Name: Dr Lawrence D. Phillips
Job title: Emeritus Professor of Decision Sciences
Organisation / Institution: London School of Economics & Political Science
Is this input submitted as an organisational or individual response? Individual
Are you happy for your response to be published by the Academy? Yes

---

1. The overarching aim of the workstream is to better understand how society uses evidence to judge the risks and benefits of medicinal products. In your view, what are the key factors underpinning this process that the Academy should consider?

Key factors have emerged from two research projects: (1) the Benefit-Risk Project at the European Medicines Agency, a project I directed in 2009-2010, and is on-going (information here), and (2) the Benefit-Risk work package of the IMI_PROTECT project, started in September 2009, and completed earlier this year (information here). A lot of information is available from those links; I would be happy to brief you on their findings.

Both projects quantified and made explicit the implicit and subjective process of judging evidence, and the final report of the PROTECT Benefit-Risk project makes clear that the judging process requires (1) assessing the extent to which a medicine performs on the favourable and unfavourable effects, (2) the clinical relevance of the effects, and (3) the trade-offs among the effects. These three types of judgements are required to translate evidence of efficacy and safety into benefits and risks. These are the key features that any decision maker should take into account to make a rational decision. However, actual behaviour may well differ (my wife would look at a patient leaflet, look at all the side effects, and refuse to take the drug—no consideration of uncertainty or trade-offs for her!) Research in cognitive psychology suggests further that there is a problem of aggregating evidence, and that applies to all decision makers. So, that's a fourth key factor.

2. When evaluating the risks and benefits of medicinal products, what are the *strengths* of evidence that originates from different sources?

Obviously, well-designed studies conducted in head-to-head random-controlled trials are the gold standard, but my work with many teams of assessors and experts always requires discussion and debate about even the best studies, especially in light of shared clinical experience (the former Chair of the CHMP once told me that clinical experience is an essential input for approving new drugs and re-considering drugs already on the market). In modelling the benefit-risk of drugs, all the sources you identified in this question come into play, for each has its strengths and limitations, but these depend on context, so your question has no absolute answers. Also, much depends on who is making the judgements about the sources of evidence, and how well they exercise the four features I listed in question 1.

3. When evaluating the risks and benefits of medicinal products, what are the *limitations* of evidence that originates from different sources?

Meta-analyses can deal with the different sources; that's what a random effects model is for, as you know. But the main limitation is the lack of evidence.

I recently modelled the benefit-risk balance of six OTC drugs, ibuprofen salts&soluble, ibuprofen acid, naproxen, diclofenac, paracetamol and aspirin. A team of experts in a workshop looked at their relative performance on 11 effects, but adequate data were lacking on seven of the effects, despite the many millions of people who have used these drugs world-wide. Even for the remaining four, participants had to discuss many studies and then agree reasonable figures for those effects. The lack of standardised methodologies for making decisions about benefits and risks is, in my view, the most serious limitation. It's the judgements that are the problem, and once the focus in shifted to how those are made, then standardised methodologies can develop. Both the EMA and PROTECT projects strongly recommended application of structured approaches to evaluating evidence and assessing the benefit-risk balance of drugs.

4. Please provide details of any further examples or case studies that it would be useful for the project to consider.

New light could be shed on the society's topic of judging risks and benefits if you would consider decisions about benefits and risks, and who makes them. For example, significance levels and confidence intervals on individual effects from even the best RCT tell you nothing about the benefit-risk balance. What a decision maker really wants to know is how likely it is that the benefit-risk balance of drug X is better than doing nothing or than drug Y. Classical statistical methods cannot answer that fundamental question. Bayesian statistical decision theoretic methods can, and they are increasingly appearing in the literature, so you might consider them.

5. Please highlight any broadly applicable principles that should govern the presentation, interpretation and weighting of evidence about medicinal products.

These principles have been made explicit in the outputs of the PROTECT Benefit-Risk project. I suggest you click here and then follow the 5-steps indicated in the blue tabs beginning with the Introduction to see our recommendations.

6. Concerns have been raised about how industry funding impacts on the validity, or the perception of validity, of evidence. For example, the ability of academic researchers funded by industry to remain impartial when evaluating evidence has come into question. How should conflicts of interest be addressed? How important is industry funding in generating and analysing evidence? Other than industry sponsorship, what are other potential sources of conflicts of interest?

It has been my unfortunate personal experience that if someone doesn't like the findings of an industry-sponsored study, it is rubbished and the authors are accused of conflicts of interest. I have also had very good, totally hands-off experience of industry-sponsored studies leading to

publication, as well. I think the best that can be done is to enquire more deeply about the process that developed the evidence; this is not mentioned in the COI statements used by most medical journals.

For example, my work modelling the benefit-risk of drugs is always conducted in a workshop setting where I provide impartial facilitation and do not contribute to the content of the discussion. Care is taken in choosing experts to ensure that all the important perspectives on the issues are represented in the room. The meeting begins with introductions and with statements of possible conflicts of interest, enabling everyone to see where everyone else is coming from. I ensure that all views are aired and debated in rational and respectful discourse. Quantitative modelling is undertaken using facilitation techniques that reduce bias by ensuring that the first (or most dominant participant) does not bias other participants. Peer review operates effectively as preference judgements and trade-offs are given quantitative form. The process concludes with extensive sensitivity analyses in which the overall results are tested to see if they would change under alternative points of view, which enables participants to agree about results, even though they may well disagree about details of judgement and be worried about uncertainties in the data.

I suggest that standards could be established for the process of generating and analysing data, and exercising judgement about it; this would go beyond experimental design and statistical analysis, and would include questions about the features of the decision making process.

7. Please outline any past, current or planned initiatives to examine how patients, citizens and healthcare professionals (and those who seek to inform them) evaluate scientific evidence about medicinal products.

I suggest you explore how the decision makers identified on the home page of the PROTECT B-R website make decisions about medicinal products. I interviewed regulators in six EU countries and asked them questions about their approach, for I could find nothing written down about this at the start of the EMA project in 2009. Our interviews revealed huge differences just in the meaning of ‘benefit’ and ‘risk’, and it seemed to us important that a first step to improve transparency and communicability of benefit-risk required agreed definitions of those concepts, which the EMA implemented.

After I and my colleagues modelled five new drugs then under consideration by the CHMP, we published our findings, as you can read in the attached Drug Discovery Today—Technology paper, the first to be published from the EMA B-R Project. We later recommended that an Effects Table should be included in every European Public Assessment Report, showing what favourable and unfavourable effects were the basis for judgement about a new medicine, and the data considered for a drug and its comparator(s) on each of the effects. That will appear from January 2016.

As for a new initiative, I have attached my current initiative for comparing all the main drugs for a given indication. The workshop method I mentioned in question 6, has now demonstrated its feasibility and usefulness, and I believe it would be possible over the next few years to develop apps, computer programs and publications that would show how teams of experts have evaluated the benefit-risk balance of groups of drugs, a sort of Which? analysis for drugs. This would greatly extend the ability of decision makers, and I am already engaged helping some pharmaceutical companies compare their drugs against others.

One fascinating result for one drug was to find that its overall benefit-risk balance was nearly identical with that of a competitor, but analysis of the differences on each of the effects found almost no overlaps: one drug was better on about six effects, while the other drug was better on a different six. Making that sort of feature explicit would enable prescribers and pharmacists to make better decisions (and, by the way, save drug costs).

8. What are the most effective ways of communicating evidence to various stakeholders and engaging with them about such evidence?

That's what my proposal is all about. See the last page of the proposal for an illustration of how the quantitative model can be translated into words.

There is increasing concern to find ways to communicate more effectively with the public. One PROTECT-BR paper, by Andrea Beyer et al, which will soon be published, measured patients' comprehension of benefit and risk of medicines using 5 presentation formats (text, table, bar graphs, pictograms and survival curves). Information about medicines for atrial fibrillation, breast cancer and type II diabetes was presented in three EU countries to 770 patients, who were asked questions about the presentations. Comprehension was best for table presentations, and patients liked that format best, though they were able to extract information from more complex formats such as survival curves and bar graphs.

I wonder how your work on judging the benefits and risk of medicines might be affected if you were to frame the question in terms of decision-making? The reason I ask is because it strikes me that benefit-risk balance is the route to better decisions, and that route requires consideration not only of evidence but also judgements about the adequacy and relevance of evidence, as well as the trade-offs among the effects resulting from taking medicines. That wider frame may well lead you to see and consider aspects of the use of medicines that are important but have hitherto not attracted much attention.

Attached Documents

- Phillips LD, *et al*. (2011). *Is quantitative benefit-risk modelling of drugs desirable or possible?* Drug Discovery Today: Technologies **8(1)**, e3-e10.
- Phillips LD. Proposal to fund a research programme for comparative benefit-risk analyses of marketed drugs.

Proposal to fund a research programme for comparative benefit-risk analyses of marketed drugs

Lawrence D Phillips
Emeritus Professor of Decision Sciences
London School of Economics and Political Sciences

For the past six years I have been actively involved in research on the quantitative modelling of the benefit-risk balance of drugs, first in leading a team co-opted to the European Medicines Agency from 2006 to 2009[1], second as a participant in the IMI PROTECT[2] project, which ended earlier this year. Both projects demonstrated the feasibility and desirability of a structured approach to explicit, quantitative modelling, which I believe could be applied to the comparative benefit-risk assessment of drugs (CBRAD) for any given indication.

This proposal seeks to start the process of exploring the potential of CBRAD, which could lead to establishing an independent, impartial organisation that can engage with medical experts and clinicians in workshops to build quantitative models incorporating publicly-available data and quantified expert judgement about the clinical effectiveness and relevance of the evidence.

Quantitative modelling provides a structured process for comparing the benefit-risk balance of drugs on each of the favourable and unfavourable effects, and identifying those differences that are clinically relevant. The clarity of the results makes it possible to then create brief descriptions of the relative advantages and disadvantages of each individual drug, accompanied by an overall benefit-risk rating. One possible output of this structured process, for six over-the-counter analgesics, is shown here on the last page.

This information, based on a scientifically-based comparison of drugs available for an indication, could be made available in the form of apps and computer software available to every clinician, pharmacist and hospital in the UK, and possibly to patients as well.

But doesn't the British National Formulary (BNF), which anyone can purchase, already do this?

Not really. Although it is an impressive and useful document, providing convenient summaries of drugs in the UK along with guidance on best practice, much of it from NICE, nearly everything is described in words. There is hardly any numerical data about the magnitude and likelihood of favourable effects, or the severity and incidence of unfavourable effects. Few data tables and no graphical summaries are given, forcing users to trawl through an often considerable amount of text before deciding.

Isn't this sufficient for most clinical interventions?

Yes, in deciding what approach to take for a given patient. For example, in the section on antidepressant drugs, the BNF points out that 'SSRIs are better tolerated and are safer in overdose than other classes of antidepressants and should be considered first-line for treating depression'. In the section on SSRIs, a page of

[1] Phillips, L. D., Fasolo, B., Zafiropoulos, N., & Beyer, A. (2011). Is quantitative benefit-risk modelling of drugs desirable or possible? *Drug Discovery Today: Technologies, 8*(1), e3-e10. doi: 10.1002/pds.3636

[2] http://protectbenefitrisk.eu/

general comments is followed by descriptions of six non-proprietary drugs, but there is no comparative assessment of the six.

What would a comparative assessment add?

First, an ordering of the drugs based on their overall benefit-risk balances, accompanied by text describing the key advantages and disadvantage of each drug compared to the other alternatives; second, a breakdown of how each favourable and unfavourable effect contributes to that balance; and third, a table showing the definitions of all the effects, the input data and the weights assigned to each of the effects. See the example for OTC analgesics in the case study beginning on page 4.

How would that assist decision making?

The benefit-risk ordering establishes priorities for exploring the drugs and shows the contributions that the individual effects make to the overall ordering, thereby identifying the benefits and risks that might be especially relevant to a given patient. The text enables a quick comparative survey of the alternatives. In addition, the weights, which represent the clinical relevance of the effects, are made explicit, so they can be over-ridden by the decision maker to be more applicable for the patient. The effects table provides decision makers with the evidence, enabling them to apply with confidence their own clinical judgements in making a decision.

How are the benefit-risk priorities established?

By a panel of five to seven medical experts, clinicians and pharmacists, who represent a diversity of perspectives on the relevant system of the body or aspect of medical care, for the set of drugs that are to be compared. The panel work in a face-to-face workshop for one day discussing the available evidence and creating a quantitative model based on decision theory, helped by an impartial facilitator who guides the modelling process but does not contribute to the content of the model. The facilitator is experienced in working with groups and specialises in modelling decisions with multiple favourable and unfavourable effects, each of which contribute in some degree to the benefit-risk balance. The participants provide the content for the model: the evidence and the trade-off weights between the effects.

Isn't this approach over-ambitious? There are thousands of drugs on the market.

Only drugs that are competing for the same medical condition need be included, and then only if there are sufficient competitors to warrant calling a workshop of experts. Guidance will be sought from practicing physicians and pharmacists on where the focus should be; perhaps the BNF could provide information on the most frequently viewed pages online. In any event, it will take several years to build up sufficient comparative analyses for users to consistently and regularly consult the CBRAD results. In the meantime, regular updates of the software as CBRAD is completed for an indication will provide short-term benefits to subscribers (which might be free for an initial period).

What organisational structure is required?

Could be an Institute in the medical or other faculty of a distinguished university; perhaps an adjunct to the BNF, which is owned by the British Medical Association/Royal Pharmaceutical Society; possibly a part of NICE; or a private, not-for-profit company.

What is needed to establish the feasibility and desirability of this proposal?

The feasibility of modelling a single drug against its comparator has already been demonstrated by research over the past six years administered by the European Medicines Agency (EMA) and the IMI-PROTECT projects. Some pharmaceutical companies are now beginning to use modelling of benefit-risk as an important contribution to decision making during drug development. This proposal builds on that experience, but the desirability of *comparative* assessment of drugs remains to be shown as potentially useful. That will be established in the proposed feasibility study, through demonstrations of CBRAD outputs, and interviews with potential users and key opinion leaders.

How can the difficulty of obtaining the data, experienced in both EMA projects, be overcome?

The EMA's Committee for Medicinal Products for Human Use (CHMP) has approved the use of the Effects Table in reporting on the approval of a new drug. This table, which will list the favourable and unfavourable effects taken into account in assessing the benefit-risk balance, as well as the data associated with each effect, will be reported beginning in January 2016. That information will provide a valuable starting point for comparative benefit-risk assessment. Data about drugs already approved will continue to be collected using the variety of sources already available, Cochrane, literature, etc.

What are the expected impacts of this project?

Discussions and interviews with several interested health professionals suggest that an understanding of the relative benefit-risk balance of drugs could become a useful input to the many joint formulary committees in the UK, enabling them to create formularies that are based on hard comparative evidence and the judgements of experts and clinicians. Cost saving would be an obvious impact. Another possibility is the saving of hospital admissions by prescribing drug A instead of drug B. The availability of an easily-accessed app or computer software with displays like those that have been found to work well in Which? magazine for consumer products, would enable prescribers to quickly choose a cost-effective drug, providing possible cost savings at no loss in effectiveness. Patients, especially, would welcome the non-technical descriptions shown on page 6.

The ready availability of a CBRAD model could impact drug development as a pharmaceutical company uses the model to test the benefit-risk balance of a drug under development, possibly helping to allocate its limited resources to more promising opportunities. This will become increasingly possible as the Effects Tables appear from January 2016 in the European Public Assessment reports for new drugs. Those tables will facilitate the rapid construction of quantitative models, providing the framework for obtaining assessments by patients of the trade-off weights associated with the favourable and unfavourable effects. Adding a drug that is currently being fast-tracked by the MHRA in an existing CBRAD model could speed up approval. Other impacts may well emerge during the feasibility study and in any subsequent implementation.

What are the next steps?

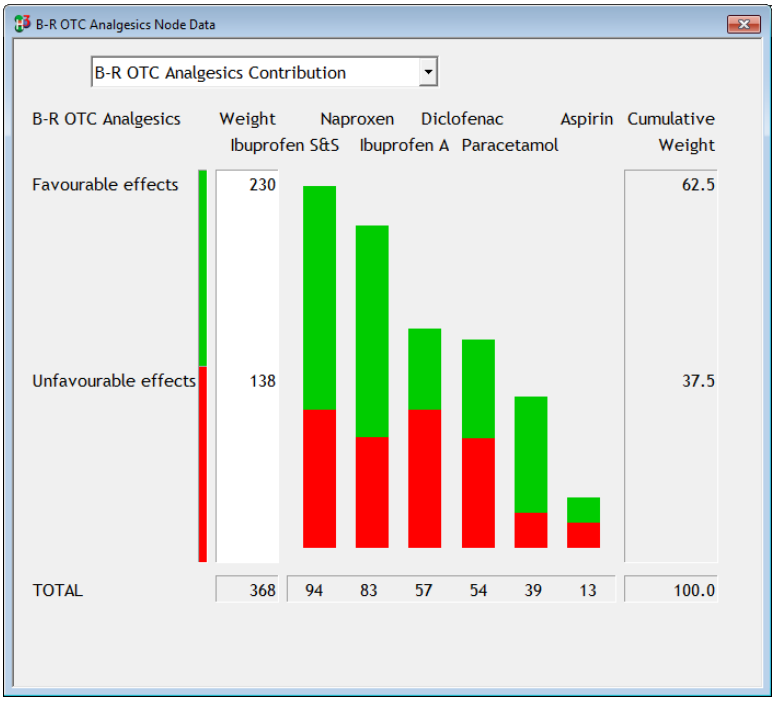
A senior advisory group consisting of Sir Alasdair Breckenridge, former Chair of the UK's Medicines and Healthcare products Regulatory Agency (MHRA), Professor David Nutt, the Edmond J Safra Professor of Neuropsychopharmacology and director of the Neuropsychopharmacology Unit in the Division of Brain Sciences at Imperial College, and Dr Desmond Fitzgerald, a former Director of ICI Pharmaceuticals and Director of Materia Medica, are supporting this initiative and guiding its implementation.

Our current concern is to find funding for a feasibility study to model the benefit-risk balance of drugs for, say, three different indications with groups of experts, develop a simple app or software program to display

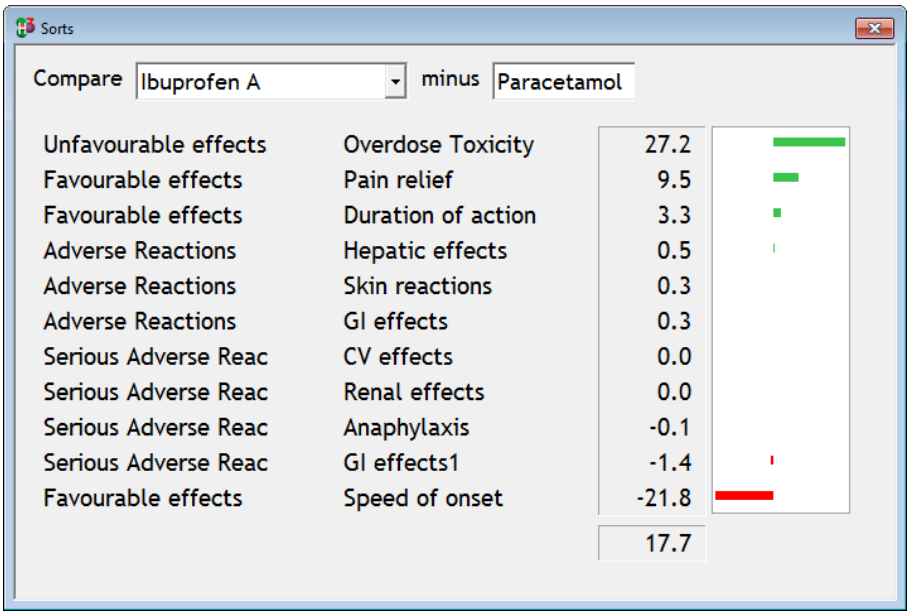
the results and then try out the app or software with select medical professionals (GPs, consultants, pharmacists, hospitals) to obtain their views about the usefulness of this approach, its delivery and its potential impact. The information gained in the feasibility study should provide information about the markets for CBRAD sufficient to support decisions about the next steps.

A case study—comparisons of OTC analgesics[3]

Seven medical experts developed in a facilitated workshop a multi-criteria decision analysis (MCDA) model of the comparative benefit-risks of six OTC analgesics: ibuprofen salts & solubilised, naproxen, ibuprofen acid, paracetamol, diclofenac and aspirin. The bar graphs show the overall results, expressed on a 0-100 preference-value scale. Longer green bars represent more benefits while longer red bars show more safety.



Note that the soluble form of ibuprofen is overall best, and both formulations of ibuprofen are safer than any of the other analgesics. As these are relative figures, it is differences, not ratios, of the numbers that are meaningful. For example, the difference between ibuprofen S&S at 94 compared to paracetamol at 39 is 55, whereas the difference between ibuprofen A at 57 and paracetamol is 18. The ratio of those differences, 55/18 = 3.1, means that the superiority of ibuprofen S&S to paracetamol is about three times the superiority of ibuprofen A to paracetamol.



What accounts for those 18-points? This is shown in the difference display, below. The effects are ordered according to the weighted differences between the two analgesics, shown in the column of numbers and displayed as horizontal bar graphs. The sum of the individual weighted differences is 17.7. Thus, the green bars show one major advantage, better overdose toxicity, and two smaller advantages, pain relief and duration of action, for ibuprofen A over paracetamol, while the red bars show one major ad-

[3] The study is currently being written up for publication in a peer-reviewed journal.

vantage of paracetamol over ibuprofen A, its speed of onset. Many effects show no differences. These displays, repeated for the differences between all 15 pairs of analgesics, were crucial in creating the text for page 6.

The Effects Table, below, provides the inputs to the MCDA model. Note that the metrics of the first four effects are based on real-world measurements, which were usually taken from a variety of publicly-available sources. The remaining inputs are preference values assessed by the groups of experts after consulting and discussing the evidence. Each row contains relative strength of preference values ranging from zero to 100, where zero represents the least preferred drug for the associated effect, and 100 represents the most preferred.

		Effect names	Units	Ibuprofen salts & solubilised	Naproxen fast acting	Ibuprofen acid	Diclofenac fast acting	Para-cetamol	Aspirin (analgesic)
Favourable Effects		Pain relief	%	63	55	48	45	33	20
		Duration of action	hours	7.0	9.0	5.5	4.5	4.0	5.0
		Speed of onset	mins.	27	30	55	45	30	50
Unfavourable Effects	Adverse Reactions	Skin reactions	No.	24	26	24	41	77	124
		GI effects	Pref.	100	100	100	100	100	0
		Hepatic effects	Pref.	100	50	100	100	0	30
	Serious Adverse Reactions	GI effects	Pref.	50	20	40	30	100	0
		CV effects	Pref.	75	80	75	0	75	100
		Renal effects	Pref.	100	0	100	100	100	100
		Anaphylaxis	Pref.	50	50	50	50	100	0
		Overdose Toxicity	Pref.	100	75	100	0	0	75

The first four rows of metrics were converted linearly to 0-100 preference values. The 11 preference scales were then weighted[4] to provide scale constants equating the units of preference across all the scales, thereby enabling the sum of weighted preferences to be calculated for each drug and then compared across the drugs.

This table represents the best collective judgements of the seven experts who took part in the exercise. Scanning across the rows shows which drug performs best and worst for each effect. Scanning down the columns is less helpful because at this stage the rows of metrics are all in different units, and the preference values are not comparable as they have not yet been weighted. However, it is also possible to show a table of the weighted preference values, which would enable comparisons of differences between any of the cells in the table.

The next page shows a possible translation of this quantitative information into readable text.

[4] The method of swing weighting was applied as the facilitator asked the group to consider the size of the real-world difference between 0 and 100 for a given effect, and to judge the clinical relevance of that difference. Comparison of pairs of swings led to ratios that were then converted to a unique weight for each effect, which preserved the original assessed ratios. The final weights are scale constants, like the 5:9 ratio for Celsius and Fahrenheit scales—5 Celsius degrees of temperature are equivalent to 9 Fahrenheit degrees.

Over-the-counter (OTC) analgesics

Seven experts compared the benefit-risk balance of six OTC pain-killers on three favourable effects, pain relief, duration of action and speed of onset. They also looked at eight unfavourable effects, but only three showed any clinically relevant differences for these six drugs: serious or non-serious gastrointestinal effects (considering incidence rate and severity), and the potential for toxic overdose, either accidental or deliberate.

The green numbers are overall weighted preference values out of 100, which take into account available data for the effects, and judgements of the experts about the clinical relevance of the effects. They do not take account of contra-indications, interactions with other drugs, or other precautions. Figures given in the descriptions are averages; individuals may experience effects that are different from these averages.

Ibuprofen soluble 94

FAVOURABLE EFFECTS
This is the best drug for pain relief and speed of onset (27 minutes). Only naproxen beats it for duration of action.
UNFAVOURABLE EFFECTS
It is safer than naproxen and aspirin for serious GI effects, about the same as diclofenac and worse than paracetamol. Both forms of ibuprofen are lowest for overdose potential.

Naproxen 83

FAVOURABLE EFFECTS
Duration of action is best for this drug, 9 hours compared to 7 hours for soluble ibuprofen, and 4 for paracetamol. Speed of onset, 30 minutes, is nearly as good as soluble ibuprofen.
UNFAVOURABLE EFFECTS
The incidence and severity of serious GI effects is better than aspirin but slightly less good than the other drugs.

Ibuprofen tablet 57

FAVOURABLE EFFECTS
Pain relief is better than diclofenac, paracetamol and aspirin. Its duration of action is somewhat better than diclofenac, paracetamol and aspirin.
UNFAVOURABLE EFFECTS
This tablet form of ibuprofen shares with soluble ibuprofen the lowest potential for overdose.

Diclofenac 54

FAVOURABLE EFFECTS
This drug is better than paracetamol and aspirin for pain relief and has a faster speed of onset than tablet ibuprofen and aspirin.
UNFAVOURABLE EFFECTS
Diclofenac is slightly better on serious GI effects than naproxen and much better than aspirin. Its potential for overdose is low.

Paracetamol 39

FAVOURABLE EFFECTS
The 30-minute speed of onset of this drug and naproxen is slightly slower than soluble ibuprofen at 27 minutes, but at least 15 minutes better than any of the others. Better than aspirin for pain relief.
UNFAVOURABLE EFFECTS
It is the safest drug for serious GI effects, but has the highest potential for overdose.

Aspirin 13

FAVOURABLE EFFECTS
Speed of onset is 5 minutes better than tablet ibuprofen. Duration of action, 5 hours, is somewhat longer than paracetamol, 4 hours, or diclofenac, 4½ hours.
UNFAVOURABLE EFFECTS
Aspirin is better than paracetamol for overdose potential. Other than the above, aspirin is poorest for all the effects