	17,2%	7,2%	9,0%	15,7%	9,2%	9,0%	3,7%
	Všechny nádory	469235	68,6%	4,7%	4,7%	2,3%	6,8%	2,8%	4,3%	5,9%



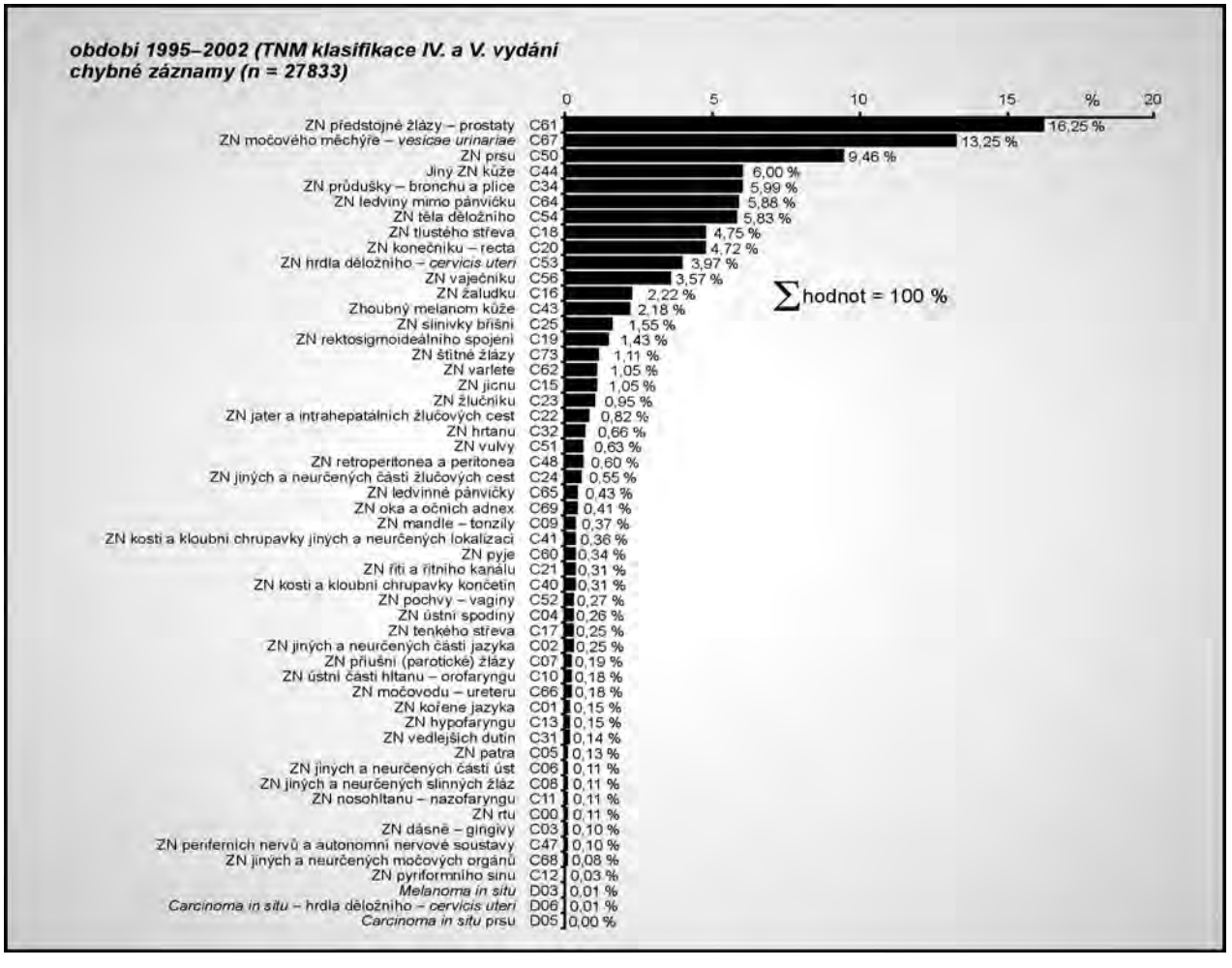
Obr. 6: Skupiny diagnóz NOR z pohledu určení klinického stadia a TNM



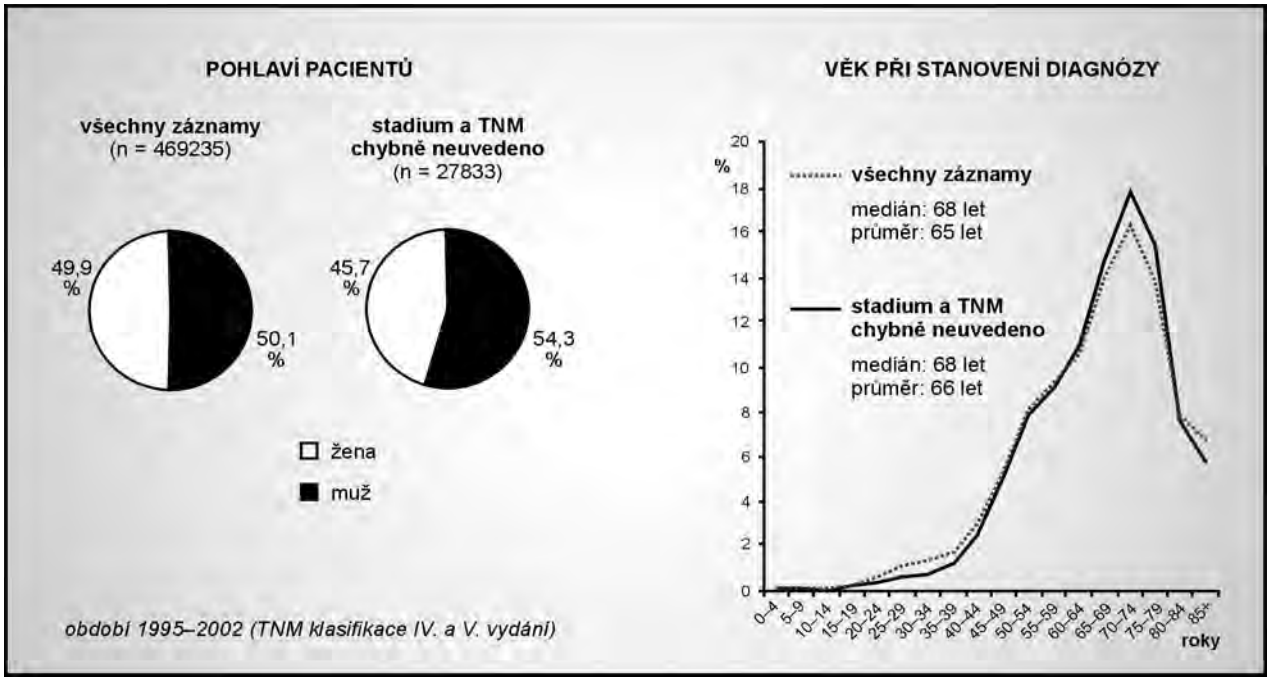
Obr. 7: Diagnózy NOR z pohledu určení klinického stadia a TNM



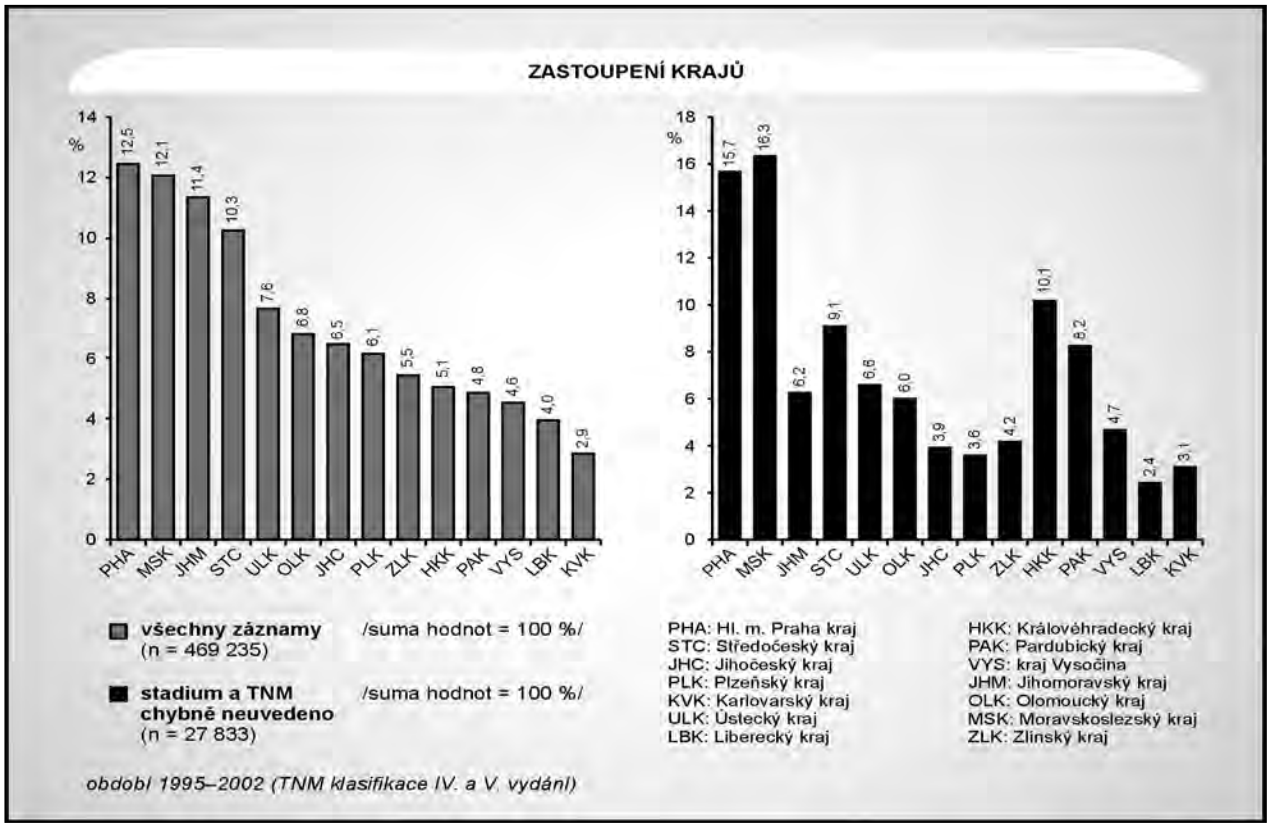
Obr. 8: Diagnózy NOR z pohledu uvedení klinického stadia a TNM



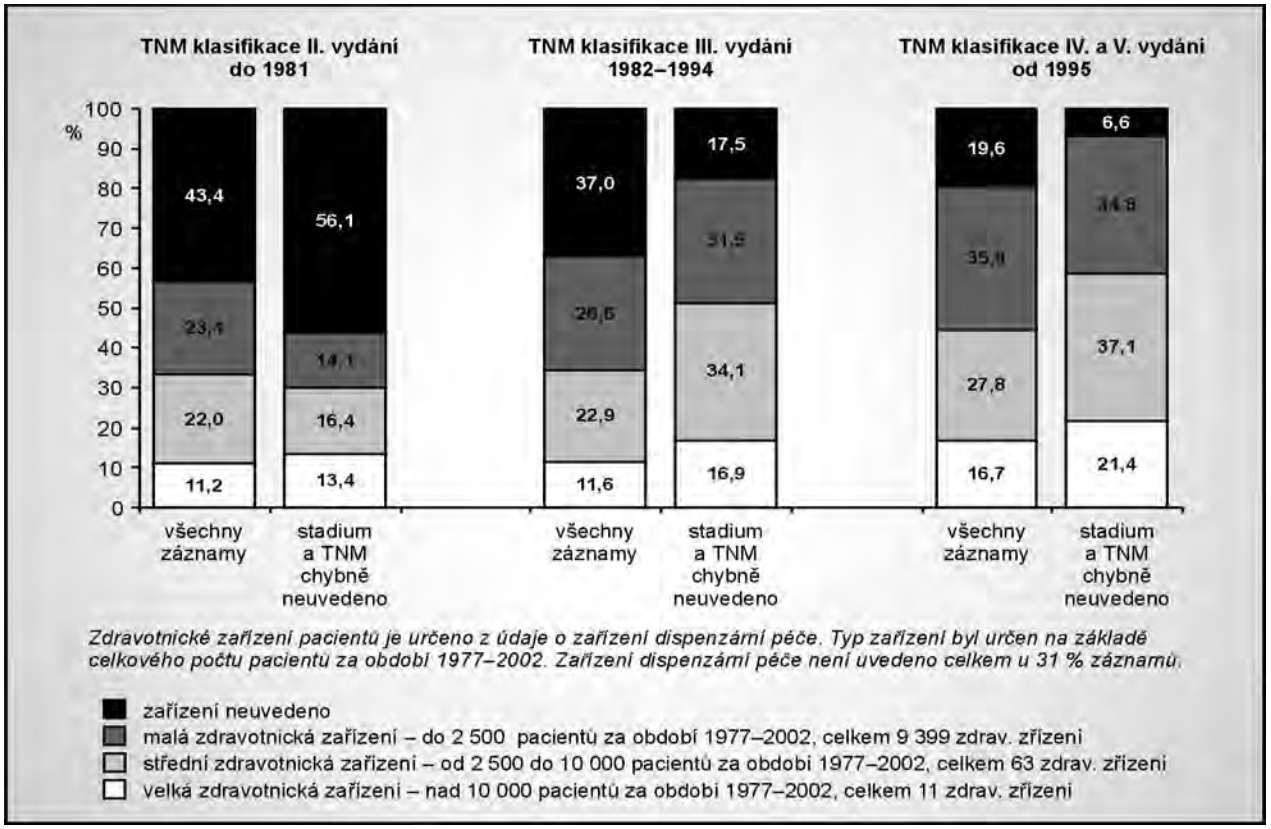
Obr. 9: Struktura záznamů NOR s chybně neuvedeným stadiem a TNM – zastoupení diagnóz



Obr. 10: Typologie záznamů NOR s chybně neuvedeným stadiem a TNM – věk a pohlaví



Obr. 11: Typologie záznamů NOR s chybně neuvedeným stadiem a TNM – zastoupení krajů ČR



Obr. 12: Typologie záznamů NOR s chybně neuvedeným stadiem a TNM – zdravotnická zařízení

Národní onkologický registr ČR: rozbor kvality údajů identifikujících nádor

Všechna uvedená pravidla lze samozřejmě aplikovat na data NOR a získat tak jasnou představu o kvalitě klíčových dat. Již ve výše uvedené diskuzi týkající se obrázku 3 bylo doloženo, že data NOR se v čase zkvalitňují a variabilita odhadů incidence a mortality se pohybuje v přijatelné rovině. Níže uvedený rozbor diagnostických záznamů má za cíl doložit diagnostickou věrohodnost databáze NOR, a tedy i její využitelnost k závažným analýzám, například celkového přežití. Jde o analýzu mimořádné důležitosti, neboť v ČR není žádný podobný soubor dat a pokud by se data NOR neukázala dostatečně kvalitní, přijde česká onkologie o jedinečnou možnost nastavit referenční standardy pro řadu epidemiologických i klinických ukazatelů. Bez významu není také fakt, že v NOR je uložena práce mnoha set lékařů a specialistů, a její zhodnocení je tedy více než žádoucí.

Obrázek 4 znázorňuje základní schéma stratifikace záznamů NOR podle úplnosti v diagnostické části. Schéma je navrženo dle výše popsaných a citačně zdůvodněných kategorií ve snaze odlišit částečně neúplné záznamy od záznamů nevyplněných z objektivních důvodů a od záznamů, které je nutno označit za chybné. Vyhodnocením získáváme na obrázku 5 a v tabulce 3 obraz jednoznačného zkvalitňování dat NOR v čase, což nelze zdůvodnit pouhým vývojem TNM klasifikace, ale i vývojem samotného registru. „Dozrávání" TNM klasifikace je patrné i z obrázků 6–8,

Tabulka 4: Typologie záznamů NOR s chybně neuvedeným klinickým stadiem a TNM – přehled dat. Období 1995-2002 (TNM klasifikace IV. a V. vydání)

	Všechny záznamy (n=469235)	Stadium a TNM neuvedeno (n=27833)
POHLAVÍ		
Muži	50,1%	54,3%
Ženy	49,9%	45,7%
VĚK		
Medián	68,0 let	68,0 let
Průměr	65,4 let	66,2 let
KRAJ		
Hl. m. Praha	12,5%	15,7%
Středočeský	10,3%	9,1%
Jihočeský	6,5%	3,9%
Plzeňský	6,1%	3,6%
Karlovarský	2,9%	3,1%
Ústecký	7,6%	6,6%
Liberecký	4,0%	2,4%
Královéhradecký	5,1%	10,1%
Pardubický	4,8%	8,2%
kraj Vysočina	4,6%	4,7%
Jihomoravský	11,4%	6,2%
Olomoucký	6,8%	6,0%
Moravskoslezský	12,1%	16,3%
Zlínský	5,5%	4,2%
ZAŘÍZENÍ DISPENZÁRNÍ PÉČE		
Velká zdravotnická zařízení	16,7%	21,4%
Střední zdravotnická zařízení	27,8%	37,1%
Malá zdravotnická zařízení	35,9%	34,8%
Zařízení neuvedeno	19,6%	6,6%

kdy u nejmladších dat NOR zaznamenáváme pouze 2,3 % neúplných záznamů z důvodu chybějící TNM klasifikace. Nutno zdůraznit, že i v současnosti existují specifické skupiny nádorů, pro které není definován TNM systém. Tyto diagnózy jsou objektivně nekompatibilní s většinových standardem registrace solidních nádorů a měly by tudíž být posuzovány odděleně. Z populačního hlediska nedojde k žádnému zkreslení, neboť jde o skupiny s relativně malou až velmi nízkou incidencí (hematoonkologické malignity, ZN mozku a CNS, nádory *in situ*, jiné kožní nádory …..) (Davis a kol., 1997; Fritz a kol., 2000; Castillo a kol., 2004). Budeme-li předpokládat, že u záznamů, kde známe pouze TNM, umíme klinické stadium doplnit, pak především data z období 1995–2002 dosahují úplnosti srovnatelné s mezinárodními databázemi (Clegg a kol., 2002; Capocaccia a kol., 2003; Eaker a kol., 2006). Incidence DCO záznamů a nálezů při pitvě je u recentních dat rovněž akceptovatelná i z mezinárodního hlediska (Capocaccia a kol., 2003; Jemal a kol., 2004). Data na obrázku 5 dále dokumentují v čase klesající podíl neléčených pacientů a pacientů s časným úmrtím po diagnóze. Pro neúplnost záznamů z důvodů časného úmrtí jsme použili hranici jednoho měsíce od diagnózy, což je kritérium paušálně používané i při analýzách v mezinárodních projektech, např. EUROCARE 3 (Gatta a kol., 2000; Capocaccia a kol., 2003; Micheli a kol., 2003). V celkovém hodnocení bylo v NOR identifikováno pouze 5,1 % záznamů, ve kterých neodůvodněně chybí jak TNM klasifikace, tak i klinické stadium. Velmi pozitivní zprávou je, že úplnost dat roste v čase a nejnovější období, které je také nejpotřebnější pro klinické analýzy, poskytuje data v dostatečné kvalitě (Obrázek 5, 8). Dále je zřetelné, že problém neúplné diagnostické identifikace se v období po roce 1995 méně dotýká nejčastějších diagnóz, jako je především karcinom prsu, kolorekta a plic (Obrázek 6,7). Ačkoli vzhledem k vysoké incidenci představují neúplné záznamy těchto diagnóz numericky velký počet (Obrázek 9), v relativním podílu neovlivňují hodnocení celkových dat (např. Obrázek 7).

Dle výše uvedených pravidel nemohou být záznamy s chybně neuvedenou diagnostikou využity například pro analýzu celkového přežití a další klinicky orientované výstupy. Typologie těchto záznamů provedená na obrázcích 10–12 a v tabulce 4 dokumentuje, že tyto záznamy přibližně odpovídají celkové databázi NOR ve všech klíčových parametrech (demografické charakteristiky pacientů, zdravotnické zařízení aj.). I vzhledem k relativně malému podílu chybných záznamů tedy není třeba se obávat systémového zkreslení výstupů databáze NOR jako celku. Možnost zkreslení existuje pouze při některých regionálních srovnáních. Relativně zvýšený podíl (> 15 %) chybně neúplných záznamů byl zachycen v Praze a v Moravskoslezském kraji (Obrázek 10), což by mohlo narušit regionální srovnání v některých diagnostických skupinách, např. u karcinomu prsu.

Výsledkem auditu správnosti údajů NOR (Obrázky 4–11, Tabulky 3–4) je soubor věrohodných záznamů s řádně zaznamenanou diagnostikou a soubor neúplných záznamů s jasně identifikovanou objektivní příčinou problémů. Výše uvedené filtry vysvětlují většinu záznamů s chybějícím TNM a klinickým stadiem. Obdobný postup filtrace je využíván u většiny populačních studií, které sledují přežití onkologických pacientů ve vztahu k diagnostice nebo léč-

bě. Celkově bylo zachyceno 5,1 % záznamů s chybně neuvedeným klinickým stadiem i TNM klasifikací. I tyto záznamy ovšem mohou být využity při běžném hodnocení incidence a mortality, neboť mají uvedenu diagnózu ZN. Při detailní analýze bylo zjištěno, že typologicky odpovídají celkové databázi NOR. Pro účely hodnocení léčebné péče lze tedy tyto záznamy vypustit, aniž bychom se dopustili systémového zkreslení anebo snížili informační hodnotu výstupů. Tyto záznamy mohou být také podrobeny tzv. statistické imputaci klinického stadia modelovým odhadem. Tento krok ale není pro řadu populačních analýz nutný, naopak standardem v literatuře je tzv. „unstaged" záznamy hodnotit odděleně a přímo pro ně kalkulovat i hodnoty přežití, které obvykle vycházejí srovnatelné s klinickým stadiem 2 až 3 (např. Sant a kol., 2003; Brenner a Arndt, 2005).

Literatura:

Bernaards C.A., Farmer M.M., Qi K., Dulai G.S., Ganz P.A., Kahn K.L.: Comparison of two multiple imputation procedures in a cancer screening survey. J. Data Sci,1, 1-20, 2003.

Brenner H., Arndt V.: Long-term survival rates of patients with prostate cancer in the prostate-specific antigen screening era: population-based estimates for the year 2000 by period analysis. J. Clin. Oncol., 23(3), 441 – 447, 2005.

Capocaccia D., Gatta G., Roazzi P., Carrari E., Santaquilani M., de Angelis R., Tavilla A., Eurocare Working Group: The EUROCARE-3 database: methodology of data collection, standardization, quality control and statistical analysis. Annals of Oncology, Supplement 5: v14-v27, 2003.

Castillo M.S., Davis F.G., Surawicz T., Bruner J.M., Bigner S., Coons S., Bigner D.D.: Consistency of primary brain tumor diagnoses and codes in cancer surveillance systems. Neuroepidemiology, 23, 85 – 93, 2004.

Clegg L.X., Feuer E.J., Midthune D.N., Fay M.P.: Impact of reporting delay and reporting error on cancer incidence rates and trends. J. Nat. Cancer Inst., 94 (20), 1537 – 1545, 2002.

Davis F.G., Bruner J.M., Surawicz T.S.: The rationale for standardized registration and reporting of brain and central nervous system tumors in population-based registries. Neuroepidemiology, 16(6), 308-3016, 1997.

Eaker S., Dickman P.W., Berquist L., Holmberg L.: Differences in management of older women influence breast cancer survival: results from a population based database in Sweden. PloS Med 3(3): e25, 2006.

Fritz A., Percy C., Jack A. a kol.: International classification of diseases of oncology. Geneva: World Health Organization, 2000.

Gatta G., Capocaccia R., Ries L. a kol.: Towards a comparison of survival in America and European cancer patients. Cancer 2000, 89: 893 – 900, 2000.

Horton N.J., Lipsotz S.R.: Multiple imputation for the fatal accident reporting system. Appl Stat – J. Royal Stat Soc ., Series C, 40, 244 – 254, 2001.

Jemal A., Clegg L.X., Ward E. a kol.: Annual report to the nation on the status of cancer, 1975 – 2001, with a special feature regarding survival. Cancer, 101(1), 4 – 27, 2004.

Micheli A., Baili P., Quinn M.: EUROCARE Working Group. Life expectancy and cancer survival in the EUROCARE-3 cancer registry areas. Ann Oncol, 14 (Suppl 5): v28-v40, 2003.

NCI (National Cancer Institute): Surveillance, Epidemiology and End Results. http://www.-seer.ims.nci.nih.gov, 2001.

Sant M., Allemani C., Capocaccia R., Hakulinen T., Aareleid T. a kol.: Stage at diagnosis is a key explanation of differences in breast cancer survival across Europe. Int. J. Cancer, 106, 416 – 422, 2003.

Sant M. and EUROCARE Working Group: Differences in stage and therapy for breast cancer across Europe. Int. J. Cancer, 93, 894 – 901, 2001.

Shetty M.R., Reinam H.M.: Tumour size and axillary metastasis, a correlative occurrence in 1244 cases of breast cancer between 1980 and 1995. Eur. J. Surg. Oncol., 23, 139 – 141, 1997.

Schenker N., Taylor J.M.G.: Partially parametric techniques for multiple imputation. Comput. Stat. Data Anal., 22, 425 – 446, 1996.