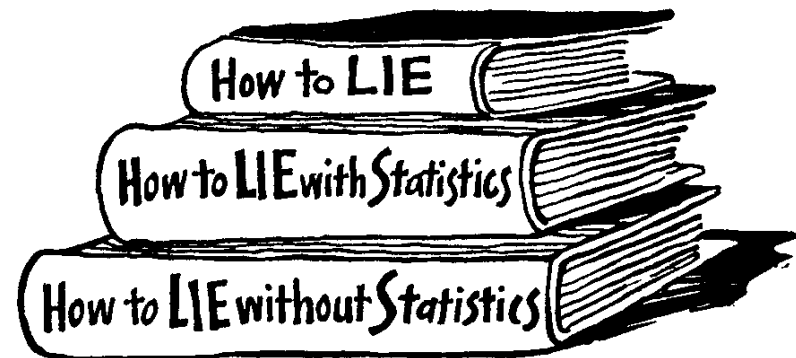


Úvod do práce s medicínskymi údajmi

Základy vedeckej práce



Úvod do metodológie

Ako vieme, že je to tak?

Pár slov na úvod o poznávaní

Metóda autority

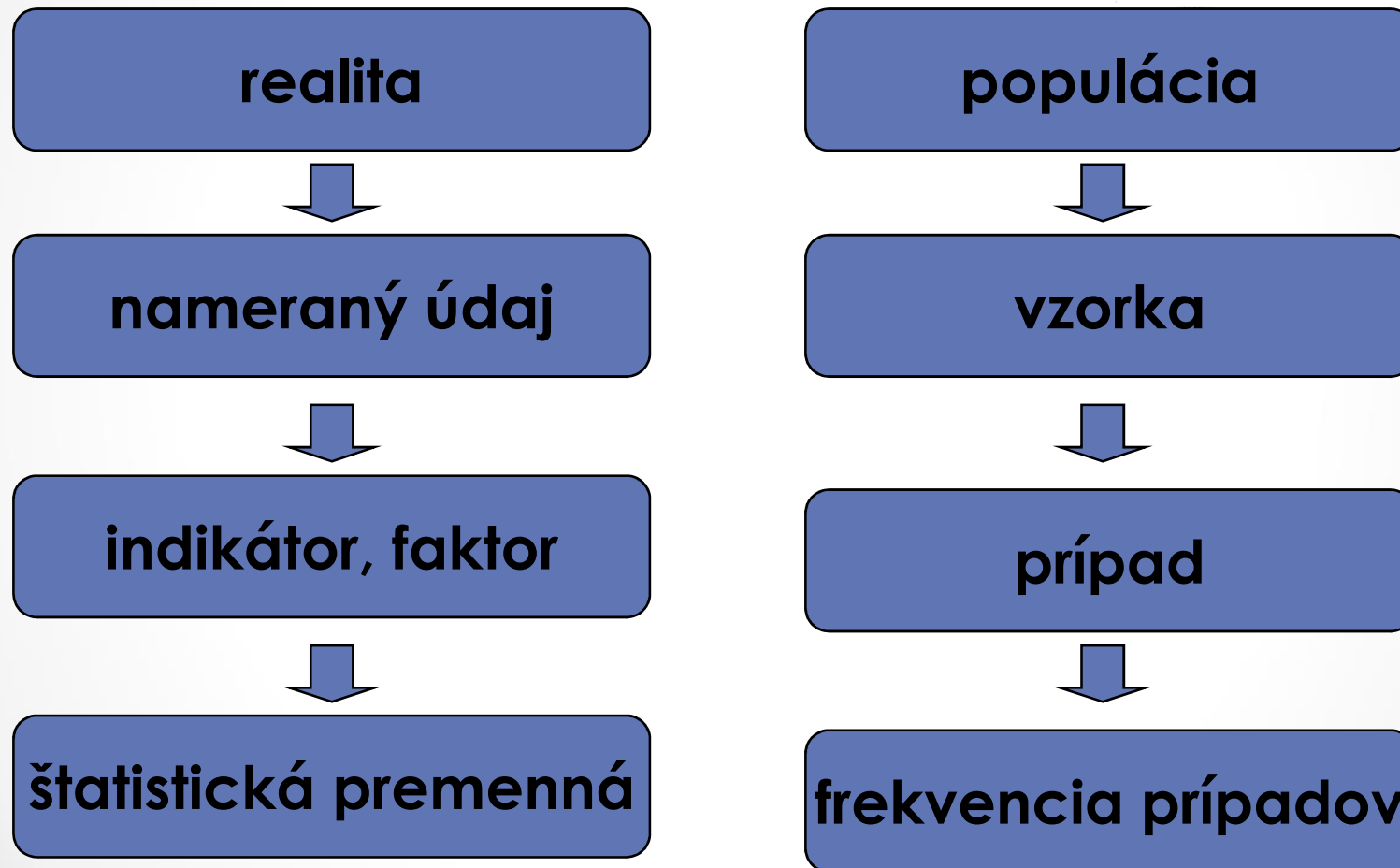
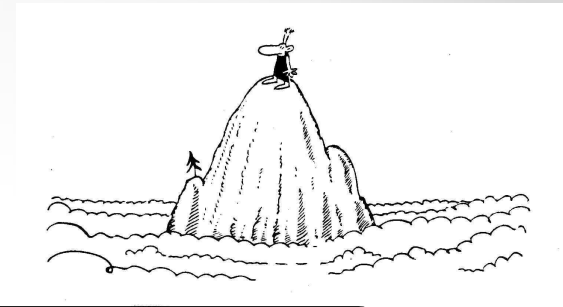
Metóda tradície

Vedecká metóda poznávania:

- (1) Systematické a organizované poznávanie
- (2) Empirická povaha
- (3) Objektívnosť a verifikovateľnosť (opakovateľnosť)
- (4) Sebakorekcia (nepretržité približovanie k pravde, nie pravda)

Cieľ = popísať, predvídať alebo vysvetliť

Osnova výskumu



Rizikové faktory výskytu HbB



Aké údaje potrebujeme?
Čo potrebujeme zmerať, zistiť?

Ako môžeme zmerať, zistiť **výskyt HbB**? Odkiaľ budú informácie? Aký formát bude mať tento **nameraný údaj**?

Čo sú **rizikové faktory HbB**? Ako môžeme zmerať, zistiť rizikové faktory HbB? Mali by sme niečo zohľadniť? **Skresľujúce premenné**?

Ako z týchto informácií urobíme **indikátor, faktor**?

Aký formát bude mať tento indikátor, faktor v štatistických analýzach? Akým spôsobom z neho urobíme **štatistickú premennú**?

realita



nameraný údaj



Indikátor, faktor



štatistická premenná

Rizikové faktory výskytu HbB

Protilátky v krvi, zdravotnícka dokumentácia, výpoveď respondenta, výpovede iných osôb, registre, iné zdroje

Výskyt HbB
Skresľujúce premenné (očkovanie, vek)
Priame rizikové faktory (rizikový sex, tetovanie, piercing, injekčné užívanie drog)
Nepriame rizikové faktory (oblasť, iné rizikové správanie, pobyt vo väzení, rómska etnicita, hodnoty, vzdelanie)

HbB (áno/nie), očkovanie (áno/nie), vek (15-35r, 35-45r), rizikové správanie (aspoň jedno/žiadne), oblasť (1-72), hodnoty (8 faktorov z FA), vzdelanie (zákl., maturitné, VŠ)

Výskumná otázka

$$A \rightarrow B$$

Zvyšuje/znižuje výskyt faktoru A

riziko výskytu (závažnosti) javu B?

Akým spôsobom sa na vzťahu $A \rightarrow B$ podieľajú **iné faktory** (skresľujú, sprostredkujú, ovplyvňujú)?

Na akú populáciu sa majú zistenia zovšeobecniť?

Typ dát

Organizácia a sumarizácia dát je prvý a najdôležitejší krok pre:

- **analýzu**
- **prezentáciu**
- **interpretáciu**

Proces štatistickej analýzy, prezentovania a interpretovania dát **závisí na type dát** a komplexite ich informácie.

Dáta sú čísla, ktoré sú výsledkom merania alebo počítania.

Typ dát

Kategorické dáta (diskrétne)

- nominálne dáta

(prípád/ne-prípád, pohlavie, diagnóza, krvná skupina)

- poradové (ordinálne) dáta

(vzdelanie, subjektívne hodnotenie zdravia)

Intervalové dáta (kontinuláne, škálové)

(vek, váha, výška, krvný tlak)



Prečo je dôležité uvažovať o type dát už pri plánovaní výskumu?

Deskriptívna štatistika

Deskriptívna štatistika – je analýza, ktorá sa používa na popis základných charakteristík dát. Ponúka jednoduchý prehľad o vzorke alebo výsledkoch meraní (**centralita, rozptyl**), ako je priemerná hodnota dát a aká je distribúcia, ktorú nadobúdajú.

Miera centrality – reflektuje centrálnu tendenciu dát:

- Priemer – súčet pozorovaní delený počtom pozorovaní
- Medián – stredná hodnota série meraní
- Modus – najčastejšie sa vyskytujúca hodnota vo všetkých pozorovaniach

Miera rozptylu – reflektuje variabilitu dát:

- Rozpätie – vzdialenosť medzi najnižšou a najvyššou hodnotou
- Percentily a kvartily – rozdelenie dát na časti (100 alebo 4)
- Variancia – miera rozdielu každej hodnoty od priemernej hodnoty (mocnina priemeru vzdialenosti každého merania od priemeru)
- Štandardná odchýlka – miera rozptylu vyjadrená ako odmocnina variancie

Priemer, medián, modus

VAR00001

	Frequency	Valid Percent	Cumulative Percent
300,00	1	20,0	20,0
450,00	4	16,0	36,0
453,00	1	4,0	40,0
455,00	2	8,0	48,0
500,00	1	4,0	52,0
521,00	1	4,0	56,0
531,00	1	4,0	60,0
534,00	1	4,0	64,0
541,00	1	4,0	68,0
623,00	1	4,0	72,0
654,00	2	8,0	80,0
724,00	1	4,0	84,0
764,00	1	4,0	88,0
3000,00	2	8,0	96,0
6000,00	1	4,0	100,0
Total	25	100,0	

Firma má 1 šéfa, 2 zástupcov, 2 vedúcich dielní a 20 zamestnancov. V tabuľke sú uvedené ich priemerné čisté zárobky.

Aká miera (priemer, medián, modus) lepšie popisuje platovú situáciu?

Priemerný zárobok je 908 € +/- 1275 €

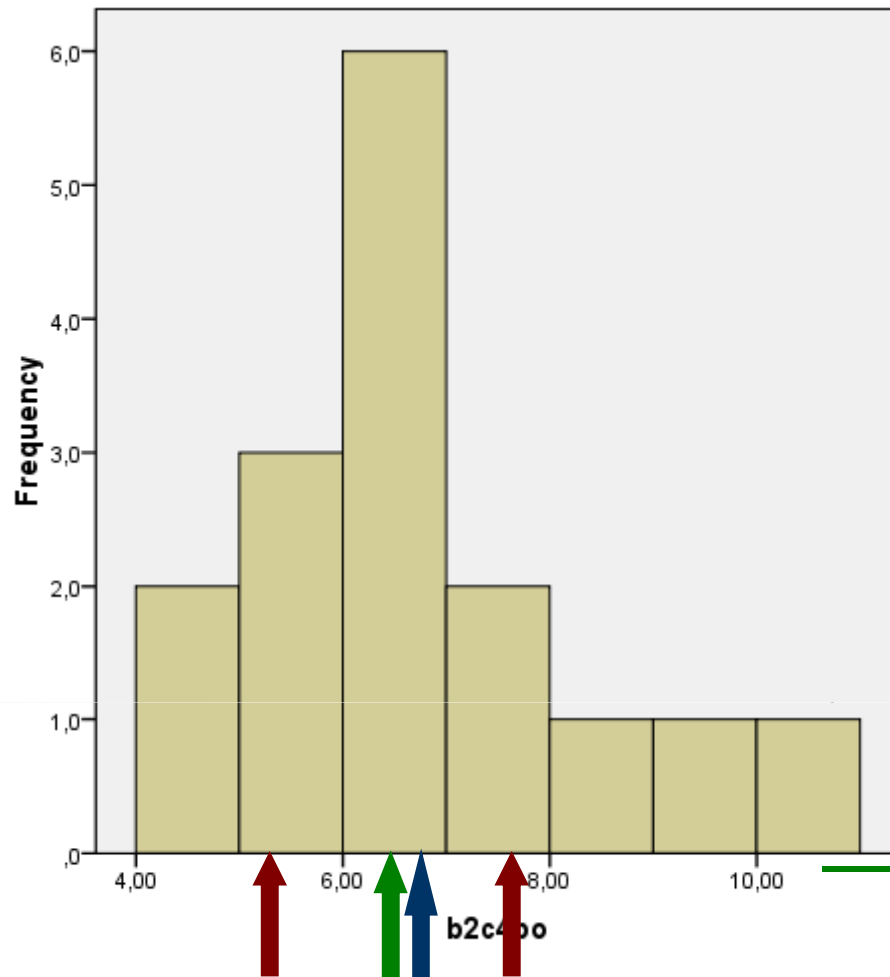
Medián = 500 € (polovica má max 500 €)

Modus = 300 € (najčastejší príjem)



Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
VAR00001	25	5700,00	300,00	6000,00	908,3600	1274,49310	1624332,657
Valid N (listwise)	25						



Medián = 6.5

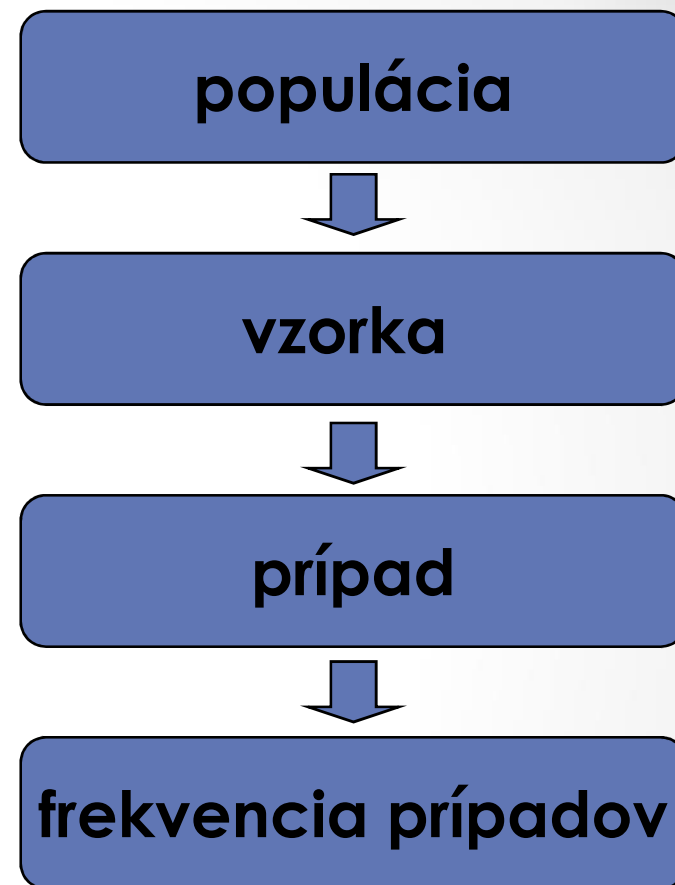
Priemer = 6.8

b2c4po

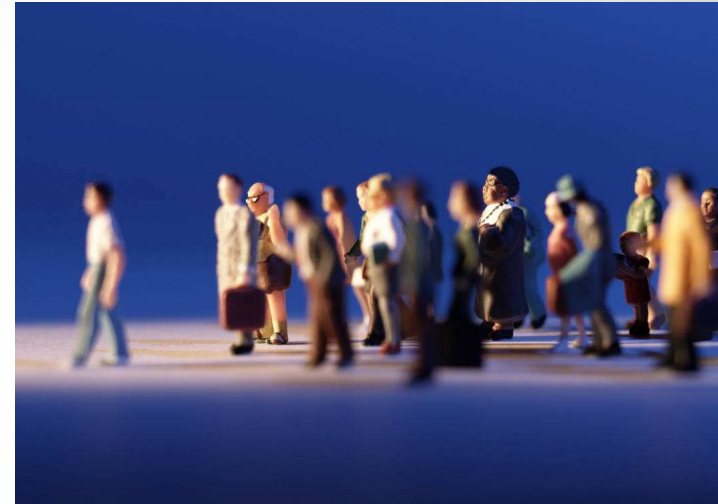
	Frequency	Percent	Valid Percent	Cumulative Percent
4,32	1	5,9	6,2	6,2
4,58	1	5,9	6,2	12,5
5,44	2	11,8	12,5	25,0
5,64	1	5,9	6,2	31,2
6,08	1	5,9	6,2	37,5
6,32	1	5,9	6,2	43,8
6,50	1	5,9	6,2	50,0
6,52	1	5,9	6,2	56,2
6,62	1	5,9	6,2	62,5
6,68	1	5,9	6,2	68,8
7,58	1	5,9	6,2	75,0
7,88	1	5,9	6,2	81,2
8,92	1	5,9	6,2	87,5
9,62	1	5,9	6,2	93,8
10,20	1	5,9	6,2	100,0
Total	16	94,1	100,0	
Missing System	1	5,9		
Total	17	100,0		



"I'm getting too old to hunt. Let's contact that professor who wanted to study us and eat the students he sends to do the research."



Populácia



Na akú skupinu ľudí chcete, aby sa vzťahovali Vaše zistenia?



Akú populáciu (ne)reprezentujú pacienti na Vašom oddelení, Vašej ambulancii?

Populácia

Populácia (základný súbor) je skupina ľudí so spoločnými charakteristikami (bydlisko, vek, príslušnosť k regiónu, udalosti) a môže byť **fixná** (definovaná na trvalo napr. po srdcovom infarkte) alebo **dynamická** (definovaná aspektom, ktorý sa môže meniť napr. ochorenie na chrípku).

Je to skupina ľudí, na ktorých by sa mali výsledky výskumu vzťahovať – dať **zovšeobecniť**.

Náhradná populácia – nie vždy je populácia na ktorej chceme niečo odhadovať tá istá ako populácia, z ktorej sa vyberá vzorka.

Vzorka



Prečo to nezmerať u všetkých?

Nesúhlas respondenta. Nedostupnosť respondenta. Obťažujúce vyšetrenie. Rizikové procedúry. Nákladné vyšetrenia. Nákladnosť pre veľkosť populácie ...



Prečo len odhadovať na základe merania v istej skupine?

Pre vedcov je reprezentatívny výber jedinou podloženou procedúrou pre to, aby výber jedincov mohli použiť na zovšeobecnenie vzhľadom k populácii.

Reprezentatívny výber nám môže nahradiť skúmanie celej populácie!

Reprezentatívny výber

Zlatým štandardom pre reprezentatívny výber je **náhodný výber** = každý jedinec relevantnej populácie má rovnakú šancu byť vybratý do vzorky.



Prečo náhodný výber?

Nevieme a nedokážeme zohľadniť všetky atribúty, ktoré reprezentujú populáciu.

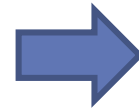


Prečo náhodný výber niekedy nestačí?

Dôvodom sú zriedkaví respondenti, ktorí sú ale dôležitý, logistika zberu údajov, ...

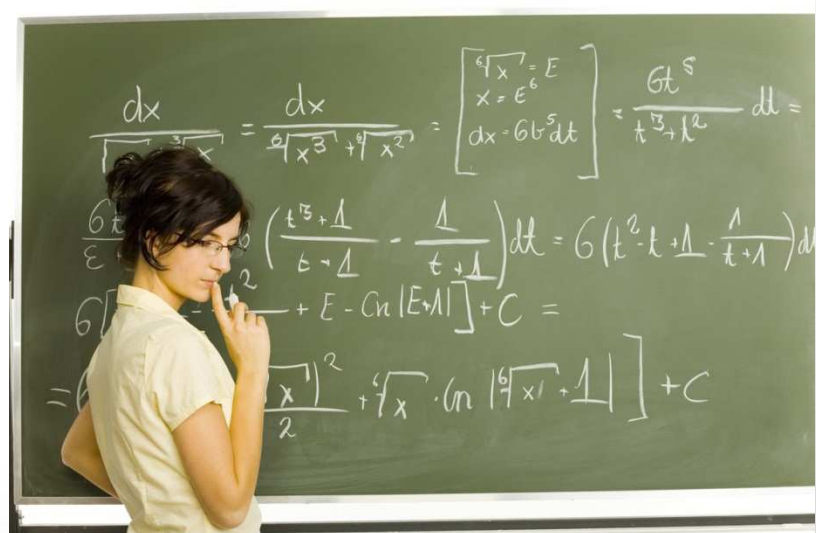
Meranie vplyvu expozície na sledovaný parameter

Expozícia
(áno/nie)

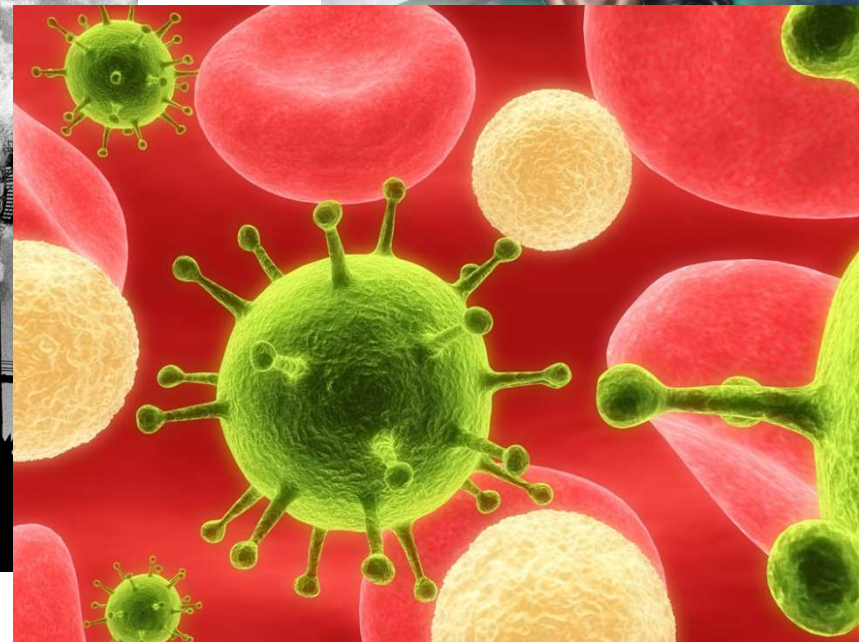


Prípád
(áno/nie)

- Prevalencia, incidencia
- Relatívne riziko
- Pomer šancí

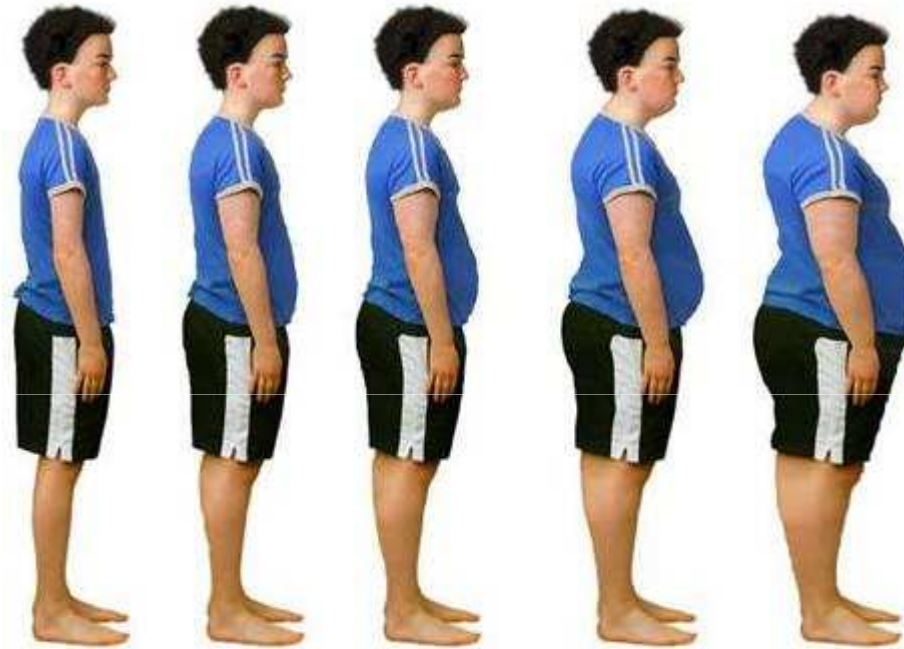


Kto bol exponovaný?



Podľa toho ako sme definovali expozíciu.

Kto je „prípád“ a kto „neprípád“?



Podľa toho ako sme definovali indikátor.

Prevalencia, incidencia

Prevalencia

= počet existujúcich prípadov/ populácia v danom čase

Incidencia

= počet nových prípadov/ populácia v ohrození

(ktorým sa to mohlo stať – napr. Japonci nežijúci v Japonsku nemohli byť ohrození ožiarením z atómovej bomby)

„Prípad“ / „Nepřípad“

Kto je „prípad“ a kto „nepřípad“?

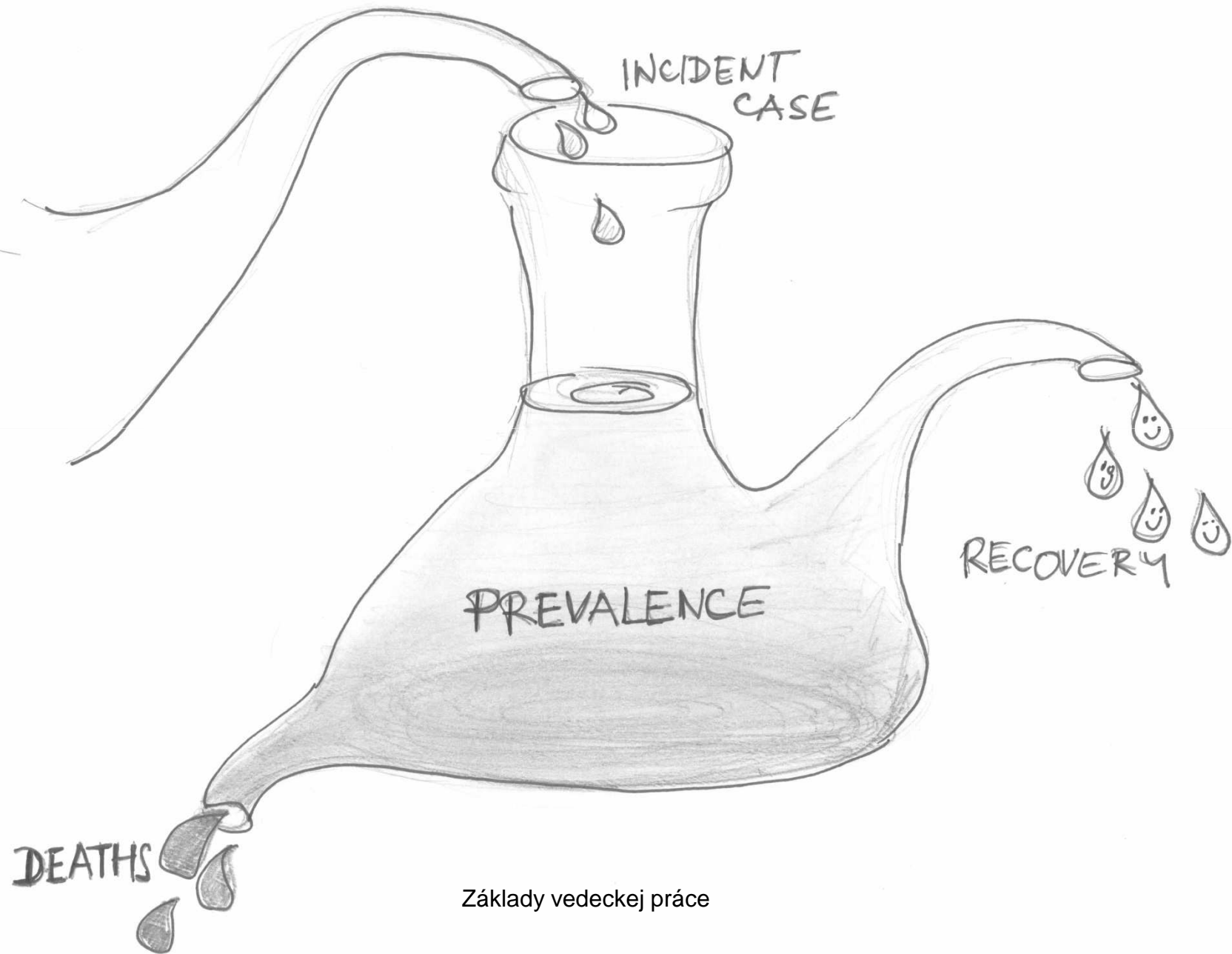
Podľa toho ako sme definovali indikátor. Napr. BMI nad 30kg/m², úmrtie, dg. depresia, atď.



V Košiciach majú 800 registrovaných bezdomovcov a z toho 4 v zime umrzli, zatiaľ čo v Trebišove majú 50 bezdomovcov a 1 umrzol. **Kde je to horšie?**

Prevalencia = počet existujúcich prípadov/populácia v danom čase, teda v KE je to $4/800 = 5/1000$ a v TV je to $1/50 = 20/1000$, z čoho vyplýva, že v Trebišove sú na tom oveľa horšie. **Určite?**

V KE môžu byť lepšie registrovaní a preto ich je viac (presnejší odhad menovateľa), zatiaľ čo v TV to nefunguje veľmi dobre – takže menovateľ máme menej presne odhadnutý a môže byť oveľa vyšší – len tí umrznutí sú naisto. Aj čitateľ nemusí byť presný, lebo príčina úmrtia môže byť klasifikovaná inak.



Základy vedeckej práce

Prevalencia

	Prípady rakoviny (numerator, čitateľ)	Obyvateľstvo (denominator, menovateľ)	N/D	na 100 000
UK	247667	60 mil	.00413	413
Belgicko	47948	10 mil	.00479	479

Incidencia

Nové prípady rakoviny prsníka v roku 2003, UK (N)	Dospelé ženy (D)	N/D	na 100 000
43000	20 mil	.00215	215

Organizácia dát v krížovej tabuľke

		Prípad			Celkovo
		áno	nie		
Vystavenie	áno	a	b		$a + b$
	nie	c	d		$c + d$
Celkovo		$a + c$	$b + d$		$a + b + c + d$

Mierky porovnania

Mierky frekvencie môžu byť porovnávané dvoma spôsobmi:

→ **odčítané** jedno druhého – **absolútne porovnanie**

→ **delené** jedno druhým – **relatívne porovnanie**

Absolútne porovnanie (odčítanie) – informácie o dopade na verejné zdravie – rozdiel medzi frekvenciou dvoch mierok (vystavení vs nevystavení)

→ rozdiel v **riziku** $a/(a+b) - c/(c+d)$

→ rozdiel v **šanciach** $a/b - c/d$

Relatívne porovnanie (delenie) – informácia o sile vzťahu medzi vystavením a výsledkom – pomer frekvencie dvoch mierok (vystavení vs nevystavení)

→ **relatívne riziko** $(a/(a+b)) / (c/(c+d))$

→ **pomer šancí** $(a/b) / (c/d)$

Relatívne riziko (RR), Pomer šancí (OR)

		prípado			riziko	šance
		áno	nie			
vystavenie	áno	a	b		$a/(a+b)$	a/b
	nie	c	d		$c/(c+d)$	c/d

$$RR = \frac{r_1}{r_2} = \frac{a / (a+b)}{c / (c+d)}$$

$$OR = \frac{O_1}{O_2} = \frac{a / b}{c / d}$$

Relatívne riziko (RR)

		Smrť		riziko	šance
		áno	nie		
Multiorgánové zlyhanie	áno	22	4	$a/(a+b)$	
	nie	5	17	$c/(c+d)$	

$$RR = \frac{r_1}{r_2} = \frac{a / (a+b)}{c / (c+d)}$$

Relatívne riziko (RR)

		Smrť		riziko	šance
		áno	nie		
Multiorgánové zlyhanie	áno	22	4	$22/(22+4)$	
	nie	5	17	$5/(5+17)$	

$$RR = (22/26) / (5/22) = 0.846 / 0.227 = \mathbf{3.72}$$

Viac než 80% (84.6%) pacientov s multiorgánovým zlyhaním zomrelo, kým iba niečo cez 20% (22.7%) pacientov, ktorí nemali multiorgánové zlyhanie zomrelo. Je to teda skoro 4x vyššie riziko úmrtia v jednej skupine oproti inej. Relatívne riziko je 3.72, inými slovami expozícia multiorgánovému zlyhaniu znamená pre pacienta 4x vyššie riziko úmrtia v danej situácii v porovnaní s pacientami, ktorí nebol exponovaný multiorgánovému zlyhaniu.

Pomer šancí (Odds ratio OR)

		Smrť		riziko	šance
		áno	nie		
Multiorgánové zlyhanie	áno	22	4		a/b
	nie	5	17		c/d

$$OR = \frac{O_1}{O_2} = \frac{a/b}{c/d}$$

Pomer šancí (Odds ratio OR)

		Smrť			riziko	šance
		áno	nie			a/b
Multiorgánové zlyhanie	áno	22	4			
	nie	5	17			c/d

Pomer šancí je pohľad na to isté z iného uhla. Kým v skupine s multiorgánovým zlyhaním bola šanca na úmrtie 22 ku 4 alebo 5.5 k 1 v skupine bez multiorgánového zlyhania to bolo naopak, teda 5 ku 17 alebo 1 ku 3.4 ($1/0.294 = 3.4$). Pomer šancí je 18.7, alebo inými slovami tí, ktorí boli exponovaní multiorgánovému zlyhaniu mali skoro 19x vyššiu pravdepodobnosť úmrtia v porovnaní s pacientami, ktorí neboli exponovaný multiorgánovému zlyhaniu.

$$OR = (22/4) / (5/17) = 5.5 / 0.294 = \mathbf{18.7}$$

Interpretácia OR

- OR 1,23: šanca fajčiara na úmrtie na rakovinu je **o 23 % vyššia** ako nefajčiara
- OR 3,5: šanca ... je **3,5-krát vyššia**
- OR 0,85: šanca ... je **o 15 % nižšia**
- OR 0,33: šanca ... je **3-krát nižšia**

Medicína založená na dôkazoch

- RCT
- Kohortová štúdia
- Case-control štúdia
- Prierezová štúdia
- Kazuistika

Ďakujem za pozornosť!