Variability in Human Drug Response
G.T. Tucker, Editor

Benefit risk evaluation of medicinal products - an epidemiological approach

Carlos Martinez

Hoechst Marion Roussel, Inc, Bridgewater, NJ, USA.

ABSTRACT

A fair comparison of the benefit/risk of drugs requires a separate estimation of both benefit and risk of each drug. Prerequisites before evaluation are: definition of the perspective of evaluation, definition of indication for drug use and definition of alternative treatment strategies. The benefit measure corresponds to the efficacy which strictly depends on the indication for drug use. Comparative estimates can only be obtained from randomized clinical trials. The evaluation of drug-attributed risks includes the following steps: define profile of ADRs of drug and comparator(s), obtain background incidence of each ADR, define common outcome measure and ADR-specific values, estimate excess incidence rate of each ADR by treatment strategy, select a common outcome unit and determine weights to permit comparison across ADRs, estimate adjusted excess incidence of each reaction with the weighting factor, sum weighed excess incidence rates for each treatment strategy, calculate net risk difference of drug-attributed risks, discuss validity of results and update evaluation when new eligible studies become published. Drug specific benefit/risk estimates could be expressed as the difference between benefits and risks, if the unit of measurement of benefits and risks is identical for all treatments only or by the benefit/risk ratio otherwise.

Key words: risk, benefit, adverse drug reaction (ADR).

Correspondence: Carlos Martínez, MD, MSc, Hoechst Marion Roussel Inc, Route 202-206, P.O.Box 6800, Bridgewater, NJ 08807-0800, USA, Tel: +1-908-231-5667, Fax: +1-908-231-4827, Email: carlos martinez@hmrag.com

INTRODUCTION

A fair comparison of benefits risks of drugs requires a separate estimation of both the benefit and the risk of each drug. A fair comparison also implies that the medicines are comparable. The indications for drug use, the patient population including age and gender distribution, and the duration of treatment need to be alike for drugs aimed at such conditions as postoperative pain, community-acquired pneumonia or type I diabetes.

The following methodological framework elaborates the means to weigh the benefits and risks of new medicines against standard treatment. This framework intends to standardize the

procedure of benefit and risk assessment from the point of view of health authorities or pharmaceutical firms, and to serve as a guidance for health authorities whose intention is to maximize health. Furthermore, the benefit-risk assessment is of primary importance to manufacturers, whose responsibility is to assess continually the benefit-risk balance of new medicines, particularly after marketing when the indication of new medicines is broadened or after new drug-attributed adverse events are identified.

This approach will result in a benefit-risk assessment that in most cases will be comparative. Since benefit-risk evaluations are only valid for drugs given for a single defined illness, this approach will not supply a benefit-risk score that will be valid for medicines used for different indications. The methodological approach will provide a score that could be used for competing drugs given for a defined indication or illness.

Although the risk of medicines can be measured, the safety of drugs is considered to be a judgment. The safety of drugs is a composite measure of benefits and risks. Our approach does not intend to change the current practice in clinical trials. It intends to use the knowledge that is already available.

BENEFIT ESTIMATION

Clinical efficacy of new compounds is usually not demonstrated by the proportion of prevented or postponed deaths, life-years gained or quality-adjusted life-years gained, i.e. direct measures of benefit. Primary endpoints in randomized clinical trials are intermediate parameters (surrogate endpoints) believed to be correlated with measures of direct benefit.

In this section the different measures of benefit as used in randomized clinical trials are reviewed and the potential benefits discussed.

1. Benefit measured as an intermediate parameter

Randomized clinical trials are the gold standard to test the efficacy of new medicines compared to placebo or commonly used effective medication. Randomization reduces bias and variability between randomized groups. Randomization generates comparable heterogeneous cohorts, i.e. the likelihood that patients in each cohort possess a given characteristic is similar. Thus, any difference in the frequency of outcomes in each group is attributable to the effect of the clinical intervention, i.e. the allocated treatment strategy.

Randomized clinical trials can have three results: (1) the new medicine is less efficacious than the comparator, (2) the new medicine is equivalent - the most common situation in real life and (3) the new medicine is more efficacious than the comparator.

Traditional measures of efficacy are intermediate parameters or surrogate endpoints that depend on the indication for drug use. Surrogate endpoints include diagnostic procedures, e.g. laboratory tests for such things as blood glucose level, cholesterol level or antibiograms; functional status such as blood pressure or ECG changes after exercise tolerance tests and forced expiratory volume at one second (FEV_1) in asthmatics; imaging procedures such as X-ray, CAT scan, ECG or ultrasound; symptom relief (particularly for pain, fever, inflammation), quality of life and other important disease-specific measures of efficacy. Advantages of cohorts can include information regarding time dependent variables such as changes in daily dose of any medication or smoking habits and total cumulative dose and the defined dose for which a clinical benefit is claimed.

For intermediate clinical endpoints an association with quality of life, morbidity or mortality is assumed; however, it is not routinely demonstrated as in morbidity or mortality studies, e.g. decreased mortality of cardiovascular events after secondary prevention of myocardial infarction with cholesterol-lowering medicines. As a consequence, long treatment duration and observation, differential loss to follow-up, patients' compliance and large sample sizes necessary to detect a significant difference would require expensive and time-consuming studies that could withhold efficacious medicines from patients.

2. Benefit measured as cumulative mortality or mortality rate

In mortality studies clinical benefit is measured by the number of prevented or postponed deaths. Mortality rates obtained in cohort studies are comparative and can either be compared with a do-nothing strategy such as aspirin in primary prevention of myocardial infarction (Physicians Health Study) [1], or with standard treatment or another medicine of interest. The measure of clinical outcome such as death is attributed to the treatment strategy as allocated by randomization. In any case, mortality studies provide strategy-specific cumulative mortalities or mortality rates, i.e. number of deaths in each treatment strategy divided either by the respective population at risk or by the accumulated person-time exposure in an intention-to-treat analysis. Total person-time exposure corresponds to the sum of person-time accumulated by individuals according to the primary assignment to a treatment strategy with the assumption that all subjects were compliant with the study protocol.

The differences and ratios of cumulative mortalities or mortality rates provide an estimate of the comparative benefit of the medicines studied. Differences provide a more informative estimate than ratios since differences inform us of the amount of prevented or postponed deaths in a defined population, i.e. provide an estimate for the number of prevented or postponed deaths attributable to a particular medicine in a defined population.

3. Benefits measured as life-years gained

Life-years gained is usually not a primary endpoint in studies submitted to registration authorities. Non-registration clinical studies have increasingly been used as a vehicle for pharmacoeconomic studies in which the clinical efficacy of a new medicine is translated into more useful natural units such as life-years gained, which are inherently built into mortality studies. Life-years gained could be calculated either from mortality data by modeling or directly from raw data (area between two Kaplan-Meier curves corresponds to the time, e.g. life-years, gained by the treatment strategy corresponding to the upper curve). The number of life-years gained attributable to a medicine could be considered as the benefit of a medicine from a societal viewpoint. The difference in life-years gained attributable to a medicine, however, does not consider the quality of life of patients whose death was prevented or postponed.

Example: In cancer therapy a new drug may postpone death, but may result in a lower quality of life because of the adverse events and repeated regimens of radiotherapy or chemotherapy.

4. Benefit measured as quality-adjusted life-years

Quality-adjusted life-years (QALYs) is a composite index that reflects the number of life-years gained, weighted by the quality of life of each of these years. Quality of life is the individual's subjective perception of his or her physical, emotional and social well-being, after

taking the effects of a disease and its treatment into account. The components of quality of life differ by the scale used but in general include physical, social, emotional and functional well-being. The evaluation of new therapies to measure health outcomes have three principle components: (1) detection of changes of health status by defining the key domains of interest (description), (2) comparative quantification of quality of life changes by determining the means of weighting elements within the descriptive system (valuation), and (3) the determination of aggregate (dis)benefit.

Quality of life measures yield a profile or an ordinal or cardinal index. Indices can be used to quantify quality of life changes, a very important measure in benefit-risk evaluation. Health utility scores rank from 0 to 100. Complete health is anchored as 100 and death as 0. The advantage of health utility scores is that drug-related mortality and drug-related morbidity are integrated into one risk score. The patient's change of quality of life is integrated as well.

Benefits can be measured in QALYs. Different methods can be used, e.g. the time-trade-off method, standard gamble, or any quality of life scale that allows the generation of an index in a cardinal scale. The underlying principle is that different health outcomes can be valued in terms of the individual's preferences under conditions of uncertainty. Individual's preferences (choice) for a hypothetical health state yields a mathematical value. The individual's preference corresponds to a ranking of health utility scores. Although standard gamble and time-trade-off methods generate ordinal utilities, there is the tendency to use them as cardinal measures. There are no alternatives or other theories.

QALY measurements inherently contain the profile, the frequency and the value of respective adverse effects of a drug. In this case one unique measure provides the benefit-risk difference of a new compound without having to estimate the benefit and risk separately. The number obtained then needs to be compared with the standard of another alternative treatment. The compound with the higher number, i.e. larger QALY gained, has the better benefit.

5. Benefit measured as a morbidity rate

In addition to primary or secondary prevention studies, morbidity can be a primary endpoint in the study of treatments for chronic diseases, e.g. non-fatal cardiovascular events (myocardial infarction or stroke) in the treatment of hypertension or retinal hemorrhage in the treatment of type I diabetes.

RISK ESTIMATION

Until the start of drug marketing, clinical trials are the exclusive source of human efficacy data, and also provide high-quality data on the risk of medicines. The primary goal of clinical trials is to collect data on drug efficacy. The frequency of drug-attributed adverse events, however, depends on their nature. Adverse events which are a result of the pharmacological action are dose-dependent and likely to occur frequently, e.g. symptomatic hypotension in antihypertensive treatment or hypoglycemia in insulin-treated diabetes. In some treatment strategies each patient may have adverse events, e.g. neutropenia in anticancer treatment.

Other adverse events such as allergic reactions are believed not to be dose dependent and therefore not to result from the intended pharmacological action of a new medicine, e.g. dry cough in ACE-inhibitor treatment (less than 10% of patients) or allergic reactions. Potentially fatal adverse events attributable to drugs, e.g. anaphylactic shock, agranulocytosis, upper

gastrointestinal bleeding or acute renal failure, are rare and occur far less than 1:1000. Since premarketing clinical trials involve about 3000 patients, potentially fatal adverse events are unlikely to occur, and if they do occur, further drug development would be questioned. Whether an adverse event can be statistically attributed to a medicine depends on the background incidence and the observed frequency of each adverse event. Potentially fatal adverse events observed in the postmarketing phase require hypothesis testing in epidemiological studies. Most data are generated by clinical studies, spontaneous reporting or epidemiological studies. Table 1 provides an overview of the origins of these data.

Table 1. Overview of the origins of data for risk estimation

	Clinical studies	Spontaneous reports	Epidemiological studies
Profile of adverse events	X	X	
Excess cumulative incidence	X ($<10^{-3}$)		X ($<10^{-3}$)
Duration of adverse event	X	X	X

Epidemiological studies are not routinely performed. They are often restricted to the investigation of certain serious adverse events. A case-control study can investigate one adverse event at a time. Therefore, epidemiological data on the frequency of other potentially serious adverse events may not exist. Published studies may therefore be biased in the sense that potential drug risks may not have been investigated.

The proposed measure of risk is drug-attributed loss of quality-adjusted life-years (DALQALY). DALQALYs can be either transient or permanent. Transient DALQALYs take into consideration (1) the drug-specific profile of adverse events, (2) the excess frequency of each adverse event, (3) the duration of each adverse event and (4) a health utility score based on each adverse event. As a consequence, the overall drug risk attributable to transient adverse events corresponds to a composite risk score synthesized from the following parameters:

$$DALQALY_{i}^{trans} = \sum_{j=1}^{n_j} U_{ij} \times d_{ij} \times CI_{ij},$$

where

n is the number of all drug-attributed adverse events of the medicines to be compared,
U is the health-state utilities for each adverse event that has a potential impact on health status,
d is the duration of each considered adverse event,

CI is the excess cumulative incidence of each adverse event,
i corresponds to each medicine to be compared, and
j corresponds to each adverse event.

In the case of permanently disabling adverse events (e.g. stroke) or death, DALQALYs result from (1) the drug-specific profile of adverse events, (2) the excess frequency of each permanently disabling adverse event, (3) the mean life expectancy of the target population and (4) a health utility score based on each adverse event. The overall drug risk attributable to permanent adverse events corresponds to a composite risk score synthesized from the following parameters:

$$DALQALY_{i}^{permanent} = \sum_{j=1}^{n_j} U_{ij} \times LE_{ij} \times CI_{ij},$$

where *LE* corresponds to life expectancy.

The mean life expectancy of the target population receiving standard therapy can depend on the underlying disease, age, gender or other factors, e.g. life expectancy in patients with heart failure NYHA IV depends on age and gender. Data on life expectancy can be retrieved from published epidemiological studies if the life expectancy is disease specific and from national statistics when an otherwise healthy population is of interest, e.g. analgesic treatment for acute and transient pain.

The total DALQALY is the sum of transient and permanent adverse events:

$$DALQALY_{i} = DALQALY_{i}^{trans} + DALQALY_{i}^{permanent} = \sum_{j=1}^{n_j} U_{ij} \times d_{ij} \times CI_{ij} + \sum_{j=1}^{n_j} U_{ij} \times LE_{ij} \times CI_{ij}$$

Data on the qualitative profile of all drug-attributed adverse events of a medicine can be obtained from clinical trial data, observational studies and spontaneous reporting. Another prerequisite is the generation of health utility values for each adverse event. Utility values represent preferences for the health state and can be generated by the general population. There are established methods for utility assessment such as visual analog scales, standard gamble or time trade-off [2-4].

Alternatively, utility values could be generated by the costs of treatment of adverse events, where costs represent the patient's need of treatment. Treatment costs due to adverse event could be translated into a utility scale from 0 to 10, where the highest treatment costs correspond to 10 and no treatment costs to 0. The highest cost corresponds to death. The costs of death could be generated by the loss of life-years, as estimated from the expected life expectancy of the target population, and the annual costs of treatment of a chronic and severe disease, e.g. renal dialysis. Both the costs of transient and permanent adverse events would be translated into the utility scale. This would circumvent the problem of who is most suited to estimate utility values: patients, a random sample of the general population or experts. In many cases common sense dictates that the general population is best suited.

HOW BENEFITS AND RISKS CAN BE CONSIDERED IN A BENEFIT-RISK EVALUATION

There are several scenarios which need to be considered: the new medicine is more efficacious, equivalent or less efficacious than the comparator. In each case the new medicine may have a lower risk, the same risk or an increased overall risk. This is shown graphically in Figure 1. It is obvious that drugs that fall into category II, i.e. less benefit and higher risk, are of no interest from a public health perspective. Drugs that fall into category IV are ideal drugs, i.e. breakthrough drugs with lower risk than standard therapy. For drugs that fall into categories I or III a benefit-risk comparison is necessary, since an increased benefit is associated with an increased risk or a decreased benefit associated with a decreased risk. Figure 1 also conveys information of three other scenarios. First, if the benefit difference is zero, the medicine with the lower risk is favorable. Second, if the risk difference is zero, then the medicine with the higher benefit is favorable. Third, if both the benefit and risk difference are zero, then the medicines have the same benefit-risk balance and the new medicine is of no particular public health value. The problem is, however, that the benefit-risk category cannot be assessed in advance, and a benefit-risk assessment is essential.

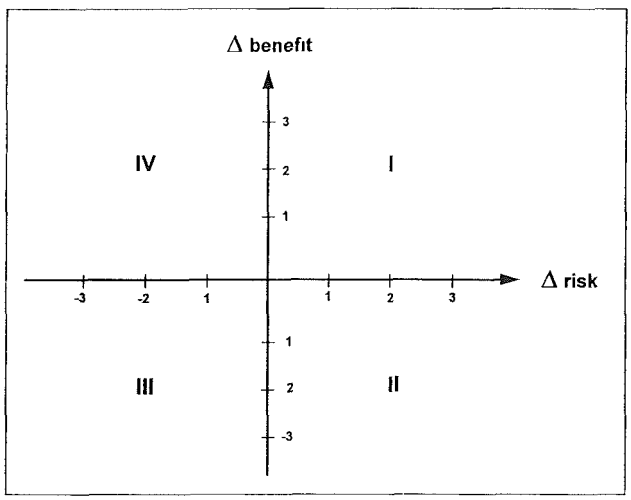

Figure 1. Benefit-risk evaluation scenarios

For the benefit-risk evaluation the following procedure is suggested:

1. Evaluate the benefit by quantifying the efficacy from randomized clinical trials.
   - Identify the comparators of the new medicine.
   - Identify the primary endpoint of registration studies.
   - Quantify the efficacy of the new compound comparatively.
2. Evaluate the overall risk from randomized clinical trials.
   - Identify all drug-attributed side effects for each drug.

- Estimate the cumulative incidence of each adverse event.
- Derive a utility value for each health state according to each adverse event.
- Estimate the duration of each adverse event.
- Estimate the life expectancy of the target population.
- Calculate the DALQALYs of each drug.

HOW BENEFITS AND RISKS OF DRUGS CAN BE BALANCED

To elaborate the benefit-risk balance evaluation, different benefit measurements are considered separately: (1) intermediate endpoints, (2) morbidity studies, (3) mortality studies, (4) life-years gained and (5) QALYs.

1. Benefits measured as an intermediate endpoint

When intermediate endpoints are measured to demonstrate clinical efficacy, then clinical efficacy and risk are not measured in the same units. To avoid the calculation of ratios such as the difference between benefit divided by the difference in risks, it is suggested that clinical efficacy be translated into a utility value, i.e. drug-attributed gain of QALYs (DAGQALYs). The medicine's benefit-risk balance could then be expressed as the difference between DAGQALYs and DALQALYs, each measured in the same units. As with risk estimation the translation of clinical efficacy into QALYs could be derived from a patient or general population.

If the new medicine and standard treatment are equivalent in clinical efficacy, then the risk difference measured in DALQALYs becomes important. The medicine with the smaller DALQALY is the more favorable.

If the new medicine and standard treatment have a different clinical efficacy but have an equivalent risk, then the medicine with the better clinical efficacy is more favorable, without translating intermediate endpoints into DAGQALYs.

2. Benefits and risks measured by morbidity frequencies

When the frequency of morbidity is the measure of clinical efficacy, then the clinical efficacy needs to be translated into a utility value, i.e. DAGQALYs. The medicine's benefit-risk balance could be expressed by the difference between DAGQALYs and DALQALYs. As with risk estimation the translation of morbidity into QALYs can be derived from a random sample of the general population.

3. Benefits and risks measured in deaths

In mortality studies new medicines are compared with standard treatment. An incremental benefit-risk analysis provides the difference in the overall cumulative death incidence. The net difference in overall mortality is already an expression of the benefit-risk balance. Non-fatal adverse events should also incorporated in the benefit risk balance by using morbidity data.

4. Benefits and risks measured in life-years gained

Information of life-years gained is inherent in mortality studies. When the overall mortality is considered, then the net difference in life-years gained is already an expression of the benefit-risk balance.

5. Benefits and risks measured in QALYs

If benefits and risks have the same unit, then the benefit-risk assessment is the difference between benefits and risks for each drug, e.g. DAGQALYs minus DALQALYs. The difference between DAGQALYs and DALQALYs corresponds to the net drug-attributed quality-adjusted life-years (DAQALYs). Any positive DAQALY is beneficial. The medicine with the higher DAQALY is more beneficial.

CONCLUSIONS AND DISCUSSION

The creation of a common measure of drug-attributed benefits and risks is proposed, i.e. drug-attributed gain of quality-adjusted life-years (DAGQALY) and drug-attributed loss of quality-adjusted life-years (DALQALY). The difference between these values can provide a comparative measure of the benefit-risk balance of medicines. The drugs to be compared, however, need to be comparable. Therefore, among other characteristics age and gender distribution, dose and duration of treatment and underlying disease have to be similar. The proposed measure of DALQALYs encompasses both transient and permanent adverse events. Transient DALQALYs take into consideration (1) the drug-specific profile of adverse events, (2) the excess frequency of each adverse event, (3) the duration of each adverse event and (4) a health-utility score based on each adverse event. For permanent adverse events information on the mean life expectancy of the target population is also necessary. For this purpose a registry of side-effect-attributed utility scores needs to be created. For each newly discovered adverse event, such a "utility register" would provide the respective utility value and allow the calculation of the updated value of the overall benefit-risk score by adding the side-effect-specific $DALQALY_{ij}$ to the previous equation. For this purpose use of an arbitrary population of 100,000 individuals as a standard is advised, i.e. $DALQALY_{ij}$ per 10^5.

The overall risk score must be compared with competing treatment strategies for which the same approach is necessary. Clinical efficacy data and adverse event data on competing treatment strategies are usually collected in randomized clinical trials.

A simplified risk analysis of non-narcotic analgesics, for example, might consider only drug-related mortality, assuming that all analgesics are equivalently efficacious and that there is only concern about potentially fatal adverse events. In this case, the drug-related mortality of each analgesic needs to be considered, thereby taking the different profiles of adverse events into account.

REFERENCES

1. Belanger C, Buring JE, Cook N, Eberlein K, Goldhaber SZ, Gordon D, Hennekens CH, Mayrent SL, Peto R, Rosner B, Stampfer MJ, Stubblefield F and Willett WC: Final report on the aspirin component of the ongoing physicians' health study. New Engl J Med 1989;321:129-135.
2. Froberg D, Kane R: Methodology for measuring health-state preferences - I: measurement strategies. Journal of Clinical Epidemiology 1989;42:345-354.
3. Froberg D, Kane R: Methodology for measuring health-state preferences - II: measurement strategies. Journal of Clinical Epidemiology 1989;42:459-471.

4. Torrance G, Feeny D: Utilities and quality-adjusted life years. International Journal of Technology Assessment in Health Care 1989;5:559-575.

Discussion: Benefit and risk evaluation of medicinal products - An epidemiological approach

J. Urquhart:
How did you ascertain drug exposure? You used the word 'on', which is a two-letter preposition that usually carries a lot of baggage in pharmacoepidemiologic studies.

C. Martínez:
These studies come from 1970s. Nowadays, the information that we are getting from the studies is more precise and complete. Overall, in all the studies, patients were asked by a questionnaire which drugs did they take last week, which dose and how often.

J. Urquhart:
I would suggest that the thing to do with studies done with meagre methods of ascertainment of exposure, is just to junk them, because they are not worth talking about. You get misled so badly by imbalanced exposure. Another point is the existence of two relevant issues when you try to do risk assessment: One is, what is the probability that the bad thing is going to occur?, and the other one, what is the bad thing? And probability runs on a common metric, but the nature of the bad things are all over the map, as you pointed out. You can find stunning differences in individual patient's preferences, about their fears, phobias, etc. Some people have a morbid fear of dying of cancer and will accept a sky-high risk of death from coronary heart disease rather than a very low risk of death from cancer for their own peculiar reasons. I think the attempt to try to homogenise the qualitative nature of adverse reactions is doomed to fail, once you confront the patient, because you find too much diversity.

C. Martínez:
I said at the beginning that a prospective is necessary, because the prospective for such an evaluation is crucial. I have done this study from a population base, from the healthcare or from a regulatory standpoint: If we had no paracetamol today, none at all, but we had to replace this drug with any other, that is what I would expect in this population. So that is crucial, because otherwise you are doing such an evaluation, and someone would say that is nonsense because my patient would also decide differently. Referring to your first point, I think that most of the reactions, besides gastrointestinal (GI) bleed, are not dose-dependent. They are idiosyncratic reactions or more of an allergic origin. And there is no difference if someone is taking probably 500 or 1000 mg of paracetamol in the last days. It would matter in GI bleed because is a type-A reaction, where there is a dose-dependence relationship. But if you see the order of magnitude at the end for this example, it is almost 80 times more. That means that even if we had patients complying fully with the drugs that were given to them, we will not really expect a higher difference. Therefore, I do not think this would invalidate this risk assessment. The order of magnitude would still be there, no matter if they had taken half of them, one-fourth or maybe double of it.

T. Salmonson:
Your approach is very interesting, and certainly a step forward to just comparing frequencies. But, did I understand you correctly when you said that you only include significant adverse

events? Is there not then a risk that you would penalise well-established drugs with one or two major adverse events, compared to a less studied drug with more severe adverse events? If that is the case, why not use for instance the upper 95% confidence interval level instead? In this way, it is possible to take into consideration the uncertainty in something that is not very well-established instead of using just the things that we know exist.

C. Martínez:

If we use the upper limit of the confidence interval, then you can imagine that for drugs that maybe have a relative risk of one, but mentioned in studies that had almost no power, the point estimate could be going from 0.1 to maybe 95, and this 95 would go incompletely and overwhelm all the other reactions. I had to make this decision when we extensively discussed this problem, and the final decision was to say I will only take those adverse reactions where we had studies with the power to detect a difference.

T. Salmonson:

That is precisely what worries me, because you are saying that with your approach there is not a risk. You are neglecting this, when you know that there is a potential, at least from the statistical point of view, that it could be 95. But now you are saying that if it does not exist, you do not take that into consideration. The more conservative approach, which I suggest, would in fact give credit to the more established drugs, where we better know the problems.

C. Martínez:

It is a quite interesting point, and if I were to recalculate all these data, it would distort the whole evaluation, going probably in a different way. And this would also harm the drugs that have been very well studied. Where we have much more power, we have thousands of cases often, and the drug which is established and known to cause for example GI bleed like diclofenac, would have a small confidence interval, where another one could not be studied because it is not that often used, just because of the power problem, it would go completely up in value and give a value to other. If you remember the incidence rates of those reactions that were very rare, they would have more value than the ones that are occurring quite frequently.

E.A. Kalso:

You referred to assessing also the benefit and, for a clinician who is looking after pain patients, assessing the risk/benefit ratio is of utmost importance. The drug doses you have analysed are not equi-analgesic. Therefore we should assess the risk for the benefit, and if the patient does not get any benefit from paracetamol, it is no use even if it has no risks, whereas ibuprofen, which is safer than diclofenac, is significantly more potent.

C. Martínez:

The reason I had to choose these doses was, as I said at the beginning, driven by what is published. If we had studies that would clearly differentiate, let's say, less than 500 mg or more than 2 g of paracetamol, it would be much easier. But I think in the publications we were happy to find anything dose-dependent, age-dependent, sometimes gender, but I could not go more in that, because the numbers were too small and the confidence intervals would go very far apart. You decide which drugs you think the patient would benefit and then among

those you look at the risk, if you have a choice. If you have no choice then there is no reason of comparing.

M. Pirmohamed:

I think you do a disservice to drugs such as aspirin, because clearly some of the reactions are dose-dependent. On the other hand, the risks that you estimated are dependent, to a great extent, on the quality of data that is available to you from the published studies. And clearly there are a lot of data out there which have not been published. For example, with aspirin you said there is a small additional excess risk of death from agranulocytosis, but aspirin is not usually associated with agranulocytosis. I was surprised to see there was no increased risk of Stevens-Johnson syndrome or toxic epidermal necrolysis with non-steroidal anti-inflammatory drugs such as diclofenac. If you look at some of the studies which have been done with non-steroidal anti-inflammatory drugs, (Roujeau JC *et al.*, N Engl J Med, 1995), you will see that NSAIDs are high up on the list of drugs causing these side effects.

C. Martínez:

If you take the Jean-Claude Roujeau's paper, you will not see an increased risk there for diclofenac or aspirin. Steroids do have, and the paracetamol-increased risk is taken from this paper. This is the only major epidemiologic study that we have so far for severe skin reactions for the drugs that we have looked at. I have not looked at all the drugs that they studied, so I have looked at those I wanted a priori to compare with each other. But the numbers are taken from there.

M.M. Reidenberg:

Is there a statistical way to correct for your risk estimate from the doses that are in the literature to what most of us would accept as equi-effective doses?

C. Martínez:

A short answer would be no. And the second answer would be, what would it mean? Even if the doses were not correct, what would you expect to change from the risk profile?

M.M. Reidenberg:

Using a different example, where the increase of risk was dose-dependent, then one could readily see that it would change.

C. Martínez:

That would be true for the drugs that have anti-inflammatory action through a COX inhibition. If the choice is between an anti-inflammatory drug and a non anti-inflammatory drug for the risk, then no matter which doses are used; if the drugs are both effective, then the drug without anti-inflammatory would be preferential. But if you are generalising your concept outside of this class of drugs, I had to go by what was published. Most studies, because of power problems, needed to make some cut point, although they are not equipotent from the activity, unfortunately. And I cannot imagine any measure to say we can adjust for, as we have no dose-dependency to certain defined adverse reactions.