Assessing the Health Economic Agreement of Different Data Sources.

ABSTRACT

A simple to use framework for assessing the agreement of cost-effectiveness endpoints generated from different sources of data is presented. The aim of this package is to enable the rapid assessment of routine data for utilisation in cost-effectiveness analyses. By quantifying the comparability of routine data with 'gold standard' trial data, decisions on the suitability of routine data for cost-effectiveness analysis are informed. 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.

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 (1, 2). There is increasing pressure to gather information and make decisions most cost-effectively. Health economists routinely use data generated from clinical trials to assess the cost-effectiveness of interventions, however it can take months or years for the 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, due to its potential for identifying cost-effective solutions rapidly and inexpensively (3, 4). 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 trials data. Whilst questions remain over what is an acceptable level of agreement, this paper introduces a simple to use tool which 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 source of information on costs and effects are available for trial participants.

Achana et al(5) introduced a methodological framework for the assessment of agreement, utilising Lin's Concordance Correlation Coefficient (CCC)(6), the difference in Incremental Net Benefit (INB) estimates, the Probability of Miscoverage (PMC) and the Probability of Cost Effectiveness (PCE). Building on this work, generalisable commands are presented 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 (i.e. difference in mean costs between the treatment and control interventions) to incremental effectiveness (i.e. the corresponding difference in mean effectiveness between treatment and control interventions):

$$ICER = \frac{Cost_A - Cost_B}{Effect_A - Effect_B}$$

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, but, as a ratio statistic, ICERs can be problematic to work with mathematically (7). The INB transforms comparisons of cost and effect instead to a linear scale and is given by:

$$INB = \lambda \left(Effect_A - Effect_B \right) - \left(Cost_A - Cost_B \right)$$

where λ is the cost-effectiveness (or willingness-to-pay) threshold 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 2010 (7) for further discussion of cost-effectiveness ratios and INBs).

In the applications that follow, the INB is used as the statistic for assessing agreement due to the mathematical convenience of working on 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 PM, the PCE and produce a simple plot assessing the cost effectiveness.

The current version of the package only allows 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 a simple two-arm randomised controlled trial data. Future development will focus on extending the routines to allow for i) RCTs with more complex designs (cluster-randomised, multiple-treatment comparisons, etc.), ii) inclusion of adjustment covariates in a regression so routines can be applied to comparisons involving non-randomised study designs and iii) allowing greater customisation 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, and 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 presented, 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/effect data, the function also requires the user to input the willingness-to-pay threshold and whether or not a higher score in the effect variable is beneficial (e.g. quality of life) or detrimental to the health of the individual (e.g. mortality). The command also requires the treatment indicator variable to be encoded 0 and 1. `heabs` stores a range of calculated values, allowing for simple use alongside Stata's built-in bootstrap command.

Lin introduced his CCC as a means of quantifying agreement between two measures(6). It follows traditional correlation coefficients, with a score of 1 suggesting perfect agreement, -1 suggesting

perfect inverse agreement and 0, no agreement. In 2005, McBride (8) suggested that moderate agreement could be identified if the lower 95% confidence interval of the CCC estimate was above 0.90. However, Cichetti (9) suggests a score as low as 0.4 can be taken as a measure of fair agreement. With such a wide range of views, it is recommended the user decide their own suitable CCC threshold for determining agreement.

Similarly, the difference in incremental net benefit has no firm interpretation in terms of its assessment of agreement. The standard errors 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, it is recommended 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 has been generated through the use of `heabs` command combined with the bootstrap function. It can also produce a scatter plot of the cost effectiveness data fitting a confidence ellipse around the data. The command is flexible, calculating only the scores specified by the user.

The PMC is recommended by Achana et al (5) to be implemented by taking the INB from the routine data source, and the confidence intervals of the gold-standard trials 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, whilst not a direct assessment of agreement, allows the user to assess what percentage of their bootstrapped INB estimates for a dataset are positive (i.e. 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 scatter plot, complete with a 95% confidence ellipse. The ellipse is generated using an incorporated version of the command `ellip` (10). This command calculates the confidence ellipse assuming the costs and effects are elliptically distributed, drawing the ellipse with `twoway line`. The fitted ellipse should contain roughly 95% of the scattered points.

2.2 SYNTAX

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

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

`cost2` and `effect2` are the variables from the second dataset, required if a comparison is to be performed. If the second pair of variables is not provided, the command. All of these variables must be numeric without any missing data.

`w2p` is the willingness to pay threshold, used in the calculation of the INB. This reflects how much the decision maker is willing to pay per unit of effect.

`intervention` is used to specify the variable which indicates which treatment arm individuals are in. It requires that 0 and 1 are used to distinguish between the two treatment arms.

`response` is used to specify whether a higher score in the effect variables should be interpreted positively (e.g. response to treatment) or negatively (e.g. death). The response variables can contain either binary (yes/no) data or be continuous measures.

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

`lci` is used to specify the variable containing the bootstrapped estimates of the lower 95% confidence interval of the INB. Needed for PMC calculation.

`uci` is used to specify the variable containing the bootstrapped estimates of the upper 95% confidence interval of the INB. Needed for PMC calculation.

`ref` is the reference INB to be used in the PMC. Achana et al (2017) (5) suggest using the INB from one dataset as the reference and comparing with the 95% confidence intervals of the other dataset.

`inb` is used to specify the variable containing the bootstrapped estimates of the INB. Needed for PCE calculation

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

`cost` is used to specify the variable containing the bootstrapped cost estimates. Needed to draw plot.

`effect` is used to specify the variable containing the bootstrapped effect estimates. Needed to draw plot.

twoway_options allow the control of titles, legends, axis and ellipse settings. See the `ellip` command for further details. (10)

2.3   SAVED RESULTS

`heabs` saves 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 saves in r():

Scalars

r(pce)	Probability of Cost Effectiveness
r(pmc)	Probability of Miscoverage

3 EXAMPLE

A quick 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 is created, loosely based on the blood pressure data. The 'before' variable is treated as the first gold standard trial data set, and the 'after' variable treated as the comparator. The full list of changes is shown through the commands below.

```
sysuse bpwide, clear
set seed 123
_strip_labels _all
drop agegrp
gen cost1 = (200-bp_before)*rnormal(50,10)
gen cost2 = (250-bp_after)*rnormal(50,10)
rename bp_before bp1
rename bp_after bp2
rename sex intervention
```

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

Firstly, 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, int(intervention) resp(detr) w2p(10)

Summary:           Int 0              Int 1
      N               60                 60
      Min Cost        98.84935           667.84314
      Max Cost        3700.2056          3592.1421
      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 additionally to the routine analysis. Here the estimates for the first data set 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 is 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 2 units difference.

```
. heabs cost1 bp1 cost2 bp2, int(intervention) resp(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       667.84314      2947.646       2947.646
   Max Cost      3700.2056      3592.1421      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 whilst 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. Whilst this appears strange, given the apparent agreement of the INB estimates and ICERs, an investigation of the cost and blood pressure scores explains why. In Figures 1 and 2, it can be seen that whilst the data is 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.

```
--------------------------------------------------------------------------+
```

	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		
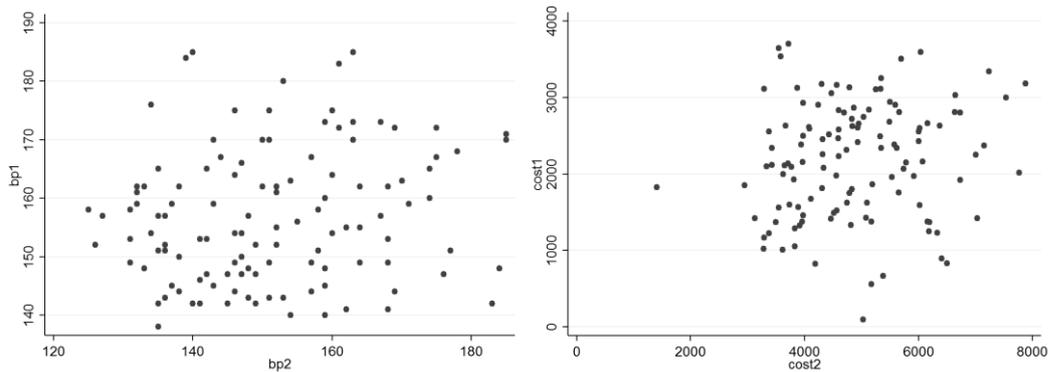
Figures 1 and 2: Scatter plots showing the relationship between the costs and blood pressure scores of the sources of data.

The command can be implemented simply within the bootstrap function 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) int(intervention) response(detr)
```

	Observed Coef.	Bootstrap Std. Err.	z	P>|z|	Normal-based [95% Conf. Interval]	
cost1	262.2021	136.2642	1.92	0.054	-4.870776	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 data set, and the first 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 only cost effective 4% of the time.

```
. heapbs, lci(NB1Lo) uci(NB1Up) ref(-316.722) inb(NB2) draw  cost(cost2)
effect(effect2)

Probability of Miscoverage = 21%
Probability of Cost Effectiveness = 4%
```

The graphical output is shown in Figure 3. Here, the scatter plot for the second source of cost effectiveness data shows 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.
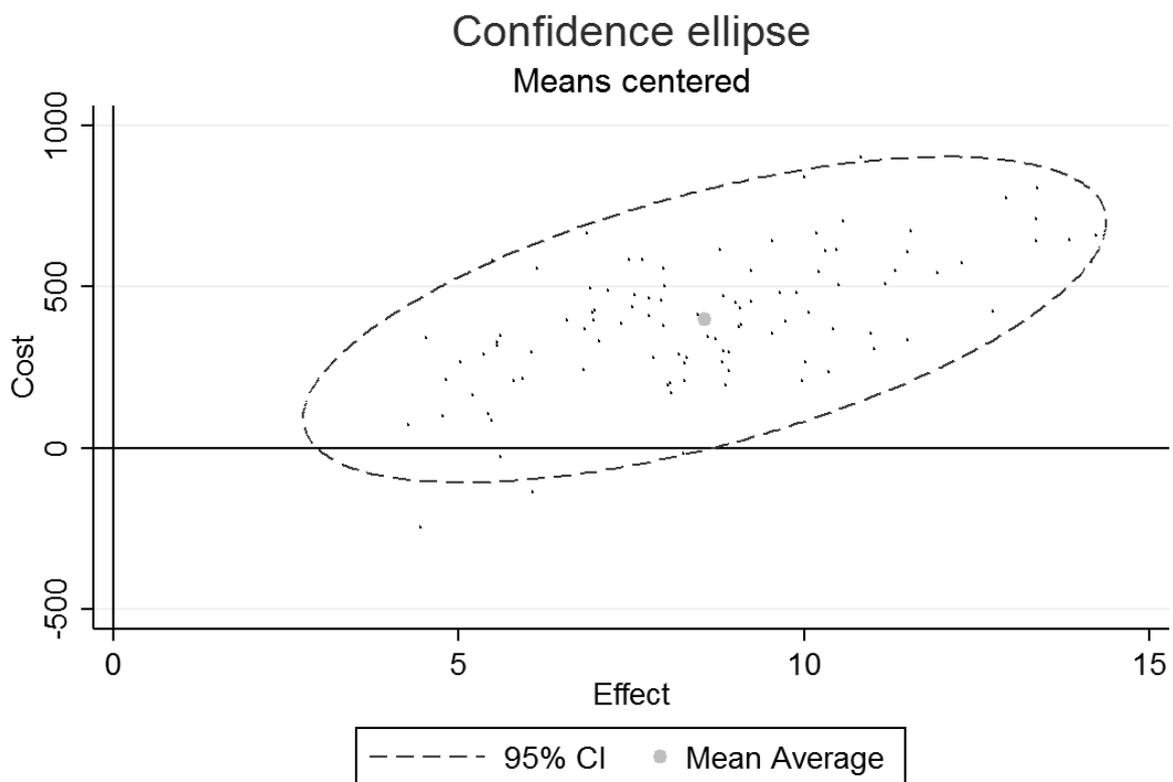

Figure 3: Plot of cost-effectiveness data for second source of data with 95% confidence ellipse.

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 the user 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 allow the user to only calculate the values they require, and to customise the graphical output accordingly.

5 ACKNOWLEDGEMENTS

6 REFERENCES

1.   Petrou S, Gray A. Economic evaluation using decision analytical modelling: design, conduct, analysis, and reporting. BMJ. 2011;342.
2.   Petrou S, Gray A. Economic evaluation alongside randomised controlled trials: design, conduct, analysis, and reporting. BMJ. 2011;342.
3.   Gates S, Lall RS, Quinn T, Deakin CD, Cooke M, Horton J, et al. 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. 2017;21(11):1-176.
4.   Raftery J, Roderick P, Stevens A. Potential use of routine databases in health technology assessment. Health Technology Assessment. 2005;9(20):1-106.
5.   Achana F, Petrou S, Khan K, Gaye A, Modi N. 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. The European Journal of Health Economics. 2017:1-12.
6.   Lawrence IKL. A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometrics. 1989;45(1):255-68.
7.   Glick HA, Doshi JA, Sonnad SS, Polsky D. Economic evaluation in clinical trials. 2014.
8.   McBride GB. A Proposal for strength-of-agreement criteria for Lin's concordance correlation coefficient. NIWA Client Report: HAM2005-062, 2005.
9.   Cicchetti DV. 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. 2001;23(5):695-700.
10.   Alexandersson A. Graphing confidence ellipses: An update of ellip for Stata 8. Stata Journal. 2004;4(3):242-56.

About the authors

Daniel Gallacher is a Research Associate at 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 Warwick Clinical Trials Unit. He provides health economic support on clinical trials and has an interest in economic evaluations on health technology appraisals.