

Multi-criteria drug benefit-risk assessment through mixed treatment comparisons

Gert van Valkenhoef (1,2), Tommi Tervonen (1),
Bert de Brock (1), Hans Hillege (2), Douwe Postmus (2)

- (1) Faculty of Economics and Business, University of Groningen (NL)
(2) Department of Epidemiology, University Medical Center Groningen (NL)

24th European Conference on Operational Research
July 2010, Lisbon



umcg



university of
 groningen

Introduction

- Pharmaceutical decision making
 - based on assessing benefits and risks of alternative drugs
 - ideally by considering all available clinical evidence
- Given outcome of clinical trials, should a new anti-depressant be allowed on the market?
- Which anti-depressant is most suited for severely depressed patients?



Problem

- Current statistical methods insufficient to get integrated overview of the alternatives and criteria:
 - Meta-analysis (evidence pooling) on single criterion
 - Only pair-wise comparisons.
- BR analysis is unstructured
 - No pre-specified criteria or models
- BR analysis is non-transparent
 - Evidence basis not (sufficiently) explicit
 - Measurements and value judgments not separated



Solution

- Multi-Criteria Decision Analysis (MCDA) methods allow
 - to evaluate multiple alternatives
 - in terms of multiple criteria
- Mixed Treatment Comparison (MTC) models enable
 - indirect comparisons
 - between ≥ 2 alternative treatments
- We put MCDA together with MTC to
 - systematically assess the (relative) benefits and risks
 - of any number of alternative treatments
 - on the relevant criteria
 - take into account + quantify uncertainty
 - explicitly based on clinical evidence

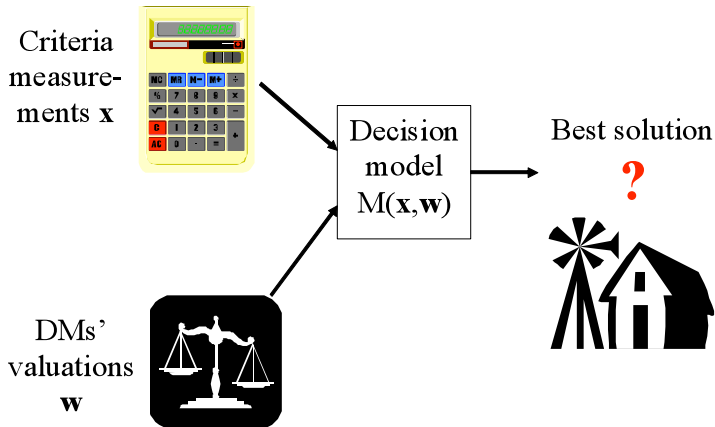


Stochastic Multi-criteria Acceptability Analysis (SMAA)

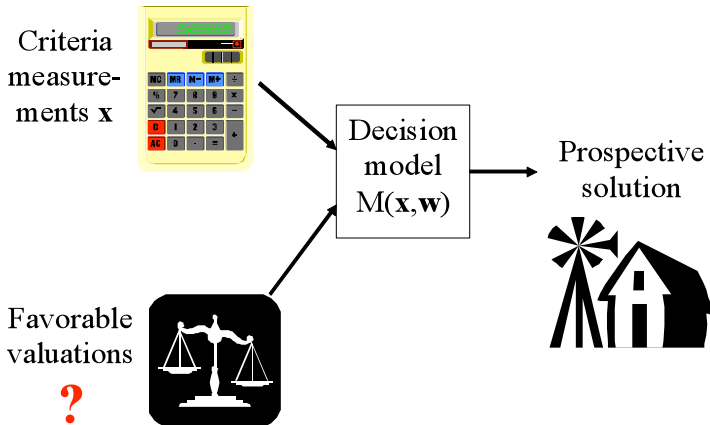
- SMAA is a multi-criteria decision aiding (MCDA) method for ranking
 - a set of m alternatives
 - based on a set of n criteria
- Evaluation of alternative x on criterion y
 - may be uncertain: specify a probability distribution
- Preference information:
 - a weight vector (optional) and
 - a value function (usually linear)
- SMAA is based on Multi-Attribute Utility Theory (MAUT)



SMAA: forward / inverse approach



SMAA: forward / inverse approach



SMAA decision aiding metrics

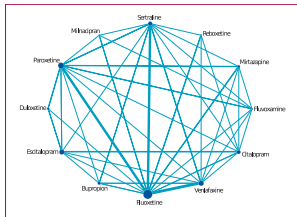
Rank acceptability index share of weights and measurements making an alternative have ranks $1, \dots, m$ (most preferred, second most, etc.).

Central weight vector center of gravity of the favourable weight space: “Which preferences support an alternative to be the most preferred one?”

Confidence factor probability for an alternative to be preferred when preferences equal its central weight vector: “Are the measurements are sufficiently precise?”

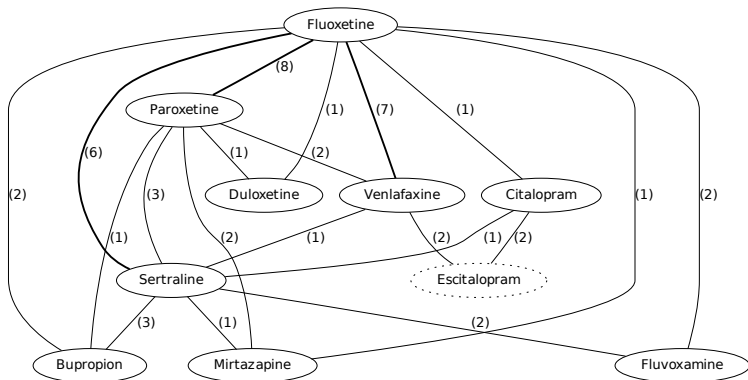


Mixed Treatment Comparison (MTC)



- A.K.A. network meta-analysis
- MTC is an extension of normal meta-analysis
- MTC allows comparison of ≥ 2 alternatives
 - Integrating direct and indirect evidence
 - While checking for (in-)consistencies

MTC: example evidence network



MTC: example results (consistency model)

Sertr	1.35 (0.83, 2.20)	1.40 (0.84, 2.32)	0.80 (0.65, 0.98)	0.81 (0.42, 1.58)	0.93 (0.52, 1.65)	1.08 (0.78, 1.49)	0.93 (0.70, 1.24)	0.92 (0.68, 1.23)	1.09 (0.83, 1.44)
0.74 (0.45, 1.21)	Escit	1.04 (0.51, 2.11)	0.59 (0.37, 0.94)	0.60 (0.28, 1.28)	0.69 (0.34, 1.39)	0.80 (0.47, 1.36)	0.69 (0.42, 1.13)	0.68 (0.39, 1.18)	0.81 (0.54, 1.22)
0.72 (0.43, 1.19)	0.97 (0.47, 1.97)	Cital	0.57 (0.33, 0.99)	0.58 (0.25, 1.34)	0.66 (0.31, 1.43)	0.77 (0.42, 1.41)	0.67 (0.37, 1.20)	0.66 (0.36, 1.18)	0.78 (0.44, 1.40)
1.25 (1.02, 1.53)	1.68 (1.06, 2.68)	1.74 (1.01, 3.01)	Fluox	1.01 (0.54, 1.90)	1.16 (0.67, 2.01)	1.35 (0.98, 1.85)	1.16 (0.92, 1.48)	1.14 (0.85, 1.54)	1.37 (1.10, 1.69)
1.23 (0.63, 2.39)	1.66 (0.78, 3.52)	1.72 (0.75, 3.97)	0.99 (0.53, 1.85)	Fluvo	1.14 (0.50, 2.64)	1.33 (0.65, 2.70)	1.15 (0.59, 2.25)	1.13 (0.57, 2.25)	1.35 (0.70, 2.60)
1.08 (0.61, 1.91)	1.45 (0.72, 2.93)	1.50 (0.70, 3.24)	0.86 (0.50, 1.50)	0.87 (0.38, 2.02)	Dulox	1.16 (0.63, 2.13)	1.01 (0.59, 1.72)	0.99 (0.54, 1.80)	1.18 (0.66, 2.10)
0.93 (0.67, 1.28)	1.25 (0.74, 2.12)	1.30 (0.71, 2.36)	0.74 (0.54, 1.02)	0.75 (0.37, 1.53)	0.86 (0.47, 1.58)	Mirta	0.87 (0.64, 1.17)	0.85 (0.56, 1.29)	1.01 (0.72, 1.43)
1.07 (0.81, 1.42)	1.44 (0.88, 2.37)	1.50 (0.83, 2.69)	0.86 (0.67, 1.09)	0.87 (0.44, 1.70)	0.99 (0.58, 1.71)	1.16 (0.85, 1.56)	Parox	0.98 (0.69, 1.40)	1.17 (0.89, 1.55)
1.09 (0.81, 1.47)	1.47 (0.84, 2.56)	1.52 (0.84, 2.75)	0.87 (0.65, 1.18)	0.89 (0.44, 1.76)	1.01 (0.56, 1.85)	1.18 (0.77, 1.79)	1.02 (0.72, 1.45)	Bupro	1.19 (0.83, 1.71)
0.91 (0.69, 1.20)	1.23 (0.82, 1.85)	1.28 (0.71, 2.29)	0.73 (0.59, 0.91)	0.74 (0.39, 1.43)	0.85 (0.48, 1.52)	0.99 (0.70, 1.39)	0.85 (0.65, 1.13)	0.84 (0.58, 1.21)	Venla

- OR (95% CI): $\frac{\text{odds}(\text{row})}{\text{odds}(\text{col})}$.
- OR > 1: higher response-rate for the column-treatment.



MTC: example results (consistency model)

Fluox	1.17 (0.89, 1.54)	1.27 (0.99, 1.62)	1.38 (1.11, 1.71)
0.85 (0.65, 1.12)	Parox	1.08 (0.76, 1.53)	1.17 (0.87, 1.59)
0.79 (0.62, 1.01)	0.93 (0.66, 1.31)	Sertr	1.09 (0.80, 1.49)
0.73 (0.59, 0.90)	0.85 (0.63, 1.15)	0.92 (0.67, 1.26)	Venla

- OR (95% CI): $\frac{\text{odds}(\text{row})}{\text{odds}(\text{col})}$.
- OR > 1: higher response-rate for the column-treatment.

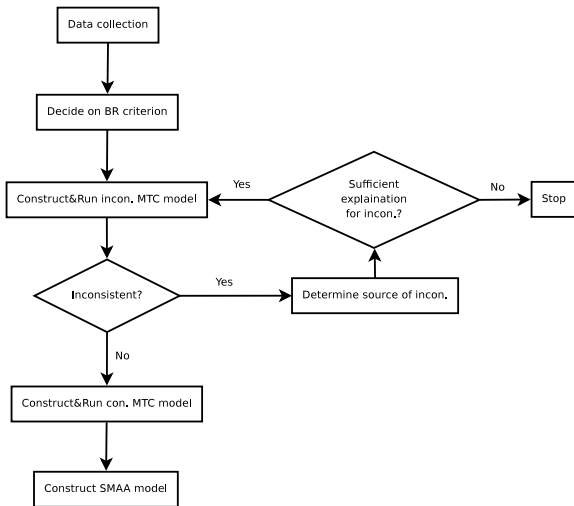


MTC-based SMAA for Benefit-Risk Assessment

- m alternative treatments are evaluated with respect to efficacy and $n - 1$ most important ADRs
- All measurements (efficacy and ADRs) are log-odds ratios compared with fluoxetine
 - Assumed normal distribution
 - Derived through mixed treatment comparison
 - Robust to choice of baseline (here: fluoxetine)
- Implemented in ADDIS (aggregate data drug information system) v0.10:
 - <http://drugis.org/addis>
 - open source software



MTC-based SMAA for Benefit-Risk Assessment



Why use MTC?

An earlier model used pair-wise meta-analysis. Problems:

- Naive pooling of ADRs not robust, leads to too precise measurements
- Choice of common comparator has unknown influence on model
 - Selection bias: arbitrary exclusion of evidence
 - Sensitivity analysis with different comparators?
- Only applicable when common comparator available
 - Not the case in many clinical domains

And we want to offer an automated solution!



ADDIS: mixed treatment comparison (efficacy)

The screenshot displays the ADDIS v0.9-SNAPSHOT software interface. The top menu bar includes File, Edit, Add, and Help. Below the menu, there are tabs for 'Add study', 'Create meta-analysis', 'Create network meta-analysis', and 'Create benefit-risk analysis'. A list of studies is shown in a table format, with columns for author/year, study description, number of patients, and a severity score. Below the table, there is a section for 'Evidence network' showing a network diagram with four nodes: Fluoxetine, Paroxetine, Sertraline, and Venlafaxine. The nodes are connected by lines representing comparisons, with numbers indicating the number of studies for each comparison: Fluoxetine to Paroxetine (5), Fluoxetine to Sertraline (5), Fluoxetine to Venlafaxine (6), Paroxetine to Sertraline (1), Paroxetine to Venlafaxine (2), and Sertraline to Venlafaxine (1). Below the network diagram, there is a section for 'Results - network inconsistency model' showing a progress bar for 'Simulating: 51%'. A yellow box at the bottom contains text explaining network inconsistency: 'In network meta-analysis, because of the more complex evidence structure, we can assess inconsistency of evidence, in addition to heterogeneity within a comparison. Whereas heterogeneity represents between-study variation in the measured relative effect of a pair of treatments, inconsistency can only occur when a treatment C has a different effect when it is compared with A or B (i.e., studies comparing

Author/Year	Study Description	Patients	Severity Score
McPartlin et al. 1998	A comparison of onc...	361	310497006 Severe d
Mehtonen et al. 2000	Randomized, double...	147	310497006 Severe d
Newhouse et al. 2000	A double-blind compa...	236	310497006 Severe d
Rudolph and Feiger, ...	A double-blind, rando...	203	310497006 Severe d
Schone and Ludwig, ...	A double-blind study ...	106	310497006 Severe d
Sechter et al. 1999	A double-blind compa...	238	310497006 Severe d
Silverstone and Ravin...	Once-daily venlafaxin...	249	310497006 Severe d
Tyee et al. 1997	A double-blind, rando...	341	310497006 Severe d



ADDIS: mixed treatment comparison (efficacy)

ADDIS v0.9-SNAPSHOT

File Edit Add Help

Add study Create meta-analysis Create network meta-analysis Create benefit-risk analysis www.drugs.org

Constitutional
Depersonalization
Diarrhea
Diminished Sexual Desire
Dizziness
Dream Abnormality
Dry Mouth
Dysmenorrhea
Dyspepsia
Emotional Indifference
Erectile Dysfunction
Failing Memory
Fatigue
Flatulence
Flu Syndrome
Gastralgia
Headache
Impotence
Increased Duration Of Sleep
Increased Salivation
Increased Sexual Desire
Infection
Insomnia
Micturition Disturbances
Nausea
Nervousness
Orgasmic Dysfunction
Orgiastic Dysfunction
Pain
Palpitation
Pharyngitis
Rash
Reduced Salivation
Reduced Sleep
Respiratory Disorder
Rhinitis
Sexual Dysfunction
Somnolence
Sweating
Tinnitus
Tremor
Vertigo

1
Venlafaxine

Results - network inconsistency model

In network meta-analysis, because of the more complex evidence structure, we can assess inconsistency of evidence, in addition to heterogeneity within a comparison. Whereas heterogeneity represents between-study variation in the measured relative effect of a pair of treatments, inconsistency can only occur when a treatment C has a different effect when it is compared with A or B (i.e., studies comparing A and C are systematically different from studies comparing B and C). Thus, inconsistency may even occur with normal meta-analysis, but can only be detected using a network meta-analysis, and then only when there are closed loops in the evidence structure. For more information about assessing inconsistency, see G. Lu and A. E. Ades (2006). *Assessing evidence inconsistency in mixed treatment comparisons*. Journal of the American Statistical Association, 101(474): 447-459. doi:10.1198/016214505000001302.

Network Meta-Analysis (Inconsistency Model)

Fluoxetine	1.22 (0.91, 1.64)	1.26 (0.97, 1.64)	1.33 (1.07, 1.67)
0.82 (0.61, 1.09)	Paroxetine	1.15 (0.75, 1.77)	1.21 (0.86, 1.71)
0.79 (0.61, 1.03)	0.87 (0.56, 1.33)	Sertraline	1.16 (0.69, 1.95)
0.75 (0.60, 0.94)	0.82 (0.58, 1.16)	0.86 (0.51, 1.45)	Venlafaxine

Inconsistency Factors

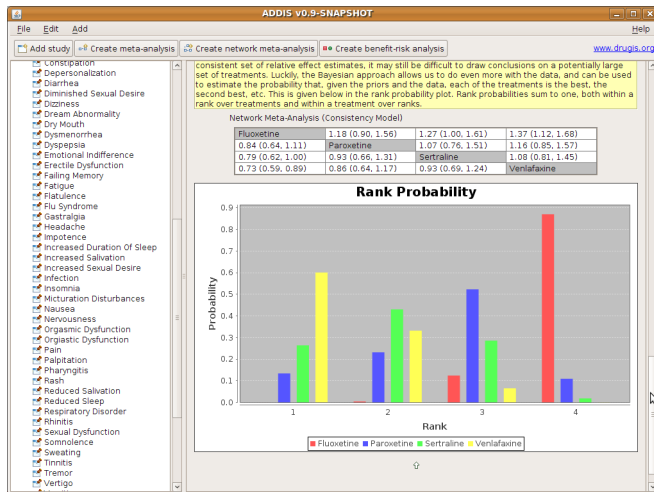
Cycle	Confidence Interval
Fluoxetine, Paroxetine, Venlafaxine	0.11 (-0.26, 0.48)
Fluoxetine, Sertraline, Paroxetine, Venlafaxine	-0.00 (-0.39, 0.38)
Paroxetine, Sertraline, Venlafaxine	-0.10 (-0.60, 0.41)

Results - network consistency model

If there is no relevant inconsistency in the evidence, a consistency model can be used to draw conclusions about the relative effect of the included treatments. Using normal meta-analysis, we could only get a subset of the confidence intervals for relative effects we derive using network meta-analysis. Network meta-analysis gives a consistent, integrated picture of the relative effects. However, given such a consistent set of relative effect estimates, it may still be difficult to draw conclusions on a potentially large set of treatments. Luckily, the Bayesian approach allows us to do even more with the data, and can be used to estimate the probability that, given the priors and the data, each of the treatments is the best, the



ADDIS: mixed treatment comparison (efficacy)



ADDIS: benefit-risk model

ADDIS v0.10

File Edit Add Help

www.drugs.org

Increased Duration Of Sle...
 Increased Salivation
 Increased Sexual Desire
 Infection
 Insomnia
 Micturition Disturbances
 Nausea
 Nervousness
 Orgasmic Dysfunction
 Orgiastic Dysfunction
 Pain
 Palpitation
 Pharyngitis
 Rash
 Reduced Salivation
 Reduced Sleep
 Respiratory Disorder
 Rhinitis
 Sexual Dysfunction
 Somnolence
 Sweating
 Tinnitus
 Tremor
 Vertigo
 Vomiting
 Weight Gain
 Weight Loss

Population characteristics

Studies

Analyses

- Hansen Diarrhea
- Hansen Dizziness
- Hansen Fluox-Venla HAM-D
- Hansen HAM-D
- Hansen Headache
- Hansen Insomnia
- Hansen Nausea

Benefit-risk analysis

- Fast BR
- Hansen BR

Benefit-Risk Analysis

ID: Hansen BR

Indication: 310497006 Severe depression

Criteria: [HAM-D Responders, Diarrhea, Dizziness, Headache, Insomnia, Nausea]

Alternatives: [Fluoxetine, Paroxetine, Sertraline, Venlafaxine]

Included Analyses

Name	Type	Indication	Outcome Measure	Drugs	Studies	Sample ...
Hansen Nausea	Markov Chain Monte ...	310497006 Severe d...	Nausea			3557
Hansen Diarrhea	Markov Chain Monte ...	310497006 Severe d...	Diarrhea			2932
Hansen HAM-D	Markov Chain Monte ...	310497006 Severe d...	HAM-D Responders			3726
Hansen Insomnia	Markov Chain Monte ...	310497006 Severe d...	Insomnia			2399
Hansen Dizziness	Markov Chain Monte ...	310497006 Severe d...	Dizziness			2653
Hansen Headache	Markov Chain Monte ...	310497006 Severe d...	Headache			3189

Measurements

Relative measurements: odds ratio or mean difference, with Fluoxetine as the common comparator.

Alternative	HAM-D Responders	Diarrhea	Dizziness	Headache	Insomnia	Nausea
Fluoxetine	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)
Paroxetine	1.17 (0.90, 1.52)	0.65 (0.42, 1.01)	1.74 (0.95, 3.19)	0.84 (0.47, 1.52)	0.95 (0.60, 1.50)	1.27 (0.93, 1.74)
Sertraline	1.27 (1.00, 1.61)	1.55 (1.00, 2.40)	0.72 (0.39, 1.32)	1.29 (0.71, 2.34)	1.14 (0.69, 1.89)	1.14 (0.81, 1.59)
Venlafaxine	1.38 (1.12, 1.71)	0.59 (0.33, 1.07)	2.93 (1.81, 4.72)	0.65 (0.35, 1.22)	1.09 (0.68, 1.73)	1.75 (1.35, 2.27)

Absolute measurements: odds or mean calculated from the assumed odds or mean for Fluoxetine. The method used to derive the assumed odds or mean are heuristic, and the absolute values should be interpreted with care.

Alternative	HAM-D Responders	Diarrhea	Dizziness	Headache	Insomnia	Nausea
Fluoxetine	1.14 (0.81, 1.62)	0.11 (0.05, 0.22)	0.08 (0.04, 0.14)	0.22 (0.12, 0.41)	0.16 (0.09, 0.27)	0.26 (0.21, 0.32)
Paroxetine	1.34 (0.87, 2.07)	0.07 (0.03, 0.17)	0.13 (0.06, 0.31)	0.18 (0.08, 0.43)	0.15 (0.07, 0.30)	0.33 (0.22, 0.46)

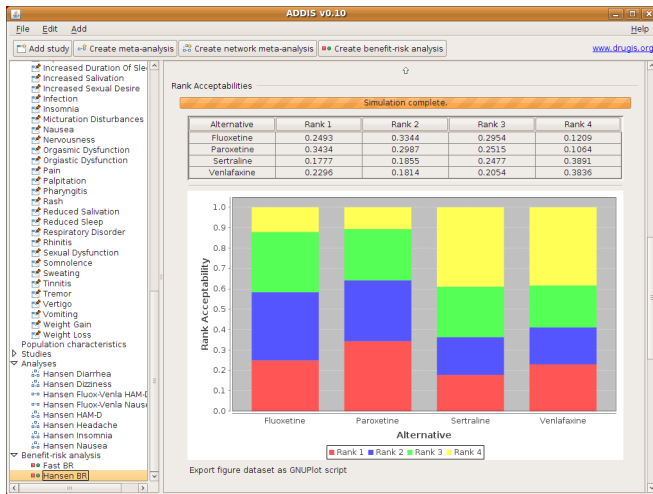


Benefit-risk scenarios

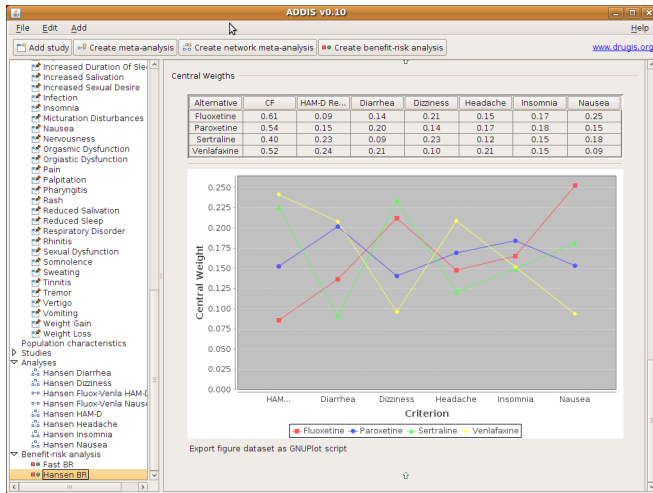
- We considered 3 scenarios:
 - ① Health policy decision making with no preferences
 - ② Prescription for mild depression
 - ③ Prescription for severe depression
- Ordinal swing weighting for prescription decisions



ADDIS: benefit-risk (no preferences)



ADDIS: benefit-risk (no preferences)



ADDIS: benefit-risk (severe depression)

ADDIS v0.10

File Edit Add Help

Add study Create meta-analysis Create network meta-analysis Create benefit-risk analysis www.drugis.org

Alternative	HAM-D Responders	Diarrhea	Dizziness	Headache	Insomnia
Fluoxetine	1.14 (0.81, 1.62)	0.11 (0.05, 0.22)	0.08 (0.04, 0.14)	0.22 (0.12, 0.41)	0.16 (0.09, 0.28)
Paroxetine	1.34 (0.87, 2.07)	0.07 (0.03, 0.17)	0.13 (0.06, 0.31)	0.18 (0.08, 0.43)	0.15 (0.07, 0.28)
Sertraline	1.45 (0.95, 2.21)	0.17 (0.07, 0.39)	0.06 (0.02, 0.13)	0.28 (0.12, 0.66)	0.18 (0.09, 0.31)
Venlafaxine	1.58 (1.05, 2.37)	0.07 (0.03, 0.16)	0.23 (0.11, 0.48)	0.14 (0.06, 0.34)	0.17 (0.08, 0.31)

Preferences

ORDINAL Preference information

Criterion	Scale	Rank
HAM-D Responders	OR: [0.90 - 1.71] Risk: [0.51 - 0.66] RD: 0.16 NNH 6.44	1
Diarrhea	OR: [0.33 - 2.40] Risk: [0.03 - 0.21] RD: 0.17 NNH 5.74	2
Dizziness	OR: [0.39 - 4.72] Risk: [0.03 - 0.27] RD: 0.24 NNH 4.20	4
Headache	OR: [0.35 - 2.34] Risk: [0.07 - 0.34] RD: 0.27 NNH 3.76	6
Insomnia	OR: [0.60 - 1.89] Risk: [0.09 - 0.23] RD: 0.14 NNH 7.03	5
Nausea	OR: [0.81 - 2.27] Risk: [0.17 - 0.37] RD: 0.20 NNH 5.10	3

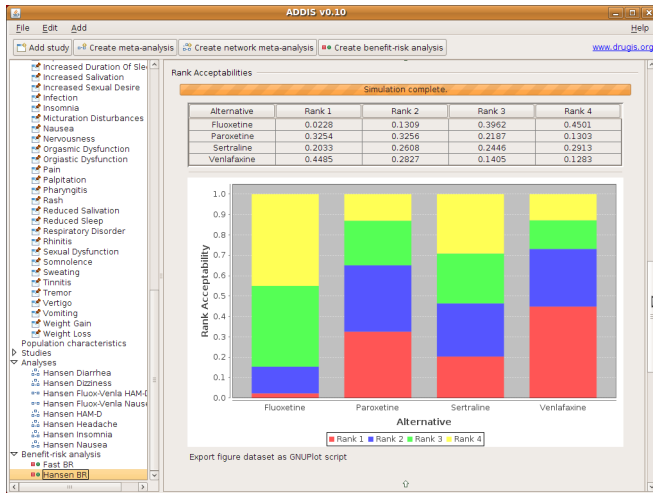
Rank Acceptabilities

Simulation complete.

Alternative	Rank 1	Rank 2	Rank 3	Rank 4
Fluoxetine	0.0228	0.1309	0.3962	0.4501
Paroxetine	0.3254	0.3256	0.2187	0.1303
Sertraline	0.2033	0.2608	0.2446	0.2913
Venlafaxine	0.4485	0.2827	0.1405	0.1283



ADDIS: benefit-risk (severe depression)



ADDIS: benefit-risk (mild depression)

ADDIS v0.10

File Edit Add Help

Add study
 Create meta-analysis
 Create network meta-analysis
 Create benefit-risk analysis
 www.drugis.org

Alternative	HAM-D Responders	Diarrhea	Dizziness	Headache	Insomnia
Fluoxetine	1.14 (0.81, 1.62)	0.11 (0.05, 0.22)	0.08 (0.04, 0.14)	0.22 (0.12, 0.41)	0.16 (0.09, 0.28)
Paroxetine	1.34 (0.87, 2.07)	0.07 (0.03, 0.17)	0.13 (0.06, 0.31)	0.18 (0.08, 0.43)	0.15 (0.07, 0.28)
Sertraline	1.45 (0.95, 2.21)	0.17 (0.07, 0.39)	0.06 (0.02, 0.13)	0.28 (0.12, 0.66)	0.18 (0.09, 0.31)
Venlafaxine	1.58 (1.05, 2.37)	0.07 (0.03, 0.16)	0.23 (0.11, 0.48)	0.14 (0.06, 0.34)	0.17 (0.08, 0.31)

Preferences

ORDINAL Preference information

Criterion	Scale	Rank
HAM-D Responders	OR: [0.90 - 1.71] Risk: [0.51 - 0.66] RD: 0.16 NNT 6.44	6
Diarrhea	OR: [0.33 - 2.40] Risk: [0.03 - 0.21] RD: 0.17 NNH 5.74	1
Dizziness	OR: [0.39 - 4.72] Risk: [0.03 - 0.27] RD: 0.24 NNH 4.20	3
Headache	OR: [0.35 - 2.34] Risk: [0.07 - 0.34] RD: 0.27 NNH 3.75	5
Insomnia	OR: [0.60 - 1.89] Risk: [0.09 - 0.23] RD: 0.14 NNH 7.03	4
Nausea	OR: [0.81 - 2.27] Risk: [0.17 - 0.37] RD: 0.20 NNH 5.10	2

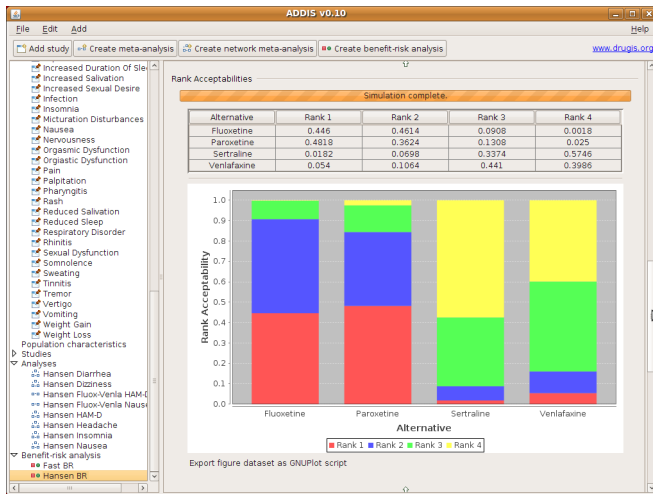
Rank Acceptabilities

Simulation complete.

Alternative	Rank 1	Rank 2	Rank 3	Rank 4
Fluoxetine	0.446	0.4614	0.0908	0.0018
Paroxetine	0.4818	0.3624	0.1308	0.025
Sertraline	0.0182	0.0698	0.3374	0.5746
Venlafaxine	0.054	0.1064	0.441	0.3986



ADDIS: benefit-risk (mild depression)



Discussion

Advantages

- Structure benefit-risk analysis
- Separate clinical data from value judgements
- Include all relevant clinical evidence
- Provide metrics for decision uncertainty
- Enable model generation for re-applicability

Limitations

- Relative scales not clinically relevant
- Data collection and extraction difficult and laborious
- Mixed Treatment Comparison is a relatively new method



Future Work

- Further development of ADDIS
 - E.g. alternative BR methods
 - 2-3 part-time developers at least next 6 months
- Robust method to present clinically relevant scales
 - Starting Aug. 2010
- Validation against existing regulatory dossiers
 - Study underway
- Applications
 - Considering several possibilities

