

**MASARYKOVA UNIVERZITA**  
**PŘÍRODOVĚDECKÁ FAKULTA**  
**STUDIJNÍ PROGRAM: EXPERIMENTÁLNÍ BIOLOGIE**

# **Diplomová práce**

**Brno 2018**

**Monika Bezděková**



**MASARYKOVA UNIVERZITA**  
**PŘÍRODOVĚDECKÁ FAKULTA**  
**STUDIJNÍ PROGRAM: EXPERIMENTÁLNÍ BIOLOGIE**

---



**IDENTIFIKACE**  
**KAUZÁLNÍCH VZTAHŮ**  
**V SÍTI HUMÁNNÍCH**  
**FENOTYPŮ**

Diplomová práce

**Monika Bezděková**

Vedoucí práce: RNDr. Jiří Jarkovský, Ph.D.

**Brno 2018**

## Bibliografický záznam

<b>Autor:</b>	Bc. Monika Bezděková Přírodovědecká fakulta, Masarykova univerzita Centrum pro výzkum toxických látek v prostředí
<b>Název práce:</b>	Identifikace kauzálních vztahů v síti humánních fenotypů
<b>Studijní program:</b>	Experimentální biologie
<b>Studijní obor:</b>	Matematická biologie
<b>Vedoucí práce:</b>	RNDr. Jiří Jarkovský, Ph.D.
<b>Konzultant:</b>	Mgr. Klára Benešová
<b>Akademický rok:</b>	2017/2018
<b>Počet stran:</b>	63+7
<b>Klíčová slova:</b>	Síť humánních fenotypů; Administrativní data zdravotnických zařízení; Národní registr hospitalizovaných; Kauzalita; Epidemiologie; Teorie grafů; Network science

## Bibliographic Entry

**Author** Bc. Monika Bezděková  
Faculty of Science, Masaryk University  
Research Centre for Toxic Compounds in the  
Environment

**Title of Thesis:** Identification of causal relations in phenotypic disease  
network

**Degree programme:** Experimental Biology

**Field of Study:** Computational Biology

**Supervisor:** RNDr. Jiří Jarkovský, Ph.D.

**Consultant:** Mgr. Klára Benešová

**Academic Year:** 2017/2018

**Number of Pages:** 63+7

**Keywords:** Phenotypic disease network; Administrative  
data of health facilities; National Register of Hospitalised  
Patients; Causality; Epidemiology; Graph theory;  
Network science

## **Abstrakt**

Práce se zabývá identifikací orientovaných vztahů v síti humánních fenotypů na základě administrativních dat. Tato data v sobě obsahují velké množství informací o pacientech a jejich onemocněních a navíc jsou dostupná z dlouhodobého časového hlediska. Identifikace příčinných vztahů na základě hospitalizační historie pacientů přispívá k pochopení progresu jednotlivých onemocnění. Datovým zdrojem pro analýzu byla data z Národního registru hospitalizovaných z let 2007–2016. Práce vyžadovala zvládnutí velkých dat a následnou aplikaci výpočetně náročných algoritmů. Výsledkem jsou sítě humánních fenotypů s orientovanými vztahy na různých úrovních Mezinárodní klasifikace nemocí. Vizualizované vztahy spadají do různých oblastí jako například onkologie, onemocnění srdce a psychiatrie. Získané vztahy jsou v souladu s literaturou. I přes limitace administrativních dat nemocnic tak lze říci, že jsou pro tento typ analýz využitelná. Do budoucna by dalším krokem mohla být analýza provedená na datech obohacených o ambulantní záznamy.

## **Abstract**

The thesis deals with identification of oriented relationships in the phenotypic disease network based on patient administrative data. These data contain a wealth of information about patients and their diseases. In addition, these data are available in the long time perspective. Identifying causal relationships based on patients' hospitalization history could help to understand the progression of the disease. The data source for the analysis was from the National Register of Hospitalised Patients from the years 2007–2016. For the work it was necessary to handle this big data and subsequently apply computationally demanding algorithms. The result are oriented phenotypic disease networks at different levels of the International Classification of Diseases. Visualized relationships fall into various areas such as oncology, heart disease and psychiatry. The relationships obtained from the networks are in agreement with literature. Despite limitations of hospital administrative data this data source can be said to be applicable to this type of analysis. The next step could be analysis performed on data enriched by outpatient records.



## ZADÁNÍ DIPLOMOVÉ PRÁCE

Akademický rok: 2016/2017

**Ústav:** Centrum pro výzkum toxických látek v prostředí

**Studentka:** Bc. Monika Bezděková

**Program:** Experimentální biologie

**Obor:** Matematická biologie

Ředitel *Centra pro výzkum toxických látek v prostředí* PŘF MU Vám ve smyslu Studijního a zkušebního řádu MU určuje diplomovou práci s názvem:

**Název práce:** Identifikace kauzálních vztahů v síti humánních fenotypů

**Název práce anglicky:** Identification of causal relations in phenotypic disease network

### Oficiální zadání:

Národní registr hospitalizovaných (NRHOSP) je populačním zdrojem dat o osobách hospitalizovaných na lůžkových odděleních včetně příslušných hospitalizačních diagnóz kódovaných podle MKN-10 klasifikace. Tato data lze využít ke zkoumání progresu onemocnění a hodnocení asociací mezi nemocemi. Síť humánních fenotypů (phenotypic disease network – PDN) vyjadřuje spojitost mezi výskytem jednotlivých onemocnění, avšak nehodnotí časovou souslednost výskytu. Úkolem studentky bude vytvořit síť nemocí rozšířenou o kauzální informaci, identifikovat zajímavé vztahy a objasnit jejich mechanismy, jsou-li známy.

### Cíle práce:

- 1) rešerše dostupné literatury zabývající se využitím administrativních dat nemocnic pro hodnocení kauzálních vztahů mezi výskytem nemocí;
- 2) seznámení se se strukturou dat NRHOSP, provedení vhodné transformace dat a výběr optimální metodiky pro jejich analýzu;
- 3) nalezení netriviálních orientovaných vztahů mezi výskytem nemocí na základě historie hospitalizací pacientů se záznamem v datech NRHOSP;
- 4) vizualizace a zhodnocení nalezených a potenciálně zajímavých vztahů;
- 5) srovnání výsledků práce se závěry jiných publikovaných studií.

### Doporučená literatura:

Hidalgo, C. A., Blumm, N., Barabási, A. L., & Christakis, N. A. (2009). A dynamic network approach for the study of human phenotypes. *PLoS Comput Biol*, 5(4).

Kannan, V., Swartz, F., Kiani, N. A., Silberberg, G., Tspiras, G., Gomez-Cabrero, D., Alexanderson, K., Tegnèr, J. (2016). Conditional disease development extracted from longitudinal health care cohort data using layered network construction. *Scientific reports*, 6.

**Jazyk závěrečné práce:** čeština

**Vedoucí práce:** RNDr. Jiří Jarkovský, Ph.D.

**Konzultant:** Mgr. Klára Benešová

**Datum zadání práce:** 25. 11. 2016

**V Brně dne:** 17. 2. 2017

Souhlasím se zadáním (podpis, datum):



.....  
Bc. Monika Bezděková  
studentka



.....  
RNDr. Jiří Jarkovský, Ph.D.  
vedoucí práce



.....  
doc. RNDr. Ladislav Dušek, Ph.D.  
zástupce ředitele Centra pro  
výzkum toxických látek v prostředí

## **Poděkování**

Na tomto místě bych chtěla poděkovat Mgr. Kláře Benešové a RNDr. Jiřímu Jarkovskému, Ph.D., za odborné vedení a cenné rady poskytnuté při vypracování této práce. Děkuji také své rodině za podporu během celého studia.

## **Prohlášení**

Prohlašuji, že jsem svoji diplomovou práci vypracovala samostatně s využitím informačních zdrojů, které jsou v práci citovány.

Brno 14. května 2018

.....

Monika Bezděková



## Obsah

1. Úvod a cíle práce .....	2
2. Kauzální vztahy v epidemiologii .....	3
2.1 Příklady aplikace na administrativní data .....	4
3. Národní zdravotnický informační systém (NZIS) .....	7
3.1 Národní registr hospitalizovaných (NRHOSP) .....	9
4. Network science a teorie grafů .....	11
4.1 Teorie grafů .....	13
5. Síť humánních fenotypů a kauzalita .....	17
5.1 Síť humánních fenotypů .....	17
5.2 Kauzalita .....	21
5.3 Kauzalita v síti humánních fenotypů .....	22
6. Praktická část .....	25
6.1 Použitý software .....	25
6.2 Datový soubor .....	25
6.2.1 Příprava datového souboru .....	25
6.2.2 Základní charakteristiky datového souboru .....	27
6.2.3 Tvorba kombinací .....	30
6.3 Síť humánních fenotypů .....	33
6.4 Síť humánních fenotypů s orientovanými vztahy .....	35
6.4.1 Kapitoly MKN-10 .....	35
6.4.2 Oddíly MKN-10 .....	38
6.4.3 Tríznačkové diagnózy MKN-10 .....	44
7. Diskuze .....	50
8. Závěr .....	52
Seznam použité literatury .....	53
Internetové zdroje .....	60
Seznam použitých zkratk .....	61
Seznam obrázků .....	62
Seznam tabulek .....	63
Přílohy .....	64
Příloha 1 .....	64
Příloha 2 .....	65
Příloha 3 .....	66

## 1. Úvod a cíle práce

Zdravotní stav pacientů je typicky charakterizován výskytem několika různých chorob a studium společného výskytu onemocnění na základě hospitalizační historie pacientů je potenciálně důležitým tématem výzkumu, který přináší nový úhel pohledu na možnou povahu zkoumaných vztahů. Přidáme-li ohodnocení o kauzální informaci, získáváme informaci o progresi onemocnění z hlediska přidružených chorob. Informace o tom, že se jedna nemoc s jistou pravděpodobností vyvine v chorobu druhou, pak může být cennou informací například pro preventivní medicínu.

Administrativní data pacientů s sebou nesou řadu výhod v podobě jejich populačního charakteru, velkého množství informací o pacientech a jejich diagnózách, často dlouhodobého časového rámce a relativně snadné dostupnosti. I přes tyto výhody je však nutné se řádně seznámit s řešenou problematikou i daty samotnými a minimalizovat možné komplikace a zkreslení při analýze. Při snaze získat orientované vztahy mezi nemocemi z administrativních záznamů pacientů je nutné se vypořádat s velkým objemem záznamů. Identifikovat smysluplné a ne zcela triviální vztahy v takto husté síti se kvůli samotnému objemu a složitosti údajů ukázalo být náročným úkolem.

Mezi cíle diplomové práce patřilo seznámit se se strukturou dat Národního registru hospitalizovaných a najít vhodnou transformaci a metodiku pro jejich analýzu. V návaznosti na to bylo kladeno za cíl nalezení netriviálních orientovaných vztahů mezi výskytem nemocí na základě hospitalizační historie pacientů a následná vizualizace potenciálně zajímavých vztahů. Výsledky práce pak byly konfrontovány se závěry jiných studií. Součástí teoretické části práce je rovněž rešerše dostupné literatury zabývající se využitím administrativních dat pro hodnocení kauzálních vztahů mezi výskytem nemocí.

## 2. Kauzální vztahy v epidemiologii

Kauzalita se do diskuze o epidemiologických výsledcích pravděpodobně poprvé dostala v padesátých a časných šedesátých letech 20. století. V té době do popředí zájmu vešlo zkoumání vztahu mezi kouřením a rakovinou plic (Kundi, 2006).

Příčinná souvislost patří mezi základní epidemiologické pojmy a kauzální výroky typu „kouření způsobuje rakovinu“ jsou již po dlouhou dobu součástí epidemiologické literatury, přesto však neexistuje žádná jednoznačná definice této oblasti. Pokud někdo tvrdí, že "kouření způsobuje rakovinu", mohlo by to totiž znamenat, že "pouze kouření způsobuje rakovinu", "u všech lidí, kteří kouří, se rozvine rakovina" nebo "alespoň u jednoho kuřáka dojde k rozvoji rakoviny" (Parascandola et Weed, 2001). MacMahon a Pugh v roce 1970 napsali, že „slovo příčina je abstraktní podstatné jméno a stejně jako krása bude mít různé významy při různém kontextu“. Aby bylo možné použít slovo "příčina" jako vědecký výraz, je důležité objasnit, co se tím v daném kontextu rozumí.

Dlouhodobé diskuze se zabývaly tím, za jakých podmínek lze dopadu různých determinantů přisuzovat příčinnou souvislost s určitým cílovým onemocněním. Takzvaná kritéria kauzality, která popsal sir Austin Bradford Hill v roce 1965, jsou často schematicky aplikována bez ohledu na skutečnost, že sám Hill varoval před jejich bezmyšlenkovitým užíváním a nebyly míněny jako kritéria ani jako kontrolní seznam pro stanovení potenciálních příčin onemocnění. Navíc se projevuje tendence nesprávně interpretovat nedostatek důkazů o příčinných souvislostech jako důkaz neexistence příčinné souvislosti. První případ, tedy nedostatek důkazů pro prokázání příčinného vztahu, pouze vyjadřuje nedostatek znalostí či informací. V druhém případě, při důkazu neexistence příčinného vztahu, jde o důležité rozhodnutí, které může mít významné důsledky pro opatření přijatá v rámci prevence. Mnohdy pak dochází ke špatnému pochopení a tvrzení, že expozice není asociována s nepříznivými zdravotními účinky. Často jsme zanecháni před řadou problémů, které nelze optimálně řešit – neexistují žádná kritéria, která by v případě naplnění vedla k tomu, aby faktor byl označen jako kauzální či nikoli. To však neznamená, že podle našich současných poznatků nemůžeme dospět k rozhodnutí o vztahu expozice a nemoci (Kundi, 2006). Toho se týká i Hillův citát: „Veškerá vědecká práce je neúplná – ať už jde o pozorování nebo experiment. Veškerá vědecká práce může být rozrušena nebo pozměněna pokrokem v poznání. To nás neopravňuje neuplatňovat znalosti, které již máme, nebo odložit práci, o níž se zdá, že je požadována v danou dobu“.

Stuart et al. v roce 2013 řešili problematiku odhadů příčinných vztahů z observačních studií, za využití elektronických zdravotních záznamů. Zmiňují, že ačkoli se randomizované studie obecně považují za nejlepší způsob, jak odhadnout kauzální účinky, samotná realizace těchto studií je mnohdy náročná a někdy dokonce neuskutečnitelná. Elektronické zdravotní záznamy oproti tomu nabízí velké množství informací, často i z dlouhodobého hlediska. Rozsáhlé údaje však samy o sobě nemohou odpovědět na kauzální otázky, proto je třeba stále přemýšlet, jak získat nejlepší odpovědi na výzkumné otázky týkající se kauzality.

Dahlem et al. se v roce 2015 věnovali problematice predikovatelných vztahů z elektronických zdravotních záznamů a zkoumali, do jaké míry lze prognózu zdravotního stavu pacienta předvídat vzhledem k jeho historii diagnóz. Schopnost predikovat progresi

onemocnění vzhledem k historii chorob daného člověka má potenciál řešit jednu z nejnaléhavějších otázek společnosti: posílení poskytování zdravotní péče a snižování nákladů. Kontrola progresu onemocnění je neodmyslitelně spojena se schopností předpovědět možná budoucí onemocnění vzhledem k pacientově anamnéze. Autoři mimo jiné představili nástroj pro studium kvality predikce, který je součástí elektronických lékařských záznamů. Podle nich by další práce za využití podrobnějších filtračních technik a při zaměření na konkrétní podmínky mohla mít potenciál poskytnout cenný vhled do medicíny založené na důkazech.

Listl et al. se zabývali příčinnými vztahy získanými z observačních dat. Cílem jejich práce publikované v roce 2016 bylo mimo jiné přispět k informované debatě o metodologických přístupech pro hodnocení kauzality v dentálním výzkumu. Jejich studie poskytla přehled nejmodernějších metod specificky určených pro hodnocení kauzality z observačních dat, které by i vzhledem k rostoucí dostupnosti rutinně sbíraných administrativních dat a elektronických zdravotních záznamů mohly být užitečné při dentálním výzkumu. Chtěli rovněž vyzdvihnout, že je možné využít širší škálu metod pro určování příčinných souvislostí než se spoléhat pouze na randomizované kontrolované studie.

Naproti tomu Lee a Yoon, kteří se ve své studii z roku 2017 věnovali problematice velkých dat v medicíně, píše, že analýza těchto dat je cílena spíše na generování hypotéz, než jejich testování a zaměřují se především na časovou stabilitu a asociace než na kauzální vztahy. Poukazují na to, že analýzy rozsáhlých lékařských dat jsou komplikovány řadou technických otázek jako chybějící hodnoty, nadměrná dimenzionalita či zkreslení a sdílejí přirozená omezení observačních studií jako neschopnost testovat kauzalitu v důsledku nadbytečných a zavádějících faktorů a reverzibilních příčin.

Administrativní data v sobě skrývají velký potenciál s ohledem na množství obsažené informace, často dlouhého časového okna a dostupnosti. Je však nutné se řádně seznámit s řešenou problematikou i daty samotnými a minimalizovat možné komplikace a zkreslení na minimum.

## 2.1 Příklady aplikace na administrativní data

Chen et al. v roce 2009 provedli studii, kde se zabývali predikcí progresu rakovinných metastáz. Prostřednictvím rozsáhlého souboru pacientů s onkologickým onemocněním zkoumali topografické modely rozvoje metastáz, a to za využití síťového přístupu. Studie ukázala, že zacházení se sekundárními metastázami jako se samostatnými komorbidními nemocemi umožňuje konstrukci sítí zaměřených na progresi rakovinových metastáz. Z těchto sítí mohli poté analyzovat jejich dynamiku, sítě mezi sebou srovnávat a také je byli schopni využít jako základ pro prediktivní algoritmy. Tyto sítě mohou být použity k identifikaci pravděpodobné sekvence metastáz u pacienta a podnítit diagnostické testování a případnou specifickou léčbu.

Himes et al. se v roce 2009 ve své studii věnovali identifikaci rizikových faktorů chronické obstrukční plicní nemoci u pacientů s astmatem za využití elektronických zdravotnických záznamů. Autoři vytvořili prediktivní model pro chronickou obstrukční plicní nemoc využívající komorbidit a demografických informací získaných ze zdravotních záznamů pacientů s astmatem. Při aplikaci modelu na nezávislou skupinu pacientů s diagnózou astmatu dospěli k dobré prediktivní přesnosti. Dle jejich závěrů je možné pro prediktivní modely použít

údaje získané ze zdravotních záznamů pacientů a při zlepšení extrakce dat a zahrnutí více proměnných mohou být dané modely klinicky užitečné a sloužit k lepšímu pochopení trendů onemocnění.

V roce 2011 se Hripesak et al. snažili ve své studii ukázat, že velká heterogenní klinická databáze může odhalit časové vzorce v klinických asociacích. Píší, že elektronické záznamy nejsou přímým odrazem zdravotního stavu pacienta, ale spíše odrazem procesu zdravotnické péče a za pomoci vhodných technik, porozumění a řádného zahrnutí časové složky z nich mohou být odvozeny interpretovatelné asociace.

Jensen et al. se v roce 2014 zabývali časovými trajektoriemi onemocnění za využití dat populačního registru se záznamy o 6,2 milionech pacientů v Dánsku. Ukázali, že velká data ve formě údajů o hospitalizacích pacientů z celé populace mohou být použita k extrakci a seskupení trajektorií, které budou novým způsobem popisovat biologickou progresi onemocnění. Mimo jiné poukazují na to, že stratifikace pacientů podle typu hospitalizací (například akutní, následná, dlouhodobá) je stejně důležitá jako stratifikace podle věku a pohlaví. Získané trajektorie dle nich mají prediktivní potenciál pro progresi onemocnění a dále mohou poskytnout příležitost k provedení systémové analýzy pro identifikaci mezer v diagnostice a léčbě onemocnění.

V roce 2015 Yadav et al. modelovali progresi u pacientů s komplikací diabetu pomocí vlastního DMCI (Diabetes Mellitus Complication Index) skóre odvozeného z elektronických zdravotních záznamů. Věnovali se modelování trajektorií vývoje diabetických komplikací u pacientů s různými diabetickými komplikacemi na počátku léčby. Rozdělili pacienty do několika subpopulací dle jejich výchozích komplikací a dospěli k závěru, že pacienti s různými počátečními komplikacemi mají různá rizika vzniku dalších potíží. Zjistili také, že definované skupiny se liší nejen rizikem, ale také jeho chováním v čase – u některých skupin roste riziko zpočátku vyšší rychlostí a poté se zpomaluje, zatímco u jiných skupin je nárůst rizika v čase konstantní. Rovněž se ukázalo, že trajektorie pacientů se liší dokonce i v rámci jedné skupiny – pacienti s dalšími komplikacemi mají jiná rizika a trajektorie. Výsledky podporují závěr, že podskupiny pacientů se liší podle úrovně závažnosti.

V práci z roku 2016 se Miotto et al. zabývali prediktivním modelováním za využití elektronických zdravotních záznamů pacientů. Představili novou aplikaci metod hlubokého učení k získání prediktivních deskriptorů z administrativních dat. Představená metoda zachycuje hierarchické zákonitosti a závislosti v datech za účelem vytvoření souboru, který lze efektivně využít při prediktivních klinických aplikacích. Aplikace pro predikci onemocnění, která byla ve studii vyhodnocena, by dle nich mohla být použita v řadě klinických úkolů směřujících k personalizované medicíně jako je vyhodnocení individuálních rizik pacientů.

Beck et al. se ve své studii z roku 2016 věnovali trajektoriím multimorbidit<sup>1</sup>, které vedou k mortalitě u pacientů se sepsí. Usilovali o predikci mortality při sepsi za využití elektronických lékařských záznamů pokrývajících všechny hospitalizace dánských nemocnic z let 1996–2014. Celkovým cílem bylo zhodnotit prediktivní hodnotu historie pacientů se sepsí za použití celého spektra předcházejících diagnóz. Píší, že pro tento typ analýz musí být analyzovaná populace dostatečně velká, aby bylo možné obdržet významný popis historie

---

<sup>1</sup> Multimorbidita je přítomnost více chorob u jednoho pacienta.

diagnóz v čase, a proto využili populační registr. Na rozdíl od většiny běžných epidemiologických postupů zaměřených pouze na několik diagnóz použili jako výchozí bod multimorbiditu. Navrhují, že by historie onemocnění pacientů mohla být dalším aspektem personalizované medicíny a dále poukazují na to, že by historie diagnóz pacientů mohla být brána jako stratifikační parametr v klinických studiích. Zdůrazňují však, že ačkoli jsou dané vztahy mezi nemocemi zkoumány i z časového hlediska, je důležité si uvědomit, že časová asociace nemusí nutně znamenat příčinnou souvislost, a nelze prohlásit, že jedna nemoc zapříčiňuje druhou, jelikož může docházet k vlivu zavádějících faktorů.

Chen et al. se v roce 2018 zabývali kauzálními vztahy v oblasti selhání ledvin. Konkrétně šlo o zjištění příčinných rizikových faktorů pro těžké akutní renální selhání. Aplikovali metodu McDSL (Multi-causes Discovery with Structure Learning) na elektronické zdravotní záznamy pacientů. Pomocí tohoto algoritmu byly odhaleny možné rizikové faktory jako příčiny dosažení třetího stadia renálního selhání. Autoři píší, že akutní renální selhání je běžná klinická událost s vysokou morbiditou a mortalitou, přičemž třetí stádium této nemoci je nejzávažnější, a odhalení příčin z elektronických záznamů pacientů je cenné pro podporu klinického výzkumu.

Výchozí studií pro diplomovou práci byl článek z roku 2016, kde se Kannan et al. věnovali extrakci příčinných vztahů z administrativních zdravotních dat a jejich vizualizaci pomocí sítí. Úspěch personalizované medicíny závisí na existenci klinicky praktických nástrojů, které mohou posoudit riziko onemocnění jednotlivce a navrhnout efektivní terapii ve správný čas. Píší, že používání údajů o celé populaci k analýze souvislostí jednotlivých onemocnění se ukázalo jako slibná cesta k extrakci vztahů, které by mohly být využity v klinických studiích, jejichž výsledky by mohly být zakódovány v systémech podpory klinického rozhodování ve zdravotnictví. Poskytli komplexní rámec pro analýzu dat s údaji o chorobách, a to pro velkou skupinu pacientů. Vytvořili systematický postup pro konstrukci orientovaných sítí nemocí a vyvození významných asociací. Vzhledem k vysoké důležitosti této studie pro praktickou část diplomové práce se jí podkapitola 5.3 věnuje blíže.

### 3. Národní zdravotnický informační systém (NZIS)

Národní zdravotnický informační systém je součástí státní statistické služby a díky němu je možno hodnotit celkový zdravotní stav obyvatelstva. Jedná se o jednotný celostátní informační systém veřejné správy, jehož zřizovatelem je Ministerstvo zdravotnictví České republiky (MZ ČR) a správou byl pověřen Ústav zdravotnických informací a statistiky ČR (ÚZIS ČR).

NZIS byl zřízen v roce 1966 a to dle zákona č. 20/1966 Sb., o péči o zdraví lidu, jenž byl později nahrazen zákonem č. 372/2011 Sb., o zdravotních službách a podmínkách jejich poskytování, platným od 1. 4. 2012. Ten již upřesňoval podmínky provozování NZIS.

Od 1. 7. 2016 dále došlo k novelizaci zákona č. 372/2011 Sb., o zdravotních službách a podmínkách jejich poskytování. Novela č. 147/2016 Sb. zpřesňuje povinnosti a pravomoci ÚZIS jakožto správce systému. Její součástí je rovněž vymezení obsahu a funkčnosti Národního registru zdravotnických pracovníků a Národního registru zdravotních služeb hrazených z veřejného zdravotního pojištění a závazek zřízení Národního diabetologického registru a Národního registru intenzivní péče (URL 1: Informace o NZIS).

Mezi součástí NZIS patří (URL 1: Informace o NZIS):

- Program statistických zjišťování Ministerstva zdravotnictví
- Národní zdravotní registry
- Národní registr poskytovatelů
- Národní registr zdravotnických pracovníků
- Národní zdravotní registry vedené podle zákona upravujícího transplantace
- Informační systém infekčních nemocí vedených podle zákona o ochraně veřejného zdraví
- Národní registr hrazených zdravotních služeb
- Další informační systémy

Údaje do systému poskytují jednotlivá zdravotnická zařízení, sbírána jsou data jak o zdravotním stavu obyvatel, tak rovněž o samotných zařízeních a jejich činnosti. Jedná se o sběr formou povinných hlášení do jednotlivých registrů a výkazů, probíhající podle ustanovené metodiky.

Konkrétní úlohy NZIS jsou definovány v již výše zmíněném zákoně č. 372/2011 Sb., o zdravotních službách a podmínkách jejich poskytování (zákon o zdravotních službách) v § 70 odst. 1 a jeho novele č. 147/2016 Sb. s účinností od 1. 7. 2016. Cílem NZIS je shromažďování a zpracování statistických údajů o zdravotním stavu obyvatelstva České republiky, poskytování údajů a statistických informací v rozsahu určeném právními předpisy, a to včetně poskytování informací pro mezinárodní instituce, dále řízení zdravotnictví a tvorba zdravotní politiky, včetně zajištění transparentnosti poskytování a financování zdravotních služeb, zajištění rovného přístupu k zdravotním službám a hodnocení indikátorů kvality a bezpečnosti zdravotních služeb. Neopomenutelnou úlohou NZIS je pak vedení Národních zdravotních registrů, Národního registru poskytovatelů, Národního registru zdravotnických pracovníků a Národního registru hrazených zdravotních služeb.



## Národní zdravotní registry

Národní zdravotní registry vznikly především z iniciativy odborných lékařských společností, které rovněž stanovily jejich odbornou náplň. Účelem národních zdravotních registrů je sběr informací k hodnocení zdravotního stavu obyvatelstva, ke sledování vývoje, příčin a důsledků onemocnění, a to včetně jejich ekonomických a sociálních dopadů. Součástí je také evidence a sledování pacientů s vybranými společensky závažnými chorobami a sledování výskytu, vývoje, příčin a důsledků těchto nemocí i v návaznosti na další péči. Dochází rovněž k evidenci a sledování pacientů s úrazy (URL 2: Národní zdravotní registry).

Mezi úkoly zdravotních registrů se také řadí statistická a vědecká zpracování jejich dat, zaměřená především na analýzy zdravotního stavu obyvatel a kvalitu a využívání zdravotní péče, s cílem zlepšovat zdraví populace. Význam registrů spočívá také v jejich schopnosti monitorovat stávající trendy v kvalitě poskytování zdravotní péče v České republice. Informace z registrů slouží pro databáze zdravotnických ukazatelů Eurostatu, Světové zdravotnické organizace (WHO), Organizace pro hospodářskou spolupráci a rozvoj (OECD) a pro další mezinárodní srovnání (URL 2: Národní zdravotní registry).

Informace ze zdravotních registrů slouží mimo jiné jako podklady pro reporty, ročenky, zdravotnické statistiky a další publikace se zpracovanými analýzami zdravotnických dat. Tyto publikace jsou poté volně dostupné na webových stránkách ÚZIS ČR. Kromě toho je spravován web Regionální zpravodajství NZIS (<http://reporting.uzis.cz>) poskytující online regionální reporting populačního zdraví. Na webu se nachází statistické výstupy navazující na Zdravotnickou ročenku České republiky a krajů a Kardexy vydávané ÚZIS ČR. Rovněž se zde prezentují indikátory zdraví a zdravotní péče a komentované analýzy vybraných témat zdravotnické problematiky v krajích.

Údaje z registrů jsou poskytovány pouze v agregované podobě – například za územní celky, skupiny diagnóz a druhy zdravotnických zařízení. Avšak pro vědecké a výzkumné účely lze poskytnout anonymizovaná neagregovaná data. Při anonymizaci se každému pacientovi jako unikátní identifikátor přidělí hash pomocí aplikace jednosměrné hashovací funkce na původní identifikační údaj (rodné číslo). Unikátní identifikátor pacienta je nezbytný pro kontrolované pořizování záznamů do registrů, sledování průchodu pacienta systémem, vyhodnocování kvality a výsledků péče a pro zdravotnické statistiky (URL 2: Národní zdravotní registry).

Při zpracování údajů v registrech, jejich ochraně a poskytování správce registru a případně zpracovatel důsledně dodržuje zákon č. 101/2000 Sb., o ochraně osobních údajů, ve znění pozdějších předpisů (URL 2: Národní zdravotní registry).

V současné době existuje 12 národních zdravotních registrů zaměřených na určitou část zdravotnického sektoru. Většina z registrů prošla za dobu svého provozu jistým vývojem, který doprovázelo rovněž rozdělování či naopak slučování jednotlivých registrů (URL 2: Národní zdravotní registry). Seznam veškerých registrů včetně informací o jejich struktuře lze nalézt na webu ÚZIS ČR.



### 3.1 Národní registr hospitalizovaných (NRHOSP)

Datový soubor poskytnutý pro praktickou část diplomové práce pochází z Národního registru hospitalizovaných, a proto se kapitola 3.1 věnuje tomuto registru důkladněji, jelikož pochopení struktury dat NRHOSP je pro navazující analýzy klíčové.

Národní registr hospitalizovaných je celoplošný populační registr, jenž slouží k evidenci osob, které byly hospitalizovány na lůžkových odděleních zdravotnických zařízení. Registr navazuje na Informační systém Hospitalizace, který ÚZIS ČR provozoval již od roku 1960. První sběr za všechna lůžková oddělení byl uskutečněn v roce 1981 a pravidelný každoroční sběr dat probíhá od roku 1992 (URL 3: Národní registr hospitalizovaných (NRHOSP)).

Smysl tohoto registru spočívá v získávání informací o zdravotním stavu hospitalizovaných osob. Poskytuje podklady pro hodnocení činnosti zdravotnických zařízení s lůžkovou péčí a výsledné informace z registru se pak dále předávají do databáze Světové zdravotnické organizace i dalším mezinárodním organizacím. Data z tohoto registru jsou také podkladem pro již zmíněné zdravotnické ročenky vydávané ÚZIS ČR – v tematické řadě Zdravotnická statistika vychází od roku 1992 každoročně publikace Hospitalizovaní (URL 3: Národní registr hospitalizovaných (NRHOSP)).

Mezi výhody dat NRHOSP patří, stejně jako u dalších národních zdravotních registrů, pevně definovaná datová struktura a informace o incidenci onemocnění. Značnou limitací je však skutečnost, že registr obsahuje pouze hospitalizační data, nikoli data o ambulanci péči. Tato nevýhoda by však postupem času měla být odstraněna díky novému Národnímu registru hrazených zdravotních služeb (NRHZS).

V NRHOSP jsou evidovány případy všech hospitalizovaných osob na území České republiky, a to včetně cizinců, bezdomovců a narozených dětí. Povinnost hlášení do tohoto registru mají všichni poskytovatelé lůžkové péče (včetně lůžkové péče jednodenní), vyjma zařízení poskytující lázeňskou a rehabilitační péči. Statistickou jednotkou je každá ukončená hospitalizace na jednom oddělení, nehledě na to, zda byl tento pobyt pacienta na oddělení ukončen jeho propuštěním, přeložením na jiné oddělení či do jiného zařízení, nebo úmrtím pacienta (ÚZIS ČR, 2016).

NRHOSP obsahuje restrukturalizovaná administrativní data, která se dělí do tří hlavních celků, tzv. datových vět. Hospitalizační věta zahrnuje údaje o pobytu na jednom oddělení, výkonová věta obsahuje všechny provedené výkony během hospitalizace na jednom oddělení a pobytová věta shrnuje celý pobyt pacienta ve zdravotnickém zařízení (Pavlík et al., 2015).

Data, která jsou v registru zpracovávána, zahrnují údaje nezbytné pro identifikaci daného zdravotnického zařízení, osobní údaje potřebné pro identifikaci pacienta a jeho sociodemografické údaje, které ovlivňují jeho zdravotní stav (věk, pohlaví, zaměstnání, ...), dále se jedná o údaje popisující samotný pobyt pacienta v lůžkovém zařízení – například datum a čas přijetí i propuštění, hlavní a vedlejší diagnózy, operační diagnózy, vnější příčiny nemoci a klasifikace pacienta podle diagnostické skupiny DRG. Hlavní diagnózy jsou diagnostikované stavy primárně zodpovědné za potřebu hospitalizace. Vedlejší diagnózy pak zahrnují především komorbiditu a komplikace, popisují onemocnění, potíže či nálezy existující současně s hlavní diagnózou (URL 3: Národní registr hospitalizovaných (NRHOSP)). Výčet veškerých údajů předávaných do registru je k nalezení na stránkách ÚZIS a dále v práci jako

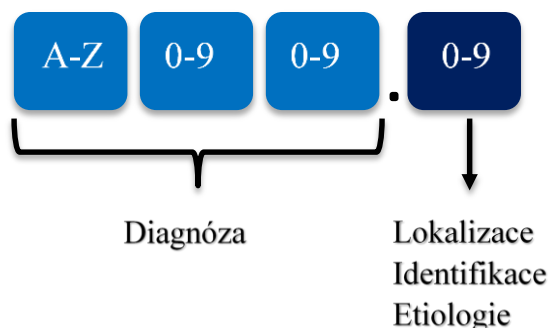
Příloha 1, kde je uveden přehled vybraných údajů vyskytujících se v jednotlivých datových větvích.

### Mezinárodní klasifikace nemocí (MKN)

Mezinárodní klasifikace nemocí je standardní diagnostický nástroj pro epidemiologii a klinické účely. Jde o metodu klasifikace nemocí založenou na převodu diagnóz do alfanumerických kódů. MKN je základem pro identifikaci zdravotních trendů a statistik v celosvětovém měřítku. Jedná se o mezinárodní klasifikační standard pro všechny klinické a výzkumné účely (URL 4: International Classification of Diseases (ICD)). V České republice se aktuálně používá 10. revize, která vešla v platnost v roce 1994 (URL 5: MKN-10).

Klasifikace se dělí do 22 kapitol (Příloha 2) a každá z těchto kapitol obsahuje další hierarchické členění. Základem MKN jsou třímístné kódy, přičemž každý z nich může obsahovat až deset dalších čtyřmístných podpoložek. Ačkoli vykazování na úrovni čtyřmístných položek není povinné pro hlášení na mezinárodní úrovni, doporučuje se tam, kde není položka dále dělena, použít písmeno „X“, aby kódy měly standardní délku pro zpracování. Vykazování na čtyřmístné úrovni je nejobvyklejší, existuje i možnost doplňkového třídění na úrovni pátého nebo dalšího místa (World Health Organization, 2014).

Kód obsahuje na prvním místě písmeno A-Z, další pozice obsahují číslici 0–9, čtvrtý znak následuje po desetinné tečce (World Health Organization, 2014). Struktura kódu na čtyřmístné úrovni je schematicky znázorněna na Obr. 1.



**Obr. 1** Struktura kódu MKN-10

## 4. Network science a teorie grafů

Network science (NS) je věda o sítích, jejíž nástroje mají svůj původ v matematice a teorii grafů a nachází uplatnění v různých oborech – od telekomunikačních sítí přes sociální výzkumy až po biologické a medicínské obory (Barabási, 2012).

Jedná se o nový vědní obor, ačkoli je složité přesně datovat jeho počátek. Zatímco studie sítí má dlouhou historii, především v matematice, teorii grafů a sociologii, NS jako samostatný vědní obor se objevil až v první dekádě 21. století. Klíčová vlastnost NS tkví v tom, že architektura sítí, které vznikají v různých vědních oblastech, si je vzájemně podobná, jelikož se řídí stejnými principy organizace. Jako důsledek této skutečnosti lze využívat společný soubor matematických nástrojů k prozkoumání těchto systémů. Je samozřejmé, že každý obor má jiné cíle, výzvy i technické detaily, které jsou důležité samy o sobě. Přesto společná povaha mnoha problémů vedla k mezioborovému obohacení nástrojů i nápadů. NS tak poskytuje univerzální jazyk, díky němuž obory mohou bez problémů vzájemně komunikovat (Barabási, 2012). Jde o perspektivní rozvíjející se oblast umožňující nový pohled na řešení složitých problémů, které se jinak mohou jevit neprůnikatelně. Zaměřuje se na vysvětlování složitých jevů ve větších měřítkách vyplývajících z jednoduchých principů tvorby síťových vazeb (URL 6: Center for Network Science).

Ačkoli NS úzce souvisí s teorií grafů, z poznatků publikovaných Albertem-Lászlóm Barabásim a jeho týmem plyne, že skutečné sítě se chovají odlišně od tradičních předpokladů teorie grafů. Sítě by měly mít většinu vrcholů s přibližně stejně velkou konektivitou – to se typicky modeluje náhodnými grafy. Avšak moderní síťový výzkum může ukázat, že velké množství vrcholů reálných sítí má velmi malý počet připojení, a naopak existuje několik vrcholů s velkou konektivitou (URL 7: Network science). Zjednodušeně řečeno, NS od teorie grafů odlišuje především zaměření na data, funkci a užitečnost a pro popis vlastností sítí není postačující využití abstraktních pojmů. I přes tyto rozdíly je nutné znát matematický formalismus stojící v pozadí a porozumět mu (Barabási, 2012).

Jak již bylo řečeno, NS nachází uplatnění v široké škále oborů a využití těchto nástrojů tak není ojedinělé ani v biologii, potažmo medicíně. Biologické systémy jsou často reprezentovány jako sítě, za nimiž se skrývají složité soubory vztahů mezi různými entitami. Každá biologická entita musí interagovat s dalšími a tyto interakce probíhají od molekulární úrovně až na úroveň samotných ekosystémů. Tyto skutečnosti vedou k příležitostem modelovat biologii pomocí rozmanitých typů sítí, kterými jsou například sítě ekologické, metabolické, neurologické či molekulární. Například protein může být modelován jako síť aminokyselin a ekosystém lze prezentovat jako síť druhů a interakcí mezi nimi. Díky aplikaci sítí je možné analyzovat biologické entity nejen jako jednotlivé komponenty, ale i jako celé interakční systémy (URL 8: Network analysis in biology).

Aplikace NS na biologii člověka vedla k publikaci zajímavých výsledků a analýza biologických sítí s ohledem na lidská onemocnění vedla k rozvoji další oblasti NS – network medicine.

## Network medicine

Během posledního desetiletí se rozvinula snaha porozumět původu a vzájemnému propojení onemocnění prostřednictvím network science. Jedná se o interdisciplinární obor vyžadující kombinaci specifických výpočetních a biologických znalostí a technik – v podstatě jde o kombinaci systémové biologie a network science, sloužící jako podpora pro identifikaci, prevenci a léčbu nemocí (Korcsmaros et al., 2017). V epidemiologii se pak pomocí těchto metod zkoumá například šíření chorob v populaci. Pojem network medicine vešel v povědomí díky článku „Network Medicine – From Obesity to the 'Diseasome'“ z roku 2007, jehož autorem je Albert-László Barabási. V tomto článku mimo jiné zmiňuje, že není postačující znát seznam genů zapříčiňujících onemocnění, abychom dokázali porozumět mechanismům nemoci, ale měli bychom se pokusit podrobně zmapovat schéma propojení různých buněčných komponent, které jsou danými geny a jejich produkty ovlivněny (Barabási, 2007).

Analýza medicínských dat za pomoci sítí a jejich vizualizace umožnila identifikaci základních molekul a modulů biologických systémů. Rovněž to napomohlo lepšímu pochopení, jak mohou změny buněčných procesů vést ke komplexním onemocněním. Network medicine totiž zahrnuje rozsáhlé síťové analýzy velkých datových souborů popisující různé interakce mezi nemocemi, fenotypy, geny, ale například i přenos a cíle léčiv a jejich možné vedlejší účinky (Korcsmaros et al., 2017). Komplexní nemoci jsou způsobeny kombinací genetických i environmentálních faktorů a například odkrývání molekulárních cest, jejichž prostřednictvím genetické faktory působí na fenotyp, není snadný úkol. Přístupy založené na sítích však představují silné nástroje pro vypořádání se s tímto problémem (Cho et al., 2012). Network medicine v tomto případě nabízí nový pohled na lidskou patogenezi, kdy jsou uvažovány především interakce mezi geny než geny samy o sobě (Chan et Loscalzo, 2012). Léčebné postupy se pak nemusí zaměřovat pouze na určitou genetickou mutaci, ale spíše na biologické cesty, kterými se onemocnění projevuje. Například u pacientů, kteří byli léčeni látkou inhibující konkrétní dráhy, melanomy ustoupily, avšak během několika měsíců se vrátily zpět v plné síle, jelikož se projevíly prostřednictvím jiné dráhy. Porozumění melanomové síti by lékařům mohlo umožnit používat léčiva, která působí na více drahách (Graber, 2014). V tomto ohledu se network medicine zaměřuje na dynamiku síťového propojení a signalizačního systému tak, aby pomohla s identifikací optimálních cílů léčiv (Korcsmaros et al., 2017).

Mezi nejznámější typy sítí patří human disease network (HDN) – síť humánních nemocí, disease gene network (DGN) – síť humánních genů, phenotypic disease network (PDN) – síť humánních fenotypů a human metabolic network (HMN) – síť humánních metabolismů.

Z výše uvedených příkladů plyne, že interakce mohou být reprezentovány pomocí sítí na různých úrovních. Jejich následným propojením pak dochází ke vzniku nových entit, jako je například diseasom,<sup>2</sup> který byl vytvořen propojením sítě humánních nemocí a sítě humánních genů (Agusti et al., 2011).

Není však nutné zacházet až na úroveň genů a jejich interakcí. Pohled na úrovni samotných nemocí/fenotypů prostřednictvím PDN nám rovněž poskytuje zajímavý a užitečný obraz

---

<sup>2</sup> Všechny poruchy a nemoci organismu zobrazeny jako celek s ohledem na genetické rysy

o nemocnosti a chorobách v populaci. Dává nám informaci o vzájemné asociaci chorob zjištěné z medicínské historie pacientů. Podrobně se této oblasti věnuje kapitola 5.

#### 4.1 Teorie grafů

NS úzce souvisí s teorií grafů a navzdory rozdílům mezi nimi je nutné porozumět matematickému pozadí, proto se podkapitola 4.2 věnuje právě teorii grafů.

V matematice pojmem graf rozumíme většinou zobrazení funkční závislosti, avšak v teorii grafů je tento termín vykládán spíše jako objekt popsáný množinou vrcholů a hran (Nešetřil, 1979). Graf je možné si představit jako zjednodušení reálného světa, kdy je zkoumaný problém znázorněn pomocí vrcholů a hran, které je spojují (Kubásek, 2015). Takové grafy mají velkou výhodu ve své názornosti při zobrazení i poměrně komplikovaných vztahů a zároveň jejich relativně snadným vykreslením. Teorie neurčuje, co dané vrcholy a hrany mají reprezentovat – jejich interpretace se definuje až samotným problémem. Pomocí vrcholů a hran mezi nimi tak lze popsat nejrůznější úlohy ze všedního života. Práce na takto obecné rovině s sebou přináší výhodu v možnosti využití obecných poznatků získaných řešením určitého problému pro jiné úlohy (Hordějčuk, 2017). Některé aplikace grafů jsou zcela zjevné a daný problém lze přímočaře přeformulovat na grafovou úlohu a získané řešení opět převést zpět „do reálného života“, jiné úlohy však vyžadují hluboké pochopení daného problému a aplikace grafů dokonce často zůstávají skryté – použitý postup pramení z teorie grafů, ale samotný graf není explicitně použit (Demel, 2002).

**Na začátek definujeme několik vybraných základních pojmů** (Vítečková et al., 2006; Hordějčuk, 2017):

- **Graf** – systém, který lze zakreslit v rovině pomocí bodů (vrcholů) a spojníc mezi nimi (hran)
- **Cesta** – posloupnost orientovaných hran, kdy mají následující hrany počátek v tom vrcholu, v němž končí hrana přechodí
- **Smyčka** (cyklus) – cesta, která začíná i končí ve stejném vrcholu
- **Rovnoběžné hrany** – hrany, které spojují stejnou dvojici vrcholů
- **Násobné hrany** – případ, kdy jsou stejné vrcholy spojené větším množstvím hran
- **Stupeň vrcholu** – počet hran spojených s vrcholem

#### Klasifikace grafů dle orientace hran

Podle orientace hran se grafy rozlišují na orientované a neorientované. Toto dělení je pro práci zásadní a níže jsou uvedeny a vysvětleny definice pro oba typy grafů.

Neorientované grafy se využívají v případech, kdy je orientace hran nepodstatná a není potřebné rozlišovat mezi počátečními a koncovými vrcholy (Demel, 2002). Jistou formou neorientovaného grafu je v základu i síť humánních fenotypů bez zahrnutí kauzální informace (kapitola 5).

#### **Definice 4.1** (Neorientovaný graf)

Nechť  $V, E$  jsou libovolné disjunktní množiny a  $\varepsilon: E \mapsto V \otimes V$  zobrazení. Neorientovaným grafem nazýváme uspořádanou trojici  $G = \langle V, E, \varepsilon \rangle$ , prvky množiny  $E$  jsou hranami grafu  $G$ , prvky množiny  $V$  vrcholy grafu  $G$  a zobrazení  $\varepsilon$  incidencí grafu  $G$  (Kolář, 2000; Demel, 2002).

Incidence grafu  $\varepsilon$  přiřazuje každé hraně neuspořádanou dvojici vrcholů. Pokud  $\varepsilon(e) = [u, v]$ , nazýváme  $u, v$  krajními vrcholy hrany  $e$ . Takové vrcholy jsou s hranou  $e$  incidentní, jinými slovy říkáme, že je hrana  $e$  spojuje. Je-li hrana incidentní pouze s jedním vrcholem, tedy pokud  $\varepsilon(e) = [v, v]$ , tak hranu  $e$  nazýváme smyčkou. Existuje-li vrchol, se kterým není incidentní žádná hrana, nazýváme ho izolovaným (Kolář, 2000; Demel, 2002).

U velkého množství úloh je žádoucí určit, ve kterém vrcholu hrana začíná a kde končí, tedy je podstatná orientace hran. V tom případě se využívá orientovaný graf. Při jeho znázornění se u hran kreslí šipky, aby bylo zřejmé, jakým způsobem jsou vrcholy uspořádány (Habala 2012, Hordějčuk 2017). Jako orientovaný graf můžeme vnímat i síť humánních fenotypů obohacenou o kauzální informaci, je-li naším cílem zjistit nejen jaké nemoci se vyskytují společně, ale rovněž příčinné vztahy mezi nimi (kapitola 5).

**Definice 4.2** (Orientovaný graf)

Nechť  $V$  a  $E$  jsou libovolné disjunktní množiny a  $\varepsilon: E \mapsto V \times V$  zobrazení. Orientovaným grafem nazveme uspořádanou trojici  $G = \langle V, E, \varepsilon \rangle$ , prvky množiny  $E$  nazýváme orientovanými hranami grafu  $G$ , prvky množiny  $V$  vrcholy grafu  $G$  a zobrazení  $\varepsilon$  incidencí grafu  $G$  (Kolář, 2000; Demel, 2002).

Stejně jako u neorientovaného grafu přiřazuje incidence  $\varepsilon$  hranám dvojici vrcholů, tentokrát však uspořádanou. Jestliže pro  $e \in E$  je  $\varepsilon(e) = (u, v)$ , nazýváme vrchol  $u$  počátečním a  $v$  koncovým vrcholem hrany  $e$ . Také říkáme, že hrana  $e$  je orientována od vrcholu  $u$  k vrcholu  $v$  (Kolář, 2000; Demel, 2002).

Na Obr. 2 jsou znázorněny jednoduché příklady neorientovaného a orientovaného grafu.

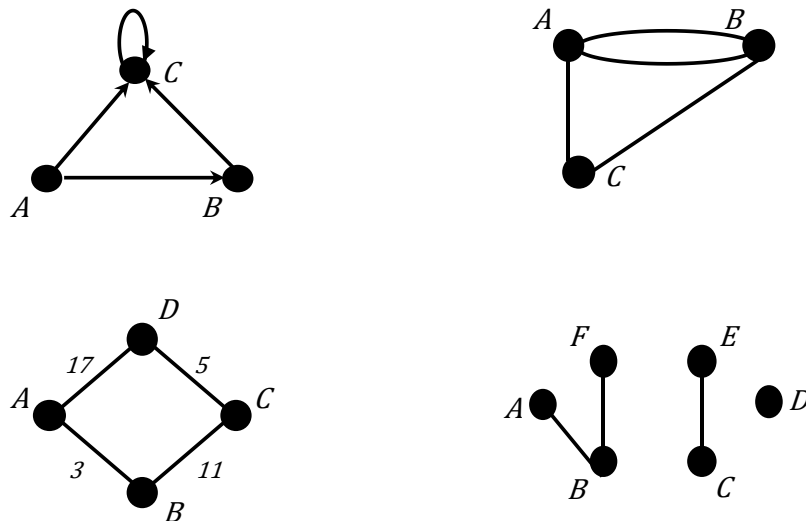


**Obr. 2** Příklad neorientovaného (vlevo) a orientovaného (vpravo) grafu

Grafy lze samozřejmě členit i dle jiných kritérií, než je orientace hran. Podle přítomnosti či nepřítomnosti smyček a rovnoběžných hran se rozlišují multigrafy, prosté grafy a jednoduché (obyčejné) grafy. Multigrafy mohou obsahovat jak smyčky, tak rovnoběžné hrany. Prosté grafy neobsahují rovnoběžné hrany, ale mohou se u nich vyskytnout smyčky. Jednoduché grafy nemají smyčky ani rovnoběžné hrany. Podle počtu vrcholů a hran pak rozlišujeme grafy prázdné (prázdná množina vrcholů a hran), konečné (konečná množina vrcholů a hran) a nekonečné (nekonečná množina vrcholů a hran). Dále můžeme mít grafy souvislé a nesouvislé. Souvislý graf je takový graf, pro který platí, že pro všechny dvojice vrcholů existuje alespoň jedna cesta, která je spojuje. V případě nesouvislého grafu naopak platí, že

neexistuje alespoň jedna cesta mezi všemi dvojicemi vrcholů. Grafy můžeme rovněž dělit dle uspořádání jejich vrcholů a hran. Pokud nastane situace, kdy lze vrcholy rozlišit na dvě disjunktí množiny, přičemž hrany spojují pouze vrcholy náležící do různých tříd, mluvíme o bipartitním grafu. Analogicky lze obecně definovat  $k$ -partitní grafy. Podle množství vrcholů a hran, respektive poměru počtu hran k vrcholům, lze grafy dělit na řídké a husté. Řídké grafy obsahují relativně málo hran vzhledem k počtu vrcholů, husté grafy naopak mají hran hodně. Toto dělení je však neurčitě a většinou se nevyužívá. Naopak jedním ze základních a často používaných ukazatelů je v tomto ohledu stupeň vrcholu podávající informaci o počtu hran incidujících s daným vrcholem. V případě neorientovaného grafu je stupeň vrcholu jednoduše počet hran do něj zasahujících. Pro orientované grafy se rozlišuje stupeň vrcholu na vstupní (počet hran orientovaných směrem do vrcholu) a výstupní (počet hran orientovaných směrem z vrcholu). Platí, že součet stupňů vrcholů je vždy sudý a rovný dvojnásobku počtu hran. Sčítáme-li totiž stupně všech vrcholů grafu, počítáme vlastně, kolik existuje konců hran. Každá hrana má dva konce, proto se výsledný počet jejich konců rovná dvojnásobku počtu všech hran (Vítečková et al., 2006; Foltýnek et Dannhoferová, 2011; Hordějčuk, 2017).

Na Obr. 3 jsou ilustrovány příklady grafů a jejich vlastností.



**Obr. 3** Orientovaný prostý graf se smyčkou v bodě C (vlevo nahoře), neorientovaný multigraf s násobnou hranou mezi body A a B (vpravo nahoře), jednoduchý graf s ohodnocenými hranami (vlevo dole), nesouvislý graf s izolovaným vrcholem D (vpravo dole)

Ohodnocení grafu je další důležitou vlastností. V případě ohodnocených grafů existuje reálná funkce definovaná na množině všech hran, přiřazující všem hranám nějakou hodnotu. Ohodnoceny mohou být kromě hran i vrcholy grafu. Přiřazení hodnot hranám, respektive vrcholům grafu je žádoucí především v praktických aplikacích. Nejčastěji jde o ohodnocení číselné, možné je však i užití textu či podmínek, kdy lze danou hranu/vrchol využít. Formálně zapsáno (Foltýnek et Dannhoferová, 2011):

**Definice 4.3** (Hranové ohodnocení grafu)

Hranové ohodnocení grafu je zobrazení  $d: E \mapsto R^+$  (někdy připouštíme i 0 a záporné hodnoty). Číslo  $d(e)$  nazýváme zpravidla délkou hrany  $e$ , také hovoříme o vzdálenosti vrcholů.

**Definice 4.4** (Vrcholové ohodnocení grafu)

Vrcholové ohodnocení grafu je zobrazení  $k: V \mapsto R$ . Číslo  $k(v)$  nazýváme klíčem vrcholu  $v$ , také hovoříme o hodnotě nesené vrcholem  $v$ .

Pro práci je rovněž vhodné zmínit termín podgraf, který v teorii grafů můžeme chápat jako jistou obdobu pojmu podmnožina. Zmínit jej je vhodné z hlediska velkého počtu nemocí (vrcholů) a především asociací (hran) mezi nimi – při zobrazení asociací mezi nemocemi tak zpravidla nevykreslujeme všechny vazby, ale pouze relevantní vztahy o určité síle, a stejně tak nemusí dojít k vykreslení málo četných onemocnění (nebo skupin onemocnění).

Podgraf vznikne vynecháním některých vrcholů a hran původního grafu. Důležité je, že podgraf musí být také grafem, tedy spolu s každou hranou v podgrafu se musí vyskytovat i oba vrcholy, které spojuje. Každý graf je zároveň sám sobě podgrafem (Demel, 2002). Formálně zapsáno (Foltýnek et Dannhoferová, 2011):

**Definice 4.5** (Podgraf)

Řekneme, že graf  $G' = \langle V', E', \varepsilon' \rangle$  je podgrafem grafu  $G = \langle V, E, \varepsilon \rangle$ , jestliže  $V' \subseteq V$ ,  $E' \subseteq E$  a zobrazení  $\varepsilon'$  je zúžením zobrazení  $\varepsilon$  na množinu  $E$ .

Rozlišujeme také speciální druhy podgrafů, mezi jeden ze základních patří faktor grafu. Faktor vznikne vynecháním pouze některých hran při současném ponechání všech vrcholů. Jako vedlejší důsledek může docházet ke vzniku izolovaných vrcholů, které incidovaly s vypuštěnými hranami (Demel, 2002). Formálně zapsáno (Foltýnek et Dannhoferová, 2011):

**Definice 4.6** (Faktor grafu)

Nechť  $G = \langle V, E, \varepsilon \rangle$  je graf. Řekneme, že graf  $G' = \langle V', E', \varepsilon' \rangle$  je faktorem grafu  $G$ , jestliže  $V' = V$ ,  $E' \subseteq E$  a zobrazení  $\varepsilon'$  je zúžením zobrazení  $\varepsilon$  na množinu  $E$ .

Sít' humánních fenotypů v provedení prezentovaném v kapitole 5 lze přirovnat k obyčejnému konečnému nesouvislému grafu s ohodnocenými hranami.



## 5. Síť humánních fenotypů a kauzalita

Zdravotní stav pacientů se typicky nevyznačuje pouze jedním onemocněním, ale často je charakterizován několika přidruženými nemocemi. Objevování souvislostí mezi nimi pak může vést k zajímavým poznatkům, či dokonce k odpovědím na různé klinické nebo epidemiologické otázky. Přidáme-li navíc ohodnocení o kauzální informaci, můžeme získat informaci o progresi onemocnění z hlediska přidružených chorob. Identifikace kauzálních vztahů mezi diagnózami by mohla být rovněž nápomocná při rozvoji rozhodovacích procesů.

### 5.1 Síť humánních fenotypů

Během posledních let bylo vydáno několik článků, jejichž cílem bylo porozumět vzájemnému propojení i původu nemocí za pomoci analytických nástrojů network science. Zkoumány však byly především genové interakce. Pohled na souvislost nemocí z hlediska fenotypové informace přišel až o něco později, a to i přesto, že pro zkoumání souvislostí z pohledu fenotypů jsou k dispozici rozsáhlé a často velmi kvalitní údaje. Základem jsou především administrativní data.

Tento datový zdroj má, i přes svá omezení, velký potenciál a nese s sebou několik výhod. Jedná se o data populačního charakteru obsahující velké množství informací o pacientech, především o jejich diagnózách, hospitalizacích i provedených výkonech. Nespornou výhodou administrativních dat je často také jejich dostupnost z dlouhodobého časového hlediska. Jako každý datový zdroj mají samozřejmě i tato data své nevýhody. Mezi největší mínusy patří závislost na správném vykazování a rozhodování při kódování podle mezinárodní klasifikace, přičemž výjimkou nejsou ani typografické chyby či chybějící údaje. Také nízká klinická specifita administrativních dat může být pro některé účely značně omezující (Ferver et al., 2009; Riley, 2009, Schulman et al., 2013). Pro zjištění asociací, potažmo sestavení sítě humánních fenotypů, jsou však administrativní data ideální, a to především díky jejich populačnímu charakteru, dobré dostupnosti v elektronické podobě a jelikož jsou data vykazována povinně, nepojí se s jejich sběrem ani žádné další náklady.

Síť humánních fenotypů, zkráceně PDN (z anglického Phenotypic disease network) může být chápána jako prostředek pro vizualizaci vztahů mezi nemocemi. PDN zkonstruovali a v roce 2009 představili Hidalgo et al. v článku „A Dynamic Network Approach for the Study of Human Phenotypes“.

Síť lze chápat jako mapu prostoru diagnóz, přičemž její struktura nám může být nápomocná při snaze pochopit souvislosti mezi jednotlivými chorobami. Obecně bychom v tomto případě mohli říci, že dvě nemoci jsou vzájemně propojeny, vyskytují-li se společně u významného množství pacientů v populaci. Velkou výhodou této metody je především možnost přehledné vizualizace daných vztahů.

Hidalgo et al. také tvrdí, že pacienti, kterým bylo diagnostikováno onemocnění vyznačující se vyšší mírou propojenosti než jiná onemocnění, vykazují větší riziko úmrtí oproti pacientům s nemocemi, které jsou v síti propojeny méně. Síť může také pomoci odhalit souvislost mezi nemocemi a jejich progresí, protože u pacientů existuje tendence k rozvoji dalšího onemocnění, které je v síti blízké tomu, jež jim bylo diagnostikováno dříve.

V případě obyčejné, neorientované PDN se jedná o síť založenou na výpočtu asociační matice z administrativních záznamů pacientů. Pro výpočet matice lze použít vhodné asociační koeficienty (Groza et al., 2015; Hoehndorf et al., 2015; Zanzoni et al., 2015), je však možné využít například i kosinovou podobnost (Roque et al., 2011) nebo Pearsonův korelační koeficient pro binární data či relativní rizika (Hidalgo et al., 2009).

Možností výpočtu vstupní matice je více, vždy je však třeba se zamyslet, zda je daný postup vhodný pro použitá data a případně jestli není nutné provést transformaci dat.

Kosinová podobnost je vhodná pro data v textové podobě, tedy v případě, kdy pacientovu anamnézu charakterizuje například vektor diagnóz zapsaných v MKN klasifikaci. Jedná se o skalární součin, tedy o součin vektorů vydělených jejich velikostmi, tedy platí:

$$S_{cos}(x_1, x_2) = \frac{x_1^T \cdot x_2}{\|x_1\| \cdot \|x_2\|} \quad (5.1)$$

kde  $\|x_1\|$  resp.  $\|x_2\|$  jsou délky vektorů  $x_1$  a  $x_2$ . Hodnoty  $S_{cos}(x_1, x_2)$  jsou pak rovny kosinu úhlu mezi oběma vektory (Janoušová et al., 2015).

Naopak v případě, kdy pacientovu historii reprezentuje vektor nul a jedniček, který určuje výskyt či nevýskyt nemocí v pacientově anamnéze, je vhodné využít například asymetrický Jaccardův koeficient.

Hodnoty binárních proměnných lze sumarizovat do frekvenční tabulky (Tab. 1).

**Tab. 1** Frekvenční tabulka

		x <sub>2</sub>		
		1	0	
x <sub>1</sub>	1	<i>a</i>	<i>b</i>	<i>a+b</i>
	0	<i>c</i>	<i>d</i>	<i>c+d</i>
		<i>a+c</i>	<i>b+d</i>	<i>p</i>

kde *a* je počet proměnných, které nabývají pro oba objekty hodnotu 1 (společný výskyt), *d* představuje počet proměnných, jež nabývají pro oba objekty hodnotu 0 (společná absence), *b* a *c* značí počet neshodujících se proměnných. Zároveň platí, že  $p = a + b + c + d$  (Haruštiaková et al., 2012).

Jaccardův koeficient lze poté zapsat následovně:

$$S_{jac}(x_1, x_2) = \frac{a}{a + b + c} \quad (5.2)$$

Maximální podobnosti Jaccardův koeficient nabývá, jestliže mají dva objekty identické hodnoty, tedy v případě  $b = c = 0$  (Haruštiaková et al., 2012). Nespornou výhodou tohoto koeficientu, pro zkoumání podobnosti založené na výskytu onemocnění v rámci pacientů, je jeho schopnost vypořádat se s problémem dvou nul. Jinými slovy řečeno, k výsledné podobnosti přispívají pouze společné výskyty a společné nevyskytování chorob není bráno v potaz.

V případě binárního vektoru se nabízí i zmíněná varianta Pearsonovy korelace pro binární data ( $\Phi$ -korelace). Matematicky ji lze zapsat následovně:

$$\Phi = \frac{ad - bc}{\sqrt{(a + b)(a + c)(b + d)(c + d)}} \quad (5.3)$$

Zde už je ale třeba uvažovat nad transformací. Pro tuto metodu totiž nastává problém, pokud mají sledovaná onemocnění výrazněji odlišné prevalence, jelikož rozdělení proměnných ovlivňuje výsledek korelace (Wuensch, 2009, Hidalgo 2009, Legendre et Legendre, 1998). Metoda se rovněž ukázala být citlivá na problém dvou nul.

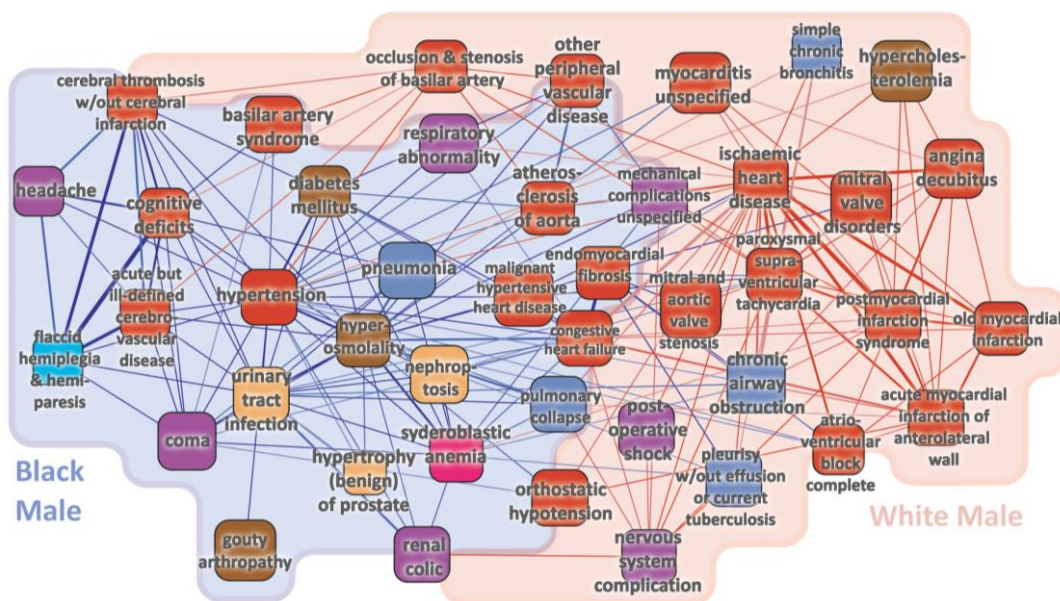
Další zmíněnou možností jsou relativní rizika. Výpočet relativních rizik je založen na srovnání pravděpodobností výskytu sledovaného jevu ve dvou skupinách relativním způsobem. Relativní riziko lze definovat vztahem:

$$RR = \frac{P_{exp}}{P_{neexp}} = \frac{\frac{a}{a + c}}{\frac{b}{b + d}} \quad (5.4)$$

Výsledkem je pak číslo, které vyjadřuje, kolikrát je ve skupině vystavené rizikovému faktoru ( $P_{exp}$ ) vyšší pravděpodobnost výskytu jevu než v kontrolní skupině ( $P_{neexp}$ ) (Pavlík, 2012).

Co se struktury sítě týče, vrcholy PDN představují konkrétní nemoc, případně nadřazenou skupinu jako například oddíly či kapitoly MKN-10 a hrany mezi nimi znázorňují jejich vztahy. Jako vhodnou aditivní informaci je možné do sítě zanést informaci o síle asociací (hranové ohodnocení) a četnosti asociace dané nemoci s ostatními (stupeň vrcholu). Co se týče vrcholového ohodnocení grafu, může být někdy přínosnější místo četnosti asociace nemoci s ostatními zanést spíše informaci o četnosti nemoci ve sledované populaci.

Na Obr 4. lze vidět příklad PDN, jak ji zkonstruovali Hidalgo et al. v roce 2009. Vrcholy zastupují skupiny onemocnění, síla spojnic mezi nimi odráží míru asociace. Jedná se o podgraf, přičemž jsou vyobrazeny spojitosti k hypertenzi a ischemické chorobě srdeční.



**Obr. 4** Síť ľudských fenotypů (Hidalgo et al., 2009; upraveno)

## 5.2 Kauzalita

Lze říci, že dvě proměnné jsou spolu asociovány (korelovány), jestliže změna jedné proměnné zapříčiní změnu proměnné druhé. Asociaci nazveme kauzálním (příčinným) vztahem, jestliže lze dokázat, že změna nezávislé proměnné (expozice rizikovému faktoru) vede ke změně závislé proměnné (efektu) (World Health Organization, 2001). Jinými slovy řečeno, kauzalita je příčinný vztah mezi expozicí rizikovému faktoru a zdravotním následkem. Při prokazování kauzality se počítá s vyloučením náhody, biasu<sup>3</sup> a zavádějících faktorů (skutečná příčina zkoumaného výsledku spojená s domnělým rizikovým faktorem). Při zvažování příčinného vztahu se bere ohled na sílu asociace, časovou souvislost, specificitu, vztah dávky a účinku a případné ověření experimentem. Absolutním požadavkem je však časová souslednost, další podmínky jsou již relativní (Šejda et al., 2005).

Možné typy proměnných v kauzálních vztazích (World Health Organization, 2001):

- Závislé a nezávislé proměnné – definuje se nezávislá proměnná (rizikový faktor) a závislá proměnná (efekt), definice proměnných závisí na hypotéze studie
- Zavádějící proměnné – nezávislá proměnná (jiná než zkoumaný rizikový faktor) mající vliv na závislou proměnnou, systematicky korelována s rizikovým faktorem
- Intermediární proměnné – proměnné zprostředkovávající efekt, v řetězci kauzality leží mezi původním rizikovým faktorem a závislou proměnnou
- Modifikace efektu – některé nezávislé proměnné mohou ovlivňovat (pozitivně i negativně) vliv rizikového faktoru, přičemž některé zavádějící proměnné mohou rovněž modifikovat výsledný efekt

Kauzální vztahy lze dělit následovně (World Health Organization, 2001):

- Přímé – nezávislá proměnná mění závislou proměnnou přímo, bez zásahu dalších proměnných
- Nepřímé – nezávislá proměnná způsobuje změny závislé proměnné prostřednictvím jiných proměnných či faktorů
- Interakce – existence interakce více nezávislých proměnných, vedoucí ke změně závislé proměnné
  - Podmíněná interakce – rizikové faktory jsou schopné způsobit změnu v závislé proměnné jen tehdy, když působí společně

Pro klinické rozhodovací procesy má kauzalita význam například pro diagnostiku, léčbu nebo prevenci. V epidemiologii se navíc soustředí i na zjištění příčinných vztahů, které nelze ovlivnit nyní, ale předpokládá se, že bude možné jejich úspěšné ovlivnění v budoucnosti, a to například imunizací či léčbou (Janout, 1998).

Důležité je uvědomit si rozdíl mezi asociací a kauzalitou – asociace je nutným, ale nikoli postačujícím požadavkem kauzality. Jinými slovy, asociace kauzalitu neimplikuje a jestliže jsou spolu dvě proměnné asociovány, nutně mezi nimi nemusí existovat příčinný vztah. Jak je

---

<sup>3</sup> Bias je zkreslení výsledků, systematická chyba v odhadu pozorovaného efektu, která vznikla jako důsledek chyby při plánování nebo provedení epidemiologické studie.

napsáno výše, mezi kritéria pro stanovení kauzality patří vyloučení biasu, náhodné chyby a zavádějících faktorů. Základní kritéria pro posuzování kauzality navrhl v roce 1965 britský statistik Hill, přičemž se tato pravidla v různých modifikacích používají dodnes (Janout, 1998). Patří mezi ně (Hill, 1965; Janout 1998; Bonita et al., 2006):

- Časová souslednost – přirozeně příčina předchází následku. Hodnocení časové souslednosti může být problematické u studií, jež hodnotí zároveň expozici i následek.
- Síla asociace – čím silnější je asociace mezi nezávislou a závislou proměnnou, tím vyšší je pravděpodobnost kauzálního vztahu.
- Konzistence – čím větší existuje shoda mezi výsledky různých studií provedených různými lidmi na různých místech, tím vyšší je pravděpodobnost kauzální asociace.
- Specificita – čím menší počet příčin stačí k vysvětlení následku, tím lépe. Tato podmínka bývá splněna většinou u infekčních onemocnění, naopak chronické choroby ji většinou nesplňují. O kauzální asociaci se může jednat i v případě, že tato podmínka splněna není.
- Reverzibilita asociace – pokud se odstraní, případně sníží expozice podezíraného faktoru, měla by se snížit frekvence následku.
- Biologická věrohodnost – existence teoretického biologického mechanismu, který dokáže vysvětlit pozorovaný vztah. Je potřeba brát ohled na to, že stav lékařské vědy v určitém čase ovlivňuje, co je bráno za věrohodné.
- Vztah dávky a účinku – vyjadřuje se tehdy, když různá dávka expozice způsobuje různou míru následku. Je-li tento vztah prokázán, podporuje kauzalitu pozorovaného vztahu. Čím větší míra expozice, tím větší míra následků.
- Analogie – existence analogického příkladu podobné expozice nebo nemoci kauzalitu podporuje. Obecně se analogie pokládá za slabý doklad kauzality.

Sám Hill však varuje před bezmyšlenkovitým užíváním jím navržených kritérií. Žádné z nich nemůže být samo o sobě důkazem kauzality a zároveň nepřítomnost kteréhokoliv z nich kauzální vztah nevylučuje (Hill, 1965).

### 5.3 Kauzalita v síti humánních fenotypů

Jak již bylo předesláno dříve, identifikace kauzální vztahů v PDN může pomoci pochopení výskytu chorob. K důvodům pro zájem o zjištění či naopak vyloučení kauzálního vztahu v epidemiologii patří snaha pochopit determinanty výskytu onemocnění a jejich šíření, možnost identifikovat příslušné vazby v řetězci kauzality, kde je možnost intervence, a také propojit výstup a dopad intervenčních programů na vstup (World Health Organization, 2001). Údaje o poskytnuté zdravotní péči tak mají mimo jiné i jistý potenciál pro možné rozhodovací procesy ve zdravotnictví. Navzdory skutečnosti, že se jedná o ambici, která vede řadu stávajících snah v oblasti translačního výzkumu, se ukázalo být velmi obtížné realizovat ji v praxi (Kannan et al., 2016).

Identifikovat kauzální vztahy v síti kvůli samotnému objemu a složitosti údajů není triviální a nepřispívá tomu ani fakt, že některé diagnózy jsou si velmi blízké, avšak zaznamenány

odlišně. I z tohoto důvodu bude v praktické části výsledná síť konstruována primárně na úrovni oddílů MKN-10 namísto jednotlivých diagnóz.

Extrakcí kauzálních vztahů z PDN se zabývali Kannan et al. v článku „Conditional Disease Development extracted from Longitudinal Health Care Cohort Data using Layered Network Construction“, který byl publikován v roce 2016. Autoři představili metodiku pro odvození vztahů podmíněného rozvoje onemocnění a vlastní postup výpočtu vztahů pro síť, který umožňuje identifikaci orientovaných vztahů.

Hypotéza kauzální nezávislosti mezi dvěma diagnózami je rovnost  $n_{i \rightarrow j}$  a  $n_{j \rightarrow i}$ , tedy počet případů, kdy nemoc  $i$  předcházela nemoci  $j$  a naopak. Zamítnutí této nulové hypotézy se však nijak nezaměřuje na sílu kauzální asociace. Skutečným cílem je zjistit domnělé příčinné vztahy mezi dvojicemi onemocnění  $i$  a  $j$  na základě toho, jak často choroba  $j$  následuje po chorobě  $i$ . Tedy testování, s jakou pravděpodobností se nemoc  $j$  rozvine po přechozí diagnóze choroby  $i$ . Tj.:

$$\begin{aligned} P(\eta_j(\tau > t) = 1 | \eta_i(t) = 1; t; \tau \in [0, T]) \\ = \frac{P(\eta_j(\tau > t) = 1, \eta_i(t) = 1 | t; \tau \in [0, T])}{P(\eta_i(t) = 1 | t \in [0, T])} \\ \approx \frac{n_{i \rightarrow j}}{n_i} \end{aligned} \quad (5.5)$$

kde  $\eta_\kappa(t)$  je 0 nebo 1 v závislosti na tom, zda nemoc  $\kappa$  byla diagnostikována v čase  $t$  nebo někdy dříve. V případě, kdy by daná pravděpodobnost byla stejná jako u výskytu v celé populaci, poměr by byl úměrný:

$$\frac{n_{i \rightarrow j}}{n_i} \propto \frac{n_j}{N} \Rightarrow \frac{n_{i \rightarrow j}}{n_{j \rightarrow i}} = 1 \quad (5.6)$$

což dává přesně nulovou hypotézu. Při předpokladu, že existuje kauzální efekt,  $n_{i \rightarrow j}/n_i$  je přímým ukazatelem síly daného efektu. Při korekci na prevalenci onemocnění v populaci získáme výsledný vztah

$$f_{i \rightarrow j} = \frac{n_{i \rightarrow j}}{n_i} - \frac{n_j}{2N} \quad (5.7)$$

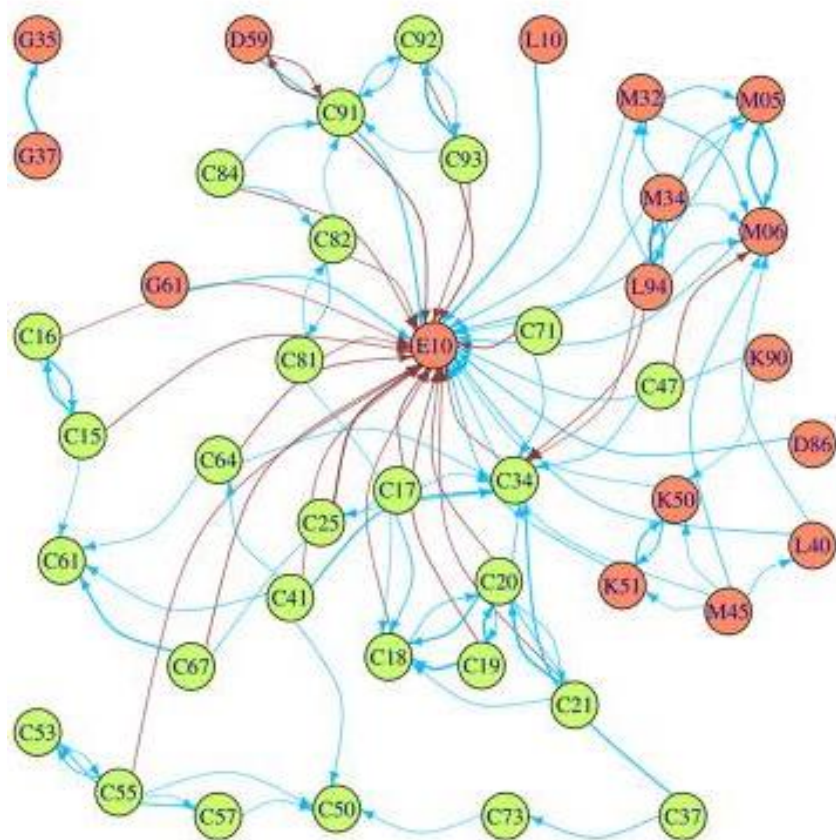
kde  $n_{i \rightarrow j}$  je počet osob s nemocí  $i$ , kterým byla později diagnostikována nemoc  $j$ ,  $n_i$  a  $n_j$  jsou prevalence nemocí  $i$  a respektive  $j$ ,  $N$  je pak velikost zkoumané populace. Tento vzorec je nazýván CIF (Causal information fraction).

I když je výsledná míra asociace silná, nemusí to nutně znamenat, že jedna nemoc zapříčiňuje druhou. Je také možné, že v pozadí existují společné příčiny obou chorob, které se projeví dřívejší manifestací jedné z nich (Kannan et al., 2016).



Ve své studii Kannan et al. zahájili analýzu na úrovni všech tříznakových MKN-10 kódů a následně sdružili podobná onemocnění do jednoho uzlu, čímž vytvořili nižší dimenzionální vrstvenou síť. Hrany obsahující významné nové informace pak dále kontrolovali rozbalením do bipartitně orientované sítě, což vedlo k předpokládaným kauzálním informacím na úrovni tříkódových diagnóz.

Na Obr. 5 je znázorněna orientovaná PDN ze zmíněného článku, zaměřena na orientované vztahy mezi rakovinou a autoimunitními chorobami na úrovni třímístných kódů MKN-10. Červené vrcholy představují autoimunitní onemocnění a zelené vrcholy nádorová onemocnění. Modré hrany spojují nemoci stejné kategorie, hnědé potom spojují onemocnění z různých kategorií. Tloušťka spojnice odráží velikost CIF indexu.



**Obr. 5** Síť orientovaných vztahů zaměřená na souvislosti výskytu rakoviny a autoimunitních chorob (Kannan et al., 20016; upraveno)



## 6. Praktická část

Praktická část práce je zaměřena na hledání a následnou vizualizaci netriviálních orientovaných vztahů v síti humánních fenotypů. Data NRHOSP poskytnutá pro praktickou část byla zpracována v souladu s platnou legislativou jako anonymizované záznamy.

### 6.1 Použitý software

Pro zpracování a analýzu dat byl primárně využit program IBM SPSS Statistics 24. V tomto programu proběhly veškeré kroky analýzy od přípravy dat přes restrukturalizaci a tvorbu kombinací až po zjištění orientovaných vztahů. Následná vizualizace zjištěných vztahů byla provedena v softwaru R – The R Project for Statistical Computing (R Core Team, 2018), a to především za využití knihovny igraph (Csardi et Nepusz, 2006).

### 6.2 Datový soubor

Jak již bylo zmíněno dříve, podkladem pro praktickou část práce jsou data z NRHOSP. Soubor obsahuje data o 22,1 milionech hospitalizací za období 2007–2016, které nastaly u celkového počtu 7,6 milionů unikátních pacientů.

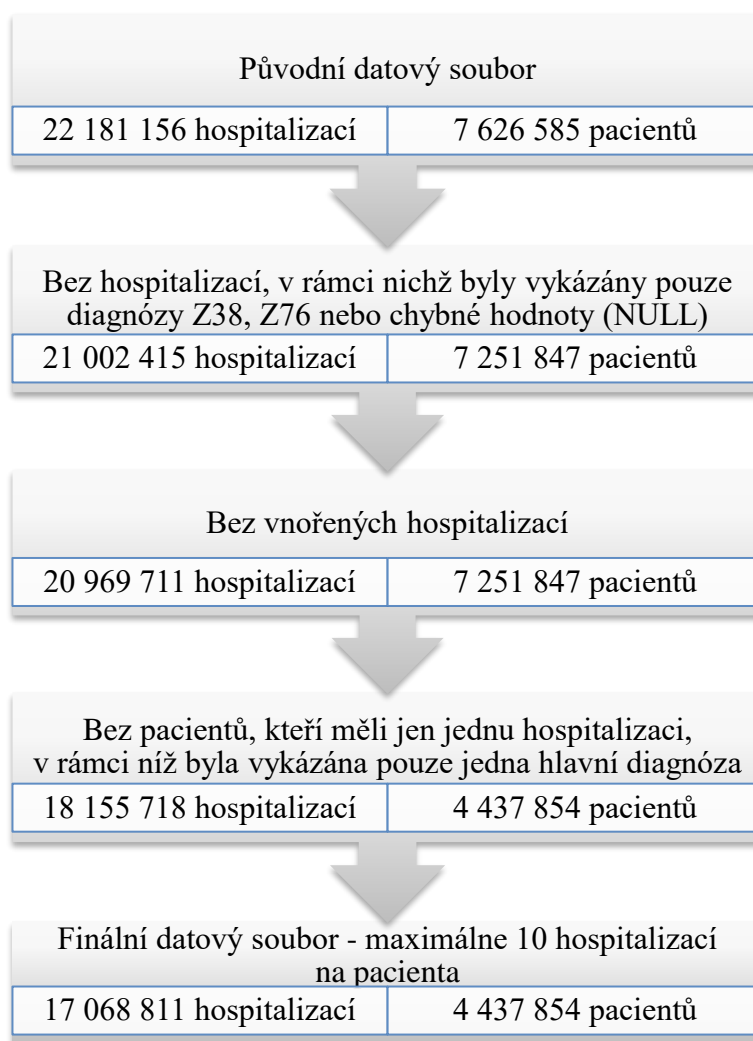
#### 6.2.1 Příprava datového souboru

Před samotnou analýzou bylo žádoucí datový soubor připravit a očistit od nepotřebných informací, a to především vzhledem k jeho velikosti a předpokladu jejího dalšího nárůstu při tvorbě kombinací. Kompletní datový soubor obsahoval 132 proměnných, z nichž však většina byla pro práci nadbytečná. Došlo tedy k jejich ořezání na 14, přičemž byly ponechány nezbytné informace týkající se diagnóz – 5 hlavních, 5 vedlejších, 1 výsledná. Operační diagnózy a vnější příčiny nemocnosti a úmrtnosti se dále neuvažovaly. Ve výsledném souboru zůstal také identifikátor pacienta, přičemž původní textový hash byl kvůli zmenšení celkové velikosti souboru překódován na číselný identifikátor. Ponechány byly rovněž informace pro určení základních charakteristik pacienta – věk a pohlaví. Aby se předešlo duplicitnímu výskytu diagnóz v rámci jednoho hospitalizačního pobytu, byla každá diagnóza pro jeden pobyt uvažována maximálně jedenkrát, přičemž prioritu měly postupně první až pátá hlavní diagnóza a následně první až pátá vedlejší diagnóza. Stejně tak výsledná diagnóza se brala v potaz pouze v případě, pokud se již nevyskytovala mezi hlavními.

Co se týče počtu záznamů, bylo žádoucí ponechat co nejvíce informací a neořezávat soubor více, než je nezbytné. Po bližším prozkoumání se však i z tohoto hlediska objevilo množství záznamů, jež by pro výsledné hledání orientovaných vztahů neměly přínos. Typicky šlo o pobyty pacientů, kteří za sledované období měli pouze jednu hospitalizaci, v rámci níž byla vykázána pouze jedna hlavní diagnóza, případně záznamy o živě narozených dětech či osobách, které se setkaly se zdravotnickými službami za jiných okolností (například zdravé osoby sloužící jako doprovod). Tedy záznamy, u nichž byla vykázána pouze jedna z diagnóz Z38 či Z76).

Schéma čištění datového souboru z hlediska počtu záznamů je blíže znázorněno na Obr. 6. Prvním krokem bylo odstranění záznamů obsahujících pouze jednu ze zmíněných diagnóz Z38, Z76, případně chybnou hodnotu. Poté proběhlo odstranění vnořených hospitalizací (hospitalizace v jiném zdravotnickém zařízení v rámci hospitalizačního pobytu), a to zejména

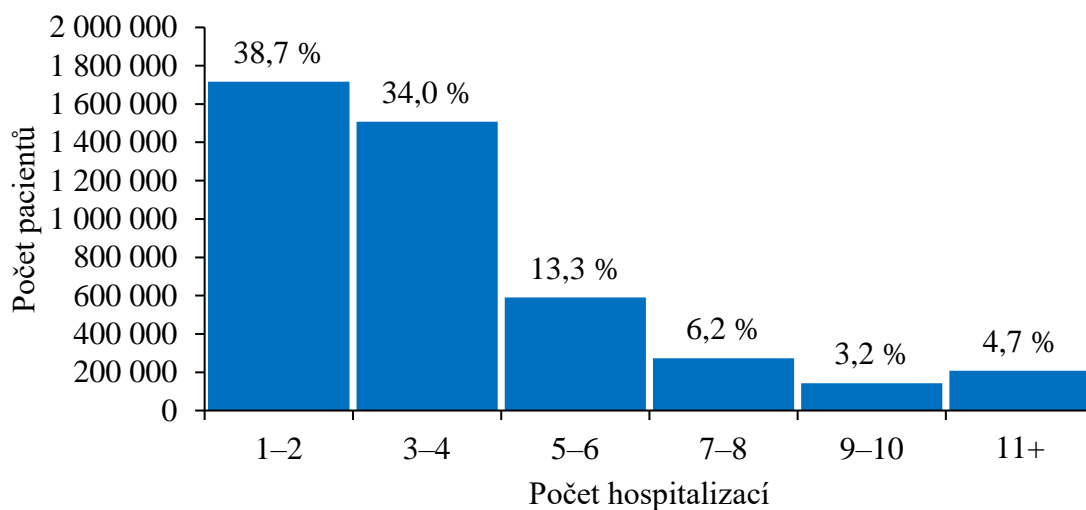
vzhledem k jejich problematickému časovému zařazení do individuálního chronologického vektoru pacientových diagnóz. Následně byly vymazány záznamy pacientů, kteří měli pouze jednu hospitalizaci, v rámci níž jim byla vykázána pouze jedna hlavní diagnóza, jelikož by tyto záznamy do výsledného tvoření kombinací v konečném důsledku neměly čím přispět. Pro každého pacienta bylo v rámci sledovaného časového období uvažováno maximálně 10 hospitalizací. Výchozí soubor pro tvoření kombinací ve finále obsahoval celkem 17,1 milionů hospitalizací u 4,4 milionů unikátních pacientů.



**Obr. 6** Schéma přípravy datového souboru

Z Obr. 6 si lze všimnout, že nejvlivnějším kritériem na výslednou velikost datového souboru byl požadavek na více vykázaných hlavních diagnóz. Jak již bylo zmíněno, pro každého pacienta bylo uvažováno maximálně 10 záznamů, jelikož následná tvorba kombinací by pro větší počet záznamů byla výpočetně velmi náročná. Podíváme-li se na počet hospitalizací před výsledným ořezáním na soubor s maximálně 10 hospitalizacemi na pacienta (Obr. 7), lze si všimnout, že téměř tři čtvrtiny pacientů (72,7 %) mají mezi 1 až 4 hospitalizacemi. Pouze 4,7 % pacientů má 11 a více záznamů. Vzhledem k průměrnému počtu hospitalizací na pacienta tak toto kritérium není nijak zásadně omezující a jde o kompromis mezi počtem záznamů

a výpočetní náročností. Medián počtu záznamů pacientů tvoří 3 hospitalizační pobyty a průměr je 4,1 hospitalizačních pobytů na pacienta. Průměr je však ovlivněn extrémními hodnotami, maximální počet hospitalizačních pobytů na jednu osobu ve sledovaném období je 731.

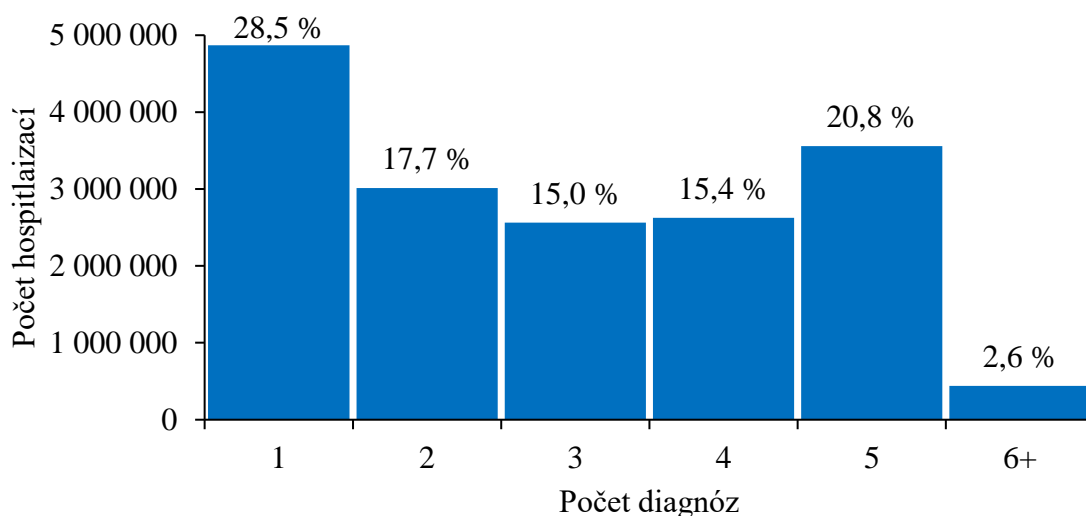


**Obr. 7** Počet záznamů o hospitalizacích na jednoho pacienta (N = 18 155 718 hospitalizací)

#### 6.2.2 Základní charakteristiky datového souboru

Nyní se už veškeré výstupy v práci budou zakládat na finálním ořezaném souboru s 17 068 811 hospitalizacemi.

Podíváme-li se na počet diagnóz na hospitalizaci (Obr. 8) vidíme, že největší podíl zaujímá případ, kdy pacienti měli za hospitalizace vykázanu pouze jednu diagnózu. Maximální počet vykázaných diagnóz během jednoho pobytu je 11 (5 hlavních, 5 vedlejších a 1 výsledná). Průměrný počet činí 2,9 diagnóz na jeden hospitalizační pobyt.



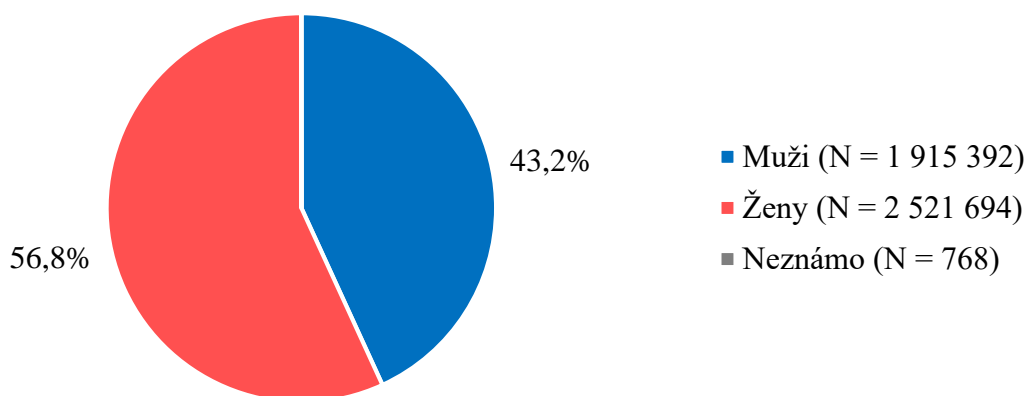
**Obr. 8** Počet diagnóz na jednu hospitalizaci

V Tab. 2 jsou uvedeny základní charakteristiky počtu diagnóz rozděleny na hlavní a vedlejší. Lze si všimnout, že zatímco průměr počtu hlavních diagnóz na jeden pobyt je pouze 1, průměr počtu vedlejších diagnóz je téměř dvojnásobný. Jedná se již o charakteristiku očištěného souboru a nejsou tak brány v úvahu duplicitní diagnózy v rámci jednoho hospitalizačního pobytu.

**Tab. 2** Počet diagnóz na hospitalizaci – charakteristiky

	<b>Průměr</b> ± směrodatná odchylka	<b>Medián</b> (5.–95. percentil)
N = 17 068 811		
<b>Hlavní diagnózy</b>	1,0 ± 0,2	1,0 (1,0–1,0)
<b>Vedlejší diagnózy</b>	1,9 ± 1,6	2,0 (0,0–4,0)
<b>Diagnózy celkem</b>	2,9 ± 1,6	3,0 (1,0–5,0)

Celkem je ve výsledném souboru 4 437 854 pacientů, z čehož je 2 521 694 žen a 1 915 392 mužů. U 768 pacientů je pohlaví neznámo. Graf se zastoupením pohlaví je znázorněn na Obr. 9.



**Obr. 9** Zastoupení pohlaví

Důvodem k mírné převaze žen jsou pravděpodobně hospitalizace spojené s těhotenstvím a porodem.

Tab. 3 ukazuje základní charakteristiky věkového složení pacientů při jejich poslední hospitalizaci

**Tab. 3** Věkové složení pacientů při poslední hospitalizaci

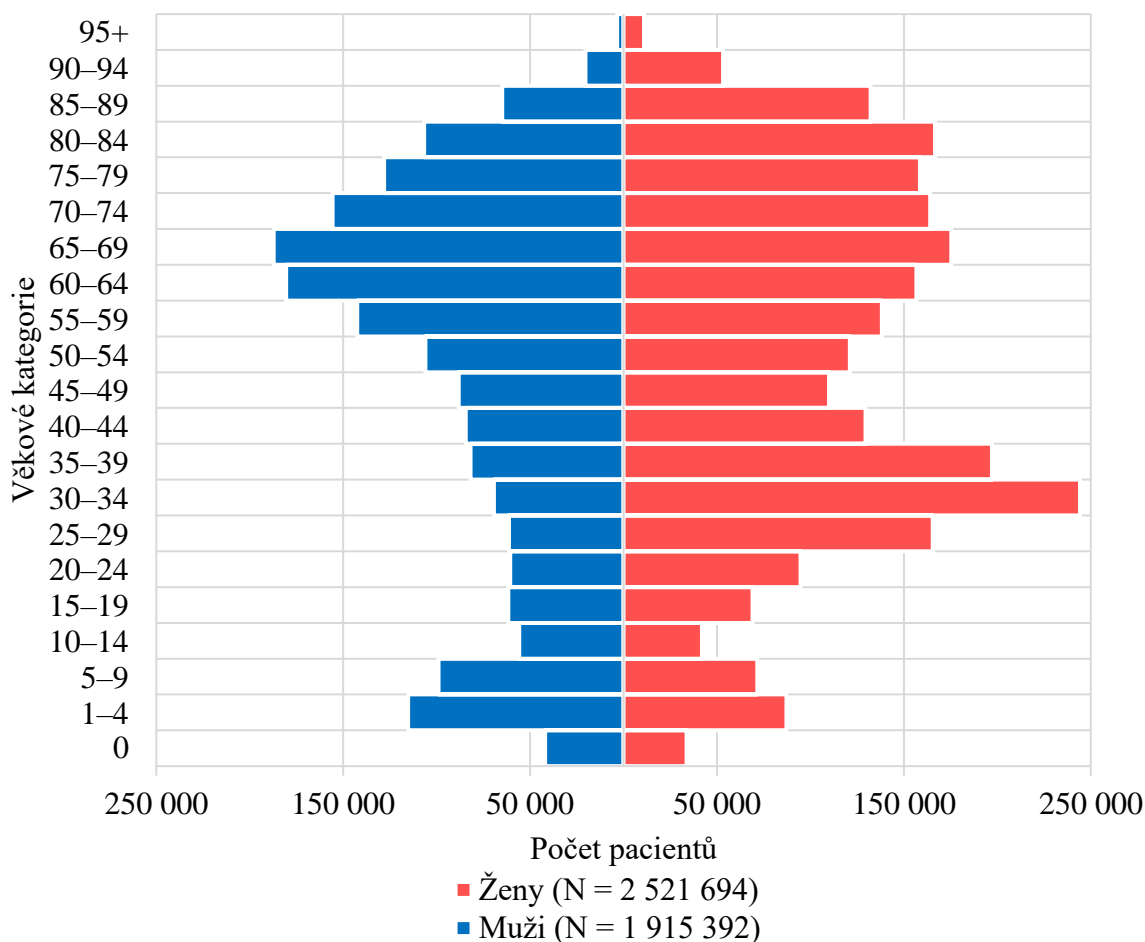
	<b>Průměr</b> ± směrodatná odchylka	<b>Medián</b> (5.–95. percentil)
<b>Muži (N = 1 915 392)</b>	48,9 ± 26,2	56,0 (2,0–84,0)
<b>Ženy (N = 2 521 694)</b>	49,7 ± 25,0	50,0 (5,0–87,0)
<b>Celkem (N = 4 437 086)</b>	49,4 ± 25,5	53,0 (3,0–86,0)

Průměrný věk všech pacientů v souboru je 49,4 let a medián 53 let. Muži mají věkový průměr 48,7 a medián 56 let. U žen je průměrný věk o něco vyšší než u mužů (49,7 let), avšak medián je o šest let nižší než u mužů (50 let).

Obr. 10 znázorňuje věkovou strukturu pacientů v souboru při poslední hospitalizaci. Věk byl kategorizován do 21 kategorií po pěti letech, vyjma poslední kategorie sdružující pacienty ve věku 95 let a starší a prvních dvou kategorií, kdy stojí samostatně věková skupina 0 a 1–4.

I zde si lze všimnout, že převaha žen v souboru je pravděpodobně spojena s hospitalizacemi pro těhotenství a porod. Sledujeme nárůst počtu hospitalizací žen ve věku 25–44, přičemž tato věková kategorie zahrnuje 29,2 % všech hospitalizovaných žen. U mužů je tento podíl téměř poloviční a do věkové kategorie 25–44 spadá 15,5 % hospitalizovaných mužů. K dalšímu nárůstu dochází u pacientů ve věku 55 až 89 let, přičemž tento nárůst je zřetelný především u mužů.

Rovněž zde vidíme možné vysvětlení rozdílu mezi muži a ženami co se týče průměrného věku a mediánu. Ženy jsou častěji hospitalizovány i ve starším věku než muži a věkový průměr pravděpodobně ovlivňuje i možné extrémní hodnoty při hospitalizacích žen pro věkovou kategorii 95+. Medián je u žen naopak nižší než u mužů, jelikož velká část pacientek spadá do mladších a středních věkových kategorií



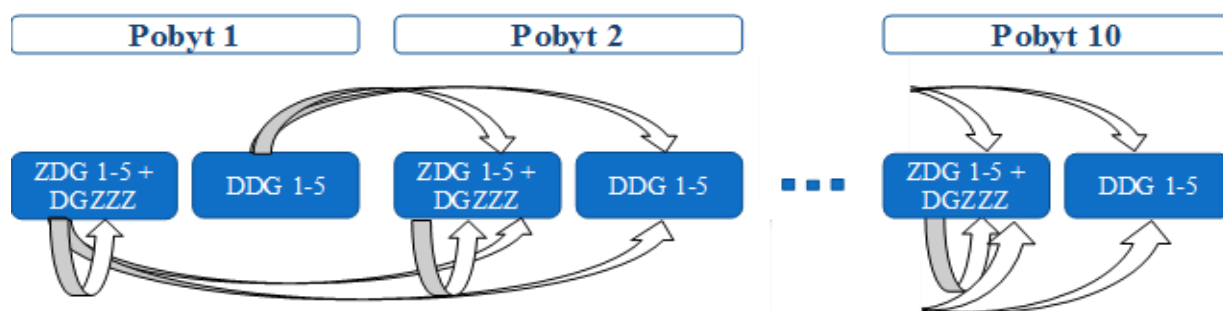
**Obr. 10** Věková struktura pacientů při poslední hospitalizaci dle pohlaví

### 6.2.3 Tvorba kombinací

Pro tvorbu kombinací bylo využito celkem 11 proměnných týkajících se diagnóz pacienta, konkrétně všech 5 hlavních, 5 vedlejších a 1 výsledná diagnóza. Výsledná diagnóza však byla ve výsledném souboru brána v potaz jen v malém množství případů. Jde totiž o diagnózu, která shrnuje celý pobyt pacienta ve zdravotnickém zařízení, a je tak ve většině případů již vykázána mezi hlavními diagnózami. Základním požadavkem před tvorbou kombinací bylo, aby pacient měl v rámci pobytu každou diagnózu maximálně jedenkrát a nedocházelo tak k nadbytečným duplikacím.

Před samotnou tvorbou kombinací byla ještě provedena restrukturalizace, kdy se pobyty každého pacienta, původně zapsané v řádcích jako jednotlivé záznamy, přeskupily na jeden řádek. Pacienty pak charakterizoval chronologický vektor jejich diagnóz. Ačkoli původním vstupem bylo 11 proměnných s diagnózami, po restrukturalizaci se ve výchozím souboru pro tvorbu kombinací tento počet zdesetinásobil, jelikož bylo uvažováno až 10 hospitalizací na pacienta.

Prioritu měly hlavní diagnózy před vedlejšími, výsledná diagnóza se v rámci každého pobytu zařadila mezi poslední hlavní a první vedlejší diagnózu. Při dalším popisu bude však výsledná diagnóza zahrnována do skupiny hlavních diagnóz. Na Obr. 11 je znázorněno schéma tvorby kombinací. V základu byly kombinace tvořeny na úrovni třímístných diagnóz a teprve poté zobecňovány na vyšší úrovni.



**Obr. 11** Schéma tvorby kombinací. ZDG 1-5 značí pětici hlavních diagnóz, DGZZZ označuje diagnózu výslednou a DDG 1-5 zahrnuje skupiny pěti vedlejších diagnóz.

Jak je zmíněno výše, každého pacienta charakterizoval vektor diagnóz chronologicky řazených dle pobytů a v rámci každého pobytu hlavní diagnózy předcházely vedlejšími. V rámci jednoho pobytu spolu byly kombinovány pouze hlavní diagnózy, vedlejší diagnózy se kombinovaly pouze s hlavními a vedlejšími diagnózami následujících pobytů. Například první hlavní diagnóza z prvního pobytu vešla do kombinace se všemi ostatními hlavními diagnózami prvního pobytu a všemi diagnózami (hlavními i vedlejšími) pobytů následujících, nikoli však s vedlejšími diagnózami téhož pobytu. Druhá hlavní diagnóza prvního pobytu byla kombinována s následujícími hlavními diagnózami prvního pobytu (nikoli však už s první hlavní diagnózou) a opět se všemi diagnózami následujících pobytů. Co se týče vedlejších diagnóz, ty spolu v rámci jednoho pobytu do kombinace nevcházely, ale byly kombinovány až se všemi diagnózami následujících pobytů. Výsledný maximální počet možných kombinací pro jednoho pacienta byl 5 600. Převážná většina kombinací však byla prázdná, protože jak lze

vidět z Obr. 7 a Obr. 8, většina pacientů měla méně jak limitních 10 hospitalizací a rovněž při hospitalizacích byl vykazován menší počet diagnóz. Vznikaly tak i prázdné či neúplné kombinace. Nicméně vzhledem k velkému množství pacientů a variabilitě počtu jejich hospitalizací i diagnóz, bylo nutné spočítat všechny možné kombinace proměnných pro celý soubor a kombinace bez informace poté vyloučit.

Soubor s vytvořenými kombinacemi byl původně vyprodukovan na úrovni tříznakových kódů klasifikace MKN-10, ale vzhledem k již zmíněným problémům s velkým množstvím těchto diagnóz a možnému vykazování podobných stavů různými kódy byl soubor dále zobecněn, a to jak na úroveň kapitol, tak také na úroveň oddílů MKN-10. Při tvorbě oddílů proběhlo detailnější rozdělení pro novotvary in situ, nezhooubné novotvary a novotvary nejistého nebo neznámého chování. Zatímco původně tyto tři kategorie tvořily samostatné oddíly, v práci byla snaha přiblížit se členění dle lokalizace zhoubných novotvarů. Seznam nadefinovaných oddílů je dodán jako Příloha 3. Zatímco práce na úrovni kapitol je příliš obecná a nabízí pouze základní vzhled, oddíly tvoří detailnější homogenní skupiny členěné především podle lokalizace a chování nemocí.

Výsledný počet záznamů s plnohodnotnými kombinacemi na úrovni tříznakových diagnóz v celém souboru činil 390 392 647, je však nutné si uvědomit, že většinu z nich tvořily duplicity v rámci pacienta. Proto bylo dále nutné protřídit soubor tak, aby pro každého pacienta byla uvažována daná kombinace diagnóz pouze jedenkrát, aby síla daného vztahu nebyla nadhodnocována. Například měl-li pacient třikrát zaznamenanou kombinaci E11 (diabetes mellitus 2. typu) → I25 (chronická ischemická choroba srdeční), do výsledného souboru kombinací byla z jeho historie započítána tato kombinace pouze jednou. Tuto agregaci však bylo nutné provádět pro každou úroveň MKN klasifikace zvlášť. Zatímco kombinace E11 → I25 a E66 (obezita) → I10 (esenciální hypertenze) jsou na úrovni třímístných kódů odlišné a započítají se tak obě, podíváme-li se na úroveň kapitol MKN-10, prezentují tyto kombinace pouze jednu totožnou, a to IV. (nemoci endokrinní, výživy a přeměny látek) → IX. (nemoci oběhové soustavy). Na úrovni kapitol by tak obě zmíněné kombinace měly být prezentovány pouze jednou zobecněnou. Budeme-li tedy brát kombinace na třech různých úrovních (kapitoly, oddíly, tříznakové diagnózy), každý soubor bude mít odlišný počet kombinací, přičemž u kapitol bude počet nejmenší, zatímco na nejpodrobnější úrovni tříznakových diagnóz bude tento počet nejvyšší.

Vzniklé finální počty záznamů pacientů zachycující kombinace bez duplikací v historii pacienta činily 170 506 499 pro úroveň třímístných kódů, 128 173 669 pro oddíly MKN-10 a 59 919 486 pro nejjobecnější soubor na úrovni kapitol MKN-10. Je třeba si uvědomit, že jsou zahrnuty i vztahy s jednou chorobou na obou stranách orientovaného vztahu, tedy například vztahy typu E11 → E11, které v sobě nesou informaci spíše o tom, že pacient byl vícekrát hospitalizován pro stejnou nemoc, než že by mohly nějak přispět ke zjištění příčinných vztahů. Přestože jsou kombinace těchto typů v základních souborech ponechány, při vizualizaci budou ignorovány. Následně proběhla agregace na jednotlivé unikátní kombinace pro samotné zjištění jejich frekvencí, které jsou dále potřebné i pro výpočet CIF indexu.

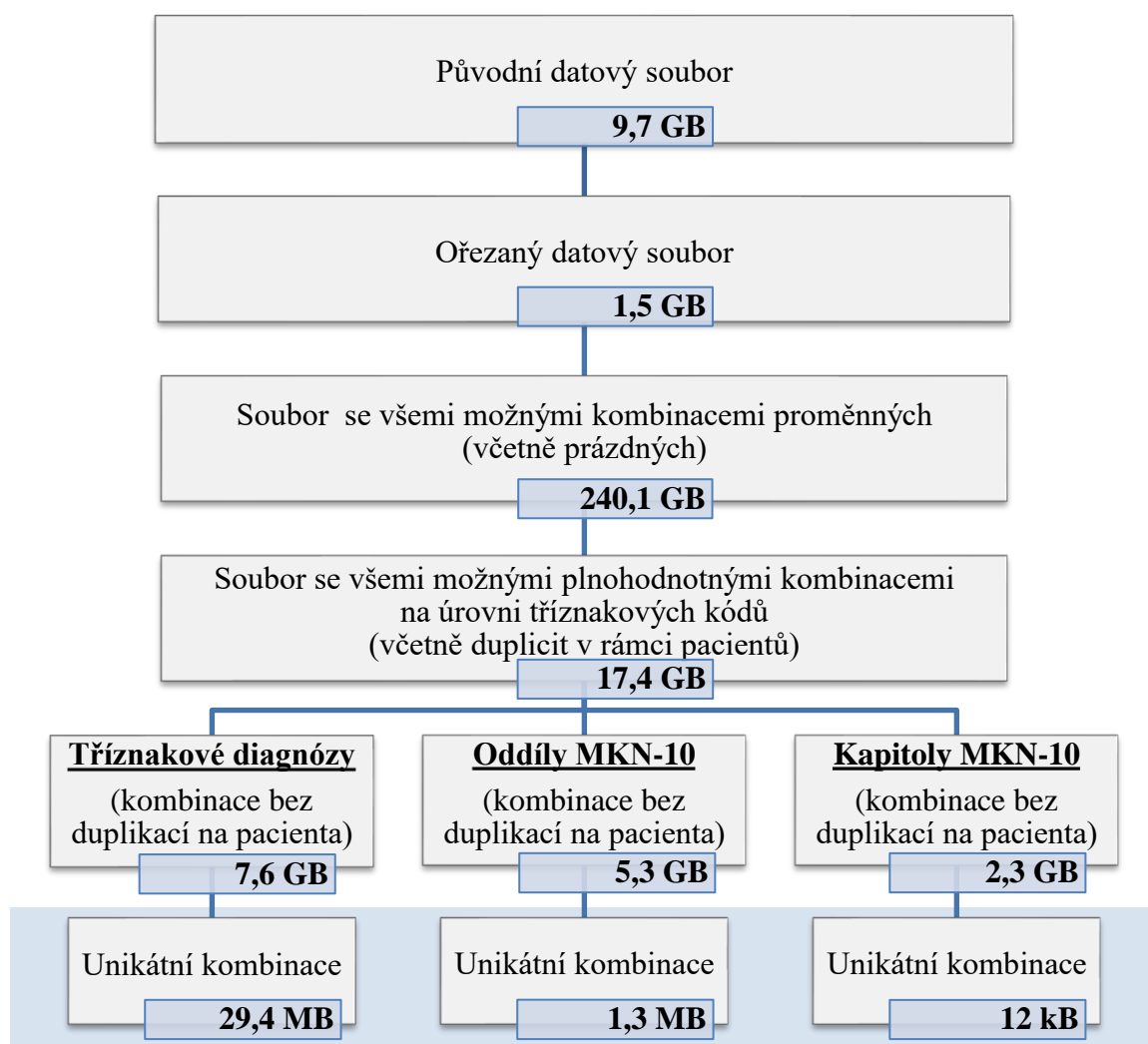
Tab. 4 prezentuje jak počet záznamů o kombinacích bez duplikací v rámci historie pacientů, tak počet unikátních kombinací po agregaci na jednotlivých úrovních MKN-10.

**Tab. 4** Počet všech kombinací (bez duplikací v pacientově historii) a počet unikátních kombinací na definovaných úrovních klasifikace

	Počet záznamů	Počet unikátních kombinací
<b>Kapitoly MKN-10</b>	59 919 486	484
<b>Oddíly MKN-10</b>	128 173 669	74 462
<b>Tříznakové diagnózy</b>	170 506 499	1 573 022

Jak si lze všimnout, bylo nutné se vypořádat s opravdu velkými soubory, aby se nakonec dospělo k třem malým výsledným souborům, které v základu obsahovaly pouze seznam unikátních kombinací a jejich frekvencí. Na tyto soubory se dále napojily ještě dva sloupce s prevalencemi (počtem pacientů) pro obě onemocnění příslušné daným kombinacím.

Postupný vývoj velikosti souboru v průběhu předpřípravy datového souboru a zjištění kombinací znázorňuje Obr. 12. Soubory týkající se kombinací v sobě nezahrnovaly textové popisy jednotlivých kapitol/oddílů/diagnóz, jelikož ty měly poměrně velký vliv na velikost souborů a pro skladování souborů nebylo ideální je ponechávat trvale.



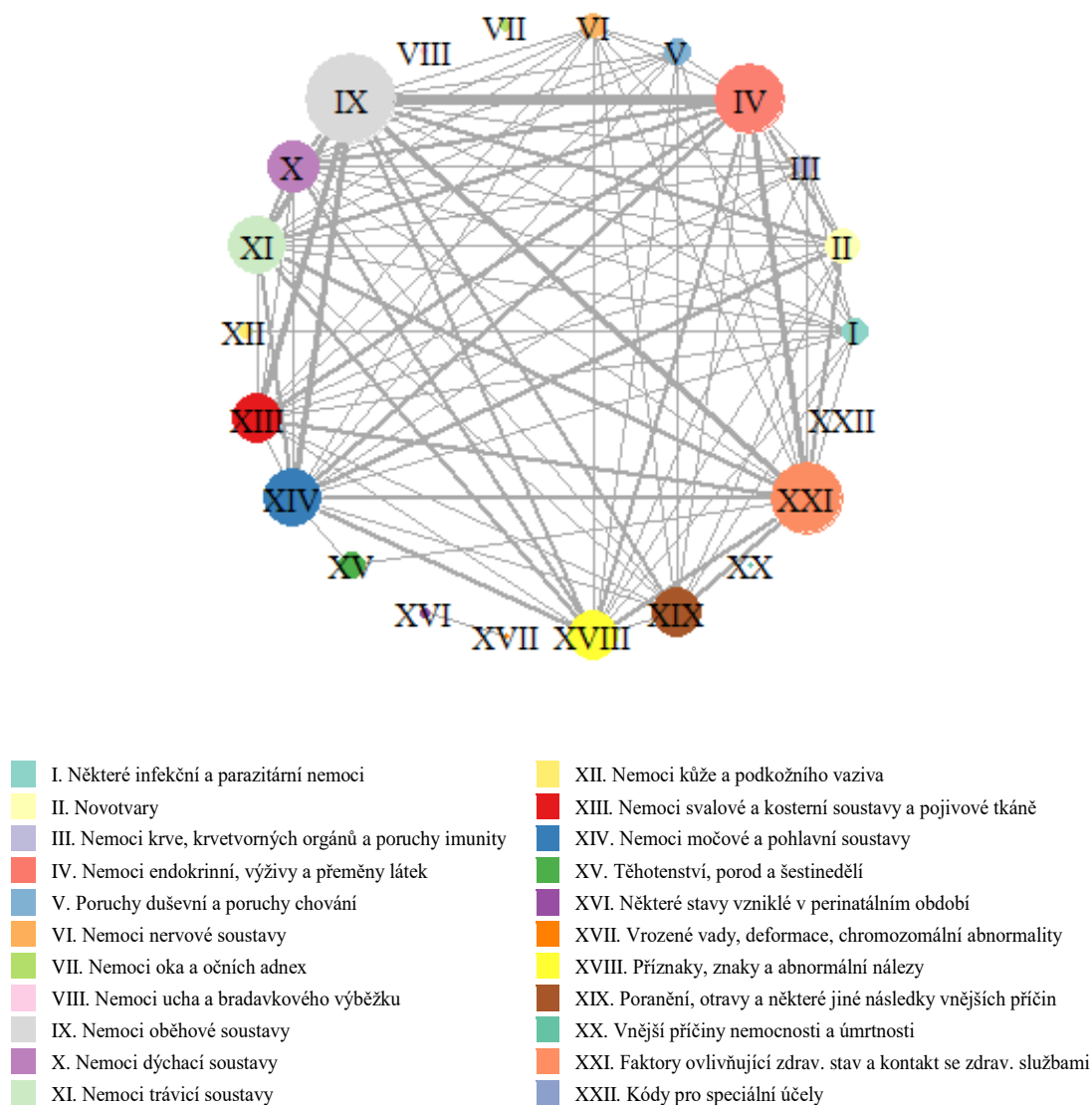
**Obr. 12** Vývoj velikosti souboru při přípravě a tvorbě kombinací



### 6.3 Síť humánních fenotypů

Před tvorbou orientované sítě byla zkonstruována i základní síť humánních fenotypů zobrazující asociace mezi nemocemi (Obr. 13). Jelikož se jedná jen o prvotní náhled na data prostřednictvím sítě a ukázkou vlastní sítě na poskytnutých datech, je konstruována pouze na obecné úrovni kapitol MKN-10.

Vstupním souborem pro tvorbu asociační matice byl, stejně jako pro následnou tvorbu orientované sítě, již ořezaný soubor s maximálně deseti hospitalizacemi na pacienta a s využitím hlavních a vedlejších diagnóz. Nejdříve proběhlo zobecnění diagnóz na úroveň kapitol MKN-10 a tvorba indikátorových proměnných, následně agregace na úroveň pacienta. Každého pacienta pak charakterizoval binární vektor, který identifikoval, zda se v jeho historii onemocnění z konkrétní kapitoly vyskytlo či nikoli. Na základě těchto vektorů byla spočítána asociační matice identifikující vztahy mezi kapitolami MKN-10. Vzhledem k binární povaze vektorů charakterizujících historii pacientů byl využit Jaccardův koeficient.



Obr. 13 Síť humánních fenotypů – kapitoly MKN-10

Velikost vrcholů grafu prezentuje četnost výskytu kapitol u pacientů a tloušťka spojnice odráží míru asociace. Tedy čím větší vrchol, tím více pacientů mělo vykázanou některou z nemocí spadající do dané kapitoly, a čím tučnější spojnice, tím silnější daná asociace je. Pro snadnější interpretaci a lepší přehlednost jsou v síti zahrnuty pouze asociace, které vykazovaly podobnost 0,1 a vyšší.

Lze si všimnout, že nejčetnější kapitola IX. týkající se nemocí oběhové soustavy se často a poměrně velkou silou pojí s ostatními. Nutno poznamenat, že četnost této kapitoly spočívá především v její povaze, jelikož do ní spadají mimo jiné dvě nejčetněji zaznamenávané diagnózy v historii pacientů, jimiž jsou hypertenzní nemoci a ischemické choroby srdeční. Především primární hypertenze je suverénně nejčetnější diagnózou vyskytující se v historii hospitalizovaných osob (výskyt u 1 665 209 pacientů). Na druhém místě, co se četnosti diagnóz u pacientů týče, se vyskytuje právě chronická ischemická choroba srdeční, i tak je však její četnost více než dvakrát menší než četnost primární hypertenze (výskyt u 786 118 pacientů). Diagnóza hypertenze, často vykazována na pozici vedlejší diagnózy, tak může poměrně výrazně ovlivňovat výsledné asociace.

Nejsilnější asociaci (s podobností 0,519) v síti spolu mají zmíněná IX. kapitola s kapitolou IV., která zahrnuje nemoci endokrinní, výživy a přeměny látek. IV. kapitola je u pacientů celkově třetí nejčetnější a zahrnuje v sobě mimo jiné diagnózu pro poruchy metabolismu lipoproteinů a jiné lipidemie a také diabetes mellitus 2. typu, což je z hlediska počtu pacientů třetí a čtvrtá nejčetnější diagnóza.

Jak již bylo zmíněno, zejména nemoci oběhové soustavy se často sdružují s ostatními. S podobností 0,319 se pojí k nemocem trávicí soustavy (kapitola XI.), s podobností 0,312 k nemocem močové a pohlavní soustavy (kapitola XIV.) a na úrovni podobnosti 0,303 jsou s nimi asociovány nemoci svalové a kosterní soustavy a pojivové tkáně (kapitola XIII.).

Nemoci endokrinní, výživy a přeměny látek se s podobností 0,292 asociují s nemocemi trávicí soustavy, s podobností 0,278 se k nim pojí nemoci močové a pohlavní soustavy (kapitola XIV.) a na úrovni podobnosti 0,265 nemoci svalové, kosterní a pojivové tkáně.

Kapitola II. zahrnující novotvary se asociuje s nemocemi oběhové soustavy na úrovni podobnosti 0,242. S podobností 0,209 se pojí k nemocem močové a pohlavní a také ke kapitole obsahující nemoci endokrinní, výživy a přeměny látek.

Lze si všimnout, že poměrně často se s ostatními kapitolami asociuje i kapitola XXI. Jedná se sice o druhou nejčetnější kapitolu, ale nezahrnuje v sobě diagnózy pro onemocnění jako taková, ale především faktory, které ovlivňují zdravotní stav a okolnosti, za kterých se pacienti se zdravotními službami setkali. Mimo jiné zahrnuje kódy pro screeningová vyšetření, radioterapie, či kódy pro identifikaci určitých chorob v osobní a rodinné anamnéze.

Tato síť představuje především prvotní náhled na data z pohledu asociací mezi jednotlivými kapitolami nemocí. Síťemi detailnějších úrovní, a především se zahrnutím orientovaných vztahů, se zabývá následující podkapitola 6.4.

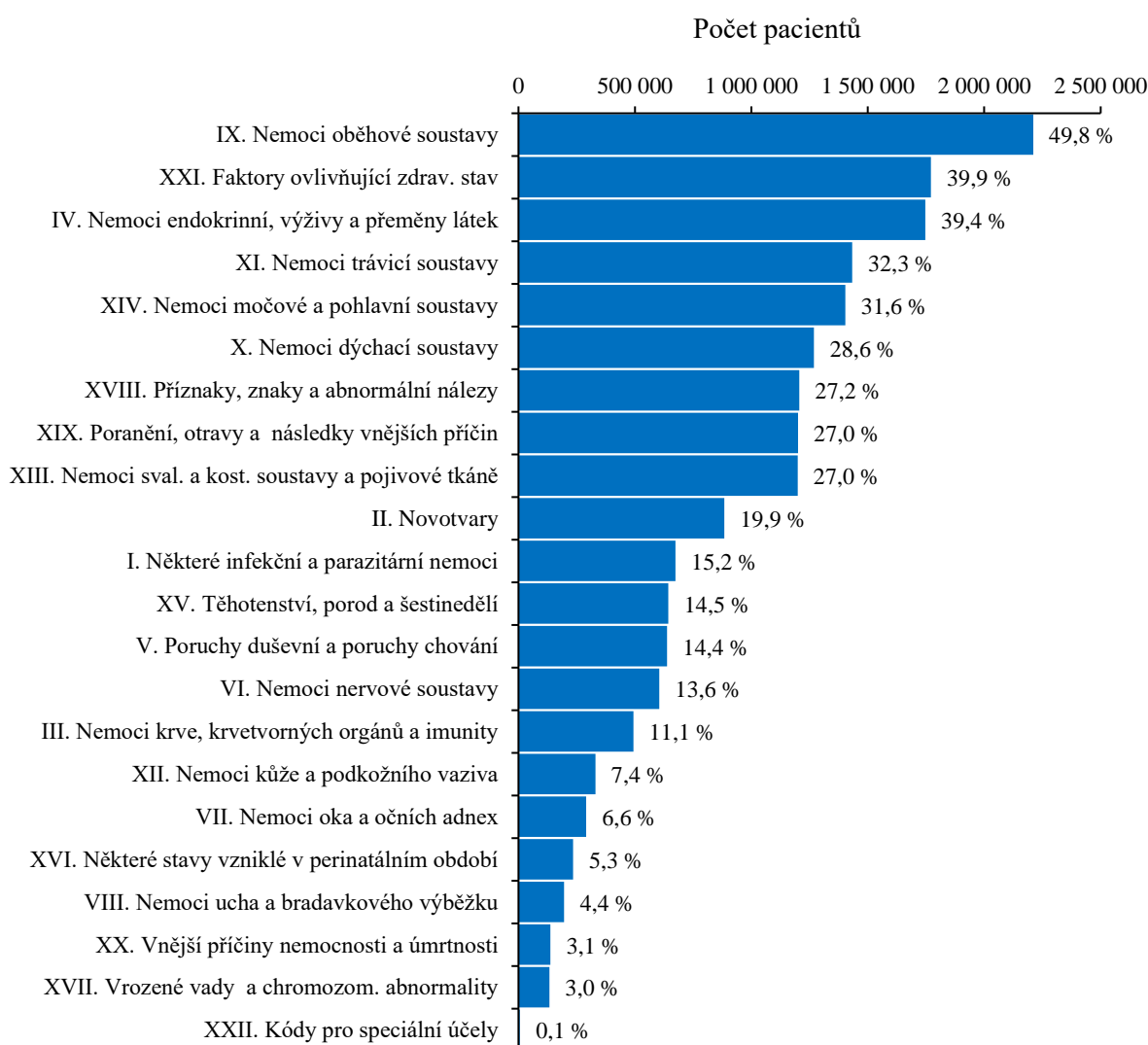
## 6.4 Síť humánních fenotypů s orientovanými vztahy

Jak již bylo zmíněno, identifikace příčinných vztahů v síti kvůli samotnému objemu a složitosti údajů není triviální a v některých případech se mohou různé kódy vztahovat k velmi podobným onemocněním. Proto budou, po prvním náhledu na orientovanou síť na úrovni kapitol, pro vybrané vztahy konstruovány sítě na úrovni oddílů MKN-10 a až poté pro detailnější popis i na úrovni tříznakových diagnóz.

Pro hodnocení vztahů byla využita varianta CIF indexu, zmíněného v podkapitole 5.3. Proces tvorby kombinací potřebných pro výpočet popisuje podkapitola 6.2.3.

### 6.4.1 Kapitoly MKN-10

Na Obr. 14 jsou uvedeny celkové četnosti jednotlivých kapitol, tedy u kolika pacientů byla někdy v historii vykázána diagnóza z dané kapitoly.

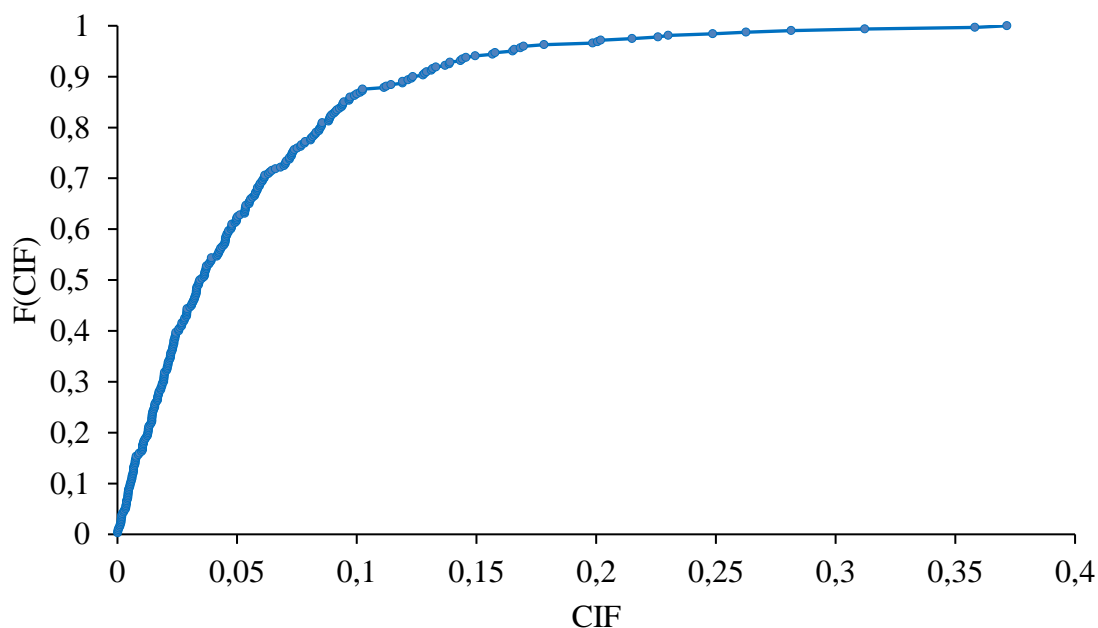


**Obr. 14** Zastoupení kapitol MKN-10 v hospitalizační historii pacientů

Z Obr. 14 je patrné, že dominuje kapitola IX. týkající se nemocí oběhové soustavy. Ve své hospitalizační historii nemoc z této kapitoly měla vykázána téměř polovina (49,8 %) pacientů, jedná se však opět o již očištěný soubor (mimo jiné bez diagnóz Z38 a Z76). K vysoké četnosti

této kapitoly velkou měrou přispívá primární hypertenze, často vykazována na místě vedlejší diagnózy. Druhou nejčetnější kapitolou je XXI. Faktory ovlivňující zdravotní stav a kontakt se zdravotnickými službami. Tahle kapitola neobsahuje diagnostické kódy onemocnění, ale zahrnuje kódy pro určení stavů ovlivňujících zdravotní stav. Nejčetnějším oddílem této kapitoly byla skupina kódů Z80–Z99 (Osoby s potenciálně ohroženým zdravím ve vztahu k rodinné/osobní anamnéze a některým podmínkám ovlivňujícím zdravotní stav). Tento oddíl zahrnuje kódy například pro implantáty, závislost na přístrojích a pomůckách, pooperační komplikace, ale také pro zaznamenání určitých diagnóz v osobní i rodinné anamnéze pacienta. Třetí nejčetnější kapitolou je IV. Nemoci endokrinní výživy a přeměny látek. Kapitola II. zahrnující novotvary je pak na desátém místě (výskyt u 19,9 % pacientů). Naopak nejméně četnou kapitolou je dle očekávání XXII. Kódy pro speciální účely. Tato kapitola slouží k pomocnému kódování a zahrnuje v sobě pouze několik kódů pro provizorní určení nových diagnóz či pro určení rezistence na antimikrobiální a protinádorové léky. Tato kapitola byla pro další práci vyloučena.

Na Obr. 15 je znázorněna kumulativní distribuční funkce CIF indexu vypočítaného pro kapitoly MKN-10 vyjma kapitoly XXII. Rovněž nejsou brány v potaz záporné hodnoty indexu. Záporná hodnota se může vyskytnout v případě, kdy četnost sledovaného vztahu  $n_{i \rightarrow j}$  příliš nízká.



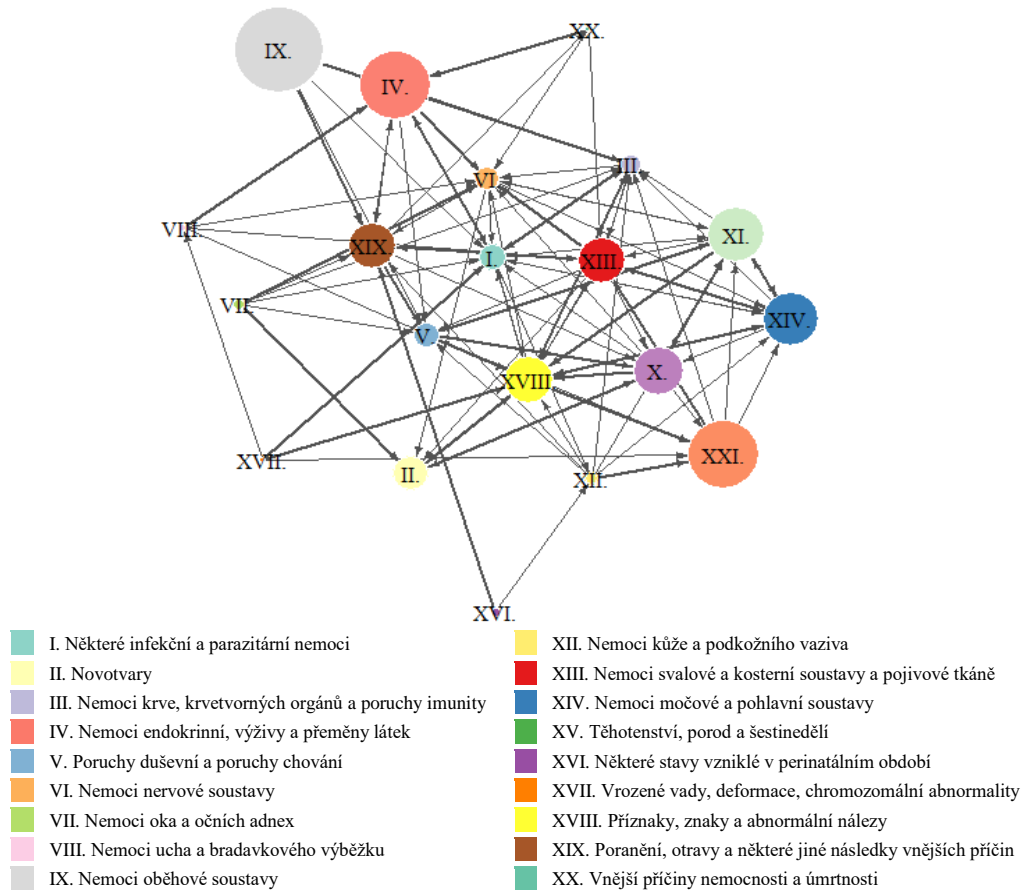
**Obr. 15** Kumulativní distribuční funkce pro CIF > 0 na úrovni kapitol MKN-10

Při vykreslování celé sítě může nastat problém při stanovení cut-off<sup>4</sup>. Příliš nízké hodnoty indexu náleží slabším vztahům, naopak vysoké hodnoty poukazují na vztahy zřejmé. Rovněž se mohou vyskytnout opačně orientované vztahy (tzv. artefakty), než by se dalo očekávat,

<sup>4</sup> Cut-off je pojem označující hraniční hladinu.

vykazující vysoké hodnoty indexu. Tato situace může nastat v případě, kdy je velký rozdíl v prevalencích (má-li první diagnóza výrazně nižší prevalenci než diagnóza druhá).

Obr. 16 zobrazuje orientovanou síť na úrovni kapitol. Aby síť nebyla příliš hustá, zobrazeny jsou vztahy pouze pro hodnoty indexu mezi 0,025 a 0,065.



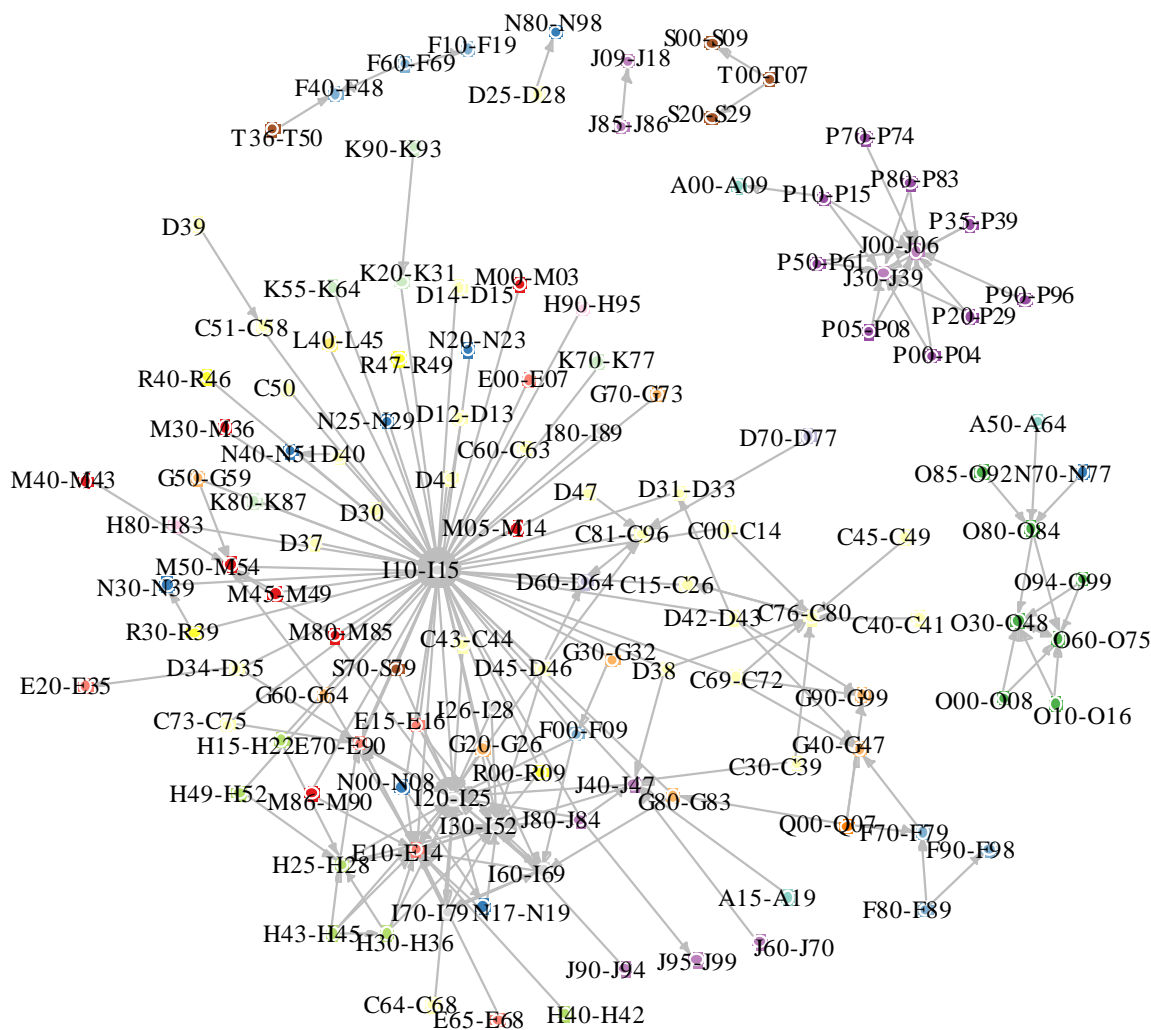
**Obr. 16** Orientovaná síť na úrovni kapitol MKN-10

Stejně jako v případě neorientované sítě (Obr. 13) i na Obr. 16 velikost vrcholu odráží četnost kapitoly a tloušťka spojnic sílu vztahu, tedy velikost CIF indexu. Pravděpodobně zde lze pozorovat vztahy, které jsou způsobeny velkými rozdíly mezi četnostmi kapitol. Na této úrovni však nelze ani s určitostí říct, zda se opravdu jedná o artefakty, a samozřejmě to neznamená, že na podrobnějších úrovních mezi nemocemi těchto kapitol nemohou existovat netriviální zajímavé vztahy. Jedná se o velmi obecnou síť, prakticky na úrovni orgánových soustav člověka, ze které nelze vypočítat žádné zajímavé orientované vztahy. Kapitoly nejsou na rozdíl od oddílů MKN-10 homogenní a obsahují velké množství rozmanitých diagnóz. Interpretace orientované sítě na této úrovni tak postrádá smysl.

## 6.4.2 Oddíly MKN-10

Další úroveň, na níž bude síť zkoumána, představují homogenní oddíly MKN-10. Přehled všech 291 definovaných skupin se nachází v Příloze 3. Vzhledem k povaze a případné nízké prevalenci byly vynechány oddíly posledních tří kapitol XX., XXI. a XXII.

Obr. 17 zobrazuje náhled na síť při intervalu CIF mezi 0,15 a 0,30, jelikož pro tyto hodnoty se začaly objevovat zajímavé vztahy.



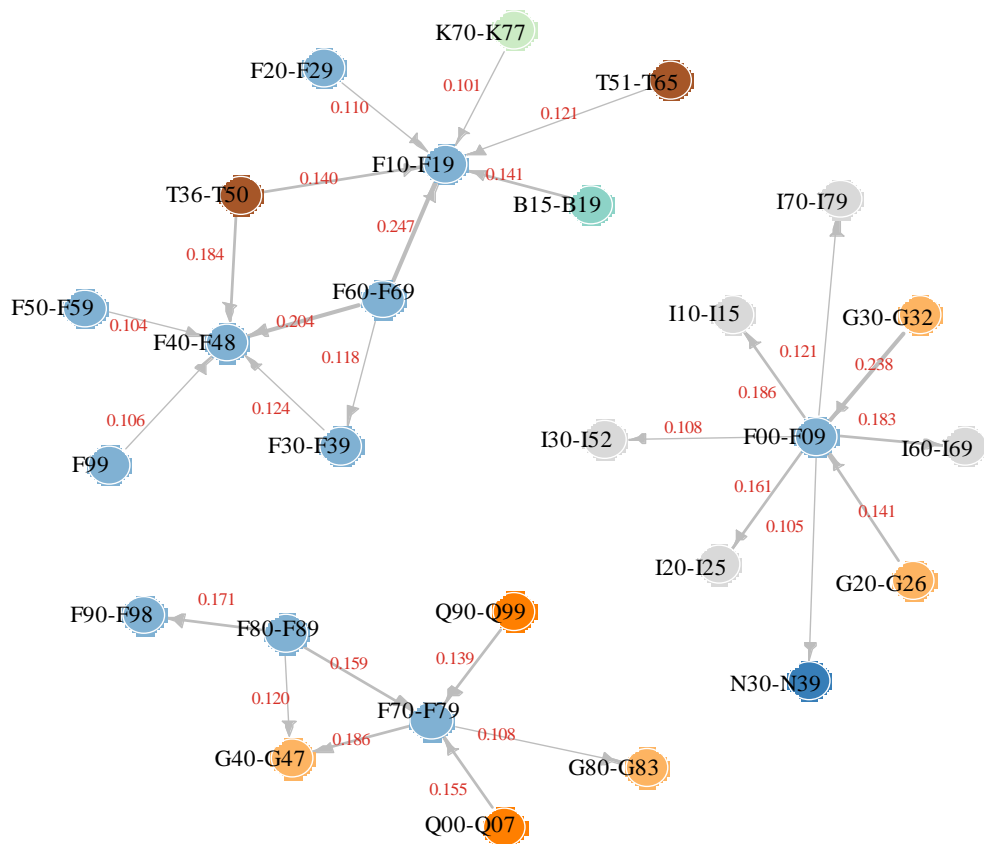
**Obr. 17** Orientovaná síť pro oddíly MKN-10

Kromě velkého shluku uprostřed si lze všimnout menších oddělených skupin. Důvodem, proč velká část oddílů směřuje k hypertenzním nemocem (I10–I15), může být jednak jejich velmi častý výskyt, ale pravděpodobně také to, že se příliš nevykazují jako primární diagnóza.

S některými stavy vzniklými v perinatálním období (fialový shluk vpravo nahoře) jsou provázány nemoci dýchací soustavy (akutní infekce horních dýchacích cest J00–J06 a jiné nemoci horních dýchacích cest (J30–J39)). Samostatně stojí také menší skupina týkající se duševních poruch (nahore), kde lze mimo jiné pozorovat vztah pro otravu léky (T36–T50) a pravděpodobně následnou diagnostiku neurotických, stresových a somatoformních poruch (F40–F48). Zajímavý může být také vztah mezi poruchami chování a osobnosti u dospělých (F60–F69) směřující k diagnóze duševních poruch a poruch chování způsobených užíváním

psychoaktivních látek (F10–F19). Vedle můžeme pozorovat osamostatněnou dvojici zobrazující vztah mezi nezahoubnými novotvory ženských pohlavních orgánů (D25–D28) a nezářlivými onemocněními ženského pohlavního ústrojí (N80–N98). Další samostatnou dvojici je souvislost mezi hnisavými a nekrotickými stavy dolní části dýchacího ústrojí (J85–J86) a chřipkou a zánětem plic (J09–J18). Další samostatnou skupinu tvoří trojice týkající se poranění. Poranění celého těla (T00–T07) pravděpodobně vede na následnou diagnostiku poranění hlavy a krku (S00–S09) a poranění hrudníku (S20–S29). Samostatnou skupinu tvoří také diagnózy týkající se těhotenství a porodu (vpravo).

Obr. 18 vizualizuje všechny vztahy spojené s duševními poruchami a poruchami chování pro CIF v rozmezí 0,10–0,25.



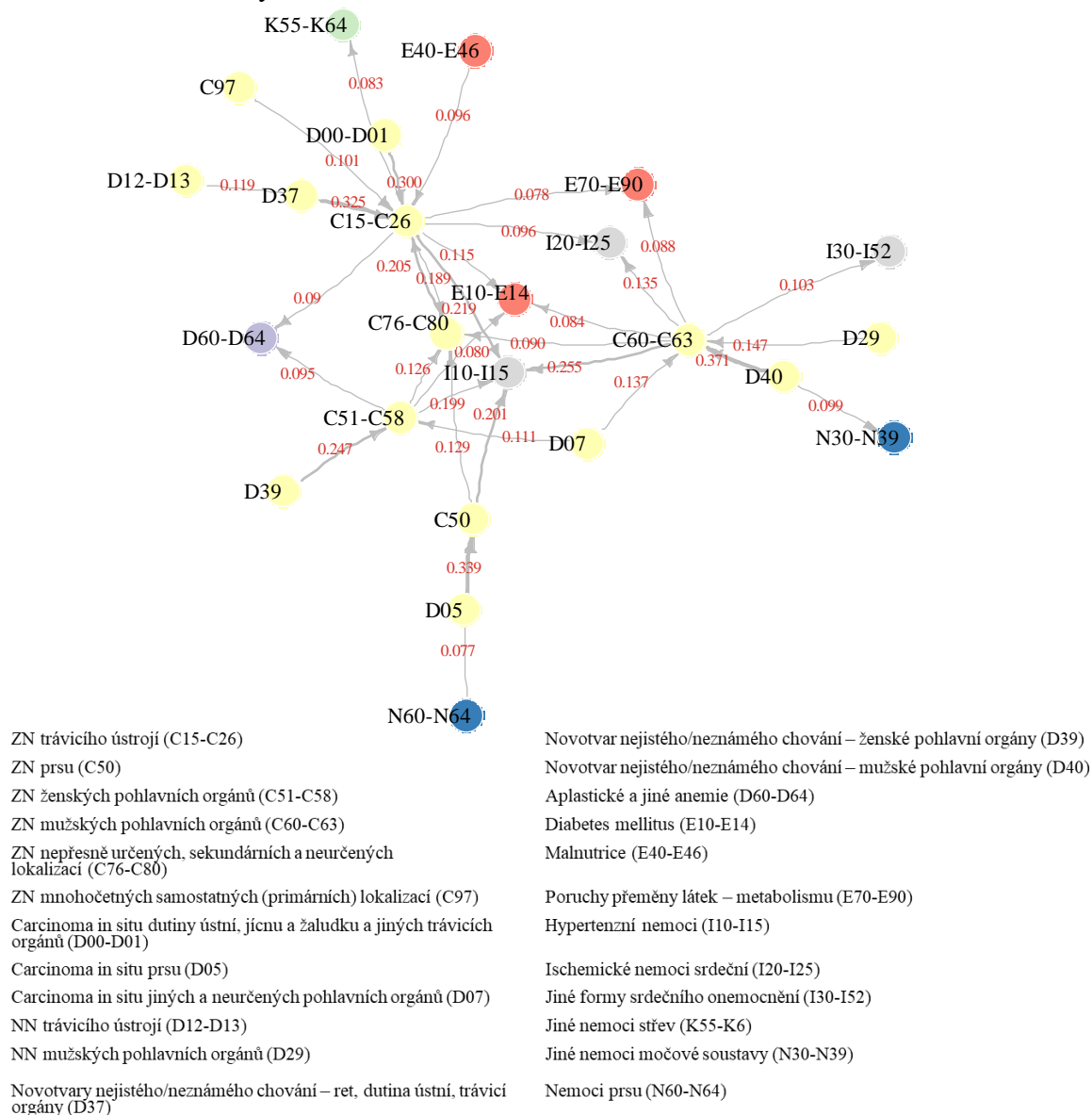
- |   |  |
|---|--|
| Virová hepatitida (B15-B19)   | Poruchy záchvatové - paroxysmální (G40-G47)                      |
| Organické duševní poruchy včetně symptomatických (F00-F09)                    | Mozková obrna a jiné syndromy ochrnutí (G80-G83)                 |
| Poruchy duševní a poruchy chování způsobené psychoaktivními látkami (F10-F19) | Hypertenzní nemoci (I10-I15)                                     |
| Schizofrenie, poruchy schizotypální a poruchy s bludy (F20-F29)               | Ischemické nemoci srdeční (I20-I25)                              |
| Afektivní poruchy (F30-F39)   | Jiné formy srdečního onemocnění (I30-I32)                        |
| Neurotické, stresové a somatoformní poruchy (F40-F48)                         | Cévní nemoci mozku (I60-I69)                                     |
| Syndromy poruch chování, spojené s fyziol. poruchami (F50-F59)                | Nemoci tepen, tepének a vlásečnic (I70-I79)                      |
| Poruchy osobnosti a chování u dospělých (F60-F69)                             | Nemoci jater (K70-K77)   |
| Mentální retardace (F70-F79)  | Jiné nemoci močové soustavy (N30-N39)                            |
| Poruchy psychického vývoje (F80-F89)  | Otrava léky, léčivy, návykovými a biologickými látkami (T36-T50) |
| Poruchy chování a emocí se začátkem obvykle v dětství a dospívání (F90-F98)   | Toxické účinky látek ze zdrojů převážně mimo lékařství (T51-T65) |
| Neurčená duševní porucha (F99)  | Vrozené vady nervové soustavy (Q00-Q07)                          |
| Extrapyramidové a pohybové poruchy (G20-G26)                                  | Abnormality chromozomů nezařazené jinde (Q90-Q99)                |
| Jiné degenerativní nemoci nervové soustavy (G30-G32)                          |  |

**Obr. 18** Orientované vztahy pro duševní poruchy a poruchy chování



Při vizualizaci vztahů spojených s duševními nemocemi lze pozorovat tři oddělené skupiny. Nahoře vidíme shluk se vztahy především pro poruchy způsobené užíváním psychoaktivních látek (F10–F19) a pro neurotické poruchy (F40–F48). Zajímavé vztahy obsahuje skupina umístěná dole, kde se duševní nemoci pojí s nemocemi nervové soustavy a vrozenými vadami. Vrozené vady a abnormality chromozomů v síti směřují k mentální retardaci. Ta je dále, stejně jako poruchy psychického vývoje, spojena se záchvatovými paroxyzmálními poruchami. Shluk vpravo je způsoben pravděpodobně spíše komorbidními vztahy než příčinnými. Zajímavě se jeví jiné degenerativní nemoci a extrapyrimidinové a pohybové nemoci směřující k organickým duševním poruchám (např. demence).

Následující síť na Obr. 19 vizualizuje vztahy u oddílů týkajících se zhoubných novotvarů, konkrétně prsu (C50), ženských pohlavních orgánů (C51–C58), mužských pohlavních orgánů (C60–C63) a zhoubných novotvarů trávicího ústrojí (C15–C26). Zobrazeny jsou vztahy s hodnotou CIF 0,075 a vyšší.

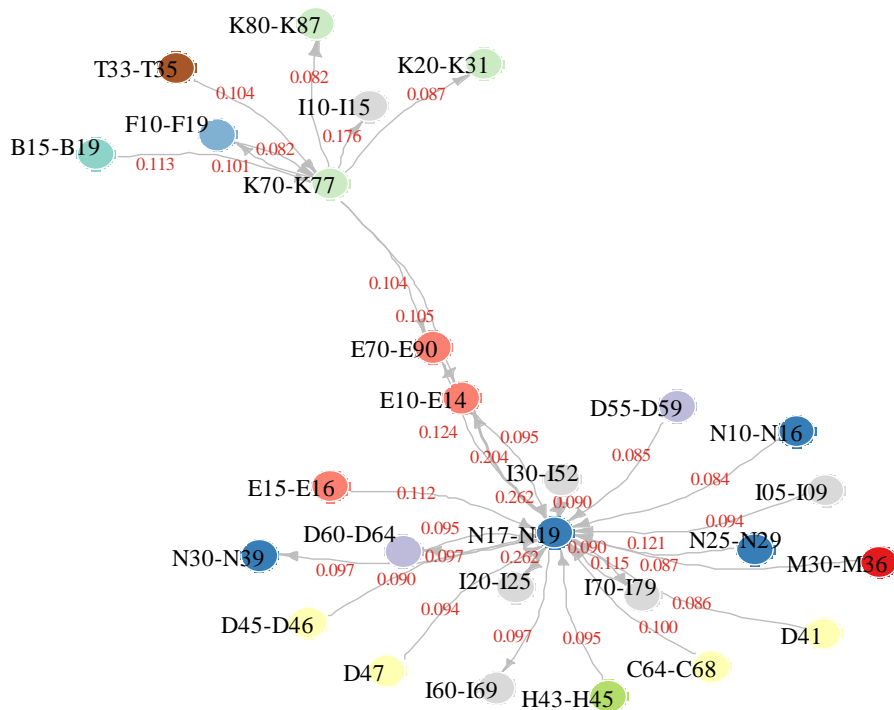


**Obr. 19** Orientované vztahy pro zhoubné novotvary prsu, ženských a mužských pohlavních orgánů a trávicího ústrojí. ZN – zhoubný novotvar, NN – nezhoubný novotvar.



Pokud bychom měli Obr. 19 pomyslně rozdělit dle definovaných spojitostí, v dolní části se nachází diagnózy spojené se zhoubnými novotvary prsu a ženských pohlavních orgánů, vpravo vztahy pro zhoubné nádory mužských pohlavních orgánů a nahoře jsou zobrazeny diagnózy spojené se zhoubnými novotvary trávicího ústrojí. Lze si všimnout častého propojení nezahobných novotvarů a carcinoma in situ se zhoubnými novotvary. Podíváme-li se například na diagnózu C50 (ZN prsu), vidíme, že dle sítě jí předcházela carcinoma in situ prsu (carcinoma in situ je karcinom lokalizovaný v místě svého vzniku bez přesahu do dalších vrstev) a nemoci prsu. ZN prsu je pak dále propojen se ZN nepřesně určených, sekundárních či neurčených lokalizací. Nejvíce vztahů se pojí pravděpodobně k ZN trávicího ústrojí. I zde lze pozorovat přechod od nezahobných novotvarů ke zhoubnému novotvaru.

Sít na Obr. 20 vizualizuje spojení pro další závažná onemocnění, jimiž jsou selhání ledvin (N17–N19) a nemoci jater (K70–K77). Zobrazené vztahy se nachází v rozmezí hodnot 0,08–0,30.



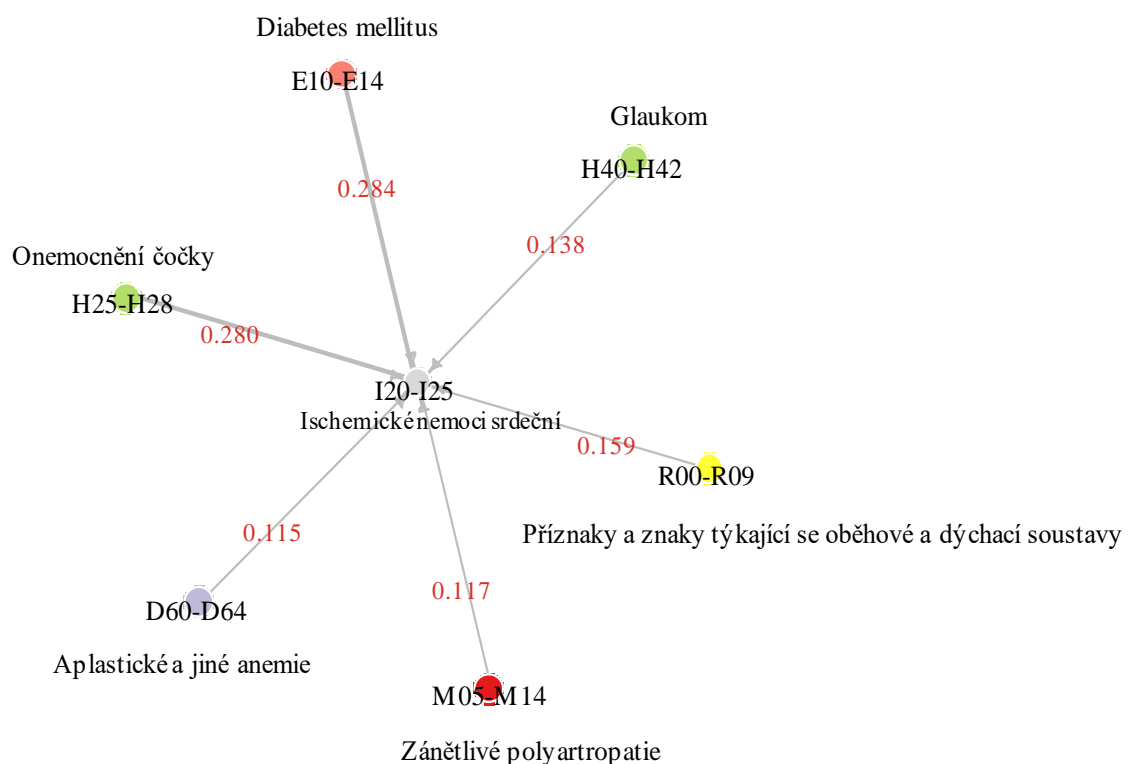
- |  |   |
|--|---|
| Virová hepatitida (B15-B19)  | Ischemické nemoci srdeční (I20-I25)                         |
| ZN močového ústrojí (C64-C68)  | Jiné formy srdečního onemocnění (I30-I52)                   |
| Novotvar nejistého/neznámého chování - močové orgány (D41)                           | Cévní nemoci mozku (I60-I69)                                |
| Polycythaemia vera, myelodysplastické syndromy (D45-D46)                             | Nemoci tepen, tepének a vlásečnic (I70-I79)                 |
| Novotvar nejistého/neznámého chování - mizní a krvetvorná tkáň (D47)                 | Nemoci jícnu, žaludku a dvanáctníku (K20-K31)               |
| Hemolytické anemie (D55-D59)   | Nemoci jater (K70-K77)                                      |
| Aplastické a jiné anemie (D60-D64)   | Nemoci žlučníku, žlučových cest a slinivky břišní (K80-K87) |
| Diabetes mellitus (E10-E14)  | Systémová onemocnění pojivové tkáně (M30-M36)               |
| Jiné poruchy regulace glukózy a vnitřní sekrece slinivky břišní (E15-E16)            | Tubulo-intersticiální nemoci ledvin (N10-N16)               |
| Poruchy přeměny látek - metabolismu (E70-E90)  | Selhání ledvin (N17-N19)                                    |
| Poruchy duševní a poruchy chování způsobené užíváním psychoaktivních látek (F10-F19) | Jiné nemoci ledvin a močodů (N25-N29)                       |
| Nemoci sklivce a očního bulbu (H43-H45)  | Jiné nemoci močové soustavy (N30-N39)                       |
| Chronické revmatické choroby srdeční (I05-I09)                                       | Omrzliny (T33-T35)  |
| Hypertenzní nemoci (I10-I15)   |   |

**Obr. 20** Orientované vztahy pro selhání ledvin a nemoci jater

Nahoře na Obr. 20 se nachází asociace s nemocemi jater, dolní část pokrývají vztahy pro selhání ledvin. Oba shluky jsou spolu propojeny přes poruchy přeměny látek a diabetes mellitus. Kromě vztahů nemocí jater s dalšími nemocemi trávicí soustavy lze pozorovat trajektorii vedoucí od virové hepatitidy – tento vztah je vzhledem k povaze nemoci logický. Ne úplně překvapivé je rovněž spojení duševních poruch a poruch chování, které jsou způsobeny požitím psychoaktivních látek, s onemocněním jater. Zvláště se však může jevit vztah omrzlin k nemoci jater. Omrzliny jsou ale asociovány právě s poruchami chování, které vznikly požitím psychoaktivních látek (mimo jiné i alkoholu). Vzhledem k povaze sítě tak trajektorii pravděpodobně chybí tento mezistupeň.

Kolem selhání ledvin se sdružuje podstatně více diagnóz. Spojnice v síti ukazují, že selhání ledvin pravděpodobně mohly předcházet nemoci glomerulů, jiné nemoci ledvin či močových, ale například i novotvary močových orgánů. Zajímavé jsou také vztahy anemií a systémových onemocnění pojivové tkáně směřující k selhání ledvin.

Obr. 21 znázorňuje vybrané vztahy pro oddíl I20–I25, který zahrnuje ischemické choroby srdeční (ICHS). Tato onemocnění patří mezi velmi rozšířené diagnózy způsobené nedostačeným prokrvením srdce. Zobrazené vztahy se nacházely v rozmezí CIF 0,10–0,30 a vzhledem k tomu, že tento oddíl má velké množství spojení s ostatními (jak lze vidět z Obr. 17), bylo do sítě vybráno pro ilustraci jen několik vztahů z daného rozmezí CIF indexu, které byly následně srovnány s literaturou.



**Obr. 21** Vybrané vztahy pro ICHS

Nejsilnějším (0,284) z vybraných vztahů je Diabetes mellitus → ICHS. Nejedná se o překvapivou informaci, diabetes je považován za jeden z rizikových faktorů rozvoje ICHS (Dresslerová a Vojáček, 2010). Tento vztah vykazoval poměrně velkou sílu (0,210) i opačným směrem. Zde je však třeba přihlídnout k možným rehospitalizacím pacientů a především ke struktuře dat, jelikož diabetes patří mezi onemocnění často vykazovaná na pozici vedlejších diagnóz.

Překvapivé jsou spojitosti mezi očními onemocněními a ICHS. Ačkoli by se mohlo zdát, že jediným důvodem těchto spojitostí je nadhodnocení způsobené rozdílem mezi prevalencemi, vztahy se tímto směrem projevují i při pohledu na pouhé frekvence bez zohlednění prevalencí (onemocnění čočky → ICHS má četnost 47 798, zatímco vztah opačný 28 742; vztah glaukom → ICHS byl zaznamenán v 18 056 případech, opačně 12 990). Ačkoli to samozřejmě neznamená, že by onemocnění spojené se zrakem byla rizikem pro rozvoj ICHS, souvislost kataraktu (součást H25–H28) i glaukomu (H40–H42) s ICHS již byla předmětem zájmu. Hu et al. se v roce 2016 zabývali zvýšeným rizikem ICHS u osob s kataraktou v tchajwanské populaci. Jejich zjištění naznačilo, že pacienti s kataraktou jsou vystaveni zvýšenému riziku následného rozvoje ICHS. Chen et al. v roce 2016 sledovali závislost glaukomu a ICHS. Jejich studie ukázala, že pacienti s primárním glaukomem otevřeného úhlu mohou trpět ICHS častěji než ostatní. Zdůrazňují však, že nedošli k závěru, že glaukom je rizikovým faktorem nebo že existuje kauzální vztah mezi těmito chorobami, ale pouze zjistili vyšší podíl vývoje ICHS u pacientů s glaukomem. To může být způsobeno podobnými základními mechanismy těchto dvou onemocnění.

Možným a pravděpodobným vysvětlením těchto vztahů může také být, že onemocnění zraku patří mezi doprovodné stavy nemocí diabetu a hypertenze, které jsou jako takové rizikovým faktorem pro ICHS.

Zajímavá může být spojitost zánětlivé polyartrózie → ICHS. Do oddílů M05–M14 spadají například i revmatoidní artritidy, přičemž pro tuto chorobu je zvýšené riziko aterosklerózy (kornatění tepen) (Kaplan, 2010).

Rovněž anemie se považuje za zhoršující okolnost ICHS. Na základě znalosti fyziologické úlohy přivádění kyslíku do myokardu může být příčinou závažnějších kardiovaskulárních onemocnění nebo markerem dalších procesů, které se vyskytují v těle a vyvolávají závažnější onemocnění (Zeidman et al., 2004).

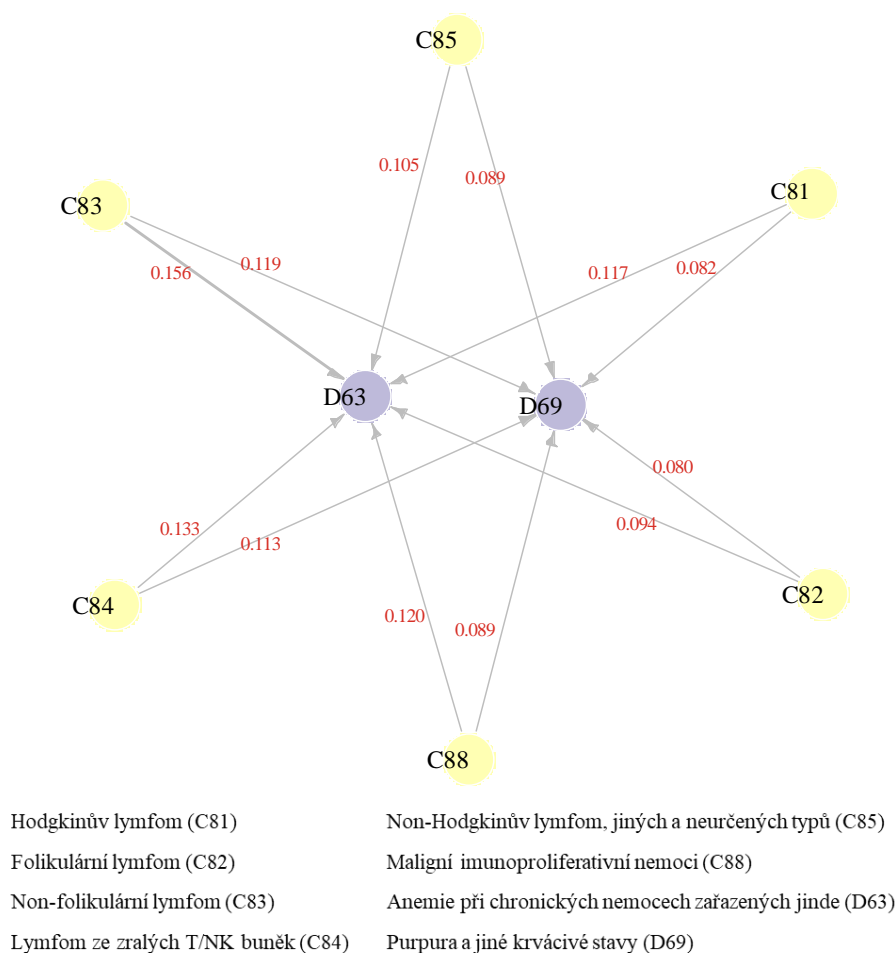
Co se týče vztahu R00–R09 → I20–I25, spíše než o kauzální vztah se může jednat o doprovodné okolnosti ICHS. Oddíl R00–R09 mimo jiné obsahuje typické projevy ICHS jako bolest na hrudi, nepravidelnost dýchání, abnormality srdeční činnosti a zjištění abnormálního krevního tlaku bez diagnózy hypertenze.

### 6.4.3 Tříznakové diagnózy MKN-10

Jak již bylo předesláno, vybrané vztahy budou zobrazeny na podrobnější úrovni tříznakových diagnóz. Pro další práci budou opět vyloučeny diagnózy z posledních třech kapitol (XX., XXI. a XIX.).

Sítě na úrovni tříznakových diagnóz jsou svým způsobem podgrafy sítě z Obr. 17, kdy je zobrazen vztah mezi dvěma oddíly, ale tentokrát na úrovni diagnóz, které do nich náležejí. Výběr zobrazených vztahů proběhl především na základě literatury.

Obr. 22 znázorňuje spojitosti mezi lymfomy (část oddílu C81–C96) a anemiemi (D60–D64), případně poruchami koagulace (D65–D69). Na úrovni oddílů se projevovaly vztahy směrem C81–C96 → D60–D64 s velikostí CIF 0,190 a C81–C96 → D65–D69 s CIF 0,128.

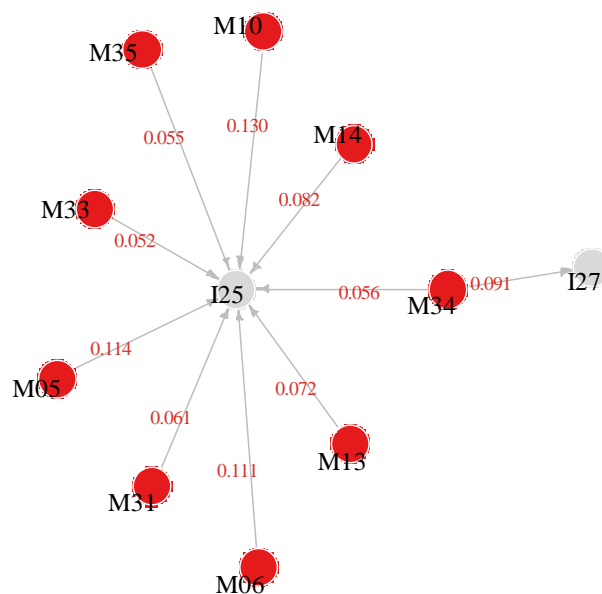


**Obr. 22** Orientované vztahy mezi lymfomy a anemiemi/poruchami koagulace

Zobrazeny jsou vztahy pro CIF 0,08 a větší. Vzhledem k tomu, že pouze 147 pacientů mělo ve své historii vykázanu diagnózu C86 (lymfom z T/NK-buněk, jiné určené typy), byla ze sítě vyloučena.

Vizualizované vztahy jsou víceméně v souladu se sítí, kterou zkonstruovali Kannan et al., 2016. Jsou však také v souladu s medicínskými znalostmi, jelikož projev anemie u pacientů s lymfomy je znám (Moulet et al., 1998; Birgegård, 2008).

Obr. 23 ilustruje spojitosti systémových onemocnění pojivové tkáně a zánětlivých polyartropatií s ischemickými nemocemi srdečními, případně kardiopulmonálními nemocemi. Spodní hranice CIF indexu byla 0,05.

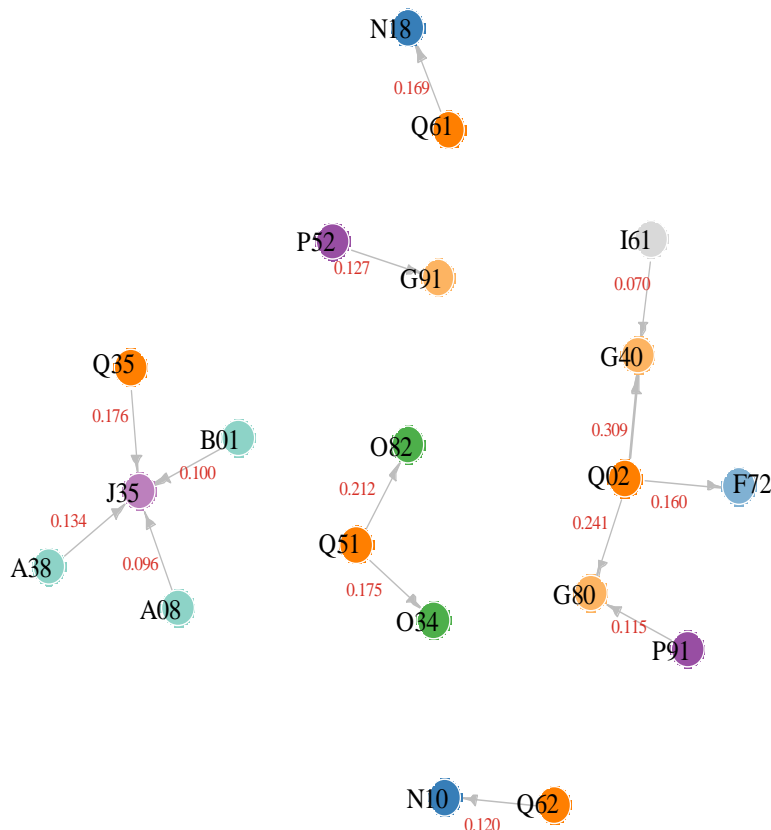


Chronická ischemická choroba srdeční (I25)	Artrópatie při jiných nemocech zařazených jinde (M14)
Jiné kardiopulmonální nemoci (I27)	Jiné nekrotizující vaskulopatie (M31)
Séropozitivní revmatická artritida (M05)	Dermatopolymyozitida (M33)
Jiná revmatická artritida (M06)	Systémová skleróza (M34)
Dna (M10)	Jiné systémové postižení pojivové tkáně (M35)
Jiná artritida (M13)	

**Obr. 23** Orientované vztahy mezi systémovými onemocněními pojivové tkáně a zánětlivými polyartropatiemi s ischemickými nemocemi srdečními, případně kardiopulmonálními nemocemi

Ze sítě si lze všimnout, že zejména diagnózy spadající do kategorie zánětlivých polyartropií pravděpodobně mohou předcházet chronické ischemické chorobě srdeční. Opět se jedná o vztahy korespondující se současnými znalostmi z medicínské praxe. Spojení mezi revmatickými artritidami, případně diagnózou dny s chronickou ischemickou chorobou nejsou neznámá (Crowson et al., 2013; Del Rincón et al., 2001, Choi et al., 2007). Systémová onemocnění pojivové tkáně vykazují oproti zánětlivým polyartropiím v síti o něco menší sílu spojení, avšak jedná se také o známé vztahy (Lizzerini et al., 2006; Al Husain et Bruce, 2010).

Obr. 24 ilustruje další vybrané vztahy v rozmezí CIF 0,07–0,31. Jedné se o spojení pro diagnózy týkající se především vrozených vad, chromozomálních abnormalit a nervové soustavy.



Střevní infekce viry a jinými určenými mikroorganismy (A08)

Spála – skarlatina (A38)

Plané neštovice – varicella (B01)

Těžká mentální retardace (F72)

Epilepsie – padoucnice (G40)

Mozková obrna (G80)

Hydrocefalus (G91)

Intracerebrální krvácení (I61)

Chronické nemoci mandlí a adenoidní tkáně (J35)

Akutní tubulo-intersticiální nefritida (N10)

Chronické onemocnění ledvin (N18)

Péče o matku pro (známou nebo suspektní) abnormalitu mateřských pánevních orgánů (O34)

Porod jediného plodu císařským řezem (O82)

Nitrolební neúrazové krvácení plodu a novorozence (P52)

Jiné poruchy stavu mozku novorozence (P91)

Microcephalia (Q02)

Rozštěp patra (Q35)

Vrozené vady dělohy a děložního hrdla (Q51)

Cystická nemoc ledvin (Q61)

Vrozené obstrukční defekty ledvinové pánvičky a vrozené vady močovodu (Q62)

**Obr. 24** Vybrané orientované vztahy pro vrozené vady, chromozomální abnormality a nemoci nervové soustavy

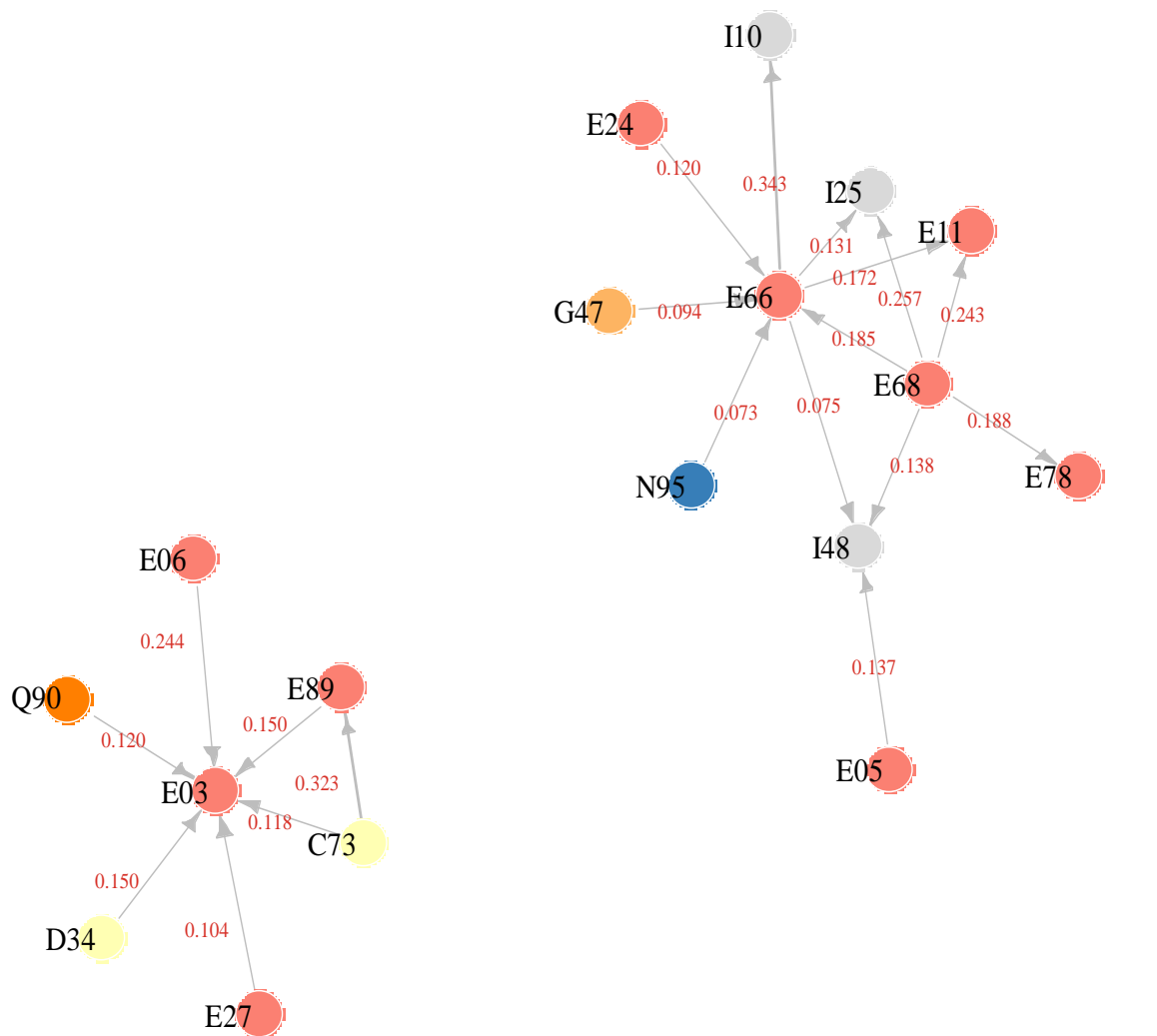
Vlevo pozorujeme shluk naznačující, že nemocím mandlí a adenoidní tkáně pravděpodobně mohou předcházet například nemoci jako spála a plané neštovice, ale také rozštěp patra, přičemž se nejedná o překvapivé souvislosti.

Také vidíme, že vrozené vady dělohy a děložního hrdla mohou vést na budoucí problémy při porodu, jako je potřeba speciální péče o matku a nutnost porodu císařským řezem. I zde se jedná o logickou souvislost spíše přispívající k věrohodnosti sítě.

Dále pozorujeme dvě osamostatněné dvojice naznačující, že vrozené obstrukční defekty ledvinové pánvičky a vrozené vady močovodu mohou předcházet tubulo-intersticiální nefritidě, a nitrolební neúrazové krvácení plodu a novorozence pravděpodobně může vést k hydrocefalu (nadměrné hromadění mozkomíšního moku v nitrolebí).

V pravé části Obr. 24 pozorujeme vztahy naznačující, že microcephalia (vývojová porucha projevující se zakrněním či předčasným ukončením růstu mozku a obvykle i celé hlavy) může vést k mozkové obrně, epilepsii nebo těžké mentální retardaci. Také tady se jedná o již zkoumané souvislosti (Woods, 2004; Abdel-Salam et al., 2000).

Obr. 25 znázorňuje vybrané spojitosti pro poruchy štítné žlázy, obezitu a jiné hyperalimentace (zvýšený příjem výživy). CIF je brán v rozmezí CIF 0,07–0,33.



- Zhoubný novotvar štítné žlázy (C73)
- Nezhoubný novotvar štítné žlázy (D34)
- Jiná hypotyreóza (E03)
- Tyr(e)otoxikóza (hypertyreóza) (E05)
- Zánět štítné žlázy – thyroiditis (E06)
- Diabetes mellitus 2. typu (E11)
- Cushingův syndrom (E24)
- Jiné poruchy nadledviny – glandulae suprarenalis (E27)
- Obezita – otylost (E66)

- Následky hyperalimentace (E68)
- Poruchy metabolismu lipoproteinů a jiné lipidemie (E78)
- Metabolické a endokrinní poruchy po operacích a jiných výkonech (E89)
- Poruchy spánku (G47)
- Esenciální (primární) hypertenze (I10)
- Chronická ischemická choroba srdeční (I25)
- Fibrilace a flutter síní (I48)
- Menopauzální a jiné perimenopauzální poruchy (N95)
- Downův syndrom (Q90)

**Obr. 25** Vybrané orientované vztahy pro poruchy štítné žlázy, obezitu a jiné hyperalimentace

Vlevo dole na Obr. 25 pozorujeme síť, jejímž středem je diagnóza zahrnující jiné hypotyreózy (snížená činnost štítné žlázy). Na sníženou produkci hormonů štítné žlázy má podle sítě vliv například zánět štítné žlázy, poruchy nadledvin a metabolické a endokrinní poruchy vzniklé po operacích a jiných výkonech. Mimo to ale snížené funkci štítné žlázy dle sítě pravděpodobně přecházely novotvary štítné žlázy, a to jak zhoubné, tak nezhoubné. Rovněž se zdá, že pacienti s Downovým syndromem mají vyšší pravděpodobnost rozvoje poruchy štítné žlázy, přičemž tato souvislost již byla zkoumána (Cebeci et al., 2013).

Vpravo na Obr. 25 je vyobrazena síť se vztahy týkajícími se především obezity a následků hyperalimentace. Směrem k obezitě se projevují hyperalimentace, menopauzální poruchy, poruchy spánku a Cushingův syndrom. Obezita pak může působit nepříznivě na rozvoj ischemické choroby srdeční, hypertenze, diabetu a fibrilace a flutteru síní. Podíváme-li se na diagnózu týkající se hyperalimentace, vidíme, že pravděpodobně může být přítěžujícím faktorem pro rozvoj diabetu, fibrilace a flutteru síní, poruch metabolismu lipoproteinů a ischemické choroby srdeční. Prezentované vztahy se opět zdají být logickými a v souladu se současnými medicínskými znalostmi.

Nakonec Obr. 26 na následující straně prezentuje spojitosti vybraných autoimunitních onemocnění ke zhoubným novotvarům. Mezi konkrétní autoimunitní choroby, které jsou v síti prezentovány, patří hemolytické anemie, zánět štítné žlázy, zánětlivá polyneuropatie, alopecie a ulcerózní kolitida (vředový zánět tračníku). V síti jsou zobrazeny vztahy s  $CIF > 0,01$  a nepromítly se do ní tak asociace k dalším autoimunitním chorobám, jako je například revmatoidní artritida či psoriáza. Obecně je třeba vzít na vědomí, že se jedná už o poměrně slabé vztahy vizualizované spíše pro zajímavost.

Nahoře na Obr. 26 můžeme pozorovat souvislost mezi zánětem štítné žlázy a zhoubným novotvarem prsu a štítné žlázy. Itoh et al., 1975 v případové studii došli k závěru, který naznačoval, že pacienti s Hashimotovou tyroiditidou jsou rizikovou populací pro rakovinu prsu. Giani et al., 1996 zjistili, že celková prevalence poruch štítné žlázy je u pacientů s karcinomem prsu zvýšena a autoimunitní onemocnění štítné žlázy, zejména Hashimotova tyroiditida, ve značné míře ovlivňují zvýšenou prevalenci onemocnění štítné žlázy u pacientů s rakovinou prsu. Chen et al., 2013 došli k závěru, že pacienti trpící Hashimotovou tyroididou mají vyšší riziko karcinomu štítné žlázy a kolorektálního karcinomu.

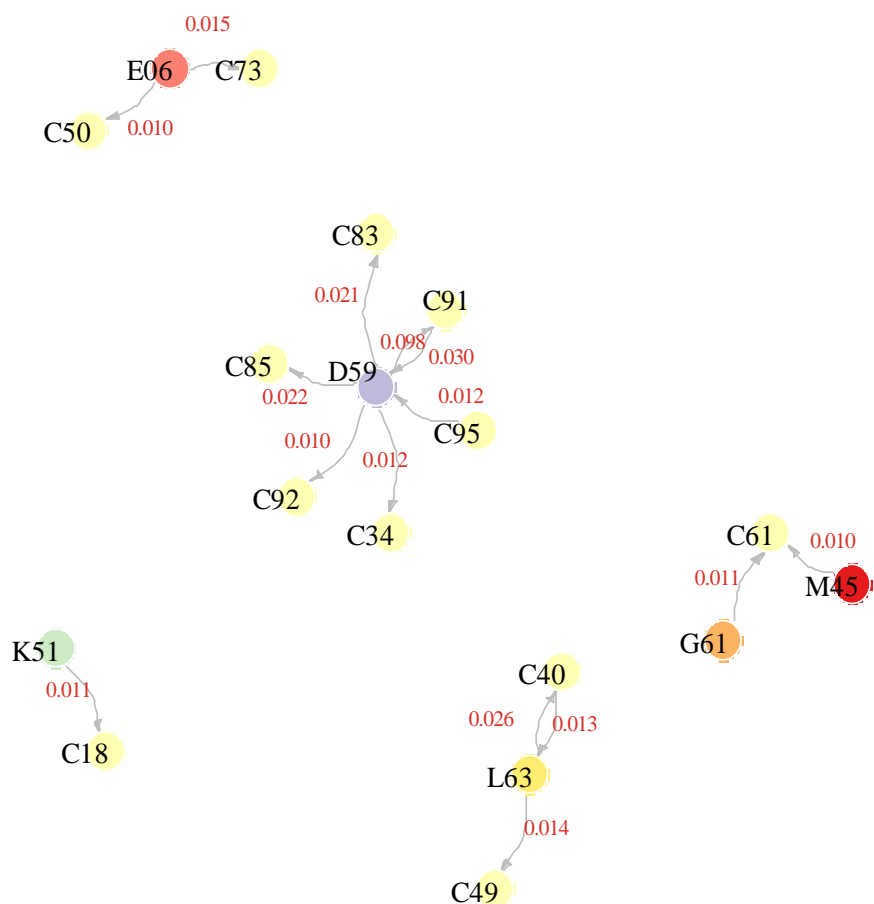
Dále pozorujeme pravděpodobný vliv ulcerózní kolitidy na zhoubný novotvar tlustého střeva. Toto spojení již bylo dokumentováno (Lakatos et al., 2008).

Dalším zobrazeným vztahem je možná souvislost mezi zánětlivými polyneuropatiemi, ankylozující spondylitidou a zhoubným novotvarem prostaty, přičemž zvýšené riziko zhoubného novotvaru prostaty u mužským pacientů trpících ankylozující spondylitidou již bylo zkoumáno (Chang et al., 2017).

Dále si lze všimnout vztahů alopecie ke zhoubným novotvarům kostí a pojivové tkáně. Vztah alopecie ke zhoubným novotvarům kostí je obousměrný.

Největší shluk se pak týká hemolytických anemií. Především vztah autoimunitní hemolytické anemie a lymfoproliferativních onemocnění byl již předmětem zájmu (Sallah et al., 2001).





Zhoubný novotvar tlustého střeva (C18)

Zhoubný novotvar průdušky – bronchu a plice (C34)

Zhoubný novotvar kosti a kloubní chrupavky končetin (C40)

Zhoubný novotvar jiné pojivové a měkké tkáně (C49)

Zhoubný novotvar prsu (C50)

Zhoubný novotvar předstojné žlázy – prostaty (C61)

Zhoubný novotvar štítné žlázy (C73)

Non-folikulární lymfom (C83)

Non-Hodgkinův lymfom, jiných a neurčených typů (C85)

Lymfoidní leukemie (C91)

Myeloidní leukemie (C92)

Leukemie neurčeného buněčného typu (C95)

Získané hemolytické anemie (D59)

Zánět štítné žlázy – thyroiditis (E06)

Zánětlivá polyneuropatie (G61)

Vředový zánět tračníku (K51)

Plešatost – alopecie (L63)

Ankylozující spondylitida (M45)

**Obr. 26** Hemolytické anemie, zánět štítné žlázy, zánětlivá polyneuropatie, alopecie a ulcerózní kolitida ve vztahu ke zhoubným novotvarům

## 7. Diskuze

V praktické části práce byla zkonstruována orientovaná síť humánních fenotypů pro vybrané vztahy na třech úrovních klasifikace MKN-10. Pro vyjádření síly vztahů byla využita varianta CIF indexu ze studie od Kannan et al., 2016, přičemž zkoumaná populace zahrnovala 4 437 854 pacientů.

Síť s vizualizací spojitostí na úrovni kapitol byla však příliš obecná a sloužila tak spíše jako ilustrace, jak může síť s orientovanými vztahy vypadat. Kapitoly MKN-10 totiž obsahují velké množství rozmanitých diagnóz a snaha interpretovat síť na takto obecné úrovni postrádá smysl. Při konstrukci sítě na úrovni oddílů MKN-10 již lze pozorovat konkrétnější souvislosti, jelikož se jedná o poměrně homogenní skupiny sdružující onemocnění členěné především dle jejich lokalizace a chování. Po prvotní vizualizaci celé sítě bylo blíže zobrazeno několik vybraných vztahů. Na základě této sítě a literatury pak byly také vizualizovány vztahy na podrobnější úrovni tříznakových diagnóz. Pro konstrukci sítě na úrovni oddílů a tříznakových diagnóz byly vynechány poslední tři kapitoly MKN-10, a to vzhledem k jejich povaze, případně nízkým prevalencím.

Při velkém rozdílu mezi prevalencemi je také vhodné výsledný vztah kriticky zhodnotit, jelikož například pro vztahy jako jsou specifická onemocnění srdce a často vykazované symptomy jako hypertenze, CIF spojuje první k druhému, přestože pořadí výskytu typicky bývá opačné. Tato situace nastává proto, že počet pacientů s hypertenzí je mnohem vyšší než pro všechna specifická srdeční onemocnění, a jedná se pak o artefakt.

Zejména na úrovni třímístných diagnóz byly vizualizované vztahy objeveny spíše manuálně. Při velkém množství diagnóz a možných kombinací činilo problém stanovení cut-off tak, aby došlo k zobrazení jen omezeného množství zajímavých vztahů. Kannan et al., 2016 také poznamenávají, že ty nejzajímavější vztahy našli spíše manuálně než za pomoci algoritmu.

Mezi zjištěné vztahy na úrovni oddílů patří například neurotické a somatoformní poruchy a poruchy způsobené psychoaktivními látkami, pravděpodobně diagnostikované poté, co pacienti byli hospitalizováni pro otravu léčivými a návykovými biologickými látkami. Dále bylo zjištěno, že mentální retardaci pravděpodobně mohou předcházet vrozené vady nervové soustavy, abnormality chromozomů a poruchy psychického vývoje. Mentální retardace pak pravděpodobně může předcházet například mozkové obrně.

Při zkoumání spojitostí pro novotvary prsu, pohlavních orgánů a trávicího ústrojí byly zjištěny vztahy naznačující, že zhoubným novotvarům mohou předcházet nezhoubné.

Při vizualizaci vybraných spojení s ischemickými srdečními nemocemi se objevily i na pohled zvláštní páry jako onemocnění čočky a glaukom směřující k diagnóze ischemických nemocí srdce. Pomineme-li možné nadhodnocení vztahu vzhledem k rozdílu mezi prevalencemi, souvislost mezi glaukomem a kataraktou s ischemickými srdečními chorobami již byla předmětem zájmu (Hu et al., 2016; Chen et al., 2016). Důvodem tohoto spojení může být mimo jiné skutečnost, že onemocnění očí jsou doprovodnými stavy při diabetu a hypertenzi, přičemž tyto dvě diagnózy jsou považovány za rizikový faktor ischemických srdečních chorob. Je zde tak pravděpodobný vliv zavádějících faktorů.

Mezi první zkoumané souvislosti na úrovni tříznakových diagnóz patřily spojitosti mezi lymfomy a aplastickými (a jinými) anemiemi a poruchami koagulace. Zjištěné vztahy byly víceméně v souladu se sítí, kterou zkonstruovali Kannan et al., 2016. Projevy anemie u pacientů s lymfomy však byly pozorovány již dříve (Moulet et al., 1998; Birgegård, 2008).

Při zkoumání souvislostí mezi systémovými onemocněními pojivové tkáně a zánětlivých polyartropatií s ischemickou chorobou srdeční se ukázaly spojnice naznačující, že systémová onemocnění pojivové tkáně i zánětlivé polyartropatie pravděpodobně předchází ischemické chorobě srdeční, přičemž se těmito vztahy různé studie již také zabíraly (Lazzerini et al., 2006; Del Rincón et al., 2007; Crowson et al., 2013). Například pro revmatoidní artritidu představují ischemické nemoci srdeční jednu z nejčastějších příčin úmrtí (Kaplan, 2010).

Mezi zkoumanými souvislostmi byly také vybrané vztahy pro poruchy štítné žlázy, obezitu a jiné hyperalimentace. Projevila se například spojitost naznačující, že pacienti s Downovým syndromem by mohli mít pravděpodobně později diagnostikovány poruchy štítné žlázy. Touto problematikou se ve své studii zabývali například i Cebeci et al., 2013.

Pozorován byl také vztah mezi zánětem štítné žlázy a zhoubným novotvarem štítné žlázy a prsu. Chen et al. v roce 2013 ve své publikaci došli k závěru, že pacienti trpící Hashimotovou tyroiditidou (autoimunitní zánět štítné žlázy) mají vyšší riziko karcinomu štítné žlázy a kolorektálního karcinomu. Itoh et al. v roce 1975 na základě případové studie došli k závěru naznačujícímu, že pacienti s Hashimotovou tyroiditidou jsou rizikovou populací pro rakovinu prsu.

Nakonec byly prezentovány spojitosti vybraných autoimunitních chorob k novotvarům. Jednalo se však už obecně o poměrně slabé vztahy znázorněné spíše pro zajímavost. Ukázala se zde ale například i trajektorie vedoucí od ulcerózní kolitidy ke zhoubnému novotvaru tlustého střeva, přičemž tento vztah rozebírali detailně v roce 2008 Lakatos et al.

Výsledky prezentované v práci byly v souladu s literárními zdroji, což potvrzuje validitu dat a jejich možnou využitelnost pro další analýzy. Je nutné však zmínit i limitace dat. Je třeba si uvědomit, že se jedná o hospitalizační záznamy pacientů a některá onemocnění (typicky například hypertenze, diabetes) hospitalizaci nutně nevyžadují a poprvé se tak v historii pacientových diagnóz mohou objevit až při hospitalizaci se závažnějším onemocněním. Do vztahů se tedy mohou promítnout jako následek i v případě, že jimi pacient trpěl dříve než před samotnou chorobou, která byla skutečným důvodem hospitalizace. Řešením by mohl být nový datový zdroj NRHZS, který obsahuje i ambulantní záznamy.

## 8. Závěr

Práce se zabývá hledáním orientovaných vztahů v síti humánních fenotypů. Podkladem pro praktickou část byly administrativní záznamy s informací o hospitalizacích pacientů, konkrétně data pocházela z Národního registru hospitalizovaných za období 2007–2016.

Úlohou práce bylo mimo jiné vypracovat literární rešerši zabývající se hledáním orientovaných vztahů z administrativních zdravotních záznamů. Tento přehled zahrnující jak studie zabývající se vhodností administrativních dat pro hledání příčinných vztahů, tak studie přímo aplikující vybrané metody na administrativní data za účelem získat orientované vztahy se nachází v kapitole 2.

Kapitola 3 se věnuje Národnímu zdravotnickému informačnímu systému a blíže se zaměřuje na Národní registr hospitalizovaných, jelikož pochopení jeho struktury bylo nezbytné pro práci s daty v praktické části. Součástí této kapitoly je také stručné představení MKN-10 jakožto klasifikace, ve které jsou kódovány veškeré vykázané diagnózy nacházející se v datovém souboru.

V kapitole 4 je nastíněna problematika teorie grafů, network science a network medicine stojící v pozadí. Dále na ni navazuje kapitola 5, kde je rozebrána síť humánních fenotypů, a to včetně orientovaných vztahů.

Kapitola 6 je již věnována praktické části práce. Na začátku je popsán proces čištění souboru a tvorby kombinací, po popisu základních charakteristik následuje prvotní náhled na data ve formě sítě humánních fenotypů. Následně byla konstruována síť humánních fenotypů již se zahrnutím orientovaných vztahů. Vzhledem k velkému počtu možných vztahů bylo prezentováno jen několik vybraných. Výběr zobrazených vztahů probíhal na základě konstrukce obecnější sítě a s ohledem na literaturu. Diskuzí zjištěných vztahů a použitých metod se zabývá kapitola 7.

Práce vyžadovala zvládnutí velkých dat NRHOSP a následnou aplikaci výpočetně náročných algoritmů na tato velká data. Zkonstruované sítě jsou v souladu s literaturou a i přes limitace administrativních dat NRHOSP tak lze říci, že jsou pro tento typ analýz využitelná. Dalším krokem by mohla být analýza provedená na datech NRHZS, jelikož ta obsahují i záznamy o ambulantní péči. Tento datový zdroj je však dostupný teprve od letošního roku.

V budoucnu by mohlo být také zajímavé zaměřit se na konstrukci sítě s ohledem na pohlaví, jelikož je známo, že některé faktory mohou být rizikovější pro jedno pohlaví více než pro druhé. Stejně tak by mohla být zajímavá stratifikace dle věku.

Na závěr je potřeba opět zdůraznit, že ze zjištěných vztahů nelze činit přímé závěry o kauzalitě a časová asociace neznamena nutně příčinnou souvislost, jelikož se v pozadí může nacházet řada zavádějících faktorů. I když je výsledná míra asociace silná, nemusí to tak nutně znamenat, že jedna nemoc zapříčiňuje druhou, protože se v pozadí mohou nacházet společné příčiny obou chorob, které se projevíly dřívější manifestací jedné z nich.

## Seznam použité literatury

ABDEL-SALAM, Ghada M H, Adrienne A HALÁSZ a Andrew E CZEIZEL, 2000. Association of epilepsy with different groups of microcephaly. *Developmental Medicine & Child Neurology* [online]. 42(11), 760-767 [cit. 2018-05-05]. DOI: 10.1111/j.1469-8749.2000.tb00040.x. ISSN 00121622. Dostupné z: <http://doi.wiley.com/10.1111/j.1469-8749.2000.tb00040.x>

AGUSTI, Alvar, Patricia SOBRADILLO a Bartolomé CELLI, 2011. Addressing the Complexity of Chronic Obstructive Pulmonary Disease. *American Journal of Respiratory and Critical Care Medicine* [online]. 183(9), 1129-1137 [cit. 2018-01-07]. DOI: 10.1164/rccm.201009-1414PP. ISSN 1073-449x. Dostupné z: <http://www.atsjournals.org/doi/abs/10.1164/rccm.201009-1414PP>

AL HUSAIN, Awal a Ian N. BRUCE, 2010. Risk factors for coronary heart disease in connective tissue diseases. *Therapeutic Advances in Musculoskeletal Disease* [online]. 2(3), 145-153 [cit. 2018-05-04]. DOI: 10.1177/1759720X10365301. ISSN 1759-720X. Dostupné z: <http://journals.sagepub.com/doi/10.1177/1759720X10365301>

BARABÁSI, Albert-László, 2007. Network Medicine — From Obesity to the “Diseasome”. *New England Journal of Medicine* [online]. 357(4), 404-407 [cit. 2017-12-09]. DOI: 10.1056/NEJMe078114. ISSN 0028-4793. Dostupné z: <http://www.nejm.org/doi/abs/10.1056/NEJMe078114>

BARABÁSI, Albert-László, 2012. Network science [online]. [cit. 2017-12-09]. Dostupné z: <http://networksciencebook.com>

BECK, Mette K., Anders Boeck JENSEN, Annelaura Bach NIELSEN, Anders PERNER, Pope L. MOSELEY a Søren BRUNAK, 2016. Diagnosis trajectories of prior multi-morbidity predict sepsis mortality. *Scientific Reports* [online]. 6(1), - [cit. 2018-04-12]. DOI: 10.1038/srep36624. ISSN 2045-2322. Dostupné z: <http://www.nature.com/articles/srep36624>

BIRGEGÅRD, G. 2008. Managing anemia in lymphoma and multiple myeloma. *Therapeutics and clinical risk management*, 4(2), 527.

BONITA, Ruth, Robert BEAGLEHOLE a Tord KJELLSTRÖM, 2006. *Basic epidemiology*. 2nd ed. Geneva: World Health Organization. ISBN 978-92-4-154707-9.

CEBECI AYŞE, Nurcan, Güven AYLA a Yıldız METIN, 2013. Profile of Hypothyroidism in Down's Syndrome. *Journal of Clinical Research in Pediatric Endocrinology* [online]. 5(2), 116-120 [cit. 2018-05-05]. DOI: 10.4274/Jcrpe.884. ISSN 13085727. Dostupné z: [http://cms.galenos.com.tr/Uploads/Article\\_6495/116-120.pdf](http://cms.galenos.com.tr/Uploads/Article_6495/116-120.pdf)

CROWSON, Cynthia S., Katherine P. LIAO, John M. DAVIS, Daniel H. SOLOMON, Eric L. MATTESON, Keith L. KNUTSON, Mark A. HLATKY a Sherine E. GABRIEL, 2013. Rheumatoid arthritis and cardiovascular disease. *American Heart Journal* [online]. 166(4), 622-628.e1 [cit. 2018-05-03]. DOI: 10.1016/j.ahj.2013.07.010. ISSN 00028703. Dostupné z: <http://linkinghub.elsevier.com/retrieve/pii/S0002870313004481>

CSARDI G a T NEPUSZ: The igraph software package for complex network research, InterJournal, Complex Systems 1695. 2006. <http://igraph.org>

DAHLEM, Dominik, Diego MANILOFF a Carlo RATTI, 2015. Predictability Bounds of Electronic Health Records. Scientific Reports [online]. 5(1), - [cit. 2018-04-13]. DOI: 10.1038/srep11865. ISSN 2045-2322. Dostupné z: <http://www.nature.com/articles/srep11865>

DEL RINCÓN I., K. WILLIAMS, M. P. STERN, G. L. FREEMAN a A. ESCALANTE. 2001. High incidence of cardiovascular events in a rheumatoid arthritis cohort not explained by traditional cardiac risk factors. Arthritis & Rheumatology, 44(12), 2737-2745.

DEMEL, Jiří, 2002. Grafy a jejich aplikace. 1. vyd. Praha: Academia. ISBN 80-200-0990-6.

DRESSLEROVÁ, I a J VOJÁČEK, 2010. Diabetes mellitus a ischemická choroba srdeční. Vnitř. Lék. 56(4): 301-306.

FERVER, Kari, Bryan BURTON a Paul JESILOW, 2009. The Use of Claims Data in Healthcare Research. The Open Public Health Journal [online]. 2(1), 11-24 [cit. 2018-03-03]. DOI: 10.2174/1874944500902010011. ISSN 18749445. Dostupné z: <http://benthamopen.com/ABSTRACT/TOPHJ-2-11>

FOLTÝNEK, Tomáš a Jana DANNHOFEROVÁ, 2011. Teorie grafů. Brno: Mendelova univerzita v Brně. ISBN 978-80-7375-500-3.

GIANI, C, P FIERABRACCI, R BONACCI, et al., 1996. Relationship between breast cancer and thyroid disease: relevance of autoimmune thyroid disorders in breast malignancy. The Journal of Clinical Endocrinology & Metabolism [online]. 81(3), 990-994 [cit. 2018-05-06]. DOI: 10.1210/jcem.81.3.8772562. ISSN 0021-972X. Dostupné z: <https://academic.oup.com/jcem/article-lookup/doi/10.1210/jcem.81.3.8772562>

GRABER, Cynthia, The Diseaseome Could Take Medicine Beyond the Genome [online]. 2014 [cit. 2018-01-26]. Dostupné z: <http://www.pbs.org/wgbh/nova/next/body/network-medicine/>

GROZA, Tudor, Sebastian KÖHLER, Dawid MOLDENHAUER, et al., 2015. The Human Phenotype Ontology: Semantic Unification of Common and Rare Disease. The American Journal of Human Genetics [online]. 97(1), 111-124 [cit. 2018-02-22]. DOI: 10.1016/j.ajhg.2015.05.020. ISSN 00029297. Dostupné z: <http://linkinghub.elsevier.com/retrieve/pii/S0002929715002347>

HABALA, Petr, 2012. Diskrétní matematika: Grafy [online]. [cit. 2017-12-29]. Dostupné z: <https://math.feld.cvut.cz/habala/teaching/dma/dmknih12.pdf>

HARUŠTIAKOVÁ, Danka, Jiří JARKOVSKÝ, Simona LITTNEROVÁ a Ladislav DUŠEK. 2012. Vícerozměrné statistické metody v biologii. Vyd. 1. Brno: Akademické nakladatelství CERM. ISBN 9788072047918.

HIDALGO, César A., Nicholas BLUMM, Albert-László BARABÁSI, Nicholas A. CHRISTAKIS a Lauren Ancel MEYERS, 2009. A Dynamic Network Approach for the Study of Human Phenotypes. PLoS Computational Biology [online]. 5(4), e1000353- [cit. 2018-02-19]. DOI:

10.1371/journal.pcbi.1000353. ISSN 15537358. Dostupné z: <http://dx.plos.org/10.1371/journal.pcbi.1000353>

HILL, Austin Bradford, 1965. The Environment and Disease: Association or Causation?. *Proceedings of the Royal Society of Medicine*. 58(5), 295-300.

HIMES, B. E., Y. DAI, I. S. KOHANE, S. T. WEISS a M. F. RAMONI, 2009. Prediction of Chronic Obstructive Pulmonary Disease (COPD) in Asthma Patients Using Electronic Medical Records. *Journal of the American Medical Informatics Association* [online]. 16(3), 371-379 [cit. 2018-04-12]. DOI: 10.1197/jamia.M2846. ISSN 1067-5027. Dostupné z: <https://academic.oup.com/jamia/article-lookup/doi/10.1197/jamia.M2846>

HOEHNDORF, Robert, Paul N. SCHOFIELD a Georgios V. GKOUTOS, 2015. Analysis of the human diseaseome using phenotype similarity between common, genetic, and infectious diseases. *Scientific Reports* [online]. 5 [cit. 2018-02-22]. DOI: 10.1038/srep10888. ISSN 20452322. Dostupné z: <http://www.nature.com/articles/srep10888>

HORDĚJČUK, Vojtěch, 2017. Teorie grafů [online]. [cit. 2017-12-27]. Dostupné z: <http://voho.eu/wiki/graf/>

HRIPCSAK, George, David J ALBERS a Adler PEROTTE, 2011. Exploiting time in electronic health record correlations. *Journal of the American Medical Informatics Association* [online]. 18(Supplement\_1), i109-i115 [cit. 2018-04-12]. DOI: 10.1136/amiajnl-2011-000463. ISSN 1527-974X. Dostupné z: <https://academic.oup.com/jamia/article-lookup/doi/10.1136/amiajnl-2011-000463>

HU, Wei-Syun, Cheng-Li LIN, Shih-Sheng CHANG, Ming-Fong CHEN a Kuan-Cheng CHANG, 2016. Increased risk of ischemic heart disease among subjects with cataracts. *Medicine* [online]. 95(28), e4119- [cit. 2018-05-01]. DOI: 10.1097/MD.00000000000004119. ISSN 0025-7974. Dostupné z: <http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00005792-201607120-00020>

CHAN, Stephen Y. a Joseph LOSCALZO, 2012. The Emerging Paradigm of Network Medicine in the Study of Human Disease. *Circulation Research* [online]. 111(3), 359-374 [cit. 2018-01-21]. DOI: 10.1161/CIRCRESAHA.111.258541. ISSN 0009-7330. Dostupné z: <http://circres.ahajournals.org/cgi/doi/10.1161/CIRCRESAHA.111.258541>

CHANG, Chih-Cheng, Cheng-Wei CHANG, Phung-Anh Alex NGUYEN, et al., 2017. Ankylosing spondylitis and the risk of cancer. *Oncology Letters* [online]. 14(2), 1315-1322 [cit. 2018-05-06]. DOI: 10.3892/ol.2017.6368. ISSN 1792-1074. Dostupné z: <https://www.spandidos-publications.com/10.3892/ol.2017.6368>

CHEN, L L, N BLUMM, N A CHRISTAKIS, A-L BARABÁSI a T S DEISBOECK, 2009. Cancer metastasis networks and the prediction of progression patterns. *British Journal of Cancer* [online]. 101(5), 749-758 [cit. 2018-04-12]. DOI: 10.1038/sj.bjc.6605214. ISSN 0007-0920. Dostupné z: <http://www.nature.com/articles/6605214>



CHEN, Y-K, C-L LIN, F T-F CHENG, F-C SUNG a C-H KAO, 2013. Cancer risk in patients with Hashimoto's thyroiditis: a nationwide cohort study. *British Journal of Cancer* [online]. 109(9), 2496-2501 [cit. 2018-05-06]. DOI: 10.1038/bjc.2013.597. ISSN 0007-0920. Dostupné z: <http://www.nature.com/articles/bjc2013597>

CHEN, Yu-Yen, Hsiao-Yun HU, Dachen CHU, Hsin-Hua CHEN, Chin-Kuo CHANG, Pesus CHOU a Tomohiko AI, 2016. Patients with Primary Open-Angle Glaucoma May Develop Ischemic Heart Disease More Often than Those without Glaucoma: An 11-Year Population-Based Cohort Study. *PLOS ONE* [online]. 11(9), e0163210- [cit. 2018-05-01]. DOI: 10.1371/journal.pone.0163210. ISSN 1932-6203. Dostupné z: <http://dx.plos.org/10.1371/journal.pone.0163210>

CHEN, Weiqi, Yong HU, Xiangzhou ZHANG, Lijuan WU, Jianqin HE, Zilin TANG, Xing SONG, Lemuel R. WAITMAN a Mei LIU, 2018. Causal risk factor discovery for severe acute kidney injury using electronic health records. *BMC medical informatics and decision making* [online]. 22 March 2018, 18(1), 13 [cit. 2018-04-12]. DOI: 10.1186/s12911-018-0597-7. Dostupné z: <https://bmcmmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-018-0597-7>

CHO, Dong-Yeon, Yoo-Ah KIM a Teresa M. PRZYTYCKA, 2012. Chapter 5: Network Biology Approach to Complex Diseases. *PLOS Computational Biology* [online]. 8(12) [cit. 2018-01-06]. e1002820. Dostupné z: <https://doi.org/10.1371/journal.pcbi.1002820>

CHOI, H. K. a G. CURHAN, 2007. Independent Impact of Gout on Mortality and Risk for Coronary Heart Disease. *Circulation* [online]. 116(8), 894-900 [cit. 2018-05-03]. DOI: 10.1161/CIRCULATIONAHA.107.703389. ISSN 0009-7322. Dostupné z: <http://circ.ahajournals.org/cgi/doi/10.1161/CIRCULATIONAHA.107.703389>

ITOH, Kunihiko a Nobuhiro MARUCHI, 1975. NREAST CANCER IN PATIENTS WITH HASHIMOTO'S THYROIDITIS. *The Lancet* [online]. 306(7945), 1119-1121 [cit. 2018-05-06]. DOI: 10.1016/S0140-6736(75)91006-5. ISSN 01406736. Dostupné z: <http://linkinghub.elsevier.com/retrieve/pii/S0140673675910065>

JANOŠOVÁ, Eva, Danka HARUŠTIAKOVÁ, Simona LITTNEROVÁ a Jiří JARKOVSKÝ, 2015. Vícerozměrné metody pro analýzu a klasifikaci dat: Metriky pro určení podobnosti mezi dvěma obrazy s kvalitativními hodnotami souřadnic. In: HOLČÍK, Jiří, Martin KOMENDA (eds.) a kol. *Matematická biologie: e-learningová učebnice* [online]. Brno: Masarykova univerzita [cit. 2018-03-22]. ISBN 978-80-210-8095-9. Dostupné z: <http://portal.matematickabiologie.cz/index.php?pg=analyza-a-hodnoceni-biologickych-dat--vicerozmerne-metody-pro-analyzu-dat--podobnosti-a-vzdalenosti-ve-vicerozmerne-prostor--metriky-pro-urceni-vzdalenosti-a-podobnosti-mezi-dvema-vektory--metriky-pro-urceni-podobnosti-mezi-dvema-obrazy-s-kvalitativni-mi-hodnotami-souradnic>

JANOUT, Vladimír, 1998. *Klinická epidemiologie-nedílná součást klinických rozhodovacích procesů*. Praha: Grada. ISBN 80-7169-353-7.

JENSEN, Anders Boeck, Pope L. MOSELEY, Tudor I. OPREA, Sabrina Gade ELLESØE, Robert ERIKSSON, Henriette SCHMOCK, Peter BJØDSTRUP, Lars Juhl JENSEN a Søren BRUNAK, 2014. Temporal disease trajectories condensed from population-wide registry data covering 6.2



million patients. *Nature Communications* [online]. 5, - [cit. 2018-04-12]. DOI: 10.1038/ncomms5022. ISSN 2041-1723. Dostupné z: <http://www.nature.com/doifinder/10.1038/ncomms5022>

KANNAN, Venkateshan, Fredrik SWARTZ, Narsis A. KIANI, Gilad SILBERBERG, Giorgos TSIPRAS, David GOMEZ-CABRERO, Kristina ALEXANDERSON a Jesper TEGNÉR, 2016. Conditional Disease Development extracted from Longitudinal Health Care Cohort Data using Layered Network Construction. *Scientific Reports* [online]. 6 [cit. 2018-03-17]. DOI: 10.1038/srep26170. ISSN 2045-2322. Dostupné z: <http://www.nature.com/articles/srep26170>

KAPLAN, Mariana J., 2010. Cardiovascular Complications of Rheumatoid Arthritis: Assessment, Prevention, and Treatment. *Rheumatic Disease Clinics of North America* [online]. 36(2), 405-426 [cit. 2018-05-01]. DOI: 10.1016/j.rdc.2010.02.002. ISSN 0889857X. Dostupné z: <http://linkinghub.elsevier.com/retrieve/pii/S0889857X10000141>

KOLÁŘ, Josef, 2000. *Teoretická informatika*. 2. vyd. Praha: Česká informatická společnost, 204 s. ISBN 8090085385.

KORCSMAROS, Tamas, Maria Victoria SCHNEIDER a Giulio SUPERTI-FURGA, 2017. Next generation of network medicine: interdisciplinary signaling approaches. *Integrative Biology* [online]. 9(2), 97-108 [cit. 2018-01-26]. DOI: 10.1039/C6IB00215C. ISSN 1757-9694. Dostupné z: <http://xlink.rsc.org/?DOI=C6IB00215C>

KUBÁSEK, Miroslav, 2015. Teoretické základy informatiky: Teorie grafů. In: HOLČÍK, Jiří, Martin KOMENDA (eds.) a kol. *Matematická biologie: e-learningová učebnice* [online]. 1. vyd. Brno: Masarykova univerzita [cit. 2017-12-27]. ISBN 978-80-210-8095-9. Dostupné z: <http://portal.matematickabiologie.cz/index.php?pg=zaklady-informatiky-pro-biology--teoreticke-zaklady-informatiky>

KUNDI, Michael, 2006. Causality and the Interpretation of Epidemiologic Evidence. *Environmental Health Perspectives* [online]. 114(7), 969-974 [cit. 2018-03-25]. DOI: 10.1289/ehp.8297. ISSN 0091-6765. Dostupné z: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1513293/>

LAKATOS, Peter Laszlo a Laszlo LAKATOS, 2008. Risk for colorectal cancer in ulcerative colitis: Changes, causes and management strategies. *World Journal of Gastroenterology* [online]. 14(25), 3937- [cit. 2018-05-06]. DOI: 10.3748/wjg.14.3937. ISSN 1007-9327. Dostupné z: <http://www.wjgnet.com/1007-9327/full/v14/i25/3937.htm>

LAZZERINI, Pietro Enea, Pier Leopoldo CAPECCHI, Francesca GUIDERI, Maurizio ACAMPA, Mauro GALEAZZI a Franco LAGHI PASINI, 2006. Connective tissue diseases and cardiac rhythm disorders: An overview. *Autoimmunity Reviews* [online]. 5(5), 306-313 [cit. 2018-05-04]. DOI: 10.1016/j.autrev.2005.11.002. ISSN 15689972. Dostupné z: <http://linkinghub.elsevier.com/retrieve/pii/S1568997205002119>

LEE, Choong Ho a Hyung-Jin YOON, 2017. Medical big data: promise and challenges. *Kidney Research and Clinical Practice* [online]. 36(1), 3-11 [cit. 2018-04-08]. DOI:

10.23876/j.krcp.2017.36.1.3. ISSN 2211-9140. Dostupné z: <http://www.krcp-ksn.org/journal/view.html?doi=10.23876/j.krcp.2017.36.1.3>

LEGENDRE, Pierre a Louis LEGENDRE, 1998. Numerical ecology. Vyd. 2. Amsterdam: Elsevier, 853 s. Developments in environmental modelling. ISBN 0444892508.

LISTL, Stefan, Hendrik JÜRGES a Richard G. WATT, 2016. Causal inference from observational data. Community Dentistry and Oral Epidemiology [online]. 44(5), 409-415 [cit. 2018-04-08]. DOI: 10.1111/cdoe.12231. ISSN 03015661. Dostupné z: <http://doi.wiley.com/10.1111/cdoe.12231>

MACMAHON, B. a T. F. PUGH. Epidemiology, principles and methods. Little Brown, Boston, p I. Zitiert in der Übersetzung von Pflanz M (1973) Allgemeine Epidemiologie. 1970.

MIOTTO, Riccardo, Li LI, Brian A. KIDD a Joel T. DUDLEY, 2016. Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. Scientific Reports [online]. 6(1), - [cit. 2018-04-13]. DOI: 10.1038/srep26094. ISSN 2045-2322. Dostupné z: <http://www.nature.com/articles/srep26094>

MOULLET, I., G. SALLES, N. KETTERER, C. DUMONTET, F. BOUAFIA., E. M. NEIDHART-BERARD, C. THIEBLEMONT, P. FELMAN a B. COIFFIER. (1998). Frequency and significance of anemia in non-Hodgkin's lymphoma patients. Annals of oncology, 9(10), 1109-1115.

NEŠETŘIL, Jaroslav, 1979. Teorie grafů. Praha: SNTL – Nakladatelství technické literatury. Matematický seminář SNTL.

PARASCANDOLA, M. a D. WEED, 2001. Causation in epidemiology. Journal of Epidemiology & Community Health [online]. 55(12), 905-912 [cit. 2018-04-05]. DOI: 10.1136/jech.55.12.905. ISSN 0143005X. Dostupné z: <http://doi.org/10.1136/jech.55.12.905>

PAVLÍK, Tomáš a Ladislav DUŠEK, 2012. Biostatistika. Brno: Akademické nakladatelství CERM. ISBN 978-80-7204-782-6.

PAVLÍK, Tomáš, Jiří JARCOVSKÝ a Ladislav DUŠEK. 2015. Metodika vícerozměrné analýzy Národního registru hospitalizovaných za účelem výběru reprezentativní sítě poskytovatelů zdravotních služeb ČR [online]. Ústav zdravotnických informací a statistiky České republiky, Institut biostatistiky a analýz Masarykovy univerzity [cit. 2017-10-10]. Dostupné z: <http://www.drg-cz.cz/res/file/metodiky/zz-metodika-typologie.pdf>

R CORE TEAM (2018): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at <http://www.Rproject.org/>.

RILEY, Gerald F., 2009. Administrative and Claims Records as Sources of Health Care Cost Data. Medical Care [online]. 47(Supplement), S51-S55 [cit. 2018-03-03]. DOI: 10.1097/MLR.0b013e31819c95aa. ISSN 00257079. Dostupné z: <http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00005650-200907001-00009>

ROQUE, Francisco S., Peter B. JENSEN, Henriette SCHMOCK, Marlene DALGAARD,

Massimo ANDREATTA, Thomas HANSEN, Karen SØEBY, Søren BREDKJÆR, Anders JUUL, Thomas WERGE, Lars K. JENSEN, Søren BRUNAK a Marylyn D. RITCHIE. 2011. Using Electronic Patient Records to Discover Disease Correlations and Stratify Patient Cohorts. *PLoS Computational Biology* [online]. 7(8) [cit. 2018-02-22]. DOI: 10.1371/journal.pcbi.1002141. ISSN 15537358. Dostupné z: <http://dx.plos.org/10.1371/journal.pcbi.1002141>

SALLAH, S., J. Y. WAN. a L. R. HANRAHAN. 2001. Future development of lymphoproliferative disorders in patients with autoimmune hemolytic anemia. *Clinical cancer research*, 7(4), 791-794.

SCHULMAN, Kathy L., Karina BERENSON, Ya-Chen (TINA) SHIH, Kathleen A. FOLEY, Arijit GANGULI, Jonas DE SOUZA, Nicholas A. YAGHMOUR a Alex SHTEYNŠHLYUGER, 2013. A Checklist for Ascertaining Study Cohorts in Oncology Health Services Research Using Secondary Data: Report of the ISPOR Oncology Good Outcomes Research Practices Working Group. *Value in Health* [online]. 16(4), 655-669 [cit. 2018-03-03]. DOI: 10.1016/j.jval.2013.02.006. ISSN 10983015. Dostupné z: <http://linkinghub.elsevier.com/retrieve/pii/S1098301513000612>

STUART, Elizabeth A., Eva DUGOF, Michael ABRAMS, David SALKEVER a Donald STEINWACHS, 2013. Estimating Causal Effects in Observational Studies Using Electronic Health Data: Challenges and (some) Solutions. *EGEMs (Generating Evidence & Methods to improve patient outcomes)* [online]. 1(3) [cit. 2018-04-09]. DOI: 10.13063/2327-9214.1038. ISSN 2327-9214. Dostupné z: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4049166/>

ŠEJDA, Jan, Zdeněk ŠMERHOVSKÝ a Dana GÖPFERTOVÁ, 2005. *Výkladový slovník epidemiologické terminologie*. Praha: Grada. ISBN 80-247-1068-4.

ÚZIS ČR. 2016. Národní registr hospitalizovaných: Závazné pokyny NZIS [online]. Praha: Ústav zdravotnických informací a statistiky ČR [cit. 2017-10-09]. Dostupné z: [http://www.uzis.cz/system/files/zpok\\_hosp\\_020\\_20140701\\_2\\_0.pdf](http://www.uzis.cz/system/files/zpok_hosp_020_20140701_2_0.pdf)

VÍTEČKOVÁ, Miluše, Petr PŘIDAL a Tomáš KOUDELA, 2006. *Základy teorie pravděpodobnosti a teorie grafů* [online]. Ostrava: Vysoká škola báňská, Technická univerzita Ostrava, Fakulta strojní [cit. 2018-01-30]. Dostupné z: <http://books.fs.vsb.cz/SystAnal/texty/download/cele.pdf>

WOODS, C Geoffrey, 2004. Human microcephaly. *Current Opinion in Neurobiology* [online]. 14(1), 112-117 [cit. 2018-05-05]. DOI: 10.1016/j.conb.2004.01.003. ISSN 09594388. Dostupné z: <http://linkinghub.elsevier.com/retrieve/pii/S0959438804000042>

WORLD HEALTH ORGANIZATION, 2001. *Health research methodology: a guide for training in research methods*. 2nd ed. Manila, Philippines: World Health Organization Regional Office for the Western Pacific.

WORLD HEALTH ORGANIZATION. 2014. *Mezinárodní statistická klasifikace nemocí a přidružených zdravotních problémů – Instrukční příručka* [online]. 10. revize, 2. edice. Praha [cit. 2017-10-20]. ISBN 978-80-7280-846-5

WUENSCH, Karl L. 2009. *Correspondence Between  $\Phi$  and Odds Ratios in a 2 x 2 Contingency Table* [online]. East Carolina University [cit. 2018-03-21]. Dostupné z: <http://core.ecu.edu/psyc/wuenschk/StatHelp/Phi-OddsRatio.docx>

YADAV, Pranjul, Lisiane PRUINELLI, Andrew HANGSLEBEN, Sanjoy DEY, Kaherine HAUWILLER, Bonnie L. WESTRA, Connie W. DELANEY, Vipin KUMAR, Michael STEINBACH a Gyorgy J. SIMON. 2015. Modeling Trajectories for Diabetes Complications.

ZANZONI, Andreas, Charles E. CHAPPLE a Christine BRUN, 2015. Relationships between predicted moonlighting proteins, human diseases, and comorbidities from a network perspective. *Frontiers in Physiology* [online]. 6 [cit. 2018-02-22]. DOI: 10.3389/fphys.2015.00171. ISSN 1664042x. Dostupné z: <http://journal.frontiersin.org/Article/10.3389/fphys.2015.00171/abstract>

ZEIDMAN, A., Z. FRADIN., A. BLECHER., H. S OSTER, Y AVRAHAMI a M. MITTELMAN, 2004. Anemia as a risk factor for ischemic heart disease. *Imaj-Ramat Gan-*, 6(1), 16-18.

## Internetové zdroje

URL 1: Informace o NZIS. ÚZIS ČR: Ústav zdravotnických informací a statistiky ČR [online]. Praha [cit. 2017-09-29]. Dostupné z: <http://www.uzis.cz/nas/informace-nzis>

URL 2: Národní zdravotní registry. ÚZIS ČR: Ústav zdravotnických informací a statistiky ČR [online]. Praha [cit. 2017-09-29]. Dostupné z: <http://www.uzis.cz/registry/narodni-zdravotni-registry>

URL 3: Národní registr hospitalizovaných (NRHOSP). ÚZIS ČR: Ústav zdravotnických informací a statistiky ČR [online]. Praha [cit. 2017-10-09]. Dostupné z: <http://www.uzis.cz/registry-nzis/nrhosp>

URL 4: International Classification of Diseases (ICD). 2017. World Health Organization [online]. Geneva [cit. 2017-10-20]. Dostupné z: <http://www.who.int/classifications/icd/en/>

URL 5: MKN-10. ÚZIS ČR: Ústav zdravotnických informací a statistiky ČR [online]. Praha [cit.2017-10-20]. Dostupné z: <http://www.uzis.cz/cz/mkn/index.html>

URL 6: Center for Network Science [online]. Central European University, Budapest, Hungary [cit. 2018-01-02]. Dostupné z: <https://cns.ceu.edu/>

URL 7: Network science [online], [cit. 2018-01-02]. Dostupné z: <http://www.network-science.org/>

URL 8: Network analysis in biology, EMBL-EBI [online]. Wellcome Genome Campus, Hinxton, Cambridgeshire [cit. 2018-01-06]. Dostupné z: <https://www.ebi.ac.uk/training/online/course/network-analysis-protein-interaction-data-introduction/network-analysis-biology-0>

## Seznam použitých zkratk

CIF	Causal Information Fraction
ICHS	Ischemické choroby srdeční
MKN	Mezinárodní klasifikace nemocí
MZ ČR	Ministerstvo zdravotnictví České republiky
NN	Nezhoubný novotvar
NS	Network science
NRHOSP	Národní registr hospitalizovaných
NRHZS	Národní registr hrazených zdravotních služeb
NZIS	Národní zdravotnický informační systém
PDN	Síť humánních fenotypů (Phenotypic Disease Network)
ÚZIS ČR	Ústav zdravotnických informací a statistiky ČR
ZN	Zhoubný novotvar

## Seznam obrázků

Obr. 1	Struktura kódu MKN-10	str. 10
Obr. 2	Příklad neorientovaného a orientovaného grafu	str. 14
Obr. 3	Příklady grafů a jejich vlastností	str. 15
Obr. 4	Síť humánních fenotypů (Hidalgo et al., 2009)	str. 20
Obr. 5	Síť orientovaných vztahů zaměřená na souvislosti výskytu rakoviny a autoimunitních chorob (Kannan et al., 2016)	str. 24
Obr. 6	Schéma přípravy datového souboru	str. 26
Obr. 7	Počet záznamů o hospitalizacích na jednoho pacienta	str. 27
Obr. 8	Počet diagnóz na jednu hospitalizaci	str. 27
Obr. 9	Zastoupení pohlaví	str. 28
Obr. 10	Věková struktura pacientů při poslední hospitalizaci dle pohlaví	str. 29
Obr. 11	Schéma tvorby kombinací	str. 30
Obr. 12	Vývoj velikosti souboru při přípravě a tvorbě kombinací	str. 32
Obr. 13	Síť humánních fenotypů – kapitoly MKN-10	str. 33
Obr. 14	Zastoupení kapitol MKN-10 v hospitalizační historii pacientů	str. 35
Obr. 15	Kumulativní distribuční funkce pro CIF > 0 na úrovni kapitol MKN-10	str. 36
Obr. 16	Orientovaná síť na úrovni kapitol MKN-10	str. 37
Obr. 17	Orientovaná síť pro oddíly MKN-10	str. 38
Obr. 18	Orientované vztahy pro duševní poruchy a poruchy chování	str. 39
Obr. 19	Orientované vztahy pro zhoubné novotvary prsu, ženských a mužských pohlavních orgánů a trávicího ústrojí	str. 40
Obr. 20	Orientované vztahy pro selhání ledvin a nemoci jater	str. 41
Obr. 21	Vybrané vztahy pro ICHS	str. 42
Obr. 22	Orientované vztahy mezi lymfomy a anemiemi/poruchami koagulace	str. 44
Obr. 23	Orientované vztahy mezi systémovými onemocněními pojivové tkáně a zánětlivými polyartropatiemi s ischemickými nemocemi srdečními, případně kardiopulmonálními nemocemi	str. 45
Obr. 24	Vybrané orientované vztahy pro vrozené vady, chromozomální abnormality a nemoci nervové soustavy	str. 46
Obr. 25	Vybrané orientované vztahy pro poruchy štítné žlázy, obezitu a jiné hyperalimentace	str. 47
Obr. 26	Hemolytické anemie, zánět štítné žlázy, zánětlivá polyneuropatie, alopecie a ulcerózní kolitida ve vztahu ke zhoubným novotvarům	str. 49

## Seznam tabulek

Tab. 1	Frekvenční tabulka	str. 18
Tab. 2	Počet diagnóz na hospitalizaci – charakteristiky	str. 28
Tab. 3	Věkové složení pacientů při poslední hospitalizaci	str. 28
Tab. 4	Počet všech kombinací (bez duplikací v pacientově historii) a počet unikátních kombinací na definovaných úrovních klasifikace	str. 32

## Přílohy

### Příloha 1

Vybrané parametry Národního registru hospitalizovaných (Pavlík et al., 2015)

Hospitalizační věta	Pobytová věta	Výkonová věta
<ul style="list-style-type: none"><li>• Identifikační číslo zdravotnického zařízení (ZZ)</li><li>• Druh zdravotnického zařízení</li><li>• Kód oddělení, kód pracoviště</li><li>• Demografické údaje pacienta</li><li>• Kraj, okres, obec bydliště pacienta</li><li>• Datum začátku příznaků</li><li>• Datum a čas přijetí do ZZ</li><li>• Doporučení, způsob a důvod přijetí</li><li>• Základní diagnóza a vnější příčina (MKN-10)</li><li>• Další diagnózy (MKN-10)</li><li>• Datum a čas operace</li><li>• Hlavní operační diagnóza</li><li>• Hlavní operační výkon</li><li>• Reoperace a pooperační komplikace</li><li>• Nemocniční nákaza</li><li>• Ošetrovací dny, ošetrovací dny na JIP</li><li>• Datum a způsob ukončení hospitalizace</li><li>• Hlavní a bezprostřední příčina úmrtí pacienta</li><li>• Zdravotní pojišťovna</li><li>• Ošetrovací dny dle jednotlivých kategorií</li><li>• Identifikační kód výkonu</li><li>• Počet výkonů</li><li>• Hlavní diagnóza za pobyt na oddělení</li><li>• DRG skupina</li></ul>	<ul style="list-style-type: none"><li>• Identifikační číslo ZZ</li><li>• Rezort zdravotnického zařízení</li><li>• Druh zdravotnického zařízení</li><li>• Demografické údaje pacienta</li><li>• Kraj, okres, obec bydliště pacienta</li><li>• Datum a čas přijetí do ZZ</li><li>• Doporučení, způsob a důvod přijetí</li><li>• Identifikátor rehospitalizace pro danou diagnózu</li><li>• Hospitalizace na oddělení (1-5 oddělení)</li><li>• Kód oddělení, kód pracoviště</li><li>• Typ lůžka</li><li>• Základní diagnóza a vnější příčina (MKN-10)</li><li>• Datum, čas a druh operace</li><li>• Hlavní operační diagnóza</li><li>• Hlavní operační výkon</li><li>• Reoperace a pooperační komplikace</li><li>• Ošetrovací dny, ošetrovací dny na JIP</li><li>• Další diagnózy (MKN-10)</li><li>• Datum a způsob ukončení hospitalizace</li><li>• Hlavní diagnóza za pobyt na oddělení</li><li>• Hlavní a bezprostřední příčina úmrtí pacienta</li><li>• Celkový počet ošetrovacích dnů</li><li>• Celkový počet ošetrovacích dnů na JIP</li><li>• DRG skupina</li></ul>	<ul style="list-style-type: none"><li>• Identifikační číslo ZZ</li><li>• Druh zdravotnického zařízení</li><li>• Kód oddělení, kód pracoviště</li><li>• Datum provedení výkonu</li><li>• Kód výkonu</li><li>• Počet provedených výkonů</li><li>• Identifikační kód výkonu</li><li>• Rok propuštění pacienta</li></ul>



## Příloha 2

### Kapitoly MKN-10

---

I	A00-B99	Některé infekční a parazitární nemoci
II	C00-D48	Novotvary
III	D50-D89	Nemoci krve, krvetvorných orgánů a některé poruchy týkající se mechanismu imunity
IV	E00-E90	Nemoci endokrinní, výživy a přeměny látek
V	F00-F99	Poruchy duševní a poruchy chování
VI	G00-G99	Nemoci nervové soustavy
VII	H00-H59	Nemoci oka a očních adnex
VIII	H60-H95	Nemoci ucha a bradavkového výběžku
IX	I00-I99	Nemoci oběhové soustavy
X	J00-J99	Nemoci dýchací soustavy
XI	K00-K93	Nemoci trávicí soustavy
XII	L00-L99	Nemoci kůže a podkožního vaziva
XIII	M00-M99	Nemoci svalové a kosterní soustavy a pojivové tkáně
XIV	N00-N99	Nemoci močové a pohlavní soustavy
XV	O00-O99	Těhotenství, porod a šestinedělí
XVI	P00-P96	Některé stavy vzniklé v perinatálním období
XVII	Q00-Q99	Vrozené vady, deformace a chromozomální abnormality
XVIII	R00-R99	Příznaky, znaky a abnormální klinické laboratorní nálezy nezařazené jinde
XIX	S00-T98	Poranění, otravy a některé jiné následky vnějších příčin
XX	V01-Y98	Vnější příčiny nemocnosti a úmrtnosti
XXI	Z00-Z99	Faktory ovlivňující zdravotní stav a kontakt se zdravotnickými službami
XXII	U00-U99	Kódy pro speciální účely

---

### Příloha 3

#### Oddíly MKN-10 definované pro práci

Rozsah diagnóz	Název oddílu
<b>I. Kapitola</b>	
A00-A09	Střevní infekční nemoci
A15-A19	Tuberkulóza
A20-A28	Některé bakteriální zoonózy
A30-A49	Jiné bakteriální nemoci
A50-A64	Infekce přenášené převážně sexuálním stykem
A65-A69	Jiné spirochétové nemoci
A70-A74	Jiné nemoci způsobené chlamydiemi
A75-A79	Rickettsiázy
A80-A89	Virové infekce centrální nervové soustavy
A90-A99	Virové horečky a virové hemoragické horečky přenášené členovci
B00-B09	Virové infekce charakterizované postižením kůže a sliznice
B15-B19	Virová hepatitida
B20-B24	Onemocnění virem lidské imunodeficience [HIV]
B25-B34	Jiné virové nemoci
B35-B49	Mykózy
B50-B64	Protozoární nemoci
B65-B83	Helmintózy – hlístové nemoci
B85-B89	Zavšivení, napadení roztoči nebo jinými členovci
B90-B94	Následky infekčních a parazitárních nemoci
B95-B98	Bakteriální, virová a jiná infekční agens
B99	Jiné infekční nemoci
<b>II. Kapitola</b>	
C00-C14	ZN rtu, dutiny ústní a hltanu
C15-C26	ZN trávicího ústrojí
C30-C39	ZN dýchacích a nitrohručních orgánů
C40-C41	ZN kosti a kloubní chrupavky
C43-C44	ZN kůže
C45-C49	ZN mezoteliální a měkké tkáně
C50	ZN prsu
C51-C58	ZN ženských pohlavních orgánů
C60-C63	ZN mužských pohlavních orgánů
C64-C68	ZN močového ústrojí
C69-C72	ZN oka, mozku a jiných částí centrální nervové soustavy
C73-C75	ZN štítné žlázy a jiných žláz s vnitřní sekrecí
C76-C80	ZN nepřesně určených, sekundárních a neurčených lokalizací
C81-C96	ZN primární, mízní, krvetvorné a příbuzné tkáně
C97	ZN mnohočetných samostatných (primárních) lokalizací
D00-D01	Carcinoma in situ dutiny ústní, jícnu a žaludku a jiných trávicích orgánů
D02	Carcinoma in situ středního ucha a dýchací soustavy
D03-D04	Melanoma in situ a carcinoma in situ kůže
D05	Carcinoma in situ prsu
D06	Carcinoma in situ hrdla děložního
D07	Carcinoma in situ jiných a neurčených pohlavních orgánů
D09	Carcinoma in situ jiných a neurčených lokalizací
D10-D11	NN úst a hltanu
D12-D13	NN trávicího ústrojí
D14-D15	NN dýchacích a nitrohručních orgánů
D16	NN kosti a kloubní chrupavky
D17-D18	Nezhoubný lipomatózní novotvar, herngiom, lymfangiom
D19-D21	NN mezoteliální a měkké tkáně
D22-D23	Melanocytové névy a NN kůže
D24	NN prsu
D25-D28	NN ženských pohlavních orgánů
D29	NN mužských pohlavních orgánů
D30	NN močového ústrojí
D31-D33	NN oka, mozku a jiných částí centrální nervové soustavy
D34-D35	NN štítné žlázy a jiných žláz s vnitřní sekrecí
D36	NN jiných a neurčených lokalizací
D37	Novotvar nejistého/neznámého chování – ret, dutina ústní, trávicí orgány
D38	Novotvar nejistého/neznámého chování – střední ucho, dýchací a nitrohruční orgány
D39	Novotvar nejistého/neznámého chování – ženské pohlavní orgány
D40	Novotvar nejistého/neznámého chování – mužské pohlavní orgány
D41	Novotvar nejistého/neznámého chování – močové orgány

D42-D43	Novotvar nejistého/neznámého chování – mozek a jiné části centrální nervové soustavy
D44	Novotvar nejistého/neznámého chování – žlázy s vnitřní sekrecí
D45-D46	Polycythaemia vera, myelodysplastické syndromy
D47	Novotvar nejistého/neznámého chování – mizní a krvetvorná tkáň
D48	Novotvar nejistého/neznámého chování – neurčená lokalizace
<b>III. Kapitola</b>	
D50-D53	Nutriční anemie
D55-D59	Hemolytické anemie
D60-D64	Aplastické a jiné anemie
D65-D69	Vady koagulace, purpura a jiné krvácivé stavy
D70-D77	Jiné nemoci krve a krvetvorných orgánů
D80-D89	Některé poruchy mechanismu a imunity
<b>IV. Kapitola</b>	
E00-E07	Poruchy štítné žlázy
E10-E14	Diabetes mellitus
E15-E16	Jiné poruchy regulace glukózy a vnitřní sekrece slinivky břišní
E20-E35	Poruchy jiných žláz s vnitřní sekrecí
E40-E46	Malnutrice
E50-E64	Jiné nutriční karence
E65-E68	Obezita a jiné typy hyperalimentace
E70-E90	Poruchy přeměny látek – metabolismu
<b>V. Kapitola</b>	
F00-F09	Organické duševní poruchy včetně symptomatických
F10-F19	Poruchy duševní a poruchy chování způsobené užíváním psychoaktivních látek
F20-F29	Schizofrenie, poruchy schizotypální a poruchy s bludy
F30-F39	Afektivní poruchy
F40-F48	Neurotické, stresové a somatoformní poruchy
F50-F59	Syndromy poruch chování, spojení s fyziologickými poruchami a somatickými faktory
F60-F69	Poruchy osobnosti a chování u dospělých
F70-F79	Mentální retardace
F80-F89	Poruchy psychického vývoje
F90-F98	Poruchy chování a emocí se začátkem obvykle v dětství a dospívání
F99	Neurčená duševní porucha
<b>VI. Kapitola</b>	
G00-G09	Zánětlivé nemoci centrální nervové soustavy
G10-G14	Systémové atrofie postihující primárně centrální nervovou soustavu
G20-G26	Extrapyramidové a pohybové poruchy
G30-G32	Jiné degenerativní nemoci nervové soustavy
G35-G37	Demyelinizující nemoci centrální nervové soustavy
G40-G47	Poruchy záchvatové – paroxysmální
G50-G59	Onemocnění nervů, nervových kořenů, pletení
G60-G64	Polyneuropatie a jiné nemoci periferní nervové soustavy
G70-G73	Nemoci myoneurálního spojení a svalů
G80-G83	Mozková obrna a jiné syndromy ochrnutí
G90-G99	Jiné poruchy nervové soustavy
<b>VII. Kapitola</b>	
H00-H06	Nemoci očního víčka, slzného ústrojí a očnice
H10-H13	Onemocnění spojivky
H15-H22	Nemoci skléry, rohovky, duhovky a řasnatého tělesa
H25-H28	Onemocnění čočky
H30-H36	Nemoci cévnatky a sítnice
H40-H42	Glaukom
H43-H45	Nemoci sklivce a očního bulbu
H46-H48	Nemoci zrkovitého nervu a zrkových drah
H49-H52	Poruchy očních svalů, binokulárního pohybu, akomodace a refrakce
H53-H54	Poruchy vidění a slepota
H55-H59	Jiné nemoci oka a očních adnex
<b>VIII. Kapitola</b>	
H60-H62	Nemoci zevního ucha
H65-H75	Nemoci středního ucha a bradavkového výběžku
H80-H83	Nemoci vnitřního ucha
H90-H95	Jiná onemocnění ucha
<b>IX. Kapitola</b>	
I00-I02	Akutní revmatická horečka
I05-I09	Chronické revmatické choroby srdeční
I10-I15	Hypertenzní nemoci
I20-I25	Ischemické nemoci srdeční
I26-I28	Kardiopulmonární nemoc a nemoci plicního oběhu

I30-I52	Jiné formy srdečního onemocnění
I60-I69	Cévní nemoci mozku
I70-I79	Nemoci tepen, tepének a vlásečnic
I80-I89	Nemoci žil, mizních cév a mizních uzlin nezařazené jinde
I95-I99	Jiná a neurčená onemocnění oběhové soustavy
<b>X. Kapitola</b>	
J00-J06	Akutní infekce horních dýchacích cest
J09-J18	Chřipka a pneumonie
J20-J22	Jiné akutní infekce dolní části dýchacího ústrojí
J30-J39	Jiné nemoci horních dýchacích cest
J40-J47	Chronické nemoci dolní části dýchacího ústrojí
J60-J70	Nemoci plic způsobené zevními činiteli
J80-J84	Jiné nemoci dýchací soustavy postihující hlavně intersticiium
J85-J86	Hnisavé a nekrotické stavy dolní části dýchacího ústrojí
J90-J94	Jiné nemoci pohrudnice
J95-J99	Jiné nemoci dýchací soustavy
<b>XI. Kapitola</b>	
K00-K14	Nemoci ústní dutiny, slinných žláz a čelistí
K20-K31	Nemoci jícnu, žaludku a dvanáctníku
K35-K38	Nemoci apendixu
K40-K46	Kýly
K50-K52	Neinfekční zánět tenkého a tlustého střeva
K55-K64	Jiné nemoci střev
K65-K67	Nemoci pobřišnice
K70-K77	Nemoci jater
K80-K87	Nemoci žlučníku, žlučových cest a slinivky břišní
K90-K93	Jiné nemoci trávicí soustavy
<b>XII. Kapitola</b>	
L00-L08	Infekce kůže a podkožního vaziva
L10-L14	Puchýřnatá onemocnění
L20-L30	Dermatitida a ekzém
L40-L45	Papuloskvamózní onemocnění
L50-L54	Kopřivka a erytém
L55-L59	Onemocnění kůže a podkožního vaziva spojené se zářením
L60-L75	Nemoci kožních adnex
L80-L99	Jiné nemoci kůže a podkožního vaziva
<b>XIII. Kapitola</b>	
M00-M03	Infekční artropatie
M05-M14	Zánětlivé polyartropatie
M15-M19	Artrózy
M20-M25	Jiná onemocnění kloubů
M30-M36	Systémová onemocnění pojivové tkáně
M40-M43	Deformující dorzopatie
M45-M49	Spondylopatie
M50-M54	Jiné dorzopatie
M60-M63	Onemocnění svalů
M65-M68	Onemocnění synoviální blány a šlachy
M70-M79	Jiná onemocnění měkké tkáně
M80-M85	Poruchy hustoty a struktury kosti
M86-M90	Jiné osteopatie
M91-M94	Chondropatie
M95-M99	Jiná onemocnění svalové a kosterní soustavy a pojivové tkáně
<b>XIV. Kapitola</b>	
N00-N08	Nemoci glomerulů
N10-N16	Tubulo-intersticiální nemoci ledvin
N17-N19	Selhání ledvin
N20-N23	Urolitáza
N25-N29	Jiné nemoci ledvin a močovodů
N30-N39	Jiné nemoci močové soustavy
N40-N51	Nemoci mužských pohlavních orgánů
N60-N64	Nemoci prsu
N70-N77	Zánětlivá onemocnění ženských pánevních orgánů
N80-N98	Nezánětlivá onemocnění ženského pohlavního ústrojí
N99	Jiná onemocnění močové a pohlavní soustavy
<b>XV. Kapitola</b>	
O00-O08	Těhotenství končící potratem
O10-O16	Edém, proteinurie, hypertenzní onemocnění (pozdní gestóza) v těhotenství, při porodu, v šestinedělí
O20-O29	Jiná onemocnění matky převážně v souvislosti s těhotenstvím
O30-O48	Péče o matku ve vztahu k plodu, amniové dutině a možným porodním problémům

O60-O75	Komplikace porodu
O80-O84	Porod
O85-O92	Komplikace spojené převážně s šestineděním
O94-O99	Jiné porodnické stavy, nezařazené jinde
<b>XVI. Kapitola</b>	
P00-P04	Postižení plodu a novorozence onemocněním matky, komplikacemi těhotenství a porodu
P05-P08	Poruchy spojené s délkou těhotenství a růstu plodu
P10-P15	Poranění za porodu
P20-P29	Respirační a kardiovaskulární poruchy specifické pro perinatální období
P35-P39	Infekce specifické pro perinatální období
P50-P61	Krvácivé stavy a hematologické poruchy plodu a novorozence
P70-P74	Přechodné poruchy endokrinní a přeměny látek specifické pro plod a novorozence
P75-P78	Poruchy trávicí soustavy plodu a novorozence
P80-P83	Stavy postihující kůži a regulaci teploty plodu a novorozence
P90-P96	Jiné poruchy vzniklé v perinatálním období
<b>XVII. Kapitola</b>	
Q00-Q07	Vrozené vady nervové soustavy
Q10-Q18	Vrozené vady oka, ucha, obličeje a krku
Q20-Q28	Vrozené vady oběhové soustavy
Q30-Q34	Vrozené vady dýchací soustavy
Q35-Q37	Rozštěp rtu a rozštěp patra
Q38-Q45	Jiné vrozené vady trávicí soustavy
Q50-Q56	Vrozené vady pohlavních orgánů
Q60-Q64	Vrozené vady močové soustavy
Q65-Q79	Vrozené vady a deformace svalové a kosterní soustavy
Q80-Q89	Jiné vrozené vady
Q90-Q99	Abnormality chromozomů nezařazené jinde
<b>XVIII. Kapitola</b>	
R00-R09	Príznaky a znaky týkající se oběhové a dýchací soustavy
R10-R19	Príznaky a znaky týkající se trávicí soustavy a břicha
R20-R23	Príznaky a znaky týkající se kůže a podkožního vaziva
R25-R29	Príznaky a znaky týkající se nervové, svalové a kosterní soustavy
R30-R39	Príznaky a znaky týkající se močové soustavy
R40-R46	Príznaky a znaky týkající se vědomí, vnímání, emočního stavu a chování
R47-R49	Príznaky a znaky týkající se řeči a hlasu
R50-R69	Celkové příznaky a znaky
R70-R79	Abnormální nálezy při vyšetření krve bez diagnózy
R80-R82	Abnormální nálezy při vyšetření moči bez diagnózy
R83-R89	Abnormální nálezy při vyšetření jiných tělesných tekutin, látek a tkání bez diagnózy
R90-R94	Abnormální nálezy při diagnostických zobrazovacích a při funkčních vyšetřeních bez diagnózy
R95-R99	Nepřesně určené a neznámé příčiny smrti
<b>XIX. Kapitola</b>	
S00-S09	Poranění hlavy
S10-S19	Poranění krku
S20-S29	Poranění hrudníku
S30-S39	Poranění břicha, dolní části zad, bederní páteře a pánve
S40-S49	Poranění ramene a paže (nadloktí)
S50-S59	Poranění lokte a předloktí
S60-S69	Poranění zápěstí a ruky
S70-S79	Poranění kyčle a stehna
S80-S89	Poranění kolena a bérce
S90-S99	Poranění kotníku a nohy pod ním
T00-T07	Poranění postihující více částí těla
T08-T14	Poranění neurčené části trupu, končetiny nebo části těla
T15-T19	Účinky cizího tělesa vniklého přirozeným otvorem těla
T20-T25	Popáleniny a poleptání povrchu těla, určené podle lokalizace
T26-T28	Popáleniny a poleptání omezené na oko a vnitřní orgány
T29-T32	Popáleniny a poleptání mnohočetných a neurčených částí těla
T33-T35	Omrzliny
T36-T50	Otrava léky, léčivy, návykovými a biologickými látkami
T51-T65	Toxické účinky látek ze zdrojů převážně mimo lékařství
T66-T78	Jiné a neurčené účinky vnějších příčin
T79	Některé časné komplikace úrazů
T80-T88	Komplikace zdravotní péče nezařazené jinde
T90-T98	Následky poranění, otravy a jiných následků vnějších příčin

---

**XX. Kapitola**

---

V01-V09	Chodec zraněný při dopravní nehodě
V10-V19	Cyklista zraněný při dopravní nehodě
V20-V29	Jezdec na motocyklu zraněný při dopravní nehodě
V30-V39	Člen osádky tříkolového motorového vozidla zraněný při dopravní nehodě
V40-V49	Člen osádky osobního automobilu zraněný při dopravní nehodě
V50-V59	Člen osádky dodávkového nebo lehkého nákladního automobilu zraněný při dopravní nehodě
V60-V69	Člen osádky těžkého nákladního vozidla zraněný při dopravní nehodě
V70-V79	Člen osádky autobusu zraněný při dopravní nehodě
V80-V89	Jiné nehody při pozemní dopravě
V90-V94	Nehody při vodní dopravě
V95-V97	Nehody při dopravě vzduchem a vesmírným prostorem
V98-V99	Jiné a neurčené dopravní nehody
W00-W19	Pády
W20-W49	Vystavení neživotným mechanickým silám
W50-W64	Vystavení životným mechanickým silám
W65-W74	Náhodné (u)tonutí a potopení
W75-W84	Jiná náhodná ohrožení dýchání
W85-W99	Vystavení elektrickému proudu, ozáření a extrémní okolní teplotě a tlaku vzduchu
X00-X09	Vystavení kouři, ohni, dýmu a plamenům
X10-X19	Kontakt s horkem a horkými látkami
X20-X29	Kontakt s jedovatými živočichy a rostlinami
X30-X39	Vystavení přírodním silám
X40-X49	Náhodná otrava škodlivými látkami a expozice jejich působení
X50-X57	Přetížení, cestování a strádání
X58-X59	Náhodné vystavení jiným a neurčeným faktorům
X60-X84	Umyslné sebepoškození
X85-Y09	Napadení (útok)
Y10-Y34	Případ (událost) nezjištěného úmyslu
Y35-Y36	Zákonný zákrok a válečné operace
Y40-Y59	Léky, léčiva, návykové, biologické látky - příčina nežádoucích účinků při léčebném použití
Y60-Y69	Nehody u pacientů při zdravotní péči
Y70-Y82	Nežádoucí příhody související s použitím lékařských přístrojů/pomůcek při diagnostice/léčbě
Y83-Y84	Operační/vyšetřovací/léčebné výkony jako příčina abnormální reakce pacienta či komplikace, bez nehody při výkonu
Y85-Y89	Následky vnějších příčin nemoci a úmrtnosti
Y90-Y98	Doplňkové faktory týkající se příčin nemoci a úmrtnosti zařazených jinde

---

**XXI. Kapitola**

---

Z00-Z13	Osoby, které se setkaly se zdravotnickými službami za účelem prohlídky a vyšetření
Z20-Z29	Osoby s potenciálně ohroženým zdravím ve vztahu k přenosným nemocem
Z30-Z39	Osoby, které se setkaly se zdravotnickými službami za okolností souvisejících s reprodukcí
Z40-Z54	Osoby, které se setkaly se zdravotnickými službami za účelem určitých výkonů a zdravotní péče
Z55-Z65	Osoby s potenciálně ohroženým zdravím ve vztahu k socioekonomickým a psychosociálním okolnostem
Z70-Z76	Osoby, které se setkaly se zdravotnickými službami za jiných okolností
Z80-Z99	Osoby s potenciálně ohroženým zdravím ve vztahu k rodinné/osobní anamnéze a některým podmínkám ovlivňujícím zdravotní stav

---

**XXII. Kapitola**

---

U00-U49	Provizorní určení pro nové diagnózy nejisté etiologie nebo nouzového použití
U80-U89	Rezistence na antimikrobiální a protinádorové léky

---