

Odhadování incidence HIV z průřezových dat

Michal Kulich, Arnošt Komárek, Marek Omelka

KPMS MFF UK

Robust 2014 Jetřichovice



- 1 Motivace – Projekt ACCEPT
- 2 Průřezové odhady incidence HIV
- 3 Vyhodnocení algoritmů pro průřezové odhady

Obsah

- 1 Motivace – Projekt ACCEPT
- 2 Průřezové odhady incidence HIV
- 3 Vyhodnocení algoritmů pro průřezové odhady

Projekt ACCEPT

Randomizovaná studie pro vyhodnocení vlivu behaviorální intervence na incidenci HIV v Africe a Thajsku

- *Aktivní intervence*: (CBVCT)
 - snadno dostupné dobrovolné testování na HIV
 - veřejná kampaň na podporu testování
 - společenské aktivity, podpora
- *Standardní intervence* (SVCT)
 - testování na klinikách, v nemocnicích atp.

Plán studie

Skupinově randomizovaná studie: Experimentální jednotky jsou geografické oblasti

- Oblasti byly spárovány podle geografické polohy, velikosti, etnického složení, dostupnosti zdravotnických služeb, atp.
- V každém páru byla jedné oblasti náhodně přiřazena intervence CBVCT, druhé oblasti intervence SVCT.

Rozsah výběru – oblasti

- Chiang Mai (Thajsko): 7 párů
- Kisarawe (Tanzánie): 5 párů
- Mutoko (Zimbabwe): 4 páry
- Vulindlela (Jižní Afrika): 4 páry
- Soweto (Jižní Afrika): 4 páry

Velikost oblastí: 5,000 – 15,000 obyvl. (Soweto 15,000 – 25,000)

Rozsah výběru – jedinci

- **Úvodní šetření:**

náhodný výběr z populace 18-32 let v každé oblasti demografický a behaviorální dotazník

Rozsah výběru: 250 – 500 na oblast; 14,203 celkem

- **Randomizace**

- **Provádění intervence:** 3 roky

- **Pointervenční šetření:**

náhodný výběr z populace 18-32 let v každé oblasti, nezávisle na úvodním šetření

demografický a behaviorální dotazník

krevní vzorek pro vyšetření přítomnosti viru HIV

Rozsah výběru: 500 – 1700 na oblast; 56,683 celkem

Časový průběh studie

- 2000–2003 plánování, grantové žádosti
- 2003 grant udělen
- 2004 přípravné práce
- 2005 úvodní šetření, randomizace
- 2006 zahájení intervencí
- 2009 ukončení intervencí
- 2009–2011 pointervenční šetření
- 2011–2012 čištění dat
- 2012–2013 analýzy
- 2013 publikace

Statistická analýza

Označme odhady incidence HIV v oblastech s aktivní a standardní intervencí i -tého páru: $\tilde{\lambda}_i^1, \tilde{\lambda}_i^0$

Efekt intervence v i -tém páru odhadneme jako logaritmus podílu odhadnutých incidencí, tj.

$$\tilde{\theta}_i = \log \tilde{\lambda}_i^1 - \log \tilde{\lambda}_i^0$$

Statistická analýza

- odhadovaným parametrem je relativní riziko $\varrho_0 \equiv \exp\{\theta_0\}$, kde $\theta_0 \equiv E \tilde{\theta}_i = E \log(\tilde{\lambda}_i^1 / \tilde{\lambda}_i^0)$
- budeme testovat hypotézu $H_0 : \theta_0 = 0$ proti alternativě $H_1 : \theta_0 \neq 0$ a konstruovat 95%-ní interval spolehlivosti pro ϱ_0

Statistická analýza

Odhad logaritmu relativního rizika

$$\hat{\theta}_W = \frac{\sum_{i=1}^n w_i \tilde{\theta}_i}{\sum_{i=1}^n w_i},$$

kde w_i jsou vhodně zvolené váhy.

Testová statistika je [Robust 2012] $T_W = \frac{\hat{\theta}_W}{S_W}$, kde

$$S_W^2 = \sum_{i=1}^n a_i^2 \hat{\theta}_i^2 - \hat{\theta}_W^2 \sum_{i=1}^n a_i^2, \quad a_i = \frac{w_i}{\sum_{i=1}^n w_i}.$$

Téma příspěvku

Jak odhadnout incidenci HIV v dané oblasti?

Obsah

- 1 Motivace – Projekt ACCEPT
- 2 Průřezové odhady incidence HIV
- 3 Vyhodnocení algoritmů pro průřezové odhady

Klasický odhad incidence

Klasický postup:

- ustanov representativní kohortu HIV– jedinců o velikosti N_-
- sleduj je po dobu τ , testuj je na HIV v pravidelných intervalech
- nasčítej počet N_I nově nakažených jedinců
- odhadni incidenci

$$\hat{\lambda}_C = \frac{N_I}{\tau(N_- - N_I/2)}$$

- ve jmenovateli je odhad celkové doby sledování všech jedinců z kohorty

Klasický odhad incidence

Nevýhody:

- Je těžké získat reprezentativní kohortu
- Sledování trvá delší dobu a je drahé
- Vyžaduje opakované testování všech jedinců \Rightarrow nelze použít v této studii

Průřezový odhad incidence

Alternativa: Odhad z průřezových dat

- získkej reprezentativní náhodný výběr z populace
- proved' jeden HIV test na každém jedinci
- u nakažených jedinců proved' dodatečná laboratorní měření, jejichž výsledky korelují s délkou infekce
- odhadni počet N_R „nedávno“ nakažených jedinců
- odhadni incidenci

$$\hat{\lambda}_C = \frac{N_R}{\tau_R(N_- + N_R/2)}$$

- τ_R (*časové okno*) odhaduje délku sledování jedince; ve jmenovateli je odhad celkové doby sledování

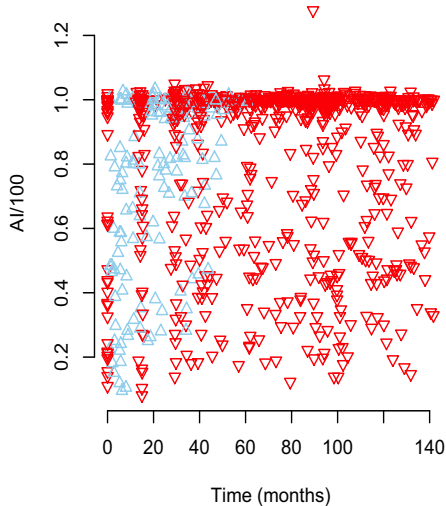
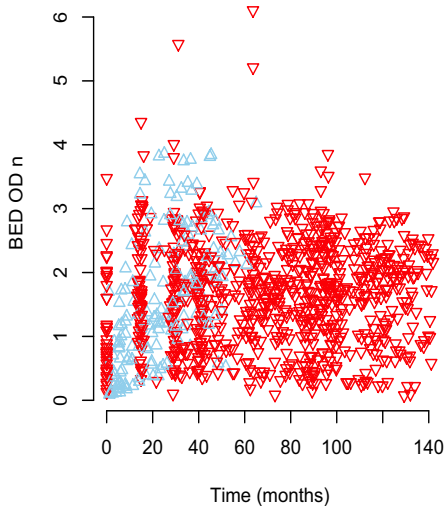
Laboratorní měření pro detekci nedávno nakažených infekcí

- Provádějí se na HIV+ vzorcích
- Měří imunitní odezvu vůči viru; hladina většinou roste s dobou od nakažení
- Příklady: BED, avidity index, ...
- Detekce nedávno nakažených: stanov hraniční hodnotu pro BED/AI, nižší hodnoty považuj za nedávno nakažené

Laboratorní měření pro detekci nedávno nakažených infekcí

- BED** koncentrace protilátek IGg, hodnoty na škále 0–6, hraniční hodnota 0.8
- AI** měří sílu vazby protilátek na antigeny, hodnoty na škále 0–1.2, hraniční hodnota 0.4
- CD4** měří koncentraci CD4 lymfocytů, hodnoty > 0
- VL** měří koncentraci virové RNA, hodnoty > 0

Závislost BED a AI na době od nakažení



Problémy průřezového odhadu incidence

- odhad není konsistentní
- vychýlení závisí na populaci (zejm. na rozdělení dob od nakažení v populaci)
- vychýlení může být kladné nebo záporné a nelze jej odhadnout, predikovat ani odstranit

Řešení:

- uvažuj procedury spojující více laboratorních testů (multi-assay algorithm, MAA)
- uvažuj různé hraniční hodnoty pro detekci nedávno nakažených infekcí
- vyber algoritmus, který minimalizuje střední čtvercovou chybu

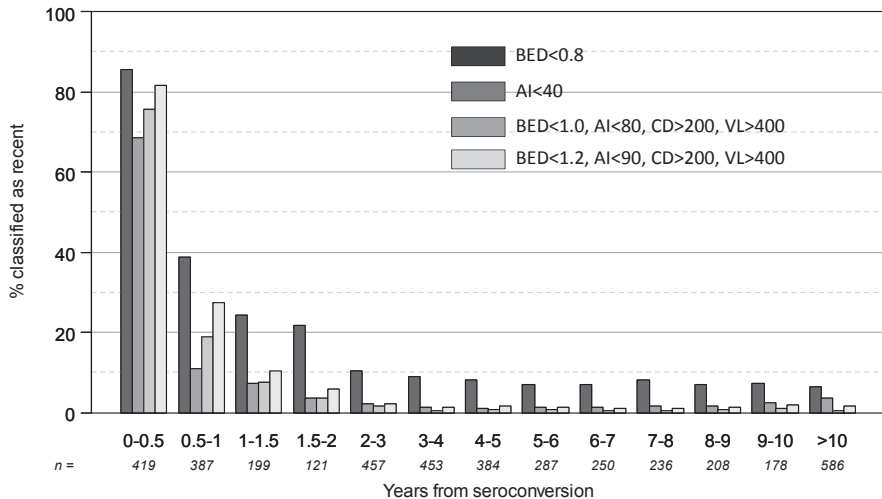
Obsah

- 1 Motivace – Projekt ACCEPT
- 2 Průřezové odhady incidence HIV
- 3 Vyhodnocení algoritmů pro průřezové odhady

Strategie

- Algoritmy jsme vyhodnocovali simulačně
- Používali jsme **validační data** obsahující hodnoty BED, AI, CD4 a VL pro $> 4,000$ vzorků HIV+ jedinců s přibližně známou dobou od nakažení
- Vyhodnotili jsme **klasifikační pravděpodobnosti** v závislosti na době od nakažení
- Spočítali jsme **MSE odhadu incidence** pro 403 algoritmů pro 3 různá rozdělení doby od nakažení
- Odhadli jsme **sílu** pro test efektu intervence v Projektu ACCEPT pro 8 vybraných algoritmů

Klasifikace – výsledky



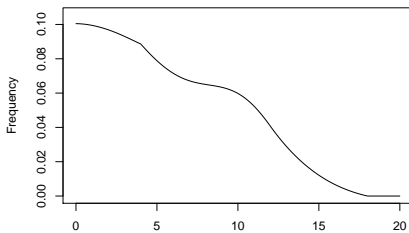
Porovnání algoritmů

Uvažuj 403 algoritmy s různými kombinacemi hraničních hodnot

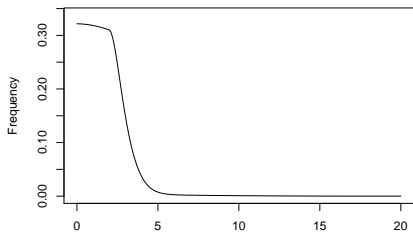
- Proveď **simulaci**: náhodně vyber 200 vzorků z validačních dat, spočítej **odhad incidence**, odhadni **vychýlení**, **rozptyl**, **střední čtvercovou chybu**.
- **Seřaď algoritmy** podle velikosti střední čtvercové chyby. Ta ale závisí na rozdělení dob od nakažení v populaci.
- Opakuj simulaci pro **tři různé scénáře** specifikující různá rozdělení dob od nakažení. Zajisti, aby v simulovaných datech rozdělení dob od nakažení odpovídalo zvolenému scénáři.

Scénáře

Stable epidemic

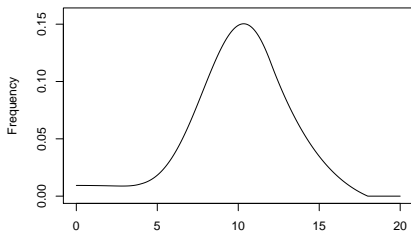


New epidemic



Time since infection [years]

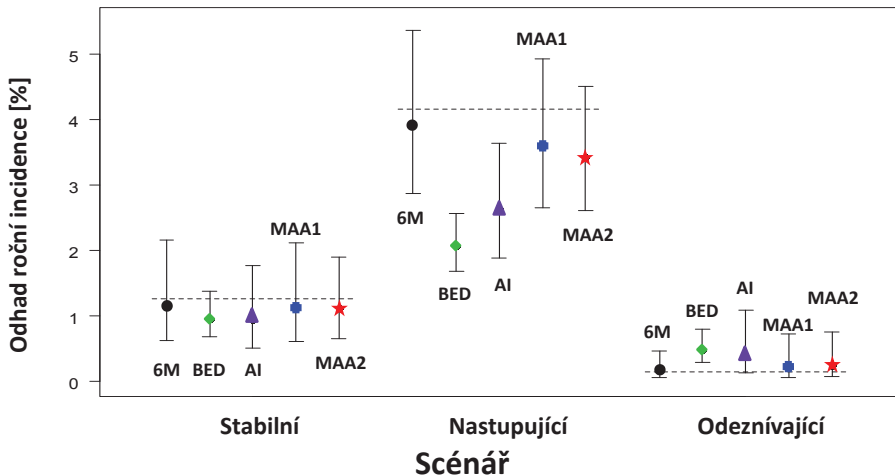
Waning epidemic



Time since infection [years]

Porovnání algoritmů – výsledky

- 6-měs. sledování
- ◆ BED<0.8
- ▲ AI<40
- BED<1.0, AI<80, CD>200, AI>400
- ★ BED<1.2, AI<90, CD>200, AI>400



Porovnání algoritmů – výsledky

	Scénář								
	Stabilní			Nastupující			Odeznívající		
	Poř.	Vych.	RM	Poř.	Vych.	RM	Poř.	Vych.	RM
12 měsíců		0.0%	0.21		0.0%	0.09		0.0%	0.56
6 měsíců		-7.9%	0.32		-5.7%	0.17		18.6%	0.51
BED<0.8	95	-23.2%	0.32	396	-49.9%	0.70	396	221.4%	1.19
AI<40	339	-24.5%	0.42	390	-37.0%	0.49	390	149.7%	1.06
MAA1	125	-9.9%	0.33	7	-13.1%	0.21	20	49.0%	0.71
MAA2	23	-11.4%	0.29	91	-17.6%	0.24	78	64.2%	0.75

RM: odmocnina ze střední čtvercové chyby odhadu log incidence

Poř.: pořadí podle RM mezi 403 algoritmy

Vych.: relativní vychýlení odhadu incidence

6, 12 měsíců: odhad incidence pomocí sledování o příslušné délce

MAA1: BED<1.0, AI<80, CD4>200, VL>400

MAA2: BED<1.2, AI<90, CD4>200, VL>400

Výběr algoritmu – 1. kolo

- Algoritmy byly **seřazeny** podle střední čtvercové chyby pro odhad logaritmu incidence pod daným scénářem
- Každému algoritmu bylo přiřazeno **celkové skóre** rovnající se součtu pořadí pro tři scénáře
- **Šest algoritmů** s nejlepším celkovým skóre bylo vybráno pro vyhodnocení síly testu na relativní riziko (podíl incidencí)

Výběr algoritmu – 2. kolo

Máme 6 uvažovaných algoritmů

- Proved' simulační studii podle situace v Projektu ACCEPT: 17 párů oblastí, v každé náhodný výběr jedinců. Velikosti výběrů a počty nakažených odpovídaly skutečnosti.
- Simuluj data výběrem vzorků z validační databáze podle stabilního scénáře, uměle vytvoř efekt intervence (redukce incidence o cca 35% v posledních 2 letech v oblastech s aktivní intervencí).
- Spočítej odhady incidence v každé oblasti a odhady efektu intervence v každém páru. Spočítej celkový efekt intervence, proved' test, spočítej interval spolehlivosti.
- Vyber algoritmus, který poskytuje nejlepší sílu testu.

Výběr algoritmu – výsledky

Algoritmus	Odh. Efekt	Empirická směr. odch.	Odhadnutá směr. odch.	Síla testu	Pokrytí 95% CI
6 měsíců	0.656	0.109	0.104	93.4	92.0
24 měsíců	0.651	0.066	0.071	100.0	91.0
BED<0.8	0.774	0.068	0.071	90.7	43.6
AI<40	0.720	0.101	0.098	83.6	82.9
MAA1	0.674	0.102	0.100	93.7	91.2
MAA2	0.684	0.094	0.092	95.4	89.8

6, 24 měsíců: odhad incidence pomocí sledování o příslušné délce

MAA1: BED<1.0, AI<80, CD4>200, VL>400

MAA2: BED<1.2, AI<90, CD4>200, VL>400

Výběr algoritmu – výsledky

Vybrali jsme algoritmus

BED < 1.2, AI < 90, CD4 > 200, VL > 400

Výsledky Projektu ACCEPT

<i>Efekt intervence (RR)</i>	0.861
<i>95% interval spolehlivosti pro RR</i>	0.725 – 1.023
<i>Vážená t-statistika</i>	–1.945 na 9.44 stupních volnosti
<i>P-hodnota</i>	0.0822

Země	Preval. [%]		Nedáv. ^a		Inc. [%]		RR ^b	Váha ^c
	C	S	C	S	C	S		
Zimbabwe	13.1	12.7	25	42	0.68	1.13	0.63	13.9
Tanzánie	6.1	5.8	27	20	0.86	0.68	1.23	10.1
Vulindlela	31.3	30.3	110	120	3.76	4.03	0.93	53.8
Soweto	12.3	15.9	52	68	1.18	1.63	0.74	22.2

^a Počet nedávných infekcí detekovaných algoritmem MAA2

^b Relativní riziko infekce (CBVCT vs. SVCT)

^c Váha při analýze

To je dnes všechno.