Meta-analysis of irinotecan +/- 5-FU chemotherapy outcomes for advanced metastatic colorectal cancer

NHMRC Clinical Trials Centre - University of Sydney, March/April 2010

Context

The DaVinci study was a randomised open label phase II trial of irinotecan alone or in combination with 5-flurouraocil (5-FU) for the treatment of pre-treated advanced metastatic colorectal cancer. It was hypothesised that the addition of 5-FU to irinotecan based therapy would reduce the percentage of patients who experienced serious diarrhoea without influencing effectiveness. Secondary outcomes were safety, tumour response, progression free survival and overall survival.

Student contribution

- Analysis of the DaVinci study, summarisation of outcomes
- Meta-analysis of published studies investigating treatment regimens used in DaVinci
- Detailed investigation of a method capable of combining evidence from randomised with non-randomised single arm studies, and its application to DaVinci.

Statistical issues

1. Perform meta-analysis of randomised trials addressing the DaVinci question.
   a. Extraction of estimates from published data (including approximation)
   b. Fixed and random effect models of treatment effects in randomised trials

2. Summarise non-randomised and single arm trials of treatment arms DaVinci.
   a. Use of meta-analysis theory to pool estimates of treatment outcomes.

3. Methods to combine non-randomised trial data into meta-analysis estimates.
Supervisor:
Prof. Malcolm Hudson
NHMRC Clinical Trials Centre, University of Sydney

Declaration
I declare this project is evidence of my own work, with direction and assistance provided by my project supervisor. This work has not been previously submitted for academic credit.

Chris Brown
30 June 2010

Supervisor Statement
This supervisor's involvement in the project was conducted over a period of three months, with Chris Brown reporting weekly on his progress. The project is related to work Chris conducted, as described in this report, but required additional research work to be squeezed into an already heavy work schedule. Chris worked progressively, in an independent manner, towards this project's completion. His efforts required review and insight into the statistical methods of random effects meta-analysis, as well as clinical trials research methodology. The synthesis of estimates from different classes of study is challenging and develops BCA training; the topic was well chosen. The mathematical content in random-effects methods raised a number of issues, such as the ML estimation of parameters, that required significant reading and research along the path to satisfactory implementation of the methods selected. A consistent engagement with these topics and responsiveness to my comments and suggestions led to a good understanding being developed and successful application. Chris exhibited strong computational skills that assisted his implementation of theoretical methods. Overall, Chris Brown demonstrated independent thought, application, perseverance, and broad skills development in bringing the project to a successful conclusion.

Malcolm Hudson
3 July, 2010

Related Publication
The work discussed in this report has since been published:

Clarke, S. J., S. Yip, et al. (2011). "Single-agent irinotecan or 5-fluorouracil and leucovorin (FOLFIRI) as second-line chemotherapy for advanced colorectal cancer; results of a randomised phase II study (DaVINCI) and meta-analysis." Eur J Cancer.
1. Project description

1.1 Cancer therapy

Colorectal cancer is treated with surgery to remove cancer, radiotherapy to directly kill cancer cells and chemotherapy, which is toxic and circulates through the entire body, systemically interrupting normal cell cycles. Often the cancer has spread from the colorectal region by the time it is diagnosed or during the initial treatments and this is known as metastatic disease.

In recent times, the most common first line chemotherapy used is called 5-flourouracil (5-FU) as this has been shown to provide good outcomes for patients with colorectal cancer. However, this cancer can also be relatively aggressive and can reappear or worsen despite the first attempts of treatment. In these cases a second line of chemotherapy treatment (sometimes referred to as salvage therapy) is usually attempted to stop tumour growth or at least delay deterioration and improve remaining quality of life.

Clinical trials are experiments conducted in humans and are the building blocks of medical research. In trials addressing cancer, participants are usually treated with new and promising treatments or in ways of administration which may improve their cancer outcomes (efficacy) or side effects (safety). In some scenarios new treatment options may be equally effective as the existing standard but may present potential improvements in safety outcomes or simply be cheaper or easier to administer.
In clinical trials side effects of treatment (adverse events / toxicities) are measured and classified against an agreed global standard developed by the US National Cancer Institute, the Common Terminology Criteria for Adverse events [1]. This enables comparison across studies within and between countries and over time.

1.2 DaVinci Study

The DaVinci study, hereafter referred to as the study, was an open-label randomised phase II trial of irinotecan chemotherapy alone (every three weeks) or in combination with 5-FU (every two weeks) for the treatment of advanced metastatic colorectal cancer previously treated with chemotherapy.

Serious (Grade 3 or 4) diarrhoea is debilitating (Table 1) and is experienced by a substantial proportion of patients (~20%) treated with single-agent irinotecan therapy. Usually the addition of a second chemotherapy would result in additional toxicity. Possibly due to the reduced dose level of irinotecan, or an interaction in the way the two drugs act together, there was anecdotal evidence that this combination would in fact reduce diarrhoea. It was therefore hypothesised that the addition of 5-FU to irinotecan alone would reduce the percentage of patients who experienced serious diarrhoea whilst not impairing survival.

Table 1 – Serious Diarrhoea

| Grade 3 | Increase of ≥ 7 stools per day over baseline; incontinence; IV fluids ≥ 24hrs; hospitalization; severe increase in ostomy output compared to baseline; interfering with ADL (activities of daily living). |
| Grade 4 | Life-threatening consequences (e.g. hemodynamic collapse). |
Irinotecan is an easy treatment to administer. In contrast, treatment with 5-FU is an inconvenient therapy because it is received more frequently and requires IV administration over a number of hours. The addition of 5-FU would be considered worthwhile only if the combination sufficiently reduced the rate of diarrhoea without compromising the other outcomes with the additional burden and risk of complications.

Secondary outcomes of the DaVinci study were therefore: general patient safety (all toxicities), tumour response rates (shrinkage or disappearance), progression free survival (PFS) and overall survival (OS). Tumour response, PFS and OS are common efficacy measure of clinical trials in this setting.[2]

This study was designed and begun prior to the availability of new biological agents which became potential treatment options for these patients. These agents were introduced and became available during DaVinci recruitment and patients and clinicians were understandably attracted to them in preference to participating in this study. The study hence had difficulty recruiting and eventually closed with 78 patients. This sample size was not the 100 required to reliably detect the smallest worthwhile difference in toxicity (12%) but reasonably accurate estimates of treatment effects were expected.

The reduced size of the trial reduced the probability of statistically detecting the difference if it did exist for diarrhoea and the other endpoints. During the course of the trial some independent studies had been published partially addressing the same question and so there was potential for a meta-analysis to pool estimates and possibly provide the additional power required to conclusively answer the DaVinci questions.
1.3 Meta-analysis

This project aims to summarise the key outcomes of the DaVinci study and other studies which have been conducted using these agents in order to determine whether it is possible to definitively answer whether irinotecan in combination with 5-FU reduces the rate of diarrhoea for these patients.

Since 1991, a large number of studies have been conducted using these treatment arms either individually or as part of randomised trials addressing various clinical questions. In some randomised studies, single agent irinotecan was used originally as the experimental arm, in other later studies as a control arm. Additionally, some studies were conducted with multiple arms of the same treatment administered with differing doses or scheduling.

All clinical trials are subject to what is known as selection bias - where the enrolment of a specific sample of patients’ generally better (or worse) prognosis patients restricts the generalisability of the results to the applicable (target) population. This is compensated for in randomised studies since (if randomisation is not compromised) the groups are identical apart from (a) chance differences and (b) the effect of treatment. This enables an unbiased comparison for the target population defined by the trial eligibility criteria.

Single arm studies lack such a randomised control group and, since many factors can affect observed results, they are difficult to interpret. They often leave unanswerable question marks about the results. Hence, for important decisions regarding changes in clinical practice, randomised evidence is preferred.
It is important to note that generalisability remains a concern in the interpretation of randomised studies. A description of the study population for which the treatment estimates are valid can be quantified by summaries of sample, demographics and baseline factors which influence treatment and outcomes. The complete impact of these factors is impossible to quantify entirely this is an even larger concern in single-arm (and non-randomised) studies where there is not control group and why they can not provide direct estimates of the treatment effect in the target population.

In the case of DaVinci’s exploration of whether irinotecan should be used alone or in combination with 5-FU, there are a substantial number of non-randomised studies which have used these treatments in the target population. These studies, despite their limitations, may still contribute valuable information to the interpretation of the randomised trial results.

2. Data Management

2.1 Trial data

The DaVinci study was an investigator led study conducted by the Australasian Gastro Intestinal Trials Group (AGITG) at the NHMRC Clinical Trials Centre (CTC). The trial was conducted at 15 sites across Australia and recruited patients between June 2005 and January 2008. The study was approved by the University of Sydney ethics committee as well as ethics committees at the participating hospitals, and all participants gave written informed consent to participate.

The data for the trial was collected on pre-specified data forms by site investigators and data managers, entered by a number of CTC staff who also performed a
comprehensive data query and quality assurance process. I performed the statistical analysis in late 2008 and updated this and the trial report in 2009. The trial manuscript is in the process of being written and the results of this project will be summarised and inform the context of the interpretation.

2.2 Systematic review

The literature search was performed by a colleague at the CTC in 2009. This comprehensive search located studies which had used irinotecan in the treatment of second line therapy for colorectal cancer. The search was restricted to studies in which 5-FU was administered via intravenous infusion (as opposed to via a bolus) as this method of administration was thought to be substantially different from the clinical perspective.

Published journal articles were located as well as a number of abstracts. After consideration it was decided that it was important to include abstracts (despite not being peer-reviewed literature) as they provide information on otherwise unpublished trials, minimising publication bias (interesting or ‘significant’ results are more likely to be published). Abstracts which were later published were excluded in favour of the full publication. Unfortunately, abstracts can be limited in size and may contain only brief summary statistics and often entirely lack survival information. The literature review was conducted with limited time and resources and no attempt was made to contact authors to obtain any missing information.

Extraction of the important summary data was set up and coordinated by me. I performed independent validation of the extraction and calculated the required summary statistics in Microsoft Excel.
3. Statistical Methods

The statistical methods of this project broadly fall into three categories:

1. The extraction (and standardisation) of published treatment outcomes
2. Use of meta-analytic theory to pool estimates of treatment effects
3. Identification and use of a valid method to combine single arm trial data into standard meta-analysis

3.1 Extracting estimates of treatment from published data

Two types of outcome data are under investigation: proportions (toxicity and response rates) and survival (times, potentially right censored). In general, sufficient detail to perform a meta-analysis can often be obtained directly from tables or manuscript text. This is particularly true for dichotomous data where the standard error of the proportions is directly related to the proportion of successes and the number of subjects contributing to the estimate. In randomised studies, odds-ratios for the comparison of the proportions in the two arms can be easily calculated directly from the number of successes/failures in each group (if not otherwise provided).

Parmar et al. [3] summarise a number of ways to extract and estimate key information from published studies. These methods were applied here.

3.1.1 Data extraction (survival data)

When the outcome of interest is the time to some event it is possible that, at the conclusion of the study, not all patients will have experienced it. One option is to exclude the patients however this can lead to biased results. A better option is to
include patients in the analysis in such a way which incorporates the fact that they have not had the event after some known period of follow-up.

Methods to incorporate the right-censored nature of time-to-event data are relatively recent [4] compared with the history of data analysis in general [5]. Since the development of the Kaplan-Meier survival estimate, the median survival time has become a meaningful summary statistic in the context of right-censored data and is widely reported. Plots of the Kaplan-Meier estimate over the relevant time range are often presented. However, it is understandable that there is still significant variation in the way time-to-event outcomes are reported especially considering the general issues in the area of study reporting quality [6]. For example, some trial still do not routinely report appropriate 95% confidence intervals (CI) for means, proportions, odds-ratios, let alone median survival times (and, if provided, rarely specify the method used in its calculation).

For meta-analysis of survival outcomes, it is an important concern that summary statistics can not be accurately replicated without access to individual patient data. The success of any meta-analysis therefore depends largely on the level of detail that is provided in trial manuscripts. However, if the authors have presented a survival curve, Parmar et al [3], show that it is possible to extract a reasonable estimate of the event times based on the co-ordinates of the curve and in turn estimate the summary statistics required. Their approach has been adapted for the purposes of this meta-analysis.

**Estimation of HR**

In randomised studies, the estimate of the unadjusted hazard ratio (HR) is widely accepted as the best comparison of the failure rates of the two arms. The Cox-
proportional hazards model is used to estimate the HR and test whether it differs significantly from 1 (no difference). Michiels et al. [7] note that, when necessary, a hazard ratio can be estimated as the ratio of two median survival times. It is also easy to show that where a p-value is reported for a Cox regression model, the standard error of the estimate log-HR (LHR) can be approximated by calculating the inverse of the test statistic and this is one of the first methods recommend by Wiebe et al. [8] when the variance is not supplied. Both these methods were used to summarise trial data for this project.

**Survival proportions**

In order for outcome estimates from single arm studies to be pooled (and possibly incorporated into a meta-analysis) a reliable method of pooling individual survival estimates is required. As discussed further in section 3.2, it is not possible to easily pool estimates of survival medians as they do not have the same additive properties as the mean.

An alternative is to consider the estimated survival at a fixed time point (e.g. 6/12 months) as this summary statistic will still provide an estimate of the survival distribution but have attractive properties for the purposes of pooled estimates. The estimate of the survival probability at a fixed time point can easily found by determining the height of the Kaplan-Meier curve at that time.
Kaplan-Meier curves may not be available for all studies. If the distribution of the survival probability is assumed to be exponential then it is possible to estimate the parameter, $\lambda$ (=0.1 above), given a survival estimate at some time point. By using this $\hat{\lambda}$ the survival at time $t$ can be estimated.

**Equation 1 – Estimation of Survival at time $t$**

Assuming $S_t = e^{-\lambda t}$, then $\lambda = -\frac{\log_e(S_t)}{t}$

Estimating $\hat{\lambda}$ with $t_{0.5}$ such that $S(t_{0.5}) = 0.5$ (median)

We have: $S_t = e^{-\frac{\log_e(0.5)}{t_{0.5}}}$

In the example above, $\hat{\lambda}$ could be estimated as $-\log_e(0.5)/6.85 = 0.1$. This can be used in equation 1 to confirm $S(3) = \exp(-0.1*3) = 0.74$. This function is used to estimate progression free survival at 3 months (PFS3 - Figure 1 – General survival distribution) and overall survival at 6 months (OS6) for the studies where Kaplan-Meier curves are not available.

The standard error of these estimates need to incorporate the degree of information contained in the curve – the number of patients “at risk” at the desired time. This number is a function of the sample size and the censoring distribution and is
estimated as part of the Parmar method. Given the nature of this project, the study sample size has been used as a crude estimate of the accuracy. This should be a reasonable substitution as it represents the initial number at risk and will be consistent across studies under the assumption that all studies under investigation will have a similar pattern of censoring.

### 3.1.2 Data standardisation (toxicity reporting)

The trials which met the inclusion criteria were published over a number of years and from a number of international regions. It is therefore expected that there will be some variation in the approach to reporting study outcomes. This is particularly the case for toxicity rates where some studies report the percentage of patients who had any toxicity and others report the percentage of total affected cycles.

The importance of the difference between these two statistics can be seen when one considers that chemotherapy is toxic and administered in cycles. If all patients received only one cycle of treatment the statistics will be identical. If all patients receive the same number of cycles the two proportions will be similar as each patient will have the same exposure or ‘risk’ of having the toxicity. In practice, patients will have varying number of cycles. Additionally, for some types of toxicity the effects of chemotherapy are cumulative. The patients who have more cycles of treatment will be more likely to experience toxicity and have more weight in an analysis based on cycles.

A simple approach to the problem would be to exclude toxicity estimates from studies which reported the outcome as the percentage of affected cycles.
However, depending on the homogeneity within these study ‘types’ it may be possible to adjust the treatment estimates in an unbiased fashion.

It is also worth noting at this point that the length of treatment cycle may contribute to the likelihood of toxicity. It is possible that the reduced size of single doses could improve side-effects but this could be offset by the higher frequency of treatment. Similar adjustments may be possible / required when considering differing treatment cycle lengths and this may be a factor here as there was experimentation with both irinotecan and 5-FU being administered as a weekly regimen by Rothenberg [9] and Tsavaris [10]. The different treatment cycle lengths will be summarised as part of the meta-analysis to identify any heterogeneity in outcomes.

### 3.2 Combining study estimates

Where a number of studies have been conducted (with common treatment arms) it is possible to pool the estimates of their treatment effects to determine the overall pooled effect. This is possible providing all studies are designed addressing the same outcome (measured under similar conditions). There is substantial development of these methodologies [11-12] and since they essentially rely only on having valid weights for use in a weighted mean, they can be applied to the pooling of estimates of any type. There are two main groups of models, fixed and random effects, which are discussed in detail below.

These methods are commonly used to pool comparative treatment effects (e.g. odds ratios) from randomised studies but, notwithstanding the prior discussion of generalisability and selection bias, could also be used to pool treatment outcome estimates of single arm studies (e.g. toxicity rates) as required here.
Meta-analysis models

The basic fixed effects meta-analysis model is a weighted mean of the study estimates where the weights are the inverse of the respective standard errors.

Equation 2 – Fixed effects model (inverse variance)

\[ \theta_{IV} = \frac{\sum w_i \theta_i}{\sum w_i}, \quad SE(\theta_{IV}) = \frac{1}{\sqrt{\sum w_i}} \quad \text{where} \quad w_i = \frac{1}{SE(\theta_i)^2} \]

\( \theta \) is the treatment effect in the i'th trial and \( \theta_{IV} \) is the inverse Variance estimate

The random effects meta-analysis model incorporates a parameter of between study variance which is set to zero if the heterogeneity statistic, \( Q \), is less than \( k-1 \) (in which case, it becomes the inverse variance fixed effects model.

Equation 3 – Random effects model (DerSimonian and Laird)

\[ \theta_{DL} = \frac{\sum w_i \theta_i}{\sum w_i}, \quad SE(\theta_{DL}) = \frac{1}{\sqrt{\sum w_i}} \quad \text{where} \quad w_i = \frac{1}{SE(\theta_i)^2 + \tau^2} \]

\[ \tau^2 = \frac{Q - (k-1)}{\sum w_i - \sum w_i^2 / \sum w_i}, \quad Q = \sum w_i (\theta_i - \theta_{IV}), \quad k = \#\text{ studies} \]

Assumptions of meta-analysis

There are few assumptions which underlie the general meta-analysis methodology. One common one, used to simplify the models required, is that the treatment effect is assumed to remain constant between studies and the differences between studies are due to random error.

This assumption can be relaxed and the treatment effect common to all studies can be considered a random variable, of fixed variance, as described by DerSimonian
and Laird [13]. This methodology can be applied to any estimates of effect where a valid variance is computed or other appropriate weighting exists.

The R program (v2.7.1) [14] was used with Thomas Lumley’s RMeta package (v2.1.6) [15] to calculate the required estimates and create necessary plots.

**Time to event outcomes**

In this review the two time-to-event outcomes are of high importance to the overall interpretation of the toxicity results. One of the aims of this project is to give consideration to the possible ways of pooling estimates of survival and to evaluate the potential for including single arm studies to improve the meta-analysis effect estimates. It is possible to quantify and compute the variance of a median but the interpretation and their impact on the meta-analysis of this is difficult to quantify.

Pooling median survival would not make sense by the standard methods. The proposed method of calculating PFS3 and OS6 probabilities enables the simple use of the meta-analysis methods and an easy interpretation.

**3.3 Combining single arm studies into meta-analysis**

The paper by Begg and Pilote [16] details an approach to combine estimates of single arm studies into meta-analysis of randomised studies in such a way that the weight of their influence is dependent on the overall weight of evidence between studies compared with that within studies.

It is very similar to the method of DerSimonian and Laird random effects model but based on the treatment outcomes in each arm of the study (e.g. x and y, with their
respective standard errors, $S_x$ and $S_y$) as opposed to the treatment effect (e.g. $x-y$, and the appropriate standard error for this difference).

**Equation 4 – Begg and Pilote formulation**

\[ x_i \sim N(\theta_i, s_i(x)^2) \quad i = 1, \ldots, n; \]
\[ y_i \sim N(\theta_i + \delta, s_i(y)^2) \quad i = 1, \ldots, n; \]
\[ u_i \sim N(\theta_i, s_i(u)^2) \quad i = n + 1, \ldots, n + k; \]
\[ v_i \sim N(\theta_i + \delta, s_i(v)^2) \quad i = n + k + 1, \ldots, n + k + m; \]
\[ \theta_i \sim N(\mu, \sigma^2) \]

Their model incorporates single arm estimates of each treatment arm $u$ and $v$ in the estimation of the mean, $\mu$, and the difference between the treatment arms, $\delta$.

There are two extremes to the weighting they propose:
- Where between study variance ($\sigma^2$) is significantly larger than the within study variances the method reverts to weights based on the standard inverse variance weights including all studies.
- If the between study variance ($\sigma^2$) is negligible the method is designed to minimise the amount of weight assigned to the single-arm studies and becomes the estimate which would be obtained based on only the randomised studies (inverse variance weighted random effects model).

The estimates of *baseline treatment outcome* ($\mu$) and *treatment effect* ($\delta$) can be obtained by solving the equations (see appendix 2) for the maximum likelihood. The method requires a starting value of $\sigma^2$, the variance of the random treatment outcome assumed to be a component of all treatment arms. The likelihood can then be solved for $\sigma^2$ based on the estimated values of $\mu$ and $\delta$ in an iterative procedure.
4. Results

4.1 DaVinci results

The DaVinci trial recruited 78 patients and its primary outcome was the rates of serious diarrhoea with the addition of 5-FU to an irinotecan based treatment.

Table 2 – Results of DaVinci

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Single agent (95%CI)</th>
<th>Combination (95%CI)</th>
<th>Comparison (95%CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serious diarrhoea</td>
<td>18.6% (10,33)</td>
<td>9.5% (4,21)</td>
<td>OR=2.16 (0.60,7.78)</td>
</tr>
<tr>
<td>Any 3/4 Toxicity</td>
<td>48.8% (35,63)</td>
<td>47.6% (33,62)</td>
<td>OR=0.95 (0.41,2.23)</td>
</tr>
<tr>
<td>Response Rate</td>
<td>11% (2,21)</td>
<td>11% (2,21)</td>
<td>OR=1.0 (0.27,3.73)</td>
</tr>
<tr>
<td>PFS (median)</td>
<td>4.0m (2.7,5.7)</td>
<td>6.2m (5.4,6.7)</td>
<td>HR=0.81 (0.52,1.25)</td>
</tr>
<tr>
<td>OS (median)</td>
<td>11.2m (8.3,13.3)</td>
<td>14.9m (8.1,19.3)</td>
<td>HR=0.72 (0.46,1.12)</td>
</tr>
</tbody>
</table>

There was a definite trend toward reduced serious diarrhoea in the combination arm (p=0.24). The rate of any serious toxicity was similar in both arms, as was the rate of tumour response. There appears to be a reduced risk of progression and slightly longer survival on the combination arm however these effects are unexpected and are unlikely to be a reflection of any true benefit in these endpoints.

Figure 2 – DaVinci Time to progression
The 3 month progression free survival estimates can be determined from the Kaplan-Meier estimates. For illustrative purposes, the respective estimates were also calculated from the median survival (Table 3). The approximation estimates are reasonably similar but will not be used as Kaplan-Meier estimates are available.

Table 3 – Estimation of PFS3/OS6 for DaVinci

<table>
<thead>
<tr>
<th></th>
<th>Kaplan-Meier</th>
<th>Estimate from median</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Single</td>
<td>Combination</td>
</tr>
<tr>
<td>PFS3</td>
<td>64%</td>
<td>80%</td>
</tr>
<tr>
<td>OS6</td>
<td>72%</td>
<td>72%</td>
</tr>
</tbody>
</table>

Despite being able to calculate the standard error of these proportions in the usual way [17] the estimates used for the meta-analysis will be based on the number at risk at the start of the trial (so they are consistent with those from the other studies).

4.2 Literature review / data extraction results

The literature review identified two studies [18-19] which were randomised comparisons similar to DaVinci. Meta-analysis of these three studies is possible but is potentially limited in its ability to conclusively convince people if the studies exhibit sufficiently differing results. There were an additional 25 single arm studies of single agent irinotecan and 3 additional studies of the combination chemotherapy.

Seymour Study

The Seymour trial [19] was a large, 5 arm randomised study in the UK, concurrently addressing a number of questions in first and second line treatment. Various combinations of 5-FU, irinotecan and Oxaliplatin were tested. The first two arms of
this study were both treated with 5-FU (first line) and each was treated with one of
the DaVinci arms following progression (second line). In the second line of these
groups, there were 364 patients in the single arm and 185 patients treated with
combination.

Table 4 – Results of Seymour

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Single agent (95%CI)</th>
<th>Combination (95%CI)</th>
<th>Comparison (95%CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serious diarrhoea</td>
<td>16% (12,20)</td>
<td>8% (4,11)</td>
<td>OR=2.32 (1.25,4.27)</td>
</tr>
<tr>
<td>Response Rate</td>
<td>11% (8,14)</td>
<td>16% (11,22)</td>
<td>OR=0.62 (0.37,1.04)</td>
</tr>
<tr>
<td>PFS (median)</td>
<td>4.3m (not avail.)</td>
<td>4.4m (not avail.)</td>
<td>HR=0.98 (0.85,1.13)</td>
</tr>
<tr>
<td>OS (median)</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
</tbody>
</table>

There was a statistically significant reduction in risk of diarrhoea in patients
receiving the combination chemotherapy (p=0.007). The study reported toxicity,
response rates and progression free survival specifically for second line treatment
but only reported overall survival (OS) with respect to the upfront treatment plan
(i.e. from first line therapy).

There were no direct comparisons made between the arms of interest however it
was possible to calculate odds ratios directly. Unfortunately, no survival curves
were available from which to estimate the Hazard ratio using the method of Parmar
et al. but, as discussed in section 3.1.1, it is possible to estimate the HR using the
ratio of the median survival times.

Equation 5 – Estimate of HR for Seymour study

\[
HR \approx \frac{S_{0.5\text{ SingleAgent}}}{S_{0.5\text{ Combination}}} = \frac{4.3m}{4.4m} = 0.977
\]
It is also possible to estimate the standard error (SE) of the HR using the supplied p-value for the treatment comparison since the p-value is based on the Wald test statistic.

Equation 6 – Estimate of SE(LHR) for Seymour study

\[ P(z > \frac{\text{Ln}(0.977)}{SE(LHR)}) = 0.75 \]

\[ 0.02299 \div SE(LHR) = 0.1015 \]

\[ \therefore SE(LHR) = 0.2265 \]

where, \( LHR = \text{Ln}(HR) \)

These estimates enable calculation of the 95% confidence interval around the hazard ratio however it is noted that there is significant error likely brought about by the combination of these estimation methods.

Graven Study

The Graven study compared the two treatment combinations but used a weekly schedule of administration of both. With a total of 55 patients it was slightly smaller than the DaVinci study. The weekly schedules may have influenced the treatment outcomes compared with the other studies however the comparison between the two arms should still be approximation of the effect of the addition of 5-FU.

Table 5 – Results of Graven

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Single agent (95%CI)</th>
<th>Combination (95%CI)</th>
<th>Comparison (95%CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serious diarrhoea</td>
<td>19% (4,33)</td>
<td>11% (0,22)</td>
<td>OR=1.89 (0,41,8,85)</td>
</tr>
<tr>
<td>Response Rate</td>
<td>11% (0,32)</td>
<td>11% (0,32)</td>
<td>OR=1.04 (0,19,5,68)</td>
</tr>
<tr>
<td>PFS (median)</td>
<td>3.7m (3.1,7.8)</td>
<td>3.7m (2.7,5.2)</td>
<td>HR=0.87 (0,48,1,55)</td>
</tr>
<tr>
<td>OS (median)</td>
<td>10.7m (8.0,12.9)</td>
<td>9.5m (6.5,12.6)</td>
<td>HR=1.31 (0,68,2,52)</td>
</tr>
</tbody>
</table>

This study also suggested a benefit in the combination arm with respect to diarrhoea. All other outcomes were approximately the same in both arms.
4.2 Meta-analysis results

Meta-analysis

The Seymour study alone provides substantial evidence that diarrhoea is reduced by using the combination treatment. However, their result is one of many presented in their analysis (no adjustment for multiple comparisons) and even they are apprehensive about making conclusive statements [19].

The sample size of the Seymour study dwarfs the others in the comparison and so their estimate has the dominant weight in the pooled estimate. Despite this the estimates from Graven and DaVinci are consistent with the Seymour result and the pooled effect (OR=2.24) is more accurate, 95% CI (1.33, 3.76). There was no significant heterogeneity (p=0.97), and the random effect variance estimate is sufficiently small that it could be excluded and considered a fixed effects model.

Figure 3 – Meta-analysis of diarrhoea - randomised studies

Summary OR= 2.24  95% CI ( 1.33,3.76 )
Test for heterogeneity: X^2( 2 ) = 0.06  ( p-value 0.9707 )
Estimated random effects variance: 0.0
Meta-analysis output and graphs from the other endpoints is supplied in Appendix1.

In summary the there was no significant difference between the two arms for any of the other endpoints. However there was a trend towards higher response rate driven by the results of the Seymour study.

**Table 6 – Meta-analysis results (randomised studies)**

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Studies</th>
<th>Comparison (95% CI)</th>
<th>Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response Rate</td>
<td>All 3</td>
<td>OR = 0.68 (0.43, 1.08)</td>
<td>p = 0.7</td>
</tr>
<tr>
<td>PFS (3 month)</td>
<td>All 3</td>
<td>HR = 0.96 (0.84, 1.09)</td>
<td>p = 0.7</td>
</tr>
<tr>
<td>OS (6 months)</td>
<td>DaVinci + Graven</td>
<td>HR = 0.92 (0.51, 1.67)</td>
<td>p = 0.1</td>
</tr>
</tbody>
</table>

**Single arm Studies**

As expected, there was a large number of studies (21) conducted investigating single agent irinotecan [9-10, 20-39] in this setting. There were fewer studies (3) of combination therapy which fitted our criteria [40-42].

**Figure 4 – Summary of single arm studies**

- Green points are combination estimates, red/orange are single agent.
- Numbers next to points relate to study identification (order of extraction).
The funnel plot of treatment outcomes (figure 4) is enables assessment of potential publication bias and visual representation of the estimates available from all studies. The estimates in the randomised studies (ID-0: DaVinci, ID-10: Graven, ID-12: Seymour) are joined to illustrate the fact that with similar sample sizes the standard error of the single agent estimates are larger, in part, due to proportions closer to 0.5 (where SE is maximised for a given sample size). The Seymour study (ID-12) had twice as many patients in the single agent arm and this created similar standard errors of the estimates in its two arms.

Observing this plot there is an obvious trend towards higher rates of diarrhoea in the single agent treatment trials. One approach to analysis is to combine the single arm studies using the usual meta-analytic methods (Figure 5). Doing this estimates the single agent percentage of serious diarrhoea to be as 23.5%, 95% CI (20%, 27%). This is significantly higher than in the combination arm where the pooled effect of the 6 studies is 8.4%, 95% CI (6%, 11%).

There is a large degree of heterogeneity between the single agent studies and this appears to be primarily driven two important factors. Firstly, effective diarrhoeal treatment became available during the early 90’s and this is reflected in the studies which were designed before then. Secondly, there appears to be a trend towards higher toxicity in studies of weekly schedules. This is discussed further below.

It is possible to compare the pooled treatments estimates based the assumption that the standard error of the difference is equal the sum of the variances. This can be used to calculate a z-statistic. The reduction is therefore estimated as 15.1%, 95% CI (10.5, 19.7), p<0.001, however this comparison ignores correlation of randomised studies is biased by heterogeneity factors already mentioned.
Equation 7 – Significance test of treatment difference

\[ SE_{diff} = \sqrt{0.0189 + 0.0138} = 0.0234, \quad Z = \frac{0.235 - 0.0839}{0.0234} = 6.45, \quad P(z > 6.45) < 0.001 \]

Figure 5 – Pooled estimates of diarrhoea rates
*Study name format: Year – Author – Irinotecan dose (mg), day of cycle – Cycle length

### Diarrhoea (Single Agent)

Study Reference
- 1996 - Rothenberg - 125 d1 - 7d
- 1997 - Pitot - 125 d1 - 7d
- 1997 - Rougier - 350 d1 - 21d
- 1998 - Ang - 350 d1 - 21d
- 1998 - Cunningham - 350 d1 - 21d
- 1998 - Rougier - 300-350 d1 - 21d
- 1999 - Aravantinos - 350 d1 - 21d
- 1999 - Frontris - 300-350 d1 - 21d
- 1999 - Rothenberg - 100 d1 - 7d
- 1999 - Rothenberg - 125 d1 - 7d
- 1999 - Van Cutsem - 350 d1 - 21d
- 1999 - Schoffski - 350 d1 - 21d
- 2000 - Schoffski - 350 d1 - 21d
- 2003 - Fuchs - 125 d1 - 7d
- 2003 - Fuchs - 350 d1 - 21d
- 2003 - Mendez - 350 d1 - 21d
- 2005 - Garcia-Giron - 250 d1 - 14d
- 2007 - Graeven - 125 d1 - 7d
- 2007 - Seymour - 350 d1 - 21d
- 2008 - Haller - 350 d1 - 21d
- 2009 - Kim - 350 d1 (300 for ecog2) - 21
- Current Study - 350 d1 - 21d

Summary

Proportion patients with Grade 3/4

### Diarrhoea (Combination)

Study Reference
- 2001 - Gil-Delgado - 180 d1 - 14d
- 2002 - Leonard - 180 d1 - 14d
- 2004 - Toumi - 180 d1 - 14d
- 2007 - Graeven - 80 d1 - 7d
- 2007 - Seymour - 180 d1 - 14d
- Current Study - 180 d1 - 14d

Summary

Proportion patients with Grade 3/4

**Single Agent**
- Summary effect = 23.5% - 95%CI (20, 27)
- Estimated heterogeneity variance: 0.006
  \( p = 0.001 \)

**Combination**
- Summary effect = 8.4% - 95%CI (6, 11)
- Estimated heterogeneity variance: 0.0
  \( p = 0.88 \)
Weekly scheduling

In the single agent studies various cycle lengths (7, 14, 21 days) had been used in the treatment administration. All studies of combination therapy with the exception of Graven used the same 14 day schedule used in DaVinci.

The American study by Fuchs 2003 directly compared weekly and three weekly schedules of single agent irinotecan in a randomised study. Despite observing similar overall toxicity rates and efficacy (PFS, OS, response) they found significantly more diarrhoea in the weekly arm (36% v 19%, p=0.002).

Figure 6 – Treatment schedule subgroups (Single agent)

Their results are consistent with the estimates obtained from our meta-analysis. For the 5 studies which administered the single-agent in a weekly cycle (with 125mg) the average toxicity rate was 30%, 95%CI (22, 38), but there was significant heterogeneity. In the remaining 15 estimates of three weekly (350mg) treatment the pooled toxicity rate was 22%, 95%CI (20, 24).

The heterogeneity of the 3 weekly estimates is still extremely significant (p<0.001) but major contributors to this are the two early studies conducted before the introduction of effective anti-diarrhoeal medications and the Seymour study which, contributing a significant (21%) percent of the total weight, has an estimate (16%) reasonably lower than the pooled one.
**Begg and Pilote method**

The method of Begg and Pilote was then used to estimate the pooled effects and treatment difference incorporating the paired nature of the randomised studies.

For this data, the derivative of the log likelihood with respect to $\sigma^2$ is 0 when $\sigma^2$, the common variance parameter, was 0.00421. This is therefore the maximum likelihood estimate of $\sigma^2$ and using the Begg equations (Appendix B) leads to estimated Single Agent rate of serous diarrhoea of 22.8% and a 12.6% reduction (to 10.3%) for combination treatment.

**Figure 7 – Begg log likelihood functions for DaVinci data**

Since the estimate of the common random effect parameter is not zero, the method also incorporates the covariance between the randomised studies. This makes this method the best unbiased estimate of for the treatment groups incorporating the single arm studies.
5. Interpretation

DaVinci Study

The DaVinci study was conducted because there was evidence in the first-line setting that the combination treatment may reduce key side-effects without having an adverse effect on the important outcomes of progression free survival and overall survival.

Due to practical constraints the DaVinci study was limited in size. Despite this, the results indicate that the combination regimen is potentially associated with a sizable reduction in diarrhoea. It was hence interesting to compare these results against other studies conducted and consider whether combining the results of these studies provided sufficient evidence of this effect.

In order to summarise the context of the DaVinci study it is worth summarising the course of research in the years proceeding. Irinotecan had become widely used in first and second line treatment for colorectal cancer [32]. As this regimen was becoming more commonly used, a range of studies addressed the question of whether the single agent could be delivered as a 7day cycle (at a lower dose than the 21day cycles) in order to reduce the toxicity. The second line-combination regimen was initially tested using this weekly schedule.

Meta analysis / Pooled estimates of key outcomes

The Graven study is one of two other studies which have compared the single agent vs. combination in randomised setting however used 7day cycles in both arms. The difference in cycle length is a confounding factor when comparing results across all studies.
The other study, by Seymour et al, was conducted in the context of first and second line treatments in attempt to determine the optimal treatment strategy for this group of patients. It evaluated a range of comparisons including whether irinotecan or oxaliplatin (another efficacious drug in this setting) was better given at first line in combination with 5-FU, or delayed until second line.

**Bias in Seymour study**

The Seymour study only reports results for patients who received second line treatment as per the randomisation. Directly comparing the single agent / combination second-line arms from this study is a potentially biased comparison. Patients were only continued on to the second-line treatment if they were considered suitable for further chemotherapy (Figure 8 – Drop out bias in Seymour study).

![Figure 8 – Drop out bias in Seymour study](image)

These patients were randomised prior to receiving their 1st line treatment (rather than immediately before 2nd line) it is therefore not possible to guarantee that
patients withdrew from 2nd line treatment independently of the treatment arm they were allocated to (patients who drop out are potentially different in the two arms).

On the single agent arm there were 251 (38%) patients who failed first line who died or were not able to receive 2nd line due to disease progression and 51 patients (12% of those receiving 2nd line) received alternative treatment to that allocated. These percentages are in different to those of the combination arm (32% and 19%) and the differences imply there is an element of bias which has been introduced (as patients and clinicians knew what the next treatment would be). Despite this, the impact on the diarrhoea endpoint is probably minor.

**Meta analysis of 3 studies**

The value of meta-analyses based on a small number of studies can be limited unless the size of the studies is sufficient to provide reliable estimates. The dominance of the Seymour study in the pooled estimates, raises the potential concerns with the interpretation of their result in our meta-analysis. This has been discussed and consensus is that these issues are not likely to affect the results.

**Summary**

The DaVinci study found that diarrhoea was reduced with the combination treatment arm however was too small for the results to be interpreted with confidence in clinical practice. While a meta-analysis of three randomised studies was possible, and supported the benefits of combination chemo in reducing diarrhoea, potential bias in the studies again reduces the confidence with which these results can be applied to individual patients by clinicians. An additional 20 single arm studies are available, but the study design limits the interpretation of results at the clinical level.
The use of meta-analysis techniques which can combine the evidence from these three sources (DaVinci, additional randomised studies and single arm trials) allows all available information to be considered when making a decision for clinical practice. The results of this meta-analysis suggest that there is a benefit in reduced diarrhoea with combination therapy.

On the basis of the results of this work, it is recommended that for patients with advanced colorectal cancer who have previously been treated with chemotherapy, the use of combination 5FU and Irinotecan can reduce the incidence of severe diarrhoea (an unpleasant and dangerous side effect) without affecting the important outcomes of overall and disease free survival. In practice, the decision would be strongly influenced by practical concerns particularly the frequency of visits required and the need for IV to receive the 5-FU.
6. Discussion of BCA project analysis

In retrospect, being my first systematic review, more data was initially extracted than was eventually necessary and this increased the difficulty and error rate of the task. The additional information did prove useful in the process of de-identification of duplicate studies and elimination of inappropriate studies from those initially selected.

There were a large number of eligible studies which were not identified during the initial search. These were discovered during validation of the extracted data from the references of some selected papers. This delayed the analysis process as articles were obtained. There were a surprising number of articles located via this method highlighting the importance of a high quality and through review process.

A great deal of time was devoted to understanding the paper by Begg and Pilote. Much of this was in setting up a program to facilitate the calculation of the necessary components of the simultaneous equation representation of the likelihood. Because the likelihood does not have simple form it was not initially clear how to calculate estimates of the variance of the common random effect. After consideration with Prof Hudson we realised that they had provided the derivative of the likelihood with respect to the parameter of interest and that it could therefore be evaluated.

Validation of their results was not straight forward as there were typographical errors in key equations relating to the derivatives of the log-likelihood. Appendix 2 contains the details a derivation of these corrected terms. This minor error means that in their example, they did not notice that in fact there exists a maximum likelihood estimate >0. A brief summary of the validation is provided in Appendix 3.
References


Appendix 1.1 - Results for secondary outcomes

Tumour Response

<table>
<thead>
<tr>
<th>Study Reference</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>DaVinci (n=88)</td>
<td>0.25</td>
</tr>
<tr>
<td>Graven (n=55)</td>
<td>0.40</td>
</tr>
<tr>
<td>Seymour (n=549)</td>
<td>0.63</td>
</tr>
<tr>
<td>Summary</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>1.58</td>
</tr>
<tr>
<td></td>
<td>2.51</td>
</tr>
<tr>
<td></td>
<td>3.98</td>
</tr>
<tr>
<td></td>
<td>6.31</td>
</tr>
</tbody>
</table>

(favours combination)   (favours single agent)

Random effects (DerSimonian-Laird) meta-analysis

Summary OR = 0.68  95% CI (0.43,1.08)

Test for heterogeneity: X^2(2) = 0.7 (p-value 0.7058)
Estimated random effects variance: 0.0

Progression Free Survival

<table>
<thead>
<tr>
<th>Study Reference</th>
<th>Hazard ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>DaVinci (n=88)</td>
<td>0.50</td>
</tr>
<tr>
<td>Graven (n=55)</td>
<td>0.63</td>
</tr>
<tr>
<td>Seymour (n=549)</td>
<td>0.79</td>
</tr>
<tr>
<td>Summary</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>1.26</td>
</tr>
<tr>
<td></td>
<td>1.58</td>
</tr>
</tbody>
</table>

(favours combination)       Hazard ratio

Random-effects meta-analysis

Summary exp(effect): 0.96  95% CI (0.84,1.09)
Estimated heterogeneity variance: 0.0  p= 0.693

Overall Survival

<table>
<thead>
<tr>
<th>Study Reference</th>
<th>Hazard ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>DaVinci (n=88)</td>
<td>0.50</td>
</tr>
<tr>
<td>Graven (n=55)</td>
<td>0.79</td>
</tr>
<tr>
<td>Seymour (n=549)</td>
<td>1.26</td>
</tr>
<tr>
<td>Summary</td>
<td>2.00</td>
</tr>
</tbody>
</table>

(favours combination)       Hazard ratio

Random-effects meta-analysis

Summary exp(effect): 0.92  95% CI (0.51,1.67)
Estimated heterogeneity variance: 0.1  p= 0.137
Appendix 1.2 - Patients with Tumour Response

Single Agent
Summary effect=0.125  95% CI (0.106, 0.144)
Estimated heterogeneity variance: 0.0012  p< 0.001

Combination
Summary effect=0.142  95% CI (0.0723, 0.212)
Estimated heterogeneity variance: 0.0049  p= 0.002

Response (Single Agent)

Response (Combination)
Appendix 1.3 - Patients Progression Free at 3 months

**Single Agent**
Summary effect=0.604  95% CI (0.548, 0.659)
Estimated heterogeneity variance: 0.014  p< 0.001

**Combination**
Summary effect=0.622  95% CI (0.511, 0.734)
Estimated heterogeneity variance: 0.012  p= 0.002

### Progression Free Survival (Single Agent)

- 1997 - Rougier - 350 d1 - 21d
- 1996 - Rougier - 300-350 d1 - 21d
- 1999 - Aravaninos - 350 d1 - 21d
- 1999 - Frontini - 300-350 d1 - 21d
- 1999 - Hoeffken - 350 d1 - 21d
- 1999 - Rothenberg - 125 d1 - 7d
- 1999 - Van Cutsem - 350 d1 - 21d
- 2002 - Tsavaris - 250/350 d1 - 21d
- 2003 - Fuchs - 125 d1 - 7d
- 2003 - Fuchs - 350 d1 - 21d
- 2003 - Mendez - 350 d1 - 21d
- 2003 - Tsavaris - 175 d1 - 10d
- 2003 - Tsavaris - 350 d1 - 21d
- 2004 - Lal - 350 d1 - 21d
- 2005 - Garcia-Giron - 250 d1 - 14d
- 2007 - Graeven - 125 d1 - 7d
- 2007 - Seymour - 350 d1 - 21d
- 2007 - Tsavaris - 100 d1 - 7d
- 2008 - Haller - 350 d1 - 21d
- 2009 - Kim - 350 d1 (300 for ecog2) - 21
- Current Study - 350 d1 - 21d

### Progression Free Survival (Combination)

- 2002 - Leonard - 180 d1 - 14d
- 2004 - Toumigand - 180 d1 - 14d
- 2007 - Graeven - 80 d1 - 7d
- 2007 - Seymour - 180 d1 - 14d
- Current Study - 180 d1 - 14d
Appendix 1.4 - Patients Alive at 6 months

Single Agent
Summary effect=0.714  95% CI (0.69, 0.738)
Estimated heterogeneity variance: 0.0012  p= 0.025

Combination
Summary effect=0.761  95% CI (0.627, 0.896)
Estimated heterogeneity variance: 0.014  p= 0.005
Appendix 2 - Calculations from Begg paper

In order to calculate estimates for this meta-analysis using the method of Begg and Pilote it was necessary to program solutions to the equations they listed on page 900 of their publication (summarised below).

\[
x_i \sim N(\theta_i, s_i(x)^2) \quad i = 1, \ldots, n; \\
y_i \sim N(\theta_i + \delta, s_i(y)^2) \quad i = 1, \ldots, n; \\
u_i \sim N(\theta_i, s_i(u)^2) \quad i = n + 1, \ldots, n + k; \\
v_i \sim N(\theta_i + \delta, s_i(v)^2) \quad i = n + k + 1, \ldots, n + k + m;
\]

\[
x = \sum_{i=1}^{n} x_i s_i(y)^2/\hat{d}_i, \quad y = \sum_{i=1}^{n} y_i s_i(x)^2/\hat{d}_i, \quad y_o = \sum_{i=1}^{n} y_i \hat{\sigma}^2/\hat{d}_i, \quad x_o = \sum_{i=1}^{n} x_i \hat{\sigma}^2/\hat{d}_i, \\
u = \sum_{i=n+1}^{n+k} u_i/(s_i(u)^2 + \hat{\sigma}^2), \quad v = \sum_{i=n+k+1}^{n+k+m} v_i/(s_i(v)^2 + \hat{\sigma}^2), \quad d_x = \sum_{i=1}^{n} s_i(x)^2/\hat{d}_i,
\]

\[
d_y = \sum_{i=1}^{n} s_i(y)^2/\hat{d}_i, \quad d_o = \sum_{i=1}^{n} \hat{\sigma}^2/\hat{d}_i, \\
d_u = \sum_{i=n+1}^{n+k} (s_i(u)^2 + \hat{\sigma}^2)^{-1}, \quad d_v = \sum_{i=n+k+1}^{n+k+m} (s_i(v)^2 + \hat{\sigma}^2)^{-1},
\]

and

\[
\hat{d}_i = s_i(x)^2 s_i(y)^2 + s_i(x)^2 \hat{\sigma}^2 + s_i(y)^2 \hat{\sigma}^2,
\]

The following two equations can be solved to find maximum likelihood estimates of \(\mu\) and \(\delta\) for a given value of \(\sigma^2\):

\[
x + y + u + v = \mu(d_x + d_y + d_u + d_v) + \delta(d_x + d_v), \\
y + y_o - x_o + v = \mu(d_x + d_o) + \delta(d_x + d_o + d_v),
\]

An updated value of \(\sigma^2\) can then be determined using the derivative of the likelihood with respect to \(\sigma^2\) (which depends on the new \(\mu\) and \(\delta\)). This cycle is repeated until satisfactory convergence is reached.
Clarification of the derivative of the log likelihood with respect to $\sigma^2$

We examine the $u$ component:

$$f(u_i \mid \mu, \delta, Su_i, \sigma^2) = \frac{1}{\sqrt{2\pi(Su_i^2 + \sigma^2)}} \exp\left(-\frac{(u_i - \mu)^2}{2(Su_i^2 + \sigma^2)}\right)$$

$$l(u_i \mid \mu, \delta, Su_i, \sigma^2) = \prod_{i=1}^{n} f(u_i \mid \mu, \delta, Su_i, \sigma^2)$$

$$L(u_i \mid \mu, \delta, Su_i, \sigma^2) = \sum_{i=1}^{n} \ln\left(f(u_i \mid \mu, \delta, Su_i, \sigma^2)\right)$$

$$= \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi(Su_i^2 + \sigma^2)}}\right) - \left(\frac{(u_i - \mu)^2}{2(Su_i^2 + \sigma^2)}\right)$$

$$= \sum_{i=1}^{n} -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(Su_i^2 + \sigma^2) - \left(\frac{(u_i - \mu)^2}{2(Su_i^2 + \sigma^2)}\right)$$

Differentiating with respect to $\sigma^2$ we find the component of the log-likelihood which the $u$ variables contribute:

$$\frac{dL}{d\sigma^2} = \frac{d}{d\sigma^2} \left\{ \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi(Su_i^2 + \sigma^2)}}\right) - \left(\frac{(u_i - \mu)^2}{2(Su_i^2 + \sigma^2)}\right) \right\}$$

$$= \sum_{i=1}^{n} \left( -\frac{1}{2(Su_i^2 + \sigma^2)} \right) - \frac{(u_i - \mu)^2}{2} \frac{d}{d\sigma^2} \left(\frac{1}{(Su_i^2 + \sigma^2)}\right)$$

$$= \sum_{i=1}^{n} \left( -\frac{1}{2(Su_i^2 + \sigma^2)} \right) - \frac{(u_i - \mu)^2}{2} \left(-\frac{1}{(Su_i^2 + \sigma^2)^2}\right)$$

$$= \sum_{i=1}^{n} \left( -\frac{1}{2(Su_i^2 + \sigma^2)} + \frac{(u_i - \mu)^2}{2(Su_i^2 + \sigma^2)^2}\right)$$

This is slightly different to that reported in Appendix1 (page 905) of Begg and Polite:
Appendix 3 - Validation of Begg program

Initial validation of program

This derivative does not cross 0 so the maximum likelihood estimate of \( \sigma_2 = 0 \).
This in turn estimates \( \mu = 0.324 \) and \( \delta = 0.134 \) which matches their results.

Eventual validation of program

Plots of the log-likelihood and corrected derivative with respect to \( \sigma_2 \):

Using the iterative procedure to estimate and plot the likelihood over various valid ranges it is found that it is maximised when \( \sigma_2 \) is 0.00126. This determines estimates of \( \mu = 0.322 \) and \( \delta = 0.148 \).
Appendix 4 - Abbreviations

5-FU  5-Flourouacil (common chemotherapy used in cancer treatment)
ADL  Activities of daily living
AGITG  Australasian Gastro Intestinal Trials Group
CI  Confidence interval
CTC  NHMRC Clinical Trials Centre, University of Sydney
CTCAE V3.0  National Cancer Institute - Common Terminology Criteria for Adverse Events (version 3.0).

Diarrhoea
  Grade 3: increase of ≥ 7 stools per day over baseline; incontinence;
  IV fluids ≥ 24hrs; hospitalization; severe increase in ostomy output
  compared to baseline; interfering with ADL (activities of daily living).

  Grade 4: life-threatening consequences (e.g. hemodynamic collapse)

HR  Hazard ratio
OS  Overall survival
OS6  Overall survival at 6 months
PFS  Progression free survival
PFS 3  Progression free survival at 3 months

Acknowledgements

Prof Malcolm Hudson – Project supervisor
Prof Val Gebski – AGITG group statistician
Prof Stephen Clark – Principal Investigator DaVinci
Sonya Yip – Assistance with literature review / data extraction