Assuring Gender Equity in Recruitment Standards for Police Officers

Roy J. Shephard and Jean Bonneau

Catalog Data

Key words: human rights, occupational fitness, police officers, test validity
Mots clés: les droits humaines, capacité du travail, policiers, validité des épreuves

Abstract/Résumé

Human Rights Tribunals require application of non-discriminatory fitness standards in the hiring, promotion, and retention of employees. This issue has become controversial for public safety officers such as police, where differences in average levels of absolute fitness between men and women cause a high proportion of female applicants to fail many entrance tests. The present review summarizes the impact on physical working capacity of commonly encountered gender differences in size, body composition, haemoglobin levels, and muscular strength. The principles applied in designing content- and construct-validity occupational fitness tests are described, and Human Rights policies are reviewed in the light of the Meteorin judgment. Criteria are indicated for establishing a bona-fide occupational fitness requirement, and description is given of the approach used in developing standards that satisfy these criteria. Requirements are based on the task to be accomplished. The potential training response of female applicants is likely at least to match that of their male peers, and the needs of female police recruits are thus best accommodated by providing every opportunity to augment fitness to the required minimum level. The main weakness of any current requirement is that most police forces do not yet apply an equivalent criterion to older incumbent officers, where similar issues may arise.

En évaluant l’engagement, la promotion et la retraite obligatoire du personnel, les Tribunals sur les Droits Humaines demandent l’application des épreuves de capacité physique

R.J. Shephard is with the Faculty of Physical Education & Health, and Graduate Programme in Exercise Sciences, University of Toronto, Toronto, ON; J. Bonneau is with the Royal Canadian Mounted Police, Ottawa, ON.
denue de toute discrimination. Cette exigence a cree des controverses, particulierement en engagent des agents de la securite publique, parce qu'une proportion elevee des candidats feminins ont rate les epreuves d'embauchage, a cause d'une difference de capacite physique assez importante entre les femmes et les hommes typiques. Cette revue examine l'influence des differences sexuelles de grandeur, composition corporelle, concentration de hemoglobine et de force musculaire sur la capacite physique des candidats. Les meres respectives des epreuves de type inductive ou de connaissance est discute, et l'influence de la decision Meiorin sur la politique de la Commission des Droits Humains est critiquee. Les criteres d'une exigence occupationnelle de bonne foi sont indiquees, et le development des mesures de capacite physique appropriees sont decrit. Les epreuves d'embauchage des policiers sont basees sur la tache typique, l'arrestation d'un jeune homme en bonne forme. La capacite d'entrainement physique des candidates feminins egale ceux des candidats masculins, et alors on peut accomoder les besoins des femmes en leur offrant les possibilites d'augmenter leur capacites physiques au niveau reclame. Neanmoins, le solidite des normes proposees est encore compromis par l'inobservance d'une proportion des titulaires.

General Considerations

The application of any occupational fitness standard is likely to have an adverse impact on the recruitment and retention of certain demographic groups, including women (Evans, 1980; Greenberg and Berger, 1983), older applicants, small men, and those with physical disabilities. Indeed, the primary purpose of most fitness tests is to discriminate against individuals with an insufficient functional capacity. The issue of employment standards for public safety officers has been reviewed previously in the context of constraints faced by the elderly employee (Shephard, 1987, 1990, 1991; note that the term officer is used in this review to signify an incumbent peace officer, irrespective of their rank within the police force). The present review focuses on appropriate standards for the recruitment and retention of female police officers, taking account of the decision of the Canadian Supreme Court in the Meiorin case, the first hearing to deal with gender discrimination and bona fide occupational requirements in Canada. The gender-related issues that arise are in many respects similar to those that affect the retention of older male employees. In both cases, test scores show a wide spread between individuals, so that the fittest women or the fittest elderly employees out-perform the average person in the comparison group. Moreover, the average difference in functional capacity between a female and a male applicant is of similar magnitude to the average difference between an elderly and a young employee. Thus, to avoid a charge of gender discrimination, it is important that elderly male incumbents meet the fitness standards expected of young female recruits.

Occupational standards are occasionally introduced with the intent of discriminating against women, the elderly, a specific ethnic group, or those with physical disability, but more commonly, application of what is thought to be a neutral and necessary standard has an adverse impact on one particular group of recruits or employees. In the case of police officers, there is an underlying public expectation that recruits should have a high level of physical fitness. In Canada, the Police Services Act (which governs the Organisation of Municipal and Provincial police services) specifies that "no person shall be appointed as a police officer unless he or she is physically and mentally able to perform the duties of the position, having
regard to his or her own safety and the safety of members of the public." Likewise, Section 9.1 of the Royal Canadian Mounted Police (RCMP) Act gives the RCMP authority not to hire applicants who are unfit.

After a brief discussion of gender differences and of the principles of occupational fitness assessment, the impact of the Meiorin case upon Human Rights standards is discussed. Bona fide occupational requirements are reviewed, and the corresponding principles are applied to the setting of entry requirements for police officers. Damage arising from failure to meet prescribed standards is discussed in the context of accommodating female applicants, and possible gender differences in training responses are explored.

**Gender Differences in Physical Working Capacity**

**EFFECTS OF SIZE AND BODY COMPOSITION**

A variety of inherent biological differences predispose to discrepancies of functional capacity between women and men: height, body mass, body fat content, haemoglobin level, cardiac dimensions, muscle dimensions, muscle composition, and vascular impedance. Nevertheless, the female disadvantage can be largely overcome by rigorous training, and a well-trained woman can out-perform a sedentary man. Stature is particularly important to the performance of both women and men, influencing body mass and thus inertia, leverage, reaching, jumping, absolute aerobic power, and absolute muscle strength. The typical gender difference in the size of serving police officers is illustrated by the data of Farenholtz et al. (1989). In this study, male officers had an average height of $1.81 \pm 0.06$ m, and females an average of $1.69 \pm 0.06$ m.

Other factors being equal, body mass is proportional to the third power of standing height, and the gender difference in standing height thus explains a large part of the corresponding difference in body mass; residual differences reflect larger muscles in the male and a higher percentage of body fat in the female. Farenholtz et al. (1989) found an average body mass of $85.9 \pm 9.5$ kg in male officers, and of $65.5 \pm 9.1$ kg in females; on the basis of height differences alone, a value of $70.0$ kg would have been expected in the women.

**GENDER DIFFERENCES IN OXYGEN TRANSPORT**

Treadmill or cycle ergometer measurements of absolute peak aerobic power show an average gender difference of 30–40%, favouring male candidates (Shephard et al., 1988). Much of this difference is size-related, but if the data are expressed as ml/min per kg of body mass in order to minimize size effects, the average female still faces a handicap from gender differences in the amount of adipose tissue (Shephard, 2000b); even in body builders who have deliberately minimized their body fat content, typical figures are 10% in women and 6% in men (Tesch, 1992). The gender discrepancy in peak aerobic power can thus be reduced further by expressing oxygen transport per litre of active leg muscle, or per kg of lean body mass.

Other factors contributing to the poorer inherent oxygen transporting ability of women include a lower haemoglobin level and a smaller heart (Gledhill and Warburton, 2000; Gledhill et al., 1999; Shephard, 2000a; Wiebe et al., 1999). On
average, a woman’s haemoglobin concentration is only about 88% of the male value, and the heart volume (and thus the likely peak systolic volume) shows at least the expected 16-20% size differential. Gender differences in oxygen transporting ability therefore persist, even if values for peak aerobic power are expressed in ml/min per kg of lean tissue. Discrepancies may become smaller when looking at tasks that involve relatively small muscle groups. Here, the pumping capacity of the heart becomes progressively less important to oxygen delivery. The force exerted by the active muscles now limits oxygen transport (Kay and Shephard, 1969). During resisted effort, this factor again places female officers at a disadvantage relative to males. The smaller muscles of the average woman must generally contract at a larger fraction of their maximal force than those of a man in order to accomplish a given task, and this tends to occlude the blood vessels supplying the active muscles.

During walking, running, and many other normal duties, a police officer must displace her or his entire body mass against gravity. Performance then depends on the peak aerobic power per kg of total body mass, and the average woman has a 10% disadvantage relative to a man of similar age and fitness. However, if aerobic effort must be exerted against some external object, as in pushing or dragging either a vehicle or a criminal, the relevant unit becomes the absolute oxygen transport, measured in L/min, and the female disadvantage rises to 30-40%. In many operational situations, the burden of the small person is also increased disproportionately by a fixed absolute weight of equipment (guns, belts, communication devices, helmets, and body armour).

Nevertheless, there are large inter-individual differences in maximal aerobic power, and some well-trained women can out-perform most men (Shephard, 1994, 1997a, 1997b, 1999, 2000a, 2000b, 2000c; Shephard et al., 1988). Thus, in a Cooper-type aerobic test, Parnholtz et al. (1989) found that some female police officers could complete the nominal 12 min run in as little as 8 min 58 sec, nearly matching the best male score of 8 min 33 sec.

**GENDER DIFFERENCES IN MUSCLE STRENGTH**

Because women generally have small hands and light bones, the arms of a woman may have a greater working capacity than a typical man, if expressed per unit of muscle volume. In most situations, however, the arms are working against some external resistance, and then the average woman is at a substantial disadvantage relative to a man because of smaller and weaker muscles. In subjects aged 20-29 years, the Canada Fitness Survey (Fitness and Amateur Sport, 1983) found that women had a handgrip force only 57% of the average male value. They also performed 27% fewer sit-ups and 19% fewer push-ups than men of similar age.

Gender differences in strength depend on the muscle groups that are examined. The largest discrepancies are for the arms and shoulders, and the smallest for the lower limb muscles (Sale, 1999; Shephard, 2000b; Tesch, 1992; Wilmore, 1974). It is thus important to look closely at the specific muscle groups used in policing, and to ensure that these are the muscle groups evaluated during entry tests (Bard et al., 1985).

The average cross-sectional area of individual muscle fibres is less in women than in men, and the peak force developed by a given cross-section of muscle
tissue is also less than in men because the muscles contain a larger amount of fat (Kanehisa et al., 1994, 1996; Maughan et al., 1984) and connective tissue (Miller et al., 1993). Women reach peak muscle force more slowly than men (Behm and Sale, 1994; Häkkinen, 1993; Komi, 1980). The peak anaerobic power and capacity of women are typically only about 70% of values found in their male counterparts (Karlsson and Jacobs, 1980). Absolute muscle endurance is greater in men than in women, but women can sustain contractions at a similar relative force to men (Maughan et al., 1986; West et al., 1995). The gender difference is smaller for eccentric than for concentric muscle contractions (Aura and Komi, 1986; Seger and Thorstensson, 1994).

It is difficult to devise any test of muscle performance where the setting of a reasonable minimum standard does not disqualify a large proportion of women. For example, Wilkie (1974) found that 50% of women but only 4% of men failed to surmount a six-foot (1.83 m) fence in 15 seconds, and 91% of women, but only 8% of men could not perform two pull-ups. Likewise, Masters athletes at all ages show a consistent 15% gender differences in long jump records (Spirduso, 1995).

Principles of Occupational Fitness Assessment

CONSIDERATIONS OF TEST DESIGN

Given the complexity and variability of police duties, it is difficult to specify a minimum standard of either aerobic or muscular fitness on theoretical grounds. From the viewpoint of Human Rights, it is desirable that test scores and associated judgments that form the basis of an occupational assessment should impact equitably on all demographic constituents of the potential labour force. Small discrepancies in average scores for sub-groups of the population are perhaps inevitable, but one commonly accepted standard of practice is that acceptance rates for any given sub-group should not fall below 80% of that for a peer group (for example, 90% of men and 72% of women should pass the test).

The characteristics of an ideal occupational fitness test are

- equal impact on all demographic segments of potential labour force
- economical to administer on a large scale
- scores objective, reliable and valid, well-correlated with actual job performance and simulated components of task
- scores independent of skills acquired "on the job"
- task items critical to successful job performance
- test environment realistic (e.g. clothing and equipment)
- test safe to perform
- test accommodates alternative task approaches.

Assessments should be economical to administer, but scores should show a good correlation with the performance of actual and simulated components of the work task. Scores should also be independent of skills acquired during experience of the job (Maher, 1984; Hogan and Quigley, 1986; Hogan, 1991; Arvey et al., 1992; Arvey et al., 1992; Anderson and Plecas, 1999). In the case of police officers, measurements should ideally be performed under emergency conditions, while candidates are wearing full emergency equipment, although safety concerns may
preclude too close a simulation of fights and extreme environmental conditions during testing. Items included in any assessment should be critical to the safety of the public, fellow officers and the individual (in that order); if the danger arising from failure to meet a particular requirement is high, it is reasonable to include it even if it arises only rarely during normal employment (Shephard, 1990).

Empirical tests that are chosen to evaluate recruits should be objective, reliable and valid. Clear instructions should be developed, with adequate guidance to appraisers, and where possible, scoring should accommodate alternative methods of performing a given task. Test candidates should also be given preliminary education sessions, support materials, training tips and opportunities to practice the required tasks. The test batteries may be designed for content or construct validity, or they may represent a hybrid of the two approaches.

CONTENT VALIDITY TESTS

Tests with content validity are based on field simulations of police work. They are well accepted by the labour force and by arbitrators, but are critically dependent on inclusion of relevant skills in the test battery. Critics have argued that in a job as varied as policing, there is a risk of omitting critical elements from the overall test battery. However, the required standards can be refined progressively to mimic actual movement patterns and to match the physical demands of the job (Pelot et al., 1999). It is important that task simulations also give opportunity for female candidates to utilize any methods of compensating for the heaviest components of job demands (Guion, 1978).

Despite the claims of Booth and Hornick (1984), one difficulty with many tests is that scores depend heavily on body size. Scores are also very susceptible to environmental factors. For example, push/pull tests on women may be invalidated because they lack sufficient body weight to avoid slipping on the smooth floor of a gymnasium (Gaul and Wenger, 1992).

Unfortunately, tests with content validity lack an obvious relationship to health, and thus have limited educational value in enhancing the lifestyle of either new recruits or incumbent police officers (Bonneau, 1986; Burelle et al., 1987). Such tests also require more equipment than simple "physiological" procedures such as a step test and sit-ups, and in theory there may be a greater potential for accidents and injuries during the field simulation of strenuous tasks (although in practice there have as yet been no serious injuries using the Police Officer Physical Ability Test (POPAT; Farenholtz and Rhodes, 1986) or the Physical Ability Requirement Evaluation of the Canadian RCMP (Health Service Directorate, 1996).

CONSTRUCT VALIDITY TESTS

Construct validity tests attempt to discover the underlying physiological demands of police work (e.g., aerobic power, muscle strength, flexibility and agility), and to base minimum test scores on an appropriately weighted combination of these fundamental characteristics (Greenberg and Berger, 1983). A suitable prediction equation is developed by stepwise multiple regression, canonical regression and/or principal component analyses relating the "physiological" measurements to performance criteria. For example, Greenberg and Berger (1983) used a regression
equation to predict the likelihood of restraining and apprehending a suspect from a combination of anthropometric and strength data. At least initially, such statistical manipulations should be gender specific, to test whether there are gender differences in the approach to essential tasks.

The tests used in making such prediction equations are generally well known. Protocols have been standardized, and age and gender specific norms are available. Information is also available allowing scores to be related to health (Tefft, 1999) and training responses (Bonneau and Brown, 1995; Byrd, 1976; Lindell, 1975; Métivier et al., 1982). Unfortunately, scores have a low face validity (that is, the potential connection to job performance is not obvious, at least at first inspection). The standard error of measurements is also quite large. The actual ability of the tests to predict job performance is poor, in part because the number of test measurements is limited. For example, strength is often assessed by a single static test such as maximal handgrip force, although such values show poor correlations with dynamic strength, or indeed with the maximal static force developed by other muscle groups. Further, there are important gender differences in the distribution of strength between individual muscle groups (Shephard, 2000b). The “physiological” tests which are chosen often fail to evaluate anaerobic performance adequately. Finally, many construct validity test batteries fail to include an assessment of movement quality, thus ignoring elements of coordination, balance and agility which may be important to job performance (Baechle, 1994).

HYBRID TESTS

Hybrid tests combine items designed for both content- and construct-validity tests (Wilmore and Davis, 1979). They add important constructs that cannot be evaluated conveniently by task simulation. For example, Wilmore and Davis (1979) added measures of cardiovascular endurance (a shuttle run), muscular strength, flexibility and body composition to two more direct measurements of policing ability—surmounting a barrier and a simulated arrest.

METHODS OF SCORING

Results are commonly presented as a global score that summarizes performance on all test items. This approach allows both individual candidates and those assessing them to make trade-offs between strengths and weaknesses on individual components of the test battery. The standard demanded of prospective police officers can be based on the performance of a target group (for example, a typical suspect) or on the scores attained by a group of incumbents who appear to be carrying out their jobs effectively.

The latter approach poses several practical difficulties. The new recruit is likely to be at a disadvantage relative to an incumbent, but yet may be capable of acquiring the necessary skills during basic training or a short period on the job. If a specific type of heavy demand is made only rarely, it may also be unclear whether job performance is indeed adequate in those officers who achieve only low test scores. Acceptance of such a standard may place both the public and fellow police officers at unwarranted risk. In principle, each officer should be capable of rescuing a member of the general public and/or a fellow officer under life-threatening
conditions. If entry standards are based on the poorest score attained by an “adequate” incumbent, the question of statistical instability arises. Those who accept this approach thus prefer to base their minimum standard on values that are two or three SDs below the population mean. Such standards exclude, respectively, 16.7% or 2.5% of the labour force who purport to be performing their jobs adequately.

IMPLEMENTATION OF RECRUIT AND INCUMBENT TESTING

A survey of 2914 law enforcement agencies (International Associations of Chiefs of Police, 1998, #2174) found that 81% imposed some type of physical fitness standard at recruitment, but only 16% required fitness testing of existing officers.

Both construct- and content-validity tests are problematic when applied to the individual, because of limited test-retest reliability and low correlations with other measures of job performance and physical fitness (Shephard, 1989). The latter point is well illustrated by an analysis which found no significant correlation between RCMP Physical Activity Requirement Evaluation (PARE) scores and values obtained on the Canadian Standardized Test of Fitness (Gaul and Wenger, 1992). The lack of relationship reflects the substantial elements of skill that are required in the PARE test, and possibly also the limited reliability of the Canadian Standardized Test of fitness.

It is widely agreed that a fair proportion of older serving officers would currently fail minimum entry standards (Bonneau and Brown, 1995; Charles, 1982; Evans, 1980; Health Service Directorate, 1996; Maher, 1984; Métivier et al., 1982; Wilmore and Davis, 1979). Normal police duties seem insufficient to maintain the fitness of serving officers (Bonneau and Brown, 1995; Charles, 1982; Evans, 1980; Health Service Directorate, 1996; Maher, 1984; Métivier et al., 1982; Wilmore and Davis, 1979). Anecdotal comments of serving officers suggest that older incumbents rely on a younger employee to catch the “bad guy”, “I would conduct a chase for a minimal amount of time, and then if not successful I would call for a back up” (Gaul and Wenger, 1992).

Almost all of the arguments for mandatory fitness testing of recruits apply even more strongly to incumbents, as the various aspects of fitness deteriorate due to the normal ageing process (Shephard, 1997a). Regular screening would be likely to encourage the maintenance of personal fitness, reducing absenteeism and premature retirement or death (Wilson and Bracci, 1982), and enhancing job performance.

A fair proportion of police officers agree with the concept of annual testing in one RCMP survey, 65% of all participants, and 53% of women reacted positively to this proposal; 70% of those surveyed also agreed that the 4 min PARE time was a reasonable and achievable goal for serving officers (although among female incumbents only 50% agreed and 17% disagreed with such a requirement; Gaul and Wenger, 1992).

**Bona Fide Occupational Requirements**

**EARLY APPROACHES**

Early Human Rights legislation in Canada and the U.S. provided that everyone had the right to freedom from discrimination in employment on the basis of "gender, age or disability". However, it did not preclude a difference in treatment
if an employer could demonstrate that a difference associated with gender, age or disability made it impossible for a specific category of individual to perform the essential duties of a given job.

Given the inherent discriminatory nature of most fitness tests, the first step is to define *bona fide* occupational requirements, abilities which if lacking would preclude the safe or effective discharge of necessary duties. The essential job demands could be psychological or physiological, the latter including minimum sensory abilities, aerobic power, muscular strength and tolerance of adverse environments.

**PRINCIPLES APPLIED IN OCCUPATIONAL MEDICINE**

The U.S. National Institute of Occupational Safety and Health (U.S. National Institute for Occupational Safety and Health, 1981) set an "action limit" for the special training and selection of workers in physically demanding occupations if (a) the average rate of energy expenditure exceeded 14.6 kJ/min, corresponding to an oxygen consumption of about 0.7 L/min; (b) the compression force on the lower back exceeded 3.4 kN; or (c) fewer than 75% of females and 99% of males could perform the task safely. The maximal permissible limit was defined as a work intensity three times the action limit. Tasks between the action limit and this ultimate ceiling were regarded as undesirably strenuous, requiring either a specific selection and training of personnel, or an ergonomic modification of the task (the latter essentially a method of accommodating employees with a limited functional capacity).

The aerobic component of task demands might be assessed by the wearing of a portable oxygen consumption monitor (Shephard, 1994); alternatively, the overall intensity of heavy tasks as seen in a film of operations could be rated by a competent assessor (Moores, 1970), or the metabolic cost of individual components of a task could be integrated on a minute-by-minute basis, using the standard principles of time-and-motion analysis (Gilbreth and Gilbreth, 1920). Employment at the action limit for an 8-hour shift required personnel with a maximal oxygen consumption of about 1.74 L/min; corresponding figures were 1.48 L/min for a 4-hour stint, and 1.10 L/min for a 2-hour period of work (Shephard, 1990).

A commonly accepted minimum aerobic standard for public safety officers (a maximal oxygen intake of 45 ml/(kg.min), Gledhill and Jamnik, 1992, or in a man 3 L/min, Shephard, 1991) would allow 8 hours of work at approximately twice the action limit. Women have a better tolerance of very prolonged aerobic effort than men (Speechly et al., 1996). However, such figures have little relevance to typical "critical incidents" in the police force, where the task continues for no more than 1–5 minutes (Anderson and Plecas, 1999; Fahrenholtz et al., 1989).

Nottrot and Celentano (1984) found that in military tasks that required heavy lifting, the most useful test in terms of both content validity (the correspondence between the laboratory test and the required task) and construct validity (here defined as test scores which ranked workers in accord with previously formulated hypotheses about the optimum type of employee) were (a) the maximum mass that could be lifted from the ground to a height of 1.83 m, as determined by an incremental lifting task; and (b) the maximum force developed when pulling against a chain dynamometer at a height of 1.1 m.

On both tests, the ability to develop a force of about 360 N provided optimal discrimination between military recruits with inadequate and adequate strength.
However, the approach of Nottrodt and Celentano placed females at a disadvantage, because it required a relatively high lift and imposed a fixed lifting technique (Shephard, 1990). Further, it had greater relevance to military duties such as the loading of aircraft than to the lifting and dragging of burdens encountered in police work.

THE MEIORIN CASE

The principles traditionally applied in recruit testing were influenced substantially by the outcome of the Meiorin case. Ms. Meiorin was a seasonal forest firefighter with a total of three year’s experience. When mandatory fitness testing was first introduced for full-time firefighters (in 1988), she passed three of the required fitness tests but failed a fourth assessment (the ability to run 2.5 km in 11 min). Her time for the run was 11 min 49.4 sec. Male test participants had achieved a pass rate of 65–70%, but only 35% of females had met the 11-min standard (Eid, 2001).

In reviewing her case, the Supreme Court of Canada found several faults on the part of the employer:

1. Occupational fitness standards must accommodate individual and group differences, to the extent that is reasonably possible (Eid, 2001). The required standards of maximal aerobic power had been based primarily on the abilities of several dozen elite, mostly male, firefighters, whereas the employer should have evaluated a wide range of people of both genders in order to determine the minimum fitness necessary to do the job. There was no objective evidence that female firefighters either needed or could achieve the 50 ml/(kg.min) standard demanded of male participants. Indeed, Ms. Meiorin had apparently performed her job satisfactorily for three years, and volunteer firefighters were still not required to pass a fitness test before engagement in front-line firefighting duties.

2. The employer showed a correlation between fitness test scores and job performance, but failed to demonstrate that a minimum aerobic standard was needed in order to discriminate between firefighters who could work safely and efficiently, and those who could not. The studies had failed to distinguish female from male test subjects, so that it remained unclear whether a woman could still perform a firefighter’s duties even if she had a lower maximal aerobic power than a man.

3. No clear evidence was given of the extent of any hardship or danger that would arise if a differing standard of occupational fitness were to be applied to women and to men.

CRITICISMS OF THE MEIORIN DECISION

A recent critique (Gillis, 2001) listed several objections to the Meiorin judgment, all of which appear to have some merit:

1. The hearing had demanded objective proof of various issues from the defendants, but had accepted hearsay evidence with regard to the physical condition of the plaintiff.

2. The plaintiff had worked for several years as a seasonal forest firefighter, but it remained unclear whether she had been exposed to physically demanding emergencies.
3. Ms. Meiorin had skills in first aid, and for this reason she may not have not been employed in a front-line position.

4. Ms. Meiorin was rated as “not the worst” of employees, but also “not the best”. Reasons for this particular classification were not clarified.

5. The tort of “negligent hiring” has particular force in certain occupations (public carriers, police and others concerned with public health, safety and welfare). Insufficient consideration had been given to the legal obligations of the employer with respect to occupational health and safety, workers’ compensation liabilities, and vicarious liability for hiring and retaining unfit employees.

6. No evidence had been sought from the union and from fellow employees on their reactions to safety hazards and/or the need to compensate for Ms. Meiorin’s lack of fitness by working harder. Consideration should have been given to the employees’ rights to know about unsafe working conditions, and to refuse such work. Account should also have been taken of difficulties in negotiating a mandatory minimum fitness standard with police unions (for instance, issues of privacy and of discipline).

7. Insufficient account had been taken of practical problems in devising and validating occupational fitness tests specific to each of many potential minority groups.

8. The initial motivation for the fitness testing of firefighters had been a coroner’s inquest. Details of this incident had not been explored, nor had there been any discussion of the prevalence of deaths and injuries in relation to the fitness of firefighters.

CURRENT APPROACH

Because of the decision in the Meiorin case (above), the Canadian Human Rights Commission has now developed a much more rigorous definition of gender discrimination in Canada:

1. If a mandatory fitness requirement is applied equally to men and women, this is judged as discriminatory because it holds women to a male standard.

2. Employers must now develop hiring standards based on substantive equality, rather than accommodating “differences” after the fact. Such accommodation is rejected, because it allows the underlying discriminatory standards to remain in place and excludes those who are not prepared or able to challenge the standards or to demand accommodation.

3. If a mandatory fitness test is to be applied, it must be justified as a bona fide occupational requirement. Evidence must be provided that inability to meet the required standards of fitness creates insuperable problems when performing essential components of a job— for example, in policing it causes a safety risk to the general public, fellow officers or the individual.

ESTABLISHING A BONA-FIDE OCCUPATIONAL REQUIREMENT

The steps in establishing a bona-fide occupational requirement are

- establish purpose of the proposed standard
- demonstrate that the purpose is rationally connected to essential requirements of the job
demonstrate that the test was not introduced with a discriminatory intent
establish that the standard is not excessive
demonstrate that non-discriminatory alternatives were reviewed with third parties such as unions
demonstrate that accommodation is impossible without undue hardship to the employer.

The process has been discussed at some length in the light of the Meiorin case (Gledhill et al., 2001a, 2001b). Based on the Supreme Court decision of September 1999, three specific steps are now required in order to demonstrate a bona fide occupational requirement:

1. The purpose of the proposed standard must be established and the employer must demonstrate that this purpose is rational and connected to performance of the job. If the purpose is not legitimate, or is not based on the essential needs of the job, then it is not a bona fide occupational requirement.

2. The employer must demonstrate that she or he had an honest belief that the proposed standard was necessary to accomplish the purpose identified in (a), and that the requirement was not introduced with any discriminatory intent.

3. It must be shown that application of the standard is necessary to accomplish the stated purpose, that the standard is not excessive, and that avoiding discrimination against a claimant is impossible without undue hardship. A reviewing tribunal would wish to consider whether non-discriminatory alternatives had been explored by the employer; whether the purpose could be accomplished by establishing different standards for different categories of worker; and whether third parties, such as unions, had assisted the employer in setting an appropriate minimum standard.

**Determination of Occupational Standards for Police Officers**

**EARLY APPROACHES**

As in some other industries (Shephard, 1974), selection of police officers was based initially on height and body mass. It was argued that such standards had greater relevance in public safety work than in other sections of the labour force (Hogan, 1991; Hogan and Quigley, 1986). However, a study of 412 police forces indicated that widely divergent criteria were being applied. Minimum standards for height ranged from 1.68 to 1.83 m, and for body mass from 54.5 to 81.8 kg; likewise, maxima for height ranged from 1.88 to 2.03 m, and for body mass from 79.5 to 122.7 kg (Baker and Danielson, 1974).

Height and weight standards were used in the selection of police officers until around 1970 in the US and 1980 in Canada (Bonneau and Brown, 1995). Given the gender differences in height and body mass noted above, such criteria inevitably had an adverse impact on female recruitment. This approach was abandoned after both North Carolina and Los Angeles police forces were found guilty of gender discrimination because they had imposed excessive standards of height and body mass (Carmean, 1984).

The RCMP demanded fitness in recruiting posters as long ago as 1893, when it asked for “applicants between the ages of 22–40, active able-bodied men of
thoroughly sound constitution” (Bonneau and Brown, 1995). Formal fitness testing was first introduced into policing some 20 years ago, when female recruits were already being accepted into police units, and in Canada task simulation and agility tests subsequently found favour in selection test batteries (Bonneau and Brown, 1995). Given the long history of fitness requirements, the current policy of testing applicants cannot be criticized as being introduced with the deliberate objective of excluding women from police work.

CONSTRUCT VALIDITY TESTING

The construct validity approach has used standard batteries of “physiological” tests, with patterns of assessment following the recommendations of Fleishman (1964) and the International Committee for the Standardization of Fitness Tests (Larson, 1974). In Canada, assessors commonly used the Canadian Standardized Test of Fitness (CSTF) (Fitness Canada, 1986). Problems were encountered when administering the CSTF, because nervousness pushed the resting heart rate of some candidates to a level that did not allow even the first stage of the progressive stepping test to be carried out (J. Bonneau - unpublished data). It was then necessary to refer such individuals for costly laboratory treadmill tests, and many of those who were referred proved to have a relatively high level of aerobic fitness.

Some police forces have required candidates to meet no more than the 25th percentile of norms for serving officers, on the assumption that most incumbents are performing their work satisfactorily. A better option is to expect all applicants to attain a relatively high score (Greenberg and Berger, 1983; Métévier et al., 1982), on the basis that the average suspect is likely to be a young adult male of at least average fitness (Bard et al., 1985), and that some of the duties involved in policing make heavy physical demands. Métévier et al. (1982) proposed setting the entry standard at the 60th percentile of norms for incumbents, apparently ignoring the implication that 60% of serving police officers would then be regarded as having an inadequate level of fitness. A proposed entry criterion at 50% of the mean for serving officers (Wilson and Bracci, 1982) must be criticized on similar grounds. Spitler et al. (1987) suggested that those with average fitness scores for their age could complete all police duties, but given the substantial gender difference in average fitness for adults aged 20-50 years, it seems unlikely that women recruited on such a basis would be able to meet the demands of policing. In contrast, Wilkie (1974) suggested that above-average fitness was needed to undertake the duties encountered during a police patrol. The Police Institute of Québec test (Côté, 1996) has applied differing fitness standards for female and male candidates, for instance a static vs. a dynamic chin-up, as have some American jurisdictions (Palmer, 1984); again, this seems hard to justify if police work requires a finite arm strength and endurance.

The Sureté de Québec test calculates officers’ scores as a weighted STEN value, based on norms for a large sample of male recruits (Côté, 1996). Tests scored in this fashion show a substantial correlation with PARE scores (Osborn, 1976).

Police officers have tended to perpetuate the image that they have a high level of fitness, although such an image does not accord with empirical observations (Babin, 1983; Gendron, 1984). Pollock et al. (1978), Stamford and associates (1978), Burkhart (1980) and Klinzing (1980) all found that on average, the physical condition of incumbent officers was less than that of jail prisoners, students,
or young adults. Wilmore and Davis (1979) tested 140 male and 16 female police officers in the laboratory and a further 300 officers in the field. They found no significant differences of fitness between this sample and the sedentary North American population of comparable age and gender. Physical condition was poorer in female than in male police officers, and in both genders scores declined progressively as they became older. Stamford and associates (1978) observed that the problem was particularly serious in older officers, due to a decrease of strength and cardiorespiratory capacity plus an accumulation of body fat. The lack of fitness in incumbents is an important potential criticism of any attempt to impose minimum fitness standards upon applicants, since it implies that serving officers in general are neither able nor expected to match the levels of fitness demanded of recruits.

Some police forces continue to use construct-validity tests, but others such as the RCMP and the California Highway Patrol now prefer content validity tests. One argument advanced by those opposed to the field tests is that they are too specific, and not all job requirements can be simulated. However, such an objection ignores the muscle group specificity of a treadmill maximal oxygen intake assessment or a determination of peak handgrip force. Specific laboratory measurements of this sort give only a limited representation of the complex tasks that a police officer must undertake (Bard et al., 1985), and they may not discriminate sufficiently between officers with a good and those with a poor performance of their duties (Wilmore and Davis, 1979).

DEVELOPMENT OF CONTENT VALIDITY TESTS

An early decision of the US Supreme Court (Dorothy vs. Rawlinson, 1977: 433 US 321) encouraged the development of content validity tests for the selection of police officers. Such tests are easier to defend than "physiological" tests (Gruber et al., 1983; Wilson and Bracci, 1982), particularly if the items have been selected on the basis of a careful task analysis (Gledhill and Jamnik, 1992) and have subsequently been validated on the job.

Some determinations of task demands have limited the discussion to the police chief and his medical adviser (for instance, Métivier et al., 1982), or only a limited number of physical incidents have been evaluated (for example, Gruber et al., 1983). But most authors have undertaken thorough task analyses (Table 1), including consultation with incumbent officers, ride-along observations, use of recall questionnaires, the measurement of task demands (ratings of perceived exertion, heart rate and/or oxygen consumption monitoring), discussions with focus groups, individual interviews, supervisor and/or peer ratings of performance and critical incident reports (Health Service Directorate, 1996; Shephard, 1974). The statements of police officers must nevertheless be interpreted with caution, since fear of disciplinary sanctions may lead to an under-reporting of physical incidents when dealing with suspects (Gruber et al., 1983).

It is important that the definition of critical tasks be based on the opinions expressed by both women and men, as job priorities may differ between the two sexes (Eid, 2001). For instance, it has been suggested that female officers are more successful than men in using verbal rationalizing to control their suspects (Gaul and Wenger, 1992), although physical confrontations were still noted in a small sample of female RCMP officers (Anderson and Plecas, 1999).
Table 1  Task Analysis for Police Officers

<table>
<thead>
<tr>
<th>Author</th>
<th>Listing in task analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bard et al. (1985)</td>
<td>Walking, climbing, descent, car patrol, high-speed drive, ticketing, office work, carrying, manual work</td>
</tr>
<tr>
<td>Côté (1996) (SQ)</td>
<td>420 m obstacle course (73% of total score), reaction time (11%), vigilance (16%)</td>
</tr>
<tr>
<td>Gruber et al. (Côté, 1996)</td>
<td>Climbing, running, lifting, pulling</td>
</tr>
<tr>
<td>Osborn (1976)</td>
<td>Climbing, running 10-25 m with 0-2 turning, jumping 1-2 m, lifting bodies of 80 kg, balance, carrying 10-35 kg, wrestling, ramping, pulling, kicking, hitting, pushing, holding</td>
</tr>
<tr>
<td>Spitler et al. (1987)</td>
<td>500 m run, 150 m obstacle run, carry casualty, climb in window, climb wall, jump, crawl, balance</td>
</tr>
<tr>
<td>Wilson &amp; Bracci (1982)</td>
<td>Climbing, running, jumping, standing, balance, pushing, pulling, wrestling, ramping, hitting, kicking</td>
</tr>
<tr>
<td>RCMP (Health Service Directorate, 1996)</td>
<td>Running, walking, pushing, pulling, carrying, jumping, vaulting and going around or clearing obstacles</td>
</tr>
</tbody>
</table>

Key task demands of policing. Based on the recommendations of the Office of Human Resources of the Government of Québec (Cimon, 1984; Turcotte, 1979), task analyses by Bard and associates (1985) identified the relative importance of selected characteristics to the performance of police work. In general, the demand upon police officers is for strength, stamina and movement ability (Hoffman, 1990; Hoover, 1992). Such characteristics as speed, strength and flexibility are more important than a high level of aerobic power, both to job performance and the avoidance of injuries (Allen et al., 1984; Barnard and Anthony, 1980; Brownlie et al., 1982; Collingwood et al., 1989; Davis and Dotson, 1987, 1989; Davis and Wilmore, 1979; Farenholtz et al., 1989; Hiley et al., 1990; Mostardi et al., 1990).

Wilkie (1974) listed wrestling as a police officer’s most frequent physical requirement, although he commented that no analysis could identify all of the multitude of duties that an officer might be required to perform. Other investigators, such as Osborn (1976), have ranked climbing somewhat higher than wrestling. Bard and associates (1985) emphasized that some tasks, although encountered relatively rarely, were nevertheless of vital importance to the safety of the public, fellow-officers and the individual. Such items must therefore play an important role in the selection of recruits (Bonneau and Brown, 1995).

There seems general agreement on the “core competencies” required of police officers throughout the developed world (Bonneau and Brown, 1995; Farenholtz and Rhodes, 1986, 1990; Gaul and Wenger, 1992; Greenberg and Berger, 1983;
Osborn, 1976; Superko et al., 1988; Wilmore and Davis, 1979). More than a half of active police duties involve climbing, running, lifting, wrestling, and dragging (Gruber et al., 1983), and most of these tasks are heavily dependent upon muscle strength. Some 57% of suspects push or pull an officer, and the officer pushes or pulls 76% of suspects during an arrest. An arrest often involves twisting movements, but these are avoided in most test batteries because of the dangers of injuring test candidates (Anderson and Plecas, 1999). 80% of physical encounters last for less than one minute (Anderson and Plecas, 1999). Further, task demands seem very similar between men and women officers (Health Service Directorate, 1996). Gruber et al. (1983) have argued that although no test battery can include all potential job demands, a 50% representation of the task provides an adequate basis of job assessment.

Police rarely run an uninterrupted distance of more than 80 m in the course of their duties; even Osborn (1976), who advocated a 366 m run, found that the actual distance covered per burst of running was only 30-80 m. Anderson and Plecas (1999) noted that 5 of 6 tasks identified in their task analysis were regarded by police officers as demanding medium to maximum effort. The heart rates of police officers are rarely greater than 125 beats/min, except when carrying (Bard et al., 1985), and there is little correlation between running speed and success either in making an arrest or in rescue missions (Davis and Wilmore, 1979). Nevertheless, running over short distances was the sixth of the job-related items identified by Booth and Hornick (1984), and a 2.4 km run has been considered “a job content behavior with a high face validity” (Davis and Dotson, 1987). Others have advocated distance runs (Collingwood et al., 1989; Johnston and Hope, 1981) and the total duration of the POPAT test (4 min) is such that it has around a 50% aerobic basis (Fahrenholtz et al., 1989). The run component of the POPAT covers a total of 400 m, with an average performance time of 2 min 13 sec. Fahrenholtz and Rhodes (1986) have concluded that officers should anticipate running a longer distance than this at least once per year. Against this reliance on aerobic testing, many of the apparently aerobic tasks undertaken by a police officer are of relatively short duration, limiting the importance of the individual’s peak aerobic power to overall performance. Whereas a demand that is only 40% of maximal aerobic power can be fatiguing if sustained over an eight-hour shift, as much as 75% of maximal aerobic power can be tolerated for one hour, and even higher fractions of peak oxygen transport can be accepted for periods of a few minutes (Shephard, 1994).

Only two task analyses have identified reaction speed as an important component of police work (Booth and Hornick, 1984; Métilier et al., 1982), although this seems necessary to the physical aspects of policing. Babin (1983) further maintained that lack of fitness and thus fatigue would reduce accuracy in the use of firearms.

Details of some of the field tests proposed for police officers are summarized in Table 2. The ability of field tests to predict the efficacy of future police officers remains to be clearly established (Bard et al., 1985). A multiple correlation analysis on 217 male and 13 female police officers (Wilmore and Davis, 1979) found that policing skills were most closely correlated with performance of a vertical jump. However, only two items in the test battery of Wilmore and Davis (1979) had significant predictive value, and the multiple correlation was around 0.5, suggesting that the test battery predicted no more than 25% of an officer’s overall policing ability.
Table 2  Field Tests Proposed for the Selection of Police Officers

<table>
<thead>
<tr>
<th>Authors</th>
<th>Proposed test battery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bard et al. (1985)</td>
<td>Choice reaction time; obstacle course including staircase ascent, crawl/jump, balance beam and dragging 75 kg sack, stabilometer; mastery of simulated criminal (arm flexion and extension) and fitting of handcuffs.</td>
</tr>
<tr>
<td>Booth &amp; Hornick (Colorado) (1984)</td>
<td>Reaction time/peripheral vision, arm strength, leg strength, grip strength, balance</td>
</tr>
<tr>
<td>Collingwood et al. (1989)</td>
<td>2.4 km run, skinfolds, bench press, sit-ups, agility run</td>
</tr>
<tr>
<td>Côté (Sûreté de Québec) (1996)</td>
<td>420 m obstacle course in segments of 50 m, jumping 1.8 obstacle, climbing and descending stairs, jumping two 0.45 m obstacles, jumping 0.9 m ramp, fall on back or stomach, push/pull and controlled falls, lift and carry 36 kg 15 m. Test repeated twice, first test 20% of score. Result calculated as weighted STEN score.</td>
</tr>
<tr>
<td>Institute of Police, Québec</td>
<td>Cooper test, dynamic chin-up (men) static chin up (women), flexion and extension on parallel bars (men only), sit-ups, vertical jump.</td>
</tr>
<tr>
<td>(Bonneau, 1986; Bonneau and Brown, 1995; Côté, 1996)</td>
<td></td>
</tr>
<tr>
<td>Davis &amp; Dotson (1987)</td>
<td>Lug wrench, drag body 15 m, push car 15 m, foot pursuit (150 m run, 1.8 m wall, simulated arrest)</td>
</tr>
<tr>
<td>Davis &amp; Wilmore (1979)</td>
<td>Get out of car, run 6 m, climb 1.5 m wall, run 30 m, climb 1.8 m fence, simulated arrest, drag 75 kg dummy 46 m</td>
</tr>
<tr>
<td>Gruber et al. (1983)</td>
<td>Obstacle course (0.4 km run, 0.15 m run and 1.8 m fence, 61 m slalom, pull 91 kg dummy 26 m</td>
</tr>
<tr>
<td>Jordan &amp; Schwartz (1986) (NY Police)</td>
<td>Stair climbing and descent, running, pulling and pushing 55 kg box, dragging mannequin between cones, obstacle course with 1.52 m wall.</td>
</tr>
<tr>
<td>Osborn (1976)</td>
<td>Carry 80 kg dummy 0.9-6.4 m, balance beam, jump from 1.2 m, push patrol car with flat tyre 3 m, crawl through openings, 366 m dash</td>
</tr>
<tr>
<td>PARE (Gaul and Wenger, 1992)</td>
<td>Obstacle course (running, jumping over and clearing an obstacle, going up and down stairs, making sharp turns, and vaulting over a 0.91 m high obstacle, with 6</td>
</tr>
</tbody>
</table>
Table 2 (Continued)

<table>
<thead>
<tr>
<th>Authors</th>
<th>Proposed test battery</th>
</tr>
</thead>
<tbody>
<tr>
<td>PARE (Gaul and Wenger, 1992, continued)</td>
<td>Repetitions of the circuit. Alternately push and pull a 32 kg weight through six 180 degree arcs without dropping the weight; between these exercises, four controlled falls. After a 30 sec rest, a 36 kg torso bag must be lifted and carried 15 m, untimed</td>
</tr>
<tr>
<td>Spitler et al. (1987)</td>
<td>Grip strength, push-ups, chin-ups, bench press 1RM, sit-ups, vertical jump, agility run, 46 m dash, flexibility</td>
</tr>
<tr>
<td>Wilkie (1974)</td>
<td>Climb 1.8 m fence, pull 80 kg 15 m, run 0.4 km, carry 82 kg stretcher 30 m</td>
</tr>
<tr>
<td>Wilson &amp; Bracci (1982)</td>
<td>Obstacle course (run, slalom, jump, ramp, balance beam)</td>
</tr>
</tbody>
</table>

The Physical Readiness Evaluation Procedure (PREP) test. In Ontario, the PREP test was advocated for the evaluation of firefighters and other public safety officers (Gledhill and Jammik, 1992). This test includes 4 circuits of a 25 m course, with candidates wearing a 4.1 kg weight belt. Stairs are climbed at each rotation, and a 1.22 m fence is scaled in the second circuit. Following completion of the 4 circuits, two arm restraint simulations are made on a body control push/pull device, and a dummy of 68 kg is dragged a distance of 15 metres. A total of 175 seconds is allowed for completion of the circuit. The required standards were derived from blended scores for women and men (Gledhill and Jammik, 1992). However, standards do not appear to be based on any clear job-related minimum requirement. Rather they were set at a level that >99.9% of men and 90-95% of female recruits could pass. The score included data for a shuttle run (Naughton et al., 1996). The result was expressed as a running pace (Stage 6.5 of the Léger run), but it could equally well have been expressed as a predicted maximal oxygen intake.

The Police Officer Physical Ability (POPAT) test. The POPAT test (Anderson and Plecas, 1999; Farenholtz et al., 1989; Farenholtz and Rhodes, 1986, 1990) was developed from a comprehensive analysis of the duties of 217 police officers (Farenholtz et al., 1989; Farenholtz and Rhodes, 1986, 1990). Scoring was based on the times required by suspects to complete the test items that had been identified.

Farenholtz reasoned that a high percentage (80–95%) of suspects would be males, aged 20–28 years, often abusing drugs and alcohol. The average physical characteristics of 32 male and 12 female suspects were a body mass of 76 kg and a height of 1.78 m, but male values averaged 1.81 m and 82.5 kg. The suspects (who were awaiting trial and did not have access to any special physical education programmes) completed the test circuit in an average of 4 min 12 s, and for ease of
administration the times for police officers were rounded to an entry requirement of 4 min 15s. One potential criticism is that the prisoners were volunteers, and for this reason they may not have been entirely typical of the criminal population.

The POPAT was performed by a sample of 79 male and 29 female police officers; 54/79 (68%) of men and 5/29 (16%) of women met the test requirement (Farenholtz et al., 1989). Scores on the POPAT test were compared with the results of "physiological" tests (vertical jump, pull-ups, push-ups, sit-ups, grip strength, treadmill aerobic and anaerobic tests, Canadian aerobic step test, a Cooper-type 2.4 km run, Wingate tests of anaerobic power and capacity, 50 m dash, sit-and-reach test, and underwater weighing. The 2.4 km run, the treadmill aerobic test and the Wingate anaerobic capacity test all showed correlations of about 0.7 with the POPAT score, indicating (as might be anticipated with a 4 min all-out test) an equal dependence of performance upon the aerobic and anaerobic abilities of applicants. However, scores for the strength tests showed a much weaker correlation (0.4–0.5) with the "fight" component of the POPAT test.

Development of the PARE test. The PARE test is a modified version of the POPAT test, developed by the RCMP because of concerns that an excessive proportion of female recruits were failing the original type of assessment. The vaulting requirement (5 vaults onto the abdomen, five onto the back) was modified on the basis that the fight component of the original test (an average duration of 1 min 52 s) was unrealistically long. The vaults were integrated into the run, and a series of four falls were added to the fight sequence, which now had an average duration of 1 min 10 s (Health Service Directorate, 1996).

The first station in the PARE test was an obstacle course (running; jumping over and clearing an obstacle; going up and down stairs; making sharp turns; jumping clear over a 1.83 m mat; leaping over 2 sticks 3 m apart and raised 0.45 m from the floor; and vaulting over a three-foot 0.91 m high railing; falling on the abdomen or the back; getting up and making 6 repetitions of the circuit). At the second station, the participant alternately pushed and pulled a 32 kg weight through six 180° arcs. Between these exercises, four controlled falls were performed (2 forwards and 2 backwards). The third station, performed after a 30 sec rest, was untimed. Here, a 36 kg torso bag was lifted and carried a distance of 15 metres without dropping it. The total running distance for the test was about 400 m.

The standards required of serving officers were more severe than for applicants. The weight for the push/pull component of the assessment was 4 kg greater, and the weight to be carried was 9 kg heavier.

A comparison of POPAT vs. PARE scores in 52 M showed that after eliminating eight individuals who completed the new circuit much faster, the average times for the modified test were 19.3 sec shorter. Hence, the standard time for the PARE of a graduating police officer was cut from 4 min 15 sec to 4 min. The content of the PARE test conformed with observations of police work, as established by careful work and critical incident surveys on a random sample of 279 serving officers, plus some 120 ride-along observations (Anderson and Piegas, 1999).

Practical experience showed that many women were failing the 6-foot (1.83 m) jump or the 18 inch (0.46 m) clearance of a hockey stick of the PARE test. Those failing a given PARE item were initially required to retake it as the test continued. They then had less energy, and were liable to fail the same item again,
or even to fail the aerobic components of the test. However, the problem of failure of the two jump items was overcome in a revised test protocol, where the long jump distance was shortened from the average suspect jump of 1.83 m to the statistically more appropriate median value of 1.52 m, and realistic time penalties were imposed for failure to complete either type of jump (based on the measured time needed to pass through a ditch, equal to five seconds, and the likely effects of a fall, assessed at 2 seconds; Anderson and Plecas, 1999; Bonneau and Brown, 1995). Further, applicants were allowed to make repeated attempts at the entire test battery. These revisions did much to reduce the risk of discrimination in terms of gender, size and age.

The success rate of candidates was found to rise dramatically with repetition of the PARE test. In men, the pass rate increased from 80% on the first to 95% on the third attempt, and in women the pass rate rose from 39% to 80% (Anderson and Plecas, 1999). Further, the times realized by applicants at their third attempt of the revised PARE did not appear to present an unreasonable barrier, given median times of 3 min 56 sec in men and 4 min 37 sec in women.

OBJECTIVE BASIS FOR MINIMUM REQUIREMENT

The RCMP established the necessity of the minimum requirements on the basis of direct observation and reports from police officers on shift in both urban and rural areas. No substantial differences in perceived duties were seen between female and male officers (Table 3), and as noted below there were occasions when unaccompanied female officers faced heavy physical requirements. Further, similar job requirements have been identified in many other analyses of police work in both Europe and North America (Bonneau and Brown, 1995). All reports have shown that officers must run, walk, push, pull, carry, jump, vault and go around or clear obstacles on a regular basis.

There is good evidence that the energy cost of running and other task demands does not differ appreciably between women and men (Daniels and Daniels, 1992; Shephard, 2000b). Thus the composition of the test (although not necessarily its minimum speed of accomplishment) is well established for both female and police officers.

One possible weakness in application of the standards is that there has been only limited study of the impact of changes in test requirements such as a shortening of the jump distance on the ability of recruited officers to perform essential police duties. A second and more serious weakness in establishing a bona fide case is that the RCMP have not yet applied the standard consistently after recruitment. There is exposure to the test during the biennial periodic health assessment, and there are plans to make retesting mandatory within a few years. But currently, there is considerable police association resistance to the mandatory retesting of older officers (Health Service Directorate, 1996).

Given that aging induces a decrease in physiological capacities as large as the average gender gap (Shephard, 1997a), it could be argued that many older male police officers would fail the PARE test if it were to be applied to them, and that it is just as important to apply minimum fitness standards to older men as to female recruits. On the other hand, the adverse effects of aging upon performance might be offset at least in part by increased experience of task demands, and thus
Table 3  Duties of Female Police Officers in a Rural Area
(Based on Observations of Health Services Directorate, 1996)

<table>
<thead>
<tr>
<th>Type of Activity</th>
<th>Frequency</th>
<th>Mean</th>
<th>Median</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walking</td>
<td>95</td>
<td>85.6 min</td>
<td>75 min</td>
<td>60 min</td>
</tr>
<tr>
<td>Running</td>
<td>45</td>
<td>392 m</td>
<td>411 m</td>
<td>457 m</td>
</tr>
<tr>
<td>Distance running</td>
<td>42</td>
<td>4.56 min</td>
<td>6 min</td>
<td>5 min</td>
</tr>
<tr>
<td>Climbing*</td>
<td>35</td>
<td>2.17 km</td>
<td>1.5 km</td>
<td>1 km</td>
</tr>
<tr>
<td>Jumping</td>
<td>34</td>
<td>1.67 m</td>
<td>1.52 m</td>
<td>1.52 m</td>
</tr>
<tr>
<td>Vaulting</td>
<td>42</td>
<td>1.39 m</td>
<td>1.22 m</td>
<td>0.91 m</td>
</tr>
<tr>
<td>Crawling</td>
<td>5</td>
<td>11.8 m</td>
<td>13.7 m</td>
<td>9.1 m</td>
</tr>
<tr>
<td>Lifting</td>
<td>36</td>
<td>41.4 kg</td>
<td>44.6 kg</td>
<td>45.5 kg</td>
</tr>
<tr>
<td>Carrying</td>
<td>36</td>
<td>26.1 m</td>
<td>13.7 m</td>
<td>22.9 m</td>
</tr>
<tr>
<td>Pulling</td>
<td>32</td>
<td>35.6 kg</td>
<td>25.0 kg</td>
<td>34.1 kg</td>
</tr>
<tr>
<td>Pushing</td>
<td>25</td>
<td>35.7 kg</td>
<td>27.3 kg</td>
<td>36.4 kg</td>
</tr>
</tbody>
</table>

Number of records: 161
Average number of shifts: 16.1
Subjects: 10
Age (years): 32.5
Years of service: 7.6
Height: 1.69 cm
Weight: 63.4 kg

*Members in the rural areas had to climb hills rather than stairs in most cases. When stairs were involved it was on the average 1 flight (i.e., 18 steps). The data was identical for males and females.

an increased mechanical efficiency of movement. A police unit may indeed recognize the limitations of some older officers, and in consequence they may be deliberately excluded from operations that have a high physical demand. In such a scenario, the “special needs” of aging officers would have been accommodated, even if they were unable to pass the recruitment test. Nevertheless, the RCMP does plan to make testing mandatory for all front-line personnel, and there is good reason to believe that graduated implementation of such a scheme is the most effective operational approach (Health Service Directorate, 1996).

There is some evidence that female officers resolve conflicts by non-physical methods more readily than would their male counterparts (Gaul and Wenger, 1992; Health Service Directorate, 1996). However, it is unclear how this type of skill could be incorporated into any practical entry assessment. Further, on a substantial number of occasions women officers have to apply the same physical skills as male officers when completing their assigned tasks. A survey of seven female police officers in Surrey, BC, reported 12 arrests in 114 shifts; in 9 of these incidents,
the male suspect was of average height and body mass, and in 3 cases the suspect resisted arrest, so that the female officer had to use physical force to restrain him (Anderson and Plecas, 1999).

The need for a common standard for all police officers, both male and female, is indicated by the fact that physical confrontations are common when officers are operating alone, with back-up personnel several minutes away. Critical incidents typically occurred without much forewarning (Anderson and Plecas, 1999). Single officer patrols, particularly in rural areas, do not permit the adoption of interchangeable work duties. A single officer may face two opponents in 20% of critical incidents, three opponents in 12%, and four opponents in 6% (Anderson and Plecas, 1999). Alternative approaches to controlling opponents, such as the use of pepper spray, are also facing increasing legal challenge.

The case for use of an entry standard such as the PARE test would be strengthened by showing a relationship between such scores and reports of success in making arrests and completing rescues. As in the case of firefighters, the minimum entry standards for police officers have tended to be set on the basis of the performance of young males, and it could be argued that there is no direct evidence that females who fail to meet these standards could not carry out policing duties safely and effectively. However, in the RCMP test battery the underlying score is based on the speed needed to catch a typical suspect, rather than the performance of the officer or recruit in question (Farenholtz et al., 1989). If a police officer cannot complete the circuit at a speed faster than such an individual, they are unable to perform a critical job-related task. In essence, the standard is based on the target, and for this reason, it is independent of gender.

ISSUES OF CARDIAC HEALTH

Questions have sometimes been raised about the risk of developing a heart attack during active police duty. Support for such concerns was voiced by Pollock et al. (1978), who observed an above average prevalence of cardiac risk factors in middle-aged police officers. Likewise, Johnston and Hope (1981) reported a high level of cardiac risk in 26 of 145 serving police officers. Finally, there have been suggestions that the cardiac mortality was higher in police officers than in many other professions (Shephard, 1991).

If substantiated, such concerns might be an argument in favour of the recruitment of female police officers. However, detailed analysis has shown that work-site heart attacks are sufficiently infrequent that they should not be a significant criterion when considering recruitment or retention of police officers (Shephard, 1991).

SAFETY OF TEST PROCEDURES

Concerns have also been raised regarding the safety of field tests, to the extent that about a quarter of personal physicians initially refused to give permission for serving police officers to take the PARE test (Gaul and Wenger, 1992). Reasons advanced for test exclusion have included high blood pressure, a previous heart attack, an old musculo-skeletal injury which might be exacerbated, and (in some instances) capitulation to demands of officers unwilling to take the test (Gaul and Wenger, 1992). Few problems arose in those serving officers who agreed to take
the test; in 2,100 assessments, there were 5 episodes of nausea and vomiting, and 6–8 minor ankle, knee and back injuries (Gaul and Wenger, 1992).

**Damage Arising From Failure to Meet Standards**

Human Rights legislation requires employers to consider methods of accommodating unsuccessful applicants, and to detail the damage which may result if minimum standards of employment are not met. In the case of policing, acceptance of candidates with a low fitness score will lead to failure in a number of critical incidents, and other costs may be incurred as detailed below.

**CRITICAL INCIDENTS**

Application of the four minute standard to RCMP graduates allows 50% of the incidents connected with the mastery of a suspect to be completed successfully (Farenholtz et al., 1989; Farenholtz and Rhodes, 1986, 1990). In general, additional training brings recruits to a yet higher standard, so that they will be able to meet many of the remaining challenges effectively. If the minimum standard were to be accommodated to 4 min 19 s, 68% of tasks would remain incomplete or aborted, and at 4 min 45 s the failure rate would rise to an alarming 95%.

By accepting a 4 min 45 s entry standard from applicants, and relying upon subsequent training to make good the 45 sec deficit, the Canadian RCMP has already introduced a major element of accommodation into its recruitment policy.

**OTHER COSTS**

Other costs of a reduced entry standard would arise from a greater susceptibility to short-term injuries and long-term disability, poorer productivity, greater absenteeism and a more rapid employee turnover (Brownlie et al., 1982; Greenberg and Berger, 1983; Shephard, 1986; Wilmore and Davis, 1979). Further, if a police officer proved unable to complete a task that might reasonably be expected of him or her, the employer might be accused of a lack of due diligence in recruitment policies, with the risk of claims for damages because of negligent hiring or vicarious liability.

**POTENTIAL TO ACCOMMODATE FEMALE RECRUITS**

One author has recently suggested that in order to claim “undue hardship”, it must be “impossible” for an employer to accommodate a plaintiff group (Eid, 2001). How far can female applicants be accommodated by changing the nature of entry requirements?

Because of limitations inherent in both “physiological” and field tests, some have argued that the only fair approach to police recruitment may be a period of probationary employment rather than the setting of minimum entry standards. The hope is that with such a form of accommodation, unsuitable candidates would drop out during basic training. Unfortunately, military experience suggests that this does not occur in practice (Nottrodt and Celentano, 1984). When a 25 kg lifting criterion was applied to military recruits, almost all men could meet the standard initially and few women could. Nevertheless, the respective drop-out rates
during 10 weeks of basic training were 34% for men and 29% for women. The monitoring of on-job performance for a longer period is also unsatisfactory, since it is difficult to document the need to dismiss someone once they have been recruited.

Most unsuccessful female RCMP recruits failed either the six foot (1.83 m), now revised downward to five foot (1.52 m), long-jump, or the 18 inch (0.46 m) clearance of a hockey stick. Efforts at accommodation of PARE test requirements have taken the form of (a) modifying the test design (as noted above), (b) allowing applicants to make repeated attempts at the entire test battery; (c) providing expert advice on techniques of successful test completion including opportunity to watch a video, provision of a training manual entitled the RCMP Circuit Training Program, and specific advice on how to overcome problems of short stature and limited upper body strength; (d) allowing opportunity to reach the required graduation standard (a 4 min circuit time) during basic training; and (e) basing judgments on the cumulative score, thus allowing the applicant or graduating officer differing options for completing the assigned tasks (Health Service Directorate, 1996).

Another potential method of accommodating candidates who cannot meet the required standard of fitness would be to find a posting that did not include heavy physical demands. This is difficult in police work. At times of high alert, as many as a third of all police officers are posted to front-line duty, and in normal rural operations, there is often only a single person in a police cruiser, with a long distance separating the officer from any potential reinforcements (J. Bonneau, personal observations).

Any potential for accommodation should consider not only the financial cost, but also the disruption of collective agreements, issues of morale, interchangeability of employees, and the magnitude of the risk incurred by employing unfit individuals.

Gender and Training Responses

RESPONSES IN WOMEN AND IN MEN

Although every citizen should have the right to apply for any type of employment, in the case of physically demanding work such as policing, this right must be earned by developing an appropriate standard of personal physical fitness.

Rigorous training can make good much of the difference in strength and aerobic power between women and men. In terms of strength, the two sexes show a similar neural response to training, but there are theoretical reasons why women may show a lesser hypertrophic response than men to an equivalent absolute volume of training (particularly low concentrations of androgenic hormones, Häkkinen et al., 1988; O’Toole et al., 1989), lack of an exercise-induced increase in testosterone (Fahey et al., 1976; Kraemer et al., 1991; Weiss et al., 1983) and a tendency to induce equal hypertrophy of Type I and Type II muscle fibres rather than a selective increase of Type II fibres (Alway et al., 1989; Bell and Jacobs, 1990). Nevertheless, a suitable training programme can protect women against the muscle wasting associated with both aging (Nuutila et al., 1995) and trauma (Yang and Birkhahn, 1993), apparently as well as men. In practice, the training-induced increase in muscle bulk may not be as large in women as in men, but women show at least 20-30% gains in muscle force when they follow an appropriate training regimen, and in some instances the observed increases in strength are larger than
those seen in men (Cureton et al., 1988; Tesch, 1992; Weltman et al., 1978; Wilmore, 1974).

In terms of the trainability of maximal aerobic power, one study found that women gained an average of 18.7%, with a range from 7 to 37% (Knowlton et al., 1978); the response was <13% in those with an initial maximal aerobic power > 42 ml/(kg.min), but greater in those with a poor initial physical condition. A second study (in West Point recruits, who were all relatively fit initially) found average gains of 7.9% in women and 2% in men (Daniels et al., 1979). The percentage increment in maximal aerobic power certainly depends on the initial fitness of the individual (Shephard, 1968), and in such units the response of women at least matches that of men (Eddy et al., 1977). Some authors have even argued for an equal absolute response in women and men (Plowman and Smith, 1997; Wells, 1991). Endurance training can increase the size of a woman’s left ventricular cavity much as in a man, and women can also develop a substantial increase in left ventricular mass (Douglas et al., 1988; Douglas and O’Toole, 1992; Pelllicia et al., 1996; Pollak et al., 1987) similar to that of their male counterparts (Pate et al., 1987).

Women can thus reduce the gender difference in strength and aerobic power by following an appropriate training programme (National Strength and Conditioning Association, 1989), and indeed some well-trained women are stronger than most men (Baechle, 1994; Shephard, 2000b). Other factors can further enhance performance. Scores on the type of field test used in recruitment evaluations have a large skill and agility component (Gaul and Wenger, 1992), and results therefore improve substantially with a preliminary period of practice. For example, results for the PARE test improved 5–12% over 2–3 sessions (Anderson and Plecas, 1999). Reduction of obesity, and cessation of smoking in smokers could augment performance further, each boosting maximal aerobic power by about 10%. Fat loss may be achieved less readily in women than in men (Murray et al., 1986), but some candidates have scope to enhance scores by stopping smoking, given that 25% of police officers currently are smokers (Gaul and Wenger, 1992). Finally, those who achieve close to the minimum entry standard will be the least fit recruits, and in them, the training response is likely to exceed 13% (Knowlton et al., 1978), since the training response depends heavily upon the individual’s initial fitness (Shephard, 1968).

EFFECTIVENESS OF TRAINING PROGRAMMES

Unfit recruits are likely to show a substantial improvement in performance over the course of basic training. Some police forces have introduced rigorous training programmes for their officers (Lilley and Greenberg, 1984), but few data have been collected concerning their effectiveness.

Stamford et al. (1978) suggested that the physical condition of police officers was enhanced by a specific training programme, but gains were quickly lost once the programme ceased. This indicated that the normal police routine was insufficient to maintain physical condition. Pohndorf and Cathey (1975) noted that 14 weeks of basic FBI training (which at that time emphasized combat and use of a baton) did little to enhance a recruit’s deficits of cardiorespiratory fitness, grip strength and balance. On average, fitness scores (which were expressed in STEN units) increased by no more than 3 units.
The RCMP cadet training programme places a heavy emphasis on personal responsibility, with a requirement that recruits demonstrate a commitment to fitness and a healthy lifestyle (Royal Canadian Mounted Police, 1999). The formal training programme emphasizes the development of aerobic fitness, muscular strength and proficiency in self-defence, and recruits also have access to a state-of-the-art fitness centre (well furnished with treadmills, stepping devices and weight-training equipment), a swimming pool, track, and hockey, football and softball teams. There is a requirement that both male and female recruits boost their time for the PARE test circuit from the entry standard of 4 min 45 s to a graduation standard of 4 min, a 20% increase in speed.

Women who pass the recruitment standard of 4 min, 45 sec generally seem able to meet the graduation standard of 4 min. In 1996/97, 209 women were admitted with an average initial time of 4 min 10 sec, and this improved over the 22 weeks of training to 3 min 36 sec, a gain of 13.6%. In 1997/98, a further 106 women were admitted with an initial average time of 4 min 24 sec. Interestingly, the gain in performance was less for men (from 3.33 to 3.10, 10.8%) than for women (Anderson and Plecas, 1999).

Some forces lack access to programmes or facilities that could maintain and/or or enhance their physical condition. Here, the solution may be mandatory fitness testing, with monetary rewards and sanctions based on the observed fitness scores (Serra, 1984; Witzak, 1984) as a stimulus to personal conditioning.

Conclusions

There seems good evidence that women who are recruited as police officers will face unassisted physical encounters with suspects, and that in order to apprehend such individuals they will need to at least match the speed of a young male completing a circuit of the type exemplified by the RCMP PARE test. Accommodation of this bona-fide occupational requirement should be limited to the provision of opportunities to develop an appropriate standard of physical condition through instruction, practice and training. Given such opportunities, well-motivated women should be capable of meeting the current Canadian entry requirements for police officers. The challenge now is to bring older incumbents, both men and women, to an equivalent standard.

References


*Received May 9, 2001; accepted in final form August 21, 2001*