

123 Proveditelnost fúze dat NOR a dat plátců zdravotní péče, její přínos a výstupy.Dušek L.¹, Brabec P.¹, Klimeš D.¹, Ráček J.¹, Klika P.¹, Andres P.²

1) Lékařská fakulta MU, Brno

2) Masarykův onkologický ústav, Brno

Úvod

Příspěvek shrnuje studii proveditelnosti, která byla v roce 2007 vypracována pro projekt „Fúze dat Národního onkologického registru a dat plátců zdravotní péče“. Projekt vstoupí v roce 2008 do fáze pilotních zkoušek. Cílem příspěvku je předložit plánovaný postup analýz k diskusi odborné veřejnosti.

Garance za Českou onkologickou společnost ČLS JEP a její odborné sekce (abecedně): prof. MUDr. J. Abrahámová, DrSc.; prof. MUDr. Miloslav Duda, DrSc.; doc. RNDr. L. Dušek, Ph.D.; doc. MUDr. J. Finek, CSc.; doc. MUDr. M. Hajdúch, CSc.; prof. MUDr. J. Petera, CSc.; prof. MUDr. L. Petruželka, CSc.; prof. MUDr. J. Vorlíček, CSc.; prof. MUDr. R. Vyzula, CSc.; prof. MUDr. J. Žaloudík, CSc..

Projekt „Fúze dat NOR a PZP“ a jeho cíle

Projekt navrhuje retrospektivní analýzu dvou hlavních populačních datových zdrojů, které jsou v ČR k dispozici v uspokojivé kvalitě. Jde o data NOR jako reprezentativní epidemiologickou databázi nesoucí především diagnostickou identifikaci zhoubných nádorů a dále o data PZP sledující objem a strukturu nákladů vynaložených na diagnostiku a léčbu onemocnění v jednotlivých zdravotnických zařízeních. Projekt navrhuje provedení fúze části těchto datových zdrojů pro jednorázovou analýzu s těmi-to cíli:

- **obohacení databáze plátců zdravotní péče** o diagnostická data zhoubných nádorů, především o klinické stadium a další ukazatele pokročilosti onemocnění
- **validace fúzovaných dat a definice referenčního souboru**, který může být využit pro **nastavení referenčních standardů nákladovosti péče**, a to se znalostí pokročilosti nádorového onemocnění
- **mapování struktury nákladů** v různých diagnostických skupinách a zjištění jejich variability v podmínkách různých typů zdravotnických zařízení
- **analýza struktury nákladů v referenčním souboru dat** dle různých modalit léčby
- **obohacení databáze Národního onkologického registru** o přesnou identifikaci pracovišť odpovědných za hlášení a různé fáze péče o onkologického pacienta
- **exaktní zmapování incidence a prevalence** onkologických diagnóz v regionech ČR.

Provedení analýz a jejich dopad na funkčnost zapojených databází

Cíle projektu jsou pouze analytické, projekt nechce měnit pravidla a procesy související se sběrem dat v zapojených datových zdrojích. Pro provedení fúze budou nadto vybrána pouze částečná a jasně definovaná data, nepůjde o obtížně kontrolovatelné spojování historických dat. Fúzovaná data musí přinést informaci využitelnou v současné onkologii. Vlastní fúze dat se provede jednorázovým exportem definovaných datových dávek z obou zdrojů a jejich propojením přes identifikátory pacienta a choroby. Tato fáze projektu musí být přísně zabezpečena a kontrolována, neboť bude pracovat s osobními daty pacientů. Po tomto spojení je možné všechny cesty a záznamy umožňující přímou i nepřímou identifikaci konkrétní osoby smazat a spojenou databázi předat ke zpracování v anonymizované podobě. Realizace projektu tedy nijak neovlivní funkčnost zdrojových databází a také dlouhodobě nezatíží odborný personál zapojených institucí.

Navržený proces fúze

Spojení dat PZP a NOR může proběhnout v následujících krocích:

1. dohodnutí způsobu šifrování a anonymizace dat na straně správců,
2. předání anonymizovaných dat,
3. nalezení odpovídajících položek a jejich sloučení,
4. nalezení nesrovnalostí v původních datech.

Nejdříve bude zapotřebí, aby si správci obou databází dohodli:

- šifrovací klíč (číselný kód pro zašifrování z důvodu obrany proti slovníkovým útokům),
- jednosměrnou funkci (hashovací algoritmus pro zašifrování dat).

Dohodnutý šifrovací klíč a hashovací algoritmus bude shodný pro všechny položky v obou databázích. To, na jakém klíči a algoritmu se oba správci dohodli, bude tajné. Tuto skutečnost správci nesdělí žádné další osobě nebo instituci. Následně bude

INFORMAČNÍ SYSTÉMY V ONKOLOGII

ke každému rodnému číslu pacienta v obou databázích připojen šifrovací klíč a za použití dohodnutého hashovacího algoritmu bude z každé dvojice [rodné číslo pacienta, šifrovací klíč] v databázi vytvořen jednosměrný otisk, kterým bude nahrazeno původní rodné číslo pacienta. Jednosměrný otisk se stane novým identifikátorem pacienta ve všech záznamech obou databází. Dále budou z obou databází odstraněny všechny další osobní údaje pacientů, jako jsou například jméno a příjmení pacienta, datum narození nebo trvalé bydliště. Tímto způsobem vzniknou z původních databází nové anonymizované databáze bez osobních údajů pacientů. Možnost rozlišovat mezi jednotlivými pacienty a jejich onemocněními však zůstane zachována.

Závěry studie proveditelnosti

(Studie vypracována v roce 2007 pro Ministerstvo zdravotnictví ČR)

Navržený projekt fúze dat NOR a PZP je proveditelný a může probíhat pod kontrolou platných zákonů a norem. Při řešení je třeba maximálně dbát na zachování práv jak subjektů, jejichž data jsou v databázích obsažena, tak i organizací odpovědných za správu těchto dat. Tomuto musí být plně podřízena smluvní agenda, která upraví vztah všech institucí zapojených do řešení projektu. Dále je nutné zajistit informační bezpečnost s využitím veškerých dostupných technických a organizačních prostředků. V současnosti nelze doporučit neanonymizované společné zpracování dat ze zdravotnické dokumentace, registrů zdravotních pojišťoven a registrů NZIS.

Studie navrhuje nastavit řešení projektu tak, aby zpracovatel dat dostal plně anonymizované podklady a neměl žádnou přímou ani nepřímou cestu k identifikaci konkrétních osob. Podle navrženého řešení citlivá osobní data pacientů neopustí databáze institucí odpovědných za jejich správu (ÚZIS ČR, PZP). Vzhledem ke struktuře údajů v databázích PZP i NOR lze fúzi provést na základě shodného rodného čísla. Je však zapotřebí, aby rodné číslo bylo ještě před převzetím dat zpracovatelem nahrazeno jednosměrným otiskem rodného čísla (tzv. hash). Nahrazení rodného čísla otiskem umožní rozpoznat, že se údaje vztahují ke stejné osobě, ale zároveň budou jednotlivé údaje z pohledu zpracovatele anonymní. Identifikace údajů bude možná, ovšem pouze na straně správců dat (VZP, NOR). Vlastní fúzi může za těchto podmínek provést kterýkoli ze správců zdrojových dat nebo přímo určený zpracovatel. Navržené řešení nijak neovlivňuje sběr dat a chod zdrojových databází.

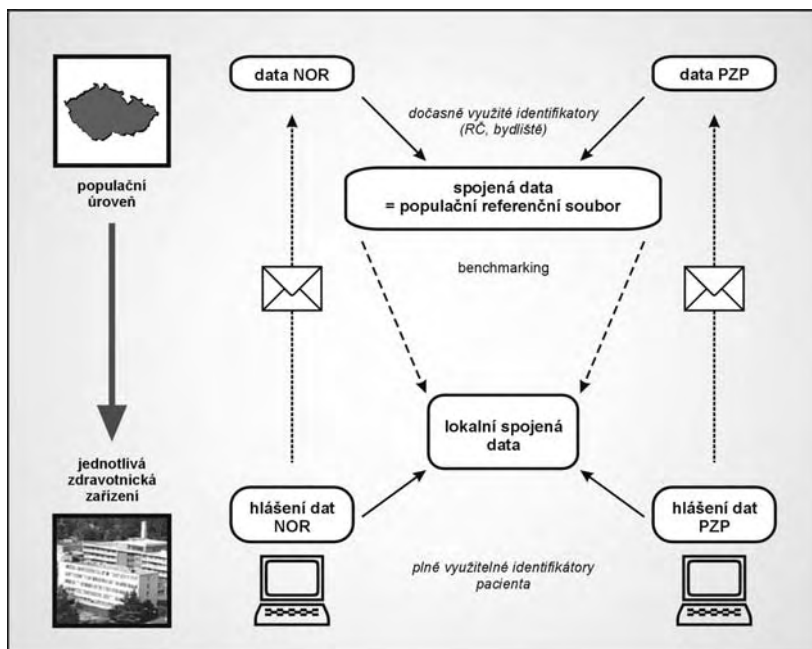
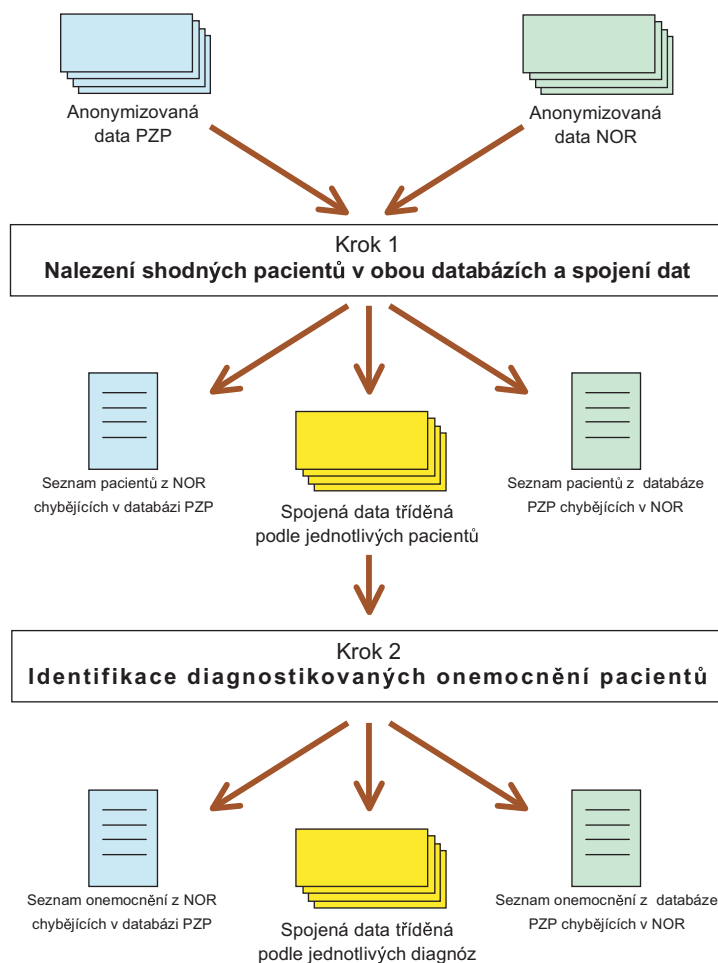
**Obr. 1**

Schéma spojení dat NOR a dat PZP na centrální a lokální úrovni

**Obr. 1**

Navržený proces fúze dat NOR a PZP

INFORMAČNÍ SYSTÉMY V ONKOLOGII

Rozbor prokázal, že spojená data NOR a PZP představují plně analyzovatelný systém, který bude základnou pro řadu výstupů, které bez takto spojených dat v ČR nejde provádět:

- hodnocení výsledků a kvality péče v onkologii, její mapování a optimální nastavení standardů
- hodnocení přínosu nejnákladnější terapie v onkologii se zřetelem na stav pacienta a onemocnění
- typologii pacientů indikovaných pro různé léčebné strategie s nejvyšším benefitem pro nemocného
- hodnocení krátkodobého i dlouhodobého přežití ve vztahu k rozsahu onemocnění a provedené léčbě
- prediktivní hodnocení nákladů potřebných pro různá onkologická onemocnění včetně různých stádií těchto onemocnění, predikce prevalence léčených pacientů
- hodnocení nákladové efektivity

Poděkování

Projekt Fúze dat NOR a PZP a zde shrnutá studie proveditelnosti jsou podporovány výzkumným grantem Ministerstva zdravotnictví ČR.