

AG Landwirtschaftliches Versuchswesen; Juni 2005 BSA Hannover

**Statistischer Wirksamkeitsnachweis
neuer Pflanzenschutzmittel
Modifikationen der EPPO Richtlinie 152**

L. A. HOTHORN , H. BLEIHOLDER

UNIVERSITÄT HANNOVER , BASF LIMBURGERHOF

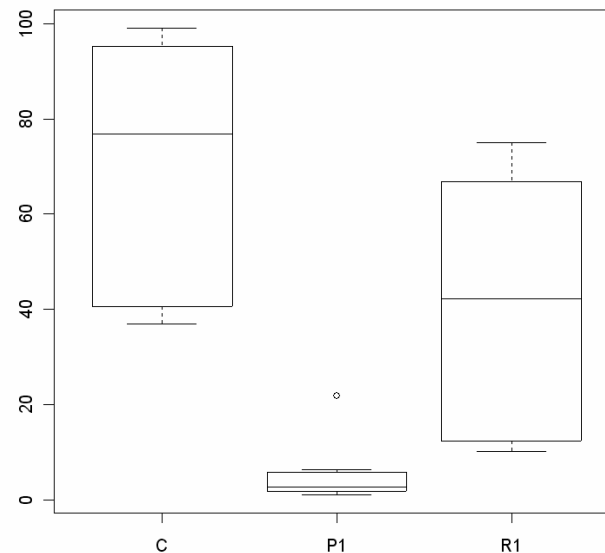
Email: hothorn@bioinf.uni-hannover.de

1) Analyse des Basis-Designs zum Nachweis der Wirksamkeit eines PSM

[Referenz, Produkt, Kontrolle]

Ein Beispiel: Percentage scores of *Septoria tritici* in winter wheat (variety RITMO) of the year 2002 of a randomly selected site (DE4).

Treat	Block	Value
Control	1	98
Control	2	68
Control	3	..
Control	4	...
Reference	1	13
Reference	2	10
Reference	3	11
Reference	4	26
Product	1	2
Product	2	1
Product	3	...
Product	4	..



EPPO:

1) *Is the trial realistic ... this will only be so if pest infestation in the untreated control is sufficiently high,*

Statistisch $H_0 : \mu_K < \delta_{\text{pest infestation}}$

2) *Are the results coherent ? Does the reference product give the expected result in comparison with the untreated control ?*

Statistisch $H_0 : \mu_R - \mu_K < \delta_{\text{coherence}}$

3 **If these two conditions are satisfied**, *it is then valid to compare the test products with the reference product and possibly to make comparisons between the products themselves. The analysis should aim primarily to **estimate the magnitude of the differences** between the various treatments and to provide an estimate of the variability of those estimates, often through a standard error or similar statistic. As a secondary aim the treatments should be compared using an appropriate statistical test.*

Statistisch :

$$H_0 : \mu_P - \mu_R \leq \delta_{\text{relevance}}$$

$$H_1 : \mu_P - \mu_R > \delta_{\text{relevance}} \quad \text{Test on non-inferiority or even superiority}$$

Idee: Nicht zwingend ein Überlegenheitsnachweis, sondern „lediglich“ ein Nichtunterlegenheitsnachweis, konditional gefolgt von einem Überlegenheitsnachweis

Falls mehrere Produkte: $H_{0_j} : \mu_{P_j} - \mu_{P_j'} = 0$ Tests zwischen den Produkten

Aber: die Relevanzkriterien $\delta_{\text{pest infestation}}$, $\delta_{\text{coherence}}$, $\delta_{\text{relevance}}$ sind weder durch die EPPO definiert, noch praktisch a-priori bekannt, sind kultur- und Befalls-abhängig,

→ Tests sind unrealistisch → Konfidenzintervalle sind das Mittel der Wahl: damit post-hoc Interpretation empirischer δ 's

Bemerkung: Nicht in der EPPO Guideline, aber praktisch : $H_0 : \mu_P - \mu_K < \delta_{\dots}$

Problem 1: Konfidenzintervalle für die Differenz(t-) oder den Quotienten (Fieller)?

KI für die Differenz ist „Software“-Standard, aber die Interpretation ist skalen-variant; Vorteil der KI for Ratio: % → skaleninvariant, d.h. interpretierbar über Kulturen, Befallsvarianten, Endpunkte, usw. möglich (aber diese KI müssen schätzbar sein)

Problem 2: Datenkondition (s. Beispiel): Annahme Gaußverteilung kühn, besser angepasst sind nichtparametrische KI, wegen der kleinen Fallzahlen sog. Exakte Methoden (man beachte Friedman ranks oder alignment for block design)

Test step	Parametric 95% confidence intervals for the difference	Non-parametric 95% confidence intervals for the difference
1) Trial realistic?	for C: [51.9; ∞] Sufficient level of infection	for C: [42.4; ∞] Sufficient level of infection
2) Results coherent?	for {R ₁ -C}: [-∞; -4.7] Coherent trial	for {R ₁ -C}: [-∞; -9.8] Coherent trial
3) Product vs. Reference	for {P ₁ -R ₁ }: [17.5; ∞] <i>Welch corrected:</i> [16.4; ∞] Product is not only non-inferior but superior to R ₁	for {P ₁ -R ₁ }: [9.1; ∞] Product is not only non-inferior but also superior to R ₁

Test step	Parametric 95% confidence intervals for the ratio	Non-parametric 95% confidence intervals for the ratio
2) Results coherent?	for {R ₁ /C}: [0; 0.91] Coherent trial	for {R ₁ /C}: [0; 0.88] Coherent trial
3) Product vs. Reference	for {P ₁ /R ₁ }: [2.1; ∞] Product is not only non-inferior but also superior to R ₁	for {P ₁ /R ₁ }: [3.6; ∞] Product is not only non-inferior but also superior to R ₁

Interpretation

Increasing effect		Decreasing effect	
Difference	Ratio	Difference	Ratio

Problem 3: Darf man für dieses multiple Testproblem (3 Hypothesen) jeweils einfach zum Niveau α testen?

Obige EPPO Sequenz stellt eine a-priori **Wichtigkeitsordnung** dar (Maurer et al. 1997): Zuerst hat man „realistic trial“ nachzuweisen, wenn dies der Fall ist, hat man sekundär die Kohärenz nachzuweisen, wenn dies der Fall ist, kann man auf Wirksamkeit (at least non-inferiority) testen (EPPO: „*If these two conditions are satisfied*“)

Damit ist ein PSM wirksam, wenn es {at least non-inferior to R} **UND** {R überlegen gegenüber K ist} **UND** {der Befall groß genug ist} **→ UND Verknüpfung → IUT**
Man beachte, daß die KI nur marginal sind, nicht simultan (vgl. Bretz et al., 2003)

Fazit:

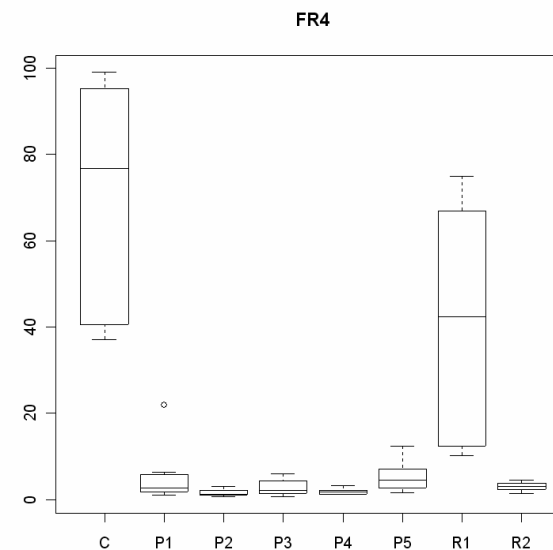
- **Man übersetze die EPPO-Sequenz in ein a-priori geordnetes Hypothesensystem,**
- **man benutze Konfidenzintervalle zur Interpretation von „Signifikanz“ und „Magnitude“, d.h. Relevanz,**
- **man bescheide sich auf einen Nichtunterlegenheitsnachweis {P vs R}**

2) Analyse des Standard-Designs zum Nachweis der Wirksamkeit eines PSM

Standard-Design mit 8 Prüfgliedern: [R₁, R₂, P₁, P₂, P₃, P₄, P₅, K]

Ein Beispiel: Infection scores in % for *Septoria tritici* on the second leaf (F-1 leaf) for the eight-armed design [R₁,R₂,P₁,P₂,P₃,P₄,P₅,C] were available from a multi-site field trial (Bleiholder, 2005) conducted on different sites in France in the year 1999 in winter wheat (variety RECITAL).

Treatment	Plot 1	Plot 2	Plot 3	Plot 4
Control	85.4	37	42.4	..
Product 1	1.52	21.9	2	..
Product 2	1.36	1.12	0.92	...
Product 3	1.8	4.72	0.72	...
Product 4	2.2	1.24	1.24	...
Product 5	2.88	3.32	1.64	...
Reference 1	13.1	11.6	10.2	...
Reference 2	2.52	1.36	2.84	..



EPPO: *Sensible contrasts to make might be:*

- *the untreated control versus the mean of the other seven treatments (1 df);*
- *reference product one versus reference product two (1 df);*
- *the mean of the reference products versus the mean of the test products(1 df);*
- *differences between the means of test products themselves (4 df).*

*Of these contrasts, the first two are designed to remove nuisance variation of relatively little biological interest, and it is the last two that help to reveal the true objectives of the trial. Each contrast provides a separate F -statistic, that may be used to test formally hypotheses of concern. In this example, *the important hypotheses might be that, on average, the test products were no better than the reference products, and that there were no differences between the test products themselves.* Interpretation of the first of these hypotheses might well be influenced by whether the contrast between the reference products themselves had exposed a large difference. If non-orthogonal contrasts are to be tested, for example *the separate five contrasts on 1 df between the mean of each test product against a specific reference product*, then these should also be done by F -test (or t -test if appropriate) using the residual mean square from the analysis of variance.*

SAS PROC GLM :

[K, P1, P2, P3, P4, P5, R1, R2]

```
proc glm data=eppo8;
class block treat;
model value=block treat;
contrast "contrast1" treat -7 1 1 1 1 1 1 1;
contrast "contrast2" treat 0 0 0 0 0 0 -1 1;
contrast "contrast3" treat 0 2 2 2 2 2 -5 -5;
contrast "contrast4" treat 0 -1 1 0 0 0 0 0,
treat 0 -1 0 1 0 0 0 0,
treat 0 -1 0 0 1 0 0 0,
treat 0 -1 0 0 0 1 0 0,
treat 0 0 -1 1 0 0 0 0,
treat 0 0 -1 0 1 0 0 0,
treat 0 0 -1 0 0 1 0 0,
treat 0 0 0 -1 1 0 0 0,
treat 0 0 0 -1 0 1 0 0,
treat 0 0 0 0 -1 1 0 0;
run;
```

p-Werte: <0.0001, 0.068, 0.153, and 0.928

Problem 1: Ignoriert selbst festgelegte Sequence, s.o.

Problem 2: F-distr. Contrasts sind inhärent 2-seitig definiert; einseitige t-distr. Contrasts sind problemadäquater

Problem 3: Erster Kontrast \bar{X}_i c_i P_1 P_2 P_3 P_4 P_5 R_1 R_2 poolt alle P_i und R_i . Das kann verzerrt sein, z.B. Daten-Manipulation für P_1 (heran an K): { 85.3, 37.1, 42.3, 38.7} $\rightarrow p < 0.0001$. Der Schluss alle Behandlungen sind sig. unterschiedlich von K ist falsch.

Lösung: Simultane CI für Vergleiche zu K (Dunnett, 1955, Dilba et al., 2004)

Comparison	Upper confidence limit
$P_{1,\text{manipulated}} - C$	19.6
$P_2 - C$	-30.1
$P_3 - C$	-29.1
$P_4 - C$	-29.7
$P_5 - C$	-28.6
$R_1 - C$	-16.0
$R_2 - C$	-29.0

Problem 4: Analyse des 2. Kontrasts $\{R_1 \text{ vs. } R_2\}$

\bar{X}_i	C	P_1	P_2	P_3	P_4	P_5	R_1	R_2
c_i	0	0	0	0	0	0	-1	1

→ wie geht man vor: i) Poolen von R_1 und R_2 falls ähnlich (äquivalent)(two-sided $(1-2\alpha)$ confidence interval) oder a-priori getrennt für R_1 und R_2 (das kann widersprüchlich sein)?

Problem 5: Analyse des 3. Kontrasts

\bar{X}_i	C	P_1	P_2	P_3	P_4	P_5	R_1	R_2
c_i	0	2	2	2	2	2	-5	-5

Nimmt die Äquivalenz von R_1 und R_2 an, und vergleicht gegen das Mittel der Produkte. Erneut verzerrt → individuelle Vergleich der Produkte gegen die Referenzen (gepoolt oder separate) UND OHNE Multiplizitätsadjustierung, da die Wirksamkeitsentscheidung eines Produktes unabhängig von der Zahl und Art anderer Produkte sein sollte.

Problem 6: Analyse des 4. Kontrasts All-Paar-Vergleiche zwischen den Produkten: nicht als Globaltest, nicht als regulatorischer Wirksamkeitsnachweis, falls gewünscht All-pairs-Comparison procedure Tukey (1953) → simultane CI

Comparison	95% two-sided simultaneous CI
P ₁ - P ₂	[-4.4; 15.4]
P ₁ - P ₃	[-5.4; 14.5]
P ₁ - P ₄	[-4.7; 15.0]
P ₁ - P ₅	[-5.9; 13.9]
P ₂ - P ₃	[-10.9; 8.9]
P ₂ - P ₄	[-10.3; 9.5]
P ₂ - P ₅	[-11.4; 8.4]
P ₃ - P ₄	[-9.3; -10.5]
P ₃ - P ₅	[-10.4; 9.4]
P ₄ - P ₅	[-11.0; 8.8]

Für die Registrierung eines ausgewählten Produktes sollte das Sub-Design [R₁,R₂,P,K] aus der randomisierten Anlage herausgeschnitten werden ODER der Kontrast

$$\bar{X}_i \quad C \quad P_1 \quad P_2 \quad P_3 \quad P_4 \quad P_5 \quad R_1 \quad R_2$$

$$c_i \quad 0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad -1 \quad 0$$

mit gemeinsamen MQ_R und df benutzt werden. Aber damit hängt die Entscheidung von den Werten anderer Produkte ab.

Fazit:

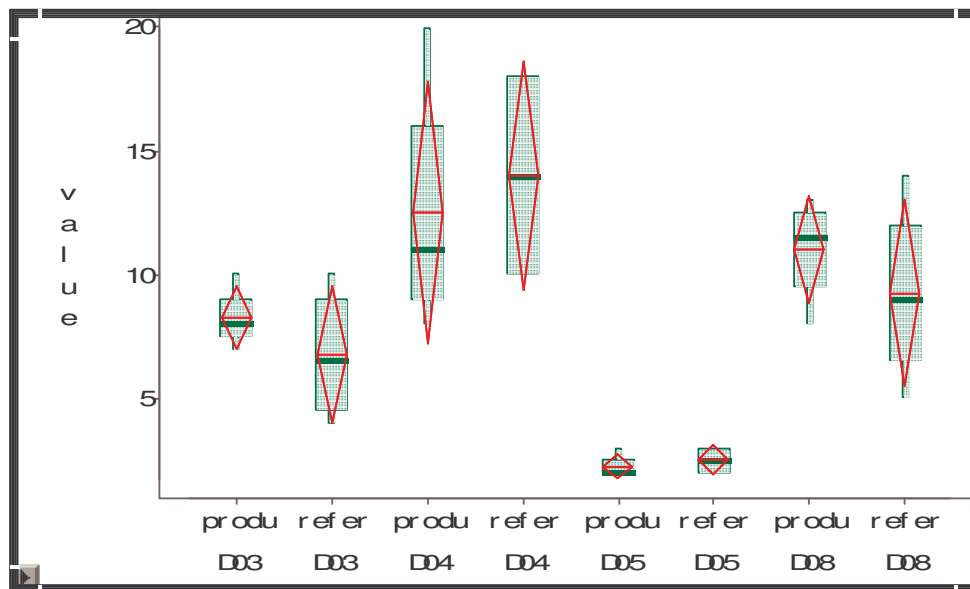
- Man nehme nur ein Referenzprodukt oder Löse den möglichen Widerspruch zwischen {R1 und R2}.
- Man schneide das Sub-Design [R1,R2,P,K] heraus (unter Verzicht auf df)

3) Analyse von Versuchsserien

Zahlreiche Wirksamkeitsstudien werden multi-zentrisch und in mehreren Jahren realisiert.

EPPO guideline “... *the analysis is to measure and test for possible interactions between products and sites, i.e. to verify that the differences between products are consistent from one site to another*”

Beispiel: Daten 1 (s.o.) für die Standorte: DE03, DE04, DE05 and DE08 (nur Standorte)



Ansätze:

- Fixed effect model → Schätzung der Interaktion (s.o. EPPO Formulierung)
- Mixed effect model → Modellierung der Interaktion und der Sites als Varianzkomponente

Problem 1: Ist der F-Test auf Interaktion im Modell mit festen Effekten adäquat?

Nein, da nur GLOBALtest, aber man will wissen welche Standorte andere Wirksamkeiten ausweisen

Lösung: Interaktionskontraste (SAS-macro *simintervals* (Westfall et al., 1999)) (1-2 α) KI auf Äquivalenz

$(\mu_{ij} - \mu_{ij'}) - (\mu_{i'j} - \mu_{i'j'})$	(1-2α) confidence intervals
$(\mu_{\text{Product,site1}} - \mu_{\text{Reference,site1}}) - (\mu_{\text{Product,site2}} - \mu_{\text{Reference,site2}})$	[-4.5; 10.5]
$(\mu_{\text{Product,site1}} - \mu_{\text{Reference,site1}}) - (\mu_{\text{Product,site3}} - \mu_{\text{Reference,site3}})$	[-5.8; 9.3]
$(\mu_{\text{Product,site1}} - \mu_{\text{Reference,site1}}) - (\mu_{\text{Product,site4}} - \mu_{\text{Reference,site4}})$	[-7.8; 7.3]
$(\mu_{\text{Product,site2}} - \mu_{\text{Reference,site2}}) - (\mu_{\text{Product,site3}} - \mu_{\text{Reference,site3}})$	[-8.8; 6.3]
$(\mu_{\text{Product,site2}} - \mu_{\text{Reference,site2}}) - (\mu_{\text{Product,site4}} - \mu_{\text{Reference,site4}})$	[-10.8; 4.3]
$(\mu_{\text{Product,site3}} - \mu_{\text{Reference,site3}}) - (\mu_{\text{Product,site4}} - \mu_{\text{Reference,site4}})$	[-9.5; 5.5]

Problem 2: . Wie schätzt man das Konfidenzintervall für den festen Faktor im gemischten Modell?

- Standard für die Differenz (covariance parameter estimates for site is 20.0 and the error is 9.2, $\mu_{\text{Reference}} - \mu_{\text{Product}}$ [-2.2; ∞].)
- Ratio-to-control one-sided $(1-\alpha)$ confidence interval $\mu_{\text{Reference}} / \mu_{\text{Product}}$ R-code in Hothorn and Oberdörfer (2005) **0.96** [0.81; ∞]

Fazit: KI für Differenz und Ratio ist im gemischten Modell inzwischen numerisch verfügbar; damit meine Empfehlung. Aber evtl. bevorzugen die Behörden und Praktiker die Interaktionsanalyse: dann lokale Interaktionskontraste statt globalem F-Test

4) Weitere Probleme

- **multiple Testen**
- **multiple Endpunkte**
- **Bestimmung einer wirksamen Konzentration**
- **Uws.**
- Draft for Ann. Appl. Biol. : Statistical aspects of efficacy evaluation of plant protection products in field trials - a comment to the EPPO PP1/152(2) guideline

5) References

- Anonymous.** 1998. OEPP/EPPO: Design and analysis of efficacy evaluation trials, EPPO Standard PP1/152(2). Bulletin OEPP/EPPO Bulletin 28, 273-348.
- Bretz F, Hothorn LA, Hsu J.** 2003. Identifying effective and/or safe doses by stepwise confidence intervals for ratios. *Statistics in Medicine* **22**: 847-858.
- Dunnnett CW.** 1955. A multiple comparison procedure for comparing several treatments with a control. *J. Am Stat. Assoc.* **50**: 1096-1121.
- Fieller E.** 1954. Some problems in interval estimation. *Journal of the Royal Statistical Society B*16: 175-185.
- Hothorn LA, Oberdoerfer R.** 2005. Statistical Analysis Used in the Nutritional Assessment of Novel Food using the Proof of Safety. *J. Agri. Food. Chem.* (subm)
- Hothorn T, Munzel U.** 2002. Non-parametric confidence interval for the ratio. Report University of Erlangen, Department Medical Statistics (www.uni-erlangen.de)
- Hsu JC.** 1996. *Multiple Comparisons: Theory and Methods*, London, Chapman & Hall.
- Maurer W, Hothorn LA, Lehmacher W.** 1995. Multiple comparisons in drug clinical trials and preclinical assays: a-priori ordered hypotheses. In: Vollmar, J. (ed.) *Biometrie in der chemisch-pharmazeutischen Industrie*, Volume 6 , Fischer Verlag Stuttgart, 3-18.
- Navarro-Llopis V, Sanchis-Cabanes J, Ayala I.** 2004. Efficacy of lufenuron as chemosterilant against *Ceratitis capitata* in field trials. *Pest Management Science* 60: 914-920.
- Perry JN.** 1986. Multiple-comparison procedures: a dissenting view. *Journal of Economic Entomology* **79**: 1149-1155.
- Seidel D, Hothorn LA.** 2000. Gerichtete Tests für geordnete kategoriale Daten bei sehr kleinen Fallzahlen. *Vortrag, Biometrische Kolloquium Rostock.*
- Westfall, P. H.; Tobias, R. D.; Rom, D.; Wolfinger, R. D.; Hochberg, Y.** 1999. *Multiple Comparisons and Multiple Tests Using the SAS System.* Cary, SAS Institute Inc.