

DOTAZNÍKY O KVALITĚ ŽIVOTA

MAREK MALÝ

ABSTRAKT. In the last decade, the assessment of quality of life became one of the important endpoints in epidemiological studies, especially in clinical trials. The aim of this paper is to point out the properties and psychometric features of questionnaires for measuring this highly subjective concept and to address the statistical issues present in analysis of quality of life data. The problems arising from the large number of variables measured, repeated measurements, and from missing data are discussed and the suitable methods for their treatment are mentioned.

Абстракт. В этой статье изучается как описывать качество жизни больных.

1. ÚVOD

V posledních 10–15 letech se v lékařské oblasti velmi zvýšil zájem o posouzení kvality života pacientů. V různých epidemiologických studiích, ale zejména v klinických pokusech je nyní značná váha přikládána nejen délce života, ale i jeho kvalitě, jejíž hodnocení se stalo jedním ze standardních výstupů. Statistické aspekty měření kvality života jsou nesmírně různorodé, protože se dotýkají jak procesu zjišťování, tak hodnocení a interpretace dat, k jejichž typickým rysům patří zejména mnohorozměrnost a opakovaná měření. Některých aspektů bychom se chtěli dotknout v tomto příspěvku.

K hlavním faktorům ovlivňujícím kvalitu života se počítají zdravotní stav, věk, sociální a ekonomické podmínky, záležitosti duchovního rázu a v neposlední řadě i přání, životní cíle a očekávání toho kterého jedince a jeho zázemí. Dá se říci, že je to pocit vznikající z rozdílu přání a míry jejich (ne)splnění v životě. Mnohé studie ukázaly, že se tato subjektivní výpověď může silně lišit od objektivní lékařské informace a že ani lékař ani nejbližší příbuzní nemocného nejsou schopni jeho odpověď dobře odhadnout. Last (1995) definuje kvalitu života jako subjektivně vnímanou úroveň, jíž lidé hodnotí své fyzické, emocionální a sociální schopnosti.

Kaplan a Bush (1982) zavedli pojem kvalita života ve vztahu ke zdraví (health-related quality of life; HRQoL), aby vymezili tu část kvality života, která je prvotně určována zdravím jedince a zdravotní péčí a která může být ovlivněna klinickými intervencemi. HRQoL vychází hlavně z oblasti sociální, tělesné a psychologické funkce jedince, úrovně chorob, schopnosti postarat se o sebe a citové pohody.

Vzhledem k povaze a obsažnosti pojmu HRQoL vykrystalizoval jako nejvhodnější nástroj ke zjišťování kvality života dotazník, nejčastěji vyplňovaný samotným pacientem při kontaktu se zdravotnickým zařízením. Dotazníky běžně obsahují cca 20-150 otázek seskupených do několika tzv. dimenzí či domén, které korespondují s výše

2000 *Mathematics Subject Classification*. Primary 62P10.

Klíčová slova. Kvalita života pacientů, klinické pokusy, výběrová šetření.

zmíněnými oblastmi lidského bytí (např. od fyzické k duševní). Dnes je známo minimálně 800 funkčních dotazníkových nástrojů, orientovaných buď obecně na velmi širokou škálu nemocí, anebo specificky na určitou nemoc, populaci, dimenzi.

Přehledný souhrn informací o nástrojích pro měření kvality života poskytuje několik základních publikací (McDowell, Newell, 1987; Fallowfield, 1990; Bowling, 1996; Bowling, 1997; Spilker, 1996). K velmi rozšířeným obecným dotazníkům patří např. SF-36 (Ware, Sherbourne, 1992), který obsahuje 36 otázek v 8 dimenzích a byl přeložen do mnoha jazyků včetně češtiny (Vondra a spol., 1999; Skalská a spol., 2000).

Kvalita života se mění v čase v závislosti na změnách samotného pacienta i jeho okolí. Je ovlivňována věkem, nemocí a její léčbou i tím, jak se jedinec vyrovná s různými stavy a stadii, jimiž prochází. Různé dimenze jsou navíc různě citlivé k různým podnětům a i rychlost, s jakou se změna kvality promítne do měření, závisí na povaze dimenze. Např. u pacientů s různými stupni chronické obstrukční plicní nemoci reagují na zhoršování stavu velmi citlivě fyzické domény, ale hodnocení duševní domény často není extrémně špatné ani u těch, kteří jsou zcela odkázáni na domácí kyslíkový přístroj. Kvalita života je tedy mnohorozměrný, subjektivní a dynamický ukazatel, v psychometrii označovaný jako konstrukt. Můžeme jej měřit pouze nepřímou a pouze v určitém (malém) počtu časových intervalů (Schumacher, Olszewski, 1991).

K oblastem, v kterých nacházejí míry kvality života uplatnění, patří

- 1) Klinická praxe - screening a monitorování psychosociálních problémů v rámci individuální péče o pacienta, výběr ošetření
- 2) Populační studie zaměřené na vnímání zdravotních problémů, dopad rizikových faktorů
- 3) Hodnocení výzkumných studií a zdravotní péče (zde často jako finální míra),
- 4) Klinické pokusy - porovnání účinnosti léků, hodnocení průběhu nemoci, vyhledání rozhodujících zdravotních činitelů
- 5) Hodnocení finanční nákladnosti a efektivity léčby, určení priorit při alokaci zdrojů.

Diskriminační vlastnosti dotazníku se uplatní v průřezové studii, kdy je potřeba rozlišit kvalitu života v různých skupinách nemocných v jednom časovém okamžiku, evaluační vlastnosti poslouží při studiu změn kvality života v čase ve studiích longitudinálního typu a při predikci jistých ukazatelů zdravotního stavu včetně délky přežívání. Pro různé účely je nutná rozdílná struktura, úrovně proměnných.

2. OTÁZKY KONSTRUKCE DOTAZNÍKŮ

Při konstrukci nových dotazníků se zpravidla vychází ze širšího počátečního okruhu otázek shromážděných na základě vyjádření expertů či cílených pohovorů s pacienty (Donner, 1997). Téměř vždy se používají otázky uzavřené, kdy respondent vybírá pouze z předepsaných variant odpovědí. Uzavřené alternativní otázky připouštějí jen dvě varianty odpovědi, selektivní nabízejí výběr z více variant - v tomto kontextu většinou z 3-7 odpovědí s logickým uspořádáním od nejmenšího po největší dopad nemoci a dávají tak vzniknout ordinální veličině.

Kromě obsahové správnosti, relativní stručnosti a snadnosti vyplňování musí každý dotazník splňovat základní psychometrické vlastnosti, zejména validitu a reliabilitu (Nunnally, Bernstein, 1994). Ty by se měly vždy znovu ověřit pro novou jazykovou mutaci i pro použití v kontextu jiného onemocnění. Validita udává, do jaké míry měří použitý dotazník skutečně to, co zamýšlí měřit. Jedním z aspektů

validity je také citlivost (ke změně). Citlivost je síla metody, míra schopnosti detekovat u konkrétního jedince klinicky významnou změnu zdravotního stavu, v čase, v průběhu studie. Spolehlivost (reliabilita) je mírou stability, tedy schopnosti dotazníku poskytovat stejné výsledky, kdykoli je měření opakováno za stejných podmínek. Říká, zda měření nebo dotazník jsou opakovatelné (reprodukovatelné). Reliabilita je nutnou, ale nikoli postačující podmínkou validity. Se spolehlivostí je úzce spjata vnitřní konzistence, která ukazuje, do jaké míry spolu otázky v rámci jedné dimenze souvisejí.

3. OVĚŘOVÁNÍ VALIDITY A RELIABILITY

V literatuře je popsáno kolem 20 typů validity. Jejich hodnocení je velmi obtížné a do značné míry individuální podle typu problému (Streiner, Norman, 1995; Říčan, 1977), často vychází z korelačních koeficientů a faktorové analýzy (Juniper a spol., 1997). Van Knippeneberg, de Haes (1988) analyzují způsoby hodnocení validity užívané v kontextu kvality života. V rámci *pojmové validity* se zkoumá, do jaké míry měří test právě to, co se očekává, že měří, tedy to, k čemu byl konstruován. Pokud se např. studovaná vlastnost mění s věkem, měla by se tato změna odrazit v měření pojmové validity. *Obsahová validita* popisuje, zda test přiměřeně pokrývá celou studovanou oblast. *Kriteriální validita* empiricky zkoumá stupeň asociace či korelace posuzovaného nástroje a určitého externího kritéria. Kritériem by, obecně vzato, měla být proměnná, kterou se testem snažíme diagnostikovat či predikovat, či která je obecně uznaným absolutním ("zlatým") standardem. V případě měření kvality života však není žádná taková jasně definovaná a přímo pozorovatelná míra dostupná, protože se jedná o subjektivní, případně experimentální fenomén. V praxi se často se na místě standardního kritéria používá některý z nejrozšířenějších déle užívaných dotazníků. Účinnou technikou, která však vyžaduje speciální plán studie, je přístup "multitrait-multimethod" [MTMM] (Campbell, Fiske, 1959), kdy jsou dvě či více různých (a pokud možno málo souvisejících) dimenzí měřeny dvěma či více metodami. Získá se tak srovnání kvalitativně různých typů korelací. Nejvyšší by měly být reliability všech nástrojů a různé míry stejné dimenze by si měly více odpovídat než různé míry různých dimenzí, mezi nimiž by vlastně neměl být vztah (Streiner, Norman, 1995). Pro výpočet citlivosti se někdy používá koncepce nejmenšího závažného rozdílu (minimal important difference), což je nejmenší změna ve skóru určité domény, která je již klinicky anebo z pohledu pacienta považována za významnou (Jaeschke a spol., 1989; Juniper a spol., 1994).

Jednou z metod zjišťování reliability (Řehák, 1998) je opakované vyplnění dotazníku (postup test-retest) po několika týdnech, kdy se již snížil efekt zapamatování původních odpovědí a ještě nedošlo ke změně zkoumané vlastnosti (zdravotního stavu). Mírou reliability je pak korelační koeficient. Při jednorázové distribuci dotazníku lze vyšetřovat vnitřní konzistenci, kdy se hodnotí stejnoměrný vztah dílčích otázek ke skóru dimenze. Je nutno si uvědomit, že délka testu zpravidla zvyšuje reliability. Základním ukazatelem je Cronbachovo alpha (Cronbach, 1951), které lze vyjádřit jako $\alpha = \frac{k\overline{cov}/\overline{var}}{1+(k-1)\overline{cov}/\overline{var}}$, kde \overline{cov} je průměrná kovariance mezi k otázkami zkoumané dimenze a \overline{var} je jejich průměrný rozptyl. Koeficient vychází z úvahy, že položky v rámci jedné dimenze mají do určité míry stejný předmět, zabývají se stejnou entitou, a průměrná korelace stanovuje do jaké míry jde o stejný předmět. Je možno jej interpretovat jako korelaci mezi pozorovaným skórem dimenze a kterýmkoli jiným testem, který by obsahoval stejný počet otázek z hypotetického souboru všech možných otázek vázaných k dané dimenzi. Podobné úvahy vedou

k hodnocení reliability pomocí Pearsonova korelačního koeficientu mezi jednotlivou položkou a skórem příslušné dimenze spočteným po jejím vypuštění ze zbývajících položek (item-scale correlation). Na základě tzv. "split-half" metody se celý dotazník ve všech dimenzích rozdělí do dvou částí (např. na sudé a liché otázky) a jako odhad vnitřní konzistence se počítá korelace mezi těmito dvěma částmi pomocí tzv. Spearmanova-Brownova vzorce (Lord, Novick, 1968). Nevýhodou je, že metoda silně závisí na způsobu rozdělení dotazníku.

4. ŠKÁLOVÁNÍ ODPOVĚDÍ

Je nezbytné, aby jednotlivé testy a dotazníky měly nejen standardizovaný obsah (Jones, 1997), ale i proces ohodnocení odpovědí, protože jedině tak jsou výsledky interpretovatelné a porovnatelné navzájem mezi studii. Jakékoli změny obsahu či vyhodnocování mohou snadno přivodit vychýlení skóre (bias) a ovlivnit validitu i reliability.

Pro kódování odpovědí se většinou používají přirozená čísla 1,2,3,... . Ta mohou sloužit i jako skóry pro zpracování, pokud je realistický předpoklad lineárního vztahu mezi skóry a jim odpovídajícími kategoriemi zdravotních problémů. Pak se lze na použitém stupnici dívat nejen jako na pořadovou (ordinální), ale i jako na intervalovou (která má stanovenou míru vzdálenosti mezi pozicemi na stupnici). Opačný případ vede k neekvidistantním ohodnocením (vlastně k vážení odpovědí) a skóry jsou odvozovány empiricky na základě nějakého zkušební vzorku pacientů pomocí různých analytických metod, např. faktorové analýzy či regrese (Olschewski, Schumacher, 1990), distanční analýzy a někdy bohužel i dosti libovolně bez racionálního podkladu. Někteří autoři doporučují přihlížet při užití čistě matematických metod i ke klinickému posouzení.

Význam přidělování skóre je naprosto zásadní. Jejich rozsah by neměl být extrémní, protože jinak může dojít k nežádoucímu potlačení či upřednostnění některých stavů a tím i k mimořádným závěrům. Je zřejmé, že různými postupy se ani zdaleka nedospěje k totožným vahám. Cox a spol. (1992) ze statistických pozic složitější metody vážení kritizují a dávají přednost prostým celočíselným skórum, případně doporučují vycházet z diskriminační analýzy, normálních skóre (pro odpovědi přibližně symetricky rozložené kolem centrální hodnoty), exponenciálních skóre (pro odpovědi koncentrované na jednom konci). Někteří autoři volí mírný odklon od celočíselných skóre ke skórum typu 1, 2.1, 2.9, 4.3, atd.

5. STATISTICKÉ OTÁZKY PŘI MĚŘENÍ KVALITY ŽIVOTA

Při řešení studií o kvalitě života je nutno preferovat statistické postupy, které jsou schopny vzít v úvahu mnohorozměrnou ordinální povahu dat, strukturu jejich závislosti v čase (sousední měření téhož pacienta, zejména když jsou nepřilíš časově vzdálená, jsou nepochybně kladně korelovaná) a poměrně častý výskyt chybějících pozorování (cenzorování).

Kdybychom se snažili posuzovat rozdíly mezi skupinami pacientů, či vývoj v čase pro každou z mnoha otázek dotazníku separátně, byly by výsledky dosti nepřehledné, ale hlavně by došlo k výraznému posunu hladiny významnosti v důsledku mnohonásobného srovnávání a metodologické problémy zůstávají i při použití korekcí typu Bonferroniho nerovnosti (Miller, 1981). Vhodný kompromis, při němž je omezen počet testovaných hypotéz a zjištěná informace kondenzována, ale zároveň zachována ve své mnohorozměrnosti, představují skóry pro jednotlivé dimenze, případně doplněné celkovým skórem ze všech otázek. Ty lze i snáze interpretovat. Celkový ukazatel

však pozbývá smyslu, pokud dochází k interakci mezi léčbou a dimenzemi, tj. pokud s pozitivní změnou v jedné dimenzi přichází negativní změna v jiné. Celkové skóre pak mnohdy není schopno zachytit jakýkoli efekt léčby. Vzhledem k tomu, že jednotlivé položky dimenze mají zhruba stejný vztah k měřenému aspektu zdravotního stavu a že žádná z položek není využita ve více než jedné dimenzi, doporučuje se vytvořit skóre dimenze jako nevážený součet či průměr skóre odpovědí a vyjádřit jej procentuálně jako podíl maximálního možného hodnocení.

Jako hlavní statistický nástroj pak často slouží (mnohorozměrná) analýza rozptylu (resp. analýza kovariance) s opakovanými měřeními, pokud u skóre dimenzí můžeme předpokládat normální rozložení, a zobecněné lineární modely v ostatních situacích (Agresti, 1989), mj. při specifikaci korelační struktury (Liang, Zeger, 1986). Flexibilní je hierarchický model s náhodnými koeficienty, který popisují Beacon a Thompson (1996). Složitější přístup, který vychází z teorie položkové analýzy, dokáže lépe zohlednit strukturu daného dotazníku, ordinální povahu proměnných a psychometrické požadavky uplatněné při tvorbě dotazníku (Douglas, 1999).

Chybějící pozorování a ztráta pacientů ze sledování (včetně neochoty či neschopnosti k další spolupráci) představují značný problém zejména proto, že mechanismus chybění často souvisí se stavem pacienta. Moderní statistická literatura rozlišuje tři hlavní mechanismy vzniku chybějících pozorování (Rubin, 1986). Údaj chybí "zcela náhodně", pokud pravděpodobnost, že k výpadku dojde, nezávisí na žádném jiném údaji, který byl nebo bude zaznamenán (např. nemocný přehlédl otázku, ošetřující sestra se na ni zapomněla zeptat). Údaj chybí "náhodně", pokud pravděpodobnost výpadku souvisí s předem známými doprovodnými údaji (věk, pohlaví) nebo závisí na dříve zjištěných skórech, ale nemá vztah k současným ani k budoucím (např. odpovědi chybí převážně u příslušníků jednoho etnika, či spíše neodpovídají pacienti, kteří měli při minulém stanovení horší HRQoL). Matematicky nejsložitější situace nastává, když data chybí "nenáhodně" a pravděpodobnost chybění může záviset nejen na dřívějších (známých) zjištěních, ale i na současných a budoucích (neznámých) skórech (např. kvalita života konkrétního pacienta je víceméně konstantní v průběhu všech dosud zaznamenaných měření, ale odmítnutí spolupráce přichází v okamžiku, kdy se HRQoL výrazně zhoršuje, takže proces zhoršování není v datech zachycen). Nenáhodné chybění může ale zároveň být z lékařského hlediska dobrým prognostickým faktorem závažné zdravotní události včetně úmrtí. Jedině v případě zcela náhodného chybění není zásadní chybou omezit zpracování jen na jedince s úplnými daty, jindy by tak snadno mohlo dojít třeba k podcenění vedlejších efektů nového léku, protože nezískáme údaje právě od probandů, kteří mají největší problémy.

Pro ošetření chybějících dat o kvalitě života bylo aplikováno a nově vyvinuto mnoho metod. Této tématice (v kontextu klinických pokusů při léčbě karcinomů) je např. věnováno celé trojčíslo časopisu *Statistics in Medicine* (Vol. 17, No. 5/6/7). Pokud u daného jedince v daném čase chybí jen některé údaje, je někdy možno je nahradit průměrným skórem z příbuzných otázek (zpravidla v rámci jedné dimenze). Troxel a spol. (1998) systematicky zpracovali situace, kdy pro určitý čas chybí celý dotazník. Shrnují řešení pro data chybějící zcela náhodně či náhodně v kontextu lineárního modelu se smíšenými efekty, zobecněného lineárního modelu a modelu GEE (generalized estimating equations - zobecněné rovnice odhadu; Zeger a spol., 1988). Nevychýlené odhady pro spojitá data s nenáhodným mechanismem chybění lze získat v modelu sdruženého rozložení kompletního datového vektoru a indikátoru chybění (Diggle, Kenward, 1994).

Alternativu k dosud uvedeným modelům představuje analýza opakovaných měření pomocí dvou typů tzv. sumárních měř (Senn a spol., 2000). V prvním případě probíhá primárně redukce dat odděleně pro každého jedince. Sumární ukazatel $S_{ik} = \sum_{j=1}^n w_j f(Y_{ijk})$ vznikne jako vážený součet n měření Y_{ijk} i -tého jedince v k -té skupině, transformovaných vhodnou funkcí f (Fairclough, 1997). Index j může probíhat buď časové okamžiky nebo jednotlivé dimenze. O'Brien (1984) navrhl neparametrický (f jsou pořadí a váhy $w_j = 1/n$) i parametrický (f jsou standardizované normální odchylky) test tohoto typu. Přístup Coxe a spol. (1992) vychází ze sumární míry představované plochou pod křivkou průběhu HRQoL v čase. Jindy lze za f zvolit rozdíl prvního a posledního pozorování jedince, průměr všech pozorování, atp. Při druhém přístupu k sumárním měřám se nejprve získají odhady mnohorozměrných parametrů odděleně pro každý z n časů či dimenzí a posléze pro každou skupinu jedinců sumární ukazatel $S_k = \sum_{j=1}^n w_j g(\beta_{kj})$, kde $g(\beta_{kj})$ jsou odhady průměru j -tého měření v k -té skupině, které jsou adjustované vzhledem k důležitým kovariantám (Fairclough, 1997).

Olschewski, Schumacher (1990) diskutují modelování kvality života vhodnými stochastickými procesy, vycházejícími z popisu průběhu nemoci několika stavy (se smrtí jako trvalým, absorbujícím stavem). Pravděpodobnosti přechodu mezi stavy lze dobře odhadovat, pokud je reálný předpoklad, že proces kvality života splňuje Markovskou vlastnost. Většinou je třeba pracovat s nehomogenním Markovským řetězcem, neboť pravděpodobnosti přechodu mezi stavy se mohou měnit v čase. Pokud je místo výpočtu pravděpodobností přechodu cílem srovnání skupin, lze propojit HRQoL s analýzou přežívání a vycházet z modelu proporciálního rizika, v němž jsou efekty ošetření zařazeny jako kovarianty (Kay, 1982).

6. KVALITA ŽIVOTA A ANALÝZA PŘEŽÍVÁNÍ

Komplexnější obraz života jedince s ohledem na jeho zdravotní stav a klady a záporů terapie poskytuje souběžné hodnocení kvality života s jeho "kvantitou", tj. délkou přežívání. Zároveň je tento přístup využíván k měření efektivity zdravotnických nákladů metodou cena-prospěch (cost-utility) (Drummond, 1991; Žáček, Holčík, 1992). Pro tyto účely byly vyvinuty jednodimensionální míry kvality života, tzv. indexy či "profily", které se pohybují v rozmezí od 1 (plné zdraví, maximální kvalita života) až po 0 (smrt) a které lze v závislosti na čase zobrazit tzv. Carlensovým vitagramem (Olschewski, Schumacher, 1990). Plocha pod touto křivkou byla použita jako vhodná sumární míra v původní definici tzv. kvalitativně upravených roků života QALY (quality-adjusted life years). Koncepce QALY vychází z myšlenky, že určitá doba optimálního zdraví má vyšší hodnotu než stejná doba života nižší kvality (Torrance, 1986). Jeden rok plného zdraví tak odpovídá dvěma rokům s indexem 0.5, což koresponduje s tím, že někdo může dávat přednost kratšímu bezproblémovému životu před delším s výraznými omezeními. Etické otázky, zvláště v souvislosti s ekonomickým využitím této charakteristiky jsou nasnadě. Ze statistického pohledu je kritizována přílišná redukce jevu, který je charakterizován vysokou interpersonální variabilitou, na jedinou hodnotu, která navíc vychází z nedostatečně přesných měření (úmrtnost, průměrných nákladů). Cox a spol. (1992) upozorňují na úskalí vznikající, když určitá léčba má protichůdné účinky z pohledu kvality života a přežívání. Pak míry těchto dvou fenoménů jednoznačně nesmí být kombinovány do jediného indexu.

Zásadní problémy se objevují, pokud v průběhu sledování dochází k cenzorování (Cox a spol., 1992). Jakou kvalitu života máme přiřadit neznámému úseku života pacienta? Jeden z přístupů, označený jako Q-TWiST (Time Without Symptoms and

Toxicity), definuje sérii úrovní zdravotního stavu, vypočte průměrný čas strávený v každé úrovni na základě "po částech" provedené analýzy přežívání a při výpočtu finální míry typu QALY váží každou úroveň odpovídající kvalitou života (Gelber a spol., 1995). Pro případy, kdy jsou přechody mezi úrovněmi stavu nezřetelné či kdy stavy adekvátně nekorrespondují s výkyvy v kvalitě života navrhli alternativní postup Glasziou a spol. (1998). Metoda je založena na separaci přežívací křivky, odhadnuté standardní Kaplan-Meierovou metodou, a tzv. funkce kvality života, odvozené z opakovaných individuálních stanovení kvality života. Ribauda a spol. (2000) navrhli sdruženou analýzu kvality života a přežívání v trojrozměrném normálním modelu.

LITERATURA

- [1] Agresti, A. (1989). A survey of models for repeated ordered categorical response data. *Stat. Med.* 8, 1209-1224.
- [2] Beacon, H.J., Thompson, S.G. (1996). Multi-level models for repeated measurement data: application to quality of life data in clinical trials. *Stat. Med.* 15, 2717-2732.
- [3] Bowling, A. (1996). *Measuring Disease: A Review of Disease-specific Quality of Life Measurement Scales*. Open University Press, Buckingham.
- [4] Bowling, A. (1997). *Measuring Health: A Review of Quality of Life Measurement Scales*, 2nd ed. Open University Press, Buckingham.
- [5] Campbell, D.T., Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psych. Bull.* 56, 81-105.
- [6] Cox, D.R., Fitzpatrick, R., Fletcher, A.E., Gore, S.M., Spiegelhalter, D.J., Jones, D.R. (1992). Quality-of-life assessment: Can we keep it simple? With discussion. *J. R. Statist. Soc. A* 155, 353-393.
- [7] Cronbach, L.J. (1951). Coefficient alpha and the internal structure. *Psychometrika* 16, 297-334.
- [8] Diggle, P., Kenward, M. (1994). Informative drop-out in longitudinal analysis (with discussion). *Appl. Stat.* 43, 49-93.
- [9] Douglas, J.A. (1999). Item response models for longitudinal quality of life data in clinical trials. *Stat. Med.* 18, 2917-2931.
- [10] Donner, C.F., Carone, M., Bertolotti, G., Zotti, A.M. (1997). Methods of assessment of quality of life. *Eur. Respir. Rev.* 7, 43-45.
- [11] Drummond, M. (1991). Economic studies. In: Holland, W.W., Detels, R., Knox, G., (eds.). *Oxford Textbook of Public Health, Vol. 2: Methods of Public Health*, 295-307. Oxford University Press, Oxford.
- [12] Fairclough, D. (1997). Summary measures and statistics for comparison of quality of life in a clinical trial of cancer therapy. *Stat. Med.* 16, 1197-1209.
- [13] Fallowfield, L. (1990). *The quality of life*. Souvenir, London.
- [14] Flanagan, J.C. (1982). Measurement of quality of life. *Arch. Phys. Med. Rehabil.* 63, 56-59.
- [15] Gelber, R.D., Cole, B.F., Gelber, S., Goldhirsch, A. (1995). Comparing treatments using quality-adjusted survival: the Q-TWiST method. *Am. Statist.* 49, 161-169.
- [16] Glasziou, P.P., Cole, B.F., Gelber, R.D., Hilden, J., Simes, R.J. (1998). Quality adjusted survival analysis with repeated quality of life measures. *Statistics in Medicine* 17, 1215-1229.
- [17] Jaeschke, R., Singer, J., Guyatt, G.H. (1989). Measurement of health status: ascertaining the minimal clinically important difference. *Controlled Clin. Trials* 10, 407-415.
- [18] Jones, P. W. (1997). Quality of life measurement: the value of standardisation. *Eur. Respir. Rev.* 7, 46-49.
- [19] Juniper, E.F., Guyatt, G.H., Willan, A., Griffith, L.E. (1994). Determining a minimal important change in a disease-specific quality of life questionnaire. *J. Clin. Epidemiol.* 47, 81-87.
- [20] Juniper, E.F., Guyatt, G.H., Streiner, D.L., King, D.R. (1997). Clinical impact versus factor analysis for quality of life questionnaire construction. *J. Clin. Epid.* 50, 233-238.
- [21] Kaplan, R.M., Bush, J.W. (1982). Health-related quality of life measurement for evaluation research and policy analysis. *Health Psychol.* 1, 61-80.
- [22] Kay, R. (1982). The analysis of transition times in multistate stochastic processes using proportional hazards regression models. *Comm. in Stat. - Theory and Methods* 11, 1743-1756.

- [23] van Knippenberg, F.C.E., de Haes, J.C.J.M. (1988). Measuring quality of life of cancer patients: psychometric properties of instruments. *J. Clin. Epidem.* 41, 1043-1055.
- [24] Last, J.M., ed. (1995). *A dictionary of epidemiology*. 3rd ed. Oxford University Press, New York.
- [25] Liang, K., Zeger, S.L. (1986). Longitudinal data analysis using generalised linear models. *Biometrics* 73, 45-51.
- [26] Lord, F.M., Novick, M.R. (1968). *Statistical Theories of Mental Tests*. Addison-Wesley, Reading, Massachusetts.
- [27] McDowell, I., Newell, C. (1987). *Measuring health: a guide to rating scales and questionnaires*. Oxford University Press, Oxford.
- [28] Miller, R.G., Jr. (1981). *Simultaneous statistical inference*. 2nd ed. Springer-Verlag, New York.
- [29] Nunnally, J.C., Bernstein, I.H. (1994). *Psychometric Theory*. 3rd edn. McGraw-Hill, New York.
- [30] O'Brien, P.C. (1984). Procedures for comparing samples with multiple endpoints. *Biometrics* 40, 1079-1087.
- [31] Olschewski, M., Schumacher, M. (1990). Statistical analysis of quality of life data in cancer clinical trials. *Stat. in Med.* 9, 749-763.
- [32] Ribaud, H.J., Thompson, S.G., Allen-Mersh, T.G. (2000). A joint analysis of quality of life and survival using a random effect selection model. *Statistics in Medicine* 19, 3237-3250.
- [33] Rubin, D.B. (1976). Inference and missing data. *Biometrika* 63, 581-592.
- [34] Řehák, J. (1998). Kvalita dat I. Klasický model měření reliability a jeho praktický aplikační význam. *Sociologický časopis* 34 (1), 51-60.
- [35] Říčan, P. (1977). *Úvod do psychometrie*. Skripta, Psychodiagnostické a didaktické testy, Bratislava.
- [36] Schumacher, M., Olschewski, M., Schulgen, G. (1991). Assessment of quality of life in clinical trials. *Stat. Med.* 10, 1915 - 1930.
- [37] Skalská, H., Sobotík, Z., Jezberová, D., Mareš, J. (2000). Use and evaluation of the Czech version of the SF-36 questionnaire self-reported health status of medical students. *Centr. Eur. J. Publ. Health* 8, 88-93.
- [38] Spilker, B., ed. (1996). *Quality of life and pharmacoeconomics in clinical trials*, 2nd ed. Lippincott-Raven Publishers, Philadelphia.
- [39] Streiner, D.L., Norman, G.R. (1995). *Health Measurement Scales*. Oxford University Press, Oxford.
- [40] Torrance, G.W. (1986). Measurement of health state utilities for economic appraisal: A review. *J. Health Econ.* 5, 1-30.
- [41] Troxel, A.B., Fairclough, D.L., Curran, D., Hahn, E.A. (1998). Statistical analysis of quality of life missing data in cancer clinical trials.
- [42] Vondra, V., Malý, M., Reisová, M. (1999). Kvalita života nemocných s chronickou obstrukční plicní nemocí, bronchiálním astmatem a s alergickou rýmou. *Respirace* 5 (3), 2-13.
- [43] Ware, J.E., Sherbourne, C.D. (1992). The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Med. Care* 30, 473-483.
- [44] Zeger, S.L., Liang, K.Y., Albert, P.S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics* 44, 1049-1060.
- [45] Žáček, A., Holčík, J. (1992). *Sociální lékařství II*. Skripta, Masarykova univerzita, Brno.