Biochemical Measurement of Injury and Inflammation in Musculoskeletal Surgeries

Dinesh Kumbhare, William Parkinson, R. Brett Dunlop and Anthony Adili
McMaster University
Canada

1. Introduction

Sufficient tissue trauma produces a temporary rise in circulating concentrations of various tissue proteins including creatine kinase (CK), myoglobin (Mb), myosin heavy chain (MHC), and collagen metabolites, as well as acute phase inflammation related analytes such as C-reactive protein (CRP), and the cytokines, interleukin-6 (IL-6) and interleukin-8 (IL-8) that mark the presence or absence of an injury or inflammatory response. Chronic, or late, elevations in biomarkers may also correlate with surgical complications. Measurement of biomarkers in orthopedic surgeries has been undertaken for 2 purposes: 1) to evaluate surgeries per se for improvement of surgical techniques and reducing adverse consequences, and 2) to examine measurement properties of biomarkers, using surgeries as models of musculoskeletal trauma and inflammation. In this chapter, some of the findings are described to illustrate both research objectives and their inter-relatedness, and to a degree, to capture aspects of the state of development of the overall biomarker research paradigm for surgical musculoskeletal injuries. The cytokine, interleukin-6 and the muscle cytoplasmic protein, creatine kinase, are emphasized partly because of known physiology and the relatively greater volume of research on these biomarkers over other candidates. However, some available research comparing these biomarkers with others is shown which has provided early evidence to substantiate better measurement characteristics for at least some applications. The survey is largely restricted to circulating blood concentrations over other possibilities (e.g. urine, joint synovial fluid, saliva) to allow comparison across studies, and given the potential for easier research and clinical application with venipuncture in comparison to joint fluid analysis, but not out of a lesser importance of investigations of other body fluids. Joint arthroplasty and lumbar surgery are the main focus since other surgeries (e.g. arthroscopy) produce much less tissue disruption and might be better studied by local tissue or fluid sampling, and because surgeries in emergency settings (e.g. fracture reduction) involve a more complex biochemical response post-surgery created by the combination of the original injury and the surgical injury. Findings from orthopedic surgeries are given priority but findings from non-orthopedic surgeries are shown when there is a difference in measurement principles or when measurement properties or clinical implications have been made clearer.
2. Measurement properties

2.1 Basic physiology

Historically, basic and applied physiology studies implicated a large number of proteins as potential markers of tissue injury, injury related inflammation, and inflammation created during injury by secondary causes such as infection or stress responses (Bauman 1994; Febbarraio & Pederson 2005; Fong et al 1990; Jakab & Kalabay 1998; Wu & Perryman 1992).

In the early, pro-inflammatory, stages of the response to trauma, monocytes and macrophages release IL-6, TNF-alpha, and IL-1beta. TNF-alpha and IL-1beta have feed forward effects on endothelial cells and fibroblasts to also release IL-6, which is produced in these cells. While all 3 proteins could be markers of trauma, IL-6 was found to be the most potent inducer of the liver acute phase proteins including CRP and has captured much of the interest in injury research since that time (Heinrich et al 1990). IL-6 was named in 1987 based on studies aimed at isolating a hepatocyte stimulating factor (Poupart et al 1987; Yasukawa et al 1987).

CK is found in the muscle cytoplasm and is involved in the conversion of adenosine triphosphate to adenosine diphosphate during the energy transfer process. Separate isoforms exist for skeletal muscle (CK-MM), cardiac muscle (CK-MB) and brain (CK-BB). This specificity for skeletal muscle has advantages over other analytes such as MB and MHC which do not have separate cardiac and skeletal muscle isomers. Musculoskeletal trauma studies have measured either the total CK or CK-MM but the total circulating CK has invariably been found to consist almost exclusively of CK-MM (Kawaguchi 1996, 1997; Strecker et al. 1999; Kumbhare et al. 2007) in non-cardiac research.

Given the large number of biomarkers being measured in clinical studies, short-listing to a manageable number of analytes that are most typically effective would be useful. Apart from the kinetics of biomarkers in circulation, other factors affecting the usefulness of biomarkers include the relative stabilities of proteins at room temperature, when frozen, and after repeated freeze thaws, as well as all of the measurement properties characteristic of the chemistry assays such as upper and lower detection limits, coefficients of variation, and the ability of the assay to recover/capture the true protein concentrations in the sample.

Studies of the rapidly changing acute stage biochemical markers have reported differences in the magnitudes of peak concentrations, the statistical significance of a protein change, and the timing of the peak. Such differences require careful analysis to establish what is consistent, as part of the development of the biomarker paradigm.

2.2 Response sensitivity

Many pioneering studies have focused partly on the simple question as to whether a reliable biochemical response is achievable with one or more proteins. The sensitivity of a response, while relevant to diagnostic accuracy, is not the same thing, the latter requiring accurate placement of patients in the clinically positive group, and the clinically negative group.

Two very early studies on lumbar surgery without fusion or instrumentation, found a serum mean CK-MM response of just over 500 U/L for single level posterior surgery, which was over 4 times the baseline CK-MM levels (Kawaguchi et al. 1996, 1997).
Injury responses to lumbar decompression with spinal instrumentation were studied including CRP and the cytokines IL-1ra, IL-6, IL-8, IL-10, TNF-RI, TNF-RII (Takahashi et al., 2001). With sampling 1, 2, and 7 days following lumbar surgeries, IL-6, IL-8, and IL-10 were elevated only in a decompression plus spinal instrumentation group (n = 7) and not in groups having decompression only (n = 7) or decompression plus fusion (n = 6) although there appeared to be trends towards increases in these biomarkers in these less invasive surgeries. Given the likelihood that the biochemical responses are smaller to less invasive surgeries, small sample sizes may have prevented injury detection. In a subsequent study, involving 12 subjects and more serial measures, a significant increase in mean IL-6 was obtained at 6 hours following lumbar decompression (Kumbhare et al., 2009) with trends toward significant elevations at 12 and 24 hours. The IL-6 response was weaker than CK, consistent with the previous research showing an easily detected CK response but more difficult to detect rise in IL-6 in the lumbar decompression model. A similar study on CK showed excellent separation of the distributions of proteins for baseline to peak, versus baseline fluctuations (Kumbhare et al., 2007) showing sensitivity to lumbar decompression. The reduced sensitivity of IL-6 to muscle trauma, in comparison to CK could be related to the degree to which IL-6 is inducible in muscle. Examination of tissue homogenates found greatest concentrations of IL-6 in bone and lung, lower concentrations in skin and adipose tissue, and lowest concentrations in skeletal muscle (Perl et al. 2003).

In the instrumentation research (Takahashi et al., 2001, 2002), IL-8 and IL-10 were found to peak earlier than IL-6 such that the injury response was largely absent by the next day, whereas IL-6 remained elevated the day after surgery and had not resolved by day 2 post-surgery. This suggested narrow diagnostic windows for IL-8 and IL-10, which could make these proteins more difficult to capture in sampling.

Myosin heavy chain (MHC) has the advantage of a wider diagnostic window than CK or the pro-inflammatory cytokines. Yet, it has been studied very little in orthopedic research. MHC is a component of the contractile apparatus of muscle. Onuoha and coworkers (2001) measured serum MHC concentrations before fracture reduction and serially after the operation in a heterogenous group having various types of fracture surgeries. Twenty-four hours post surgery, group mean (n = 12) MHC concentrations showed an approximate 3-fold increase over mean baseline concentration which was about 100 ug/L. This amounted to an effect size of about 3 standard deviations. MHC was argued to be a better marker than the traditional use of CK or myoglobin because the latter can increase in circulation with a leaky membrane only, whereas elevated MHC reflects destruction of the contractile mechanism and therefore, is more consistent with the concept of injury. In short, CK-MM has the advantage of skeletal muscle specificity, whereas MHC of specificity to the state of injury may be more specific to the state of injury.

Munoz et al (2005) compared biomarker responses following spinal instrumentation surgery with serial measures at pre-operation, post-operation, 1 day, 2 days, and 7 days. They measured a wide range of potential biomarkers including CRP and IL-6 as well as IL1b, IL-2, IL-4, IL-5, IL-8, IL-10, IL-12, TNF-alpha, and interferon-gamma. Only CRP and IL-6 showed sensitivity to the surgical trauma. Similarly, a study of total knee replacement (TKR) in Britain, found no serum injury responses for TNF-alpha (Andres et al., 2003) at 6, 24, or 48 hours following total knee arthroplasty and IL-1beta was elevated at the 24 hour time point only. This team did obtain stronger elevations in these proteins.
in joint drain fluid. Given that TKR produces biochemical changes that are quite large in comparison to other surgeries, analysis of circulating TNF-alpha and IL-1beta have not been supported as general indices of the pro-inflammatory response with the current chemistry assays when circulating concentrations are used. TNF-alpha and IL-1beta may be useful for analyses on joint fluid. TNF-alpha was found to be elevated in circulation for short bursts among patients hospitalized for sepsis (Damas et al., 1992). IL-1 beta was not. However, even in that study, persisting IL-6 elevations were not only easier to measure by virtue of their longer time course in fluctuations, but allowed greater prediction of shock than TNF-alpha.

In a British study (Hall et al., 2000) a factor analysis was used to examine the temporal pattern of a number of proteins including IL-6, CRP, and cortisol. The cortisol response started earliest and had returned to baseline by about 48 hours and was only about 50% higher than baseline at 24 hours showing less response magnitude and a shorter time course than IL-6 or CRP. Separate independent factors were created by the distinctly different time magnitude profiles for these 3 proteins. This contrasted with the use of epinephrine and norepinephrine, which were inter-correlated implying redundancy in measurement. The findings demonstrated a need for separate time series protocols when studying these different proteins.

The routine availability of a biomarker is a practical advantage and must be weighed against measurement properties of proteins that might be expensive to measure or measurable only in specialty laboratories. For example, IL-6 has frequently been considered to be preferable to CRP for measurement of early phase inflammation. However, while this seems more evident for some applications, a more general superiority of IL-6 over other candidates cannot yet be concluded. The fact that CRP is routinely available in many clinical laboratories, whereas IL-6 requires purchasing or developing immunological assays forces a need for evidence of sufficient superiority. The fact that clinical research might become applied more quickly and easily in the case of CRP is relevant. One example is a study on post-surgery infection (DiCesare et al., 2005) following TKR and THR. The statistical effect size for group differences were just over 1 standard deviation for all 3 significantly elevated proteins (IL-6, CRP, ESR) showing no relative advantage of IL-6 for separating the infected and non-infected distributions with the exception of a slightly weaker group separation by CRP in the THR patients created by a high standard deviation in the infected group. In an earlier study, that also involved hip and knee arthroplasties, the acute phase responses for IL-6 and CRP had similar effect sizes, amounting to roughly 2.5 standard deviation increases over baseline (Wirtz et al., 2000).

Outside of orthopedic surgery research, superiority of IL-6 over CRP has been reported. For example, IL-6 and procalcitonin levels were higher in the presence of episodes of bacteremia in neutropenic patients with hematological malignancies. CRP was elevated in the population but showed no correlation with the presence of bacteremia (von Lilianfield-Toll et al., 2004). The quicker return to baseline by IL-6 over CRP is a particular advantage in such situations, where the clinical condition fluctuates over time.

### 2.3 Time course

Surgery models offer the advantage of a pre-surgery baseline. In knee and hip arthroplasty, typical periods of hospitalization allow for control while sampling in the sensitive time windows for most of the biomarkers studied to date.
Evaluation of biomarker time courses requires multiple serial measures. Studies differ substantially on serial measurement, ranging from a single post-injury sample, to many time points, and all studies do not include pre-surgery or late enough time points to establish baselines and return to baseline. Therefore, there has been a limited ability to clarify the sensitive time windows and more precision is required.

Two early studies provided multiple serial measures to capture biomarker time-sensitive windows for inflammation related proteins. Wirtz et al (2000) included pre-operative and post-operative sampling at 6 and 12 hours and at daily intervals to 1 week following hip (n = 20) and knee (n = 10) arthroplasties. They compared the IL-6 time course with CRP. This revealed a slower rise in CRP relative to IL-6 and a much slower decline such that group peak CRP was reached between 1 to 2 days after surgery and remained about two-thirds of peak at 4 days and had not returned to baseline by day 7. IL-6 showed re-established baseline by about 3 days. The authors noted that other research showed the return to baseline for CRP is about 3 weeks, and ESR requires about 3 months. Therefore, an unusual persistent post-surgical elevation would be evident sooner with IL-6.

A second study (Hall et al 2000) included many serial data points and a relatively large sample (n = 158) of subjects who had either TKR or THR. Samples were taken before and after surgery prior to tourniquet release, and at 2, 4, 8, 12, and 24 hours and then daily to 7 days. Peak IL-6 occurred at 24 hours, whereas the CRP peak occurred at 48 hours. IL-6 returned to baseline more quickly than CRP. However, smaller but significant elevations in both proteins were still present on day 7. In other words, the pattern of relative changes in IL-6 and CRP were consistent with other research.

However, these 2 studies showed outcome differences of potential importance. First, the mean protein concentrations do not coincide. Wirtz et al obtained a peak IL-6 of about 400 pg/ml in each of the TKR and THR groups. Peak CRP was about 150 mg/L in each group. In the Hall study, peak IL-6 was 125 pg/ml after THR and 175 pg/ml after TKR. The teams used different chemistry assays for both proteins. Whether these dramatic differences are related to assay sensitivities, differences in drugs used in the different study locations, or other factors is not known. The use of different assays and different clinical protocols in different jurisdictions may be sources of variability in biochemical studies.

Comparisons between proteins on peak timing need to take individual differences into account. In a protocol aimed at defining the time course for CK and IL-6, serial measures were made before and immediately after lumbar decompression surgery and at 6, 12, 24, 48, at and 72 to 96 hours, and at 7 days (Kumbhare et al., 2007, 2009). The highest CK concentration occurred at different time points between 12 to 48 hours across subjects and the highest IL-6 concentration had a range of 6 to 24 hours. Time related individual differences create enough variance that this could affect various conclusions. For example, differences between 2 equally effective biomarkers on their prediction of clinical outcomes could be created by how closely sampling is done to the average peaks for each protein, or variability in peak timing within proteins in the subject sample.

### 2.4 Validity

The search for validity has been somewhat of a challenge because of the difficulty finding criterion measures. Some research suggests that biomarker responses might largely be all or
none. For example, unaccustomed exercise produces a well documented rise in CK (Critz & Cunningham 1972) and IL-6 (Petersen & Pederson 2005; Steensberg 2003). Bicep curls with both arms produced no greater change in CK than exercise at the same weight load, in one arm (Nosaka et al., 1992, 2002). Such findings could reflect a relative lack of biomarker sensitivity to injuries of different magnitude. Findings from some orthopedic studies involving direct validity quantification have produced more promising results. In a similar demonstration (Kugisaki et al., 2011), unilateral TKR was compared with bilateral TKR. The serum peak IL-6 for bilateral TKR was 246% of the peak for unilateral TKR. While a perfect outcome would be 200% difference, the design was between subjects by necessity and therefore individual differences might account for the relatively greater than double biochemical response in the bilaterals. Bilateral TKR produced only a 146% increase in CRP compared to unilateral TKR providing some evidence that CRP may be less sensitive creating less validity for detecting varying amounts of injury.

In studies of orthopedic and non-orthopedic surgeries, circulating IL-6 concentration was correlated with length of operation (Sakamoto 1994), amount of blood loss (Sakamoto 1994), and patients with higher blood loss requiring transfusions had higher IL-6 as a group in comparison to those with blood loss under 400ml and not requiring transfusions (Avall et al 1999). Studies described later in this chapter showing correlation of IL-6 responses with body temperature and post-operative pain provide further convergent type validity evidence (Lisowska et al. 2006).

In lumbar surgery, 7 patients with decompression plus instrumentation had higher post-surgery biomarker concentrations in comparison to 7 subjects who had decompression alone. Decompression plus fusion (n = 6) showed an intermediate biochemical response (Demura et al., 2006). In an older study (Kawaguchi et al., 1996), the CK-MM response was higher in patients with multiple level surgery (n = 14) in comparison to single level (n = 10). In the same study, dividing the group into greater retraction forces (retraction pressure x retraction time) and lesser forces revealed a higher CK-MM in the greater force subgroup. However, this could have been related to a higher proportion of multilevel surgeries in the greater force group given longer surgery times. Accordingly, in this team’s subsequent report on a larger sample (n = 47) multilevel surgery continued to produce a higher CK-mm response than single level, but separate analysis of surgery time showed no increased CK-MM in the subgroup having longer surgery times compared to shorter duration of muscle strain (Kawaguchi et al., 1997). They also found no difference between 2 subgroups separated on the basis of the length of the incision.

To further explore the potential for developing a criterion measure in lumbar surgery, the length and depth of the muscle isolated and retracted (i.e. surface area) was used as a criterion measure for validity examination in 18 subjects who underwent lumbar decompression (Kumbhare et al., 2008). Rather than divide the sample into subgroups, the correlation was examined continuously against individual peak serum CK. Muscle surface area and peak CK response had a Pearson r correlation of .60, falling in a moderate range.

2.5 Summary

Studies of basic measurement properties have largely involved relatively small samples. This is partly related to the need for many serial samples creating many data points.
Therefore, the extent to which the findings on kinetics of biomarkers generalize across subjects partly remains to be determined. Practical barriers make it impossible to sample precisely time-wise. Data presentations frequently use such increments as “Day 1, Day 2”, acknowledging such imprecision. It is therefore difficult to know exactly when samples were taken, and therefore difficult to extract the true time course. Muscle proteins have been studied less often than inflammation proteins and with smaller sample sizes, so relatively less is known about measurement properties in application to surgeries. Use of some proteins as general markers is weakened by relatively low sensitivity or short time courses. CK appears to be quite sensitive to surgical muscle trauma being consistently elevated even with low trauma surgery. Among the inflammation related markers, IL-6 and CRP have shown robustness in comparison to other proteins, in that they consistently are elevated. Some evidence suggests that CRP is less sensitive than IL-6 as a general proinflammation marker but the differences between IL-6 and CRP have not been large. Time courses have not been entirely consistent across studies. This could be related to difficulties timing sampling or to individual differences in time to peak. It seems likely that validity is in a moderate rather than high range in general. This could be related to variance created by: i. variability in sample timing within and between studies, or ii. the choice of criterion measures which may only approximate the construct of interest. Not discussed in this chapter are confounds and patient characteristics that may vary across individuals such as gender, age, body mass, anesthetics and other pharmacological agents, comorbidity, dehydration, activity levels, and other lifestyle factors. In fact, it may be unrealistic to expect better than moderate correlations between biomarkers and clinical outcomes.

These basic measurement issues must be kept in mind when one is considering the reasons for various findings of studies that examine clinical applications of the biomarker paradigm, and the challenges faced when attempting to obtain precise prediction.

3. Surgery evaluation, complications, and outcomes

Non-orthopedic studies have included a range of surgeries including: cardiac (Antunes et al., 2008; Booth et al., 2010), aortic (Baigrie), stomach (Aguilar-Nascimento et al., 2007; Herrera et al., 2010), bowel (Herroeder et al., 2007), liver (Gravante et al., 2010; Sato et al., 1996), hernias (Correla et al., 2001), choleystectomy (Chambrier et al., 1996; Kristiansson et al, 1999) and various cancers (Veenhof et al., 2011). Among orthopedic studies, there are now a large number of reports involving the use of biochemical injury measures for risk screening including their use in attempts to identify post-operative fever, and infection. Smaller literatures focus on technique comparisons for amount of tissue disruption or amount of inflammation, application in predicting hardware failure, and shorter term post-surgical recovery and hospital stay.

3.1 Post-operative fever

Post-operative fever has been of some interest in the biomarker paradigm. This is despite uncertainty about its clinical significance, leading to claims that it may simply be a normal response to surgeries (Andres et al., 2003; Pile 2006; Shaw & Chung 1999). It has been argued that unusually high or prolonged post-operative fever is a risk factor for complications (Munoz et al 2004; Pile 2006). There are numerous previously suggested
causes (Fanning et al., 1998) including: urinary tract infections, pneumonia, blood transfusions, infections at the wound, surgical site, intravenous catheter or created by the prosthesis, deep venous thrombosis, and C-deficile related to peri-operative antibiotics.

Andres and coworkers (2003) sought to provide evidence that post-operative fever was caused by the acute phase inflammatory response after arthroplasty. Half of their sample of 20 had early (first 3 days) post-operative temperatures in the febrile range (>38.5°C). The mean IL-6 concentration was higher in this group than the afebrile group at 24 and 48 hours, supporting the hypothesis that fever reflects an unusually high inflammatory biochemical response in the post-surgery phase. This separation based on fever, was not evident in circulating measurements of TNF-alpha or IL-1beta. Positive correlations between IL-6 and post-operative fever were also reported by others (Frank et al., 2000; Miyawaki et al., 1998). However, none of the febrile patients in the Andres et al study had complications, implying that IL-6 levels might correlate with fever, but neither IL-6 nor fever is specific for complications. Moreover, Andres et al. (2003) cite studies showing poor associations between fever and various post-operative complications. Fever may not have sufficient risk stratification potential on its own. One idea is that the fever is more typically related to extravasated blood and cellular debris (Shaw & Chung 1999) following arthroplasty. These early findings suggested that a correlation between IL-6 in the early post-operative phase and complications might exist, but prove not to be very strong.

3.2 Post-operative pain

The inflammatory response has been thought to contribute to pain in the early post-operative period leading to the hypothesis that treatment by antiinflammatories will reduce the acute phase response, reduce pain, and possibly improve return to functioning. There has been a consistent improvement in post-operative pain when various antiinflammatories are administered prior to operations. For example, intravenous diclofenac reduced post-operative pain and use of analgesics in a heterogeneous sample having various surgeries (Claeys et al.1992). Naproxen reduced postoperative pain following knee arthroscopy (Code et al., 1994) and following spinal surgery (Munoz et al, 2004).

In contrast, the effect of anti-inflammatories on the biochemical acute phase response has not been consistent. Using spinal surgery as a model, patients undergoing spinal fusion with instrumentation were assigned alternately to 2 groups of 20 subjects each, one that received 500 mg/day of naproxen combined with 40 mg/day famotidine in an attempt to blunt the inflammatory response, or to a no anti-inflammatory treatment control (Munoz et al. 2004). Following the operation, patients who received the anti-inflammatory treatment had lower body temperatures and requested less analgesic. CRP was reduced substantially from a group peak of 18.2 mg/ml to 6.7 mg/ml. This pattern was similar to that found in a study with spinal surgery without instrumentation (Takahashi et al. 2001). However, there was no significant difference between the groups on the IL-6 response, despite the anti-inflammatory pre-treatment. This contrasts with the finding of a reduction in the IL-6 response, with no reduction in CRP, to peri-operative administration of ibuprofen in cholecystectomy (Chambrier et al., 1996). Finally, Diclofenic produced no reduction in the CRP response despite pre-treatment and 24 hours continuous infusion of the drug in their sample having various major surgeries (Claeys et al., 1992).
3.3 Infection, osteolysis and hardware loosening

Studies examining the relationship between post-operative fever and infection have been reviewed for non-orthopedic surgeries (Fanning et al., 1998) and studies on knee and hip surgeries (Shaw & Chung 1999). The consistent result is that fever is an insufficient predictor of infection in non-orthopedic studies, which may be at least partly related to the high incidence of fever yet low rate of infection (Pile 2003). This has prompted the question as to whether biomarkers measured in the acute, early post-operative period, will generate the desired prediction.

Studies on chronically elevated biomarkers, typically measured at the time of resection arthroplasty, have had some success. Serum IL-6 has been evaluated for association with periprosthetic infection specifically. Recent studies have included sensitivity and specificity evaluations, which become important if there is to be a clinical application. For example, DiCesare and coworkers (2005) compared the predictive relationships of IL-6, C-reactive protein (CRP), white blood cell count (WBC), and erythrocyte sedimentation rate (ESR). Fifty-eight patients requiring re-operation were separated into 2 groups, those with, and those without infection confirmed histologically. Blood sampling was done at a mean of 84 months after the 1st operation, well after the acute elevations related to surgical injury would have resolved and baseline would be restored. Those with infection (n = 17) had higher concentrations of all 4 markers although WBC was not statistically significant. Using 10 pg/ml or higher as the cut-off/criterion, IL-6 had perfect sensitivity (1.0) and close to perfect specificity (.95) meaning 1 subject was misclassified as infected who was not. CRP had poorer performance with a sensitivity of 94 specificity of .78. In the second study, Bottner and coworkers (2007) studied a larger sample of 78 subjects. Using a cut-off score of 12 pg/ml (which is remarkably close to the cut-off in the previous study) IL-6 had a sensitivity of .95 but had a specificity of .87. CRP had the same sensitivity and .96 specificity. The combination of IL-6 > 12 pg/ml and CRP > 3.2 pg/ml correctly classified all patients. Such impressive accuracy has not always been obtained. In another study, IL-6 had a specificity of 1.0 when the cut-off was set to minimize false positives, but sensitivity was only modest at .57 (Buttarro et al, 2009).

A recent review has focused on the association between biomarkers measured at clinical follow-ups following arthroplasty (Mertens& Singh 2001) and the incidence of hardware loosening, osteolysis, and infection at the site of prosthesis. In general, synovial fluid analyses produced stronger and more consistent separation of those with and without complications. In serum analyses, prosthesis loosening had the lowest potential for prediction. IL-6 showed no significant relationship in 2 studies. TNF-alpha and IL-1beta showed significant prediction in 1 of 2 and 1 of 3 studies respectively. Collagen I had better than chance prediction in 2 of 5 studies, with the suggestion that this might have been related to the different assays used. In osteolysis, serum collagen I was significantly higher in the affected group in 1 of 2 studies. No significant relationship was obtained for IL-6, CRP, TNF-alpha, or IL-1b in any of the studies. Performance of serum biochemistry was best, in general, in the case of infected prostheses. Significant correlations were obtained for IL-6 in 2 of the 2 available studies described individually earlier, CRP in 3 of 3 studies, ESR in 3 of 3 studies, and TNF-alpha in 1 of 1 study. Analyses on the degree of sensitivity and specificity for complication versus no complication were not attempted.
In a second review, sensitivity and specificity analyses were performed on data from multiple studies (Berbari et al., 2010) creating a pooled sensitivity, with focus on circulating protein concentrations. The review included three studies in which IL-6 was measured, and 23 on CRP, 25 with ESR measurements, and 15 having measured WBC. Pooled sensitivity and specificity were as follows: WBC .45 and .87, ESR .75 and .70, CRP .88 and .74, and IL-6 .97 and .91, suggesting that IL-6 produces the best association with infection. However, as described earlier, examination of individual studies, revealed that IL-6 had modest sensitivity in 1 study (Buttar 200) in the presence of high specificity, and CRP had better specificity than IL-6 in another study despite equal sensitivity (Bottner et al. 2007). Sensitivity and specificity are influenced by the effect size of a group difference, which is a ratio of group mean differences to standard deviation. In small sample sizes, standard deviations can vary substantially from one study to another. Therefore, it is somewhat premature to conclude a superiority of IL-6 over CRP in the identification of infection.

The possibility that biomarkers could help to time staged surgeries has been of recent interest. One model is repeat TKR following infection associated with the first procedure, whereby a 2nd surgery is performed to remove the hardware and cement, debride tissue, and insert a spacer block. The infection is then treated and re-implantation is performed when the infection is resolved. Given the promising evidence for an association between inflammation markers and infection, the question is raised as to whether these proteins could be used to judge when inflammation has been sufficiently controlled. Ghanem et al (2009) found no accurate relationship between either serum or joint fluid concentrations of C-reactive protein or erythrocyte sedimentation prior to re-implantation, and the rate of persisting infection following re-implantation. Investigation of other biomarkers was suggested. In a summary of conditions for re-implant timing, the authors do recommend including a return to normal CRP, which persists after discontinuing antibiotics, as one criterion for re-implantation. They also emphasize that this is insufficient and that other criteria should be used including, effective treatment of the potential source of the infections including resolution of urinary tract infections, cellulitis, poor dentition, or skin ulceration. Treatment of otherwise unwell patients (e.g. immunocompromised, malnourished) should be optimized before re-implantation, and there should be good healing of the soft tissues and resolution of erythema.

3.4 Relative tissue disruption

It has been argued that biomarkers may help to evaluate minimally invasive techniques against more traditional surgical approaches (Cohen et al., 2009 ; Hartzband et al., 2004) or against varying minimal trauma approaches under development.

Chimento and coworkers (2005) compared their minimally invasive total hip arthroplasty procedure to the standard posterolateral approach. With the minimally invasive technique, the incision was 8 cm instead of 15 and smaller retractors were used. The technique also spared the quadratus femoris and the femoral insertion of the gluteus maximus. Circulating IL-6 concentrations did not differ between the 2 groups after surgery. However, the exact post-surgical time point was not reported. Given that the peak IL-6 concentrations were very low falling between about 5 pg/ml to 6.5 pg/ml in both the minimally invasive and
Cohen and coworkers (2009) compared 3 approaches for hip arthroplasty, 1. minimally invasive Watson Jones technique, 2. miniposterior transmuscular approach, and 3. minimally invasive -II (MIS-II) incision. The first technique involves an anterior approach with no muscle detachment but muscle is retracted creating strain forces. The posterior technique involves a small incision and minimizes, but does not eliminate, muscle detachment. In MIS-II there is an anterior incision to access the acetabulum, and a posterior incision for passage of instruments. Muscle is separated or bypassed, but not cut or detached. The 3 techniques had mean incision lengths of about 10 cm, 9cm and 9cm (combined anterior + posterior) respectively. Serum Mb and CK showed injury responses in the first 24 hours that were just over double and 4 times the normal respectively. The 3 techniques were not different in the magnitude of the biochemical response. However, group sizes of 10 were compared leading to the question as to whether larger samples could reveal subtle differences.

3.5 Other adverse effects

In their study described earlier, Cohen et al. found only 1 case in 30 of an elevated cardiac troponin following their minimally invasive hip replacement procedure. Cytokine responses were higher in patients who developed pancreatitis associated with spinal fusion surgery (He et al 2004). In a very recent study on mortality risk after hip fracture (Sun et al., 2011) IL-6, TNF-alpha, and IL-10 were measured before surgical repair, 1hour after the operation, and at 1, 3 and 5 days. Thirty-one of the 127 patients had died by the 6 or 12 month follow-ups. All 3 biomarkers showed significantly higher concentrations in the early post-surgical period in non-survivors. The best separation of the outcome groups was with IL-6 at 1 day post-surgery. Sensitivity and specificity analyses found IL-6 to be the best predictor. Setting the cut-off to minimize false negatives, IL-6 had a sensitivity of .935, but lower specificity at .635. This moderate level of prediction was also evident by the fact that the effect size of the group difference was about 1 standard deviation. This would imply a sizeable overlap in distributions. Nevertheless, the findings suggest that almost all of those who died could be predicted from the acute biochemical response while still reducing the total sample to a smaller high risk group that might be followed clinically.

3.6 Early post surgical status and hospital length of stay

Among non-orthopedic surgery studies, IL-6 on the day of haemopoietic stem cell transplantation predicted hospital length of stay (Tegg et al., 2001). Peri-operative lidocaine has been found to affect hospital length of stay for some abdominal surgeries. McCarthy (2010) reviewed studies in bowel surgery involving a total 395 treated patients and 369 controls. Lidocaine suppressed IL-6, improved post-operative pain, and reduced length of
stay by an average 1.1 days. In contrast, lidocaine had no effect on length of stay following hysterectomy (Bryson et al., 2010).

In a very recent study with 68 subjects having hip or knee arthroplasty (Koppensteiner et al., 2001), both CRP and IL-6 in the acute phase, predicted the absence of complications during hospital stay. However, because IL-6 normalized faster, it provided the better criterion for safe discharge.

3.7 Summary

In general, circulating levels of biomarkers have been found to have considerably less sensitivity than measurement of these proteins in joint or drain fluids. Circulating chronic elevations in IL-6 and CRP have been found to be correlated with the development of infections with some promising sensitivity and specificity findings. There is some promising evidence for a predictive relationship between acute phase protein elevations and complications during hospitalization outside of orthopedic surgeries with an implication that proteins might be helpful in discharge decisions. The potential for affecting length of stay in orthopedic surgeries needs to be explored. Biomarker changes in the acute or chronic period have not been sufficiently sensitive to identify loosening of surgical hardware. However, a biomarker could increase in a similar manner to a range of pathologies, or increase only to some complications thereby helping with differential diagnosis. Accordingly, lower IL-6 might rule out infection leaving hardware loosening as a possibility thereby providing useful information. The potential for sensitivity to osteolysis has met with equivocal results but has not been ruled out.

4. Summary and future directions

The study of biomarker measurement properties with surgery models and their applications in surgery issues can be argued to be a relatively young paradigm, with the introduction of IL-6 occurring only just over 20 years ago. Commercially available assays have taken time to develop but are now available for many biomarkers, allowing an expansion of this research in recent years. Early research often involved small samples and there was a need to compare proteins and establish time courses for the rapidly changing concentrations in the early post-surgery phase. Nevertheless basic measurement properties have not been sufficiently worked out. The reasons for some differences in time courses in studies on the same proteins remain unknown. The reasons why response magnitudes vary across studies are also not known. In research on muscle injury, advantages and disadvantages of cytoplasmic versus contractile apparatus proteins have been argued but need to be studied. There are a large number of patient characteristics and potential confounds including various drugs that could influence the different proteins in different ways. All of these issues require future research. Insufficient understanding of these variables could be creating weaker associations with clinical criterion variables than might be achievable.

In applying biomarkers to clinical issues, the most promising findings have been in identifying risk of infection with many studies on CRP but few studies of IL-6. While it is tempting to conclude that IL-6 is the superior infection marker, more research is needed comparing IL-6 to CRP and on the possibility of their combination to improve risk
stratification. Further improvement in assays might reveal applications for osteolysis and hardware failure but diagnostic accuracy has not been evident thus far. It seems unlikely that improvement in blood sampling will change this given that chronic levels are measured which should be much less variable than protein concentrations in the early post-surgery period. Nevertheless, adjustment for within subject fluctuations to improve baseline estimation could be attempted.

In the early post-operative period, the rapidly changing nature of biomarker concentrations is a major challenge. This becomes quite relevant in application to identifying early complications that affect length of stay. Given the paucity of research on this issue, there is a need for further research. A caveat is that insufficient understanding of the time courses or reasons for differences in time courses across studies, and the sources of variance outside of injury and inflammation, will reduce the efficiency of that line of research. This probably forces such studies to make many serial measures and measure a wide range of potential confounds or covariates before simpler models can be trusted. What appears to be emerging from the drug research is that there are substantial differences in the effects of different pharmacological agents on different proteins. Different antinflammatories do not affect inflammation biomarkers equally. This is worthy of investigation itself, but becomes a problem when the absence of a change in a biomarker is interpreted to reflect no clinical effect. The use of multiple biomarkers and multiple clinical outcomes will continue to be necessary in this line of research.

5. References


Orthopaedic surgery is the widest and the strongest growing surgical specialty. It is clear, that the process of improving treatments and patients care, requires knowledge, and this requires access to studies, expert opinion and books. Unfortunately, the access to this knowledge is being materialized. As we believe that access to the medical knowledge should be reachable to everyone free of charge, this book was generated to cover the orthopaedic aspect. It will provide the reader with a mix of basic, but as well highly specialized knowledge. In the process of editing this book, my wife Jurgita has been, as usual, the most supportive person. I would like to thank her for being in my life. I would like to thank Mr. Greblo, the Publishing Process Manager, for all his help and last but not least thanks to our readers, as without them this book would have no meaning.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following: