

Szent István Egyetem
Állatorvos-tudományi Doktori Iskola

**Epidemiológiai és ökológiai mutatók
statisztikai és valószínűségelméleti
modellezése,
térbeli szerkezetének vizsgálata**

PhD értekezés tézisei

Lang Zsolt

2014

Témavezető és témabizottsági tagok:

.....
Prof. Dr. Reiczigel Jenő
Szent István Egyetem,
Biomatematikai és Számítástechnikai Tanszék
témavezető

Prof. Dr. Rózsa Lajos
MTA-MTM Állatökológiai Kutatócsoport
témabizottság tagja

Dr. Harnos Andrea
Szent István Egyetem,
Biomatematikai és Számítástechnikai Tanszék
témabizottság tagja

Dr. Franz Rubel
Vet. Med. Univ. Wien,
Dept. für Naturwiss.
témabizottság tagja

Készült 8 példányban. Ez a n. sz. példány.

.....
Lang Zsolt

Tartalomjegyzék

1	Bevezetés.....	4
2	Konfidenciaintervallum szerkesztése valódi prevalenciára független mintákból becsült szenzitivitás és specifikusság alapján.....	6
2.1	Módszerek	6
2.2	Eredmények.....	9
2.2.1	Alkalmazás: <i>Trichomonas gallinae</i> prevalenciája házi pirókban (<i>Carpodacus mexicanus</i>)	11
2.3	Megvitatás	12
3	Szeropozitivitás és prevalencia összefüggései hazai BHV-1 mentesítés adatainak járványtani elemzése alapján	14
3.1	Anyag és módszer	15
3.1.1	Statisztikai módszerek.....	15
3.2	Eredmények.....	18
3.3	Megvitatás	20
4	Zsúfoltság és diverzitás	22
5	Egyedek és csoportjaik valószínűségelméleti modellezése	26
6	Új tudományos eredmények	27
7	A doktori kutatás eredményeinek közlései	39
8	Köszönetnyilvánítás.....	40

1 Bevezetés

Az epidemiológiában központi szerepet játszik egy betegség, kockázati tényező, demográfiai jellemző, stb. populációbeli részaránya, prevalenciája. Annak pontos megállapítása, diagnosztizálása, hogy egy kiválasztott egyed rendelkezik-e a vizsgált tulajdonsággal, sokszor nem lehetséges. A diagnosztikai eljárások adhatnak téves pozitív és téves negatív eredményt. Célkitűzéseim közül az első olyan új statisztikai eljárás kidolgozása, amivel megbízható intervallumbecslés adható a valódi prevalenciára olyankor, ha a diagnosztikai eljárás érzékenységet (a pozitív diagnózis részaránya a valóban pozitívak körében) és fajlagosságát (a negatív diagnózis részaránya a valóban negatívak körében) előzetes kísérletekkel megbecsülték. A célok sorában a második az érzékenységet (szenszitivitás) és fajlagosságot (specifikusság) ismertnek tekintő, korábban közölt egzakt statisztikai módszerek alkalmazása a magyarországi szarvasmarha-állományok BHV-1 mentesítésének statisztikai értékelésére, a szeropozitivitás aránya helyett a téves pozitív és téves negatív eseteket kiküszöbölő valódi prevalencia felhasználásával.

A prevalencia térbeli és időbeli változására hatással lehet a populáció egyedeinek aggregált eloszlása. A populáció egyedei sokszor csoportokba rendeződnek. Ilyen csoportok lehetnek pl. állatállományok, területi, térbeli társulások, paraziták gazdánkénti mikroközösségei. Az aggregáltságot, zsúfoltságot a csoportok egyedszámával jellemezhetjük. Észrevettem, hogy a zsúfoltság mérőszámai formai hasonlóságot mutatnak a biológiai diverzitásra általánosan használt indexekkel. Harmadik célom ennek az analógiának a feltárása, a két terület elméleti eredményeinek ötvözése, az így nyert új módszerek alkalmazása. Kimutattam, hogy a zsúfoltság és biológiai diverzitás nem csak fogalmi szinten, hanem közvetlenül is kapcsolatba hozható, egyszerre modellezhető, ha a csoportok több faj egyedeiből állnak. Az elméleti eredményeket tovább általánosítva valószínűségelméleti modellt vezettem be csoportokba szerveződő egyedek jellemzőinek egyedszemponitú és csoportszintű eloszlásaira, regresszió típusú modellezésére. Az elméleti matematikai eredményeket vetési és dolmányos varjak parazitáltságának összehasonlítására, ill. gímszarvasokat parazitáló szarvaskullancslegyek ivararányának regressziós modellezésére alkalmaztam.

2 Konfidenciaintervallum szerkesztése valódi prevalenciára független mintákból becsült szenzitivitás és specifikusság alapján

Betegségek prevalenciáját általában olyan diagnosztikai tesztekkel állapítják meg, amelyek téves eredményt is adhatnak. Rogan és Gladen (1978) módszert közölt a valódi prevalencia becslésére a diagnosztikai eljárás szenzitivitásának és specifikusságának figyelembe vételével. Reiczigel et al. (2010) egzakt konfidenciaintervallumot konstruált a valódi prevalenciára a pontos szenzitivitás és specifikusság ismeretét feltételezve. Új módszert ismertetünk a valódi prevalencia közelítő konfidenciaintervallumának megszerkesztésére abban az esetben, amikor a szenzitivitást és a specifikusságot független mintákból becsüljük (Lang és Reiczigel 2014).

2.1 Módszerek

Feltételezzük, hogy három egymástól független minta áll rendelkezésre:

- az első olyan egyedekből áll, amelyekről ismeretes, hogy a vizsgált betegségben szenvednek (ebből becsüljük meg a diagnosztikai teszt szenzitivitását),
- a másodikba olyan egyedek tartoznak, amelyek a betegségben nem szenvednek (ebből becsüljük a teszt specifikusságát),
- a harmadik egy véletlen minta a célpopulációból, melynek egyedeit a diagnosztikai teszttel vizsgáljuk (az egyedek valódi betegségstátusza nem ismert, csak a teszt eredményét – pozitív vagy negatív – ismerjük).

Rogan és Gladen (1978) e három mintából pontbecslést adott a valódi prevalenciára, emellett közelítő formulát közölt a becslés varianciájára. Ezekből Wald közismert módszerével konfidenciaintervallum készíthető a valódi prevalenciára. Rogan és Gladen módszerét Agresti és Coull (1998) eljárásához hasonlóan úgy módosítottuk, hogy a szenzitivitás és specifikusság mintáiban a helyesen és tévesen diagnosztizált esetek számát 1-gyel, a célpopulációból származó mintában a pozitív és negatív esetek számát $z_{crit}^2/2$ -vel növeltük, ahol z_{crit} a standard normális eloszlás előírt megbízhatósági szinthez tartozó kritikus értéke. Ha a megbízhatósági

szint 95%, akkor $z_{crit} = 1.96$. A valódi prevalencia ilyen módon adjusztált pontbecsléséből és varianciabecsléséből új típusú konfidenciaintervallumot szerkesztettünk (Lang és Reiczigel 2014).

Számítógépes szimulációval kiértékeltek a korrigált és az eredeti Wald-Rogan-Gladen konfidenciaintervallum lefedési valószínűségeit az alábbi scenáriókra.

- A valódi szenzitivitás és specifikusság vizsgált értékei 1, .99, .95, .90, .70.
- A szenzitivitás, specifikusság és prevalencia becslésére szolgáló minták nagysága 30, 100, 300, 1000, 3000.
- Az előírt, nominális megbízhatósági szint 90%, 95% és 99%.
- A lefedést a valódi prevalencia .005, .01, .02, .03, .05, .10, .20, .30, .50 értékeire állapítottuk meg.

A fenti beállítások minden kombinációjára 20000 véletlen mintát (valójában 20000 mintahármas) generáltunk és meghatároztuk belőlük a lefedési valószínűségeket.

Megvizsgáltuk a pontosság romlását mintákból becsült szenzitivitás és specifikusság mellett olyan konfidenciaintervallum szerkesztési eljárásoknál, amelyek ismert szenzitivitást és specifikusságot feltételeznek. Szimulációs kísérleteket végeztünk Blaker és Clopper-Pearson módszerével szerkesztett konfidenciaintervallumokkal (Reiczigel et al. 2010, Blaker 2000, Clopper és Pearson 1934). A szimulációk során a szenzitivitás és specifikusság becslésére szolgáló minták nagysága a prevalencia becsléséhez használt minta nagyságával megegyező, ill. kétszer, háromszor, ötször és tízszer nagyobb volt. A lefedési valószínűségeket 10000 véletlen mintahármas alapján határoztuk meg, különböző mintanagyságok, nominális megbízhatósági szintek, továbbá a valódi szenzitivitás és specifikusság .99-es, .95-ös és .90-es értékei mellett.

2.2 Eredmények

A normális eloszlással történő közelítésre támaszkodó és a korrigálatlan Wald-Rogan-Gladen varianciabecslésből számolt konfidenciaintervallum lefedési valószínűsége túlságosan alacsony, számos esetben nominális 95% mellett kevesebb, mint 90%,

nominális 90% mellett kevesebb, mint 80%, különösen olyankor, ha a szenzitivitás és a specifikusság értéke nagy, 1 közeli. Az általunk javasolt módszer alkalmazása lényegesen javította a lefedési valószínűséget, ami a korrekció után minden vizsgált paraméterkombinációra, nominális megbízhatósági szintre elfogadhatóvá vált. A 95%-os nominális szint mellett a lefedés minimuma ritkán került 94% alá és mindig 93% fölött maradt. A 99%-os nominális szint mellett a lefedési valószínűség soha nem volt kisebb 98%-nál, és 90%-os nominális szint esetében nem lett kisebb 88%-nál.

Ismert szenzitivitást és specifikusságot feltételező konfidenciaintervallum konstrukciókra a pontosság romlása jelentős, ha a szenzitivitást és a specifikusságot mintákból becsüljük. Ez csak akkor válik elhanyagolhatóvá, ha a szenzitivitás és specifikusság becslésére kiválasztott minták nagysága ötször-tízszer nagyobb, mint a prevalencia becslésére szolgáló minta.

2.2.1 Alkalmazás: *Trichomonas gallinae* prevalenciája házi pirókban (*Carpodacus mexicanus*)

Anderson et al. (2009) trichomonas protozoonok előfordulását tanulmányozták vadon élő énekesmadarak körében. Példaként vizsgáljuk meg a *Trichomonas gallinae* prevalenciáját házi pirókban (*Carpodacus mexicanus*). A mintába került 2971 madáregyed közül 51 esetben mutatták ki a parazitát, ezért a pozitív esetek aránya .017. A diagnosztikai eljárás szenzitivitását és specifikusságát a paraziták kitenyésztésével határozták meg. A szenzitivásra .97 (32/33), a specifikusságra 1 (20/20) adódott eredményül. Ezekkel számolva, a valódi prevalencia pontbecslése .018.

Elfogadva, hogy .97 és 1 a szenzitivitás és specifikusság ismert pontos értékei, a Reiczigel et al. (2010) módszerével számított Blaker konfidenciaintervallum .013 – .023. Figyelembe véve azt is, hogy a szenzitivitást és specifikusságot nagyon kis, 33 és 20 elemszámú mintából becsülték, a valódi prevalencia 95%-os konfidenciaintervalluma 0 – .053. Ez a CI sokkal szélesebb, mint az ismert szenzitivitás és

specifikusság feltételezése mellett kapott konfidenciaintervallum.

2.3 Megvitatás

Vizsgálataink során azt tapasztaltuk, hogy a szenzitivitás és a specifikusság becsléséből származó bizonytalanság lényeges hatással lehet a prevalencia becslésének pontosságára. Amint azt vártuk, ha a szenzitivitás és a specifikusság becslésére szolgáló minták nagysága sokkal (5-10-szer) nagyobb, mint a prevalencia becslésére vett minta, akkor a CI tényleges lefedése a nominális szintet közelíti ($>.93$, ha a nominális $.95$). Ha viszont a szenzitivitás és specifikusság mintáinak nagysága hasonló a prevalencia mintájának nagyságához (legfeljebb 3-szor nagyobb), akkor a CI tényleges lefedési valószínűsége túlzott mértékben lecsökkenhet a nominális szinthez képest ($<.90$ is lehet nominális $.95$ mellett).

Az általunk javasolt CI jó lefedési valószínűséget biztosít tág határok között mozgó prevalencia, szenzitivitás és specifikusság értékek mellett, beleértve azt is, amikor a prevalencia közel van 0-hoz vagy 1-hez és a szenzitivitás, specifikusság értéke 1-hez közeli. Az

adjusztálás miatt a CI kissé konzervatív, amikor a prevalencia 0 vagy 1 közelében van. A lefedés pontossága akkor is megmarad, ha a három minta nagysága jelentősen különbözik.

Széles körben elterjedt vélekedés, hogy a populációs arány $\bar{p} \pm 1.96 \cdot \text{var}(\bar{p})^{\frac{1}{2}}$ formulával értelmezett 95%-os Wald-féle konfidenciaintervalluma elfogadható pontosságú, feltéve, hogy a $\bar{p} \pm 3 \cdot \text{var}(\bar{p})^{\frac{1}{2}}$ próbaintervallum nem tartalmazza sem a 0-t, sem az 1-et, továbbá a becsléshez felhasznált minta nagysága „elég nagy”. Greiner és Gardner (2000a) javasolják ennek alkalmazását a Wald-Rogan-Gladen konfidenciaintervallumra. Számítógépes szimulációval vizsgáltuk ezt a konstrukciót és azt tapasztaltuk, hogy ez az eljárás még az eredeti, próbaintervallumot nem alkalmazó Wald-Rogan-Gladen intervallumnál is pontatlanabb.

3 Szeropozitivitás és prevalencia összefüggései hazai BHV-1 mentesítés adatainak járványtani elemzése alapján

A magyarországi szarvasmarha-állományok BHV-1 mentesítését gE-negatív markervakcinákkal végzik 2002 óta. A szeropozitivitásban 2006-ig bekövetkezett telepi változások alapján az adatok újrafeldolgozásával, a BHV-1 ellenanyagok kimutatása során alkalmazott diagnosztikai tesztek szenzitivitását és specifikusságát felhasználva, a téves pozitív és téves negatív arányok figyelembe vételével a szeropozitivitásból megbecsültük a BHV-1 fertőzöttség valódi előfordulási arányát, prevalenciáját. A becslést Rogan és Gladen (1978) módszerével végeztük el. A valódi prevalencia ismerete lehetővé tette a mentesítés hatásának pontosabb értékelését, a fertőzés korcsoportok, megyék, megyecsoportok és telepek közötti heterogenitásának és terjedési lehetőségének statisztikai vizsgálatát.

A fertőzés telepek közötti terjedésének jellemzésére az előbbieket mellett egy ökológiai mutató, a zsúfoltsági index alkalmazását is javasoljuk. A bemutatott

statisztikai módszerek más fertőző betegségek elleni mentesítések előrehaladásának nyomon követésére is alkalmasak, és segítségükkel a mentesítési programokban esetlegesen szükséges módosítások idejében meghozhatók.

3.1 Anyag és módszer

A 2002–2006 közötti vizsgálatok során a minták összesen 155 telepről származtak, de a felmérő és ellenőrző vizsgálatokban csak részben ugyanazok a telepek szerepeltek (Pálfi et al. 2007). Az adatok megismételt elemzéséhez azokat a teszteredményeket használtuk, ahol a teszt típusa, szenzitivitása, specifikussága és a kórisme minden mintára megállapítható.

3.1.1 Statisztikai módszerek

A valódi prevalenciát Rogan és Gladen (1978) módszerével becsültük. A valódi prevalencia 95%-os kétoldali megbízhatósági tartományát a Blaker-féle egzakt próbából Reiczigel módszerével számítottuk ki (Blaker 2000, Reiczigel et al. 2010).

A prevalenciákat korcsoportonként, megyénként, megyecsoportonként és telepenként is meghatároztuk. A mentesítés előtti és utáni prevalenciák szignifikáns eltérését korcsoportonként egyoldali 97,5%-os megbízhatóságú Clopper-Pearson CI-vel teszteltük (Clopper és Pearson 1934, Reiczigel et al. 2010).

Korcsoportonként megvizsgáltuk, mekkora szórással ingadozik a prevalencia megyék között és megyéken belül. Ehhez a korcsoportos prevalencia teljes varianciáját megyék közötti és megyén belüli varianciák összegére bontottuk, a szórássok az átlagos varianciák négyzetgyökei (Reiczigel, Harnos, Solymosi 2007). A szórássok becslésénél figyelembe kell venni, hogy a telepek prevalenciái a kis mintaszámok miatt pontatlanul becsülhetők. Az ebből a pontatlanságból adódó torzítás azonban korigálható (Lang et al. 2013, www.univet.hu/users/zslang/korrekciok.pdf).

Meghatároztuk az átlagos prevalenciát és a prevalencia csoporton belüli szórást három földrajzilag összefüggő, állománynagyság szerint többé-kevésbé homogén megyecsoportra (a Dunántúl Komárom-Esztergom megye nélkül, az Alföld, valamint az északi megyék:

Komárom-Esztergom, Pest, Nógrád, Heves, Borsod-Abaúj-Zemplén).

Az ismert létszámú 115 állományban megvizsgáltuk a létszám és a prevalencia kapcsolatát. Nagyobb létszámú telepen több az állatok közötti kontaktus, ezért a fertőzés elvileg gyorsabban terjedhet, és nagyobb szintet érhet el, tehát a prevalencia nagyobb lehet. E hipotézis vizsgálatához a telepeket létszám szerint növekvő sorrendbe rendeztük, majd öt, egyenként 23–23 telepből álló csoportba, kvintilisbe osztottuk. Mindegyikhez kiszámítottuk felméréskor és ellenőrzéskor a BHV-1 prevalenciát. A statisztikai összehasonlításhoz szimultán CI-eket szerkesztettünk a prevalenciákra Bonferroni módszerével.

Az egyedszám telepek közötti nagy szórása miatt az átlagos létszám nem mutatja az állatok között lehetséges kapcsolatokat, érintkezések számát, ami a fertőzés elterjedésének, maximális mértékének, a prevalenciának jellemzője. Az átlagos érintkezésszámot ezért a zsúfoltsági indexszel mértük (Lloyd 1967, Reiczigel et al. 2005, 2008).

3.2 Eredmények

A valódi prevalencia minden korcsoportban és összességében is kisebb volt a mentesítés utáni ellenőrzéskor a mentesítést megelőző felméréshez képest. A valódi prevalenciák felméréskor nagyobbak, ellenőrzéskor pedig kisebbek, mint a szeropozitivitások. A vizsgált periódusban (2002 és 2006 között) a mentesítés valójában eredményesebb volt, előbbre jutott, mint azt a szeropozitivitási arányok mutatják.

A megyéket külön-külön vizsgálva nagy eltérések tapasztalhatók mind a felméréskori, mind pedig a mentesítés utáni prevalenciák vonatkozásában. A legtöbb megyében jelentős a csökkenés. A prevalenciák megyék közötti és megyéken belüli ingadozását vizsgálva megállapítható, hogy a telepek között, megyétől függetlenül, nagy a heterogenitás mind a felmérő, mind az ellenőrző vizsgálatok során. A földrajzilag összefüggő homogén megyecsoportokba sorolt megyék esetében felméréskor nincs lényeges különbség a csoportok átlaga és szórása között. Mentésítés után a Dunántúlon és az északi megyékben valamivel jobban sikerült csökkenteni a fertőzés

prevalenciáját, mint az Alföldön. A heterogenitás is csökkent, leginkább az északi megyékben.

Az állománylétszám és a prevalencia kapcsolatát létszám szerint kvintilisekbe sorolt telep csoportokkal vizsgáltuk meg azon a 115 telepen, ahol ismertük a létszámot. Azt tapasztaltuk, hogy felméréskor a nagyobb létszámú telepek csoportjához növekvő tendenciájú BHV-1 prevalencia tartozik. Szimultán konfidencia intervallumok alkalmazásával kimutattuk, hogy a kb. 450 egyedet meghaladó méretű telepeken nagyobb a prevalencia, mint a kisebb egyedszámú telepeken. A mentesítés utáni ellenőrzésnél fordított a trend, a nagy létszámú telepeken kisebb a prevalencia. Korcsoportonként vizsgálva, felméréskor a borjak és tehének esetében figyelhető meg a telep állatszama és a prevalencia pozitív kapcsolata. A növendék és vemhes üszőknél nem vehető észre kapcsolat. Ellenőrzéskor nincs korcsoportonként kimutatható kapcsolat a telepek létszáma és a prevalencia között.

A zsúfoltsági index felméréskor 867, ellenőrzéskor 817. Az átlagosan lehetséges kontaktusszám 50–80%-kal nagyobb, mint a telepek átlagos létszáma, a nagyobb

állatszámú telepeken található ugyanis az egyedek nagyobb hányada.

3.3 Megvitatás

A valódi prevalenciára felméréskor nagyobb, mentesítéskor kisebb értéket kaptunk, mint a szeropozitivitás aránya. Ennek a paradox jelenségnek elsősorban az a magyarázata, hogy a mentesítés jelentősen csökkenti a szeropozitivitást, és a Rogan-Gladen formulából kitűnik, hogy nagy szeropozitivitásnál a valódi prevalencia még nagyobb, kis szeropozitivitásnál még kisebb lesz.

Milyen diagnosztikai tesztet érdemes választani? Mentés előtt, amikor a fertőzés aránya feltehetően nagy, elsősorban a téves negatív minták ronthatják el a valódi prevalencia becslésének pontosságát. Ezért mentés előtt fontos, hogy a választott teszt szenzitivitása minél jobb legyen. Sikeres mentés után, kis prevalencia mellett a téves pozitív minták növelhetik a prevalencia becslésének standard hibáját. Mentés után ezért minél jobb specifikusságú tesztet érdemes választani, a szenzitivitásnak ekkor kisebb a jelentősége.

A nagy zsúfoltsági index a nem egyformán hatékony mentesítés miatt heterogén prevalenciával együtt visszafertőzést okozhat. Magyarországon a vizsgált időszakban a zsúfoltsági index nagy, emellett a mentesítés sikere megyéken, megyecsoportokon belül nagyon heterogén, főleg a kisebb létszámú telepeken gyengébb. Ezért számottevő annak a kockázata, hogy egy-egy nagyobb telep mentesített állományát a közelében lévő kisebb telepek mentesítetlen állománya visszafertőzi.

Mentesítés előtt a nagyobb egyedszámú telepeken nagyobb a BHV-1 prevalencia, ezért a nagy zsúfoltsági index, kontaktusszám a BHV-1 fertőzés prevalenciáját mentesítés nélkül növelheti.

A bemutatott módszereket a leghatékonyabban országos vagy legalábbis több telepet felölelő vizsgálati program keretében lehet alkalmazni. Ilyenkor lehetőség van pontos becslésekre, korcsoportonkénti, területi, időbeli eltérések összehasonlítására.

4 Zsúfoltság és diverzitás

Egyedek csoportokban élő közösségein, például gazdák parazitái körében a zsúfoltságot, az egyedszempon-tú csoportméretet a csoportok monoton nö-vő függvénnyel skálázott egyedszámának egyedekre vett átlagával értelmezzük (Lloyd 1967, Reiczigel et al. 2005, 2008). Szoros fogalmi kapcsolatot, analógiát tártam fel a zsúfoltság és a fajok biológiai diverzitása között. A biológiai diverzitás (D) Patil és Taille (1982) szerint a fajok skálázott ritkaságának egyedekre vett átlaga. Egy faj ritkaságát olyan $R(p)$ függvénnyel mérik, ami a faj közösségen belüli p részarányának monoton csökkenő függvénye és $R(1) = 0$. A skálázó függvény megadásával a zsúfoltság és a diverzitás egymással kifejezhetők. Rögzített c konstans mellett a $c - R(p)$ függvény a faj egyedszámának monoton nö-vő függvénye és az ezzel kiszámolt zsúfoltság éppen $c - D$. Az egymásnak megfelelőetett legfontosabb zsúfoltsági és diverzitási indexek lineáris skálafüggvényre a Lloyd- és Reiczigel-féle lineáris zsúfoltság és a Simpson-féle diverzitás, logaritmus függvényre a logaritmusos zsúfoltság és a Shannon-féle diverzitási index,

hiperbolikus függvényre a hiperbolikus zsúfoltság és a fajgazdagság (a fajok száma mínusz 1).

A diverzitások elméletében értelmezett effektív fajszám megfelelőjeként a zsúfoltság fogalmkörében tetszőleges monoton növekvő skálafüggvény mellett bevezettem az effektív csoportszám fogalmát. Ez olyan egyenlő egyedszámú csoportok számát jelenti, ahol a csoportok egyedeinek zsúfoltsága a vizsgált közösség zsúfoltságával egyezik meg. Erre a fogalomra építve, a tényleges csoportszám és az effektív csoportszám hányadosaként tetszőleges monoton növekvő skálafüggvényre általánosítottam a Bez-féle aggregáltsági indexet (Bez 2000).

Kimutattam, hogy a zsúfoltság és a diverzitás között közvetlen, nem csak a fogalmi analógia szintjén megnyilvánuló kapcsolat is van olyankor, ha a vizsgált közösség csoportjaiban az egyedek több fajhoz tartoznak. A fajok együttesének zsúfoltsága felbontható a fajonkénti átlagos zsúfoltság és az átlagos csoporton belüli, szubjektív diverzitás összegére. Az additív kapcsolaton túl multiplikatív összefüggés is létezik. Ismeretes, hogy több fajból álló csoportok közösségének

együttes, ún. gamma valódi diverzitása (azaz effektív fajszáma) a csoportokon belüli átlagos valódi diverzitás és a csoportok közötti változatosságot kifejező béta diverzitás szorzata (Whittaker 1960, 1972, Hill 1973, Jost 2006, 2007, Tuomisto 2010a). A zsúfoltság–diverzitás fogalmi megfeleltetésnél a valódi diverzitásnak vagy más néven effektív fajszámnak az effektív csoportszám, ill. az aggregáltsági index felel meg. Bevezetem a teljes közösség aggregáltságára a gamma aggregáltság, a fajonkénti átlagos aggregáltságra az alfa aggregáltság elnevezéseket. A béta aggregáltságot az alfa és gamma aggregáltságok hányadosaként értelmeztem. A béta aggregáltság az aggregáltság fajok közötti változatosságát számszerűsíti.

A zsúfoltságra és a diverzitásra kapott elméleti eredményeket felhasználtam vetési varjak (*Corvus frugilegus*) és dolmányos varjak (*Corvus cornix*) tetűparazitáltságának összehasonlítására. A paraziták zsúfoltsága vetési varjakon nagyobb, emellett a zsúfoltság szubjektív diverzitásos összetevőjének is nagyobb a százalékaránya. A varjófajok közötti eltérést magyarázhatja, hogy a fertőzött dolmányos varjak

körében nagyobb arányú az egyetlen tetűfajjal fertőzöttek, így 0 diverzitásúak aránya (50%), mint a fertőzött vetési varjakon (21%).

A tetvek valódi alfa és gamma diverzitásai a vetési varjak esetében valamivel nagyobbak, mint a dolmányos varjaknál. A béta diverzitások a két varjufajnál nagyjából megegyeznek, értékük 1.5-2.5 között van. A tetűfajok átlagos, azaz alfa aggregáltsága a dolmányos varjak körében nagyobb, mint a vetési varjaknál. Ez azt jelenti, hogy a dolmányos varjak tetűfajai átlagosan aggregáltabban oszlanak meg a gazdaállatok között, mint a vetési varjak tetűfajai. A béta aggregáltságok értéke mindkét varjufajnál nagyjából 1 és 2 közötti, ez alapján a vetési varjak génuszait két, gazdákon kevésbé átfedő csoportra, a *Brueelia*, illetve a *Myrsidea* és *Menacanthus* együttesére bontottam blokk klaszterezéssel és korrespondanciaelemzéssel. A dolmányos varjakon a két domináns génusz, a *Myrsidea* és a *Menacanthus* alkot a gazdákon csak kismértékű átfedést mutató csoportot. Érdekesség, hogy a vetési varjakon – a dolmányos varjakkal ellentétben – a *Myrsidea* és *Menacanthus* génuszok nem különülnek el, gyakran előfordulnak közös gazdán.

5 Egyedek és csoportjaik valószínűségelméleti modellezése

Általános valószínűségelméleti modellt vezettem be csoportokba szerveződő egyedek jellemzőinek egyedszemponitú és csoportszintű eloszlásaira. Formulákat adtam a két szemléletmód közötti alapvető transzformációkra. Több elégséges és szükséges és elégséges feltételt fogalmaztam meg arra, hogy egy általános regressziós modell milyen feltételekkel egyezik meg a két szemléletmód mellett. Ezek egyike az, hogy a csoportok egyedszáma szerepeljen a regressziós modell magyarázó változói között. A többi ennek az egyszerű feltételnek különböző irányú matematikai finomítása.

Az említett valószínűségelméleti modellt és a hozzá kapcsolódó elméleti eredményeket alkalmaztam gímszarvasokat (*Cervus elaphus*) parazitáló szarvaskullancslegyek (*Lipoptena cervi*) ivararánya és a parazitáltság mértéke közötti összefüggés vizsgálatára. Az adatokra kevert hatásos logisztikus regressziót illesztettem úgy, hogy a gímszarvasokon talált paraziták száma a modell egyik magyarázó változója. Ez biztosította, hogy a regressziós modell ugyanaz legyen a

gímszarvasok és a szarvaskullancslegyek szempontjából. Kimutattam, hogy a hím szarvaskullancslegyek részaránya szignifikánsan kisebb az erősebben parazitált gímszarvasokon ($p = 0.0001$). Az eredmény érvényes mind a gímszarvasokra, mint egymástól független entitásokra, megfigyelési egységekre, mind pedig az egyes kullancslegyek saját gazdáján élő mikropopulációkra. A regressziós modell eredményét felhasználva levezettem, hogy míg egy tipikus gímszarvason az ivararány nem különbözik szignifikánsan az 1:1 aránytól, egy tipikus kullancslégy gazdakörnyezetén az ivararány a nőstények irányába tolódik el. A paradoxon magyarázata az, hogy egy tipikus kullancslégy az átlagosnál jobban parazitált gímszarvason él.

6 Új tudományos eredmények

1. Új módszert ismertettünk a valódi prevalencia közelítő konfidenciaintervallumának megszerkesztésére abban az esetben, amikor a diagnosztikai teszt szenzitivitását és specifikusságát független mintákból becsüljük.

2. Számítógépes szimulációs kísérletekkel igazoltuk, hogy az új konfidenciaintervallum a jelenleg elfogadott Wald-Rogan-Gladen konfidenciaintervallumnál pontosabban tartja meg az előírt megbízhatósági szintet még kicsi, 30 elemű mintákra is; a lefedés legalább 88%, 93% és 98%, ha az előírt szint rendre 90%, 95% és 99%.
3. Elméleti úton bizonyítottuk, hogy a valódi prevalencia Rogan-Gladen pontbecslése az új konfidenciaintervallumba esik, ha a diagnosztikai eljárás becsült szenzitivitása és specifikussága legalább 0.5, és a mintanagyság a szenzitivitásra ≥ 26 , a specifikusságra ≥ 26 , a célpopuláció esetében ≥ 16 .
4. Az új konfidenciaintervallum-szerkesztő eljárás előnyeit valós megfigyelések publikált adatainak újbóli feldolgozásával illusztráltuk.
5. Szimulációs kísérletekkel vizsgáltuk a pontosság romlását mintákból becsült szenzitivitás és specifikusság mellett Blaker és Clopper-Pearson módszerével szerkesztett

konfidenciaintervallumoknál. (Ezek az eljárások ismert szenzitivitást és specifikusságot feltételeznek.) Azt tapasztaltuk, hogy a pontosság romlása csak akkor válik elhanyagolhatóvá, ha a szenzitivitás és specifikusság becslésére kiválasztott minták nagysága legalább ötször nagyobb, mint a prevalencia becslésére szolgáló minta. A pontosság romlásának mértéke ezen túlmenően függ a prevalencia mintájának nagyságától és a szenzitivitás és specifikusság valódi értékétől is.

6. Számítógépes szimulációs kísérletekkel igazoltuk, hogy a Wald-Rogan-Gladen konfidenciaintervallum javítását célzó, széles körben alkalmazott és publikált $\hat{P} \pm 3 \cdot \text{var}(\hat{P})^{\frac{1}{2}}$ próbaintervallum alkalmazása az eljárás pontosságát nem javítja, emellett az esetek jelentős részében nem eredményez intervallumbecslést.

7. A magyarországi szarvasmarha-állományok BHV-1 szeropozitivitásából megbecsültük a BHV-1 fertőzöttség valódi prevalenciáját. Egzakt

konfidenciaintervallumokkal kimutattuk, hogy a valódi prevalencia minden korcsoportban és összességében is kisebb a mentesítés utáni ellenőrzéskor a mentesítést megelőző felméréshez képest. Megállapítottuk, hogy a valódi prevalenciák felméréskor nagyobbak, ellenőrzéskor pedig kisebbek, mint a szeropozitivitások, tehát a vizsgált periódusban (2002 és 2006 között) a mentesítés valójában eredményesebb volt, mint azt a szeropozitivitási arányok mutatják.

8. A prevalenciák megyék közötti és megyéken belüli ingadozását vizsgálva megállapítottuk, hogy a telepek között, megyétől függetlenül, nagy a heterogenitás mind a felmérő, mind az ellenőrző vizsgálatok során.
9. Az országot három földrajzilag összefüggő homogén megyecsoportra osztva megvizsgáltuk a prevalencia átlagát és csoporton belüli szórását. Kimutattuk, hogy felméréskor nincs lényeges különbség a csoportok átlaga és szórása között. Mentésítés után a Dunántúlon és

az északi megyékben valamivel jobban sikerült csökkenteni a fertőzés prevalenciáját, mint az Alföldön. A heterogenitás is csökkent, leginkább az északi megyékben.

10. Az állománylétszám és a prevalencia kapcsolatát szimultán konfidenciaintervallumokkal vizsgálva, Bonferroni módszerét követve kimutattuk, hogy felméréskor a nagyobb létszámú telepek csoportjához növekvő tendenciájú BHV-1 prevalencia tartozik, 5%-os szimultán tévedési valószínűség mellett. Megállapítottuk, hogy a mentesítés utáni ellenőrzésnél fordított a trend, a nagy létszámú telepeken 5%-os szinten szignifikánsan kisebb a prevalencia.

11. Korcsoportonként vizsgálva, felméréskor a borjak és tehenek esetében állapítottunk meg a telep létszáma és a prevalencia között 5%-os szinten szignifikáns pozitív kapcsolatot. A növendék és vemhes üszőknél nem találtunk szignifikáns összefüggést. A mentesítést követő ellenőrzéskor nem mutattunk ki korcsoportonként

szignifikáns kapcsolatot a telepek létszáma és a prevalencia között.

12. Kimutattuk, hogy a zsúfoltsági index felméréskor 867, ellenőrzéskor 817. Az átlagosan lehetséges kontaktusszám 50–80%-kal nagyobb, mint a telepek átlagos létszáma.
13. Fogalmi analógiát tártam fel a zsúfoltság és a fajok biológiai diverzitása között. A biológiai diverzitást Patil és Taille (1982) mint a fajok skálázott ritkaságának egyedekre vett átlagát értelmezték. Kimutattam, hogy a skálázó függvény megadásával a zsúfoltság és a diverzitás egymással kifejezhetők. Az egymásnak ilyen módon megfeleltetett legfontosabb zsúfoltsági és diverzitási indexek lineáris skálafüggvényre a Lloyd- és Reiczigel-féle lineáris zsúfoltság és a Simpson-féle diverzitás, logaritmus függvényre a logaritmusos zsúfoltság és a Shannon-féle diverzitási index, hiperbolikus függvényre a hiperbolikus zsúfoltság és a fajgazdagság (a fajok száma mínusz 1).

14. A diverzitások elméletében értelmezett effektív fajszám megfelelőjeként a zsúfoltság fogalomkörében tetszőleges monoton növekvő skálafüggvény mellett bevezettem az effektív csoportszám fogalmát. Ez olyan egyenlő egyedszámú csoportok számát jelenti, ahol a csoportok egyedeinek zsúfoltsága a vizsgált közösség zsúfoltságával egyezik meg. Erre a fogalomra építve, a tényleges csoportszám és az effektív csoportszám hányadosaként tetszőleges monoton növekvő skálafüggvényre általánosítottam a Bez-féle aggregáltsági indexet.
15. Általános feltételek mellett bebizonyítottam, hogy az aggregáltsági index ≥ 1 , és értéke pontosan akkor 1, ha az egyedek számának csoportok közötti megoszlása teljesen egyenletes.
16. Kimutattam, hogy a zsúfoltság és a diverzitás között közvetlen, nem csak a fogalmi analógia szintjén megnyilvánuló kapcsolat is van olyankor, ha a vizsgált közösség csoportjaiban az egyedek több fajhoz tartoznak. A fajok együttesére vett zsúfoltság felbontható a fajonkénti átlagos

zsúfoltság és az átlagos csoporton belüli, szubjektív diverzitás összegére.

17. Ha a vizsgált közösség csoportjaiban az egyedek több fajhoz tartoznak, akkor ismeretes, hogy a közösség együttes, ún. gamma valódi diverzitása (azaz effektív fajszáma) a csoportokon belüli átlagos valódi diverzitás (ún. alfa diverzitás) és a csoportok közötti változatosságot kifejező béta diverzitás szorzata. A zsúfoltság–diverzitás fogalmi analógia mentén a valódi diverzitásnak az aggregáltsági indexet feleltetem meg. A teljes közösség aggregáltságát gamma aggregáltságnak, a fajonkénti átlagos aggregáltságot alfa aggregáltságnak neveztem el. A béta aggregáltságot az alfa és gamma aggregáltságok hányadosaként értelmeztem, ez az aggregáltságnak a fajok közötti változatosságát kifejező tényezője.
18. A zsúfoltságra és a diverzitásra kapott elméleti eredményeket felhasználtam vetési varjak (*Corvus frugilegus*) és dolmányos varjak (*Corvus cornix*) tetűparazitáltságának

összehasonlítására. A paraziták zsúfoltsága vetési varjakon nagyobb, emellett a zsúfoltság szubjektív diverzitásos összetevőjének is nagyobb a százalékaránya. A varjófajok közötti eltérésre többek között azt a magyarázatot adtam, hogy a fertőzött dolmányos varjak körében nagyobb arányú az egyetlen tetűfajjal fertőzöttek, így 0 diverzitásúak aránya (50%), mint a fertőzött vetési varjakon (21%).

19. A valódi diverzitások és aggregáltságok multiplikatív felbontásával kimutattam, hogy a tetvek valódi alfa és gamma diverzitásai a vetési varjak esetében valamivel nagyobbak, mint a dolmányos varjaknál. A béta diverzitások a két varjófajnál nagyjából megegyeznek, értékük 1.5-2.5 között van. A tetűfajok átlagos, azaz alfa aggregáltsága a dolmányos varjak körében nagyobb, mint a vetési varjaknál. Ez azt jelenti, hogy a dolmányos varjak tetűfajai átlagosan aggregáltabban oszlanak meg a gazdaállatok között, mint a vetési varjak tetűfajai.

20. A béta aggregáltságra mindkét varjófajnál 1 és 2 közötti értékeket kaptam. Ebből kiindulva blokk klaszterezéssel és korrespondanciaelemzéssel kimutattam, hogy a vetési varjak génuszai két, gazdákon kevésbé átfedő csoportra, a Brueelia, illetve a Myrsidea és Menacanthus együttesére bonthatók. A dolmányos varjakon ugyanezekkel a statisztikai módszerekkel azt találtam, hogy a két domináns génusz, a Myrsidea és a Menacanthus alkot a gazdákon csak kismértékben átfedő csoportot. Érdekesség, hogy a vetési varjakon – a dolmányos varjakkal ellentétben – a Myrsidea és Menacanthus génuszok nem különülnek el, gyakran előfordulnak közös gazdán.

21. Általános valószínűségelméleti modellt vezettem be csoportokba szerveződő egyedek jellemzőinek egyedszemponitú és csoportszintű eloszlásaira. Formulákat adtam a két szemléletmód szerinti eloszlások közötti alapvető transzformációkra.

22. A bevezetett modellen belül valószínűségelméleti jelentést rendeltem az általános skálafüggvényű zsúfoltsághoz.
23. Elégséges és szükséges és elégséges feltételeket adtam arra, hogy egy általános regressziós modell milyen feltételekkel egyezik meg az egyedek és a csoportok szempontjából. Ezek egyike az, hogy a csoportok egyedszáma szerepeljen a regressziós modell magyarázó változói között. A többi ennek az alkalmazási feltételek és körülmények szerint kissé módosított változata.
24. Az említett valószínűségelméleti modellt és a hozzá kapcsolódó elméleti eredményeket alkalmaztam gímszarvasokat (*Cervus elaphus*) parazitáló szarvaskullancslegyek (*Lipoptena cervi*) ivararánya és a parazitáltság mértéke közötti összefüggés vizsgálatára. Az adatokra kevert hatásos logisztikus regressziót illesztettem úgy, hogy a gímszarvasokon talált paraziták száma a modell egyik magyarázó változója. Kimutattam, hogy a hím szarvaskullancslegyek

résaránya szignifikánsan kisebb az erősebben parazitált gímszarvasokon ($p = 0.0001$). Az eredmény érvényes mind a gímszarvasokra, mint egymástól független entitásokra, megfigyelési egységekre, mind pedig az egyes kullancslegyek saját gazdáján élő mikropopulációkra. A regressziós modell eredményéből levezettem, hogy míg egy tipikus gímszarvason az ivararány nem különbözik szignifikánsan az 1:1 aránytól, egy tipikus kullancslégy gazdakörnyezetén az ivararány a nőstények irányába tolódik el.

7 A doktori kutatás eredményeinek közlései

Lang Zs., Földi J., Ózsvári L., Reiczigel J.: Szeropozitivitás és prevalencia összefüggései hazai BHV-1-mentesítés adatainak járványtani elemzése alapján. *Magy. Állatorv. Lapja* 135, 525–534. 2013. IF₂₀₁₂: 0.146

Lang Z., Reiczigel J.: Confidence limits for prevalence of disease adjusted for estimated sensitivity and specificity. *Preventive Veterinary Medicine* 113, 13–22. 2014. IF₂₀₁₂: 2.389

Reiczigel J., Lang Z., Rózsa L., Tóthmérész B.: Properties of crowding indices and statistical tools to analyse crowding data. *Journal of Parasitology* 91, 245–252. 2005. IF₂₀₁₁: 1.320

Reiczigel J., Lang Z., Rózsa L., Tóthmérész B.: Measures of sociality: Two different views of group size. *Animal Behaviour* 75: 715–721. 2008. IF₂₀₁₂: 3.068

Vágó E., Lang Zs., Kemény S.: Overdispersion at the Binomial and Multinomial Distribution. *Periodica Polytechnica* 55, 17–20. 2011. IF₂₀₁₂: 0.269

8 Köszönetnyilvánítás

Szeretnék köszönetet mondani témavezetőmnek, Dr. Reiczigel Jenőnek és Dr. Rózsa Lajosnak, hogy kutatási témáikba bevontak és munkámat nagy hozzáértéssel segítették. Köszönetet mondok Dr. Földi Józsefnek és Dr. Ózsvári Lászlónak a közös kutató munkáért, lelkes támogatásukért. Köszönet illeti Dr. Kemény Sándort a kitartó buzdításért és Dr. Vágó Emesét hármunk közös kutató munkájáért. Köszönöm opponenseimnek, Dr. Harnos Andreának és Dr. Singer Júliának, hogy a dolgozatot elolvasták és észrevételeikkel annak színvonalát javították. Köszönetet mondok Dr. Boda Krisztinának és Dr. Földvári Gábornak, hogy doktori szigorlati vizsgáztatásomat elvállalták. Köszönöm Dr. Pásztory-Kovács Szilviának, hogy a fokozatszerzéssel összefüggő tapasztalatait megosztotta velem. Köszönöm Fehérvári Péternek és Dr. Vas Zoltánnak segítő megjegyzéseiket, javaslataikat.

Köszönöm feleségemnek, Annának és fiamnak, Péternek a szeretetet és türelmet, és szüleimnek, hogy mindig mellettem álltak.