

Informační technologie pro současnou onkologii – sběr a analýza dat

INFORMAČNÍ TECHNOLOGIE PRO SOUČASNOU ONKOLOGII – SBĚR A ANALÝZA DAT

INFORMATION TECHNOLOGIES FOR REGISTRATION OF CLINICAL DATA IN ONCOLOGY

KLIMEŠ D., ŠMÍD R., PAVLIŠ P., SVOBODNÍK A., DUŠEK L.

CENTRUM BIOSTATISTIKY A ANALÝZ, LF MU BRNO

Souhrn: Nabídka informačních technologií je v oblasti sběru a analýzy medicínských dat v dnešní době opravdu široká. Z technologického pohledu lze vybudovat komplexní systém, který zajistí nejen kvalitní sběr dat, ale i jejich vysoce sofistikované zpracování a prezentaci, systém řídicí, systém automaticky produkující z dat nové informace. Na začátku je však nutné zvládnout vlastní proces sběru dat. Ten je v reálném provozu značně zatěžující a pokud nepřináší okamžité znatelné přínosy, je obecně velmi špatně přijímán. Navíc pořizovací náklady nejnovějších informačních technologií jsou značně vysoké a pro zdravotnická zařízení často limitují rozvoj. Je třeba hledat dostupnější řešení, která využijí již existující datové zdroje (NIS, LIS, data pojišťoven) a která vyčtí maximum z jejich propojení. V oblasti nákladné onkologické péče může průběžné vyhodnocování existujících dat doplněných o několik stěžejních parametrů přinést nejen zkvalitnění lékařské péče, ale i nemalé finanční úspory.

Klíčová slova: elektronická dokumentace pacienta, lékařské databáze, analýza medicínských dat, HTA

Summary: Nowadays, there is a wide variety of IT solutions dealing with medical data collection and analysis. From technological point of view, it is possible to build a complex system that will handle all of the following tasks: data collection, sophisticated data analysis and presentation, management tasks, automatic extraction of new information from existing data, quality control. When building such system, it is necessary to make up the process of data collection and data management as key step for the whole process. This is very demanding in real clinical practice, and if it does not bring immediate benefits, it is very poorly accepted. Moreover, the expenses for novel IT technologies are very high and therefore it is necessary to find more accessible solution that can utilize existing data sources (HIS, LIS, insurance companies data, health care data) and gain maximum from the their connection. Continuous evaluation of existing data with a few important parameters appended can bring in the area of very expensive care in oncology both improving of quality of medical care and money saving.

Key Words: electronic medical record (EMR), medical databases, assessment of medical data, HTA

Tato kapitola by měla čtenáři přiblížit problematiku sběru dat a nahlédnout do kuchyně informatiků a databázových specialistů. Zcela záměrně se vyhneme odborným termínům, které by mohly být pro neodborníky z oblasti informatiky nesrozumitelné. Místo toho se zaměříme na vlastní problematiku sběru klinických dat, na problémy, které s ním souvisejí, a na možnosti a dostupnost technologií, které kvalitní sběr umožňují. I když v nadpisu hovoříme přímo o onkologických datech, nejsou data z oboru onkologie z pohledu informatika nijak odlišná od dat například hematologických. Specifická jsou pouze obecně data medicínská, a to jednak svou strukturou a svou značnou šíří a jednak svou obtížnou parametrizací (parametrizace = jednoznačné zvolení jedné z nabízených možností namísto nejednoznačného textového popisu skutečnosti). Proto v následujícím textu budeme hovořit spíše obecně o datech medicínských.

Hned na začátek je nutné upozornit na dvě základní věci. Za prvé, že veškeré statistické a analytické postupy vedoucí k získání nových poznatků či ověření stávajících závisí na kvalitním a dostatečném množství dat. A za druhé, že sběr kvalitních dat (tedy úplných a s minimem chyb), je velmi finančně i časově náročná záležitost. To platí obecně, v oblasti medicínských dat pak zvláště. Proč, to nastíníme v průběhu této kapitoly.

Nejprve si shrňme, jak může sbírat data jedna samotná osoba (např. za účelem vlastní publikace). Samozřejmě nejdostup-

nější je tužka a papír. I laik snadno pochopí, že chce-li spočítat například průměrnou hodnotu, nesmí jeho sbíraná data obsahovat texty typu „mírně lepší než včera“. Jakmile si toto uvědomí, je následujícím krokem příprava papírového formuláře s nachystanými kolonkami. Možná bude pro čtenáře překvapující, že tímto krokem startuje i návrh celonárodních registrů založených na velmi moderních informačních technologiích. Další možností, kterou má „běžný uživatel“, v našem případě lékař, ke sběru dat k dispozici je nějaký program z rodiny tabulkových kalkulátorů, což bývá nejčastěji MS Excel. Tento typ programu umožňuje celkem snadno převést data z kolonek papírového formuláře do jednotlivých buněk listu programu. Provádí se tak tzv. digitalizace, tedy převod dat z podoby uchopitelné pouze člověkem do podoby zpracovatelné počítačem. Odměnou za tuto náročnou a zdlouhavou práci (která má mimochodem i více než 5% chybovost) je možnost spočítat si základní statistické údaje a vykreslit je v grafech. Takto lze získat bez utrácení za informační technologie ucelený soubor dat, který je možné předat ke zpracování. Avšak jsou tu významná omezení. Programy typu MS Excel jsou navrženy jako jednoduživatelské aplikace, tedy programy, se kterými pracuje pouze jeden člověk. V momentě, kdy na sběru dat spolupracuje více osob, je nutné se porozhlédnout po jiném řešení. Pokud se i v tomto případě chceme obejít bez odborníka-informatika, je jediným řešením individuální sběr jednotlivými osobami a následné spojování vzniklých dato-

vých fragmentů do jednoho celku. Toto spojování je pak prováděno jednou osobou a je samozřejmě opět zatíženo určitou chybou.

V momentě, kdy se rozhodneme do sběru dat vložit kromě času i určitou finanční částku, měli bychom dopředu znát, co od pořizované technologie očekáváme (pojem technologie v tomto případě zahrnuje vše od počítačů přes programové vybavení po instalační, záškolovací či analytický servis). V následujících bodech jsou shrnuty možnosti dnešních technologií, seřazeny jsou podle rozšířenosti a finanční náročnosti.

- Validace vkládaných údajů
- Základní grafické a statistické výstupy
- Vyhledávací funkce
- Víceuživatelský přístup
- Zabezpečení dat
- Sběr obrazových dat
- Analytické funkce
- Řídící funkce
- Pokročilé technologie – analýza textu, analýza obrazu, doložení dat

K jednotlivým bodům blíže. Validací údajů rozumíme kontrolu správnosti a to na několika úrovních. Ta nejjednodušší je kontrola správnosti formátu a typu zadávaných dat, tedy zda je do číselné položky zadáváno číslo a ne text a podobně. Další úrovní je kontrola na mezní hodnoty u číselných (např. laboratorních) položek. Ještě vyšší a už náročnější je úroveň kontroly logické správnosti. Klasickým příkladem je kontrola, zda datum vyšetření nenásleduje za datem úmrtí či před datem narození pacienta. Takovýchto pravidel najdeme v medicínských datech samozřejmě více, bývají však podstatně složitější a jejich tvorba a zavádění do praxe vyžaduje úzkou spolupráci mezi lékařem a poskytovatelem technologie. Ve složitějších případech se už dostáváme k řídicím funkcím (technologie sama nabízí řešení problému), o kterých bude řeč později.

Jak už bylo zmíněno, základní statistické a grafické výstupy je možné připravit i v aplikacích typu MS Excel, které bývají standardem ve všech typech na zakázku tvořených jednouuživatelských programů i v komerčně dodávaných informačních systémech. Šíře poskytovaných výstupů záleží jednoznačně na domluvě mezi zadavatelem a dodavatelem technologie. Není však dobré problematiku výstupů řešit až někdy na závěr, lechse se můžete dostat do situace, kdy pro požadované výstupy nemáte (nesbíráte) potřebná data. Je proto dobré mít o výstupech jasno dříve, než zahájíte vlastní sběr dat.

Vyhledávací funkce využijeme v případě, kdy data sbíráme nejen pro jednorázový výstup (publikace), ale kdy naopak chceme s daty průběžně pracovat. Opět lze i vyhledávací funkce rozdělit na několik úrovní. Za základní funkci lze považovat vyhledání konkrétního záznamu (pacienta) podle základních (osobních) údajů. Na vyšší úrovni lze požadovat vyhledávání záznamů podle např. libovolné nadprůměrné či extrémní hodnoty. Stejně jako v případě validačních funkcí lze i po vyhledávacích funkcích požadovat zobrazení logicky „podezřelých“ záznamů. Tato úroveň se však už dotýká funkcí analytických a řídicích.

Pokud budeme požadovat souběžný sběr dat z více míst více osobami, musíme počítat s více problémy a s nimi související vyšší cenou technologie. Opět je rozdíl, pokud chceme sbírat data ve dvou sousedících místnostech nebo v celém areálu nemocnice nebo na celostátní úrovni. Základním problémem a otázkou víceuživatelského přístupu je, zda si budou všichni uživatelé rovni, či zda by bylo vhodné investovat do hierarchické struktury oprávnění, kdy má každá osoba pracující se systémem přesně určeno, ke kterým údajům má přístup, které může vkládat, měnit či mazat.

S víceuživatelským přístupem souvisí i problematika zabezpečení dat. Pod tímto pojmem se skrývají dva odlišné problémy. Prvním je zabezpečení dat před zničením díky technické závadě (zhroucený disk). Tento problém se řeší standardně

zálohováním dat na více místech (discích) a záleží pouze na nastavení jak často a jakým způsobem se toto zálohování bude provádět. Druhým problémem, který se skrývá pod pojmem zabezpečení dat, je zabezpečení před neoprávněným přístupem a zneužitím dat (van der Haak *et al.* 2003). Toto je už mnohem složitější problém a bylo o něm napsáno celé knihy. Stručně lze jen říct, že dokonale zabezpečená data jsou jen data, která neexistují. Vždy totiž existuje určité riziko, že k datům získá přístup někdo neoprávněný, a zabezpečovací technologie toto riziko pouze snižují (samozřejmě za odpovídající cenu). Zabezpečovat proces sběru dat znamená vynaložit úsilí hned několika směry. Je třeba zabezpečit přístup k uživatelskému rozhraní systému, tedy zabránit, aby k aplikacím, které zprostředkovávají vlastní sběr dat, měl přístup pouze oprávněný uživatel. Nejběžnějším řešením je přidělení přihlašovacích jmen a hesel. Jelikož hesla mohou být zcizena nebo prolomena (uhádnuta), bývá toto základní zabezpečení v případě požadavku na vyšší bezpečnost rozšířeno o pokročilejší prvky. Těmi může být ověření identity uživatele přes čipovou kartu či mobilní telefon. S takovými bezpečnostními prvky se můžeme setkat například v bankovním sektoru. Zabezpečit však nestačí pouze „oficiální“ přístup k datům, je třeba zabezpečit veškeré transfery dat a samozřejmě samotné úložiště dat. Celý proces zabezpečování pak vyčerpá značnou část z pořizovacích nákladů technologie. Při rozhodování o množství prostředků, které chceme vložit do zabezpečení dat, je vhodné zvážit následující hlediska. Jak jsou sbíraná data cenná, jak závažné by byly následky zneužití dat a jaké je množství potenciálních útočníků a jak velká může být jejich motivace. Třetí hledisko může být pro čtenáře záležející, v praxi však může být často rozhodující. Je nutné si uvědomit, že nabourávání informačních systémů je časově i finančně náročná záležitost a skutečně nebezpečí tak hrozí pouze od silně motivovaného útočníka.

Obrazová data (snímky RTG, CT apod.) jsou speciálním typem dat, která můžeme shromažďovat. Speciální jsou jednak svým objemem, jednak technikami zpracování. Velikost obrazových dat je mnohonásobně větší než velikost dat pořizovaných ručním přepisem. Z toho vyplývají i vyšší nároky na ukládací diskový prostor a kapacitu přenosové sítě při víceuživatelském přístupu. Hodnocení tohoto typu dat, pokud má probíhat strojově, pak patří mezi velmi pokročilé metody a zabývá se jím samostatná oblast umělé inteligence.

Analytické funkce již patří k nadstandardním funkcím, se kterými se setkáváme až u komplexnějších technologií jako jsou nemocniční informační systémy. Mezi tyto funkce můžeme řadit pokročilejší statistické metody včetně metod vícerozměrných, funkce umožňující vyhledávání rozhodovacích pravidel, techniky modelování a podobně.

Na analytické funkce pak navazují funkce řídicí. Od těch se očekává, že z nasbíraných dat a vložených tzv. znalostí vyprodukují pokyny (doporučení), tedy že dosud pasivní technologie převezme aktivní roli v procesu řízení a rozhodování. V lékařské praxi to znamená, že systém začne sám nabízet možné postupy léčby, upozorňovat na možné kontraindikace léků, na rizika nežádoucích účinků. Tato oblast je intenzivně studována, což dokládají články v odborných časopisech lékařské informatiky (Haug *et al.* 2003; Smith *et al.* 2003; Takeda *et al.* 2003). Bylo totiž zjištěno, že předcházení chybám v léčbě péči a nežádoucím účinkům vede nejen ke zlepšení zdravotního stavu pacientů, ale i ke značné úspoře finančních prostředků zdravotnického zařízení (Wang *et al.* 2003).

Mezi nejméně rozšířené a zatím intenzivně vědecky zkoumané analytické metody patří analýza textu, již zmíněná analýza obrazu či analýza biosignálů (např. EKG, EEG). Analýzou textu rozumíme vyhledávání konkrétních faktů ve volně psaném textu (lékařské zprávě, chorobopisu) (Mikkelsen and Aasly 2002, 2003; Tange *et al.* 1997). Analýza obrazu se v oblasti medicíny zabývá počítačovým rozbořením lékařských snímků (RTG, CT) a hledáním patologických změn na nich. Oblast

tzv. data mining technologií, stejně jako předchozí patří též do oblasti umělé inteligence. Jejím cílem je automatické hledání vztahů a pravidel v datovém souboru, generování nových hypotéz či tvorba rozhodovacích pravidel (Jerez-Aragones *et al.* 2003; Lavrac 1999; Ngan *et al.* 1999; Perner *et al.* 2002; Setiono 1996). Všechny zmíněné pokročilé metody mají za cíl napodobit lidské schopnosti a přenést určitý objem dosud výhradně lidské práce na výpočetní techniku.

Vyjmenovali jsme si funkce, které dnes můžeme po technologiích vyžadovat, pokud se pustíme do shromažďování dat. Uvedené funkce si můžeme označit za funkce výkonné, tedy funkce, které nám usnadňují nebo zkvalitňují sběr dat nebo s daty pracují a vytváří pro nás zajímavé výstupy. Velmi důležité jsou však i další vlastnosti technologie. Mezi ně patří např. uživatelská přívětivost, flexibilita, kompatibilita s existujícími systémy. Co znamená uživatelská přívětivost a srozumitelnost systému asi není třeba rozvádět, ví to každý, kdo na libovolné úrovni pracuje s počítačem. Na obranu autorů informačních systémů poznamenejme, že tato oblast není zdaleka věcí jednoduchou. Pojem flexibilita má v oblasti technologií pro sběr dat nemalý význam. U medicínských dat je totiž prakticky nemožné definovat přesný výčet položek, které budou jednou provždy sbírány bez potřeby přidání nové či modifikace existující položky. Takto lze postupovat pouze v případě jednorázových sběrů za jediným konkrétním cílem. V případě, kdy začínáme sbírat data pouze s obecnou představou, co od toho očekáváme, musíme na technologii, kterou pro sběr využíváme, vyžadovat schopnost pružně reagovat na naše požadavky ohledně počtu či struktury sbíraných položek. Tento problém z uživatelského pohledu triviální bývá na úrovni samotné technologie často velmi komplikovaný, vyžadující zásah správce systému či programátora. Takový zásah pak může trvat od několika hodin až třeba po několik týdnů. Flexibilita je také důležitá v případě přidávání či modifikace analytických výstupů.

Nově pořizovaná technologie by měla také být schopna komunikovat s technologiemi stávajícími (NIS, LIS, statistické programy). Mezi základní požadavky na technologii pro sběr dat je, aby sesbíraná data měla v nějakém formátu vyexportovat. Exportních formátů (datových rozhraní) existuje několik, existují i standardy pro výměnu lékařských dat mezi systémy (např. HL7, datový standard MZČR), jejichž použití však u nás zatím není příliš rozšířené. Převod dat mezi dvěma technologiemi či jejich vzájemnou komunikaci je pak nutné řešit případ od případu.

Samotné propojování různorodých datových zdrojů (například dat laboratorního informačního systému s daty hlášení pro pojišťovnu či libovolným národním registrem) za účelem společné analýzy má dvě základní řešení. Prvním je fyzické propojení všech dat a vytvoření tzv. datového skladu. Všechna data tak kromě původního umístění existují v kopii i v datovém skladu. Druhou možností je vytvořit analytický systém, který má přímý přístup ke všem potřebným primárním datům, data se nekopírují (Sujansky 2001). Pro první řešení lze využít zmíněná standardní datová rozhraní, kdy z každého datového zdroje vyexportujeme potřebná data a vložíme do datového skladu. Výhodou tohoto řešení je, že nevyžaduje od primárních systémů (NIS, LIS, atd.) žádné nové funkce a nemusí studovat jejich vnitřní datovou strukturu. Nevýhodou je, že datový sklad (kopie dat) zabírá nezanedbatelný diskový prostor, a navíc přesun většího objemu dat je i časově náročný. Datový sklad je také třeba neustále aktualizovat o recentní data.

Oproti tomu druhé řešení žádné kopie dat nevytváří a data ze zdrojů získává teprve ve chvíli, kdy jsou potřebná pro konkrétní analýzu. Data jsou tedy vždy „čerstvá“ a okamžitě k dispozici. Je nutné však vyřešit, jak se k těmto primárním datům přímo dostávat. Vnitřní struktura informačních systémů je velmi různorodá a nepomůžou nám zde žádné standardy. Je proto nezbytná přímá komunikace s autory těchto systémů. Ti doplní systémy o tzv. rozhraní, tedy sadu funkcí, které je možné využít pro

přímý přístup k datům. Vše je samozřejmě nutné ošetřit po bezpečnostní stránce, aby se k datům nemohla dostat neoprávněná osoba. Druhé řešení je tedy ve výsledné práci s daty efektivnější, na druhou stranu je však jeho zprovoznění podstatně dražší a zdoluhavější (Mendonca *et al.* 2001).

Doposud jsem hovořili obecně o medicínských datech, o možnostech jejich sběru a analýzy. Nyní je vhodné zmínit některá specifika dat onkologických. Data onkologických pacientů dostávají novou dimenzi, pokud se záznamy o provedených výkonech a aplikacích léků doplní o aktuální stav nádorového onemocnění (fáze onemocnění). Vývoj nádorového onemocnění lze ve většině případů velmi obecně rozdělit na následující základní fáze:

- Fáze diagnostiky
- Fáze primární terapie
- Fáze relapsu či progresu

U konkrétních onkologických diagnóz či v speciálně zaměřených studiích lze navrhnout i detailnější strukturu fázování, např. fáze neoadjuvantní terapie, fáze pooperační a podobně. U hematologických onemocnění je k uvedeným fázím nutné přidat také fázi transplantační. Uvedené fázování onemocnění je naprosto nezbytné zohledňovat při analýzách aplikované léčby. V opačném případě můžeme dojít k značně zavádějícím závěrům. Toto jednoduché ale velmi přínosné rozčlenění vývoje onkologického onemocnění je k uvedeným fázím v dnešní době standardně zabudováno do nemocničních informačních systémů a ani datový standard MZČR s ním ve stávající verzi nepočítá. Pokud chceme údaje o fázích onemocnění získat, nezbyvá nám než namáhavě prohledávat lékařskou dokumentaci a jednotlivé lékařské zprávy, což je ovšem v praktickém provozu ve větším objemu nerealizovatelné.

Technologicky však samozřejmě není problém implementovat fázování onemocnění či léčby přímo do nemocničního informačního systému. Tato více či méně nákladná modifikace by umožnila vytvořit systém, který by automaticky prolínal dostupná data (zmíněná data LIS a pojišťoven) a poskytoval lékařům nový, ucelený pohled na pacienta. Lékař (ale i manager zdravotnického zařízení) by mohl dostat do rukou nástroj, který by mu značně usnadňoval rozhodování už jenom tím, že by mu poměrně obsáhlý soubor nejrůznějších údajů o pacientovi zobrazil v prakticky navrhnutých prezentačních oknech. I bez aplikací náročných analytických funkcí má prosté, ale efektivně rozvržené zobrazení dat nemalý přínos pro proces rozhodování.

Pokud mluvíme o rozhodování na základě sbíraných dat, nemáme na mysli pouze rozhodování na úrovni konkrétního lékaře či managementu konkrétního zařízení, ale v dalších stupních se už dostáváme k procesu rozhodování na úrovni nastavování standardů léčby či porovnávání dostupných postupů (technologií). A to již není nic jiného než proces HTA, jehož významnou složkou sběr a hodnocení dat nepochybně je. Samozřejmě pro proces HTA je nutné zajistit sběr dat odpovídající kvality a objemu. Z pohledu informačních technologií to sice může být o stupeň vyšší úroveň než sběr dat pro lokální potřebu, rozhodující je však vůle sběr dat zahájit a podporovat.

Závěr

V tomto textu jsme se pokusili čtenáře seznámit s problematikou sběru dat, s možnostmi dnešních technologií, které sběr dat umožňují, nastínilí jsme problémy, které sběr medicínských dat provázejí. Je jasné, že podrobnější rozbor tohoto tématu by mohl zaplnit celou publikaci. Cílem textu bylo pouze uvést čtenáře do problematiky, upozornit zájemce o zprovoznění sběru dat či o nákup vhodné technologie na možné problémy a úskalí. Zároveň byl zmíněn i možný směr sběru a vyhodnocování onkologických dat, který by byl realizovatelný v prostředí dnešní onkologie a který by umožnil nastartování na data náročného procesu HTA.

Literatura

- Haug, P. J., Rocha, B. H. S. C., and Evans, R. S. (2003). Decision support in medicine: lessons from the system. *International Journal of Medical Informatics* **69**, 273-284.
- Jerez-Aragones, J. M., Gomez-Ruiz, J. A., Ramos-Jimenez, G., Munoz-Perez, J., and Alba-Conejo, E. (2003). A combined neural network and decision trees model for prognosis of breast cancer relapse. *Artificial Intelligence in Medicine* **27**, 45-63.
- Lavrac, N. (1999). Selected techniques for data mining in medicine. *Artificial Intelligence in Medicine* **16**, 3-23.
- Mendonca, E. A., Cimino, J. J., Johnson, S. B., and Seol, Y.-H. (2001). Accessing Heterogeneous Sources of Evidence to Answer Clinical Questions. *Journal of Biomedical Informatics* **34**, 85-98.
- Mikkelsen, G., and Aasly, J. (2002). Manual semantic tagging to improve access to information in narrative electronic medical records. *International Journal of Medical Informatics* **65**, 17-29.
- Mikkelsen, G., and Aasly, J. (2003). Narrative electronic patient records as source of discharge diagnoses. *Computer Methods and Programs in Biomedicine* **71**, 261-268.
- Ngan, P. S., Wong, M. L., Lam, W., Leung, K. S., and Cheng, J. C. (1999). Medical data mining using evolutionary computation. *Artificial Intelligence in Medicine* **16**, 73-96.
- Perner, P., Perner, H., and Muller, B. (2002). Mining knowledge for HEp-2 cell image classification. *Artificial Intelligence in Medicine* **26**, 161-173.
- Setiono, R. (1996). Extracting rules from pruned networks for breast cancer diagnosis. *Artificial Intelligence in Medicine* **8**, 37-51.
- Smith, A. E., Nugent, C. D., and McClean, S. I. (2003). Evaluation of inherent performance of intelligent medical decision support systems: utilising neural networks as an example. *Artificial Intelligence in Medicine* **27**, 1-27.
- Sujansky, W. (2001). Heterogeneous Database Integration in Biomedicine. *Journal of Biomedical Informatics* **34**, 285-298.
- Takeda, H., Matsumura, Y., Nakajima, K., Kuwata, S., Zhenjun, Y., Shanmai, J., Qiyan, Z., Yufen, C., Kusuoka, H., and Inoue, M. (2003). Health care quality management by means of an incident report system and an electronic patient record system. *International Journal of Medical Informatics* **69**, 285-293.
- Tange, H. J., Hasman, A., de Vries Robbe, P. F., and Schouten, H. C. (1997). Medical narratives in electronic medical records. *International Journal of Medical Informatics* **46**, 7-29.
- van der Haak, M., Wolff, A. C., Brandner, R., Drings, P., Wannenmacher, M., and Wetter, T. (2003). Data security and protection in cross-institutional electronic patient records. *International Journal of Medical Informatics* **70**, 117-130.
- Wang, S. J., Middleton, B., Prosser, L. A., Bardou, C. G., Spurr, C. D., Carchidi, P. J., Kittler, A. F., Goldszer, R. C., Fairchild, D. G., and Sussman, A. J. (2003). A cost-benefit analysis of electronic medical records in primary care. *The American Journal of Medicine* **114**, 397-403.