Workers’ compensation

John F. Burton, Jr

1 Introduction
Every developed country has multiple programs providing cash benefits, medical care and rehabilitation services to disabled workers. Workers’ compensation programs limit benefits to workers who experience work-related injuries. Williams (1991, p. 1) argued that workers’ compensation programs share two basic characteristics. First, workers are eligible for benefits without having to establish fault by their employers. Second, workers’ compensation benefits cover only ‘tangible economic losses’, principally loss of earnings and the cost of medical care. However, as Williams indicated, benefits are sometimes paid for impairments that do not result in economic losses. To the list of shared characteristics should be added that benefits are prescribed by statute (although employers sometimes supplement these benefits). Common but not universal features are that the program is entirely financed by employers and that workers’ compensation benefits are the exclusive remedy of injured workers against their employers.

Williams (1991) described several models of workers’ compensation that provide variations on these features. Germany enacted the first modern workers’ compensation law in 1884 while Otto von Bismark was Chancellor. The German model relies on collective responsibility of Industrial Injuries Institutes (Berufsgenossenschaften), which are non-profit corporations, to administer the program subject to limited statutory coverage and benefits requirements and government supervision. The costs are shared among employers and employees. The United Kingdom’s workers’ compensation law of 1897 was similar to the German model in providing no-fault benefits, but differed in not providing for medical care or rehabilitation services, assigning responsibility to individual employers rather than industrial associations and not requiring employers to insure their risks. The 1897 law also gave the employee the option of accepting workers’ compensation benefits or suing the employer for damages. The UK model was modified in 1948 to allow the injured worker to both receive workers’ compensation benefits and sue the employer for damages, although any recovery in a tort suit reduced the workers’ compensation benefits.

1 Unless otherwise indicated, ‘injuries’ includes injuries and diseases.
Workers’ compensation programs began in the United States\(^2\) and Canada\(^3\) in the early decades of the 20th century. There are similarities: both countries make workers’ compensation the exclusive remedy for workplace injuries (with limited exceptions) and rely on a federal model where workers’ compensation coverage and benefits are also entirely controlled by subnational governments (states or provinces).\(^4\) One major difference is that in Canada workers’ compensation insurance is provided collectively through provincial funds (plus self-insuring employers to a limited extent), while in the US employers are individually responsible for insuring the benefits, which are provided by a mixture of private carriers, state funds and self-insurance.\(^5\)

An alternative model is represented by New Zealand, the Netherlands and Switzerland, although there are differences among these countries. The common feature is that no distinction is made between work-related and non-work-related injuries. The Netherlands adopted a set of programs in 1967 that provide the same medical, disability and death benefits for all injuries and diseases whether work-related or not. In 1974, New Zealand replaced the common law system based on fault with a comprehensive compensation program that provides benefits for all accidental injuries on a no-fault basis, but maintained the distinction between work-related and other diseases.\(^6\)

The economics literature in English primarily examines the workers’ compensation programs in the US, which is reflected in the scope of this chapter, although the Canadian programs are also discussed. Subsequent sections correspond to the five major objectives for a modern workers’ compensation program endorsed for the US by the National Commission

---

\(^2\) The emergence of workers’ compensation in the US is examined from a legal perspective by Witt (2004) and from an economic perspective by Fishback and Kantor (2000).


\(^4\) The states are dominant because when the workers’ compensation movement emerged in the early 20th century, the Supreme Court’s interpretation of the Constitution precluded federal regulation of most private sector employees (Burton and Mitchell 2003). By the time constitutional law changed in the 1930s to allow federal action, the programs were so entrenched at the state level that efforts to impose federal standards on the states were unsuccessful (Burton 2004).

\(^5\) Australia is another country where the states control most aspect of workers’ compensation programs (Productivity Commission 2004).

\(^6\) In addition to Williams (1991), multi-country studies include the examination of occupational diseases by Barth (1998) and the comparison of US and European disability policies by Aarts et al. (1998). Social security programs, including programs for work injuries, are briefly described for individual countries in US Social Security Administration (2008).
on State Workmen’s Compensation Laws (National Commission 1972, p. 15): broad coverage of employees and of work-related injuries and diseases; encouragement of safety; cash benefits providing substantial protection against interruption of income; sufficient medical care and rehabilitation services; and an effective delivery system.

2 Coverage

A Coverage de jure

(i) Coverage of workers and employers One aspect of coverage pertains to which employers and workers are encompassed by workers’ compensation statutes. Each state has a workers’ compensation program that covers private sector and state and local government employees in the jurisdiction. Federal programs cover federal employees and a few private sector employees such as longshoremen. Employers are required to provide workers’ compensation benefits in all states but Texas. Most states have a few exceptions to universal coverage, such as exclusion of workers in firms with few employees, farm workers, household workers, employees of some units of state and local governments, or corporate officers. Independent contractors and self-employed persons generally are not required to be covered, although some jurisdictions allow these persons to elect coverage. The National Academy of Social Insurance (NASI) estimated that about 97 per cent of all wage and salary workers in the US were nominally covered by workers’ compensation statutes in 2005, with the range from 77 per cent in Texas to 100 per cent in 32 jurisdictions (Sengupta et al. 2007, table A1). The Association of Workers’ Compensation Boards of Canada (2008) estimated that about 83 per cent of the workforce in Canada was nominally covered in 2006, with the range from 69 per cent in Manitoba to 100 per cent in the Northwest Territory and in the Yukon Territory.7

(ii) Coverage of injuries and diseases Another aspect of coverage pertains to which injuries and diseases qualify for benefits. The ‘traditional’ legal rules for determining whether an injury qualified for workers’ compensation benefits required the worker to satisfy four legal tests: (1) there must be a personal injury (2) resulting from an accident that (3) arose out of employment and (4) in the course of employment (Willborn et al. 2007, pp. 894–937).

7 Coverage of employment in US and Canadian jurisdictions is summarized in US Chamber of Commerce (2007, charts II and III).
The legal rules for determining which diseases involving workers qualified for workers’ compensation benefits evolved over time. Many diseases could not meet the four tests for injuries and, in particular, the accident test. As a result, almost all states adopted special provisions for diseases, which typically included a list of diseases for which compensability was presumed. These lists soon became obsolete in most jurisdictions. The statutory provisions for disease also typically contained a safety valve allowing coverage of ‘all occupational diseases’. However, this exception was often interpreted narrowly to preclude coverage of diseases not characteristic of the worker’s occupation or ordinary diseases of life, even when the disease was clearly caused by the workplace and disabling. In addition, statutes of limitations often precluded a claim for a disease when there was an extended period between the worker’s exposure to a substance and the onset of the resultant disability (Willborn et al. 2007, pp. 937–42).

Many states amended their workers’ compensation laws in recent decades to limit compensability of workplace injuries and diseases (Spieler and Burton 1998). One change limited the compensability of claims involving particular medical conditions, such as psychological injuries resulting from a mental stimulus in the absence of a physical injury (‘mental-mental’ injuries). A number of states also limited coverage when the injury involved aggravation of a pre-existing condition. The most significant change of this type denies compensation when the current workplace injury is not the sole or major cause of the disability. In addition, there were procedural and evidentiary changes in claims processing that restricted compensability. For example, some statutes now require that the medical condition caused by a workplace injury be documented by ‘objective’ medical evidence. This requirement excludes claims based on subjective reports of patients that cannot be substantiated by objective evidence, including musculoskeletal injuries that involve soft tissue damage. In addition, some workers’ compensation programs imposed on workers a stricter burden of proof or a greater quantum of proof.

**B Coverage de facto**

(i) **Coverage of workers and employers** While NASI estimates that about 97 per cent of workers in the US are covered by workers’ compensation, several recent studies found that the actual coverage of workers is less than the nominal coverage required by law because of evasive strategies.
such as misclassifying employees as independent contractors or not reporting employees. In New York after an investigation of 117 companies, the state’s labor commissioner said she ‘wouldn’t doubt that 10 percent of the state’s workers are either misclassified as independent contractors or work off the books’ (Greenhouse 2008). Some employers also misclassify employees into insurance classes with lower premiums. Neuhauser and Donovan (2007) estimated that between 1997 and 2002 California employers underreported payroll by 6 to 23 per cent by misclassifying employees or by illegally forgoing the purchase of workers’ compensation insurance altogether.

(ii) Coverage of injuries and diseases The result of the ‘traditional’ legal rules for determining eligibility for workers’ compensation in combination with the ‘modern’ restrictive provisions, together with other factors, is that substantial proportions of workers with workplace injuries and diseases do not receive workers’ compensation benefits. Sengupta et al. (2008) surveyed the recent studies documenting the lack of workers’ compensation coverage. An example is Leigh and Robbins (2004) who compared the number of occupational diseases shown in epidemiological data with the number of workers’ compensation claims involving diseases in 16 states in 1999. They found that at least 90 per cent of all deaths resulting from occupational diseases and at least 80 per cent of all medical costs caused by occupational diseases were missed by workers’ compensation programs. Two studies with national samples found that a high percentage of both work-related injuries and illnesses do not receive workers’ compensation benefits. Reville and Schoeni (2003/4) found that, among people aged 51 to 61 whose health limited the amount of work they could do, 36.3 per cent reported that an injury, accident or illness at work caused their disability, but only 5.3 per cent ever received workers’ compensation benefits. Lakdawalla et al. (2007) found that about half of workers in the National Longitudinal Survey of Youth who reported suffering a work-related injury in the previous year did not receive workers’ compensation benefits.

An interesting recent development is the use of capture-recapture methodology to estimate the number of injured workers who do not receive workers’ compensation benefits. This method uses the number of workers who received workers’ compensation benefits and the number of workers identified in the Survey of Occupational Injuries and Illnesses (SOII) survey conducted by the Bureau of Labor Statistics (BLS) to estimate the universe of injured workers, including those who are not identified in either the workers’ compensation or SOII data. Rosenman et al. (2006) estimated that only about two-thirds of workers suffering workplace injuries and
illnesses in Michigan between 1999 and 2001, resulting in more than seven
days of lost work, received workers’ compensation benefits and that only
about one-third of both work-related injuries and diseases were included
in the BLS data. Boden and Ozonoff (2007, table 4) extended the capture-
recapture method to six jurisdictions. Using the conservative assumption
of source independence (which means that decisions to report injuries and
diseases to workers’ compensation and to BLS were made completely
independently), they estimated that, depending on the state, workers’ com-
pen sation only compensated from 65 per cent to 93 per cent of all lost-time
injuries and diseases and that the BLS data base only included from 51 to
76 per cent of these incidents.

While there is thus considerable evidence that a substantial proportion of
workers suffering workplace injuries and diseases do not receive workers’
compensation in the US, it is not clear whether this undercompensation
has become worse in recent years. The frequency of temporary total dis-
ability claims, which involve disabilities exceeding the waiting period for
benefits, declined 44 per cent between 1992 and 2002 in the 41 states with
data from the National Council on Compensation Insurance (NCCI),
while BLS data on the frequency of injuries that resulted in days away from
work declined by 47 per cent over those years (Sengupta et al. 2007, p. 30).
This comparison suggests that workers’ compensation undercounting has
not worsened since the early 1990s. However, there is evidence indicating
that the previously described changes during the 1990s in legal standards
determining which injuries are compensable (Spieler and Burton 1998)
had an effect. Between 1987 and 1995, the Oregon legislature enacted a
series of amendments that constricted the eligibility rules. Thomason and
Burton (2001) estimated that by 1996 these changes reduced the number
of claims by 12 to 28 per cent and the benefits for workers (and costs to
employers) by 20 to 25 per cent below what the amounts would have been
if the laws had not been enacted. Guo and Burton (2008) also found that
the frequency of workers’ compensation claims as measured by the NCCI
as a proportion of the frequency of workplace injuries reported by the BLS

There is also evidence that a significant proportion of workplace injuries
do not lead to workers’ compensation benefits in Canada. Shannon and
Lowe (2002) found that 40 per cent of workers who had incurred an eligible

---

9 An earlier study of Michigan by Biddle and Roberts (2003) found that almost
40 per cent of workers with at least seven days of lost work did not file for cash
benefits and 27 per cent did not file for any sort of workers’ compensation benefits,
including medical care.
injury did not file a workers’ compensation claim, including 30 per cent of workers with lost time from work. A study of sawmill workers hospitalized by work-related injuries found that from 10 to 15 per cent did not receive workers’ compensation benefits, with older and non-white workers less likely and seriously injured workers more likely to receive benefits (Alamgir et al. 2006). Thompson (2007) referenced a survey of Canadian physicians that found 54 per cent of work-related injuries were inappropriately billed to the public health care system rather than to provincial workers’ compensation boards.\(^\text{10}\)

As in the US, a controversy exists in Canada about whether workers’ compensation benefits are being provided to a declining proportion of workers disabled by workplace injuries. Mustard et al. (2003) compared three Ontario data series for 1993 to 1998 and found that the 28.8 per cent decline in lost-time workers’ compensation claims closely matched the drop in self-reported work-related injuries and the incidence of self-reported absences from work of at least a week. These results suggest the decline in workers’ compensation claims in Ontario was caused by improved workplace safety rather than other factors such as more restrictive eligibility rules.

(iii) Who pays for the coverage? The nominal cost of workers’ compensation is paid by employers in the form of insurance premiums or benefits paid directly by self-insuring employers. But who actually pays for the program? The dominant view among lawyers is found in the first paragraph of the leading legal treatise (Larson and Larson 2007, p. 1-1):

> Workers’ compensation is a mechanism for providing cash-wage benefits and medical care to victims of work-connected injuries, and for placing the cost of these injuries ultimately on the consumer, through the medium of insurance, whose premiums are passed on in the cost of the product.

An alternative view – suggesting that employers bear at least a portion of the costs of the workers’ compensation program – is found in Justice Pitney’s 1917 opinion in New York Central Railroad Co. v. White (1917):

> Who is to bear the charge? It is plain that, on grounds of natural justice, it is not unreasonable for the State, while relieving the employer from responsibility for damages measured by common-law standards . . . to require him to contribute

\(^{10}\) Part of the explanation for the failure to file for workers’ compensation claims discussed in the Shannon and Lowe (2002) and Thompson (2007) articles may be that these studies focused on Ontario where only 73 per cent of the workforce was covered by workers’ compensation in 2006 (AWCBC 2008).
a reasonable amount, and according to a reasonable and definite scale, by way of compensation for the loss of earning power incurred in the common enterprise . . .

There is yet another view about who pays for workers’ compensation. This position, largely espoused by economists, is that workers pay for much of the program in the form of lower wages than they would have received in the absence of workers’ compensation. The essence of the modern economists’ position about who pays for workers’ compensation was also anticipated by Justice Pitney:

And just as the employee’s assumption of ordinary risks at common law presumably was taken into account in fixing the rate of wages, so the fixed responsibility of the employer, and the modified assumption of risk by the employee under the new system, presumably will be reflected in the wage scale.

Chelius and Burton (1994) postulated that the share of the costs of the program borne by workers in the form of lower wages depends on factors such as the elasticities of supply and demand in the labor and product markets and the extent of workers’ aversion to risk. Leigh et al. (2000, pp. 177–9) offer ‘merely a suggestion’ that employers bear 40 per cent of the costs of workers’ compensation, consumers bear 20 per cent and workers bear the remaining 40 per cent. Several empirical studies provide evidence about the tradeoffs between workers’ compensation benefits and wages at different periods in the history of the program. Fishback and Kantor (1995, table III) analyzed data from the early 1900s and found that in coal mining wages declined by 1.72 times the expected present value of compensation and by 1.04 times the expected benefits for lumber mill workers (both results statistically significant) and increased by a statistically insignificant amount in unionized building trades. Gruber and Krueger (1991) found that 87 per cent of the employers’ costs of workers’ compensation was shifted to workers in the form of lower wages. Moore and Viscusi (1990, p. 66) provided a number of estimates of the effects of the tradeoff between workers’ compensation benefits and wages. Their ‘most favorable evidence on the desirability of workers’ compensation’ estimated that the US$12 billion increase in workers’ compensation premiums between 1976 and 1983 actually saved employers US$14.0 billion because of the reduction in wages. However, when Viscusi (2004, pp. 38–41) re-examined the wage-benefits tradeoff as of 1997, he calculated that (given the probabilities of injuries and markup between benefits and premiums) an additional dollar of benefits should result in a 0.040 reduction in wages for the entire sample while the empirical results showed a 0.032 reduction. He concluded that ‘these estimates would imply that the level of workers’ compensation
benefits is above the efficient insurance amount’. He explained the higher rate of tradeoff of wages and benefits found by Moore and Viscusi (1990) as reflecting ‘a different era of workers’ compensation’.

3 Safety

A Trends in workplace safety and health

The long-term trends in workplace safety and health in the US were examined by Burton and Chelius (1997, pp. 254–6). The essence of their review was that workplace fatalities from accidents declined throughout the 20th century, but trends for workplace injuries showed no such continuing improvement. Periods of substantial decline in injury frequency rates were mixed with periods of relative stability in frequency rates and with periods when injury frequencies increased (the 1950s and arguably the period from 1983 to 1992). The evidence on occupational diseases was too fragmentary to allow an overall assessment of long-term trends. Burton and Chelius concluded (1997, p. 256) that ‘at best there is no apparent overall improvement in workplace safety and health in the last twenty-five years’.


A serious obstacle to identifying trends is that the accuracy of the BLS data has been challenged. Burton and Chelius (1997, p. 255) catalogue several reasons for skepticism about the data through the early 1990s, such as the possibility that changes in the Occupational Safety and Health Administration’s (OSHA) reporting policies explain the apparent increase in the total case frequency rate from the mid-1980s until 1992. A plethora of more recent studies challenge the accuracy of the BLS data.11 Leigh et al. (2000, table 1) estimate that in 1992 workplaces accounted for 13.3 million nonfatal injuries and 1.1 million nonfatal diseases plus 6371 deaths from injuries and 60 290 deaths from diseases, considerably higher than

---

11 One positive finding for the BLS data on recent trends reported in this section was provided by Layne (2004) who found that the CFOI, which was established by BLS in 1992, provided a more comprehensive measure of deaths resulting from workplace injuries from 1992 to 1997 than did the National Traumatic Occupational Fatalities surveillance system compiled by the National Institute of Occupational Safety and Health.
the BLS estimate noted above. Leigh et al. (2004) estimate that the BLS missed between 33 and 69 per cent of all workplace injuries in part because of the exclusion of government workers and the self-employed from the BLS Annual Survey. And as previously discussed, Rosenman et al. (2006) concluded that the BLS data only counted 32.7 per cent of the work-related injuries and 30.8 per cent of occupational diseases in Michigan, while the conservative estimates of Boden and Ozonoff (2007) were that the BLS data only captured from 51 to 76 per cent of the injuries and diseases in the six states they examined.

Two studies suggest interesting reasons for the underreporting of injuries and diseases. Pransky et al. (1999) identify the safety incentives programs of employers that typically reward supervisors for reducing workplace injury rates. Boden and Ruser (2003) argue that the changes in workers’ compensation laws during the 1990s that made it more difficult for workers to file claims induced employers to record fewer injuries, which accounted for between 7.0 and 9.4 per cent of the decline in BLS injury rates between 1991 and 1997.

B Approaches to workplace safety
The approaches to achieving workplace safety and health include the labor market, safety and health laws, tort suits and workers’ compensation. The first two approaches are examined in detail in Shapiro (2008) and will only be briefly examined here, while tort suits and workers’ compensation receive more extended treatment.

Shapiro (2008) presents the theory and evidence on the role of labor markets in promoting workplace safety and health. The essence of the theory is that workers require a compensating wage differential (risk premium) to perform jobs with risks of injury or death. Employers devote resources to improving workplace safety in order to reduce the risk premium they are required to pay in order to attract workers. Equilibrium is achieved when the marginal expenditure on safety equals the marginal reduction in the risk premium. The theory has been challenged for its assumptions such as workers’ knowledge and mobility. Most empirical studies by economists find evidence of compensating wage differentials

---

12 Other approaches to workplace safety and health include government promulgation of information about risks and remedies and collective bargaining and other policies that empower workers at a local level, examined in Burton and Chelius (1997), and employer safety programs, discussed by Burton and Schmidle (2002).

13 Burton and Chelius (1997) and Ruser (2008) also provide an analysis of this approach.
for workplace risks of injury or death.\textsuperscript{14} However, compensating wage differentials are not found in all studies and Dorman (1996) presents a very skeptical assessment of the theory and empirical evidence. Moreover, as Ehrenberg (1988) persuasively argued, empirical evidence of a risk premium does not mean the payments fully compensate workers for the extent of their expected losses.

Safety standards represent the earliest form of protective legislation in the US. Most states enacted safety laws between 1877 and 1910 (Somers and Somers 1954, p. 200), although they were generally ineffective (Chelius 1974, pp. 700–29) or had limited impact (Fishback 1998, p. 754). The perceptions that injury rates were increasing in the 1960s and that state efforts were inadequate help explain the adoption of the Occupational Safety and Health Act of 1970 (OSH Act), which established a regulatory scheme controlled by the federal government. Shapiro (2008) describes five types of approaches examining whether the activities of OSHA improved workplace safety and concludes that the examinations have produced inconsistent results.\textsuperscript{15} He also discusses attempts to explain OSHA’s limited impact on occupational diseases. Shapiro provides a variety of proposals to reform both OSHA’s standard setting and enforcement activities, but concludes that proponents of reform have not reached a consensus concerning some key proposals, including whether OSHA should be subject to cost-benefit analysis.

\textit{(i) Tort law} The theoretical stimulus of tort law on safety and the empirical evidence on the stimulus were examined by Burton and Chelius (1997, pp. 272–4). When negligence is the legal standard used for tort suits, an injured worker may sue his employer for damages when the employer

\textsuperscript{14} Two particularly interesting studies are Fishback (1998, table 4), who examined labor markets at the turn of the 20th century and found risk premiums in several, but not all, industries, and Viscusi (2004), who used a measure of fatality risk by both occupation and industry, in contrast to earlier studies, which were limited to industry fatality risk data. Viscusi also included measures of non-fatal job risks and the value of workers’ compensation benefits. In most regressions, the coefficients on all three measures were highly significant, with higher risks associated with higher wages and higher benefits associated with lower wages.

\textsuperscript{15} Burton and Chelius (1997, p. 276) asserted ‘the evidence suggests that the OSH Act has done little to improve workplace safety’. Ruser (2008) also reported that most studies concluded that OSHA has had little or no effect in reducing occupational injuries and illnesses. However, Tompa et al. (2007) conducted a systematic review of the literature and found limited to mixed evidence that inspections offer deterrence and that citations and penalties aid deterrence, and strong evidence that actual citations and penalties reduce injuries.
is at fault. If the employer has not taken proper measures to prevent accidents and thus is at fault, the employer will be liable for all of the consequences of the injury.\textsuperscript{16} The standard for the proper prevention measure as developed by Judge Learned Hand and restated by Posner (1972, p. 32) involves three terms: the magnitude of the loss if the accident occurs, the probability of the accident occurring and the cost of taking precautions to prevent the accident. If the product of the first two terms exceeds the cost of the precautions, failure to take the precautions is negligence.

Posner argued that proper application of this standard will result in economically efficient incentives to avoid accidents.\textsuperscript{17} As Chelius (1977, pp. 34–5) noted, the added costs of determining liability in a court may appear to be inconsistent with achieving an efficient use of resources since legal fees are usually a significant proportion of the total award. The benefits of legal proceedings, however, may outweigh their costs if the incentives created by such a system are more accurate than those present under alternative systems.

The evidence on the stimulus to workplace safety from tort law is mixed. The generally accepted view is that tort suits were largely ineffective as a remedy for workplace injuries in the late 1800s and early 1900s. Not only were workplace injuries and fatalities increasing, but employees were generally unsuccessful in law suits, in large part because of legal defenses available to employers, such as the contributory negligence defense, which eliminated any recovery if the worker was negligent, even if the employer was negligent to a greater degree. The leading legal treatise on workers’ compensation (Larson and Larson 2007, section 2.06) concludes that ‘the precompensation lost-adjustment system for industrial accidents was a complete failure . . . ’. However, Berkowitz and Berkowitz (1985, p. 160) indicated that workers were starting to enjoy considerable success with tort suits at the beginning of the workers’ compensation era. Perhaps the tort system if left in place for workplace injuries would have evolved and produced a major stimulus to safety.

There are nonetheless two types of empirical evidence that indicate skepticism is warranted about the stimulus to workplace safety from tort suits. First, the liability rules for workplace safety evolved in three stages in the US in the early 20th century. Initially, injured workers were required to establish negligence by the employer and employers had extraordinary

\textsuperscript{16} This formulation of the negligence standard assumes that the employee is not also negligent.

\textsuperscript{17} The negligence standard must be formulated in terms of the marginal expected accident costs and prevention costs in order to achieve proper allocation decisions, as discussed by Posner (1992, pp. 163–67).
defenses even if negligent, which essentially meant that most costs of industrial injuries were borne by workers. Then, in many jurisdictions, employer liability acts were enacted that required employees to prove the employer was negligent, but that abrogated some or all of the extraordinary defenses which shifted some of the costs of industrial accidents to employers. Most states soon abandoned this approach, although railroads are still subject to the Federal Employers Liability Act (FELA). Eventually all states enacted workers’ compensation laws which require employers to provide benefits on a no-fault basis, thereby shifting most of the nominal costs of workplace injuries to employers. In the absence of costs involved in the carrying out of market transactions, the Coase Theorem demonstrates that changing the legal rules about who is liable for damages will not affect decisions involving expenditures of resources that will increase the combined wealth of the parties including expenditures on accident prevention (Burton and Chelius 1997, pp. 270–71). However, if there are transaction costs, such as those involved in drafting and enforcing contracts, then changing the legal rules about initial liability can affect the allocation of resources. In the Landes and Posner model (1987), which assumed that transactions costs were substantial and that workers are the least-cost avoiders of accidents, the shift of more liability for workplace injuries to employers should result in a decline in industrial safety.

Fishback (1998, pp. 752–9) and Butler and Worrall (2008) provided excellent reviews of the empirical studies of this topic. Chelius (1977, pp. 44–5) examined a panel of 26 states from 1900 to 1940 and found that workers’ compensation laws and employer liability acts each lowered the accident rate. This result is consistent with the existence of significant transaction costs but indicates that employers are the least-cost avoiders of accidents. Subsequent studies have generally found that the changes in legal rules affected workplace safety, but differ on whether the changes from the common law approach of torts to employer liability acts or workers’ compensation have increased or decreased workplace accidents. For example, Fishback (1992) examined fatality rates in bituminous coal mining and found that employer liability and workers’ compensation acts both raised fatality rates, suggesting that employees in this industry were the least-cost accident avoiders. Butler and Worrall (2008) examined the effects on railroad workers in New Jersey of the enactment of the FELA in 1908 (which applied to railroad workers in interstate commerce) and the state’s workers’ compensation law in 1911 (which applied to other railroad workers). They found that as liability shifted to railroad companies after the enactment of the FELA, accident rates fell for three occupational groups who worked outdoors but rose for craft employees who worked indoors. During the period when FELA and workers’ compensation were both in
effect, they found the same general pattern as in the period after the initial shift to FELA. Butler and Worrall indicated that it appeared that workers in craft occupations were the least-cost avoiders (similar to the coal miners studied by Fishback), but that employers were the least-cost avoiders for the three outdoor worker groups. Their overall conclusion was that ‘those who argue that the tort system generated better safety incentives than did workers’ compensation are wrong’ (Butler and Worrall 2008, p. 195).

The second reason for skepticism about the stimulus to safety from reliance on tort suits is the major controversy among legal scholars about whether the theoretical incentives actually work. One school of thought is exemplified by Landes and Posner (1987, p. 10), who state that ‘although there has been little systematic study of the deterrent effect of tort law, what empirical evidence there is indicates that tort law . . . deters, even where, notably in the area of automobile accidents, liability insurance is widespread . . . ’.18

An opposing view on the deterrent effects of tort law is provided by Priest (1991), who found almost no relationship between liability payments and the accident rate for general aviation and stated that ‘this relationship between liability payouts and accidents appears to be typical of other areas of modern tort law as well, such as medical malpractice and products liability’.

A survey of the deterrent effects of tort law by Schwartz (1994, pp. 378–9) distinguishes between a strong form of deterrence (as postulated by Landes and Posner) and a moderate form of deterrence in which ‘tort law provides a significant amount of deterrence, yet considerably less that the economists’ formulae tend to predict’. As to workers’ injuries, Schwartz (1994, pp. 391–3) cited the Chelius and Fishback studies and concluded: ‘it is unclear whether a tort system or workers’ compensation provides better incentives; in an odd way . . . neither study is out of line with the general rule that a properly designed set of liability rules can produce beneficial results’.19

---

18 Just in case tort law does not have a deterrent effect, Landes and Posner (1987, pp. 13–14) add this rationale for their economic analysis of tort law: ‘Even if tort law does not have a significant effect on behavior, the theory advanced in this book is not refuted. Ours is a theory of the rules of tort law rather than of the consequences of those rules for behavior. . . . The doctrinal structure would still be economic even if the social function of tort law was to assuage feelings of indignation and avert breaches of the peace rather than to promote an efficient allocation of resources to safety’.

19 What constitutes a properly designed set of liability rules? Rubin and Shepherd (2007) examined the dramatic increase in tort reform that began in many states in the mid-1980s with the purpose of limiting the number of tort cases and
Dewees et al. (1996, p. 421) examined the empirical evidence pertaining to the three major goals of tort law: deterrence, compensation and corrective justice. They concluded that the deterrence effect of tort liability for workplace injuries is limited. More generally, after reviewing the evidence from several fields involving personal injury, including accidents involving automobiles, medical practice and products, they concluded that ‘the tort system performs unevenly in deterring the causes of personal injury, so its scope should be restricted to situations where its effect seems likely to justify its high cost’ (Dewees et al. 1996, p. 413). They recommended expanding the no-fault approach exemplified by workers’ compensation to the motor vehicle and to medical malpractice fields, each with a system ‘with features similar to existing workers’ compensation schemes’.

Based on the ambiguous historical experience of the impact on safety of moving from tort suits to employer liability acts to workers’ compensation and the controversy over the deterrent effect in other areas of tort law, Burton and Chelius (1997, p. 274) concluded ‘that we would resist the use of tort suits to deal with work injuries unless much more compelling evidence of the deterrent effect is produced’. A decade later, especially in light of the comprehensive review of the empirical evidence concerning the deterrent effects of torts by Dewees et al. (1996), that conclusion deserves affirmation.

**(ii) Workers’ compensation** Workers’ compensation programs rely on two levels of experience rating to promote safety (Burton 2005b). Industry-level experience rating establishes a pure premium (or manual) rate for each industry that is largely based on prior benefit payments by the industry. The resulting differences in labor costs and prices among industries should shift the composition of national consumption towards safer products. Firm-level experience rating determines the workers’ compensation premium for the damages awarded. Theoretically, these reforms may have had opposing effects. If the costs for tortfeasors are reduced, they have less incentive to reduce the risks of accidents. However, if the costs for victims are increased, they have more incentives to take precautions. Moreover, if the costs of the tort system are reduced, potential injurers and victims have more resources to buy risk-reducing products and services. Rubin and Shepherd explored accidental, non-motor-vehicle death rates for each state from 1981 to 2000 and found that some reforms, such as a higher evidence standard for punitive damages, were associated with decreases in accidental death rates, while reforms to the collateral source rule were associated with increased deaths. Overall, the tort reforms in the states were associated with an estimated 24,000 fewer accidental deaths over the 20 years. This study suggests that changes of the legal rules within the tort system can have a major impact on prevention.
each firm above a minimum size by comparing its prior benefit payments to those of other firms in the industry. In order to remain competitive, firms have an incentive to improve safety in order to reduce premiums.

The effects of the workers’ compensation program on safety in general, and firm-level experience rating in particular, have been debated by a number of scholars. The essence of the ‘pure’ neoclassical economics approach is that the introduction of workers’ compensation will lead to reduced incentives for workers to avoid injuries, assuming that they did not purchase private disability insurance plans prior to the introduction of workers’ compensation, since the adverse economic consequences of the injuries for workers are reduced by workers’ compensation benefits. The disincentive to avoid injuries is an example of the ‘moral hazard’ problem. This economic theory also argues that the introduction of workers’ compensation will lead to smaller risk premiums and thus reduce the incentives for employers to prevent accidents unless perfect experience rating is used to finance the program, which is another variant of ‘moral hazard’.

In contrast, economists who do not endorse ‘pure’ neoclassical economics argue that the introduction of workers’ compensation with experience rating should improve safety because in the absence of the program the limitations of knowledge and mobility and the unequal bargaining power of employees mean that the risk premiums generated in the labor market are inadequate to provide employers with the safety incentives postulated by the pure neoclassical economics approach. These economists argue that experience rating should improve safety by providing stronger financial incentives to employers to avoid accidents than the muted incentives provided by risk premiums.

A number of studies of the workers’ compensation program provide evidence that should be helpful in evaluating the virtues of experience rating. However, the evidence is inconclusive. A survey of the literature by Boden (1995, p. 285) concluded that ‘research on the safety impacts has not provided a clear answer to whether workers’ compensation improves workplace safety’. In contrast, Thomason (2005, p. 26) asserted that most (11 of 14) studies he surveyed found that experience rating improves safety and health and that the studies failing to detect the relationship were methodologically weaker than the other studies. Thomason concluded (2005, p. 26): ‘Taken as a whole, the evidence is quite compelling: experience rating works’. Tompa et al. (2007, p. 91) also surveyed the literature and found moderate evidence that the introduction of experience rating reduces the frequency of injuries (although severity may increase) and moderate evidence that the degree of experience rating reduces the frequency or severity of injuries. Some estimates of the magnitude of the safety effect from industry-level and firm-level experience ratings are substantial: Durbin and
Butler (1998, pp. 78–9) suggested that a 10 per cent increase in workers’
compensation costs countrywide between 1947 and 1990 was associated
with a 12.9 per cent decline in workplace fatalities.

In Canada, workers’ compensation programs have only recently intro-
duced experience rating and in most provinces there are restrictions on its
application which substantially limit its effectiveness (Thomason 2005, p.
28). However, a recent study of the British Columbia workers’ compensa-
tion program found that the frequency for medical-only and short-term
disability claims was reduced following the introduction of experience
rating in 1986 (Campolieti et al. 2006).

Thomason (2005, p. 196) cautioned that experience rating may, in addi-
tion to encouraging employers to improve workplace safety and health,
also lead to increased claims management by employers, including the
denial of legitimate compensation claims. Thomason and Pozzebon (2002)
surveyed 450 firms in Quebec and concluded that experience rating encour-
aged employers to improve workplace safety, to speed injured workers’
return to work and to challenge workers’ claims (which represents an
example of ‘moral hazard’ involving employers). While the evidence sug-
gests that on net experience rating is associated with accident prevention
activities by employers, further studies of the conflicting reactions by
employers would be useful.

There is also evidence that the presence of workers’ compensation ben-
efits leads to changes in worker behavior consistent with the concern of
neoclassical economists about moral hazard. Thomason and Burton (1993,
p. S8) summarized a number of studies that found the reported frequency
and severity of workers’ compensation claims increase in response to
higher benefits, which suggests that a moral hazard problem exists. Recent
studies include Neuhauser and Raphael (2004), who found that for injured
workers affected by benefit increases in California there was some evidence
they were more likely to file for benefits and clear evidence the duration
of their benefits was increased, and Bolduc et al. (2002), who found that,
among Quebec construction workers, increases in benefits resulted in
more reports of injuries, especially for difficult-to-diagnose injuries such as
lower-back disorders.

Caution is needed in interpreting these studies, however, since the
increased frequency or severity reported in the claims can result from a
‘true injury effect’ (workers take more risks as a result of higher benefits
and as a result actually experience more injuries) or from the ‘reporting
effect’ (workers report claims that would not have been reported as a
result of the higher benefits, and/or extend their period of reported dis-
ability because of the higher benefits). Most studies of the relationship
between workers’ compensation benefits and the frequency and severity
of claims have not distinguished between the true injury and the reporting effects. Durbin and Butler (1998, p. 67) concluded that the latter effect dominates, which implies that the concerns of some economists that the use of workers’ compensation benefits to provide *ex post* compensation for injured workers will lead to more injuries is exaggerated. As a result, even though the gross effect of higher workers’ compensation benefits is to increase the frequency and duration of workers’ compensation claims, once the contribution of the reporting effect is subtracted, the net effect of higher benefits appears to be an improvement in workplace safety since the favorable effects of experience rating on employer safety efforts dominate the deleterious true injury effect for workers.

C. General assessment of safety

Burton and Chelius (1997, p. 283), after analyzing the contributions of labor markets, regulation, tort suits and workers’ compensation to promoting workplace safety, concluded: ‘there is [no] single prevention approach (such as laissez-faire economics or workers’ compensation) that is sufficient to produce [an] optimal level of safety . . . [Rather] what appears to work best to improve safety and health is a rather eclectic mix of approaches . . . ’. They also attempted to relate the various approaches to prevention to several variants of economic or legal theories. Thus, reliance on the labor market to generate risk premiums was examined using pure neoclassical economics theory; government promulgation of information about risks and remedies was analyzed using modified neoclassical and old institutional economics theories; changes in liability rules were considered using new institutional economics; tort suits by use of law and economics theory; and the use of government standards by government mandate theory (which is basically a legal theory). Burton and Chelius (1997, p. 282) were least impressed by the arguments and evidence pertaining to pure neoclassical economics and the government mandate theories and concluded that, among the remaining theories, no single theory would suffice as the basis for designing government policies that would optimize the amount of workplace safety. Instead, they suggested an eclectic mix of theories was needed, an adage that has not lost its luster over time.

---

20 Is there a moral difference between increases in frequencies or durations of claims resulting from the ‘true injury effect’ and from the ‘reporting effect’? Arguably, more actual injuries are deplorable while more reporting of existing injuries is acceptable or even desirable from an economic approach since higher prices (higher benefits) should result in more supply (reports of existing injuries). However, Bolduc et al. (2002, p. 650) describe the reporting effect as ‘ex post moral hazard (that is, insurance fraud)’.
4 Cash benefits
Cash benefits paid in 2005 in US workers’ compensation programs were US$29.1 billion, which represented 52.6 per cent of the US$55.3 billion of total benefits (cash, medical and rehabilitation) paid that year.\(^{21}\) Cash benefits were $0.56 per US$100 of covered wages in 2005, down significantly from the peak of $0.99 per US$100 of payroll in 1991 (Sengupta et al. 2007, tables 1 and 12). The national average of cash benefits incurred in 2003 was US$25,646,442 per 100,000 workers, with a range from US$9,155,375 in Indiana to US$44,617,707 in California (Burton and Blum 2007, table 1).\(^{22}\) Cash benefits (including survivor benefits and other benefits) in 2005 in Canadian workers’ compensation programs were C$5.245 billion, which represented 77.5 per cent of the C$6.767 billion of total benefits (cash, medical, rehabilitation and other) paid that year (Institute for Work and Health 2008).

A Operational approaches
There are three basic operational approaches to cash benefits used in workers’ compensation programs (Burton 2008). The impairment approach determines the amount of benefits by measuring or rating the extent of the worker’s medical impairment. The loss of earning capacity approach determines the amount of benefits by measuring or rating the extent of the loss of earning capacity based on the worker’s medical impairment, as well as other factors such as the worker’s education and job experience and the availability of suitable jobs. A worker with no actual loss of earnings can nonetheless receive workers’ compensation benefits under the impairment approach or the loss of earning capacity approach. In contrast, under the wage loss approach, the worker must have an actual loss of wages in addition to an impairment and a loss of earning capacity. The wage loss approach varies the amount of cash benefits over time as the extent of the worker’s actual loss of wages varies. Burton argues that in US workers’ compensation programs, the purpose of all three operational approaches is to compensate workers for their actual loss of earnings and that the impairment ratings or the loss of earnings capacity ratings, which can be termed ‘disability ratings’, are serving as proxies for the presumed losses of actual earnings associated with those ratings.

\(^{21}\) Paid benefits in 2005 are all benefits paid in that year including benefits for injuries that occurred in earlier years.

\(^{22}\) Incurred benefits in 2003 include estimates of all the benefits that will eventually be paid for injuries that occurred in 2003. The national average for incurred benefits is based on data from 46 states, excluding the five states with exclusive state funds (North Dakota, Ohio, Washington, West Virginia and Wyoming).
B Types of cash benefits

(i) Temporary total disability benefits\textsuperscript{23} Temporary total disability (TTD) benefits were received in 61 per cent of all cases in which workers received cash benefits, but accounted for only 15 per cent of all payments for cash benefits in the US (Blum and Burton 2007, tables 1 and 5). Temporary total disability benefits are based on the wage loss approach. The TTD benefits are paid in most jurisdictions between the date of injury and the date of maximum medical improvement (date of MMI). All states have a waiting period of three to seven days before the TTD benefits begin and have a retroactive date after which the benefits for the waiting period are paid if the worker is still not back to work. In Canada, the waiting period in most provinces is one day. The waiting period/retroactive payment feature is a deductible. In most states, TTD weekly benefits are two-thirds of the worker’s pre-injury gross weekly wage. In six states, TTD weekly benefits are either 75 per cent or 80 per cent of the worker’s pre-injury spendable earnings, which are defined as gross wages minus deductions for state and federal income taxes and the worker’s contribution for the social security program. In the Canadian provinces, TTD weekly benefits generally range from 80 to 90 per cent of net earnings. The loss of wages absorbed by the worker is co-insurance. All states and provinces have a maximum weekly benefit. Some states also have a maximum duration for TTD benefits, even if the worker is still totally disabled and has not reached the date of MMI.\textsuperscript{24}

(ii) Permanent partial disability benefits Permanent partial disability (PPD) benefits were received in 37 per cent of all cases in which workers received cash benefits, but represented almost 73 per cent of all cash benefits in the US (Blum and Burton 2007, tables 2 and 5).\textsuperscript{25} PPD benefits, which are paid to workers whose injuries have consequences that are permanent but not totally disabling, constitute the most complex type of

\textsuperscript{23} Some but not all states also provide temporary partial disability (TPD) benefits which are paid to workers after they return to work prior to the date of maximum medical improvement with earnings less than their pre-injury wages. The TPD benefits replace a portion of the worker’s lost earnings subject to a maximum weekly amount.

\textsuperscript{24} Benefits for temporary total disability in US and Canadian jurisdictions are summarized in US Chamber of Commerce (2007, charts VI and IX).

\textsuperscript{25} The 73 per cent of all cash benefits in cases paying PPD benefits includes any TTD benefits paid in those cases. Likewise, the 9 per cent of all cash benefits in cases paying PTD benefits includes any TTD and PPD benefits paid in those cases, and the 3 per cent of all cash benefits in cases paying death benefits includes any TTD, PPD and PTD paid in those cases.
benefit provided by workers’ compensation. Depending on the jurisdiction, PPD benefits are based on the impairment approach and/or the loss of earning capacity approach and/or the wage loss approach. One feature of the systems of PPD benefits that rely on the wage loss approach is that cases are often closed with compromise and release (C&R) agreements in which workers release their claim to future benefits in exchange for a lump-sum settlement which essentially transforms the wage loss approach into the loss of earning capacity approach (Burton 2005a, p. 95). The increasing reliance on C&R agreements has been documented by Torrey (2007).

The predominant purpose for PPD benefits in the US is to compensate for work disability: the presumed or actual loss of earnings resulting from the work-related injury. Another possible purpose is given little homage in the US, namely to compensate for non-work disability. Only a few jurisdictions, notably Florida from 1979 to 1993, provided benefits for non-work disability in addition to the benefits paid for work disability (Burton 2005a, pp. 93–4). This dual benefits approach has, however, been adopted by several Canadian provinces, including Ontario, which has a separate track of benefits for non-economic loss.

(iii) Permanent total disability benefits Permanent total disability (PTD) benefits were received in less than 1 per cent of all cases in which workers received cash benefits, but represented 9 per cent of all cash benefits in the US (Blum and Burton 2007, tables 3 and 5). In most states, some PTD benefits are based on the impairment approach, since specified serious injuries (such as loss of both arms) are conclusively presumed to constitute permanent total disability. Most states also use the wage loss approach and provide PTD benefits when the worker has no earnings and a serious loss of earning capacity due to the workplace injury is established.

PTD weekly benefits are two-thirds of the worker’s pre-injury gross weekly wage in most states, although six states and the provinces base the PTD weekly benefits on the worker’s pre-injury spendable or net earnings. All jurisdictions have a maximum weekly benefit. Most states pay PTD benefits so long as the worker is totally disabled, although some states have a maximum duration even if the worker is still totally disabled. The provinces pay PTD benefits for life or until the worker reaches a specified age.27

26 The 1979 Florida workers’ compensation statute includes wage-loss benefits which represented the wage-loss approach, and impairment benefits which were designed to compensate for the consequences of a workplace injury other than loss of earnings (Berkowitz and Burton 1987, chapter 9.)

27 Benefits for permanent total disability in US and Canadian jurisdictions are summarized in US Chamber of Commerce (2007, chart VI).
(iv) **Death benefits**  Death benefits were received in less than 1 per cent of all cases receiving cash benefits, but represented 3 per cent of all cash benefits in the US (Blum and Burton 2007, tables 4 and 5). Weekly benefits are two-thirds of the deceased worker’s pre-injury gross weekly wage in most states, although some states vary the weekly amount depending on the number of dependants and five states and the provinces base the weekly benefits on the deceased worker’s pre-injury spendable or net earnings. All jurisdictions have a maximum weekly benefit. Most states pay death benefits so long as the surviving spouse is alive and unmarried together with shorter durations for other dependants such as children. The provinces provide either benefits for life or until the survivor reaches age 65.28

C Adequacy

The National Academy of Social Insurance (Hunt 2004) discussed various meanings of adequacy of workers’ compensation cash benefits, of which three approaches were examined in Hunt (2005). The statutory benefits approach compares the benefits specified by a state’s statute to the benefits provided by other states or to the income required by some absolute standard. For example, the national average of TTD benefits rose from 80 per cent of the poverty level in 1972 to about 107 per cent of the poverty level in 1998 (Hunt 2005, figure 1). The second approach was to compare the costs of the benefits prescribed by a state’s statute for the different types of benefits (TTD, PPD, PTD and Fatal) to the costs of the benefits prescribed by the Model Act (Revised) published by the Council of State Governments. Between 1972 and 1998, for example, PPD benefits provided by the states on average only increased from 43 per cent to slightly over 50 per cent of the Model Act’s standard (Hunt 2005, figure 2).

The third approach to assessing adequacy are wage loss studies, which compare the workers’ compensation benefits received by a sample of injured workers to the actual wage losses experienced by those workers. Wage losses are calculated by comparing the wages actually received by workers after their injuries with estimates of the potential wages the workers would have received if they had not been injured, using uninjured workers as a control group to produce estimates of these potential wages. One issue in the wage loss studies concerns the value of the replacement rate (benefits divided by earnings losses) that is considered adequate. The National Commission (1972, pp. 18–20) indicated that replacement of two-thirds of lost earnings was the appropriate standard for TTD, PTD

---

Workers' compensation and death benefits, but did not specify an adequacy standard for PPD benefits. A Study Panel of the National Academy of Social Insurance, while acknowledging the decision was somewhat arbitrary, endorsed ‘the historical standard of replacing two-thirds of gross wages as a measure of benefit adequacy’ for PPD benefits (Hunt 2004, p. 128). In the last decade, there have been a series of wage loss studies which are summarized by Boden et al. (2005, table 3.4). The replacement rates for the ten years after injury were 46 per cent in New Mexico, 41 per cent in Washington, 37 per cent in California, 36 per cent in Oregon and 30 per cent in Wisconsin. The authors concluded the ‘replacement rates do not approach the benchmark for adequacy’ (Boden et al. 2005, p. 60).  

There is a fourth possible test for adequacy which is implicit in the earlier discussion of who pays for the coverage. The tradeoff between higher workers’ compensation benefits and reductions in wages was interpreted by Moore and Viscusi (1990) and Viscusi (2004) as evidence of the desirability of workers’ compensation. Arguably, benefits are adequate from the standpoint of workers when the marginal dollar increase in expected benefits is associated with a dollar decrease in pre-injury wages. While benefits were thus inadequate using this approach between 1976 and 1983, benefits were more than adequate (‘above the efficient insurance amount’) as of 1997. However, because there was a substantial decline in cash benefits per 100,000 workers in 1982–4 dollars between 1989 and 1999 (Guo and Burton 2008, figure 2), the implication of the Viscusi and Moore results that workers’ compensation benefits have become more adequate over time needs to be used with caution.

D Equity

The concepts of horizontal and vertical equity have a long history in the public finance literature and were added to the criteria for evaluating workers’ compensation benefits (Musgrave 1959, p. 20) provides an example: ‘A proper definition of income is important, not only to establish equity in a vertical sense – that is, to plan taxes and transfers so as to adjust relative positions; it is important also to establish

---

29 Berkowitz and Burton (1987, p. 372) conducted a wage loss study of workers injured in 1968 whose wages and benefits were tracked through 1973: the replacement rates were 75 per cent in Wisconsin, 59 per cent in Florida and 46 per cent in California. Because they used a more lenient test of adequacy that benefits should replace at least 50 per cent of lost wage, they concluded that for these three states as of 1968 the benefits were generally adequate. They indicated: ‘A more lenient test for adequacy, which may be particularly appropriate for 1968–1973, a period when wages probably grew more rapidly than was expected as of 1968, is that at least 50 per cent of earnings losses calculated on the basis of potential earnings should be replaced by benefits’.

30 Musgrave (1959, p. 20) provides an example: ‘A proper definition of income is important, not only to establish equity in a vertical sense – that is, to plan taxes and transfers so as to adjust relative positions; it is important also to establish
workers’ compensation benefits in the 1972 report of the National Commission.  

A recent analysis (Burton et al. 2007) elaborated on several variants of equity and used data from the Berkowitz and Burton (1987) wage loss study of Wisconsin workers injured in 1968 to illustrate the concepts.

(i) Equity for disability ratings

A ‘disability rating’ is the impairment rating or the loss of earnings capacity rating which is serving as a proxy for the presumed losses of actual earnings associated with the rating. *Vertical equity for ratings* requires that actual wage losses increase in proportion to the increase in disability ratings. For all Wisconsin workers, there is a very close relationship between rating categories and percentage earnings losses, which meant that at this level of aggregation the rating system did an excellent job of providing vertical equity (Burton et al. 2007, figure C-7). *Inter-injury horizontal equity for ratings* require that the actual wage losses for workers with the same disability ratings, but different types of injuries, should be the same or similar. There were, however, significant differences in Wisconsin among the types of injuries in the relationships between disability ratings and lost earnings (Burton et al. 2007, figure C-6). *Intra-injury horizontal equity for ratings* requires that the actual wage losses for workers with the same disability ratings and the same type of injury should be the same or similar. However, there were substantial variations in earnings losses among workers with upper extremity injuries with the same disability rating (Burton et al. 2007, figure C-5).

(ii) Equity for replacement rates

The horizontal and vertical equity tests can also be applied to replacement rates (the proportion of lost wages replaced by benefits). *Vertical equity for benefits* has two possible meanings. A narrow view is that all levels of severity should have the same proportion of earnings losses replaced by benefits. An alternative view is that more serious injuries should have a higher replacement rate. This view corresponded to the general results found in Wisconsin (Burton et al. 2007, figure C-10). *Inter-injury horizontal equity for benefits* requires that the replacement rates for workers with the same disability ratings and different equity in a horizontal sense – that is, to give equal treatment to people in equal positions’.

31 The National Commission’s glossary contained this entry (1972, p. 137): ‘equitable: delivering benefits and services as judged by the program’s consistency in providing equal benefits and services to workers in identical circumstances and its rationality in providing benefits and services in proportion to the impairment or disability for those with different degrees of loss’.
types of injuries should be the same or similar. In Wisconsin, there were significant lapses for this test (Burton et al. 2007, figure C-9). *Intra-injury horizontal equity for benefits* requires that the replacement rates for workers with the same disability ratings and the same type of injury should be the same or similar. The Wisconsin results on this test were so dismal they were not even reported.

(iii) Assessments of equity Berkowitz and Burton (1987, p. 378) in their examination of workers injured in 1968 concluded that ‘the deficiencies on the equity criterion are much more serious than the deficiencies on the adequacy criterion’. Subsequent studies have paid more attention to adequacy than equity, although there are some important exceptions. Park and Butler (2000) examined the ability of the impairment ratings used in the Minnesota workers’ compensation program to predict earnings losses in each of the four years after injury. In regressions that included the impairment rating, the square of the rating and a set of control variables, only from 2.5 to 5 per cent of the variations in wages among workers could be explained. The Minnesota rating schedule appears at best to provide a minimal degree of vertical equity, but no semblance of horizontal equity. Reville et al. (2005, pp. 52–7) examined the California PPD rating schedule and found that earnings losses for four types of injuries increased with the disability rating, but that disparities existed in the earnings losses for the four types of injuries with similar ratings. These results suggest that the California PPD rating schedule met the test for vertical equity but failed the test for inter-injury horizontal equity for ratings.32

These various studies of rating systems suggest that lack of equity is probably a pervasive property of workers’ compensation programs. Reville et al. (2005) argue that the results of wage loss studies should be used to recalibrate the disability ratings to produce a better match between the ratings and the subsequent loss of earnings, and this is a fruitful topic for further research.

32 Sinclair and Burton (1995) examined the relationship between the permanent impairment (PI) ratings from the AMA Guides to the Evaluation of Permanent Impairment, third edition (Engelberg 1988) (AMA Guides) for 78 conditions and the quality-of-life ratings assigned to those conditions by a large sample of injured Ontario workers (as well as a smaller group of Ontario residents who were not injured). The results indicated that the AMA Guides lack both vertical equity and inter-injury horizontal equity tests for ratings used as proxies for non-work disability.
5 Medical care and rehabilitation services

Medical benefits (including rehabilitation services) paid in 2005 in US workers’ compensation programs were US$26.2 billion which represented 47.4 per cent of the US$55.3 billion of total benefits (cash plus medical) paid that year. Medical benefits were only 36.3 per cent of total benefits in 1987. Medical benefits were $0.50 per US$100 of covered wages in 2005, down from the peak of $0.69 per US$100 of payroll in 1992 (Sengupta et al. 2007, tables 1, 4 and 12). Medical and rehabilitation benefits in 2005 in Canadian workers’ compensation programs were C$1.522 billion, which represented 22.5 per cent of the C$6.767 billion of total benefits paid that year (Institute for Work and Health 2008).

A Medical benefits in theory

In economic theory, an employer would devote resources to medical care and rehabilitation services in order (1) to hasten the worker’s return to work and thus reduce the payment of cash benefits, (2) to ameliorate the worker’s medical condition and thus reduce further medical expenses, and (3) to limit the extent of the worker’s disutility resulting from the permanent impairment and thus reduce the risk premium required to attract workers to the firm. Equilibrium is reached when the marginal expenditures on medical care and rehabilitation services equals the marginal reductions in costs for these three factors.

The assumptions underlying this theory can be challenged. Most PPD benefits are based on an impairment rating or a loss of earning capacity rating rather than on actual wage loss, which means that, in general, employers do not reduce benefit payments by returning a worker with permanent consequences of his or her injury to work. In addition, the widespread use of compromise and release agreements means that once the worker agrees to a settlement, the employer is likely to lose interest in the worker’s subsequent medical recovery. Despite the limitations of the theory, research on the medical and rehabilitation components of workers’ compensation would benefit from the use of an analytical framework similar to that used for the study of compensating wage differentials.

B Medical benefits in practice

Workers’ compensation in the US essentially operates its own health care system, which is separate from the health care system used by workers for non-occupational injuries. There are overlaps between the providers and facilities used by the two systems, but workers’ compensation has its own rules for access and financing of health care. Medical benefits in workers’ compensation are provided without deductibles or co-insurance payments and, with very limited exceptions, workers do not pay any of the nominal
premiums. The employer is also required in most states to provide medical benefits so long as they are medically justified, which can be for the rest of the worker’s life. Medical rehabilitation services must also be provided by the employer in most states, while vocational rehabilitation requirements vary considerably among jurisdictions. Cost containment mechanisms used in workers’ compensation to attempt to slow the growth of health care costs, including medical fee schedules, limits on employees’ choice of treating physicians and managed care, are discussed below.

In Canada, workers’ compensation programs provide full medical benefits without cost to the injured worker. The worker has the right to choose the treating physician, who is normally the worker’s regular provider of health care. In some provinces, such as British Columbia, the physicians submit all bills directly to the workers’ compensation agency. In Ontario, the physician is reimbursed by the provincial single-payer health care plan, which, in turn, bills the workers’ compensation agency for reimbursement. The expenses of care provided by other health care professionals, hospital care and pharmaceuticals are paid directly by the Ontario agency (Mustard and Sinclair 2005, pp. 200–01).

C Performance studies
Numerous studies have examined the performance of the workers’ compensation health care system. Many of these studies compare the medical outcomes for workers’ compensation beneficiaries with other patients and usually find poorer outcomes for the former. Harris et al. (2005) conducted a meta-analysis of 211 studies of surgery outcomes and found only one that described a better outcome for workers’ compensation patients. Margoshes and Webster (2000) also reviewed a number of health care studies and found that cases covered by the workers’ compensation system have quantitatively and qualitatively worse outcomes for cost and duration of disability than cases covered by other insurance systems. A study illustrating the importance of proper methodology is Johnson et al. (1998), who emphasized that data showing the proportion of injured workers who returned to work can be misleading because a substantial proportion of those workers had one or more subsequent work absences caused by the effects of their original injuries.

Several studies by economists have shown that the costs of medical care are higher in workers’ compensation than in other insurance arrangements. Johnson et al. (1993) and Baker and Kruger (1995) found that

---

33 Rehabilitation benefits in US and Canadian jurisdictions are summarized in US Chamber of Commerce (2007, chart X).
total medical charges for equivalent medical conditions were higher for a
workers’ compensation carrier than for non-occupational injuries covered
by Blue Cross/Blue Shield because charges per unit of health care service
were higher in workers’ compensation. However, in a subsequent study of
California by Johnson et al. (1996), the source of the higher total charges
in workers’ compensation was greater use of resources per case rather than
higher charges for particular services.

D Policies to improve performance
Burton (1997) examined the surge in policies to control medical costs that
was triggered by the rapid increase in those costs after the mid-1980s. Fee
schedules for services specifying maximum reimbursement rates for serv-
ces provided by hospitals, physicians and other health care providers were
used in most states. Another strategy widely adopted by states is the limi-
tation on employee choice of physician. Neumark et al. (2007) examined
the impact of variants of provider choice in California, Massachusetts,
Pennsylvania and Texas on workers injured in the late 1990s. When
workers (as opposed to employers) chose the provider, the results were (1)
higher costs including cash benefits and medical payments per claim; (2)
worse return-to-work outcomes; (3) the same rate of physical recovery as
measured by workers’ responses to questions about their health status; and
(4) higher satisfaction with their medical care. These results are sensitive to
the particular variant of employee choice: for example, the higher costs and
worse return-to-work results arose primarily when employees chose a new
provider rather than one with whom they had a previous relationship.34

Another response to surging health care costs in workers’ compensa-
tion was the importation of some of the cost-containment approaches that
have been widely used in the general health care system, including health
maintenance organizations (HMOs) and preferred provider organizations
(PPOs), in place of the traditional fee-for-service arrangements. Borba et
al. (1994) studied two managed care pilots in Florida and found that an
HMO reduced total medical costs by 54 per cent and a PPO reduced costs
by 23 per cent. Johnson et al. (2006) studied the use of health care net-
works in California, Connecticut and Texas and found that network cases
had lower health care costs, shorter durations of work absence and lower
cash benefits than non-network cases. Bernacki and Tsai (2003) reported

34 As described by Neumark et al. (2007), the 2004 California law allows
employers to establish networks composed of occupational and non-occupational
physicians and to control which physicians are in the network. The employee can
only select a physician out of the network who was previously the worker’s primary
provider under an employer-provided group health plan.
that the use of an integrated workers’ compensation management system at Johns Hopkins University reduced the frequency of claims as well as the medical costs, cash benefits and administrative costs per US$100 of payroll, even though workers had free choice of their medical providers.

The research on the use of managed care in workers’ compensation suggests that considerable savings in health care costs are possible; the duration and, in some cases, the frequency of lost-time cases can be reduced, resulting in lower payments for cash benefits; the savings usually (but not always) require some limitations on workers’ choices of treating physician; and workers are less satisfied with health care arrangements that constrain their choice of treating physician. These research results can be used to fine-tune policies such as the suggestion by Neumark et al. (2007, p. 140) that an employee should be permitted to choose the health care providers for workplace injuries when the providers previously treated the worker.

Another use of research results as a basis for policy changes involves the finding that return-to-work results are enhanced when injured workers return to the employer where they were injured. Galizzi and Boden (2003) examined workers injured in Wisconsin and found that returning to the pre-injury employer is one of the main determinants of the speed of return to work. Campolieti and Krashinsky (2006) found that workers in Ontario who returned to their time-of-accident employers earned 27 per cent more than other workers. These results support the analysis of Boden et al. (2005, p. 61), who endorsed policies such as the Oregon Employment-at-injury Program which provides subsidies to employers who offer modified or light-duty jobs to employees before they have fully recovered from their workplace injuries.

E Separate health care systems
As described earlier in this section, the workers’ compensation health care system is largely separate from the general health care system in the US, while in Canada there is more integration, particularly in the use of the same physicians to treat work-related and other sources of disorders. One issue is the extent to which costs that should be paid for by the workers’ compensation program are borne by the general health care system and vice versa. In the US, Burton and Spieler (2001, p. 74) argued that the growing restrictions on both compensability and medical care in the 1990s were likely to transfer health care costs from workers’ compensation to Medicare and Medicaid. The argument could be extended to include shifting of work-related medical costs to other health care systems. However, Butler et al. (1997) found that doctors in HMOs have a greater tendency to classify claims as compensable under workers’ compensation than do other physicians and
the rapid expansions of HMOs from 1980 to 1990 resulted in a significant increase in workers’ compensation frequency. So cost shifting both from and to workers’ compensation may have occurred in the US. In Canada, Koehoorn et al. (2006) found that hospital workers in British Columbia with work-related musculoskeletal disorders who received workers’ compensation benefits were associated with increased use of the general health care system both prior to and following their workplace injury.

In addition to the issue of whether the costs of work-related and other injuries are being shifted to the ‘wrong’ health care system, there is also the problem of the transactions costs associated with deciding whether the injury is work related, especially since the legal rules used to decide whether some medical conditions, notably back disorders, are basically arbitrary and unscientific (Burton 1985). Concern over the issues of cost sharing in conjunction with the rapid increase in health care costs beginning in the late 1980s help explain the emergence in the early 1990s of an interest in 24-hour coverage, which was defined by Burton (1997) as ‘efforts to reduce or eliminate the distinctions between benefits and services provided to disabled workers for work-related injuries and diseases, and benefits and services provided for non-work-related conditions’.

Burton catalogued four primary variants of 24-hour coverage, which ranged in terms of the complexity of the legal and regulatory environment in which they must operate from Integration within Self-Insuring Employers to Integration of Benefits and Services through General Health Care Reform. An example of the latter was the health care reform proposal made by the Clinton administration in the early 1990s which would have essentially folded the workers’ compensation health care system into the general health care system. One argument made by those opposed to the Clinton proposal was that carriers and employers would lose their ability to control the treatment of injured workers and to provide the care necessary to expedite the worker’s rehabilitation and return to work. As a result, any possible savings in health care costs would be more than offset by higher expenditures on cash benefits. A counter-argument can be made based on the experience in Ontario, where medical care provided by physicians is essentially the same for work-related and non-work-related injuries and where, in comparison to most US jurisdictions, workers’ compensation cash benefits achieved a higher wage replacement rate, medical care expenditures were a smaller proportion of premium revenue, and insurance costs for employers were in the lowest quartile of North American jurisdictions (Thomason and Burton 2002; Mustard and Sinclair 2005, pp. 201–02). While the Ontario example may not conclusively prove that an integrated health care system for work-related and other medical conditions can reduce the costs of both medical care and cash benefits compared to a
system that separates work-related medical care from other types of cases, at least the argument that an integrated system would be more expensive because of the increased expenditures on cash benefits can be challenged. This topic warrants additional research.

6 Delivery system
The benefits and services in workers’ compensation are provided by a delivery system comprised of employers, insurers, administrative agencies, attorneys, doctors and other participants. Berkowitz and Burton (1987, pp. 26–8) evaluated the efficiency of this delivery system by examining the relationship between two variables. One variable measures the administrative costs of providing benefits incurred by the participants in the delivery system. The other variable measures the quality of the benefits provided by the program, where quality is assessed on the basis of criteria such as the adequacy and equity of cash benefits. They argued that the appropriate test for delivery system efficiency, which they called ‘panoramic efficiency’, is that benefits of a particular level of quality are provided with the least possible administrative costs.35

A Delivery system models
Berkowitz and Burton (1987) and Burton (2005a, pp. 98–101) identified three delivery system models used to provide workers’ compensation benefits. One model relies on an active government agency that makes many decisions itself, closely supervises the operation of employers and carriers and limits the roles for attorneys. Wisconsin and the Canadian provinces are examples of jurisdictions using this approach. A second model relies on private parties, particularly attorneys, to make most of the decisions about benefit payments. The agency is essentially passive, although it will resolve disputes brought to it by the private parties. A third and intermediate model involves a state agency that conducts a minimal review of decisions made by the private parties and that resolves disputes in a relatively high proportion of cases, but that nonetheless relies on extensive attorney involvement to make the delivery system operate. Most states currently appear to use either the second or third model.36

35 An inappropriate meaning of delivery system efficiency, which they termed ‘myopic efficiency’, is only concerned with reducing administrative costs without concern for the quality of the program (Berkowitz and Burton 1987, pp. 27–8).

36 Roberts (2003) is one of the few studies of the relationship between the activities of state workers’ compensation agencies and program outcomes. She found, for example, that the greater the level of state approval required during the claims process, the better the benefit levels for workers.
The extensive involvement of attorneys is the distinguishing feature of the second and third models. Whether attorneys help achieve the equity and adequacy of benefits is not clear \textit{a priori}. The limited evidence suggests that the extensive reliance on attorneys undermines delivery system efficiency.\textsuperscript{37} Berkowitz and Burton (1987) concluded that Wisconsin had a better record on delivery system efficiency than Florida and California because Wisconsin combined lower administrative costs, including attorneys’ fees, with benefits that were more adequate and equitable. Thomason and Burton (1993) studied the effect of attorney involvement on the outcomes of cases paying non-scheduled PPD benefits in New York and found that attorneys increased the probability of lump-sum settlements, reduced the amounts of those settlements and had no statistically significant effect on the size of litigated awards. Torrey (2007) has documented the increasing reliance of states on lump-sum settlements (or C&R agreements) to resolve disputed claims, which is troublesome given the limited evidence on the deleterious effects of such settlements. This is another area where additional research is warranted.

\textbf{B Insurance arrangements}

In the US, employers who are covered by workers’ compensation statutes can insure their responsibilities in two or three ways depending on the state (Sengupta et al. 2007, pp. 13–14). Private carriers can sell insurance in all but five states and accounted for 50.8 per cent of benefit payments in 2005. Five states have exclusive state funds and there are 21 competitive state funds that compete with private carriers. These 26 state funds plus the federal programs accounted for 25.3 per cent of benefit payments in 2005. Employers are allowed to self-insure if they have financial capacity in all but two states, an insurance arrangement that accounted for 23.8 per cent of benefits payments in 2005. In Canada, employers’ only insurance options are provincial funds and, to a limited degree, self-insurance.\textsuperscript{38}

One of the significant developments in the US workers’ compensation insurance market in the last 30 years was the emergence of several new competitive state funds, beginning with Minnesota in 1984 and followed

\textsuperscript{37} Wickizer et al. (2004, p. 337) found that workers who were less satisfied with the handling of their claims were more than three times as likely to retain an attorney or to file an appeal.

\textsuperscript{38} In both the US and Canada, some self-insurers handle the administrative functions, such as paying claims, while others hire a third party administrator (TPA) to perform these functions. In most provinces there is also self-insurance without self-administration where the employer is financially responsible for all of its own claims but the workers’ compensation agency administers the claims.
by seven other states in the 1990s. However, the Michigan competitive state fund was privatized in 1994, the Nevada exclusive state fund was privatized in 1999, and West Virginia is in the process of closing its state fund and admitting private carriers. Presumably a primary motive for these changes was to reduce the costs of workers’ compensation insurance. However, the cost-saving motive does not appear to be evidence based. Thomason et al. (2001) found no differences in insurance rates between states with exclusive state funds and states with private carriers after controlling for other factors that influence interstate differences in costs, such as injury rates and benefits levels. They also found that states with competitive state funds have insurance costs that are nearly 19 per cent higher than the costs in states that only have private carriers.

Another significant development in workers’ compensation insurance arrangements in the US in recent decades has been the deregulation of the markets in which private carriers operate. Rate setting in workers’ compensation was highly regulated – relying on administered pricing – until the early 1980s, when an initial phase of deregulation began. By the time California deregulated its market in 1995, almost all states had adopted some form of deregulation. Thomason et al. (2001) found that comprehensive deregulation – the use of loss costs (instead of manual rates that included loadings for overhead) that were not subject to prior approval by the state before carriers could establish the rates they would charge – reduced the costs of workers’ compensation insurance by about 11 per cent below the rates that would have been charged if states had continued to rely on administered pricing. They also found that partial deregulation – for example, states that continued to rely on manual rates but allowed carriers to deviate from those rates – resulted in higher insurance rates than would have been paid by employers under administered pricing. Thomason et al. (2001, p. 267) also concluded that workers’ compensation insurance rate regulation had little, if any, effect on injury rates. This suggests that deregulation of private carriers represents a gain in delivery system efficiency since costs to employers are lower without jeopardizing worker safety. This suggestion is reinforced by Barkume and Ruser (2001), who found that comprehensive deregulations reduced premiums by 13.7 per cent and reduced injury rates by up to 8.2 per cent.

7 Conclusions
There has been an impressive increase in the research of workers’ compensation programs in the last 35 years. For example, the National Commission on State Workmen’s Compensation Laws (National Commission 1972, p. 96) noted there had been few systematic attempts to evaluate the effect of experience rating on safety. Subsequently there have been so many studies
that several surveys of the literature have been published. And in recent decades there have been many studies of such important topics as the lack of workers’ compensation coverage for many injured workers and the effects of policies designed to limit the costs of medical care.

Nonetheless there are topics where research is sparse. There are only a few states in which the adequacy and equity of cash benefits have been examined using data on actual earnings losses and benefits received by injured workers. There is evidence indicating that the disability ratings used to determine PPD benefits in most states have serious equity problems, but almost no research to identify factors in addition to impairment, such as age or education, that would improve the ability of disability ratings to predict earnings losses. Much research focuses on a single jurisdiction, thereby ignoring the considerable differences in workers’ compensation programs among states and provinces that may hold the key to identifying policies improving the overall operation of the program. And very few studies have examined both workers’ compensation and other programs that provide cash and medical benefits for disabled workers as well as the relationships among these programs.

There are obstacles to conducting research. Unlike unemployment insurance, consistent data for state workers’ compensation programs are not required by the US federal government. Moreover, most workers’ compensation data collected by the states are incomplete or incompatible across jurisdictions. In addition, most state workers’ compensation agencies do not have research units, which limits the ability of researchers to collaborate with persons who know the intricacies of the data maintained by the states. The workers’ compensation agencies in Canadian provinces generally have been much more willing to support research by their own staff or by outside scholars. For both the US and Canada perhaps the most encouraging conclusion is that, despite the obstacles to research, our understanding is much more complete now than a few decades ago of the accomplishments and deficiencies of workers’ compensation programs in achieving broad coverage, substantial income maintenance, sufficient medical care and rehabilitation services, encouragement of safety, and an effective delivery system.

Bibliography


Workers' compensation


