Pairwise comparisons of means with unequal variances in Stata

Daniel Klein & Felix Bittmann
German Stata Conference 2025
March 28 | Hamburg

LIfBi
LEIBNIZ-INSTITUT FÜR BILDUNGSVERLÄUFE

DZHW
Deutsches Zentrum für Hochschul- und Wissenschaftsforschung

The Problem

The ANOVA model

$$y_{ij} = \mu_j + \epsilon_{ij}$$

with

$i = 1, 2, \ldots, n_j$ observations
$j = 1, 2, \ldots, k$ groups

implies $k^\star = k(k-1)/2$ pairwise comparisons

$k^\star$ independent t-tests have family-wise type I error rate

$$_{FWER}\alpha = 1 - (1-\alpha)^{k^\star} \geq \alpha$$

Proposed solutions: Adjust α and p

Bonferroni

$$ {}_{b}\alpha = \alpha / k^{\star} $$

$$ {}_{b}p = \min(1, pk^{\star}) $$

Šidák

$$ {}_{si}\alpha = 1 - (1 - \alpha)^{1/k^{\star}} $$

$$ {}_{si}p = 1 - (1 - p)^{k^{\star}} $$

LIfBi
LEIBNIZ-INSTITUT FÜR BILDUNGSVERLÄUFE

DZHW.

Proposed solutions: Swap distributions

Scheffé

$${}_{sc}c(\alpha) = \sqrt{{}_{inv}F_{k-1,\nu,\alpha}(k-1)}$$

$${}_{sc}p = F_{k-1,\nu,|t_0|^2/(k-1)}$$

Tukey

$${}_{t}c(\alpha) = {}_{inv}q_{k,\nu,\alpha}/\sqrt{2}$$

$${}_{t}p = q_{k,\nu,|t_0|\sqrt{2}}$$

LIfBi LEIBNIZ-INSTITUT FÜR BILDUNGSVERLÄUFE

DZHW.

Proposed solutions: Adjust Std. err. and df

Tamhane

$$_{t2}c(\alpha) = \quad _{inv}t_{\hat{\nu},[1-\{1-\alpha\}^{1/k^{\star}}]/2}$$

$$_{t2}p = 1 - \left(1 - 2t_{\hat{\nu},|t_0|}\right)^{k^{\star}}$$

Games & Howell

$$_{gh}c(\alpha) = \quad _{inv}q_{k,\hat{\nu},\alpha}/\sqrt{2}$$

$$_{gh}p = q_{k,\hat{\nu},|t_0|\sqrt{2}}$$

LIfBi LEIBNIZ-INSTITUT FÜR BILDUNGSVERLÄUFE

DZHW.

Proposed solutions: Adjust Std. err. and df

Dunnett

$$ {}_{c}c(\alpha) = \left[\frac{{}_{inv}q_{k,\nu_l,\alpha}\left\{\frac{s_l^2}{n_l}\right\} + {}_{inv}q_{k,\nu_m,\alpha}\left\{\frac{s_m^2}{n_m}\right\}}{s_l^2/n_l + s_m^2/n_m} \right] / \sqrt{2} $$

$$ {}_{c}p = \frac{q_{k,\nu_l,|t_0|\sqrt{2}}\left(\frac{s_l^2}{n_l}\right) + q_{k,\nu_m,|t_0|\sqrt{2}}\left(\frac{s_m^2}{n_m}\right)}{s_l^2/n_l + s_m^2/n_m} $$

LEIBNIZ-INSTITUT FÜR BILDUNGSVERLÄUFE

Title

[COMMUNITY-CONTRIBUTED] **pwmc** — Pairwise multiple comparisons of means with unequal variances

Syntax

Pairwise multiple comparisons of means

pwmc *varname* [*if*] [*in*] **,** **over(***varname***)** [*options*]

options	Description
Main	
* **over(***varname***)**	compare means over the levels of *varname*
Reporting	
mcompare(*method***)**	adjust for multiple comparisons; default is **mcompare(gh)**
se(*se_type***)**	type of standard error; default is **se(hc2)**
df(*df_method* | **#)**	degrees of freedom for computing confidence intervals and *p*-values; default is **df(satterthwaite)**

method	Description
gh	Games and Howell's method; synonyms **games** or **howell**; the default
cochran	Dunnett's C method
tamhane	Tamhane's method; synonym **t2**
noadjust	do not adjust for multiple comparisons

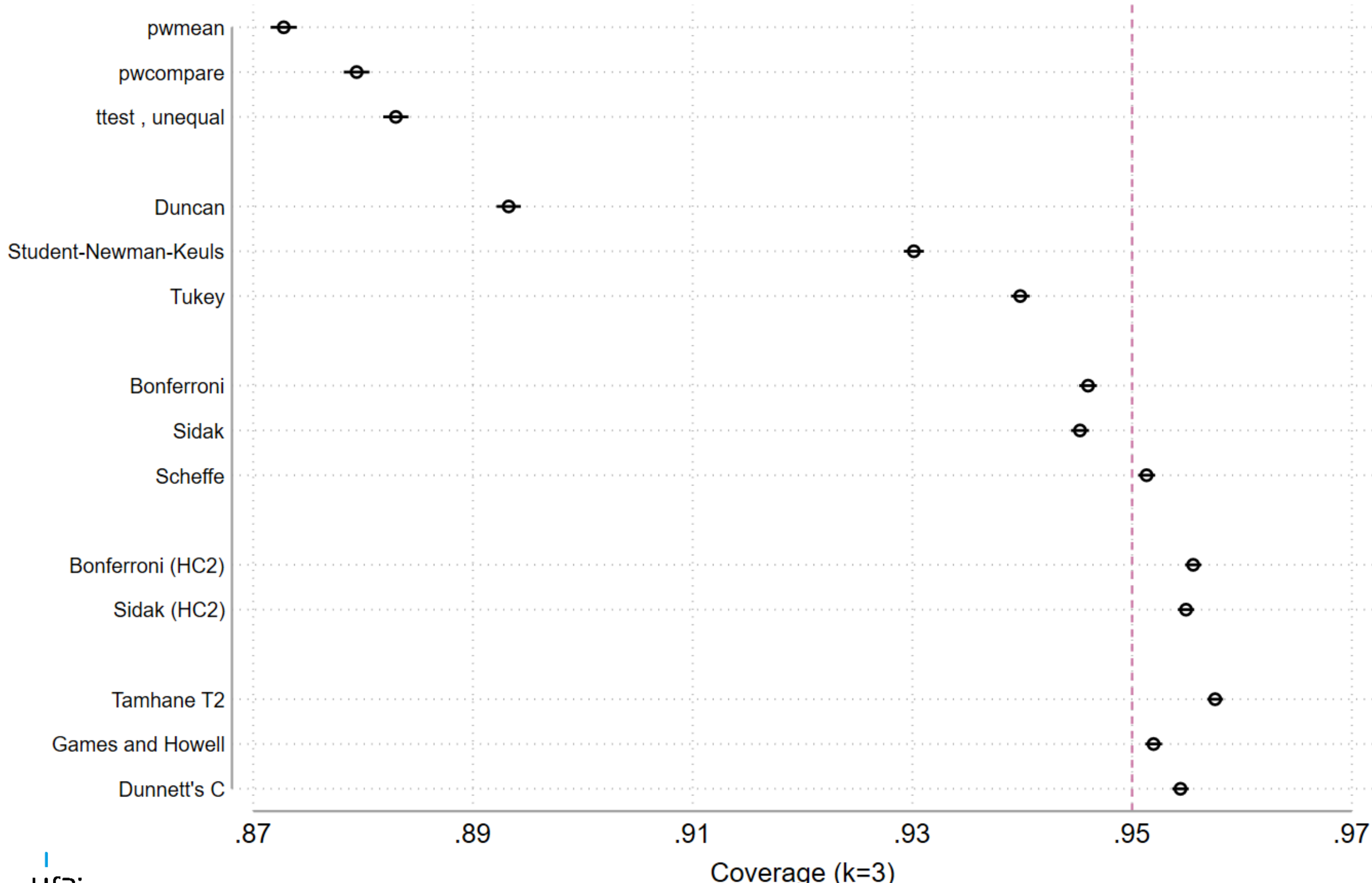
se_type	Description
hc2	robust HC2 standard errors (see regress); the default
hc3	robust HC3 standard errors (see regress)
ols	ordinary least-squares standard errors (see regress)

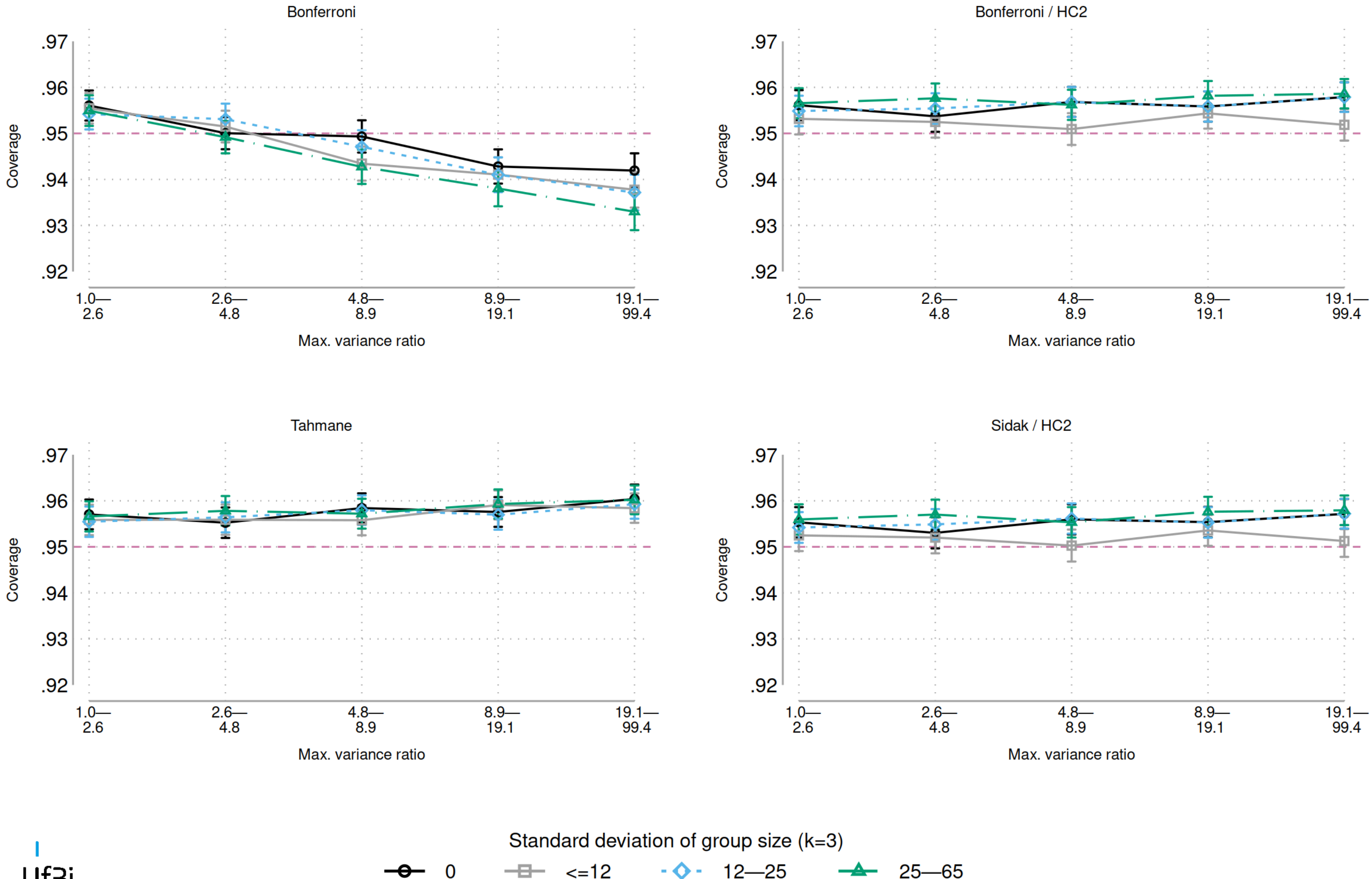
df_method	Description
satterthwaite	Satterthwaite's approximation; the default
welch	Welch's approximation
bm	Bell and McCaffrey's adjustment (see regress)
residual	residual degrees of freedom

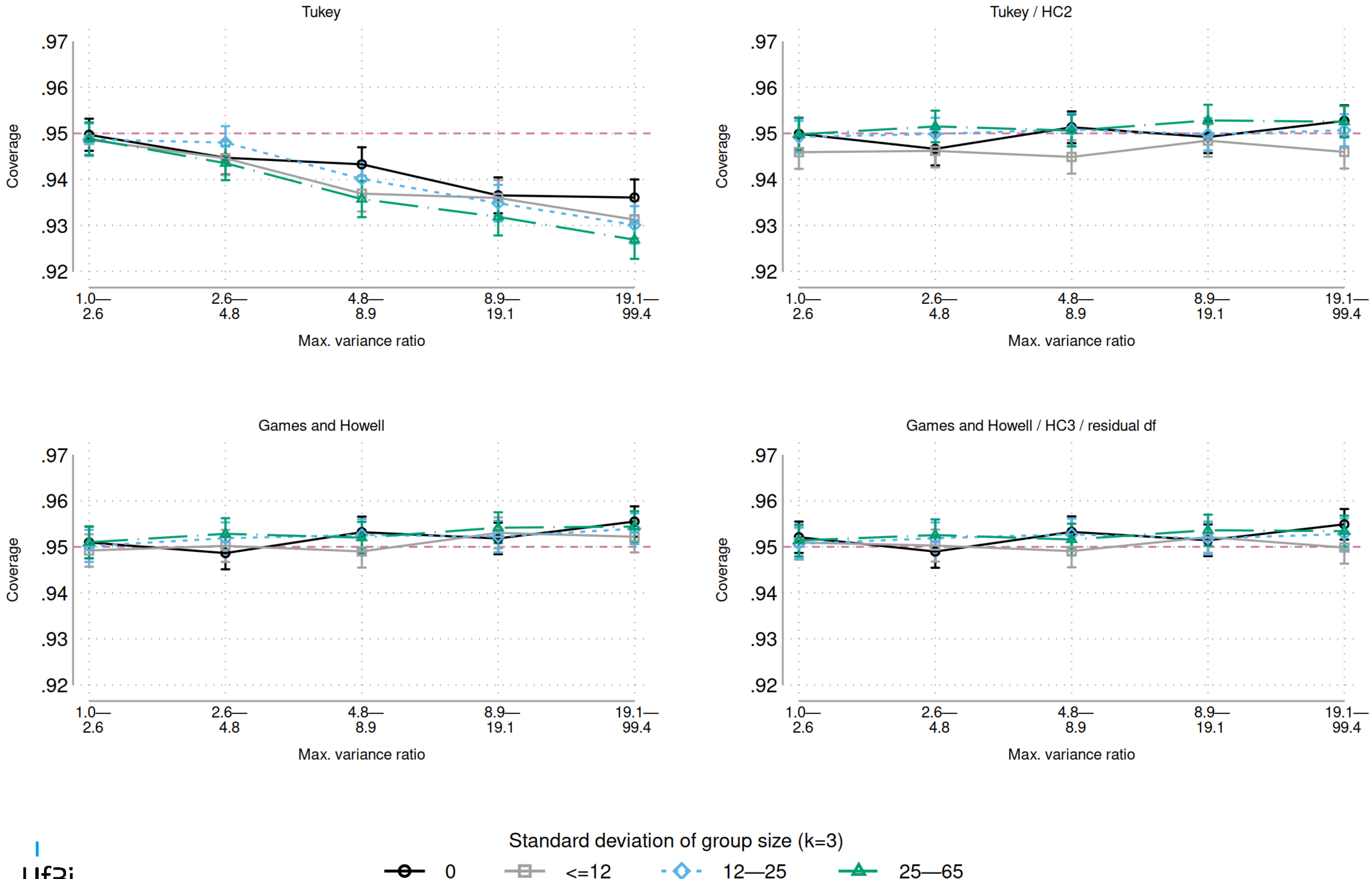
Simulation

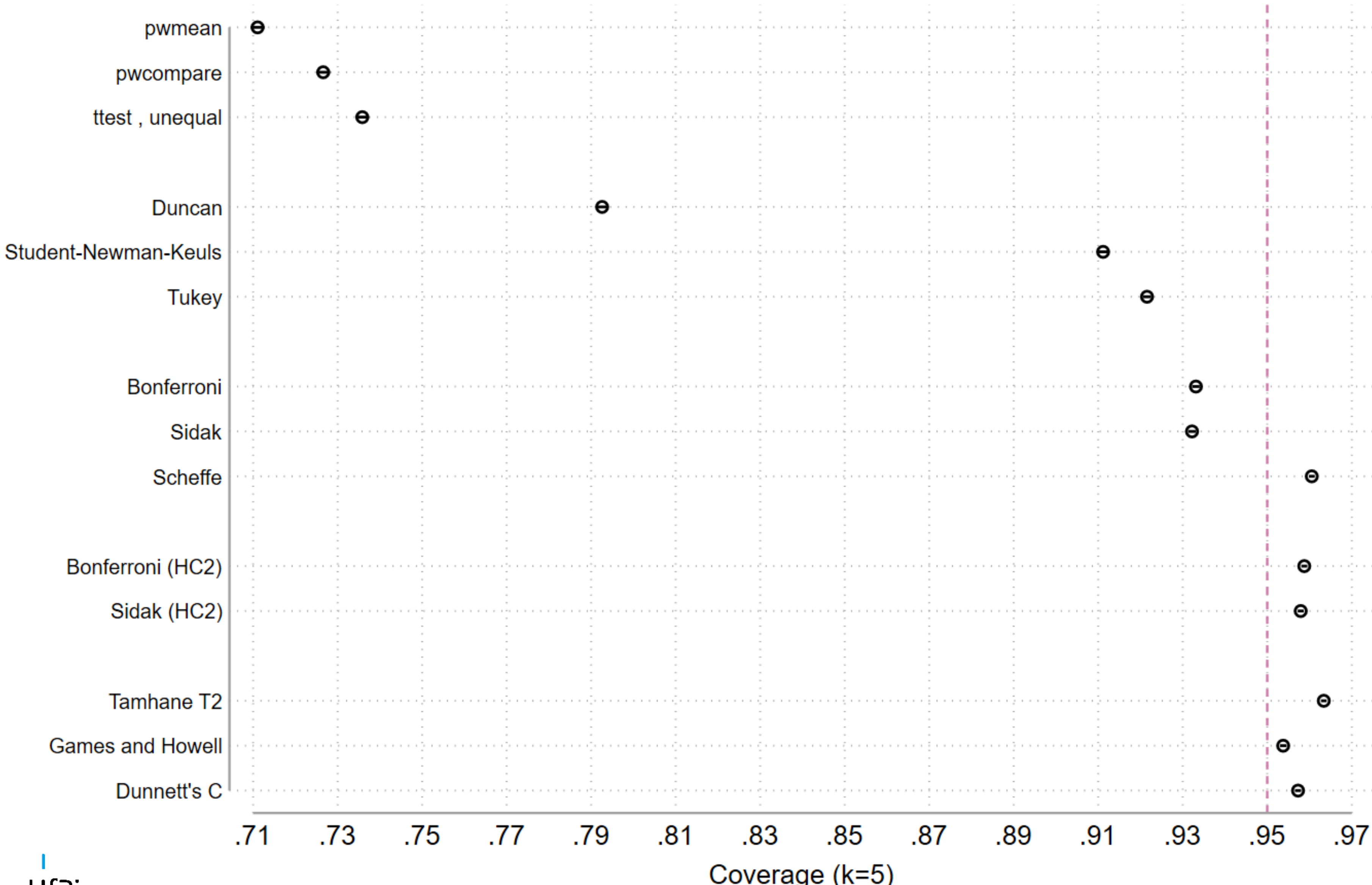
Groups:	$k = 3, k = 5$
Outcomes:	$N(0,\ \sigma_j^2)$
Std. deviation:	$[0.3, 3.5]$
Variance ratio:	$[1, \sim 100]$
Group sample sizes:	$[10, 200]$[a]
N simulations / datasets:	300,000[b]
Method / Std. err. / df:	$6 \times 4 \times 4$[c]
Coverage:	all $k^\star$ CIs include 0

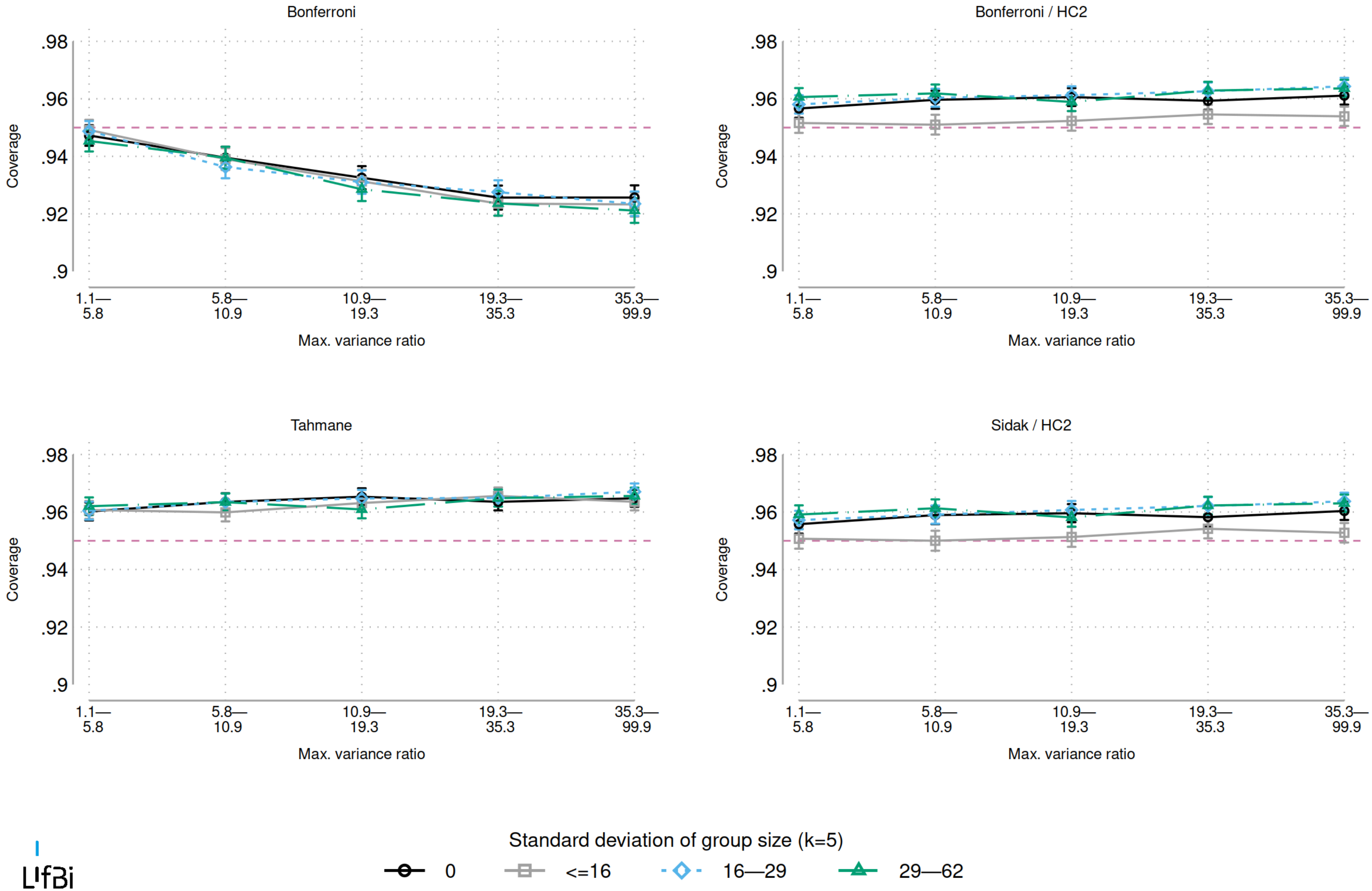
[a] restricted to range between $\mathrm{n}_1 \times 0.5$ and $n_1 \times 1.5_1$
[b] $\mathrm{N} = 75.000$ (equal variances), $\mathrm{N} = 225.000$ (unequal variances) using `parallel` (https://github.com/gvegayon/parallel)
[c] do not differentiate between Šidák and Tamhane, and Tukey and Games & Howell additionally, SNK and Duncan

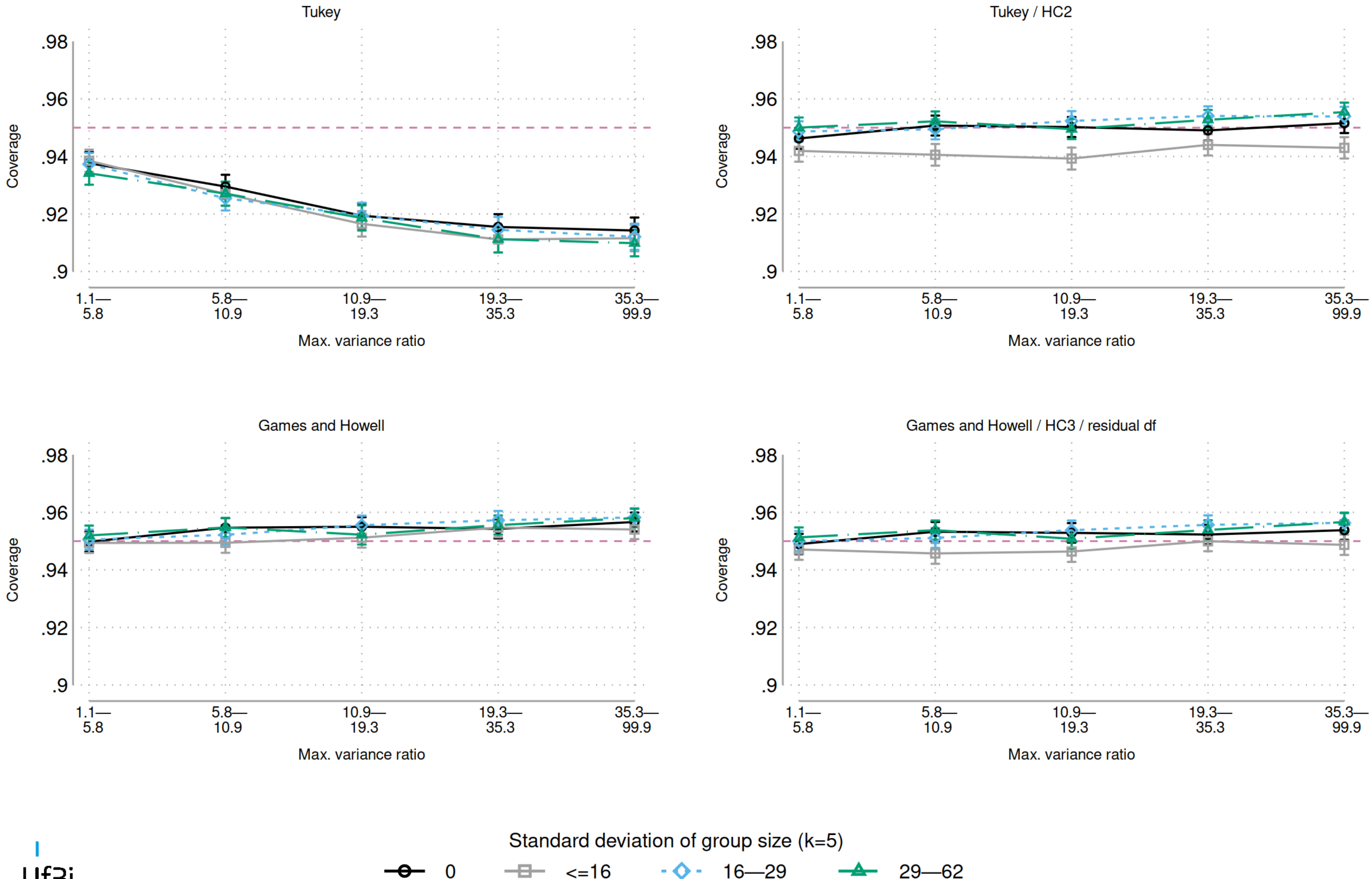

pwmean
pwcompare
ttest , unequal
Duncan
Student-Newman-Keuls
Tukey
Bonferroni
Sidak
Scheffe
Bonferroni (HC2)
Sidak (HC2)
Tamhane T2
Games and Howell
Dunnett's C
.87
.89
.91
.93
.95
.97
Coverage (k=3)


Bonferroni
Bonferroni / HC2
Tahmane
Sidak / HC2
Coverage
.97
.96
.95
.94
.93
.92
1.0—2.6
2.6—4.8
4.8—8.9
8.9—19.1
19.1—99.4
Max. variance ratio
Standard deviation of group size (k=3)
0
<=12
12—25
25—65


Tukey
Tukey / HC2
Games and Howell
Games and Howell / HC3 / residual df
Coverage
.97
.96
.95
.94
.93
.92
1.0—2.6
2.6—4.8
4.8—8.9
8.9—19.1
19.1—99.4
Max. variance ratio
Standard deviation of group size (k=3)
0
<=12
12—25
25—65


pwmean
pwcompare
ttest , unequal
Duncan
Student-Newman-Keuls
Tukey
Bonferroni
Sidak
Scheffe
Bonferroni (HC2)
Sidak (HC2)
Tamhane T2
Games and Howell
Dunnett's C
.71
.73
.75
.77
.79
.81
.83
.85
.87
.89
.91
.93
.95
.97
Coverage (k=5)


Bonferroni
Bonferroni / HC2
Tahmane
Sidak / HC2
Coverage
.98
.96
.94
.92
.9
1.1—5.8
5.8—10.9
10.9—19.3
19.3—35.3
35.3—99.9
Max. variance ratio
Standard deviation of group size (k=5)
0
<=16
16—29
29—62


Tukey
Tukey / HC2
Games and Howell
Games and Howell / HC3 / residual df
Coverage
.98
.96
.94
.92
.9
1.1—5.8
5.8—10.9
10.9—19.3
19.3—35.3
35.3—99.9
Max. variance ratio
Standard deviation of group size (k=5)
0
<=16
16—29
29—62

Conclusion

Best approach depends on specific scenario

But …

… prefer Šidák over Bonferroni

… use robust standard errors

… generally, don’t use Tukey

… until StataCorp allows robust std. err.

… use Games & Howell instead

Download `pwmc` from SSC
or GitHub (https://github.com/kleindaniel81/pwmc)

LIfBi LEIBNIZ-INSTITUT FÜR BILDUNGSVERLÄUFE

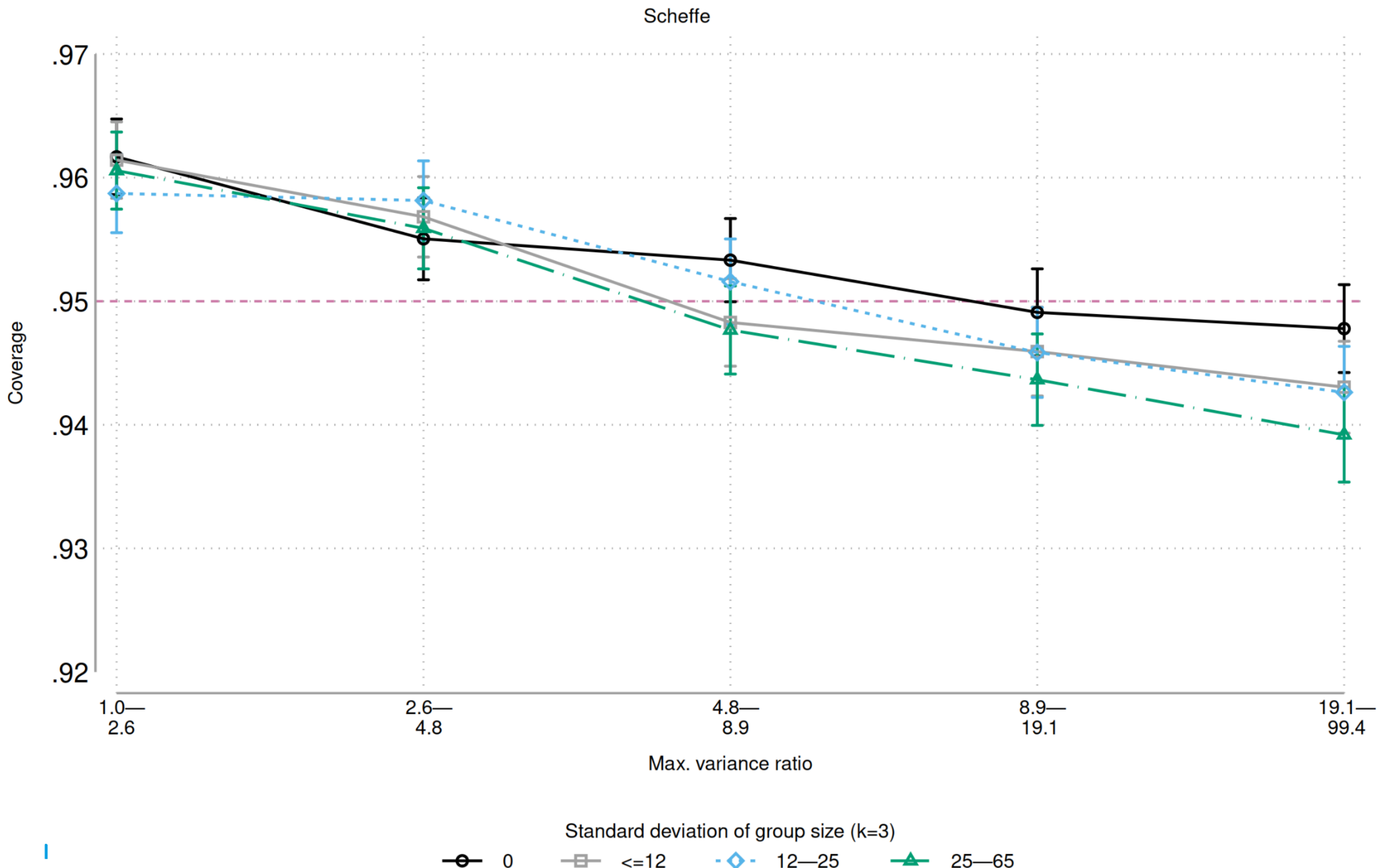
DZHW.

Contact

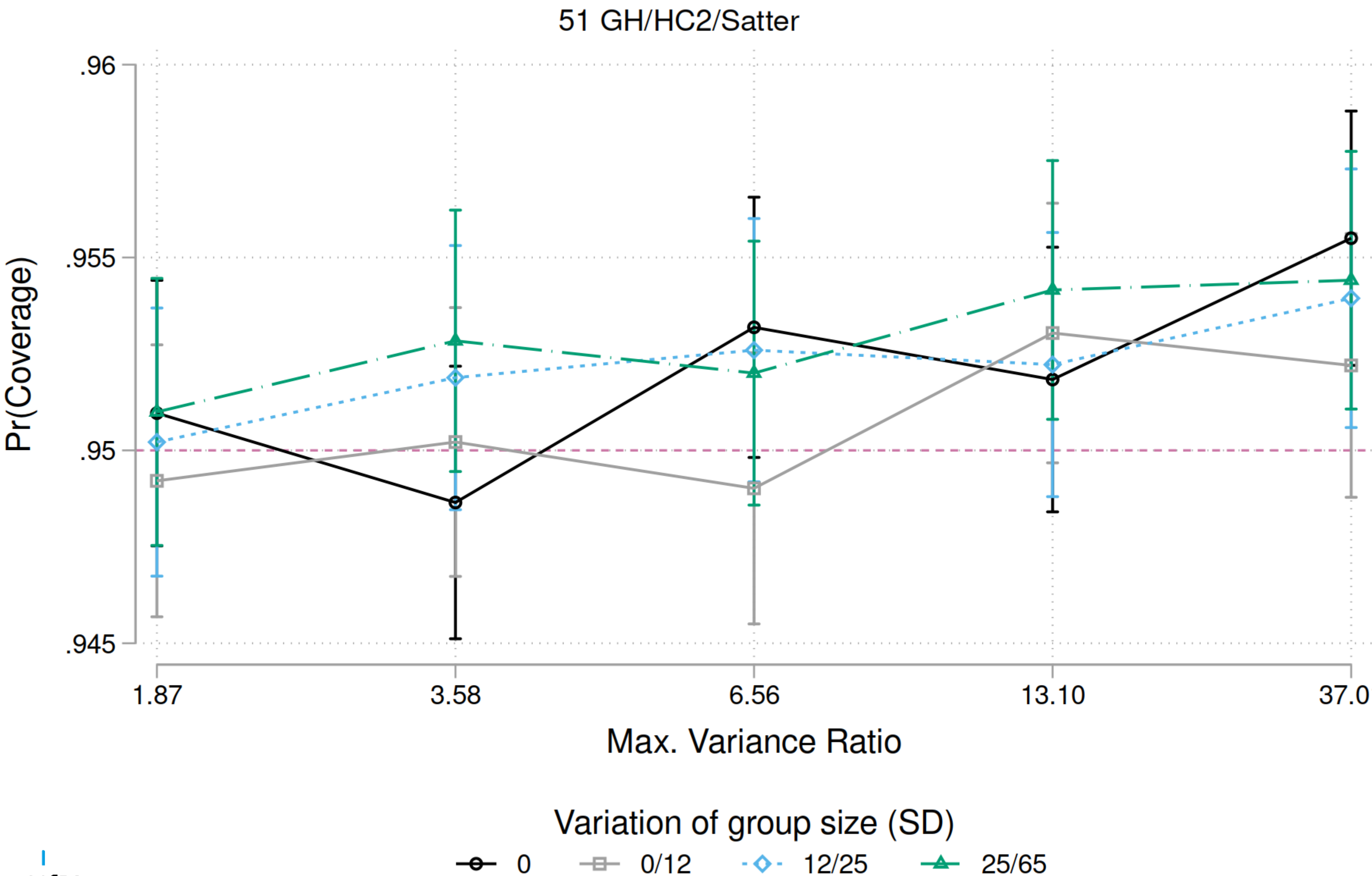
klein@dzhw.eu

felix.bittmann@lifbi.de

Backup (k=3)

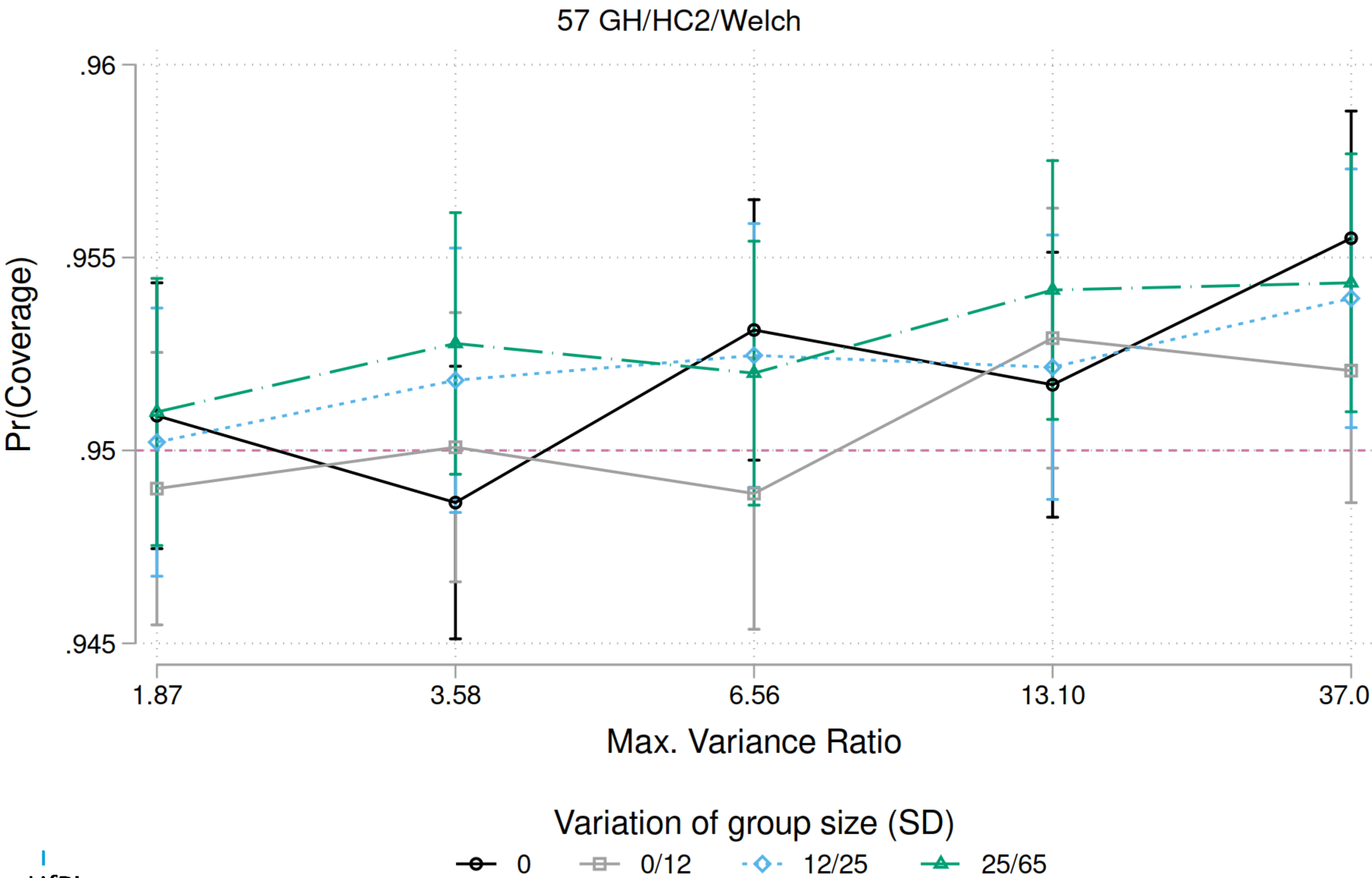


Scheffe
.97
.96
.95
.94
.93
.92
Coverage
1.0—
2.6
2.6—
4.8
4.8—
8.9
8.9—
19.1
19.1—
99.4
Max. variance ratio
Standard deviation of group size (k=3)
0
<=12
12—25
25—65



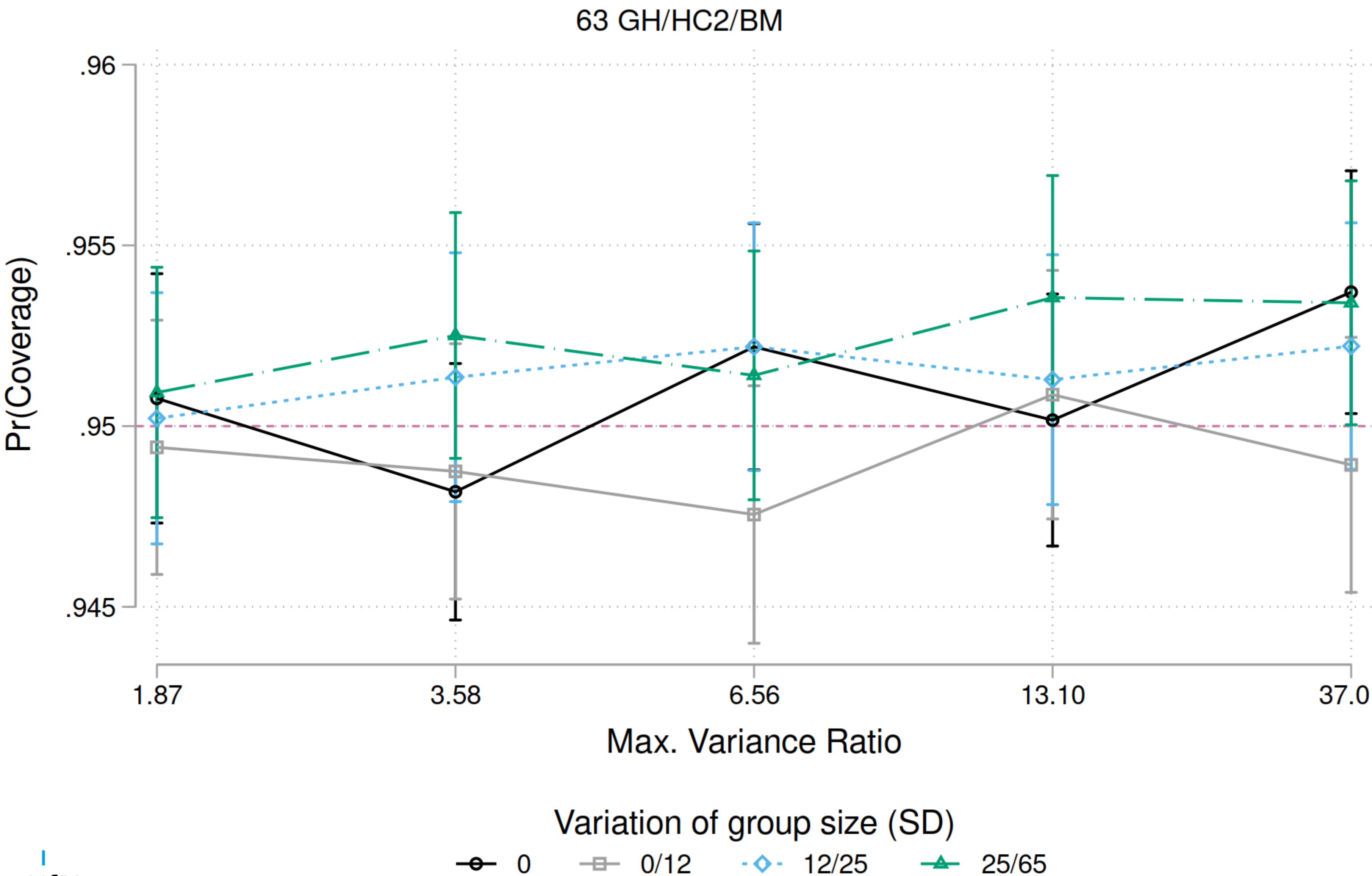
LIfBi LEIBNIZ-INSTITUT FÜR BILDUNGSVERLÄUFE

DZHW.



LIfBi LEIBNIZ-INSTITUT FÜR BILDUNGSVERLÄUFE

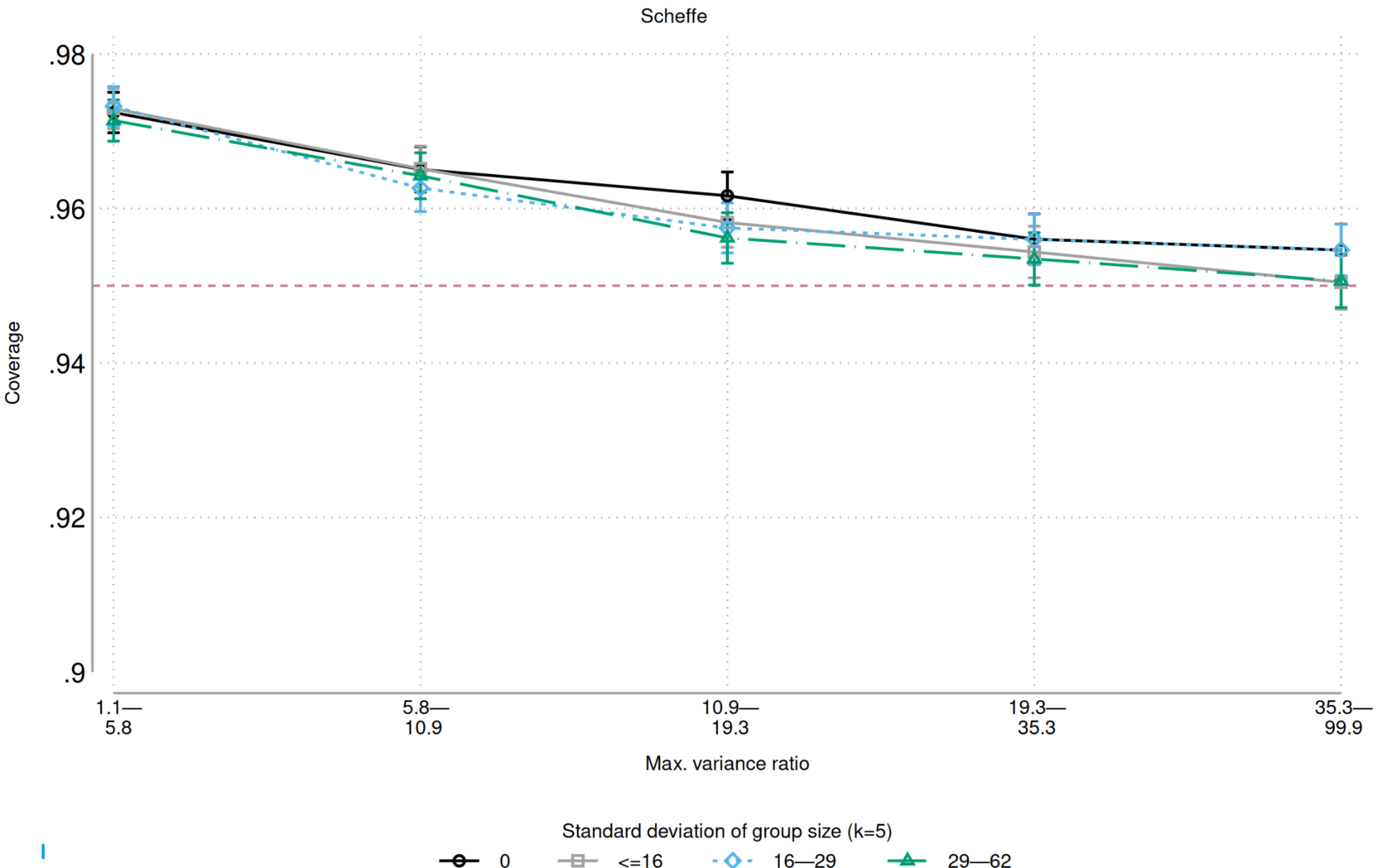
DZHW.

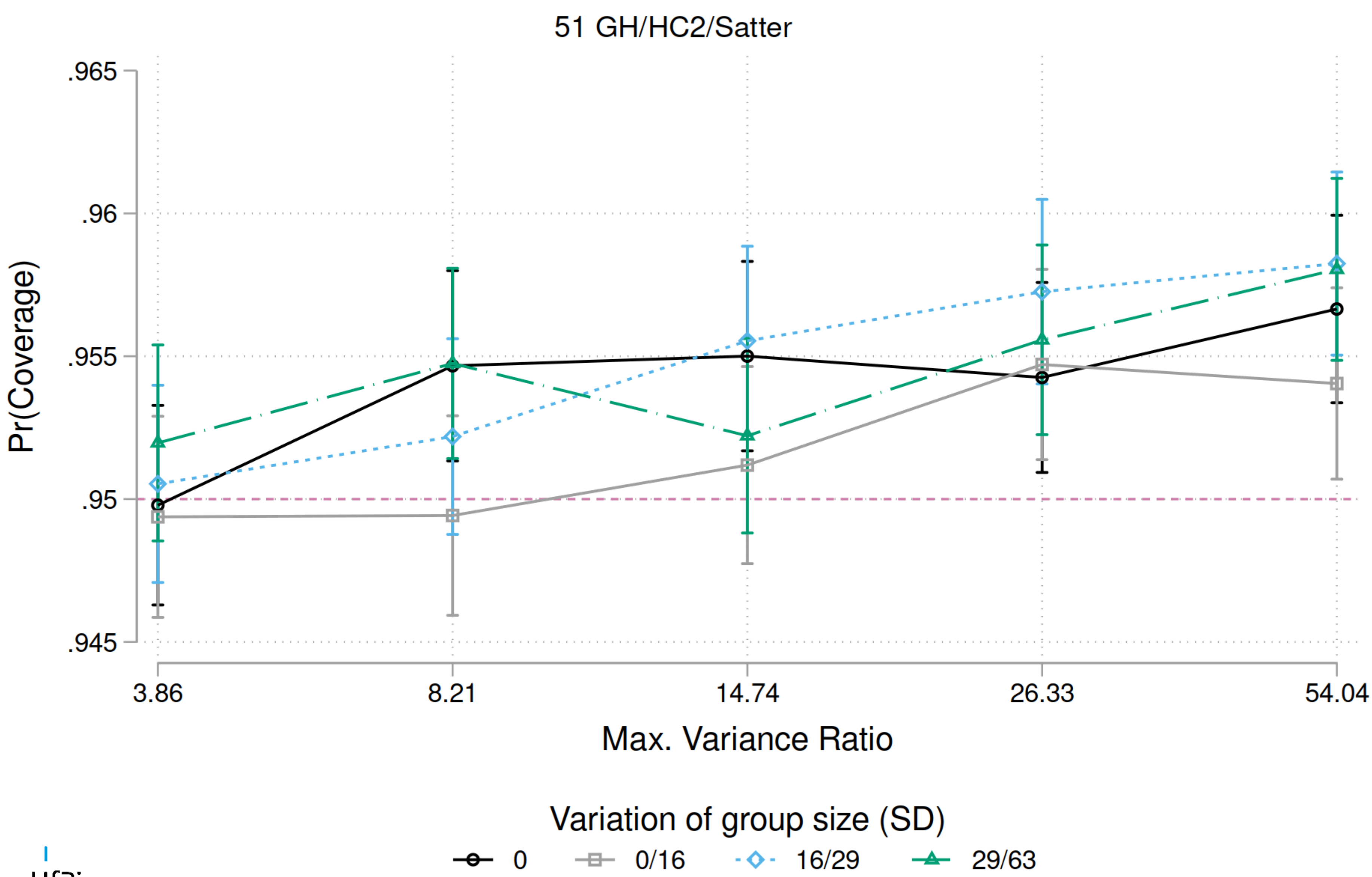


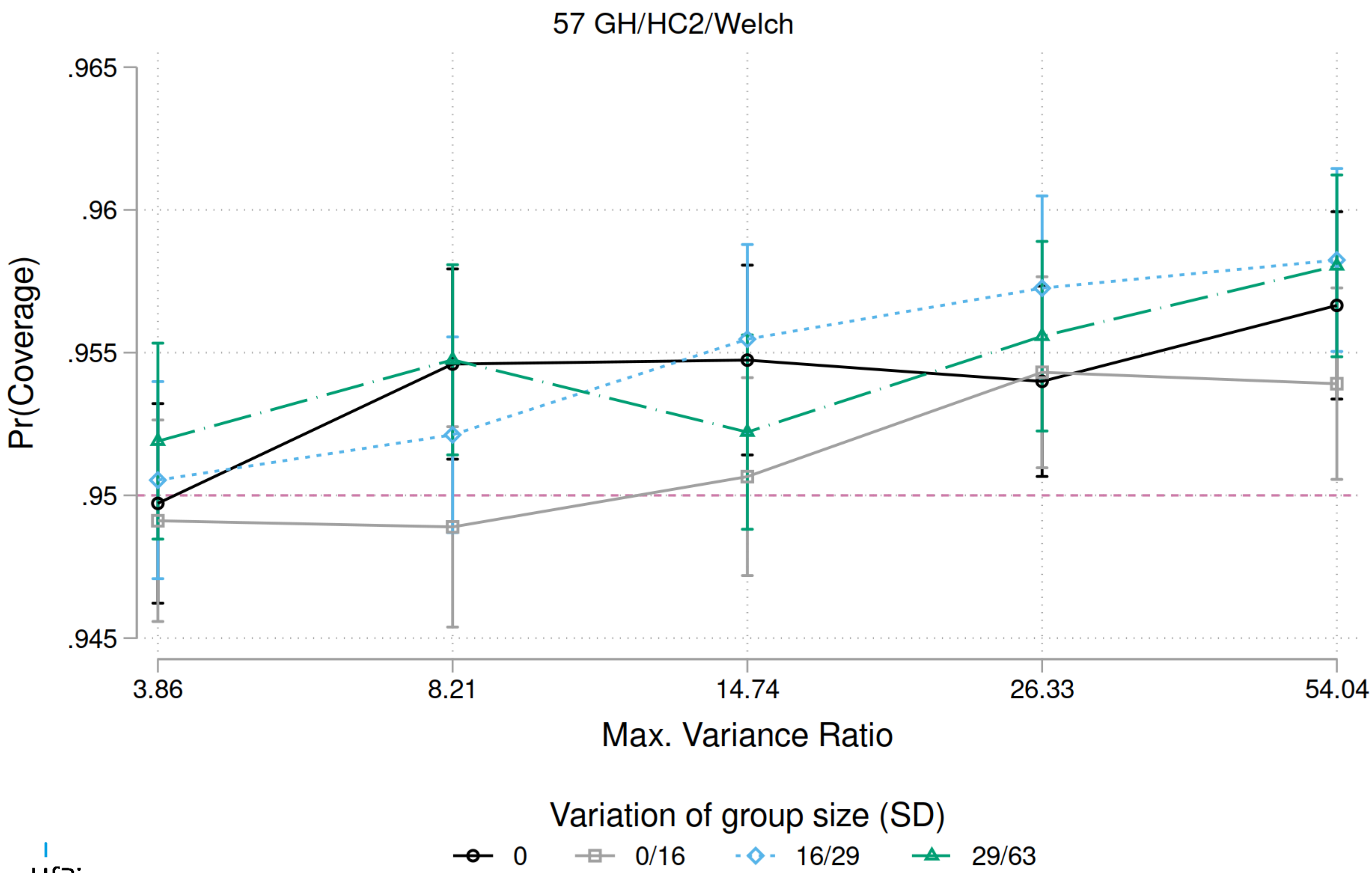
LIfBi LEIBNIZ-INSTITUT FÜR BILDUNGSVERLÄUFE

DZHW.

Backup (k=5)

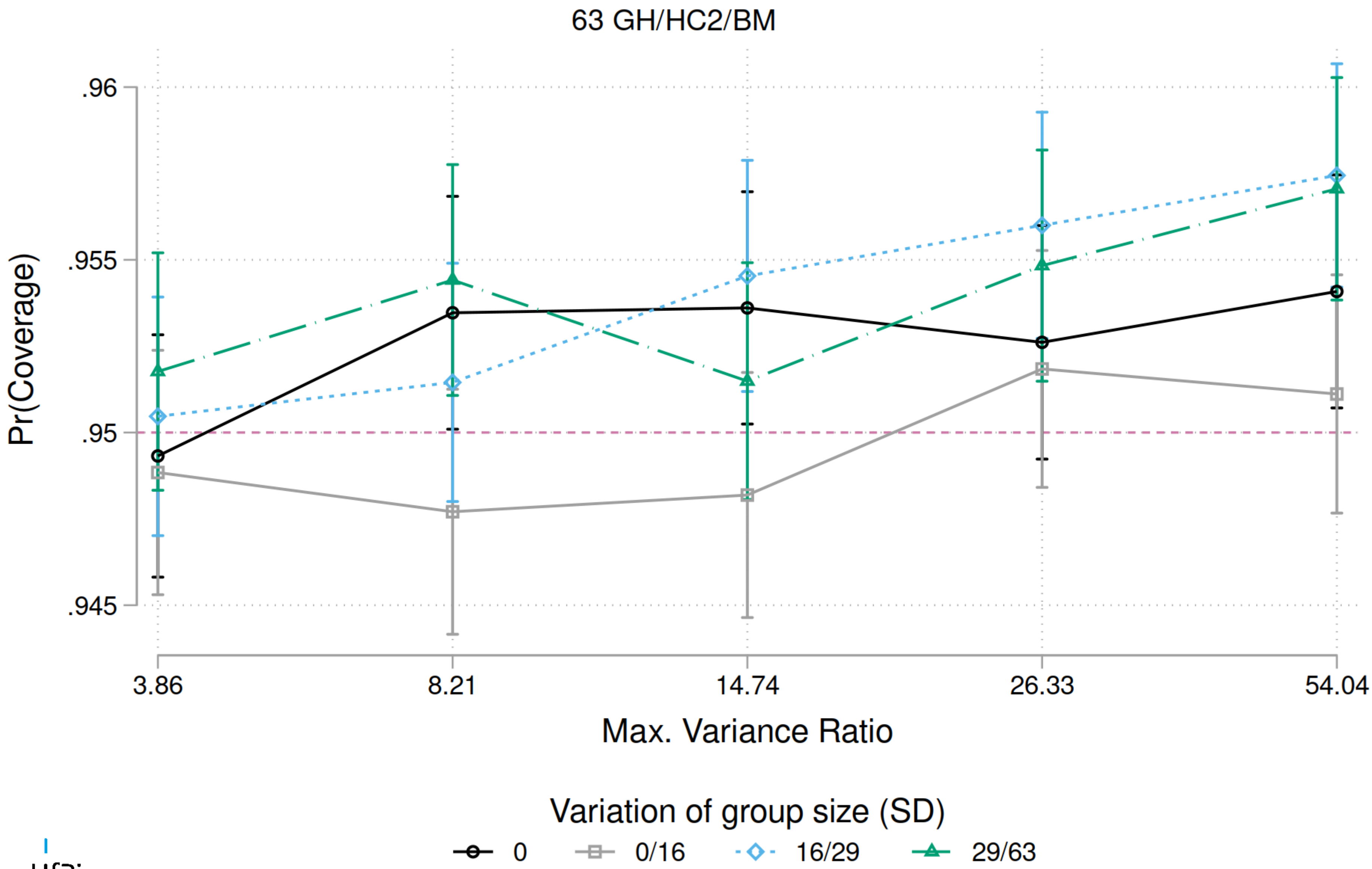

Scheffe
.98
.96
.94
.92
.9
Coverage
1.1—
5.8
5.8—
10.9
10.9—
19.3
19.3—
35.3
35.3—
99.9
Max. variance ratio
Standard deviation of group size (k=5)
0
<=16
16—29
29—62

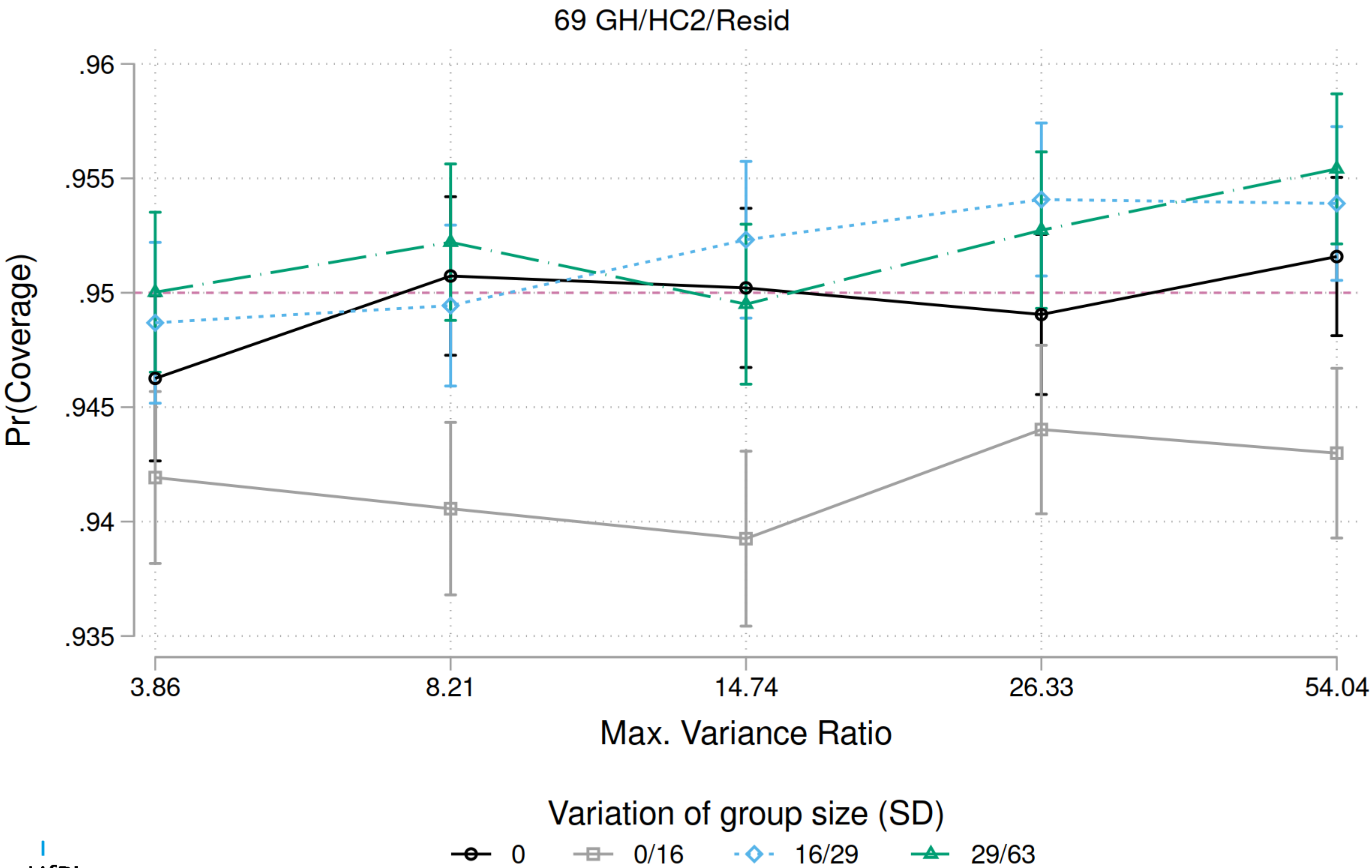

51 GH/HC2/Satter
Pr(Coverage)
.965
.96
.955
.95
.945
3.86
8.21
14.74
26.33
54.04
Max. Variance Ratio
Variation of group size (SD)
0
0/16
16/29
29/63



LIfBi LEIBNIZ-INSTITUT FÜR BILDUNGSVERLÄUFE

DZHW.


63 GH/HC2/BM
Pr(Coverage)
.96
.955
.95
.945
3.86
8.21
14.74
26.33
54.04
Max. Variance Ratio
Variation of group size (SD)
0
0/16
16/29
29/63


69 GH/HC2/Resid
Pr(Coverage)
.96
.955
.95
.945
.94
.935
3.86
8.21
14.74
26.33
54.04
Max. Variance Ratio
Variation of group size (SD)
0
0/16
16/29
29/63

Quick refresher on unadjusted t-test

Mean difference: $\hat{\delta} = \hat{\mu}_l - \hat{\mu}_m$

$100(1-\alpha)$ CI: $\hat{\delta} \pm c(\alpha)\,\widehat{se}(\hat{\delta})$

Critical value: $c(\alpha) = \quad {}_{inv}t_{\nu,\alpha/2}$

Degrees of freedom: $\nu = \sum_k (n_j - 1) = \sum_k n_j - k$

Standard error: $\widehat{se}(\hat{\delta}) = \sqrt{\frac{\sum_k s_j^2 \nu_j}{\nu}\left(\frac{1}{n_l} + \frac{1}{n_m}\right)}$

p-value: $p = 2t_{\nu,|t_0|}$

LIfBi LEIBNIZ-INSTITUT FÜR BILDUNGSVERLÄUFE

DZHW.