AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

Check for updates

The Stata Journal (2018)
18, Number 1, pp. 223–233

Assessing the health economic agreement of different data sources

Daniel Gallacher
Warwick Clinical Trials Unit
University of Warwick
Coventry, UK
d.gallacher@warwick.ac.uk

Felix Achana
Warwick Clinical Trials Unit
University of Warwick
Coventry, UK
F.Achana@warwick.ac.uk

Abstract. In this article, we present a simple-to-use framework for assessing the agreement of cost-effectiveness endpoints generated from different sources of data. The aim of this package is to enable the rapid assessment of routine data for use in cost-effectiveness analyses. By quantifying the comparability of routine data with "gold-standard" trial data, we inform decisions on the suitability of routine data for cost-effectiveness analysis. The rapid identification of informative routine data will increase the opportunity for economic analyses and potentially reduce the cost and burden of collecting patient-reported data in clinical trials.

Keywords: st0521, heabs, heapbs, health economic agreement, cost-effectiveness analysis

1 Introduction

With healthcare budgets under increasing scrutiny, the economic analysis of clinical decision making is of growing importance. This extends to the collection of evidence (Petrou and Gray 2011b,a). There is increasing pressure to gather information and make cost-effective decisions. Health economists routinely use data generated from clinical trials to assess the cost effectiveness of interventions. However, it can take months or years for patient follow-up to be completed. This delay, alongside the cost and burden placed on patients to complete lengthy study questionnaires, provides opportunities to consider alternative approaches.

There is increasing interest in the use of routine data to inform clinical decision making because of its potential for identifying cost-effective solutions rapidly and inexpensively (Gates et al. 2017; Raftery, Roderick, and Stevens 2005). However, the utility of routine data for this purpose remains uncertain. This issue can be informed by identifying the level of agreement between routine data and a "gold standard" such as existing trial data. While questions remain over what is an acceptable level of agreement, this article introduces a simple-to-use tool that quantifies the agreement between final economic endpoints generated using alternative sources of cost-effectiveness data. The routines implemented within this tool are suitable for use in a wide variety of decision-making contexts. For example, they can be used to compare and validate routine data for use in trial-based economic evaluations when alternative sources of information on costs and effects are available for trial participants.

© 2018 StataCorp LLC

Achana et al. (2018) introduced a methodological framework for the assessment of agreement using Lin's (1989) concordance correlation coefficient (CCC), the difference in incremental net benefit (INB) estimates, the probability of miscoverage (PMC), and the probability of cost effectiveness (PCE). Building on this work, we present generalizable commands allowing these analyses to be performed with relative ease. The commands are designed for use when data similar to those generated by a typical two-arm clinical trial are available.

Using individual patient data, the commands assist with the calculation of the incremental cost-effectiveness ratio (ICER) and INB. Briefly, the ICER is the ratio of incremental costs (that is, the difference in mean costs between the treatment and control interventions) to incremental effectiveness (that is, the corresponding difference in mean effectiveness between treatment and control interventions),

$$\text{ICER} = \frac{\text{Cost}_A - \text{Cost}_B}{\text{Effect}_A - \text{Effect}_B}$$

where Cost_A and Effect_A represent the means of the cost and effect in treatment A, and Cost_B and Effect_B represent the equivalent in treatment group B (where B is the control intervention). The ICER is normally the main summary measure of interest in most economic evaluations. However, as a ratio statistic, ICERs can be problematic to work with mathematically (Glick et al. 2007). The INB transforms comparisons of cost and effect instead to a linear scale and is given by

$$\text{INB} = \lambda(\text{Effect}_A - \text{Effect}_B) - (\text{Cost}_A - \text{Cost}_B)$$

where λ is the cost-effectiveness (or willingness-to-pay) threshold, which is the maximum threshold at which a decision maker is willing to pay per unit of effectiveness gained. INBs can be framed in net monetary terms (as given by the equation above) or net health terms, which are equivalent (see, for example, Glick et al. [2007] for further discussion of cost-effectiveness ratios and INBs).

In the applications that follow, the INB is used as the statistic for assessing agreement because of the mathematical convenience of working on a linear scale. If two sources of data are available, Lin's CCC of the two INB estimates will be calculated alongside the difference between the two INB estimates with a 95% confidence interval. Compatible with bootstrapping, the commands allow the calculation of the PMC and the PCE, and they produce a simple plot assessing the cost effectiveness.

The current version of the package allows only for assessment of two datasets where the comparison involves analysis of individual participant data on costs and effects (binary or continuous measures) such as that from simple two-arm randomized controlled trial data. Future development will focus on extending the routines to allow for i) randomized controlled trials with more complex designs (cluster-randomized, multiple-treatment comparisons, etc.); ii) inclusion of adjustment covariates in a regression so that routines can be applied to comparisons involving nonrandomized study designs; and iii) greater customization of the graphical output.

2 The commands

2.1 Description

The `heabs` command calculates the ICER and INB for up to two pairs of cost-effectiveness data. The command is flexible; if just one pair of cost and effect variables are entered, then only the ICER and INB will be calculated. However, if two sets of cost and effect variables are entered, then the command calculates the ICER and INB for both data sources and performs a simple comparison of the INB scores, calculating the CCC and the difference in INB estimates and relative confidence interval. In addition to cost and effect data, the function also requires the user to input the willingness-to-pay threshold and whether a higher score in the effect variable is beneficial (for example, quality of life) or detrimental to the health of the individual (for example, mortality). The command also requires the treatment indicator variable to be encoded as 0 and 1. `heabs` stores a range of calculated values, allowing for simple use alongside Stata's built-in `bootstrap` command.

Lin (1989) introduced his CCC as a means of quantifying agreement between two measures. It follows traditional correlation coefficients, with a score of 1 suggesting perfect agreement, −1 suggesting perfect inverse agreement, and 0 suggesting no agreement. McBride (2005) suggests that moderate agreement could be identified if the lower 95% confidence interval of the CCC estimate was above 0.90. However, Cicchetti (2001) suggests a score as low as 0.4 can be taken as a measure of fair agreement. With such a wide range of views, we recommend that the user decide his or her own suitable CCC threshold for determining agreement.

Similarly, the difference in INB has no firm interpretation in terms of its assessment of agreement. The standard errors (SEs) will often be large, reflecting the wide range of costs associated with a typical health economic analysis. Hence, it is unlikely that a significant difference will be found between the two sources of data. Again, we recommend that the user establish an acceptable level of difference between the two sources of data and examine the results of the bootstrap accordingly.

The `heapbs` command calculates the PMC and PCE for a set of cost-effectiveness data that have been generated through the `heabs` command combined with the `bootstrap` prefix. It can also produce a scatterplot of the cost-effectiveness data by fitting a confidence ellipse around them. The command is flexible, calculating only the scores specified by the user.

The PMC is recommended by Achana et al. (2018) to be implemented by taking the INB from the routine data source and the confidence intervals of the gold-standard trial data. This will then yield the probability that the confidence intervals do not contain the INB estimate, allowing for assessment of the agreement between the sets of data.

The PCE, while not a direct assessment of agreement, allows the user to assess what percentage of the bootstrapped INB estimates for a dataset are positive (that is, cost effective). A comparison of the PCE for both sets of cost-effectiveness data could yield further insight into the agreement of the two sources.

Finally, the optionally generated graph allows for the visual interpretation of a set of cost-effectiveness data through a scatterplot, complete with a 95% confidence ellipse. The ellipse is generated using an incorporated version of the command `ellip` (Alexandersson 2004). This command calculates the confidence ellipse assuming the costs and effects are elliptically distributed, drawing the ellipse with a twoway line. The fitted ellipse should contain roughly 95% of the scattered points.

2.2 Syntax

`heabs` *cost1 effect1* [*cost2 effect2*], `intervention(`*varname*`) response(`*string*`)`
[`w2p(`#`)`]

cost1 and *effect1* represent the cost and effect variables obtained from the first dataset.

cost2 and *effect2* are the variables from the second dataset, which are required if a comparison is to be performed. If the second pair of variables is not provided, the command performs a simple routine cost-effectiveness analysis. All of these variables must be numeric without any missing data.

`heapbs` [`, lci(`*varname*`) uci(`*varname*`) ref(`#`) inb(`*varname*`) draw`
`cost(`*varname*`) effect(`*varname*`)` *twoway_options*]

2.3 Options

`intervention(`*varname*`)` specifies the variable that indicates which treatment arm individuals are in. It requires that 0 and 1 be used to distinguish between the two treatment arms. `intervention()` is required.

`response(`*string*`)` specifies whether a higher effect score is positive (`bene`) or negative (`detr`). `response()` is required.

`w2p(`#`)` is the willingness-to-pay threshold, which is used in the calculation of the INB. This reflects how much the decision maker is willing to pay per unit of effect. The default is `w2p(0)`.

`lci(`*varname*`)` specifies the variable containing the bootstrapped estimates of the lower 95% confidence interval of the INB. This option is needed for PMC calculation.

`uci(`*varname*`)` specifies the variable containing the bootstrapped estimates of the upper 95% confidence interval of the INB. This option is needed for PMC calculation.

`ref(#)` is the reference INB to be used in the PMC. Achana et al. (2018) suggest using the INB from one dataset as the reference and comparing it with the 95% confidence intervals of the other dataset.

`inb(`*varname*`)` specifies the variable containing the bootstrapped estimates of the INB. This option is needed for PCE calculation.

`draw` specifies if the user would like to generate a plot of the cost-effectiveness bootstrapped data with a confidence ellipse.

`cost(`*varname*`)` specifies the variable containing the bootstrapped cost estimates. This option is needed to draw the plot.

`effect(`*varname*`)` specifies the variable containing the bootstrapped effect estimates. This option is needed to draw the plot.

twoway_options allow the control of title, legend, axis, and ellipse settings. See the `ellip` command for further details (Alexandersson 2004).

2.4 Stored results

`heabs` stores the following in `r()`:

Scalars

`r(cost1)`	incremental cost from the first set of cost-effectiveness data
`r(outcome1)`	incremental effect from the first set of cost-effectiveness data
`r(cost2)`	incremental cost from the second set of cost-effectiveness data
`r(outcome2)`	incremental effect from the second set of cost-effectiveness data
`r(NB1)`	INB from the first set of data
`r(seNB1)`	SE of the INB from the first set of data
`r(loCINB1)`	lower 95% confidence interval of the INB from the first set of data
`r(upCINB1)`	upper 95% confidence interval of the INB from the first set of data
`r(NB2)`	INB from the second set of data
`r(seNB2)`	SE of the INB from the second set of data
`r(loCINB2)`	lower 95% confidence interval of the INB from the second set of data
`r(upCINB2)`	upper 95% confidence interval of the INB from the second set of data
`r(diffNB)`	difference in the INB estimates of the two sets of data
`r(ICER1)`	ICER from the first set of data
`r(ICER2)`	ICER from the second set of data
`r(cccNB)`	CCC estimate of the two sources of data
`r(zcccNB)`	z score of CCC estimate; can be used in hypothesis testing

`heapbs` stores the following in `r()`:

Scalars

`r(pce)`	PCE
`r(pmc)`	PMC

3 Example

An example of the commands is shown through manipulation of the `bpwide.dta` built-in dataset, which can be reproduced using the do-file included in the package. Blood pressure is the outcome of interest, with a higher score associated with poorer health. Gender is recoded as the intervention indicator, and artificial cost data are created based loosely on the blood pressure data. The "before" variable is treated as the first gold-standard trial dataset, and the "after" variable is treated as the comparator. The full list of changes is shown through the commands below:

```
. sysuse bpwide
(fictional blood-pressure data)
. set seed 123
. _strip_labels _all
. drop agegrp
. generate cost1 = (200-bp_before)*rnormal(50,10)
. generate cost2 = (250-bp_after)*rnormal(50,10)
. rename bp_before bp1
. rename bp_after bp2
. rename sex intervention
```

Once the dataset is prepared, the functions can be applied as follows.

With no second source of data indicated, the command performs a simple routine cost-effectiveness analysis. The output displays a summary of the data, followed by estimates of the ICER, INB, and INB SE.

```
. heabs cost1 bp1, intervention(intervention) response(detr) w2p(10)
   Summary:            Int 0           Int 1
      N                60              60
      Min Cost         98.84935        667.84313
      Max Cost         3700.2055       3592.1422
      Min Effect       140             138
      Max Effect       185             185
```

	Cost	Effect	Inc Cost	Inc Effect	ICER
Int 0	2049.930	159.267	262.202	5.633	46.545
Int 1	2312.132	153.633			

	INB	INB SE
INB Results	-205.869	118.961

If a second set of data is added, then a comparison is performed with the routine analysis. Here the estimates for the first dataset are unchanged from above, but the corresponding estimates for the second data source are also displayed. In addition, the CCC estimate and the difference of the INB estimates are shown. Note that if a

willingness-to-pay threshold of £10 per unit decrease in blood pressure is used, both sources of data agree that the intervention is not cost effective, yielding negative INB scores. The CCC suggests very weak agreement between the two sources of data, and we can see the difference in the INB estimates is 110.85 units. However, the ICERs are very similar, with less than two units' difference.

```
. heabs cost1 bp1 cost2 bp2, intervention(intervention) response(detr) w2p(10)
Summary:              Data 1                          Data 2
               Int 0          Int 1          Int 0          Int 1
N              60             60             60             60
Min Cost       98.84935       667.84313      2947.6459      2947.6459
Max Cost       3700.2055      3592.1422      7765.9072      7765.9072
Min Effect     140            138            125            127
Max Effect     185            185            185            178
```

	ICER	INB	INB SE	CCC	Diff INB
DATA 1	46.545	-205.869	118.961	0.075	-110.853
DATA 2	48.083	-316.722	197.863		

The relationships between the CCC estimate and the INB estimate can be shown by changing the willingness-to-pay threshold. Below the threshold is increased to £100, and while the ICERs and INB estimates suggest that the treatment is now cost effective, the CCC is negative, and the difference between the INB estimates has increased. A negative CCC implies that the datasets are closer to drawing opposite conclusions than perfect agreement. While this appears strange given the apparent agreement of the INB estimates and ICERs, an investigation of the costs and blood pressure scores explains why. In figure 1, we see that while the data are paired, there is very little correlation between the two sources of data for both the costs and blood pressure scores. It just so happens that the populations agree. It is reasonable to expect a higher correlation between real paired individual-level data.

```
. heabs cost1 bp1 cost2 bp2, intervention(intervention) response(detr) w2p(100)
```
(*output omitted*)

	ICER	INB	INB SE	CCC	Diff INB
DATA 1	46.545	301.131	123.958	-0.041	130.647
DATA 2	48.083	431.778	208.976		

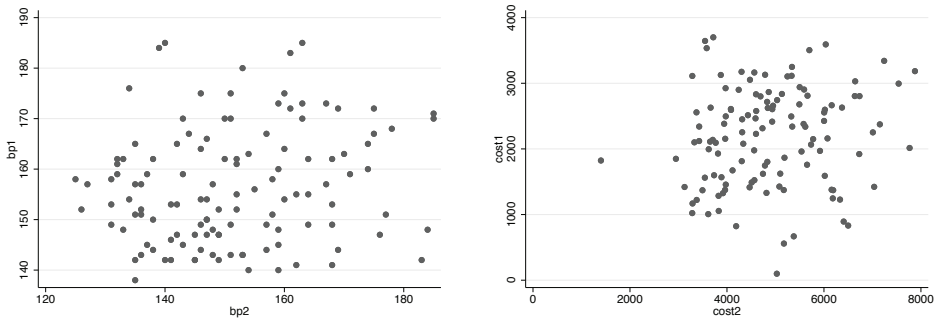

Figure 1. Scatterplots showing the relationship between the costs and blood pressure scores of the sources of data

The command can be implemented simply within the `bootstrap` prefix of Stata, as demonstrated below. The key variables used for the `heapbs` command are shown; however, additional outputs can be added.

```
. bootstrap cost1=r(cost1) cost2=r(cost2) effect1=r(outcome1)
> effect2=r(outcome2) NB1=r(NB1) NB2=r(NB2) NB1Lo=r(loCINB1) NB1Up=r(upCINB1)
> NB2Lo=r(loCINB2) NB2Up=r(upCINB2), saving(dummybpbs, replace) reps(100)
> seed(24): heabs cost1 bp1 cost2 bp2, w2p(10) intervention(intervention)
> response(detr)
```

(*output omitted*)

	Observed Coef.	Bootstrap Std. Err.	z	P>|z|	Normal-based [95% Conf.	Interval]
cost1	262.2022	136.2642	1.92	0.054	-4.870768	529.2751
cost2	399.8885	203.18	1.97	0.049	1.663059	798.1139
effect1	5.633333	1.824368	3.09	0.002	2.057638	9.209029
effect2	8.316667	2.338167	3.56	0.000	3.733945	12.89939
NB1	-205.8688	122.2152	-1.68	0.092	-445.4063	33.66862
NB2	-316.7218	190.0119	-1.67	0.096	-689.1382	55.69465
NB1Lo	-439.0318	120.8765	-3.63	0.000	-675.9454	-202.1182
NB1Up	27.29416	124.7917	0.22	0.827	-217.293	271.8813
NB2Lo	-704.5338	186.0972	-3.79	0.000	-1069.278	-339.79
NB2Up	71.09025	197.8324	0.36	0.719	-316.6542	458.8347

Once a bootstrapped dataset has been created, stored, and loaded into Stata, the `heabps` command can be applied. Here the second dataset is treated as our routine dataset, and the first dataset is treated as the gold-standard data. The text output from the command shows that the PMC is 21%, meaning that the INB estimate from the second source of data fails to appear within the 95% confidence interval from the bootstrapped dataset for the first source of data 21% of the time. The PCE estimate suggests that for the willingness-to-pay threshold selection during the `bootstrap` run, the drug is cost effective only 4% of the time.

```
. use dummybpbs.dta, clear
(bootstrap: heabs)
. heapbs, lci(NB1Lo) uci(NB1Up) ref(-316.722) inb(NB2) draw cost(cost2)
> effect(effect2) graphregion(color(white))
Probability of Miscoverage = 21%
Probability of Cost Effectiveness = 4%
```

Figure 2 shows the graphical output. Here the scatterplot for the second source of cost-effectiveness data depicts the intervention showing the treatment to be more effective and more expensive in the majority of `bootstrap` runs, with a 95% confidence ellipse and mean values clearly indicated.

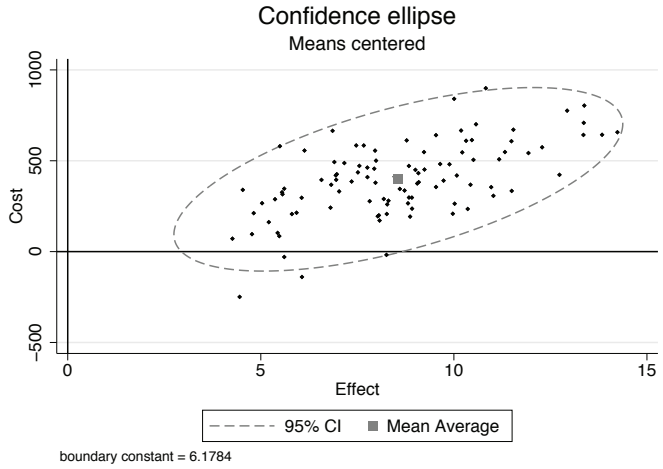

Figure 2. Plot of cost-effectiveness data for second source of data with 95% confidence ellipse

4 Conclusion

The `heabs` and `heapbs` commands described and demonstrated in this article are simple tools to aid with the evaluation of individual-level cost-effectiveness data. They also give users the opportunity to compare two sources of cost-effectiveness data with the aim of enabling more efficient clinical trial designs in the future. The flexibility of the commands allows users to calculate only the values they require and to customize the graphical output accordingly.

5 Acknowledgments

This work built on ideas from Felix Achana and Stavros Petrou (Warwick Clinical Trials Unit). Daniel Gallacher is funded by the National Institute of Health Research (Research Methods Fellowship).

In this article, we present independent research funded by the National Institute for Health Research. The views expressed are those of the authors and not necessarily those of the NHS, the National Institute for Health Research, or the Department of Health.

6 References

Achana, F., S. Petrou, K. Khan, A. Gaye, and N. Modi. 2018. A methodological framework for assessing agreement between cost-effectiveness outcomes estimated using alternative sources of data on treatment costs and effects for trial-based economic evaluations. *European Journal of Health Economics* 19: 75–86.

Alexandersson, A. 2004. Graphing confidence ellipses: An update of ellip for Stata 8. *Stata Journal* 4: 242–256.

Cicchetti, D. V. 2001. Methodological commentary: The precision of reliability and validity estimates re-visited: Distinguishing between clinical and statistical significance of sample size requirements. *Journal of Clinical and Experimental Neuropsychology* 23: 695–700.

Gates, S., R. Lall, T. Quinn, C. D. Deakin, M. W. Cooke, J. Horton, S. E. Lamb, A. M. Slowther, M. Woollard, A. Carson, M. Smyth, K. Wilson, G. Parcell, A. Rosser, R. Whitfield, A. Williams, R. Jones, H. Pocock, N. Brock, J. J. Black, J. Wright, K. Han, G. Shaw, L. Blair, J. Marti, C. Hulme, C. McCabe, S. Nikolova, Z. Ferreira, and G. D. Perkins. 2017. Prehospital randomised assessment of a mechanical compression device in out-of-hospital cardiac arrest (PARAMEDIC): A pragmatic, cluster randomised trial and economic evaluation. *Health Technology Assessment* 21: 1–176.

Glick, H. A., J. A. Doshi, S. S. Sonnad, and D. Polsky. 2007. *Economic Evaluation in Clinical Trials*. Oxford: Oxford University Press.

Lin, L. I.-K. 1989. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 45: 255–268.

McBride, G. B. 2005. A proposal for strength-of-agreement criteria for Lin's concordance correlation coefficient. NIWA Client Report HAM2005-062, National Institute of Water and Atmospheric Research. https://www.medcalc.org/download/pdf/McBride2005.pdf.

Petrou, S., and A. Gray. 2011a. Economic evaluation alongside randomised controlled trials: Design, conduct, analysis, and reporting. *British Medical Journal* 342: d1548.

———. 2011b. Economic evaluation using decision analytical modelling: Design, conduct, analysis, and reporting. *British Medical Journal* 342: d1766.

Raftery, J., P. Roderick, and A. Stevens. 2005. Potential use of routine databases in health technology assessment. *Health Technology Assessment* 9: 1–92.

About the authors

Daniel Gallacher is a research associate at the Warwick Clinical Trials Unit. His work includes providing statistical support on clinical trials and assisting with health technology appraisals.

Felix Achana is a research fellow at the Warwick Clinical Trials Unit. He provides health economic support on clinical trials and has an interest in economic evaluations on health technology appraisals.