CORRECTING FOR NON-ADHERENCE IN A RANDOMIZED STUDY OF HIP
PROTECTORS TO PREVENT FRACTURES

Karminder Singh Gill, MSPH

A dissertation submitted to the faculty of the University of North Carolina at Chapel
Hill, in partial fulfillment of the requirements for the degree of Doctor of Philosophy in
the Department of Epidemiology of the School of Public Health.

Chapel Hill
2010

Approved by:

Stephen R. Cole, PhD

Charles L. Poole, Sc.D

Sheryl Zimmerman, PhD

Christianna Williams, PhD

Stephen W. Marshall, PhD

ABSTRACT

KARMINDER GILL: Correcting for non-adherence in a randomized study of hip
protectors to prevent fractures
(Under the direction of Stephen Cole)

Between October 2002 and October 2004 the Hip Impact Protection Project (HIP PRO) cluster randomized 1042 nursing home residents to wear hip protectors on either the left or right hip; residents were followed for 676 person-years of observation. The intent-to-treat (ITT) incidence rate ratio, comparing protected to unprotected hips, was 1.23 (95% confidence limit (CL): 0.65, 2.34); overall adherence was 74%. When non-adherence is substantial an ITT analysis estimates the effectiveness of treatment in a mixed population comprised of both compliers and non-compliers and, therefore, under-estimates the etiologic effect of treatment to the extent that the study population is comprised of non-compliers.

Because of the problems inherent to ITT analyses of studies with non-trivial amounts of non-adherence, there have been calls to supplement the ITT effect estimate with adherence corrected effect estimates. Three relatively new methods in the epidemiology literature correct for non-adherence in randomized studies, and provide unbiased effect estimates: marginal structural models (using inverse probability-of-censoring weights (IPCWs)), structural nested models, and instrumental variable analysis.

We employed IPCWs to correct for non-adherence in the HIP PRO study under an assumption that we had measured and correctly modeled all important joint determinants of adherence and hip fracture, and obtained a hazard ratio of 0.55 (95% CL: 0.13, 2.40). Under a structural nested modeling approach, we employed a rank-preserving structural failure time model to identify the survival difference that would have been observed had all participants adhered to their assigned treatment. The factor by which time to a hip fracture was expanded under continuous exposure to hip protectors was 2.41 (95% CL: 0.31, 18.7), assuming a Weibull distribution for time to hip fracture. The estimated hazard ratio under constant exposure was 0.46 (95% CL: 0.07, 2.84).

Using data from the HIP PRO study, we found apparent differences in results between the ITT analysis and analyses correcting for non-adherence. We do not take the adherence corrected results as a complete reversal of the prior analysis; rather, we see these results as supplementing the ITT analysis.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

xi

CHAPTER 1

CRITICAL REVIEW OF THE LITERATURE

GENERAL BACKGROUND

A hip fracture is any break in the proximal portion of the femur, including the head of femur or neck of femur (cervical fractures), and the trochanteric region of the femur (trochanteric fractures) (see Figure 1.1). A third type of hip fracture, the sub-trochanteric fracture, is located anatomically distal to the lesser trochanter of the femur, and occurs much less often than the other two types.[1] Such fractures often result in a disruption of the blood flow to the femoral head, and cause debilitating pain and nerve impairment. The sciatic nerve, one of the major nerves of the body, innervating the muscles of the leg and foot, is often impacted.

neck of femur
pelvis
cartilage
greater trochanter
head of femur
pubic ramus
THE HIP JOINT
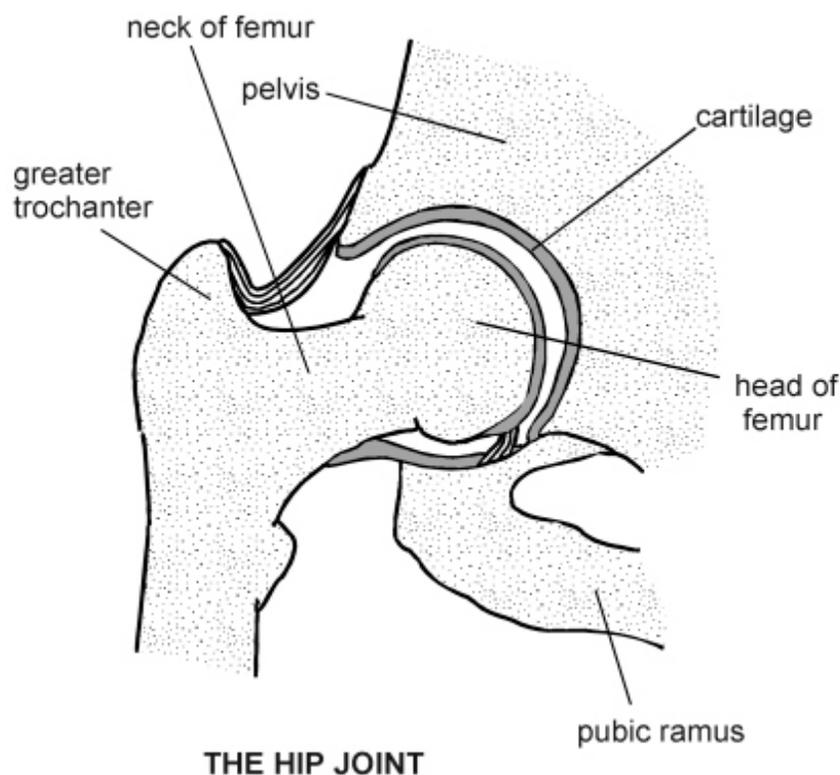
Figure 1.1. The Hip Joint[2]

The number of annual hip fractures has increased substantially over the past 25 years, owing mostly to the aging of the U.S. population.[3] These hip fractures result in a substantial cost to the individual whose hip is fractured, as well as to society.[4] The long-term consequences, in terms of death and disability, are substantial.[5-8] Consequently, much effort has been devoted to interventions that might potentially reduce the incidence of hip fractures. Exercise programs to improve balance and strength[9-11], medications, hormonal treatment, and vitamin supplementation programs to reduce the resorption of bone[4, 7, 12-14], environmental modification interventions, or some combination of these approaches[5, 11, 15] have all

noted significant but small reductions in the incidence of hip fractures. The relatively small reductions in the incidence of hip fractures resulting from these interventions prompted investigators to examine the potential benefit of hip protectors. Hip protectors are designed to prevent hip fractures by either absorbing the energy from a fall directly on the hip, or dispersing the energy of the fall to the soft tissue surrounding the hip. In this dissertation we will be particularly interested in evaluating the effect of the hip protector which was used in the Hip Impact Protection Project (HIP PRO).[16] Before embarking on an evaluation of the effectiveness of this hip protector, however, we review in greater detail the epidemiology of hip fractures, the risk factors for hip fractures, and a systematic review of previous randomized studies of hip protectors.

EPIDEMIOLOGY OF HIP FRACTURES

From 1993 to 2003, the number of hip fracture hospitalizations increased 19% in the United States, from 261,000 to 309,500,[3] and projections by researchers suggest that, with the aging of the population, the total number of hip fractures could increase to 512,000 by 2040.[4] However, because the size of the elderly population has been growing faster than predicted, and because hip fracture rates rise dramatically with age, the number of elderly experiencing a hip fracture in the United Sates could total 840,000 by 2040.[17] Worldwide projections paint an even bleaker picture, with the total number of hip fractures forecast to increase from 1.26 million in 1990 to 4.5 million by 2050, assuming no increases in age-specific hip fracture

rates.[18] If the increase in age-specific hip fracture rates seen over the past two decades is taken into account, the worldwide annual number of hip fractures could rise to 6.26 million by 2050.[19] The number of hip fractures is expected to increase most dramatically in Asia, where aging populations are expected to contribute more than two million hip fractures annually by 2050.[18]

The cost of hip fractures to society is substantial, with hip fractures accounting for more than 20% of the usage of orthopedic beds in several countries.[19] In the United States, hip fractures account for more than half of all osteoporosis-related hospital admissions among women 45 years old and over.[20] The direct costs for inpatient and outpatient medical services and nursing home care are expected to increase the total annual cost of hip fractures in the United States (in 1984 dollars) from approximately $7.2 billion to $16 billion by 2040. If a 3% or 5% rate of inflation is taken into account, the total annual costs of hip fractures could reach $82 billion or $240 billion, respectively.[4] The loss in quality of life for those impacted by hip fractures adds another important dimension to the costs of hip fractures.

The consequences of a hip fracture can be severe. Studies from across the globe reveal surprisingly consistent outcomes for individuals suffering hip fractures. One UK study found that 15% of patients presenting with a fractured neck of the femur die in the hospital and 33% are dead within one year.[21] In the United States, one year survival after hip fracture has ranged from 12-25% in the general population.[6] In nursing homes, where the incidence of first hip fracture can be as high as 44 (95% confidence interval: 34-54) per 1,000 person years, mortality rates

are even higher, at 39% for one year.[7] Age is an important predictor of death. Individuals 80 years and older have a mortality rate that is eight times greater than persons 60 and under.[22] Most of the increase in mortality attributed to hip fracture occurs within the first six months after hip fracture.[8] While overall life expectancy is reduced by 6 years in both men and women,[23] men generally experience markedly higher mortality from hip fractures when compared to women.[24]

Hip fractures result not only in high mortality rates, but, for those who survive a hip fracture, can also have severe long-term consequences. Many individuals never regain their pre-fracture level of physical functioning: most have significant reduction in their ability to function in daily life and in their ability to walk.[25] Mossey et al noted that while 80% of elderly individuals could walk independently before their hip fracture, only 21% could walk independently after the hip fracture.[26] Fear of falling, loss of confidence, and functional deterioration all contributed to the reduction in mobility. Up to 66% will ultimately have impaired mobility and 27-50% will have continued post-operative pain.[27-28] Long term follow-up results suggest that almost half of all patients with hip fracture will have become more dependent and 20-25% will have been referred to a nursing home.[29] Individuals with cognitive impairment, in particular, find it difficult to recover from hip fracture surgery.[30]

RISK FACTORS FOR HIP FRACTURES

A hip fracture usually occurs as a result of the interaction of two unique processes: a fall event resulting in direct impact to the proximal portion of the femur

and an underlying weakness in the bone caused by osteoporosis.[6] Each of these processes has underlying risk factors that have been reasonably well studied. For example, demographics, anthropometric factors, and co-morbidities have been shown to impact bone health. With falls, much of the literature has focused on the first phase of the fall (i.e., where an individual's normal protective responses fail). Environmental factors, such as loose carpeting and icy sidewalks, as well as individual co-morbidities, such as postural hypertension, have been shown to contribute to the first phase of a fall. However, a fall can be conceptualized as occurring in four stages: (1) an instability phase that results in a loss of balance due to host and environmental factors; (2) a descent phase; (3) an impact phase (i.e., the impact of the fall occurs near the hip, the soft tissue around the hip demonstrates poor energy absorption/distribution, and reduced bone strength fails to withstand the energy transmitted to the proximal femur), and (4) a post-impact phase during which the subject comes to rest.[10, 31-33] More recent interest in the later three stages of a fall have shown that the direction of the fall, the subcutaneous fat covering the greater trochanter, and other factors play an important role in the incidence of hip fractures. In short, there are a number of host and environmental risk factors that contribute to a fall, and, ultimately, result in a hip fracture.

Demographics

The incidence of hip fractures rises exponentially with increasing age.[4, 25, 34] The increase in age-specific incidence rates is often explained by muscular

6

weakness, gait and balance disorders, functional impairment, cognitive impairment, and the side effects of drugs.[6] In addition, secular increases in height (and perhaps hip axis length), and changes in environmental factors such as the hardness of surfaces on which falls occur, have been proposed as reasons for the increase in age-specific incidence rates.[34] While the increase in age-specific incidence rates occurs in both men and women, the incidence of hip fractures is approximately two times higher in women, when compared to men.[25, 34] A study by Jacobsen et al in the U.S. showed age-adjusted incidence rates in white males and females as 4.3 and 8.1 per 1,000 per year, respectively.[24] A woman 60 years of age with a median life expectancy of 81 years has an estimated residual lifetime risk of hip fracture of 14%, while a man 60 years of age with a median life expectancy of 77 years has an estimated fracture risk of 6%.[35] This relationship changes only in the very old, when men and women have nearly the same risk of hip fracture.[35] In both men and women, the age-adjusted incidence of hip fracture is lower in nonwhites than whites, and among nonwhites, the incidence is lower in Asian races than in black races.[36] The lifetime risk of hip fracture is about 6% in black women, much lower than the risk in white women, but similar to white men.[37] Higher bone mass and increased obesity may contribute to lower incidence rates in blacks and Hispanics, and shorter hip axis length in Asians, black, and Hispanics may also play a role in the lower incidence of hip fractures in these populations. [24-25]

Anthropometric Factors

Using almost any measure of body size, including weight, fat mass, percent body fat, and body mass index (BMI), the risk of hip fracture in those with a smaller body size is higher than in those with a larger body size. For example, Ensrud et al[38] found that older women with smaller body size had an approximately two-fold increased risk of hip fracture when compared with those with larger body size, whereas risks of hip fracture in women with average and larger body sizes were not significantly different from each other. The incidence rate of hip fracture was 9.35 per 1,000 years in women in the lowest quartile of total weight, compared with 4.63 per 1,000 years in the highest quartile of body weight. Another study showed that a 20% gain in weight between age 25 and old age was associated with a 40% reduction in risk of hip fracture.[39]

There are several explanations for why a larger body size might be protective against hip fractures. In women, the conversion of androstenedione to the more active estone in adipose tissue is the major source of estrogen (a hormone that is protective against hip fracture) in post-menopausal women. Most evidence suggests that greater estrogen acts primarily to reduce bone resorption in post-menopausal women.[40] In men and women, weight increases the transmitted gravitational forces on bone, stimulating bone re-modeling, and, ultimately, resulting in higher bone mineral density. Also, low body weight can be a marker of poor health, itself a risk factor for falls and fractures. Finally, increased weight leads to more soft tissue overlying the greater trochanter, which reduces the forces applied to

the proximal femur in a fall.[25, 38, 41] Lauritzen et al found a difference between cases and controls in the thickness of the subcutaneous tissue covering the hip: women with hip fractures had an average 22 mm thick soft tissue cover of the hip compared with 32 mm in healthy women.[29, 42] Bolstering the importance of the protective effect of soft tissue covering the hip are biomechanical experiments, which show that a porcine soft tissue layer of 29 mm could absorb 60% more energy than a 20 mm thick layer of the same material.

Height, independent of weight, has been demonstrated as an increased risk for hip fracture.[25, 43] Height is strongly correlated with hip axis length, which is associated with an increased risk of hip fracture. After adjustment for age, each standard deviation increase in hip axis length appears to double the risk of hip fracture.[44]

Co-morbidities

Studies have consistently demonstrated that impairment of gait and balance, and lower and upper body strength are associated with an increased risk of falls and hip fractures.[45-47] Muscle weakness, lower limb dysfunction, skeletal impairments, difficulty or dependence in activities of daily living, and use of walking aids (a marker of neuromuscular impairment) have been associated with increased hip fracture risk.[48-50] Individuals who have difficulty with gait and balance also tend to have lower reaction times, are less likely to break a fall with an outstretched arm, and are more likely to fall in a position that results in impact on the hip.[43] Especially among the

9

elderly, who have stiffer, less coordinated, and more dangerous gait than younger people, and who have poorer posture control, body-orienting reflexes, and muscle strength, the ability to avoid a fall is reduced.[43, 51] Among elderly individuals with Parkinson's disease, who have particular difficulty with gait and balance, the reported relative risk of hip fracture can be as high as 10 or more, when compared to individuals who do not have Parkinson's disease.[50, 52]

Individuals with cognitive impairment are at higher risk of hip fracture than individuals with no dementia.[49] In general, patients with Alzheimer's disease demonstrate a greater risk of falling and higher rate of serious injury than those without Alzheimer's disease. Buchner and Larson[53] noted that the risk of hip fractures among people with Alzheimer's disease was three times higher than in those without hip fractures. In another 4-year study, 36% of 44 subjects diagnosed with Alzheimer's disease fell and sustained five hip fractures; non-demented controls had four falls and only one hip fracture.[54] Individuals with dementia have a harder time recognizing environmental hazards and have poorer neuromuscular control than individuals without dementia.[25]

The presence of osteoporosis at the hip is defined as a bone mineral density measurement that is 2.5 standard deviations or more below the young healthy population. A substantial proportion of the elderly population has some degree of osteoporosis and many fall into the high-risk range. About 25% of white women in the United States who are age 65 or older have hip bone mass that is more than 2.5 standard deviations below the mean value of young normal women, and they have an approximately five-fold greater risk of hip fracture than women whose bone mass

is above that level.[55] Over a lifetime, bone density of the femoral neck declines an estimated 58% and 39% in women and men, respectively.[56] For the 50-year old woman with osteoporosis, the lifetime risk of hip fracture is greater than 60%.[34] Osteoporosis does contribute to hip fractures. Data suggest that 70% of hip fractures are at least partially attributable to osteoporosis.[57] However, there is not a great difference in bone mineral density between individuals experiencing hip fractures and age-matched controls; indeed, there is a substantial overlap in bone density values between these two groups. Most studies suggest that elderly men and women have lost sufficient bone for the hip to fracture in case of impact during an unprotected fall.[29]

Those elderly individuals who are malnourished or undernourished are at particular risk of osteoporosis and hip fracture. Deficiencies in micro-nutrients, such as calcium, vitamin D, and vitamin K, have been strongly implicated in the pathogenesis of hip fracture in the osteoporotic elderly, because these micronutrients play a major role in the control of bone remodeling and bone integrity.[58-60] In addition, deficiency in macronutrients such as proteins has been associated with lower bone mineral density (BMD) of the femoral neck.[61] Poor nutrition contributes to an increased risk of hip fracture not only through its impact on the bone, but also due to the fact that it increases the propensity to fall by impairing movement coordination and reducing muscle strength, and reduces the protective layer of soft tissue padding over the greater trochanter.[12]

A history of prior fragility fractures, particularly prior hip fractures, is associated with an increased risk of subsequent hip fractures. This increased risk

may be due to risks that are similar to the prior fractures (e.g., osteoporosis, family history), or may be due to the fact that individuals with prior fractures may be more vulnerable or accident-prone for some time after the index fracture. The later hypothesis is supported by the fact that, after a prior fracture, the greatest risk of hip fracture is in the one year time period immediately after the index fracture.[29] Women have an increased risk of hip fracture after either a fracture of the knee or a fracture of the ankle[62-63], and women who reported that their mothers had had a hip fracture had twice the risk of hip fracture of women without this family history.[39] For all individuals, previous fractures of the ribs and of the lumbar spine result in a 3-4 times increased risk of subsequent hip fractures.[56] People who sustain one hip fracture are 2-10 times more likely to fracture their second hip, when compared to controls.

Environmental Factors

Three fourths of femoral neck fractures occur in the home and approximately two thirds of these are due to a fall precipitated by an environmental hazard**.** Factors associated with falls at home include unstable furniture and appliances, creaky stairs with poor rails, change in surface level, throw rugs and frayed carpets, poor lighting, low beds and toilets, pets, and objects on the floor.[51, 64] There appears to be no seasonal variation in hip fractures, with most hip fractures occurring indoors.

The Institutional Setting

The risk of hip fractures in elderly residents who reside in institutional settings are substantially greater than those living in private homes. Butler et al[65] found that the risk of hip fracture (after adjustment for age and sex) was 10.5 times greater for those living in institutions. The incidence rate of hip fracture among the institutionalized population is high, such that one in every 25 individuals who live in an institution is likely to sustain a hip fracture annually. The difference in risks is greater for the 'younger old,' those aged 60 to 69, with differences in risk declining with age.[65] Another study by Baudoin[66] et al noted that, for patients living in community homes and who were aged 60-69, the risk of hip fracture was 15 times greater than the risk of individuals living at home. Norton et al[67] noted somewhat lower risks (age- and gender-adjusted OR=3.8 (3.0,4.8)) of hip fracture in individuals living in institutions, when compared to individuals living in private homes.

Falls

Falls have been identified as the proximal cause of a hip fracture in the elderly in 90% of cases. The elderly are especially prone to falls, with the major causes of falls in the elderly being: accidents, syncope (i.e., sudden loss of consciousness), drop attacks (i.e., sudden falls without loss of consciousness), orthostatic hypertension, dizziness/vertigo, neurology dysfunction, vision impairment, and drugs.[51, 64] Underlying disease symptoms, such as cardiac dysrhythmia,

orthostatic hypertension, and cerebrovascular disease all contribute to an increased

risk of falls in the elderly. Many medications to treat diseases in the elderly have

been associated with an increased risk of falls and subsequent hip fracture: anti-

psychotics (RR=2.0), long-acting hypnotics (RR=1.8), tricyclic anti-depressants

(RR=1.9), and opiod analgesics (codeine and/or propoxyphene) (RR=1.6) were

associated with an increased risk of hip fractures.[68-70] Over-medication, in particular,

has been associated with an increased risk of falls and hip fracture.

A prior fall, in the 9-12 months before the hip fracture, doubles the risk of hip

fracture. Fall incidence in the elderly increases with age, with 25% of elderly in their

sixties falling at least once annually, and 55% of elderly in their nineties falling at

least once.[71] Among those living in the community, 33% of individuals aged 65 and

older experience one or more falls during the year.[72] In institutional settings, with

populations that are in poorer health, the annual incidence of falls is about

1,500/1,000 resident-years. The rate of falling among those living in institutions

(excluding acute hospitals) has been estimated at 50%, with 10-25% suffering

severe consequences.[73]

Early research focused almost entirely on the instability phase of the fall, that

interaction of host and environmental factors that result in a loss of balance. More

recent research has started to focus on the orientation of the fall and the point of

impact, which also play an important role in the etiology of hip fractures. Falling

sideways is associated with an increased risk of hip fractures, while hitting the knee,

hitting the hand, and falling backward are inversely associated with the risk of hip

fracture, and falling backward is not associated with the occurrence of a hip

fractures.[43] Greenspan et al[11], in a study of community-dwelling elderly noted that sideways falls resulted in an increased risk of hip fracture when compared to a fall in any other direction (OR=5.7 (2.3,14.0)). When the point of impact after the fall is directly on the hip, the odds ratio for hip fracture in institutionalized elderly populations has been documented to range from 5.7 to 21.7.[74] Another study of elderly women living in the community who had previously fallen at least once noted a 33-fold increased risk of hip fracture when individuals landed on the hip, side of the leg, or buttocks.[43] Parkkari et al[33], using a fresh subcutaneous hematoma on the greater trochanter of the proximal femur as evidence of falling on the hip, in a case-control study, found a hematoma in 56% of hip fracture patients, but in only 6% of controls. In addition, biomechanical studies indicate that the energy available in a simple fall on to the hip is 15 times greater than the energies required to fracture the elderly hip.[31] One fourth of falls on the proximal femur result in hip fractures, whereas fewer than 2% of all falls cause this injury. These studies suggest that a typical hip fracture is the result of a fall to the side and a subsequent impact on the greater trochanter of the proximal femur, where the energy-absorbing mechanisms of the host (e.g, outstretched hands to break the fall, reduced soft tissues at the site of impact) are inadequate to prevent fracture.[75]

Hip Protectors

Over the past twenty years, many types of hip protectors have become available and are currently marketed around the world. These protectors generally

fall into two distinct categories: those made with an energy-absorbing material and those that use a semi-rigid plastic shield to divert force from the trochanteric region to the soft tissue of the thigh.[76] The hip protector that has been used in the majority of previous studies is the Safehip protector, made of stiff polypropylene on the outside and an inner lining of soft plastozote, and is designed to divert impact energy away from the greater trochanter.[77]

The hip protector in the first Hip Impact Protection Project (HIP PRO) study, which provided data for our primary analyses, was made of a 0.32 cm outer layer of 2.7-k g/m^3 polyethylene vinyl acetate foam backed by a 0.95-cm layer of 0.9-kg/m^3 polyethylene vinyl acetate foam.[16] The protectors, measuring 11.43 cm x 11.51 cm x 1.91 cm, were placed in the undergarments on only one hip and had previously been tested for bio-mechanical efficacy.[78] The undergarments came in different types to accommodate the needs of individuals with dementia and incontinence, and were made of a Lycra/cotton material.

After termination of the first HIP PRO study, a second study of hip protectors was initiated using the FallGard FG-04 hip protector. This hip protector was slightly more flexible than the hip protectors used in the first study and was composed of a dense (10 lbs / ft^3), 0.57 inch layer of polyvinyl chloride (PVN) rubber foam, backed by a softer 4 lbs / ft^3 PVN foam layer 0.18 inches thick. The dimensions of this new hip protector were similar to the hip protectors used in the first study and used the same cotton Lycra undergarment.

Adherence to Wearing Hip Protectors

Older individuals are often unwilling to wear hip protectors, such that even

initial acceptance of hip protectors is low. The most important reasons for initial

refusal include problems of fitting and discomfort, the perceived extra effort needed

to wear the protectors (including difficulty wearing them when going to the

bathroom), appearance, and the personal belief that they are not at risk for hip

fractures.[73, 79-81] The various studies that have measured initial acceptance of hip

protectors (or initiation of use) have found that primary acceptance was low to

moderate, ranging from 35% to 72%,[41, 73, 77, 81-87] with initial acceptance higher

among women (57%) than men. Initial acceptance varied a great deal across

studies but was highest in those randomized, controlled trials that devoted additional

resources, such as nurses focused on improving adherence, and educational

classes and materials about falls and hip fractures, to improve adherence.[2, 88-93]

Studies in real-world settings, where, for example, hip protectors were mailed to

community-dwelling residents, along with brochures on hip fractures and falls,

showed the lowest level of initial acceptance.

Even among the highly selected group of individuals who initially find the idea

of wearing hip protectors acceptable, adherence declined rapidly after initiation of

most studies. For example, in one study by Hubacher et al[82], 38% of all initial

wearers had stopped wearing the hip protectors by the end of the first month, by the

end of the second month more than half had stopped (59%), and by the end of the

third month more than two-thirds of initial wearers had stopped wearing the hip

17

protectors. Although the declines were not as steep in other studies, adherence

generally declined over time, with much of the decline occurring in the first few

months of the study.[87, 94-99] Comparison across studies is made difficult by the fact

that adherence has been defined and measured in a variety of ways. Some of the

definitions included: average wearing time on active days and during waking hours,

number of user-days per all available follow-up days, percentage of falls with hip

protectors on, percentage of participants who were wearing the hip protector on

most days, and percentage of participants who were wearing the hip protector at a

certain moment.[100] To complicate matters, the duration of follow-up varied

significantly across studies. Nevertheless, the impression persists of fairly low

adherence by the end of the study period, in both observational and randomized

studies. In five of eight randomized controlled trials examining the effectiveness of

external hip protectors, adherence with wearing hip protectors was lower or equal to

50%.[100] In some of the randomized studies, adherence rates were as low as 20%.[89]

Many of the observational studies that have examined adherence have found

adherence rates that were lower than 40% by the end of the study.[7, 73-74, 82-83]

The characteristics of institutional homes that impact long-term adherence

include having health professionals dedicated to promoting adherence, the size of

the institutional setting, and the type of institution (i.e., for-profit and not-for-profit).

Having a dedicated health professional who promotes adherence with hip protectors

increased adherence, when compared with institutional settings where no staff

member was dedicated to improving adherence.   In one study by Parkkari et al[101],

the motivation of the institution staff, who were deeply concerned about the frequent

18

fall injuries among residents in the past, resulted in relatively high adherence rates. There were also differences in adherence, according to the type of person dedicated to promoting adherence. In nursing homes where a nurse was the contact person for adherence, the probability of continued use was lower than when a physical therapist was the contact person.[102] In general, larger nursing homes show a greater probability of continued use with the wearing of hip protectors than small homes.[102] For-profit institutions have been noted to have lower long-term adherence than not-for-profit institutions.[97, 99] The impact of other nursing home characteristics on adherence have not been vigorously examined. However, anecdotal observations suggest that enthusiasm, workload, and other working conditions of the nursing staff may have a large influence on adherence.[97] In a study by Cameron et al, relatively high long-term adherence of 76% during the day and 84% at night was noted in one home for the elderly. This adherence was markedly different from other homes that were participating in the study, and was felt to have occurred mainly because of the enthusiasm and dedication of the nursing staff.[76]

The type of hip protectors, design, fit and appearance all play a role in the long-term adherence to the wearing of hip protectors. Individual hip protectors, such as the Safehip protector, have been found to not be sufficiently comfortable to be worn at night by thin, severely disabled and cognitively impaired older women.[103] When comparing different types of hip protectors in two different studies, the comparison was between a soft hip protector and a hard hip protector. Both studies noted that participants using the softer type were more compliant.[84] In contrast, one study[104] found no difference in the use of the Safehip design, which has a rigid,

19

concave shell that is designed to disperse the force of a fall away from the greater trochanter, when compared to the HipSaver design, which has a flexible foam protector that is designed to absorb the force of a fall. While the Safehip garment was closer fitting and appeared less obvious under clothing, the HipSaver garment was looser and the protector was softer and more flexible, but was more easily visible under clothing.   Across numerous studies, poor fit was often identified as the reason for discontinuation.[73, 81-82, 84, 97, 102, 105] For example, in one study, 70% of the drop-outs felt that wearing the protector was very uncomfortable.[82] The problem of the tightness of the hip protectors cannot be completely resolved because hip protectors need to fit snugly over the greater trochanter to be effective. However, improved design and the availability of additional sizes, could make hip protectors more comfortable and easier to remove,[84] and could reduce the concerns about hip protectors being 'too hot,' 'uncomfortable,' and 'irritating to the skin'.[106] Appearance of the hip protectors was a concern to many residents,[73, 81-82, 84] but especially women.[105] More wearers who found the appearance of the hip protector as un-attractive discontinued wearing the protectors.[84] Conversely, a protective garment that did not show and looked well was associated with positive adherence. [81]

Adherence to the wearing of hip protectors is influenced by a number of individual characteristics, including gender, age, cognitive status, history of falls and fractures, continence problems, self-efficacy, and physical status. Individuals with mild or moderate urinary incontinence find the hip protector underwear acceptable, but difficulties are perceived for those residents who are severely incontinent.[105] Those individuals with urge incontinence report additional incontinent episodes

related to the use of the undergarments and require additional help with dressing and un-dressing.[85]   Both issues lead to increased staff support and often result in decreased adherence with the wearing of hip protectors. Cognitive impairment results in increased agitation with the wearing of protective garments and, thus, decreased adherence for some individuals.[10] Many individuals with cognitive impairment, however, once they acquire the habit of wearing the hip protector, wear it more habitually.[91, 104, 107] The probability of wearing hip protectors also increases for individuals with an increasing number of fall risk factors (two fall risk factors: OR=1.47 (0.83, 2.60); three or more risk factors: OR=2.02 (1.10, 3.71).[82] Individuals who have a history of actually falling also are more likely to be adherent with the wearing of hip protector undergarments.[10, 84, 86-87] A history of fractures or other significant injury in the past twelve months, subsequent to a fall, was predictive of higher adherence.[81, 104] Being female was associated with increased adherence, possibly because women are more susceptible to adherence encouragement by staff, who are also largely female[81-82, 102];increasing age, in contrast, resulted in lower adherence.[84] Finally, individuals with a lower perceived self-efficacy for hip protector use and lower self-rated health were more likely to be adherent with the wearing of protective undergarments.[108] These individuals are more physically dependent and are more susceptible to the influence of staff who encourage them to wear protectors.[84, 104] Individuals with physical and mental deficiencies are relatively powerless, especially in institutional settings, where their satisfaction depends largely on staff.[104] In institutional settings with a high percentage of individuals with

21

dementia and/or physical disabilities, adherence is significantly impacted by the

commitment of nursing and personal care staff in the institution.[76]

Search Strategy and Inclusion Criteria

A literature search of all hip protector studies was conducted using the PubMed, CINAHL, Cochrane, ISI Web of Science, and EMBASE databases on November 13, 2006, and was repeated on August 5, 2009. The specific search criteria for each of the databases are identified in Appendix I. Briefly, the search was limited to randomized control trials (RCTs) in humans, found in full publications, abstracts, and non-English language articles, in which the intervention was a hip protector of any type, and the outcome was a primary or secondary hip fracture. This initial review of the literature resulted in the identification of 48 articles from PubMed, 30 from CINAHL, 43 from Cochrance, 11 from ISI Web of Science; from EMBASE we restricted the search to articles not already identified from PubMed and found 4 additional publications.

In addition, we reviewed all references of any selected articles from the database search, as well as all references in the systematic review of hip protectors by Waldegger et al[7], Sawka et al[79], and Parker et al[109]. When reviewing the titles of articles found in the reference section of previously identified articles and systematic reviews, we retained for further review only articles that had in the title at least one item from each of the two sub-headings below:

Heading 1	Heading 2
Effectiveness	Fall
Intervention	Fracture
Trial	Hip
Randomized	Hip Protector
Prevent	Hip Pad
Harmful	

Other titles could be selected only if highly suggestive of a study of hip protectors to prevent hip fractures.

A total of 697 articles were initially identified by conducting the database search and title search. Of the total 697 articles initially identified for review, 43 articles which met the inclusion criteria for the database search and 7 articles which had in the title at least one item from each of the two sub-headings above were selected for a full abstract review. Abstract review revealed 16 randomized studies of hip protectors in which the outcome was a hip fracture. These studies are described in greater detail in Appendix 2, with particular emphasis on demonstrated efficacy and adherence.

Summary

Of the sixteen studies included in these analyses, ten were carried out in Europe, two in Australia, two in Japan, and two in the United States (Table 1.1). Most of the studies were presented as full manuscripts; however, data for five studies were available only in abstracts, short reports, or letters to the editor. The majority of hip protector studies after 2000 used a protector whose bio-mechanical

properties had been clearly documented, the Safehip protector. For the remaining

studies, the type of hip protector was either specifically designed for the randomized

study or used an alternate commercially available hip protector such as the KPH hip

protector or the JOFA AF hip protector. Many of these alternate hip protectors also

provided evidence of bio-mechanical effectiveness, with the notable exception of the

Ekman and Villar studies.[87, 110] There was variability in the design of hip protectors.

The Safehip protector had an outer shield of polypropylene with an inner plastozote

lining, sewn into a special cotton-lycra underwear. The two hip protectors used in

the HIP PRO studies have been described previously and shown to be different from

the Safehip protector. The protector used in the Ekman[110] study used closed-cell

polyethylene foam, which was similar to the material used in knee protectors of

hockey players. In contrast, the Villar[87] protector was made of polypropylene which

was encased in compressed polystyrene. The remaining non-Safehip protectors

were not described in greater detail. A total of 12 studies occurred in long-term care

settings, while the remaining four studies were carried out in community settings. Of

the studies undertaken in long-term care settings, all studies were cluster

randomized except for the Cameron 2001[103] and van Schoor studies.[107]

Not all studies provided information on the demographic characteristics of the

study population, particularly those randomized studies which presented results in

abstracts, short reports, or letters to the editor. The median or mean age of study

participants ranged from 78 to 86 years, indicating that most study participants were

elderly (Table 1.2). The study inclusion/exclusion criteria of five studies limited the

study population to females. The remaining studies with non-missing gender

suggested that the majority of participants were female, reflecting the fact that women tend to survive longer than men. We will contact the authors of the studies with missing information on gender and age to completely describe the study population of included studies. The total follow-up time ranged from 3 months in the Villar[87] et al study to 28 months in the Birks et al study.

Adherence ranged from a low of 20% in the O'Halloran et al[83] study to a high of 88% in the Koike et al study[111] (Table 1.3). The adherence in the Koike study is surprisingly high, relative to most of the other studies. This high level of adherence may have been the result of cultural differences. The adherence of the Koike et al[111] and the Harada et al[91] study (70%) were both at the high end of the range and both occurred in Japan. The way that adherence was defined was unique in almost all of the studies, making comparison across studies difficult. In five studies, adherence was not defined or not clearly explained. Four studies defined an adherent participant as one who wore the hip protector daily during waking hours. For the remainder of the studies, adherence was defined on a continuum, ranging from such definitions as 'wears protectors for a minimum of one hour per day' or wears protectors 'most of the time' to more strict definitions noting that a participant was adherent only if they wore the protector 24 hours a day. Despite the variability in how adherence was defined, there did not appear to be a relationship with overall adherence. We would expect that more lenient definitions of adherence would result in higher overall percent adherence – however, this was not always true.

Other factors that may have influenced overall adherence included such items as who checked adherence, the frequency of the adherence checks, and whether

26

the adherence check occurred at random times. There were three studies in which

adherence was assessed by self-report, three studies in which it was evaluated by

caregivers, five studies in which adherence was reported by research staff, and

three studies where this information was not provided. Based on previous studies of

adherence, one would expect that individuals providing self-reports and caregivers

may have inflated adherence, while the adherence reported by research staff may

have been the most objective and least prone to error. However, no clear

relationship was evident in the included studies between overall adherence and who

checks the adherence. Note that only in four out of five studies where adherence

was evaluated by research staff had random adherence checks. For a few studies

adherence was checked daily or three times a week; for the remainder of the studies

adherence was infrequent, with the Birks 2004[112] study having the greatest time

between adherence checks, evaluating adherence every six months.

Only a few of the studies met the majority of criteria that denote a high quality

randomized study in which the eligibility criteria are clearly described, baseline

demographic characteristics are tabulated, it is clear that those evaluating the

outcome were blinded to the exposure, the attrition from the study is described, the

study achieved the proposed sample size, and an ITT analysis was undertaken

(Table 1.4). In particular, the earliest two studies of hip protectors (which showed a

strong protective effect of hip protectors) were of poor quality (Lauritzen[88] and

Ekman[110]). But even the two highest quality studies of Kannus et al[89] and Kiel et al[16]

either failed to indicate whether the outcome assessor was blinded or failed to

accrue sufficient study participants because of an over-optimistic assumption of hip

fracture rates.

In evaluating efficacy, we excluded from the analysis two studies in which the

hip protector protected only one hip (Birge[113], Kiel[16]) and one study in which the

outcome was a second hip fracture (Birks 2004).[112] The mechanism of biologic

action was perceived to be different in these three studies because of differences in

design or because the study population was at a much higher risk of hip fracture.

This left 13 studies for the evaluation of efficacy (Figures 1.2-1.5). For one study in

which there were no events in the protected and unprotected hips (Villar),[87] we used

the simple continuity correction of Sweeting et al[114], so that this study could be

included in the analysis. We add 0.5 to each of the four cells in the 2 X 2 table.

With this method, the data equivalent prior for the analysis of the Villar study[87]

(which received the continuity correction) was based on a hypothetical previous

study with N=2, a balanced design ($N_1=N_0=1$), and OR=1.0.

For the included studies, we evaluated publication bias by evaluating funnel

plot asymmetry with the trim and fill method of Duval and Tweedie.[115] Funnel plot

asymmetry was not indicated by visual inspection, and the trim and fill method did

not impute data for any hypothetically missing studies. This suggested that there

was no publication bias for the included hip protector studies.

Heterogeneity between studies was examined by calculating a p-value for the

Cochran Q, which was defined as $Q = \sum_{i=1}^{13}(1/\sigma_i^2)[(RD)_i - (\overline{RD})]^2$, where $\overline{RD}$ is the

inverse variance weighted average risk difference, and is distributed as χ^2 with (I − 1

= 12) degrees of freedom. Pronounced heterogeneity was observed in the 13

included hip fracture studies (Cochran Q=41.09, p<0.0001). This heterogeneity is

explored in Figures 1.2-1.5, which present forest plots of the risk difference and 95%

confidence intervals of each of the individual studies, as well as a random effects

summary risk difference and its associated 95% confidence interval. Each figure

presents the studies in order, by four study characteristics that are hypothesized to

cause the observed between-study heterogeneity. Figure 1.2 presents the studies in

order of total follow-up time. Shorter duration studies, those with follow-up of

months or less, appear to show a stronger protective effect of hip protectors, but this

relationship does not appear strong. In figure 1.3 we present data for included

studies, by overall adherence. We would expect that studies with lower adherence

would be more likely to be biased toward the null. However, this does not appear to

be the case, possibly because of the great variability in how adherence is defined

and measured. The three studies in Figure 1.4 which were in community settings

appear to be different from the studies that occurred in long-term care settings.

Importantly, we excluded one study by Birks et al[112] which was in the community and

revealed a precise risk difference of 0.04. This study was excluded because the

outcome that was evaluated was a second hip fracture. Finally, figure 1.5 orders the

included studies by whether the intervention was a Safehip protector or some other

protector. The non-Safehip protector studies appear to have a somewhat stronger

protective effect, when compared to the studies of Safehip protectors. While some

of these study characteristics seem to explain some of the heterogeneity in the

included studies by visual inspection, a more thorough analysis requires the use of

meta-regression. We have chosen to defer a meta-regression until we are able, if possible, to fill in missing information on gender and age for some of the included studies.

Discussion

The earliest studies to investigate the protective effect of hip protectors were founded on extensive biomechanical testing of the hip protectors, as well as on mounting evidence that hip fractures occurred mainly as a result of a fall on to the greater trochanter, where the skin and subcutaneous fat were insufficient to absorb and disperse the force of the fall, and the underlying bone had deteriorated as a result of osteoporosis. Lauritzen et al[88] randomized nursing homes in a 1:2 ratio to either receive hip protectors or not receive hip protectors; 215 individuals received hip protectors, and 418 did not. The relative risk for hip fracture was 0.44 (0.21,0.94), when comparing the intervention hip–protector group to the controls.[88] Other, smaller randomized studies in the late nineties also hinted at a strong protective effect of hip protector.[87, 110, 116] The design and analysis of these early studies, however, was questioned, bringing into doubt the findings from these early studies. In the Lauritzen et al[88] study, for example, individuals who died were replaced by new arrivals from a waiting list, leading to the possibility of selection bias (i.e., individuals might have gained admission preferentially to a nursing home that was randomized to wearing hip protectors). Also, in the Lauritzen study, randomization occurred at the nursing home level, but analysis was conducted at the

individual level. The main consequence of adopting a cluster design is that the outcome for each patient can no longer be assumed to be independent of that for any other patient. Patients within any one cluster are more likely to have similar outcomes.[117] If the clustering effect is ignored, confidence intervals might be over-narrow, increasing the chances of spuriously significant findings and misleading conclusions.[117]

Additional studies of hip protectors, some of which attempted to address the design flaws and analysis flaws of earlier studies, were conducted in the first part of the new millennium, in both institutional settings and in the community.[83, 89, 91, 94-96, 103, 112, 118] Systematic reviews of the efficacy of hip protectors included different studies, according to the strictness of their inclusion/exclusion criteria, and came to conflicting conclusions.[7, 79, 109] The systematic review by Waldegger et al included only five studies which were conducted in institutional settings.[88-89, 91, 110, 119] Their results indicated a relative risk of hip fractures of 0.35 (0.23, 0.51), when comparing the hip protector intervention group against controls and noted no significant heterogeneity between studies. [7] Another analysis by the Cochrane group[109], included many more studies that were carried out in nursing or residential care settings, as well as a few studies that were conducted in the community.[82-83, 88-89, 91-92, 94-96, 110, 112, 116, 118-119] They noted that hip protectors are an ineffective intervention for those living at home (RR=1.16 (0.85, 1.59)), but are marginally protective in individuals living in institutions who have access to care staff that encourage adherence and provide help with the protective undergarments (RR=0.77 (0.62, 0.97)). Sawka et al, in their systematic review, used much stricter

inclusion/exclusion criteria and excluded cluster randomized trials that did not

account for the clustering in the analysis, as well as trials that allowed individuals

who died or were lost to follow-up to be replaced by other individuals from a waiting

list. In the four trials that were included in their analysis[95-97, 112], they found no

protective effect of hip protectors (RR=1.07 (0.81, 1.42)).[79] These results from the

systematic reviews of protective undergarments, as well as the results from the

individual studies, provide conflicting evidence of the effectiveness of hip protectors

to prevent hip fractures.

Table 1.1 Characteristics of 16 Studies of the Risk of Hip Fractures Among Elderly Individuals Wearing Hip Protectors When Compared with Individuals not Wearing Hip Protectors.

First Author	Year	Country	Publication Type	SafeHip Protector	Long-term Care Setting	Cluster Randomization
Lauritzen	1993	Denmark	Manuscript	No	Yes	Yes
Ekman	1997	Sweden	Short Report	No	Yes	Yes
Jantti	1998	Finland	Letter	No	Yes	No
Villar	1998	U.K.	Manuscript	No	No	No
Kannus	2000	Finland	Manuscript	No	Yes	Yes
Birge	2001	U.S.	Abstract	No	Yes	Yes
Cameron	2001	Australia	Manuscript	Yes	Yes	No
Harada	2001	Japan	Manuscript	No	Yes	Yes
Birks	2003	U.K.	Short Report	Yes	No	No
Cameron	2003	Australia	Manuscript	Yes	No	No
Meyer	2003	Germany	Manuscript	Yes	Yes	Yes
van Schoor	2003	Netherlands	Manuscript	Yes	Yes	No
Birks	2004	U.K.	Manuscript	Yes	No	No
O'Halloran	2004	Ireland	Manuscript	Yes	Yes	Yes
Koike	2005	Japan	Abstract	Yes	Yes	Yes
Kiel	2006	U.S.	Manuscript	No	Yes	Yes

Table 1.2. Summary Patient Characteristics of 16 Studies Evaluating the Risk of Hip Fractures

First Author	Year	Total Follow-up (Months)	Mean or Median Age	% Female
Lauritzen	1993	11	.	67
Ekman	1997	11	84.0	.
Jantti	1998	12	.	.
Villar	1998	3	81.4	100
Kannus	2000	24	82.0	78
Birge	2001	12	.	.
Cameron	2001	18	85.0	100
Harada	2001	12	83.2	100
Birks	2003	14	80.5	87
Cameron	2003	24	83.0	100
Meyer	2003	18	86.0	86
van Schoor	2003	17	85.0	89
Birks	2004	28	77.8	100
O'Halloran	2004	18	.	.
Koike	2005	10	.	100
Kiel	2006	20	85.3	79

Table 1.3. Adherence Characteristics of 16 Studies Evaluating the Risk of Hip Fractures.

First Author	Year	% Adherence	Adherence Definition[a]	Who Checks Adherence?	Frequency of Adherence Checks	Random Adherence Checks?
Lauritzen	1993	24.0	.	.	.	.
Ekman	1997	44.0	.	.	.	.
Jantti	1998	44.0	.	.	.	.
Villar	1998	26.7	9	Research Staff	Fortnightly	Yes
Kannus	2000	48.0	8	Caregiver	Daily	No
Birge	2001	63.0	7	.	.	.
Cameron	2001	57.0	1	Research Staff	Day 18, 67, 321, 544	No
Harada	2001	70.0	5	Caregiver	Daily	No
Birks	2003	22.0	1	Self-report	.	No
Cameron	2003	53.0	1	Self-report	Month 1, 3, 12, 18, 24	No
Meyer	2003	67.0	3	Caregiver	.	No
van Schoor	2003	37.0	9	Research Staff	Month 1, 6 12	Yes
Birks	2004	31.0	2	Self-report	Every 6 Months	No
O'Halloran	2004	19.9	6	Research Staff	Weeks 2, 4, 8, 12, 18	No
Koike	2005	87.5	1	.	.	.
Kiel	2006	77.8	4	Research Staff	3 X Week	Yes

[a]Participant was deemed to be adherent according to different definitions: (1) wears daily during waking hours, (2) if wears the protector 'most of the time', (3) when wearing protector during a fall, (4) wearing protector at time of adherence visit , (5) wears protectors 24 hours a day, (6) wearing protector at time of nurse facilitator visit, (7), during time out of bed, (8) wearing protector for minimum of one hour per day, (9) not clear

Table 1.4. Quality of 16 Studies Evaluating the Risk of Hip Fractures.

First Author	Year	Eligibility Criteria Described?	Outcome Assessors Blinded?	Attrition Described?	ITT Analysis?	Achieved Sample Size?	Baseline Variables Tabulated?
Lauritzen	1993	No	.	No	Yes	.	No
Ekman	1997	No	.	No	Yes	.	Yes
Jantti	1998	Yes	.	Yes	Yes	.	No
Villar	1998	Yes	.	Yes	.	.	No
Kannus	2000	Yes	.	Yes	Yes	Yes	Yes
Birge	2001	Yes	.	No	Yes	.	No
Cameron	2001	Yes	No	Yes	Yes	No	Yes
Harada	2001	Yes	.	No	Yes	.	Yes
Birks	2003	Yes	No	Yes	Yes	No	Yes
Cameron	2003	Yes	Yes	Yes	Yes	No	Yes
Meyer	2003	Yes	.	Yes	Yes	Yes	Yes
van Schoor	2003	Yes	No	Yes	Yes	No	Yes
Birks	2004	Yes	No	Yes	Yes	No	Yes
O'Halloran	2004	Yes	.	Yes	Yes	Yes	Yes
Koike	2005	Yes	.	Yes	Yes	.	No
Kiel	2006	Yes	Yes	Yes	Yes	No	Yes

Figure 1.2. Summary Risk Differences and 95% Confidence Intervals for the Effect of Hip Protectors, by Follow-up

Study ID	ES (95% CI)	% Weight	Total Follow-up (in Months)
Villar	-0.01 (-0.04, 0.03)	7.51	3
Koike	-0.04 (-0.07, -0.01)	8.78	10
Ekman	-0.03 (-0.05, -0.00)	10.54	11
Lauritzen	-0.04 (-0.08, -0.01)	8.05	11
Harada	-0.09 (-0.17, -0.02)	3.16	12
Jantti	-0.17 (-0.31, -0.03)	1.02	12
vanSchoor	-0.00 (-0.04, 0.04)	6.64	17.4
Cameron	0.01 (-0.07, 0.10)	2.52	18
Meyer	-0.04 (-0.07, -0.00)	8.58	18
OHalloran	0.00 (-0.01, 0.02)	11.99	18
Cameron2	-0.00 (-0.05, 0.04)	6.60	24
Kannus	-0.04 (-0.06, -0.02)	11.63	24
Birks	0.00 (-0.01, 0.01)	12.97	28
Overall (I-squared = 71.2%, p = 0.000)	-0.02 (-0.04, -0.01)	100.00	

NOTE: Weights are from random effects analysis

-.307   0   .307

Figure 1.3. Summary Risk Differences and 95% Confidence Intervals for the Effect of Hip Protectors, by Adherence

Study ID	ES (95% CI)	% Weight	Overall Adherence
OHalloran	0.00 (-0.01, 0.02)	11.99	19.9
Lauritzen	-0.04 (-0.08, -0.01)	8.05	24
Villar	-0.01 (-0.04, 0.03)	7.51	26.7
Birks	0.00 (-0.01, 0.01)	12.97	31
vanSchoor	-0.00 (-0.04, 0.04)	6.64	37
Ekman	-0.03 (-0.05, -0.00)	10.54	44
Jantti	-0.17 (-0.31, -0.03)	1.02	44
Kannus	-0.04 (-0.06, -0.02)	11.63	48
Cameron2	-0.00 (-0.05, 0.04)	6.60	53
Cameron	0.01 (-0.07, 0.10)	2.52	57
Meyer	-0.04 (-0.07, -0.00)	8.58	67
Harada	-0.09 (-0.17, -0.02)	3.16	70
Koike	-0.04 (-0.07, -0.01)	8.78	87.5
Overall (I-squared = 71.2%, p = 0.000)	-0.02 (-0.04, -0.01)	100.00	

NOTE: Weights are from random effects analysis

-.307   0   .307

Figure 4. Summary Risk Differences and 95% Confidence Intervals for the Effect of Hip Protectors, by Setting

Study ID	ES (95% CI)	% Weight	Setting
Birks	0.00 (-0.01, 0.01)	12.97	0
Cameron2	-0.00 (-0.05, 0.04)	6.60	0
Villar	-0.01 (-0.04, 0.03)	7.51	0
Cameron	0.01 (-0.07, 0.10)	2.52	1
Ekman	-0.03 (-0.05, -0.00)	10.54	1
Harada	-0.09 (-0.17, -0.02)	3.16	1
Jantti	-0.17 (-0.31, -0.03)	1.02	1
Kannus	-0.04 (-0.06, -0.02)	11.63	1
Koike	-0.04 (-0.07, -0.01)	8.78	1
Lauritzen	-0.04 (-0.08, -0.01)	8.05	1
Meyer	-0.04 (-0.07, -0.00)	8.58	1
OHalloran	0.00 (-0.01, 0.02)	11.99	1
vanSchoor	-0.00 (-0.04, 0.04)	6.64	1
Overall (I-squared = 71.2%, p = 0.000)	-0.02 (-0.04, -0.01)	100.00	

NOTE: Weights are from random effects analysis

-.307   0   .307

Figure 5. Summary Risk Differences and 95% Confidence Intervals for the Effect of Hip Protectors, by Hip Protector Type (SafeHip vs Other)

Study ID		ES (95% CI)	% Weight	safehip
Ekman		-0.03 (-0.05, -0.00)	10.54	0
Harada		-0.09 (-0.17, -0.02)	3.16	0
Jantti		-0.17 (-0.31, -0.03)	1.02	0
Kannus		-0.04 (-0.06, -0.02)	11.63	0
Lauritzen		-0.04 (-0.08, -0.01)	8.05	0
Villar		-0.01 (-0.04, 0.03)	7.51	0
Birks		0.00 (-0.01, 0.01)	12.97	1
Cameron		0.01 (-0.07, 0.10)	2.52	1
Cameron2		-0.00 (-0.05, 0.04)	6.60	1
Koike		-0.04 (-0.07, -0.01)	8.78	1
Meyer		-0.04 (-0.07, -0.00)	8.58	1
OHalloran		0.00 (-0.01, 0.02)	11.99	1
vanSchoor		-0.00 (-0.04, 0.04)	6.64	1
Overall (I-squared = 71.2%, p = 0.000)		-0.02 (-0.04, -0.01)	100.00	

NOTE: Weights are from random effects analysis

-.307   0   .307

40

CHAPTER 2

METHODS

To date, a number of studies have been conducted into the efficacy of hip protectors to prevent hip fractures: the results have been mixed.[82-83, 88-89, 91, 94-96, 103, 110, 112, 116, 118-119] Notably, in all of these studies, those that demonstrated the efficacy of hip protectors and those that did not, adherence to the wearing of protective undergarments has been low. No study has noted an adherence rate above 80%, and most have overall adherence rates that are below 50%. This is true in observational studies, as well as in randomized controlled trials. Despite adherence rates that are below 80%, sometimes substantially below 80%, the randomized studies report intention-to-treat (ITT) results as if they were causal effect estimates. Such low adherence rates, however, eliminate the benefits of randomization as a tool to estimate etiologic effects and may result in biased effect estimates. We propose to correct for the poor adherence in on one randomized trial with three relatively new methods for compliance correction, instrumental variable analysis[120-121], inverse probability of censoring weights (IPCWs)[122], and structural nested models (SNMs).[123-124]

STUDY SETTING AND CLINICAL PROCEDURES

For these analyses, we will use data from the Hip Impact Protection Project (HIP PRO), a multi-center, randomized, controlled trial of the efficacy of an energy absorbing hip protector for the prevention of hip fractures in nursing homes.[16, 125] This study was conducted in three states, Massachusetts, Maryland, and Missouri, with data collection beginning in October 2002 and continuing through August 2006. HIP PRO was designed to eliminate some of the design flaws of previous studies. Hip protectors were designed to be worn on only one hip, so that only one hip would be protected, and each individual would serve as his/her own control. Randomization of the side of the hip to be protected occurred at the level of the nursing home. As in other studies, significant efforts were expended to improve adherence. Unlike previous studies, adherence was assessed three times a week at randomly timed visits by study staff. The hip fracture outcome was evaluated by at least two members of a Clinical Endpoints Committee; all members of the Clinical Endpoints Committee were blinded to the side of the protected hip.

HIP PRO was a study designed to investigate the efficacy and safety of hip pants in the prevention of hip fractures in an elderly nursing home (NH) population.[16] With a unique under-garment design, it attempted to address some of the problems of earlier hip protector studies in NHs. Residents wear a hip protector (of the energy absorbing and shunting type) on one hip, so that each resident serves as his or her own control. This design feature eliminates the problems of selection bias that were present in previous studies. Randomization as to the side that the hip protector

42

would be worn on was done at the nursing home level, mainly for logistics and convenience. The main research aims of the HIP PRO study were to: (1) determine whether a trochanteric pad, inserted into undergarments with a pocket on one side, will reduce the incidence of hip fracture on the protected side by 50%, when compared to unprotected side, in elderly NH residents; (2) evaluate those resident-level and facility-level factors that impact adherence with the wearing of the protective undergarments.

To address these aims, intermediate and skilled nursing facilities that had a minimum of 100 beds, had participated in other research projects, employed charge nurses who are not from agencies, had a good reputation among health care providers and families, and were geographically close to a HIP PRO Clinical Center were recruited into the study. These screening criteria were designed to screen out those facilities that might provide poor quality data or who might drop out early from the study. Within each enrolled facility, a complete census of residents was taken and residents who were older than age 65, had gotten out of a bed or chair without human assistance in the past 4 weeks, and were in a non-Medicare, non-acute, non-rehabilitation bed, were identified as potential candidates for inclusion in the study. Residents with any of the following criteria were excluded from the HIP PRO study: (1) refused consent or a responsible party refused to give informed consent, (2) had a hip circumference greater than 48 inches, (3) had pressure sores, skin tears, or skin shearing in readily viewable areas over bony prominences covered by the hip protector undergarment, (4) had prior bilateral hip fractures or hip replacement, (5) had bilateral surgical scars over both hips, (6) had a terminal illness expected to

result in death in the next 6 months, (7) had signs or symptoms of a hip fracture at the time of enrollment, (8) refused to have their hips examined at the time of enrollment, (9) generally refused to wear clothing during the day, (10) were in isolation for a contagious disease, or (11) the nursing home staff strongly recommended not approaching.

Hip Fracture Outcome

The majority of hip fractures require surgery to stabilize the fracture, reduce the pain associated with the hip fracture, and allow the healing process to begin. Consequently, most of the hip fractures in HIP PRO were evident and readily identified. However, to make sure that HIP PRO identified all hip fracture outcomes, a fracture surveillance protocol was initiated. The purpose of the surveillance program was not only to identify all fractures, but also to ascertain whether the protective underwear and protector were being worn at the time of the fall and the protector was properly positioned. The main source of hip fracture identification in the fracture surveillance program was the fracture reporting hot line. Nursing home staff who became aware of a hip fracture were expected to call the fracture reporting hot line. Once the hot line had been activated the following items were completed: (1) a Fall Report Form, and (2) a Resident Hospitalization Form. When a hip fracture was not reported to the study hotline and the study staff became aware of the hip fracture during site visits, the research assistant was expected to complete the relevant forms. Also, any transfer of a participating member to a hospital

resulted in a request for all hospitalization records, so that it could be determined whether a hip-fracture did or did not occur. At times, hip fractures occurred that did not require hospitalization and were not apparent to nursing home staff. Study staff were trained to also check for signs and symptoms of clinically unapparent hip fractures in participating residents. These signs and symptoms included: (1) new pain in the hip region that was not present at previous visit, (2) new and acute inability to bear weight or transfer, (3) acute change in mobility not due to other obvious reason, (4) externally (occasionally internally) rotated leg, and (5) bruise or skin tears over the femoral trochanter. Any hip fractures that were suspected but clinically unapparent were confirmed by radiologic examination.

All hip fractures were evaluated by at least one orthopedic surgeon and one geriatrician, using information from ER visits, hospital discharge summaries, x-ray reports, and operating notes. If these two physicians were not able to agree about the hip fracture outcome, a second orthopedic surgeon was contacted to act as the tie-breaker. Particularly difficult cases were discussed and a consensus for these cases was reached at annual meetings by the full five-member HIP PRO Clinical Endpoints Committee (CEC).

Adherence

Two-Week Run-In Period

      After the resident was deemed eligible for the study and had signed informed consent, he/she went through a two-week adherence run-in period. During this two week run-in period, HIP PRO staff made three unannounced visits per week to the nursing home. If the resident demonstrated adequate adherence (undergarment on forward with the waistband in the proper position, protector in proper placement over the greater trochanter) during the run-in period in four out of the six total visits, that resident could then continue to participate in the main HIP PRO study. All six adherence checks in the two week run-in period were required for subsequent data analysis.

Main Study

      One of the unique features of HIP PRO and one of its main strengths was the monitoring of adherence by study staff. Study staff monitored adherence during three random, unannounced visits per week. These visits occurred at any time during the week, including weekends and all shifts, with the exception that residents were not disturbed between the hours of 10 p.m. and 6 a.m. Although fractures do occur at night, the vast majority occur during waking hours and the adherence visits were designed to occur during the hours when residents were most at risk for hip

fractures. During the adherence visits, the staff noted the date, time, and results of

the visit, including, if the resident was non-adherent, the reason for the non-

adherence. When a hip fracture was suspected and the resident was not available

in the nursing home, adherence was reported by the resident or responsible family

member. These adherence reports were marked as self-reports and were obtained

from the resident if she had a Short Blessed Test (SBT)[126] score less than 12 or

from the family member if the resident had an SBT score greater than 12. Residents

not contributing any adherence data for more than five weeks (due to hospitalization,

transferred to another institutional setting, etc.) were withdrawn from HIP PRO. The

adherence monitoring data for any particular HIP PRO staff were compared to the

adherence monitoring data of a second observer who independently assessed

adherence. These checks on the quality of the adherence data were done at the

beginning of the study and every six months thereafter. Nursing homes that

demonstrated poor overall adherence for a period of two months or more were

provided with additional training. Continued poor adherence resulted in recruitment

being stopped at that nursing home.

Factors Affecting Adherence

There are many individual-level and facility-level variables that could

potentially impact adherence with the wearing of protective under-garments. The

individual-level characteristics include demographics (age, educations status,

gender, etc.), past fall/fracture history, other medical history, assessment of physical

functioning, level of agitation, etc. All of these factors were assessed at the baseline

visit. For those residents who are not cognitively impaired (i.e., have a Short

Blessed Test score less than 12), additional information about fear of falling, body

image rating, and perceived need for hip protectors was assessed because these

factors have been shown to impact adherence in other studies. Facility-level factors

that could potentially impact adherence, including facility size, ownership, affiliation,

staff turnover, number of temporary staff, and others, were collected during baseline

interviews with the director of nursing and the nursing home administrator.

Study Results – HIP PRO Study I

HIP PRO started enrolling nursing homes and residents in October 2002.

The hip protector that was initially used in HIP PRO was made of several different

materials, including polyethylene and ethylene vinyl acetate. The hip protector, with

dimensions of 4.5" x 6.5" x 0.79" was placed in an undergarment, which was

produced and distributed by FallGard Co. of Napaville, Illinois, of high-quality cotton

Lycra. After 20 months of follow-up, the first study was terminated due to lack of

efficacy. In this first study there were a total of 1042 residents who were deemed

eligible for the study. Of these, 192 did not complete the two-week adherence run-in

period; however, all individuals were included in the final analysis. The study

population was characterized as being primarily female (79%), of white ethnicity

(86%), were severely cognitively impaired (71%) and had a mean age of 85 years

(s.d. 7.4). Residents exhibited a range of physical functioning, and were, for the

most part, at least partially incontinent of bladder and bowel. There were a total of 69 (7%) individuals with a history of hip replacement, 163 (16%) with a history of hip fractures, and 301 (29%) with a history of falls in the past 30 days. The nursing homes which contributed data to this study were large (>100 beds) and equally likely to be for-profit and not-for-profit.

Each resident could contribute only one hip fracture and was removed from the study after the index hip fracture. As a result, we do not worry about correlation within individuals, despite the fact that the same individual contributed a protected and unprotected hip. Clustering within nursing homes was ignored for the main analysis since the intra-class correlation (ICC) was small. After 676 hip-years of follow-up, 21 hip fractures occurred in protected hips and 17 hip fractures occurred in unprotected hips. Out of a total of 1042 enrolled NH residents, 38 (4%) experienced the primary outcome (Table 2.1). The first HIP PRO study was terminated early due to a lack of efficacy. The intent-to treat analysis found a matched rate ratio of 1.24 (95% confidence interval of 0.65 to 2.34), when comparing hips which were randomly assigned to be protected with the contra-lateral hips which were not protected.

Table 2.1.Hip Fractures on Protected versus Unprotected Hips (HIP PRO[1] Study 1)

	Intent-to-Treat (N=1,042)	
	Protected	Unprotected
Number of Hip Fractures	21	17
Hip-Years of Observation	676	676
Incidence Rate	3.1 (1.8, 4.4)	2.5 (1.3, 3.7)

[1]Hip Impact Protection Project.

Adjusting for clustering within nursing homes did not meaningfully change the results. Kiel et. al.[16] also noted that in residents who had an overall adherence of 80% or greater (32% of all residents), 15 hip fractures occurred in protected hips and 10 hip fractures occurred in unprotected hips, across a total of 284 hip-years of observation. If we assume in this subset of 'adherent' residents, as was true for the total study population, that protected and unprotected hips had equivalent follow-up, that no individual had more than one hip fracture, and that the intra-class correlation between nursing homes is small, the matched per-protocol rate ratio, comparing hip fractures in protected and unprotected hips, would be 1.50 (0.67, 3.34).

The overall adherence for all residents, who were followed up over an average of 7.8 months, was 73.8%. At the beginning of the study adherence was initially 60%, but increased to approximately 80% at the 6-month point, before starting to decline down to approximately 70% by the end of the study.

Study Results – HIP PRO Study II

After the first HIP PRO study was terminated, a second study, using similar methods, but incorporating a different hip protector, was conducted. The hip protector used in the second study was the FallGard FG-04 hip protector. In the second study, a total of 1445 residents contributed 1004 hip years of observation between October 2004 and August 2006. Some nursing home residents were switched to the new hip protector after termination of the first study, while 988 entirely new residents were also enrolled into the second hip protector study. As in the first study, participating residents had a mean age of 85 ± 8 years, and were primarily women (78%), white (86%), and had severe cognitive impairment (72%). All but four nursing homes participating in the first study also contributed data to the second study; in addition, 18 new nursing homes were recruited. These nursing homes were large ($\geq$ 100 beds), mainly not-for-profit institutions (55%).

In the second HIP PRO study, after 1004 hip-years of follow-up, 17 hip fractures occurred in protected hips and 15 hip fractures occurred in unprotected hips (Table 2.2). The intent-to-treat analysis in this second study found a matched rate ratio of 1.13 (95% confidence interval of 0.57, 2.27), when comparing protected hips to unprotected hips. Again, adjusting for clustering within nursing homes did not meaningfully change the results.

Table 2.2.Hip Fractures on Protected versus Unprotected Hips (HIP PRO[1] Study 2)

	Intent-to-Treat (N=1,445)	
	Protected	Unprotected
Number of Hip Fractures	17	15
Hip-Years of Observation	1004	1004
Incidence Rate	1.7 (1.1, 2.7)	1.5 (0.9, 2.5)

[1]Hip Impact Protection Project.

Causal Inference and the HIP PRO Study.

As has been long established, causal effect estimates can be consistently

estimated in randomized experiments under certain ideal situations. Assuming a

sufficient sample size, no loss to follow-up, and full adherence with treatment, the

observed conditional risk (Pr[Y=1|A=a]) in a randomized study is equal to the

marginal counterfactual risk (Pr[Y_a=1]); therefore, the associational risk ratio (or risk

difference or odds ratio) from a randomized study is equal to the causal risk ratio (or

risk difference or odds ratio).[127] In essence, the two groups produced by

randomization are comparable or, more precisely, exchangeable. As a result, which

particular group receives the treatment is irrelevant. This simple consequence of

randomized studies means that under ideal conditions, the associational effect

estimates obtained from randomized studies are causal effect estimates. It should

be mentioned that, among other things, an ideal experiment would also have no

measurement error of the exposure and outcome, and there would be no biases

arising from the fact that the participant, the healthcare workers responsible for the participant's treatment, or the person responsible for measuring or evaluating the outcome know which treatment has been allocated. Such an ideal experiment may not be feasible in a real world setting; however, it provides a conceptual framework in which determination of a 'causal' relationship would be feasible.[128]

Unfortunately, as we move away from an ideal randomized study, the effect estimates obtained from randomized studies can no longer be viewed as causal effect estimates. Specifically, in the presence of non-adherence, an intention-to-treat (ITT) analysis guarantees only exchangeability of two groups that are now defined by a misclassified exposure (the original treatment assignment). It has sometimes been suggested, incorrectly, that the as-treated analysis or the per-protocol analysis resolves this problem. While these analyses do guarantee a correct classification of exposure, they do not guarantee exchangeability of the groups defined by this exposure. The advantage of the ITT analysis, when compared to the as-treated analysis and the per-protocol analysis, is that it provides an unbiased associational measure (and, thus, a causal effect estimate) if the casual null hypothesis holds for the exposure that was actually observed under non-adherence. (As an aside, note that in randomized trials with a survival endpoint the treatment effect can be evaluated with the following proportional hazards model: $\lambda_i(t)$ = $\lambda_0(t) \exp(\psi_H R_i)$. The estimation of the risk ratio parameter $\exp(\psi_H)$ is based on the partial likelihood of Cox. The logrank test, which is equivalent to the partial likelihood score test, can be used for testing the causal null hypothesis of no treatment effect.[129]

What this means in practical terms for the HIP PRO analysis is that, if there is truly no causal (protective or harmful) effect of hip protectors on hip fractures, then the ITT analysis provides an unbiased estimate of this effect. But what if hip protectors actually do provide protection against hip fractures in the true study population? In that case, under non-adherence, the ITT analysis and the as-treated analysis will both fail in estimating the true causal effect of hip protectors on hip fractures, because, as noted above, neither approach guarantees exchangeability of the study groups. Three relatively new methods in the epidemiology literature correct for non-adherence in randomized studies, and can provide unbiased effect estimates: instrumental variable analysis (which comes from the econometric literature)[120-121, 130], marginal structural models (using inverse probability of treatment weights) by Robins et al[122, 124, 131-132], and structural nested models, also by Robins et. al.[123-124, 133]

EFFICACY VERSUS EFFECTIVENESS

Our motivation to correct for the non-adherence in the HIP PRO study is to estimate the effect of treatment for all persons who receive the therapeutic agent to which they were assigned (i.e., hip protectors or no hip protectors). This estimate, which measures the biologic action of hip protectors among adherent participants, has been conceptualized as estimating 'biologic efficacy' by Sommer and Zeger[134] and 'method-effectiveness' by Sheiner and Rubin.[135] We will consider this estimate as simply defining the 'efficacy' of treatment. An alternative estimate which is often of interest in randomized studies is the effect of treatment assignment, sometimes referred to as an estimate of 'use-effectiveness' [135] or 'programmatic effectiveness'[134], but which we will consider simply as an estimate of 'effectiveness.' In a randomized study, when adherence is complete, the estimate of efficacy equals the estimate of effectiveness.

In the presence of important non-adherence investigators have traditionally undertaken a number of alternative strategies to analyze data from randomized studies. One somewhat simplistic approach that is sometimes undertaken to account for non-adherence is to adjust for the non-adherence. This adjustment apparently should allow for the estimation of the effect of treatment (i.e., randomization) on the outcome. Unfortunately, this approach is incorrect because stratification on intermediate variable (e.g., adherence, which is on the causal pathway) can lead to biased estimates of the direct effect of randomization on the

outcome. To understand why, consider that adherent patients may be different from non-adherent patients on characteristics that affect both the self-selection for treatment and the outcome. For example, in the hip protector studies, individuals who are sicker and at higher risk for hip fracture may be more likely to wear hip protectors. When this is true, a comparison of adherent and non-adherent patients will be invalid. Because of problems of this sort, investigators often rigidly follow ITT analysis in randomized studies.[121]

An ITT (as randomized) analysis compares observed outcomes according to initial group assignment. ITT analyses provide valid estimates and associated tests and confidence intervals for the effect on outcome of assignment to therapy (i.e., for effectiveness).[135] However, for estimating efficacy, the ITT estimate suffers bias toward the null in proportion to the degree of non-adherence.[136] This happens, for example, when hips assigned to wear protectors do not wear them (for whatever reason), so that the group originally assigned to wear protectors are now comprised of adherent and non-adherent hips. The dilution effect of non-adherent hips makes it difficult to estimate efficacy with conventional ITT analyses and makes it particularly difficult to answer the question of how large the treatment effect might be.[136]

Despite the limitation of ITT analysis to estimate efficacy under non-adherence, the ITT estimate has become the gold standard for analyzing data from randomized studies. One of the strengths that is often posited by advocates of the ITT analysis is that only the original randomized assignment is guaranteed, in expectation, to be un-confounded, and so the only inferences that can be trusted

must be based on the original assignment.[135] This is an important point that is worth further consideration. However, it must be considered in the context of what question the ITT analysis is designed to address. Recall that the ITT analysis is designed to evaluate outcome differences associated with the prescription of alternative treatments, not of the treatment themselves. Sheiner et al[135] argue convincingly that the ITT analysis, in the presence of non-adherence, can only validly determine whether, under usual conditions of use, the new treatment produces net benefit. So, an ITT analysis does not answer the question of whether the therapy (e.g., hip protectors) could help if utilized; rather, it answers the question of whether the therapeutic 'policy' of providing hip protectors actually helps. Even in the limited subset of studies which evaluate the effectiveness of the intervention, however, an ITT analysis is only valid if the trial design ensures that most of the features potentially affecting adherence resembles those prevalent in practice.[135] When this is not true, the results of the trial will only have internal validity (i.e., comparisons between the two treatment arms are unbiased for a population that is similar to the one that has been particularly selected for a given trial); however, it will not have external validity (i.e., the results of the trial do not apply broadly to the general population).[137]

While acknowledging that the ITT analysis does not provide a valid estimate of the strength of treatment (under non-adherence), some scientists argue that it does provide evidence of the existence of efficacy. This is because the ITT estimate is unbiased for both efficacy and effectiveness (because they are both zero), and an alpha-level test of effectiveness is equal to an alpha-level test of efficacy. Following

this logic, in the scenario where non-adherence is substantial and the ITT estimate is downward biased for efficacy, an ITT analysis would suffer some loss of power since it would fail to reject the null hypothesis when it is false. However, this loss of power could easily be overcome by choosing a larger study size.[135]

The problems with the above line of reasoning are two-fold. First, in many situations, the extent of non-adherence may be unknown *a priori* and it may be difficult to determine an adequate sample size to account for non-adherence. In this situation, confusion can result about what future course of action should be undertaken if the null hypothesis of no treatment effect cannot be rejected. If the intervention is not efficacious, then it should be abandoned. Alternatively, if the intervention failed because of non-adherence then efforts should be taken to discover and address the reasons for non-adherence. The ITT estimate alone provides no guidance about which course of action is best. Second, even if the level of non-adherence in the proposed study is correctly anticipated and an adequate sample size is determined, an ITT analysis will only provide evidence for the existence of efficacy. But, in most situations, we are also concerned about an unbiased estimate of the strength of treatment. For example, if we are interested in choosing among alternative treatments, we would obviously be interested in which treatments are more or less beneficial.[135] The downwardly biased estimate for efficacy which is provided by an ITT analysis (under non-adherence) prevents us from making an informed decision among alternative treatments.

When non-adherence is appreciable, some investigators have undertaken a comparison of participants according to treatment received. The two most

commonly used methods include 'as-treated' and 'per-protocol' analyses. Both of these methods are widely distrusted because they do not use the correct assumption of initial randomization and because their effect estimates and tests of the null are arbitrarily biased when adherence is non-random (e.g., when the baseline risk of the outcome is different for adherent and non-adherent participants).[138] However, as Sheiner and Rubin[135] and Korohonen et. al.[129] have pointed out, both of these methods correctly model the actual assignment mechanism, and, under additional assumptions, may yield valid estimates of efficacy for some subset of the study population. As such, both of these methods deserve closer examination.

Both the as-treated and per-protocol analyses group participants according to their received treatment, but in slightly different ways. The per-protocol analysis discards entire records of patients who go off treatment. The main problem with this approach is that the comparison groups are now based on post-randomization events affected by treatment and prognostic factors, including non-adherence.[136] One can imagine a situation in which participants with poorer prognosis are removed from the treatment group because they are less able to adhere to the assigned intervention. Under this typical scenario for a per-protocol analysis, the effect estimate will be biased away from the null (i.e., the intervention will appear to be more efficacious than it truly is). Because the per-protocol analysis is extremely sensitive to even small sources of bias[136], it should generally be avoided. However, it is worth considering what question the per-protocol analysis seeks to answer. Sheiner et al suggest that the per-protocol analysis answers the question: "What are

the differences between average outcomes among adherent participants in the treatment arm and average outcomes among adherent participants in the control arm?"[135] Rather than contrasting two groups that are deemed exchangeable (because of randomization), the per-protocol analysis contrasts the expected outcomes in two different populations. If one can imagine scenarios where non-adherence occurs at random (e.g., if all the non-adherence were attributable to pharmacy dispensing errors) then the per-protocol analysis addresses efficacy and the non-adherence simply decreases study power by eliminating some outcomes in the treatment group.[135] In typical randomized studies with non-adherence, however, it is difficult to justify random non-adherence. In such situations, the effect estimates from a per-protocol analysis will be biased and tests of the null may be non-zero even if there is no effect of treatment on the outcome.[136]

The as-treated analysis, in contrast to the per-protocol analysis, uses all of the study data by redefining non-adherent participants in the treatment arm as control participants. This approach attempts to deal with the loss of power seen in the per-protocol analysis, but at the cost of blurring the definition of adherence to treatment. In addition, it does nothing to clarify the confusion in inference that occurs because of the post-randomization redefinition of treatment groups. The as-treated analysis compares outcomes in two different populations: adherent participants receiving treatment and those who are in the control group because of assignment or because of non-adherence.[135] The as-treated approach can be misleading when outcome dependent non-adherence is present.[129] The direction of the bias will depend on the characteristic of non-adherent participants who are

reclassified as controls. If worse prognosis participants tend to be non-adherent then the bias will be away from the null; conversely, if 'switchers' are better prognosis participants, then the effect estimates will be attenuated. Korohonen et. al. provide simulation results showing that an as-treated analysis is valid when non-adherence occurs at random.[129] As we have pointed out earlier, however, this rarely occurs in real world settings. Consequently, scientists should be cautious when using an as-treated analysis to estimate efficacy.

We have seen that the three standard approaches to estimate efficacy are often biased under non-adherence because they rely on incorrect assumptions. The ITT analysis assumes that all individuals assigned to treatment are continuously exposed to the intervention during follow-up. The per-protocol and as-treated estimators avoid this pitfall by modeling actual assignment to treatment; however, they make a strong assumption that adherence is random. The ITT is preferred when estimating efficacy because of the apparent lack of an acceptable alternative approach to avoid selection bias.[134] We contend that valid alternatives to the ITT analysis do exist, but these alternative methods require an accurate measurement of adherence, of relevant concomitants, or both. In the following section we describe one of these methods, an instrumental variable (IV) approach, which estimates the magnitude of the treatment effect (efficacy) among adherent patients in a randomized study setting. This approach respects the original randomization and is valid even when adherent participants have a different baseline risk than non-adherent participants.[139]

NON-ADHERENCE CORRECTION USING INSTRUMENTAL VARIABLES

Introduction

Instrumental variable (IV) analysis has had limited application in the epidemiology literature because this technique requires the existence of one or more variables that are at least modestly associated with the treatment variable but have no direct effect on the outcome variable of interest.[137] Here, we describe a non-parametric IV method which can be applied to randomized studies with a dichotomous exposure and a dichotomous outcome. In the context of a randomized clinical trial, treatment assignment Z (i.e., randomization) can provide an almost perfect instrumental variable for confounder control. Z is a variable that, we hope, is highly correlated with the receipt of treatment (X) and it's only effect on the outcome, Y, is through X (i.e., there is no direct effect of Z on Y, or effects that are mediated by variables other than X).

The goal of the analysis is to compare the adherent participants in the intervention arm to an inferred control group chosen to eliminate bias.[134] The resulting estimate of efficacy in adherent participants has been variously called a local average treatments effect (LATE) or the complier-average causal effect (CACE) of treatment.[140] In the current analysis, we specifically seek to apply the IV method to the HIP PRO study to estimate efficacy in an adherent sub-group of the total study population. Because of the historical development of terminology in the

IV literature, we will refer to 'adherent' participants as 'compliant' participants for the duration of this discussion.

For the HIP PRO study, the directed acyclic graph (DAG)[127, 141-142] below shows a causal diagram that relates treatment assignment (Z) to treatment received (X) and to the outcome (Y). A variable (Z) is an instrumental variable for the causal effect of (X) on (Y) if its average effect on X is nonzero, it satisfies the exclusion restriction and the monotonicity assumption, is randomly assigned, and the stable unit treatment value assumption holds.[140] These assumptions are discussed in greater detail below:

Z (random treatment assignment)

U (Unmeasured Confounders)

X (actual wearing of hip protectors)

Y (hip fracture during follow-up)
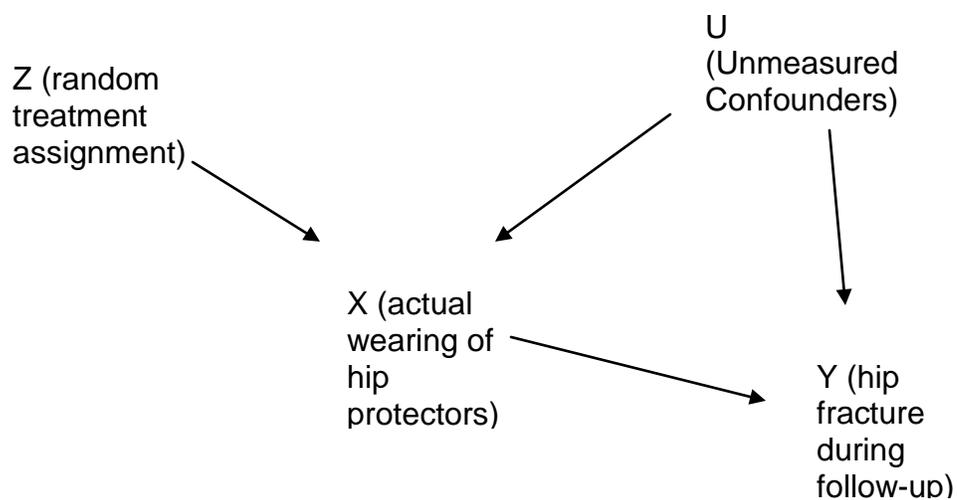
Figure 2.1 – Directed acyclic graph (DAG) relating treatment assignment (Z) to treatment received (X) and to the outcome (Y)

The above causal diagram allows us to conveniently discuss the assumptions of an IV analysis. First, we begin with the assumption that the potential outcomes for each participant *i*'s hip *j* are unrelated to the treatment status of other hips. This assumption, also known as the stable unit treatment value assumption (SUTVA) is

important for most epidemiologic analyses and it ensures that the outcomes in the analysis are independent. Because no individual could have more than one hip fracture in HIP PRO, we do not concern ourselves with clustering of hip fractures within individuals. And, because the intra-class correlation coefficient (ICC) between nursing homes was small, we posit that the impact of clustering within nursing homes was negligible and that the SUTVA assumption was approximately met in HIP PRO. The satisfaction of this assumption is important because the relatively straightforward IV analysis outlined below is not valid when SUTVA is violated.[140]

The second assumption for an IV analysis is that treatment assignment (Z) is randomized and that any effect of (Z) on the outcome (Y), must be through effect on treatment actually received (X) (see the causal diagram above). This assumption, also known as the exclusion restriction, has certain implications. First, this assumption implies that randomization (Z) is independent of any measured and unmeasured set of covariates (U) that might affect both the receipt of treatment (X) and the outcome (Y). Because treatment assignment is randomized we would expect that there are no confounders that would confound the (Z)-(Y) relationship. In practice, this is unverifiable from the data but is assured to be true (in expectation) as the sample size of the trial becomes large. The exclusion restriction also implies that the treatment that was actually received is substantially associated with randomization, but not fully determined by assignment Z. In the HIP PRO study this appears to be true because only hips randomized to treatment actually wore the hip protectors. Finally, the exclusion restriction implies that the assignment to treatment (Z) is independent of the outcome (Y), given treatment actually received (X) and (U).

This suggests that randomization (Z) has no direct effect on the outcome (Y), but, again, this cannot be verified from the data.[120-121]

The final two assumptions for an IV analysis relate to the composition of the study population. Under SUTVA and the exclusion restriction, the study population from a randomized study can be conceptualized as comprising four potentially latent class of participants (these latent classes have also been labeled as principal strata by some authors): (1) Would-be compliers are those individuals who would take whatever treatment is assigned, (2) Always-takers are those who would take the treatment no matter what their assignment, (3) Never-takers are those who would not take the treatment no matter what their assignment, and (4) Defiers are those who would take the treatment opposite from their assignment regardless of the treatment they were assigned.[143] Under this classification, an IV analysis assumes that there are at least some would-be compliers. Finally, we assume that there are no individuals who do the opposite of their treatment assignment. With these last two assumptions we are left with only compliant participants, always-takers, and never-takers, and so there are six possible combinations of participants, three in each treatment arm.

It should be noted here that even under all four of the assumptions outlined above, we are unable to identify specific compliant participants for whom we can identify the average treatment effect. We can only identify the local average treatment effect for 'would-be' compliers. This effect is not the average treatment effect for the entire study population or even a sub-population identifiable from observed values. Rather, this is the effect for a hypothetical population of

participants who would have complied with treatment if they had been assigned to treatment.[120]

If we are interested in linear relationships, then, under the conditions outlined above for an IV analysis, we can write the relationship between the randomization (Z) and the outcome (Y) as a product of the (Z)-(X) and the (X)-(Y) associations[121] $Assoc_{ZY} = Assoc_{ZX} * Assoc_{XY}$, and then solve this equation for the (X)-(Y) association. This association is important when considering non-adherence in the HIP PRO study – note that although the observed (X)-(Y) association can be confounded, the (Z)-(X) and the (Z)-(Y) associations cannot be confounded because (Z) is randomly assigned. As a result, we can obtain an unbiased estimate of the effect of (X) on (Y) with the following ratio:

$$E[Y \mid Z=1] - E[Y \mid Z=0] \quad / \quad E[X \mid Z=1] - E[X \mid Z=0]$$

This is the estimated effect of treatment received (i.e., $Assoc_{XY}$) and has sometimes been referred to as Bloom's IV estimator.[128] In essence, what we are doing here is taking the entire effect of (Z) on (Y) and factoring out the effect of (Z) on (X), to arrive at the average effect of (X) on (Y). Looked at another way, the denominator of the estimator above can be thought of as the intention-to-treat effect. If adherence is 100%, the denominator of the estimator will equal 1.0 and the effect of (X) on (Y) will equal the effect of Z on Y. However, as non-adherence increases, the denominator allows for the unbiased adjustment of the ITT effect.[121]

There are several alternative but equivalent strategies to estimate the effect of treatment received (X) on the outcome, hip fractures (Y). This effect estimate can be obtained by implementing a two-stage least squares (2SLS) algorithm (i.e., first regressing X on Z and then regressing Y on the predicted value of X obtained from the first regression).[128] An alternative strategy is a two-stage procedure described by Nagelkerke.[144] These authors also begin by regressing X on Z, but, rather than using the predicted values in the second stage, they use the residuals. In the second stage, a regression of Y on both Z and the residuals from the first stage will yield an unbiased estimate of the effect of treatment received.

Although the two methods described above are attractive options, we seek to produce a measure of effect that is comparable to the rate ratio obtained from the HIP PRO study. To that end, in the next section we implement a non-parametric IV method developed by Cain et. al. which estimates the rate instead of the risk of hip protectors.[145]

Methods and Results

For the IV analysis, we use the following notation. Subscript i indexes 1 to N=1042 participants in the first HIP PRO study, and subscript j indexes either the left (=0) or right (=1) hip. Adherence visits occurred three times a week, across all nursing shifts. For the purpose of the current analysis, and for consistency across subsequent analyses, we collapse the adherence visits into two week increments. Defined in this manner, we used subscript k to denote the 1 to $K_{i,j}$ two-week

adherence visits. The maximum number of visits was 54. Subscript *x* indexes the actual wearing of the hip protectors during a two-week period, where a protected hip was considered compliant when wearing the hip protector in more than 75% of adherence visits in that two-week time period. Once a protected hip became non-compliant in any two-week time period, then that hip was considered as non-compliant in all subsequent visits. Note that unprotected hips were always compliant (*x*=1) because no hip assigned to the control arm of the study wore hip protectors. Subscript *z* denotes the randomization arm, where 1 indicates assignment to the hip protector arm and 0 the control arm. $D_{ijkxz}=1$ indicates that hip *i,j* experienced a hip fracture between visits *k* and *k*+1 while using therapy *x* in the randomization arm *z*, 0 otherwise. T_{ijkxz} is the number of person-years that hip *i,j* contributed between visits *k*-1 and *k* while using therapy *x* under randomization arm *z*.

Under this notation, the total number of hip fractures occurring while using therapy *x* when randomized to treatment *z*, summed over hips and adherence visits, is:

$$D_{xz} = \sum_{i=1}^{1042} \sum_{j=1}^{2} \sum_{k=1}^{K_{i,j}} D_{ijkxz}$$

The total number of person-years contributed while using therapy *x* under randomized treatment *z*, again summed over hips and adherence visits, is:

$$T_{xz} = \sum_{i=1}^{1042} \sum_{j=1}^{2} \sum_{k=1}^{K_{i,j}} T_{ijkxz}$$

We estimate the conditional probability of using therapy x under randomized treatment arm z, $\alpha_{xz} = P(X=x|Z=z)$, by the proportion of person-time spent using therapy x under randomized treatment z, $\alpha_{xz} = T_{xz} / T_{+z}$, where $T_{+z} = \sum\limits_{x=0}^{1} T_{xz}$.

Figure 2.2 depicts a tree diagram which shows the division of hip fractures and person-time, by randomized treatment arm, actual use of hip protectors, and potential hip protector use in the HIP PRO study.[16] Our goal is to use the observed data in the second and third rows of the diagram to estimate, in the fourth row, the potential hip fractures and potential person-time that would have been observed in the alternate treatment arm. We proceed by using the notation above to describe the data in the second and third rows. There were a total of 38 hip fractures in HIP PRO during 1354 person-years of follow-up. A total of $D_{+1}=21$ hip fractures occurred during $T_{+1}=677$ person-years in the protected arm of the study, while $D_{+0}=17$ hip fractures occurred during $T_{+0}=677$ person-years in the control arm. In the treatment arm, there were $D_{11}=2$ hip fractures during $T_{11}=146.54$ compliant person-years; the remaining $D_{01}=19$ hip fractures occurred during $T_{01}=530.46$ non-compliant person-years. In the control arm hips were always compliant; as a result, the corresponding hip fractures and person-times were apportioned as follows: $D_{00}=17$, $T_{00}=677$, $D_{10}=0$, $T_{10}=0$. The conditional probability of using therapy $x=1$ under randomized treatment $z=1$, as noted previously, was estimated by the proportion of person-time spent using therapy $x=1$ (i.e., T_{11}) under randomized treatment $z=1$ (i.e., T_{+1}): $\alpha_{11} = 146.54 / 677 = 0.22$. In a similar manner, we calculated $\alpha_{01} = 0.78$, $\alpha_{00} = 1.00$, $\alpha_{01} = 0.00$.

The standard intention-to-treat analysis compares rates between the two treatment arms, without considering that overall adherence in HIP PRO was only 74%. The ITT estimator of the average effect of treatment can be calculated from Figure 2.2, according to the following formula:

$$\beta_{ITT} = \frac{\alpha_{11} \times \left(\dfrac{D_{11}}{T_{11}}\right) + \alpha_{01} \times \left(\dfrac{D_{01}}{T_{01}}\right)}{\alpha_{10} \times \left(\dfrac{D_{10}}{T_{10}}\right) + \alpha_{00} \times \left(\dfrac{D_{00}}{T_{00}}\right)} = \frac{\left(\dfrac{D_{+1}}{T_{+1}}\right)}{\left(\dfrac{D_{+0}}{T_{+0}}\right)} = \frac{21/677}{17/677} = 1.23$$

The 95% confidence interval for this point estimate is calculated in the usual way, using the standard error of the rate ratio:

$$s.e(\ln(RR)) = \sqrt{\frac{1}{21} - \frac{1}{677} + \frac{1}{17} - \frac{1}{677}} = 0.32$$

The 95% confidence for the ITT effect of treatment assignment was (0.65, 2.32).

Note again that the ITT effect estimate is biased for treatment received. What we are interested in is an IV estimator which compares rates between the two treatment arms among 'would-be compliers', those who would have used hip protectors if assigned to wear protectors and those who would not have used hip protectors if not assigned to wear them. The distinction between an 'observed complier' and a 'would-be' complier is an important one. A 'would-be complier' is an individual who would comply under both treatment regimes, while an 'observed complier' is compliant only under the treatment actually assigned. The crux of the problem is that we only observe the compliance status with the assigned treatment for each individual; we don't know the compliance status that would have occurred if assigned to the alternate treatment.[143] The appeal of the IV method is that it allows

70

us to estimate the compliance status in the 'would-be compliers.' Specifically, for the HIP PRO study, the IV approach that we use allows us to estimate the rate of hip fractures in the 'would-be compliers' in both treatment arms.

Next, we describe how to obtain the ITT effect estimate from data in Figure 2.2. The following heuristic explains why this approach works. The rate of hip fractures in the treatment group who use the treatment is an unbiased estimate of the rate of hip fractures in 'would-be compliers' and the 'always-takers.' The 'always-takers' are actually non-compliers because, if they had been assigned to the control arm, they would have worn hip protectors. Note that the rate of hip fractures in the control group who adopt the treatment is an unbiased estimate of the rate of hip fractures in these 'always-takers.' By using information in the control arm of the study for 'always-takers', we are able to factor out the effect of 'always-takers' in the treatment arm. The difference is an unbiased estimate of the rate of hip fractures in the 'would-be compliers.'[143]

Following Cain et al[145] we calculate the person-time in the fourth row of Figure 2.2 before calculating the number of events. We use the conditional probability of using therapy (i.e., α_{xz}) in the alternate treatment arm to determine the potential person-time in 'would-be compliers.' Consider the 146.54 person-years of those individuals who were randomized to wear hip protectors and actually wore them. We use the conditional probability of using therapy (α_{01}) and not using therapy (α_{00}) in the control arm of the study to calculate the potential person-time. Recall that there were no hips that were non-compliant in the control arm; as a result, $\alpha_{01}=0.00$ and $\alpha_{00}=1.00$. The potential person-time for the 'would-be

compliers' (i.e., the $X^{z=0}=0$ group) is then 146.54 = 1.00 * 146.54. This represents

the amount of person-time in the treatment arm of the study (for those who actually

used therapy) that would have been compliant person-time if assigned to the control

arm. The remaining person-time is apportioned to the 'always-takers' (the $X^{z=0}=1$

group). In the case of the HIP PRO study, this was 0 person-years. The person-

time for those who were non-compliant does not need to be partitioned and is simply

carried forward to the fourth row. This person-time (530.46 person-years) is

allocated to the 'never-takers' – the third hypothetical (potential) group in the

randomized arm of the study. We proceed in a similar manner to allocate person-

time to the 'would-be compliers' ($X^{z=1}=1$) as α_{11} * 677 PYs = 0.22 * 677 PYs =

146.54 PYs, in the 'never-takers' ($X^{z-1}=0$) as α_{10} * 677 PYs = 0.78 * 677 PYs =

530.46 PY. For the 'always-takers' the number of person-years are carried forward

to the fourth row.

To determine the number of hip fractures which should be allocated to each

of three groups of patients, we use the rate of hip fractures in the alternate treatment

arm. For example, to determine the total number of hip fractures allocated to the

'always-takers' group in the treatment arm of the study, we multiply the total person-

time occurring in compliant patients in the treatment arm by the rate (0 / 0 PYs) in

the non-compliant ' always-takers' in the control arm. This is equal to 0 = (0 / 0 PYs)

* 0 PYs. The remaining hip fractures (2-0=2) are allocated to the 'would-be

compliers.' In the treatment arm, the hip fractures occurring during non-compliant

person-time (i.e., 19 hip fractures) are not partitioned; they are all allocated to the

'never-takers.' Similarly, in the control arm, the number of hip fractures in the 'never-

takers' is equal to the rate of hip fractures in the 'never-takers' in the treatment arm multiplied by the person years in the 'never-takers.' Note that this is equal to 19 = (19 / 530.26 PYs) * (530.46 PYs). However, this is greater than the number of hip fracture in compliant hips in the control arm of the study. This is not possible, and, because we cannot determine the number of hip fractures for 'never-takers,' we also cannot determine the number of hip fractures in the 'would-be compliers.' Consequently, we cannot determine the IV estimate of the rate ratio, which is a comparison of the rate of hip fractures in 'would-be compliers' in the treated versus untreated groups. Note that the potential outcomes in the 'would-be compliers' in the control arm remained missing; as a result, the IV rate ratio was also missing because: RR = (2 / 146.55 PYs) / (? / 146.55 PYs) = ?

Discussion

One of the strengths of the HIP PRO study is that it collected detailed information on compliance during three weekly unannounced visits by compliance monitors. This is in contrast to many randomized studies where the non-compliance is measured with much error. When treatment received for study participants is not known with some level of accuracy, it can be difficult, if not impossible, to estimate treatment effects. Also, treatment effects can be difficult to estimate if study participants can be non-compliant because they take an intervention that is not under study. In the current IV analysis, we have an accurate measure of compliance in HIP PRO and we assumed that there were no interventions available other than

the hip protectors. By study design, study participants were allowed to switch treatments only in one direction, from hip protectors to the control (i.e., no treatment) arm.[144] In similar scenarios, the IV approach has been known, in other study settings, to provide a valid estimate of the effect of treatment received in the 'would-be compliers.'

In the HIP PRO study a simple IV method designed to estimate the rate ratio in 'would-be compliers' failed. This may have been because of the HIP PRO study design, which resulted in equivalent person-time being allocated to both treatment arms and which had no observed non-compliant individuals in the control arm. We attempted to obtain a rate ratio in the current analysis, to compare our results with the ITT rate ratio. It is not clear whether the IV approach is feasible in all situations, on the rate ratio scale; this requires further investigation. Beyond the fact that the IV approach failed to yield an effect estimate in the HIP PRO study, it should be noted that both the approach that we adopted and other standard IV approaches are poorly equipped to address this time-varying exposure.[144] We now turn our attention to another approach for non-adherence correction, using inverse probability-of-censoring weights, which account for the time-varying nature of exposure to hip protectors in HIP PRO.

Figure 2.2 Tree diagram showing division of hip fractures and person-time, by randomized treatment arm, actual use of hip protectors, and potential hip protector use, in the Hip Impact Protection Project (HIP PRO).
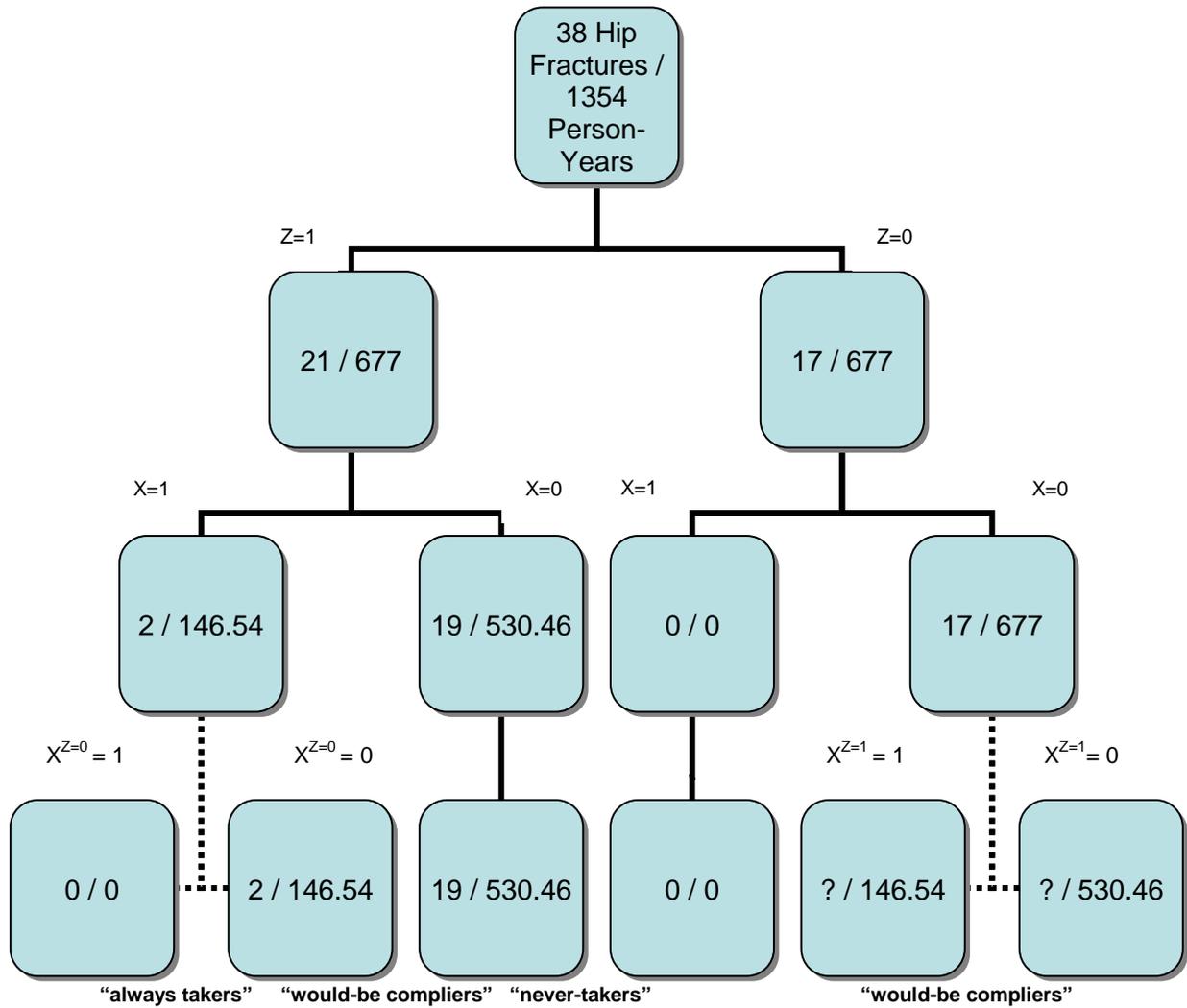
CHAPTER 3

EFFECT OF HIP PROTECTORS USING INVERSE PROBABILITY OF

CENSORING WEIGHTS (IPCWs)

Introduction

In the United States more than 300,000 people aged 65 years or older are

hospitalized as a result of hip fractures every year.[6] The one-year mortality rate for

individuals suffering a hip fracture can be as high as 33%[29, 34, 94] and up to 40% are

severely disabled or unable to walk independently two years after the event.

Effective interventions aimed at reducing hip fractures, such as bisphosphonates,

should prevent a substantial number of fractures.[146] In practice, however, adherence

to these agents has not been optimal. Alternative approaches to hip fracture

prevention, such as hip protectors, which fit over the trochanteric prominence, were

designed to attenuate the peak impact force entering the proximal femur during a fall

to the side, and, thus, address the proximate cause of hip fracture. As with other

interventions,[15, 30, 60] randomized studies of hip protectors have demonstrated mixed

results, possibly due in part to varying adherence.

In this paper, we describe the application of inverse probability of censoring

weights (IPCWs)[122, 132, 147-148] to correct for non-adherence in the Hip Impact

Protection PROject (HIP PRO), a multi-center, randomized, controlled clinical trial of hip protectors to prevent hip fractures.[16] The methods presented here are applicable to any intervention studies that suffer from less than complete adherence.

Methods

The Hip Impact Protection PROject (HIP PRO)

As described in greater detail previously, a total of 1042 residents from 37 nursing homes (NHs)[16, 125] were enrolled into the HIP PRO study between October 2002 and October 2004 and contributed 676 person-years of observation over a maximum of 24 months. HIP PRO used a cluster randomization strategy, whereby all residents in a given NH were designated to wear an energy-absorbing and distributing hip protector on either the left or right hip. A dynamic enrollment procedure, taking into account differences in NH size, was designed to achieve an approximately equal allocation of residents into each arm of the study. After randomization, the left hip was protected for 587 (56%) residents and the right hip was protected for 455 (44%) residents. Written informed consent was obtained from individual NH residents with a Short Blessed Test (SBT)[126] score less than 12 (no more than mild cognitive impairment) and from a designated responsible party otherwise. Each participating NH obtained Federal Wide Assurance from the Office of Human Research Protections, US Department of Health and Human Services. Institutional Review Board approval for the study was obtained by the three research

sites and the coordinating center. Resident safety was monitored every six months by an independent data and safety monitoring board.

At baseline and every six months after randomization, residents provided data on affect, cognitive, functional, and mobility status. Cognitively intact residents also provided information about fear of falling, perceived need for hip protectors, and body image. Research staff, on a weekly basis, reviewed resident charts, evaluated changes in resident behavior, and tracked calls made to a hip fracture hotline telephone number to evaluate falls and the presence of an incident hip fracture. Hip fractures were confirmed by a fracture adjudication committee masked to hip protector side. Research staff evaluated adherence in three unannounced visits each week, across all work shifts, including weekends. Residents who were non-adherent in more than two of six such visits during the first two weeks after randomization (the run-in period) were withdrawn from the study (N=148). Overall adherence for the study was 73.8%, with adherence defined as the number of visits during which a participating NH resident was found to be correctly wearing the garment and protector divided by the number of research staff visits to the resident. For the current analysis, a resident was censored after the first instance in which her adherence was less than 75% in any two-week time period after randomization. Sensitivity analyses substituted values of 67% and 50% for the value of 75% in the definition of non-adherence.

Statistical Analysis

IPCWs were calculated by modeling the probability of ceasing to comply with the protocol-dictated wearing of hip protectors, given the subject's observed data up to that point in time. Weighting the ITT analysis with IPCWs corrected for the variability in adherence over time.

To calculate the IPCWs, we used the following nomenclature. Let subscript i index 1 to N=1042 patients and let j denote an individual's left (=0) or right (=1) hip. Further, because we measured time in two-week increments from the start of follow-up, let k index 1 to a maximum of k=52 two-week periods after randomization. Let $X_{ij}=1$ indicate that hip j for a given NH resident was randomized to wear a hip protector, and 0 otherwise. Let D_{ijk} be an indicator of first hip fracture in the ith NH resident's hip j, during time period k, so that $D_{ijk}=1$ if a NH resident's hip j experienced a hip fracture during time period k, and 0 otherwise. Let $C_{ik}=1$ denote that a NH resident stopped wearing a hip protector and was censored during time period k, and 0 otherwise. Let V_{i0} denote a set of individual-level and NH-level covariates measured at baseline. For the present analysis, V_{i0} consisted of resident age, race (white/other), marital status (never married, married, widowed, divorced/separated), osteoporosis history (yes/no), cognitive status (based on the SBT[126]: normal/minimal impairment (SBT=0-8), moderate impairment (SBT=9-19), severe impairment (SBT≥20)), NH size, NH for-profit status, and an indicator of whether the NH allowed incontinent residents to wear hip protectors at night. In addition to all individual and NH baseline characteristics, let V_{ik} also include the

individual-level, dichotomous, time-varying covariates bowel incontinence, bladder incontinence, cane use, and indicator variables for whether resident resisted care, was verbally abusive, displayed disruptive behavior, or had a fall. These time-varying covariates were assessed every six months and missing values were filled in with previous non-missing values, if available. The time-invariant and time-varying variables were chosen based on *a priori* knowledge about the common causes of hip fractures and adherence, as well as on a concurrent analysis of the predictors of adherence in the HIP PRO study.[125]

For protected hips the weights were defined as follows:

$$SW_{ijk} = \prod_{m=0}^{k} P[C_{ijm}=0|\bar{C}_{ijm}=0,X_{ij}=1] / P[C_{ijm}=0|\bar{C}_{ijm}=0,X_{ij}=1,V_{im-1}]$$

where V_{im} represented a collection of covariates thought to be common causes of non-adherence and hip fractures based on *a priori* knowledge. Informally, SW_{ijk} represented the ratio of a subject's probability of remaining uncensored up to time period k, calculated as if there were no time-fixed or time-varying determinants of censoring, divided by the subject's conditional probability of remaining uncensored up to time period k.[149] The true weights were unknown, but we estimated them from the observed data by fitting separate parametric regression models for the numerator and the denominator.[150] For individuals censored due to non-adherence (i.e., $C_{ij}=1$), zero weights were given to all follow-up time on or after censoring. The stabilized weights SW_{ijk} for unprotected hips were set to 1 for all individuals, at all

time points, because these hips were never exposed to hip protectors and therefore were always adherent with the assignment to be an unprotected hip.

The conditional probabilities for the numerator and denominator of the stabilized weights were fit using pooled logistic regression models for the discrete-time hazard of censoring. Pooled logistic regression approximated the Cox model when the risk of hip fracture was small in any given person-time interval (in this analysis, the risk of hip fracture in any time period k was always less than one percent).[151] The models for the numerator and the denominator of the weights, respectively, were of the form:

$$\text{logit } P(C_{ik}=0|\bar{C}_{ijk}=0,X_{ij}=1) = \alpha_{0k} + \alpha_1 X_{ij},$$

and

$$\text{logit } P(C_{ik}=0|\bar{C}_{ijk}=0,X_{ij},V_{ik}) = \beta_{0k} + \beta_1 X_{ij} + \beta_2 V_{ik},$$

where α_{0k} and β_{0k} represented terms for the time specific intercepts (which we modeled as restricted cubic splines,[152] with four knots at the 5th, 35th, 65th and 95th percentiles for the number of time periods since the baseline visit), and β_2 was the transpose of the column vector of log hazard ratios for the components of the covariate history matrix V_{ik}.

We considered three alternative censoring mechanisms. In the first mechanism C_{ij}^1, we censored NH residents in any time period after their adherence

was less than 75%. This cut-point was less than full adherence but was felt to be appropriate because the overall adherence for the HIP PRO study was 74%. In practice, because there were on average 6 adherence visits in any given time period, individuals were censored when their adherence was less than 5 out of 6 visits (83%). The second mechanism C_{ij}^2 censored individuals after their adherence first became less than 67% in any given time period (in general, adherent at 3 or fewer out of 6 adherence visits). The final censoring mechanism C_{ij}^3 censored individuals after their adherence first became less than 50% in any given time period (in general, adherent at 2 or fewer of 6 visits).

We considered four alternative sets of time-invariant and time-varying covariates for inclusion in the denominator model of the stabilized weights. In the main analysis and in the three sensitivity analyses, we included the time-fixed covariate of baseline age, race, marital status, osteoporosis history, and cognitive status, as well as the following time-varying dichotomous variables which were measured every six months: bowel incontinence, bladder incontinence, cane use, resisting care, verbally abusive, disruptive behavior, and recent fall. Covariates at the NH level were nursing home size, for-profit status, and a dichotomous variable indicating whether NH allowed incontinent residents to wear hip protectors at night. In sensitivity analyses, we separately considered adding body mass index (BMI) and educational status, because these covariates were considered potential common causes of non-adherence and hip fractures. In addition, in the third sensitivity analysis, we considered adding both BMI and educational status to the denominator model of the stabilized weights.

Educational status, a time-invariant variable that was measured at baseline, was missing for 26% of participating NH residents. We constructed inverse probability weights for missing education in a fashion similar to that used for constructing IPCWs. Specifically, we defined stabilized weights to correct for missing education as follows:

$$\text{SW}_{ijk} = \prod_{m=0}^{k} P[M_{ijm}=0|X_{ij}] / P[M_{ijm}=0|X_{ij},Z_{im}] \, ,$$

where $M_{ijm}=0$ denoted that educational status was not missing for subject i's hip j at time k, X_{ij} indicated that a hip was protected or not, and Z_{ik} represented the set of known characteristics that theoretically explain the 'missingness' of education. This set included all the individual-level covariates specified earlier, excluding osteoporosis history and cane use. In sensitivity analysis, where we included educational status as a common cause of non-adherence and hip fractures, we multiplied censoring weights by the weights for missing education and ran the primary analysis in the subset of individual time periods where the resident was continuously adherent and had non-missing educational status. For all other covariates with less than 3% missing data, we replaced the missing data with the mode from the non-missing distribution of that variable.

We fit weighted pooled logistic models for the risk of hip fractures, for all different censoring mechanisms and sets of covariates. These models had the form:

$$\text{logit } P(D_{ijk}=1|\overline{D}_{ijk-1}) = \beta_{0j} + \beta_1 \, X_{ijk,}$$

where $D_{ijk}=1$ indicated that subject i incurred a hip fracture in hip j, during time period k, β_{0j} again represented time specific intercepts, and β_1 represented the effect of hip protectors. Weighting each subject's contribution by the stabilized weights created a pseudo-population that accounts for non-adherence. β_1 provided an estimate of the log causal rate ratio, comparing the rate of hip fractures in protected hips, had, possibly contrary to fact, all protected hips remained adherent, to the rate of hip fractures in unprotected hips. Confidence intervals for the inverse probability of censoring weighted estimators were based on conservative robust variance estimates.[153] All analyses were conducted using SAS version 9.1 software (SAS Institute, Inc., Cary, North Carolina).

Results

The majority of the 1042 men and women making up the study population were composed of older white females with severe cognitive impairment. In addition, most of the NH residents were widowed at baseline, while one-fifth had never been married. A minority of the residents had experienced a fall in the 30 days before the baseline visit, and a smaller proportion had a history of osteoporosis at baseline (Table 3.1). The NHs participating in the study were equally likely to be for-profit and not-for-profit, and most of the NHs allowed residents to wear hip protectors some or all of the time at night (Table 3.2).

After 676 person-years of follow-up, out of a total of 1042 enrolled NH residents, 38 (4%) experienced the primary outcome; 21 hip fractures occurred in protected hips and 17 hip fractures occurred in unprotected hips. The HIP PRO study was terminated early based on a recommendation from the data and safety monitoring board, due to a lack of efficacy and feasibility to detect a protective effect. The ITT analysis produced a matched rate ratio of 1.24 (95% confidence interval (C.I.) of 0.65 to 2.34), when comparing hips that were randomly assigned to be protected with the contra-lateral hips that were not protected. Adjusting for clustering within NHs did not meaningfully change the results (RR=1.24 (95% C.I.: 0.64, 2.37).[16] Because NH residents were withdrawn after their first hip fracture, consideration of clustering of observations within individuals was ignored. [154]

We replicated the results from the original published ITT analysis using pooled logistic regression and estimated a hip fracture incidence rate ratio of 1.24 (95% C.I.: 0.65, 2.34), when comparing protected hips to unprotected hips.[16] We present the results of our replication of the ITT analysis and three inverse probability-of-censoring weighted analyses in Table 3.3.

When we used the first censoring mechanism c_{ij}^1 to account for non-adherence, in which we censored individuals the first time their adherence fell below 75% in any given two-week period, and used all covariates previously specified, we found that the adherence corrected incidence rate ratio estimate was 0.55 (95% C.I.: 0.13, 2.40). More relaxed censoring mechanisms, where individuals were censored when adherence fell below 67% and 50%, also estimated a similar protective effect

of hip protectors in this study. These results suggest that the overall non-adherence in the study biased the ITT rate ratio.

In sensitivity analysis, we also included the individual-level, time-fixed covariate educational status in the set of V_{ij} covariates when calculating stabilized weights. We included this covariate based on *a priori* knowledge that education was related to adherence and hypothesized that higher education status would have a direct negative effect on hip fracture incidence. In the HIP PRO study population, educational status was related to adherence but was only weakly related to the clinical endpoint. Inclusion of educational status resulted in incidence rate ratios that were similar to weighted models that did not include educational status, but had less precision. With education added to the model, the estimated rate ratio was 0.62 (95% C.I.: 0.14, 2.81) under censoring mechanism C_{ij}^1, 0.61(95% C.I.: 0.13, 2.74) under C_{ij}^2, and 0.51 (95% C.I., 0.16, 1.63) under C_{ij}^3. When we separately included BMI in the set of V_{ij} covariates we found no meaningful difference when compared to models that did not include BMI. With BMI added to the model, the estimated rate ratio was 0.55 (95% C.I.: 0.13, 2.41) under C_{ij}^1, the rate ratio was 0.52(95% C.I.: 0.12, 2.30) under C_{ij}^2, and was 0.72 (95% C.I., 0.28, 1.85) under C_{ij}^3. When we added both education and BMI to the models used to calculate the stabilized weights, the estimated rate ratios were similar to those obtained from adding education alone, but the confidence intervals were wider.

Discussion

Using data from the HIP PRO study, we found substantial differences in results between the ITT analysis and analyses correcting for non-adherence with IPCWs. The ITT analysis produced a relatively more precise estimate of the rate ratio in the direction of an adverse effect of hip protectors. The results from the analysis employing IPCWs produced a much less precise estimate in the direction of a beneficial effect.

Some have suggested that the HIP PRO study design, in which only one hip was protected, allowed residents to fall preferentially onto the protected hip. However, it is unlikely that frail, elderly NH residents (71% with severe cognitive impairment) would have had the physical or mental ability to alter the direction of a fall. Rather, the ITT results are biased toward and possibly beyond the null under by non-adherence. The imprecise estimate provided by the inverse probability of censoring weighted approach, which corrects for non-adherence, supports the hypothesis that a biomechanically tested hip protector, if worn correctly, may reduce the incidence of hip fractures.

Recent meta-analyses of hip protector clinical trials, including only high-quality randomized studies, reported reductions of hip fracture incidence in nursing home residents only, but even the aggregated results were imprecise.[79, 109] The more restrictive meta-analysis by Sawka and colleagues of NH residents included only three randomized studies in institutional settings and reported a RR=0.56 (95% confidence interval: 0.31, 1.01), comparing hip fractures in individuals wearing hip

87

protectors and hip fractures in individuals not wearing hip protectors.[79] In contrast,

the meta-analysis by Parker and colleagues examined cluster randomized studies of

hip protectors[83, 88-89, 110, 118, 152], and produced an overall RR=0.77 (95% confidence

interval: 0.62, 0.97).[109] Although defined and measured in a variety of ways, the

overall non-adherence of the studies included in the meta-analyses was substantial,

ranging from 30% to 80%. Our adherence corrected results were similar to those

obtained by Sawka. However, it should be noted that correction for non-adherence

in the studies included by Sawka would have likely produced even stronger effect

estimates than was noted in the meta-analysis.

As with most modeling techniques, we make certain assumptions when

employing IPCWs to correct for non-adherence. First, in constructing IPCWs we

make a strong, un-testable assumption that there are no unmeasured confounders

of adherence and hip fractures (this is also known as the exchangeability

assumption). The HIP PRO study collected a large number of individual-level and

NH-level covariates that are risk factors for both non-adherence and for hip

fractures.[125] If these variables represent a close approximation of the full set of joint

determinants of non-adherence and hip fractures then our final estimated effect

estimate will be a valid approximation of the true effect estimate.

Personal perceived health status and perceived need for hip protectors (e.g.,

because of a fear of falling) have been shown to be predictors of both adherence

and hip fractures.[84] Covariates that captured information on perceived health status

and perceived need for hip protectors were collected in the HIP PRO study, but only

for cognitively intact residents. These covariates were not included in our final

model because the HIP PRO study population was overwhelmingly cognitively impaired; however, they should be included when correcting for non-adherence in hip protector studies of community-dwelling, cognitively intact individuals. In sensitivity analyses, we included education in our final model based on the theory that education would be a common cause of non-adherence and hip fractures. In the HIP PRO data, however, education was predictive of non-adherence, but only weakly related to hip fractures. Including variables that are related to non-adherence but unrelated to the outcome results in an increase of the variance of the final effect estimate, without decreasing bias.[155] In contrast, including variables that are related to the outcome, but only weakly related to non-adherence, decreases the variance of the final effect estimate, without increasing bias. Impaired vision has been shown to be related to hip fracture incidence. Vision status was not collected for a majority of NH residents in the HIP PRO data. If we assume that impaired vision is only weakly related to non-adherence, then exclusion of vision status from our final model would result in a less precise effect estimate. Alternatively, if vision status is a strong confounder of non-adherence and the outcome, our final effect estimate would be biased because the exchangeability assumption would be violated.

Second, we assume that the consistency assumption has been met.[156-158] This assumption requires that it should be clear how a certain level of exposure could hypothetically be assigned to a person actually exposed to a different level of exposure. In randomized studies consistency is guaranteed, and, for the current analysis, we contend that the consistency assumption has been met since we have

a clear causal contrast, comparing hip fractures in hips that were protected against hips that were not protected.

Finally, we assume that data on baseline educational status, race, history of osteoporosis, history of fall in last 30 days, marital status, and cognitive impairment are missing at random. We used two approaches to deal with missing data. For covariates with a low percent missing (<3%) we replaced the missing data with the mode of the non-missing data. For one potential common cause of non-adherence and hip fractures with a high percentage (26%) of missing data, education status, we used inverse probability of missing weights to account for missing data.[159] This approach to dealing with missing data was employed because it is consistent with the methods used to construct IPCWs for non-adherence.

Adherence has been a problem in randomized studies of hip protectors in NHs despite efforts to improve adherence, such as employing staff to encourage residents to wear hip protectors[96]. In the HIP PRO study, with arguably the most intense efforts to improve adherence, the overall adherence was only 74%. It is unlikely that future randomized studies of hip protectors will achieve overall adherence much greater than 74%. The ITT analysis of these future, hypothetical studies will compare those assigned to one treatment against those assigned to another treatment, without regard to treatment actually received, and incorrectly interpret the ITT results as the actual treatment effect. Under non-adherence, this ITT analysis will be null-biased unless researchers obtain data on all important common causes of non-adherence and hip fractures and employ methods to correct for the non-adherence. Anticipation of using methods such as IPCWs to correct for

non-adherence should also encourage investigators to increase the sample size requirements, when conceptualizing the design of studies.

To our knowledge, the current analysis is the first application of IPCWs to correct for non-adherence in a gerontologic setting; however, the methods presented here can be widely employed to correct for non-adherence in any intervention study. In particular, the use of IPCWs could be valuable to data and safety monitoring boards when they are faced with difficult decisions regarding the termination of trials with sub-optimal adherence to an intervention like hip protectors.

Tables and Figures

Table 3.1. Selected Baseline Characteristics of 1042 Men and Women Randomized
to Wear Hip Protectors on Either the Left Hip or the Right Hip in the Hip Impact
Protection Project (HIP PRO).

Characteristic	Nursing Home Residents (N=1042)[b]
Women	821 (78.8)
White[a]	895 (85.9)
Age, Mean (SD), year	85.3 (7.4)
History of osteoporosis[a]	202 (19.4)
History of fall in last 30 days[a]	301 (29.2)
Marital status[a]	
Never	193 (18.5)
Married	153 (14.7)
Widowed	631 (60.6)
Divorced/Separated	65 (6.2)
Cognitive status[a,c]	
Normal / Mild impairment	151 (14.5)
Moderate impairment	147 (14.1)
Severe impairment	744 (71.4)

[a]Missing data (<3% of total residents) replaced with mode of distribution.

[b]Represents total N(%), unless otherwise noted

[c]Cognitive status categories based on Short Blessed Test (SBT): normal / minimal
impairment (SBT = 0 – 8), moderate impairment (SBT = 9 – 19), severe impairment
(SBT ≥ 20).

Table 3.2. Nursing Home Characteristics of 37 Nursing Homes Participating in the

Hip Impact Protection Project (HIP PRO).

Characteristic	Nursing Home (N=37)[a]
For Profit	18 (48.7)
Total Number of Beds	
< 100 Beds	4 (10.8)
100 – 149 Beds	13 (35.1)
150 – 199 Beds	12 (32.4)
>= 200 Beds	8 (21.6)
Nursing Home Allows Hip Protectors to be	
None	6 (16.2)
Some	16 (43.2)
All	15 (40.5)

[a]Represents total N(%).

Table 3.3. Intent-to-treat (ITT) and Inverse Probability of Censoring Weighted (IPCW) Analysis of 1042 Men and Women Randomized to Wear Hip Protectors on Either the Left Hip or the Right Hip in the Hip Impact Protection Project (HIP PRO).

Analysis	Hip Status	No. Hip Fractures	Person-Years (PYs)	Rate / 100 PYs	Rate Ratio	95% Confidence Interval	Confidence Limit Ratio
ITT	Protected	21	676	3.11	1.24	0.65, 2.34	3.6
	Unprotected	17	676	2.52	1	-	
IPCW1[a]	Protected	2.1	146.5	1.41	0.55	0.13, 2.40	18.5
	Unprotected	17	676	2.52	1	-	
IPCW2[b]	Protected	2.3	159.8	1.41	0.52	0.12, 2.29	19.1
	Unprotected	17	676	2.52	1	-	
IPCW3[c]	Protected	5.6	331.3	1.69	0.67	0.26, 1.73	6.7
	Unprotected	17	676	2.52	1	-	

[a]Individuals censored in first time period in which adherence is less than 75%.

[b]Individuals censored in first time period in which adherence is less than 67%.

[c]Individuals censored in first time period in which adherence is less than 50%.

CHAPTER 4

EFFECT OF HIP PROTECTORS ON FRACTURES USING A STRUCTURAL

NESTED MODEL

Introduction

Randomized studies of hip protectors have been plagued by an issue that

affects many intervention studies, non-adherence with assigned therapy.[88, 91, 93, 95-96,

100, 112] These randomized studies have often been evaluated with intention-to-treat

(ITT) analyses, which compare patients based on their assignment to treatment,

regardless of their actual exposure to the intervention of interest. ITT analysis has

qualities that make it appealing to scientists. First, under full adherence and no drop-

out, the ITT analysis provides a valid test of the null hypothesis that the treatment is

ineffective and yields an unbiased estimate of a causal effect (i.e., a contrast in

potential outcomes).[127] Second, under non-adherence, an ITT analysis is

conservative in the sense that it still provides a valid test of the null hypothesis that

the treatment is ineffective, albeit at reduced power (which is a function of the

amount of non-adherence). However, by ignoring information on post-randomization

non-adherence, ITT analyses estimate the effectiveness of treatment in a mixed

population comprised of both compliers and non-compliers and, therefore, will

under-estimate the etiologic effect of treatment to the extent that the study population is comprised of non-compliers.[138]

Because of the problems inherent to ITT analyses of studies with non-trivial amounts of non-adherence, there have been calls to supplement the ITT effect estimate with adherence corrected effect estimates.[136, 160] Structural nested models preserve the validity of the test of the null hypothesis and involve direct treatment group comparisons, with randomization used as an instrument for treatment.[129, 136, 161-162] In the current analysis, we employ this method to correct for non-adherence in the Hip Impact Protection PROject (HIP PRO).[16]

Methods

Study Population

HIP PRO was a multi-center, randomized, controlled clinical trial which evaluated the effect of hip protectors on hip fractures.[16, 125] Randomization occurred at the nursing home (henceforth, home) level, with all residents of a home designated to wear the hip protector on either the right or left hip. Between October 2002 and October 2004, 1042 residents from 37 homes contributed 676 person-years of observation. Participants were seen weekly in homes that were in close proximity to three clinical coordinating centers, which were located in Boston, Massachusetts, St. Louis, Missouri, and Baltimore, Maryland. At baseline and every six months after randomization, residents provided data on affect, cognitive,

functional and mobility status. Institutional review boards approved the study protocol, and written informed consent was obtained for all residents. Resident safety was monitored by an independent data and safety monitoring board every six months. Each participating home obtained Federal Wide Assurance from the Office of Human Research Protection, US Department of Health and Human Services.

The primary study endpoint in HIP PRO was a hip fracture, which was identified by weekly chart reviews, changes in resident mobility, and calls made to a hip fracture hotline telephone number. All hip fractures were confirmed by an independent fracture adjudication committee. Research staff evaluated the use of undergarments that were specifically designed for HIP PRO, as well as the correct positioning of the hip protector, in three weekly unannounced visits. These visits occurred across all work shifts and during weekends. Total adherence in HIP PRO was based on these visits and was defined as the number of visits during which residents were found to be correctly wearing the undergarment and the hip protector, divided by the total number of research staff visits. For the purpose of the current analysis, we followed residents in two week increments from the start of follow-up until they experienced a hip fracture, were lost to follow-up, or were administratively censored at the end of the study. For each two week period, we defined a resident as being adherent if she was correctly wearing the garment in at least 75% of research staff visits; otherwise, the resident was considered non-adherent.

Statistical Analysis

Individuals were indexed by i and took values from $i = 1$ to 1042, while j indexed hips and took values of 0 and 1. We used capital letters to represent random variables and lower letters to represent possible realizations of random variables. The treatment assignment indicator was denoted by $R_{i,j}$ and took the value of 1 if an individual hip was randomized to wear the hip protector and 0, otherwise. $T_{i,j}$ represented the time from randomization to hip fracture. The time on active treatment (adherent time) was denoted by $D_{i,j}$. In the placebo arm, $D_{i,j} = 0$ for all i, j because hips in the placebo arm were never exposed to hip protectors.

The "potential failure time" was essential to our use of a structural nested model. The potential failure time for an observed participant was the time from randomization to hip fracture under a given treatment plan. One such potential failure time for each participant was the failure time that would have occurred if the participant had been unexposed throughout follow-up, say $U_{i,j}$. This potential failure time was considered a pre-randomization variable that existed for all participants but which was only partially observed. Indeed, we only observed the potential failure time for the treatment plan that the participant actually took (and that only under the consistency assumption and if the participant was not censored). To identify the point estimate for the effect of treatment, g-estimation of a structural nested failure time model leveraged a key assumption about the relationship between these potential failure times and randomization. The key assumption was that the potential

failure times $U_{i,j}$ were independent of the treatment arm indicator $R_{i,j}$. Under this assumption, using a model that linked observed treatment and failure times with potential failure times, g-estimation identified the parameter value (Ψ) at which the distribution of potential failure time given no treatment in the experimental arm was equal to the distribution of potential failure time given no treatment in the placebo arm. This parameter value was an estimate of the effect of treatment.

To link the potential failure time $U_{i,j}$, which was only observed for those never treated, with the observed treatment history and failure time, we used the accelerated failure time (AFT) model of Cox and Oakes.[163-164] We defer the complication of censoring. The AFT model, a linear model for log failure times, allowed us to estimate the effect of treatment by parameterizing a relationship between a partially observed baseline quantity and observed event times, thereby allowing inference based on randomization. The measure of association in a structural nested failure time model was the survival time ratio. We estimated the survival time ratio by comparing the potential failure time corresponding to always exposed with the potential failure time corresponding to never exposed.[165-166]

For individual hips that experienced a hip fracture, we used their observed failure time to determine their potential failure time by using the following strong version of the structural AFT model[129, 167]:

$$U_{i,j}(\Psi^*) = \int_0^{T_{i,j}} \exp[\Psi^* \times R_{i,j}(t)] \, dt$$

This deterministic AFT model assumed that, for each hip, the potential failure time under no treatment, $U_{i,j}$, may be computed as a function of the randomization indicator, $R_{i,j}$, the observed failure time $T_{i,j}$, and a scalar parameter Ψ^*. Ψ^* was an unknown parameter to be estimated, as described above. The structural AFT model quantified how treatment-free survival time was contracted or extended by a factor $\exp(-\Psi^*)$, the survival time ratio. A positive value of Ψ^* would indicate that constant exposure to hip protectors would decrease the time to hip fractures by $\exp(-\Psi^*)$ (i.e., treatment is harmful), when compared to the unexposed, while a negative value of Ψ^* would denote that constant exposure to hip protectors would increase the time to hip fractures (i.e., treatment is beneficial). When $\Psi^*=0$, there is no effect of treatment on survival time.

We estimated Ψ^* with g-estimation, a test-based procedure. This procedure checked for an association between $R_{i,j}$ and a hypothesized value $U_{i,j}(\Psi)$ of the true but unknown potential treatment-free failure time.[123, 168] Note that we were able to compute $U_{i,j}(\Psi)$ for a range of hypothesized, plausible values of Ψ (e.g., -3 < Ψ < 3, by increments of 0.02) using the observed data $\{T, R\}_{i,j}$ in the AFT model. Using a finer search grid, say in increments of 0.01 or 0.001, would be computationally intensive but yield more precise effect estimates. As noted previously, because of randomization, we assumed that the transformed failure time $U_{i,j}(\Psi)$ was independent of the randomization indicator, $R_{i,j}$, when $\Psi = \Psi^*$. This randomization assumption allowed us to evaluate whether a particular value of Ψ equaled Ψ^* by

testing β=0 in the following proportional hazards model, where time was the potential failure time, using score tests:

$$h_{U_{i,j}(\Psi)} = h_0(t)\exp(\beta * R_{i,j})$$

These tests were similar to the standard ITT test of the effect of treatment because they preserved the original randomized group assignment. The estimate $\hat{\Psi}$ of Ψ^*, obtained when β=0, was the value of Ψ for which the distribution of $U_{i,j}(\Psi)$ in the treated arm was approximately equal to the distribution $U_{i,j}(\Psi)$ in the control arm. The test-based 95% confidence interval for $\hat{\Psi}$ was {Ψ :Z(Ψ)<1.96}.

The strong version of the AFT model outlined above would only be relevant for studies in which all patients are followed until they experience the outcome. When using this version of the AFT model in the HIP PRO study, potential failure times could only be determined for individuals who experienced a hip fracture. To accommodate the balance of patients who were censored, we sought to modify the parameters of the AFT model above. There were two types of censoring in the current analysis, censoring by drop-out and censoring due to end of follow-up (administrative censoring). Initially, we considered only censoring by the planned end of study, where 440 (42%) individuals were administratively censored and experienced no hip fractures by the end of follow-up. To compute a potential time for these individuals, we defined the potential censoring time at end of follow-up, $C_{i,j}$, as the time from randomization to the end of follow-up. Then, we replaced $U_{i,j}(\Psi)$,

which was unobserved for individuals who did not experience a hip fracture, with a function of $U_{i,j}(\Psi)$ and $C_{i,j}$ that was observed for all individuals.[138] Note that because $U_{i,j}$ was independent of $R_{i,j}$, the function of $U_{i,j}(\Psi)$ and $C_{i,j}$ was also independent of $R_{i,j}$ since the censoring time $C_{i,j}$ was a baseline, pre-randomization covariate. Let

$\Delta_{i,j}(\Psi) = 1$ if $U_{i,j} < C_{i,j}(\Psi)$, and

$\Delta_{i,j}(\Psi) = 0$ if $U_{i,j} \geq C_{i,j}(\Psi)$,

where $C_{i,j}(\Psi) = C_{i,j}$ if $\Psi \geq 0$ and $C_{i,j}(\Psi) = C_{i,j} * \exp(\Psi)$ if $\Psi < 0$. With this definition of $C_{i,j}(\Psi)$, we obtain that when a participant was censored ($T_{i,j} > C_{i,j}$), $U_{i,j} = C_{i,j}(\Psi)$, so that $\Delta_{i,j}(\Psi)$ was always observed. In the Cox model defined above, instead of using $\{T, R\}_{i,j}$ to fit the model, we fit the model using $*T(\Psi) = \min(U_{i,j}, C_{i,j})$, the endpoint indicator $\Delta_{i,j}(\Psi)$, and the treatment indicator $R_{i,j}$.[123-124, 138, 168]

In the HIP PRO study 564 (54%) study participants withdrew consent, died, or were lost to follow-up before experiencing a hip fracture or being administratively censored at the planned end of follow-up. While censoring due to administrative reasons can be viewed as being non-informative (because the date of study completion is a pre-randomization variable), censoring resulting from drop-out cannot be assumed to be completely at random. For such drop-out, post-treatment dependencies between censoring and outcome could arise. The methods described

earlier to deal with administrative censoring could not be used to deal with censoring by drop-out; an alternative strategy was required.

We used inverse probability weights, which are similar to the Horvitz-Thompson device in sampling theory, to account for the potential selection bias introduced by non-administrative censoring.[169] Under this approach, we assumed that censoring due to drop-out was independent of the time to hip fracture, conditional on treatment status, age (< 80 years, ≥ 80 years), race (white/other), gender, hip fracture history (yes/no), marital status (never married, married, widowed, divorced/separated), osteoporosis history (yes/no), and cognitive status (based on the Short Blessed Test (SBT)[126]: normal/minimal impairment (SBT=0-8), moderate impairment (SBT=9-19), severe impairment (SBT≥20)). We specified a survival model, conditional on these covariates, to estimate a patient-specific weight w_i that equals the inverse of the probability that the participant remained uncensored until the occurrence of hip fracture or the administrative end of follow-up, whichever came first. For individuals who dropped out of the study, the weights were set to zero, $w_i = 0$. The individuals with non-zero weights created a pseudo-population in which censoring due to drop-out was ignorable given the set of covariates, and the final g-estimation was applied to this pseudo-population. To account for the fact that we used estimates of the weights, we used robust variance estimates in our Cox model.[170]

The g-estimation procedure estimated the survival time ratio, which described the association between exposure and survival using the accelerated failure time parameterization. Some have argued that such measures of change in survival are

often of greater public health interest than hazard ratios.[124, 171-172] For the current analysis, however, we sought to compare the ITT hazard ratio with a hazard ratio obtained from g-estimation. Therefore, we used a Weibull model (which can be expressed in either parameterization) to express the survival time ratio as a hazard ratio. We assumed that the underlying potential survival times followed a Weibull distribution and used the Weibull shape parameter $\hat{\kappa}$ (estimated using maximum likelihood) to calculate the desired hazard ratio as follows: hazard ratio = exp($\hat{\kappa}$ * $\hat{\Psi}$).

We replicated the ITT analysis from the main HIP PRO study[16] using a Cox model to compare hip fractures in hips that were randomly assigned to be protected with hip fractures in hips that were not protected. In addition, we conducted a per-protocol analysis in which we removed from the treatment arm any protected hips with an overall adherence less than 75%. This is in slight contrast to the per-protocol analysis in the main HIP PRO study, which used an overall adherence cut-point of 80% to remove individuals from the analysis. In an as-treated analysis, protected hips were re-classified an unexposed after any two week time period in which adherence was less than 75%.

Results

At baseline, the 1042 men and women comprising the study population were elderly, with a mean (standard deviation) age of 85.3 (7.4) years, and were mostly female and white. A majority was severely cognitively impaired, and approximately

104

one-third had a fall within 30 days prior to the baseline visit (Table 4.1). The nursing homes participating in the study were approximately equally likely to be for-profit (49%) and not-for-profit, and most of the nursing homes allowed residents to wear hip protectors some or all of the time at night (83%). Overall adherence for protected hips was 74%.

A total of 564 (54%) study participants dropped out of the study before experiencing a hip fracture and before the administrative end of the study. Treatment assignment did not affect the likelihood of drop-out. Drop-out was 18% higher for participants < 80 years when compared to participants ≥ 80 years, and was 23% higher for males when compared to females (Table 4.2). Conversely, the drop-out was 20% lower for participants with a hip fracture history when compared to participants with no hip fracture history, and 25% lower for participants with severe cognitive impairment when compared to participants with no/mild cognitive impairment. For the 478 patients with non-zero drop-out weights, the mean (standard deviation) was 2.18 (0.89) and the range was (1.01, 11.33).

In table 4.3 we present the replication of the ITT analysis, as well as four other estimates of the effect of hip protectors. Replicating the ITT analysis using Cox regression yielded a 24% increased hazard of hip fractures, when comparing protected hips to unprotected hips. Controlling for clustering within nursing homes did not meaningfully change the ITT results, with only a light loss in precision (hazard ratio=1.24 (95% CI: 0.64, 2.37)). For the per-protocol analysis, where we removed from the analysis any protected hips with an overall adherence less than 75%, the hazard of hip fractures was 40% higher in protected hips when compared

to unprotected hips. The as-treated analysis suggested a 16% increased hazard of hip fractures when compared protected hips to unprotected hips. In contrast to the ITT, per-protocol, and as-treated analyses, which suggested a harmful effect of hip protectors, the analysis using structural nested models suggested a protective effect of hip protectors. When we assumed that drop-out was completely at random, we obtained a hazard ratio of 0.53 (95% CI: 0.10, 2.74) when using a structural nested model; the results did not meaningfully change when we assumed that drop-out was random, conditional on the set of baseline covariates.

In Figures 4.1 and 4.2 we plotted a set of β's by the test statistic. These figures served to demonstrate the association between the potential treatment-free survival time and treatment assignment for many choices of β. Note that the function of (β) was not smooth in both figures because the test statistic jumped suddenly as β was varied. β=0 represented the null hypothesis of no treatment effect, while the β with an associated test statistic that was closest to zero was our estimate of the treatment effect. In Figure 4.1, where we assumed that censoring due to drop-out was completely at random, this minimum occurred at -0.91, yielding a time ratio ($e^{-\hat{\beta}}$) of 2.48. In Figure 4.2 we assumed that drop-out was random conditional on a set of baseline covariates, and obtained a time ratio of 2.41 ($e^{-(-0.88)}$).

We assessed modification of the effect of hip protectors by gender and age. The effect of hip protectors was similar for the 821 females in the HIP PRO study (hazard ratio = 0.54 (95% CI: 0.07, 4.29)) and the 221 males (hazard ratio = 0.56 (95% CI: 0.04, 8.10)). There was a suggestion of stronger but very imprecise protective effect of hip protectors in participants less than 80 years of age (hazard

ratio = 0.31 (95% CI: 0.01,12.53)), when compared to participants greater than 80 years of age (hazard ratio = 0.60 (95% CI: 0.08, 4.08)).

Discussion

Using a structural nested model, we found that constant exposure to hip protectors expanded the time to a hip fracture by approximately 2.5; this corresponds to an approximate 50% reduction in the hazard of hip fractures. In contrast to the ITT effect estimate, the adherence corrected hazard ratio demonstrated a protective but less precise effect of hip protectors.

To estimate an adherence-corrected effect of hip protectors on survival time we made several assumptions. First, we assumed that the structural accelerated failure time model was correctly specified and it captured the biological treatment action of hip protectors on hip fractures. Intrinsic to this assumption was that adherence and hip fractures had been measured without error. Second, we assumed that there was no modification of the treatment effect across time or with respect to patient characteristics. This non-interaction assumption obviated the need to check for time-by-treatment and time-by-covariate interactions.[129] However, we also fit two-parameter structural nested models to allow for modification by age, and gender. Our results did not vary greatly by these sub-groups, suggesting that the one-parameter structural nested model was reasonable. Third, we assumed that the potential treatment-free failure time $U_{i,j}$ for hip i, j was unaffected by the treatment assignment or the potential survival time of other individuals. Finally, we assumed

that the exclusion restriction holds, such that there is no effect of treatment assignment other than through its impact on received treatment.

The g-estimation approach used in this analysis does have certain potential weaknesses. One potential weakness is that the extra censoring introduced at certain levels of Ψ by accounting for end of follow-up results in estimates of lowered precision. This artificial censoring is necessary to avoid selection bias but it can discard potentially useful failure-time information.[138] When the number of outcomes is small prior to re-censoring, as was true with the HIP PRO study, the final effect estimate obtained from g-estimation is likely to be imprecise. Another potential weakness of the g-estimation approach is that it assumes that drop-out is non-informative. Except in studies where all participants are followed for an event that can be assessed by external data sources (e.g., total mortality), accounting for censoring due to drop-out will be necessary. In HIP PRO, 54% of study participants dropped out of the study. For the main adherence corrected effect estimate, we accounted for this substantial drop-out by using weights and assumed that drop-out was random conditional on a set of measured baseline covariates that comprised the weights. In addition, we also obtained an adherence corrected effect estimate where we assumed that drop-out was completely at random. The similarity of the results with these two approaches suggest that drop-out in HIP PRO may be non-informative, unless there exists an unmeasured covariate that was strongly predictive of drop-out but which we did not include in our weights.

In a previous analysis, we employed inverse probability of censoring weights (IPCWs) to correct for non-adherence in the HIP PRO study and obtained a hazard

ratio of 0.55 (95% CI: 0.13, 2.40). The IPCW approach considered participants as dependently censored when they wore hip protectors in fewer than 75% of adherence visits, in any two-week period. This dependent censoring was then adjusted, using information on time-invariant and time-varying covariates that were thought to be common causes of randomization and non-adherence.[132] An advantage of using the IPCW approach to non-adherence correction is that no participant information is needed after the date of non-adherence. One disadvantage, however, is that the IPCW approach assumes that information on all important prognostic factors of non-adherence and hip fractures has been accurately measured and modeled. The structural nested modeling approach used in the current analysis makes no such assumption. If we assume that censoring from drop-out is non-informative, then this approach relies only upon the relationship between pre-randomization variables and potential outcomes.

As-treated and per-protocol analyses have often been employed in randomized studies with poor adherence (due to ease of implementation), sometimes with the mistaken belief that these analyses resolve the problem inherent in an ITT analysis. While these analyses do guarantee a correct classification of exposure, they do not guarantee comparability (or, more precisely, exchangeability)[170] of the groups defined by this exposure. With per-protocol analysis only those individuals who adhere to their assigned treatment are included in the analysis. The as-treated analysis uses all of the data but compares individuals who were randomized to treatment and complied, with a mixed group of compliers and non-compliers who received no treatment.[127, 135-136] When compliers and non-

compliers have a different prognosis at baseline (i.e., when outcome dependent non-adherence is present), both the as-treated analysis and the per-protocol analysis will give biased estimates of the effect of treatment. In contrast, the g-estimation approach, as noted previously, relies on the original randomization and is valid even when there is outcome dependent non-adherence, as long as the underlying structural model is correct.[129, 160]

The assumptions of structural nested models were similar to those used in the instrumental variable (IV) approach for non-adherence correction, where randomization was considered the instrument, randomization was associated with treatment actually received, but randomization was only associated with the outcome (e.g., hip fractures) through its association with treatment received (i.e., there were no common causes of randomization and the outcome).[121] We attempted to implement an alternative IV approach to correct for non-adherence in the HIP PRO study, by comparing rates between treatment arms among "would-be compliers," those who would have worn hip protectors if assigned to wear them and those who would not have worn the hip protectors if not assigned to wear them. This method requires borrowing information from a subset of patients ("always takers" or "never takers") in the alternate treatment arm to inform the rates in the "would-be compliers."[134, 139, 145] In the HIP PRO study, the number of "never takers" could not be calculated in the no-treatment arm, and so the rate for those who would have worn the hip protectors if assigned to wear them could not be calculated. While an IV effect estimate could not be calculated for the HIP PRO study, this approach would be relevant for non-adherence correction in studies where accounting for follow-up

time is not relevant. One advantage of structural nested models, when compared to the simple IV approach outlined above, is that structural nested models account for the timing of outcomes as well as non-adherence.[124]

To correct for non-adherence in the HIP PRO study, we have employed several competing methods to estimate adherence-corrected treatment effects. Results from both IPCWs and the structural nested models in the current analysis suggest that hip protectors may be protective. As with any modeling strategies, these methods rely on additional assumptions beyond those needed for an ITT analysis. Nevertheless, the treatment effects revealed by these analyses account for post-randomization non-adherence and should serve to supplement the findings from an ITT analysis.

Tables and Figures

Table 4.1. Selected Baseline Characteristics of 1042 Men and Women Randomized to Wear Hip Protectors on Either the Left Hip or the Right Hip in the Hip Impact Protection Project (HIP PRO).

Characteristic	Nursing Home Residents (N=1042)[b]
Women	821 (78.8)
White[a]	895 (85.9)
Age, Mean (SD), year	85.3 (7.4)
History of osteoporosis[a]	202 (19.4)
History of fall in last 30 days[a]	301 (29.2)
Marital status[a]	
Never	193 (18.5)
Married	153 (14.7)
Widowed	631 (60.6)
Divorced/Separated	65 (6.2)
Cognitive status[a,c]	
Normal / Mild impairment	151 (14.5)
Moderate impairment	147 (14.1)
Severe impairment	744 (71.4)

[a]Missing data (<3% of total residents) replaced with mode of distribution.

[b]Represents total N(%), unless otherwise noted

[c]Cognitive status categories based on Short Blessed Test (SBT): normal / minimal impairment (SBT = 0 – 8), moderate impairment (SBT = 9 – 19), severe impairment (SBT ≥ 20).

Table 4.2. Hazard ratios and 95% confidence intervals for covariates used to fit a survival model for the probability of drop-out.

Variable[a]	Category	Hazard Ratio	95% Confidence Interval
Treated Hip		0.99	0.88, 1.12
Age > 80 Years		1.18	1.02, 1.38
Non-white Race		1.07	0.91, 1.28
Male Gender		1.23	1.05, 1.43
Hip Fracture History		0.79	0.63, 1.00
Osteoporosis History		0.90	0.77, 1.05
Marital Status	Single	1.00	
	Married	0.95	0.71, 1.25
	Widowed	0.93	0.70, 1.25
	Divorced	1.20	0.92, 1.52
Cognitive Status	No / Mild Imp.	1.00	
	Moderate Imp.	0.98	0.80, 1.22
	Severe Imp.	0.75	0.64, 0.89

[a] All variables were measured at baseline.

[b] Cognitive status categories based on Short Blessed Test (SBT): normal / minimal impairment (SBT = 0 – 8), moderate impairment (SBT = 9 – 19), severe impairment (SBT ≥ 20).

Table 4.3. Hazard ratios and 95% confidence intervals for the effect of hip protectors on time to hip fractures among 1042 nursing home residents.

Analysis	Hazard ratio	95% Confidence interval
Intention-to-treat	1.24	0.65, 2.34
Per-protocol[a]	1.40	0.69, 2.85
As-treated[b]	1.16	0.59, 2.30
Structural nested model: drop-out completely at random[c]	0.53	0.10, 2.74
Structural nested model: drop-out not at random[c,d]	0.46	0.07, 2.84

[a] Hips with overall adherence < 75% were removed from the analysis (note that unprotected hips were 100% adherent).

[b] Exposure status was time-varying, and, for exposed hips, was re-defined as unexposed after any two week time period in which adherence was less than 75%.

[c] Assumed that the underlying potential survival times followed a Weibull distribution.

[d] Assumed that drop-out was random conditional on the following baseline covariates: treatment status, age, race, gender, hip fracture history, marital status, osteoporosis history, and cognitive status.

Figure 4.1. Plot of β by the Wald Chi-square test, assuming that censoring due to drop-out was completely at random.
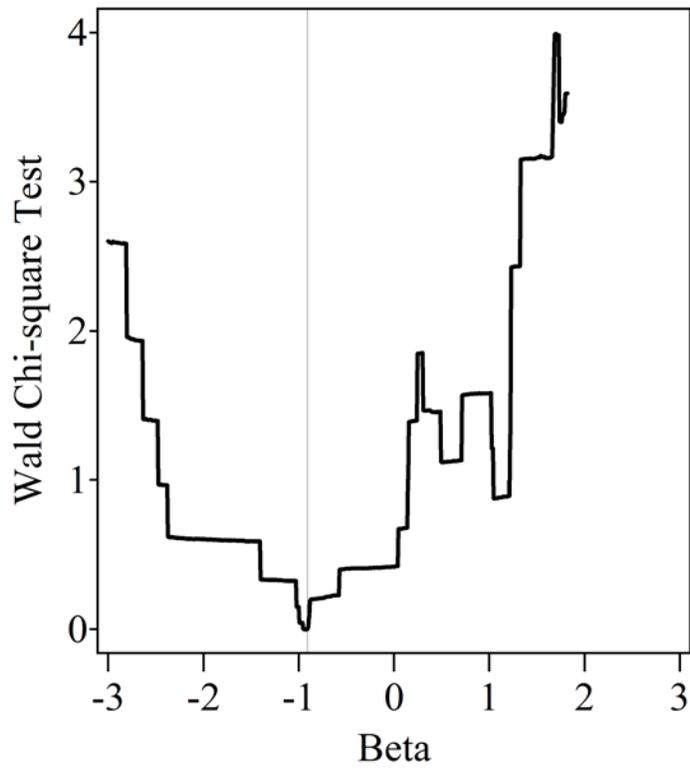
Figure 4.2. Plot of β by the Wald Chi-square test, assuming that censoring due to drop-out was ignorable given a set of baseline covariates.
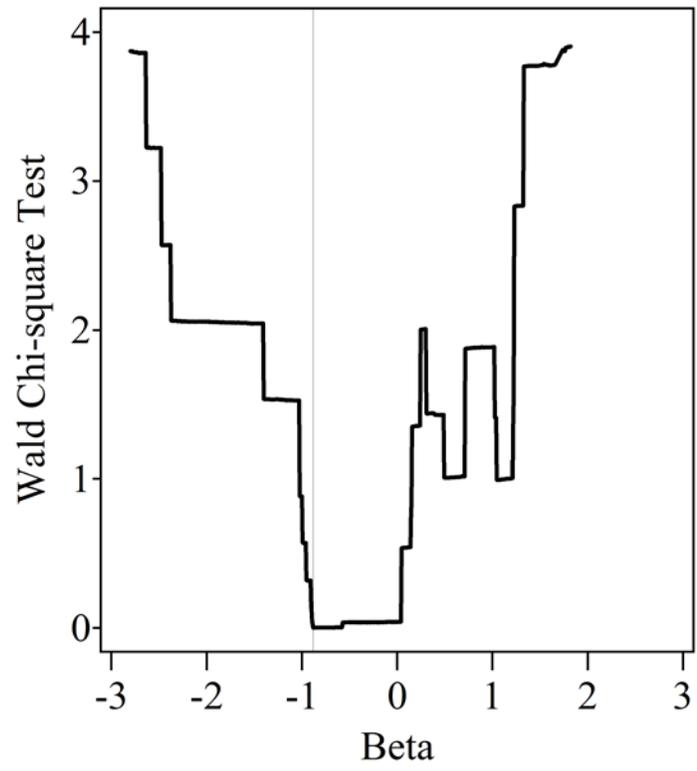
Chapter 5

CONCLUSION

FUTURE METHODOLOGICAL DIRECTIONS

We have seen that, in the presence of non-adherence, there are several valid

alternative methods to estimate efficacy in randomized studies. Despite the

existence of these methods, current regulatory practice still largely relies on the ITT

analysis. There appear to be two main reasons for this preference. First, the ITT

analysis simplifies decision-making. There is no need to consider post hoc modeling

and analyses to account for non-adherence. In contrast to complicated and,

perhaps, subjective model-based analyses, a simple ITT analysis with easily

understood conclusions is preferred. However, an ITT analysis may not answer the

main question of interest when non-adherence is substantial (i.e., does the receipt of

treatment have an effect on the outcome). We contend that if non-adherence is

present in more than a trivial proportion of participants, we must adopt more

elaborate models to estimate efficacy. In this situation, a simple ITT analysis will not

be adequate. As long as the assumptions are clearly defined and understood, the

use of more complex analyses should not be abandoned altogether. A second

reason for the preference of ITT analysis is the belief that any analysis that is not

based on the original randomization is biased.[135] This is not always true, as

demonstrated with the analysis using inverse probability-of-censoring weights in Chapter 3.

If we accept the need for alternatives to the ITT analysis in the presence of non-adherence, as we have done, there still remains a need to better understand the strengths and weaknesses of each approach. Instrumental variable (IV) analysis and analyses using structural nested models (SNMs) do not require information on the common causes of non-adherence and the outcome, but do require that individuals are followed until the planned end of the study. Analyses using inverse probability of censoring weights (IPCWs), in contrast, require no follow-up information after the study participant first becomes non-adherent, but do require that all important concomitants of non-adherence and the outcome are measured. Thus, we do have good understanding of when a particular method to correct for non-adherence may be most applicable. In many situations more than one method may be appropriate. When this is true and the analyses reveal similar results we may be more confident in the findings because each of the methods rely on different approaches and assumptions.

Still, questions remain. For example, it appears that the simple IV approach that we attempted to use is guaranteed to result in an adherence-corrected effect estimate that is stronger than the ITT effect estimate and which remains on the same side of the null. Is it reasonable that this is always true? We have seen with the IPCW and SNM analyses that non-adherence can result in a biased estimate that goes through the null. Another question that remains is 'to which study population does the effect estimate apply?' The target population of an ITT or IPCW

118

analysis is the total study population. In contrast, the target population of the IV approach is the 'would-be compliers' (i.e., those who would comply if assigned either treatment). For the SNM analysis, the effect estimate appears to be applicable to the actual complies, which, in certain circumstances (e.g., a randomized control trial), will be the same as the 'would-be compliers' because of randomization. We explore the target populations of an ITT analysis, an IV analysis, and an SNM analysis, using indirect evidence under a simulation exercise. The findings from the simulation exercise are presented in Appendix 4.

IMPROVING NON-ADHERENCE IN FUTURE HIP PROTECTOR STUDIES

Non-adherence is a multi-factorial problem which is influenced by the study setting, the treatment regimen, the characteristics of the disease, and patient characteristics, some of which are stable and others that are time-varying. With the hip protector intervention, we have seen that adherence may be different in institutional settings versus community settings, that the appearance and design of the hip protector impacts the use of the intervention, and that a patient's gender, age, cognitive status, continence, and other factors play a role in adherence. The complexity in the factors influencing adherence can usually be ignored when non-adherence is low. In hip protector studies, however, non-adherence has been pervasive and high, and, so, cannot be ignored. Consequently, the goal of this dissertation has been to implement methods to account for non-adherence in one particular randomized study of hip protector, the HIP PRO study. These methods

are important because they provide alternative tools for the epidemiologist to estimate effects, even in situations were non-adherence is unavoidable. However, the goal of future hip protector studies should not be simply to account for non-adherence, but also to improve adherence.

With the exception of the Meyer et al[118] and Kiel et al[16] studies, pervious efforts at improving adherence have often been limited to one of the following strategies: (1) providing pamphlets or videos that discuss the risk of hip fractures and the benefits of hip protectors, (2) using nursing or study staff to measure and encourage adherence, (3) improving the design of the hip protectors to decrease discomfort and improve convenience. In contrast, Meyer et al[118] used a more holistic approach to improve adherence, using social learning theory.[173-174] They used educational sessions to identify a target population that was interested in learning about the risk of hip protectors, believed that hip protectors could reduce the risk of fractures, and felt susceptible to hip fractures. After enrollment in the study, caregivers used verbal persuasion and encouragement to improve self-efficacy, and thus, adherence. Other competing theories, such as contingency contracting[175], which gives patients greater control over their own medical care, also provide a more comprehensive approach to improving adherence. The attraction of these alternate approaches is that they have a multi-factorial approach that incorporates information on how individuals conceptualize, cope, and appraise their illness state and address the psychological processes underlying a patient's non-adherence.[176] Unfortunately, such approaches necessarily restrict the population and limit the ability to generalize study findings to the broader elderly population.

For example, cognitive impaired individuals would be excluded under both approaches above.

Some would argue that it would be inappropriate to exclude individuals with dementia from hip protector studies. As evidenced by the characteristics of the HIP PRO study population, many individuals in long-term care settings will be cognitively impaired. If these individuals are included in efficacy trials of hip protectors, the main driver of adherence will be the caretakers. As with study participants, the strategies to improve adherence through caretakers include knowledge about the risk and consequences of hip fractures for patients. This knowledge can be provided by trained personnel, videos, pamphlets, or any combination of these strategies; group discussions with peers can fill in gaps of knowledge and provide the social support that fosters an environment in which encouragement of adherence becomes an important institutional goal. The purpose of such training is to overcome the educational, emotional, and behavioral barriers in caretakers that may hinder or limit adherence in their patients. Factors that can be especially important in the interaction between caretakers and patients with cognitive impairment include practitioner warmth and concern, sensitivity to patient's non-verbal behavior, and using appropriate non-verbal cues to respond to patients.[177] Caretakers should be encouraged to practice the skills needed to aid in improving adherence. Improving these skills with cognitively impaired patients will result in improved interaction with all patients.

In the past, the biomedical framework has viewed non-adherence as resulting from aberrant behavior on the part of the patient who refuses and disobeys the

prescription of treatment regimens. This model focused on the influence of patient

characteristics on adherence, and did not address factors and cognitive processes

that can also influence adherence. The focus had been on identifying groups at risk

of non-adherence, but has had limited influence in improving adherence. The

alternative theoretical frameworks outlined above have strengths and weaknesses

and may not be appropriate in every situation. In the context of future hip protector

studies, it is the investigators responsibility to decide which approach, or

combination of approaches, might best help to improve adherence in the particular

study population that she is studying. Such thoughtful approaches should help to

improve adherence.

Appendix I

Search Criteria for Systematic Review

PubMed Search statement:

(hip fracture* OR (femor* AND neck[tw] AND fracture*)) OR "hip fractures"[MeSH Terms] AND ("protective devices"[MeSH Terms] OR protective device* OR protective cloth* OR hip protector* OR hip pad*) AND (randomized controlled trial[pt] OR controlled clinical trial[pt] OR "randomized controlled trials"[MeSH Terms] OR "random allocation"[MeSH Terms] OR "double-blind method"[MeSH Terms] OR "single-blind method"[MeSH Terms] OR (random* AND control* AND trial*)) AND "aged"[MeSH Terms]

CINAHL Search Statement:

((((MH "Hip Protectors")) or (MH "Protective Clothing") or (protective cloth*) or (protective device*) or (hip pad*)) and ((MH "Hip Fractures") or (hip fracture*) or (femor* and neck and fracture*))) and ((MH "Clinical Trials+") or (random* and control* and trial*)) and (MH "Aged+")

Cochrane Search:

Use Cochrane Phase 1 search to limit results as closely as possible to RCTs. Rationale: PubMed RCT publication type has been in use since 1991. The earliest pertinent article retrieved using the broader All Phases search was from 1993. Since this is a relatively recent topic, we do not need all the possible text words related to RCTs in the All Phases search which aid in finding articles older than 1991 that were not well indexed to RCTs. Phase 1 results provide an adequate balance of precision and retrieval.

Cochrane Clinical Trials Search Statement

#1	MeSH descriptor Hip Fractures explode all trees	669
#2	MeSH descriptor Protective Devices explode all trees	1168
#3	(hip near fracture* OR femor* near neck near fracture*)	1304
#4	(protective near device* OR protective near cloth* OR hip near protector* OR hip near pad*)	549
#5	((#1 OR #3) AND (#2 OR #4))	43

ISI Web of Science Statement:

(TS=(hip fracture* OR (femor* AND neck[tw] AND fracture*)) AND (protective device* OR protective cloth* OR hip protector* OR hip pad*))

DocType=Meeting Abstract OR Meeting Summary OR Meeting-Abstract; Language=All

languages; Databases=SCI-EXPANDED, SSCI, A&HCI; Timespan=1955-2009)

EMBASE Search

Use the search strategy used for the PubMed data base, keeping only references that

were not identified in the PubMed data base.

Appendix 2

Description of Selected Publications Identified in a Systematic

Review of Randomized Hip Protector Studies

Lauritzen and colleagues[88] conducted the first randomized study of hip

protectors in Denmark, using a hip protector design that would later become standard

for the SafeHip hip protector. Within one nursing home 10 wards (247 participants)

were randomized as hip protector wards, while 18 wards (418 participants) were

provided no protectors. During eleven months of follow-up 8 hip fractures occurred in

the hip-protector group and 33 hip fractures occurred in the unprotected group, yielding

a rate ratio (95% CI) of 0.44 (0.21, 0.94). Adherence was evaluated only in a fall sub-

study in two wards from the treatment arm and it was noted that hip protectors were

being used only during 14/45 (21%) of registered falls.

As the first study of hip protectors with a precise and beneficial effect of hip

protectors, the study of Lauritzen et. al. was influential in prompting additional interest in

the use of hip protectors to prevent hip fractures. Unfortunately, the study was of poor

quality, with questionable design issues, and suffered from inappropriate analyses. No

data was provided to demonstrate balance of ward characteristics among the

randomization arms, or to demonstrate the balance of patient demographic

characteristics and medical history. Also, during the course of the study, patients who

dropped out of the study were replaced from a waiting list. In the presence of non-

random drop-out such post-randomization replacement could negate some or all of the

benefits of randomization. Finally, in the analysis, the study authors did not adjust for clustering within wards of the nursing home. Failure to account for the clustering would have resulted in overly precise effect estimates.

Ekman et al[110] conducted the next randomized study of hip protectors using protectors (JOFA AB, Malong, Sweden) that were similar to those employed by Lauritzen, where one nursing home (302 residents) was randomized to the hip protector arm and the remaining three nursing homes (442 residents) were assigned to the no hip protector arm. After 11 months of follow-up, 4 hip fractures occurred in the treatment arm and 17 in the control arm. A relative risk estimate of 0.33 (95% C.I.: 0.11, 1,00) supported the earlier findings of Lauritzen et al, and suggested that hip protectors may prove a strong protective effect. The overall "average" adherence was better at 44%; however, adherence measurement was not clearly defined. The major reasons for non-adherence included skin irritation and being bed-ridden. As with the Lauritzen trial, there was a lack of adjustment in the analysis for cluster randomization. Because the study by Ekman et al was a brief report, potentially important information on the timing of hip fractures and drop-out was not reported.

Another brief report, in a letter to the editor by Jantti et al[116], presented results in 1998 for seventy-two residents, residing in a municipal nursing home in Finland, who were equally randomized to wear hip protectors and not to wear hip protectors. The hip protector used in the study was designed by one of the authors and employed a polyethylene foam padding that was also used in knee protectors by hockey players. Participants were followed for one year and during this time one hip fracture occurred in

a protected hip, while seven hip fractures occurred in the control arm (RR (95% CI): 0.14 (0.02, 1.10)). The authors used an unusual definition of adherence, noting that of the 19 participants who were still living in the nursing home at the end of follow-up, 13 (68.4%) were still wearing the hip protectors. Also, the method by which adherence was evaluated was not clear. For those who continued to wear hip protectors, the reasons for adherence included the fact that the protectors were warm, the protectors increased the feeling of safety, and diminished the fear of falling. The authors failed to account for the background rate of hip fractures and provided insufficient power to determine a meaningful estimate of the effect of treatment.

A pilot study by **Villar and colleagues**[87] in the United Kindom was designed to evaluate adherence of nursing home residents to the wearing of hip protectors. The study randomized 101 rest home patients to wear hip protectors with an outer shield of polypropylene and inner plastozote lining; 40 rest home residents were randomized to the no treatment arm. During twelve weeks of follow-up no hip fractures occurred in either arm of the study. Adherence was monitored by randomly timed visits every 14 days. At the end of the study only 27 residents (27%) wore the hip protectors for the full 12 weeks. The reasons given for non-adherence included discomfort and poor fit; other reasons included physical difficulty, changed mind, illness, and forgetfulness. Although the authors clearly state that the purpose of the study was to evaluate adherence and not to demonstrate efficacy, the study nevertheless demonstrates that a study of short duration and small sample size would be unable to evaluate the efficacy of hip

protectors. The risk of hip protectors, even in an institutionalized setting, is small during

such a short follow-up time.

One of the largest randomized studies of hip protectors was undertaken by

Kannus and colleagues[89] in 22 community-based health centers in Finland.

Randomization occurred at the facility level in a 1:2 ratio to either a hip protector group

or to a control group. Although the number of facilities in each treatment arm was not

clear, 653 individuals were designated to wear the hip protectors and 1148 were

assigned to the control group. The investigators used the KPH hip protector, one of the

first hip protectors to be worn inside pockets of an undergarment. In the treatment

group there were 13 hip fractures, while in the control arm there were 67 hip fractures;

the relative hazard (95% C.I.) was 0.4 (0.2, 0.8). Adherence, defined as the number of

days the hip protector was worn as a percentage of all follow-up days, was 48%. In the

treatment arm 31% of participants completely refused to wear the hip protectors after

randomization. As with previous studies, there were only minor adverse events

associated with wearing the hip protectors, including skin irritation, abrasions, and

swelling of the legs. During follow-up four subjects in the treatment arm experienced a

hip fracture, despite wearing the hip protectors at the time of the fall. Because of

variable patient demographics and medical history, as well as variation in the

characteristic of the fall, hip protectors were not always protective, even when worn

correctly. In the analysis, the authors failed to account for clustering within facilities;

nevertheless, the results from the study contributed to mounting evidence that hip

protectors were strongly protective against hip fractures.

Birge et al[113] presented results from a randomized study of hp protectors in an abstract. Randomization occurred at the nursing home level in 18 nursing homes (317 residents). The number of nursing homes and the number of residents in each treatment arm was not provided. Four hip fractures occurred in protected hips and six hip fractures occurred in unprotected hips. Because follow-up time was provided only for protected hips, a measure of effect could not be estimated. Overall adherence in the study was 63%. However, it was not clear how adherence was defined and measured. The failure to include important randomization information in the abstract limited the contribution of the study to the literature.

The first study of hip protectors in which randomization occurred at the individual level was undertaken by Cameron and colleagues[103] in Australia. Inclusion criteria limited the study population to women aged 75 years and older who were at high risk of falling, and lived in hostels or nursing homes. Eighty-six participants were randomized to wear the SafeHip hip protector and eighty-eight participants were randomized to the control group. During eighteen months of follow-up eight hip fractures occurred in the intervention group and seven hip fractures occurred in the control group; the hazard ratio (95% CI) was 1.46 (0.53, 4.51). A unique feature of the study was that an adherence nurse encouraged participants and facility staff to increase the use of hip protectors in the treated group; the nurse evaluated adherence at two weeks, and at 2, 10, and 18 months. Adherence, defined as the use of hip protectors for at least half of the day, was 70% during the first two months of the study, but declined to approximately 50% by the end of the study. The small study size resulted in an imprecise effect

130

estimate. Nevertheless, this was the first study to suggest that hip protectors may not be as beneficial as previously reported. Randomization at the individual level may have been problematic for institutionalized participants because staff members, with limited time for non-care activities, may have not been able to implement distinct treatment plans for individuals in the two unique treatment groups.

Harada et al[91] conducted another small randomized study in elderly female residents in Japanese nursing homes, where the unit of randomization was the individual. There were 164 participants who were not wheel-chair bound and had the ability to stand unaided. Of the 164 eligible nursing home residents, 88 were randomized to wear "shell-shaped, polypropylene" hip protectors and 76 were randomized to not wear the protectors. The average follow-up time was 360 (S.D. 255) days for the treatment group and 397 (S.D. 244) days for the control group. During this time one hip fracture occurred in protected hips and eight hip fractures occurred in the untreated group. Despite a successful randomization which adequately balanced patient characteristics in the randomization arms, the investigators conducted an adjusted Cox regression analysis, controlling for patient age, body weight, height, grip strength, tricep skin-fold thickness, and number of falls per subjected. The adjusted hazard ratio for the effect of hip protectors was 0.08 (0.01, 0.75). Adherence was evaluated daily by care staff and was measured in three categories: (1) complete 24 hour wearing, (2) incomplete wearing, and (3) not wearing the protector at all. The investigators reported a surprisingly high rate of complete wearing of 70% at the end of the study and 17% incomplete wearing. The high adherence in this study relative to

previous studies of hip fractures may be a result of cultural factors, patient case-mix, and, as suggested by the authors, highly motivated staff members in the nursing homes.

One study by **Birks and colleagues**[95] was conducted among community dwelling individuals who had a history of a previous hip fracture, presumably because these individuals were at high risk of a second hip fracture and would be highly motivated to prevent a second hip fracture. Participants were recruited by television and newspaper advertisements, as well as from orthopedic wards, when individuals were to be discharged back home. Eligible participants were 70 years or older, were not bed or chair-bound, and had a single previous hip fracture. The 182 individuals randomized to the treatment arm were issued 3 SafeHip hip protectors, while 184 controls received no hip protectors. The main outcome was a second hip fracture, which was ascertained by contacting the general practitioner for each study participant. Six hip fractures occurred in the treatment group and two hip fractures occurred in the control group. Logistic regression analysis yielded an unadjusted odds ratio (95% C.I>) of 3.10 (0.62, 5.58). Self-reported adherence indicated that 60% of individuals in the treatment group still occasionally wore the hip protectors at the end of the study, but only 34% always wore the garments. Self-reported adherence tends to be higher than adherence evaluated by external monitors. As a result, it may be that the true adherence is even lower than suggested by self-report. Details for outcome ascertainment were sparse (e.g., there was no information provided about the response rate for general practitioners who were contacted about the occurrence of a second hip

fracture); if the response rate was differential for the two treatment arms then the study results would be biased. Also, it is not clear whether control patients were prohibited from wearing hip protectors. If there was important hip protector use in the control group then the estimate of the effect of hip fractures would also be biased.

Another study which evaluated the efficacy of community dwelling individuals occurred in Australia. **Cameron et al[96]** randomized cognitively intact women who were 74 years of age or older, lived in their own homes, and were at a high risk for falls, to receive SafeHip hip protectors and contact with an adherence nurse, or to receive no treatment. Out of a total of 600 eligible women, 302 were randomized to the treatment arm and 298 were randomized to the control arm. Planned follow-up was for two years, and, after randomization, residents were allowed to continue wearing hip protectors even if they no longer resided in their own homes. There were twenty-one hip fractures in the treatment arm and twenty-two hip fractures in the controls. The intention-to-treat estimate of the relative risk (95% CI) was 0.93 (0.51, 1.69), indicating no effect of treatment. As in the Birks 2003 study, the outcome was ascertained by self report, with follow-up of radiography reports and hospital records. Adherence was encouraged by nurses during in-home visits and by telephone calls. Self-reported adherence was defined as the amount of time that hip protectors were worn during the day. During the first year of follow-up 57% of participants reported wearing the hip protectors at least half of the day; by 18 and 24 months, however, adherence had fallen to 50% and 42%, respectively. As in the Birks et al study, the attempt to determine the effectiveness of

hip protectors in a community setting was admirable. However, inadequate exposure and outcome ascertainment make the estimate of efficacy difficult to interpret.

A study by **Meyer and colleagues** in Germany[118] used social learning theory to design a randomized study in which individuals in the treatment arm (as well as the staff of the nursing homes) received a structured education program, in addition to the SafeHip protector. In the control arm participants were offered a short introduction to hip protectors and they then had the option to also wear hip protectors. Randomization was at the level of the nursing home, with 25 nursing homes assigned to the intervention arm and 24 nursing homes to the control arm. The 459 participants in the intervention arm and 483 residents in the control arm had a high risk for falling, were 70 years of age or older, were not bed-ridden, and had lived in a nursing home for more than three months at baseline. During eighteen months of planned follow-up, 21 hip fractures occurred in the intervention group and 42 hip fractures occurred in the control group – the relative risk (95% CI) of hip fractures was 0.57 (0.19, 0.92), when comparing protected to unprotected hips. The authors reported only adherence during falls and noted that participants in the treatment arm were wearing the hip protector during 54% of falls; in the control group, hip protectors were worn in 8% of falls. Because individuals in the control group were not prohibited from wearing hip protectors, 40/483 (8%) of the participants in the control group used a hip protector during the study. The unusual study design, where control participants were also allowed to wear hip protectors, probably attenuated the true effect of hip protectors in this study.

A large, well-designed study was undertaken by **van Schoor and colleagues**[107] in the Netherlands, where residents of apartment houses for the elderly and residents of nursing homes were randomly assigned to receive SafeHip hip protectors or not. In both the intervention arm and the control arm, participants and nurses were provided education about the risk for hip fractures among institutionalized individuals and about the causes and consequences of hip fractures. The study population was restricted to individuals aged 70 years of age or older who had a high risk for falling, but who were mobile and did not have a history of bilateral hip fractures. Of 561 individuals who met the eligibility criteria, 276 were randomized to the intervention arm and 285 were assigned to the control arm. Total follow-up in the intervention arm was 357 person-years, during which 18 hip fractures occurred. In the control arm 20 hip fractures occurred during a total follow-up of 398 person-years. The hazard ratio (95% CI) for the effect of hip protectors was 1.05 (0.55, 2.03). Adherence was assessed during unannounced visits by a research assistant at 1, 6, and 12 months after inclusion in the study (this was the first study in which adherence was evaluated by unannounced visits). Newsletters were sent to the nurses of the elderly care homes and nursing homes to encourage and emphasize the importance of adherence. Despite these efforts, overall adherence in the study was similar to other studies. Adherence was 61% at 1 month, 45% at six months, and 37% at twelve months. In addition to the intention to treat analysis, the authors also presented results from a per-protocol analysis where, in the treatment arm, they kept only participants who were adherent during all three unannounced adherence visits. The per-protocol hazard ratio (95% CI)

135

was 0.77 (0.25, 2.38). When adherence is low investigators often rely on the per-protocol analysis to estimate the effect of treatment in only adherent patients. However, because the per-protocol analysis relies on post-randomization decisions to keep patients in the intervention arm, the results of such an analysis can be severely biased.

Birks et al[112] conducted another randomized study of hip protectors among women living in the community in the United Kingdom. Volunteers were recruited from the age-sex registers of general practitioners and through media advertisements. Volunteers were eligible for the study if they were female, older than 70 years of age, at high risk for a hip fracture, living in their own homes, and no history of bilateral hip fractures. Recruitment letters were sent to more than 70,000 women and approximately 19,000 (27%) responded. A total of 4,169 women met all eligibility criteria and, of those, 1,388 were randomized to the treatment arm and 2,781 to the control arm. Participants in the intervention arm were mailed three SafeHip hip protectors and a leaflet describing methods to reduce the risk of hip fractures; the control arm received only the leaflet. Hip protector adherence and hip fracture outcomes were assessed by self-report via questionnaires mailed to all participants every six months. Hip fractures were confirmed by the patient's general practitioner. General practitioners were also contacted during the two year follow-up period, if participants failed to respond to the mailed questionnaire, mainly to ascertain a hip fracture outcome. During the two year follow-up period 39 hip fractures occurred in the treatment arm and 66 in the control arm – the intention-to-treat effect of hip protectors was an odds ratio (95% CI) of 1.17 (0.78,1.75). Self-reported adherence was low at 6 months (38%), and dropped even lower to 31% at

12 months. As in the previous study by Birk et al, exposure and outcome were ascertained through self-report and hip fracture was confirmed by follow-up with the patient's general practitioner. A smaller validation study to evaluate the performance (i.e., sensitivity, specificity) of self-reported exposure and self-reported outcome in this setting would been valuable in evaluating the validity of the main study.

A cluster-randomized study of hip protectors was conducted by **O'Halloran and colleagues[83]** in 146 nursing and residential homes in the greater Belfast area of Northern Ireland. There were 40 homes (representing 1366 occupied beds) allocated to the intervention arm and 87 homes (representing 2751 occupied beds) in the control arm. The intervention was a policy of providing SafeHip hip protectors to residents of the nursing homes and supporting the implementation of the policy through education and training. Specifically, a nurse facilitator provided on-going support to the care staff and nursing home residents, including one hour workshops focusing on the risks and consequences of hip fractures, provision of manufacturer leaflets promoting the use of hip protectors, and provision of a videotaped presentation on the use of hip protectors. Adherence was assessed as a percentage of occupied beds using the hip protectors – initial acceptance was 37% but decreased to 20% by the end of the 72 week follow-up period. There were 85 hip fractures in the intervention arm and 163 hip fractures in the control arm. Using the total number of occupied beds to calculate the hip fracture rate in each arm of the study, the authors calculated a rate ratio (95% CI) of 1.05 (0.76, 1.45) for the effect of hip protectors. The negative finding in the study is not surprising, given the very low adherence over the course of the study. Also, there were 24

individuals in the control group who wearing hip protectors at the start of the study and they were not excluded from the analysis. A study of this type may be useful at evaluating the effectiveness of hip protectors, but does not provide useful information about efficacy.

Another cluster-randomized study of hip protectors in Japan was reported by **Koike and colleagues**[178] in an abstract. There were a total of 76 aged-care facilities that were randomized in a 1:3 ratio to control and protector groups (the total number of facilities in each treatment arm was not indicated). The intervention involved the provision of the SafeHip hip protector to women older than 65 years age that were mobile and at high risk for hip fractures. In addition, monthly newsletters, posters, illustrated books, and visits by the research team were provided to improve the motivation of the nursing staff in the nursing homes. There were a total of 308 participants in the treatment arm and they experienced 6 hip fractures during follow-up; the 306 participants in the control arm experienced 17 hip fractures. The relative risk (95% CI) of hip fractures was 0.35 (0.14, 0.86), when comparing the treated group to the control group. In the abstract, the authors indicated that 21 residents were excluded from the intention to treat analysis if they were bed-ridden, hospitalized or died during the 10 months of follow-up. Such exclusion would invalidate the intention to treat analysis and it is not clear why the authors excluded these patients. In addition, randomization occurred at the nursing home level in a 1:3 ratio, but the total number of residents in the two treatment arms is approximately equal. This apparently conflicting information demands further explanation.

Finally, a study by **Kiel et al**[16] reported a multi-site randomized study of hip protectors in U.S. nursing homes. This study was the one used for all analyses in this dissertation and is described in greater detail later. Briefly, the study randomized 37 nursing homes to have eligible residents wear hip protectors on either the left hip or right hip. There were a total of 1042 participants who provided 676 person-years of follow-up. A total of 21 hip fractures occurred in protected hips and 17 hip fractures occurred in unprotected hips. The estimate for the effect of treatment was a rate ratio (95% CI) of 1.24 (0.65, 2.34). Overall adherence was 74%.

Appendix 3

The Effect of Hip Protectors in HIP PRO II

After termination of the first HIP PRO study, a second study of hip protectors was initiated using the FallGard FG-04 hip protector. This hip protector was slightly more flexible than the hip protectors used in the first study and was composed of a dense (10 lbs / ft^3), 0.57 inch layer of polyvinyl chloride (PVN) rubber foam, backed by a softer 4 lbs / ft^3 PVN foam layer 0.18 inches thick. The dimensions of this new hip protector were similar to the hip protectors used in the first study and utilized the same cotton Lycra undergarment. In this second study, a total of 1445 residents contributed 1467 hip years of observation between October 2004 and August 2006. Some nursing home residents were switched to the new hip protector after termination of the first study, while 988 entirely new residents were also enrolled into the second hip protector study. As in the first study, participating residents had a mean age of 85 ± 8 years, and were primarily women (78%), white (86%), and had severe cognitive impairment (72%) (Table A3.1). All but four nursing homes participating in the first study also contributed data to the second study; in addition, 18 new nursing homes were recruited. These nursing homes were large (≥ 100 beds), mainly not-for-profit institutions (55%) (Table A3.2).

Our replication of the intent-to-treat (ITT) analysis, using pooled logistic regression, found a rate ratio of 1.13 (95% confidence interval (C.I.): 0.57, 2.27), comparing hip fractures in protected hips against hip fractures in unprotected hips. As

with protectors in the first HIP PRO study, the ITT analysis for the second study

suggests that the FallGard FG-04 hip protectors caused hip fractures, resulting in a non-

significant 13% increase in the incidence of hip fractures (compared to a 23% increase

in the first study). Employing the marginal structural models of Robins et al, we again

constructed inverse probability of censoring weights (IPCWs) using a joint set of

covariates that were determinants of non-adherence and hip fractures. The set of

covariates that resulted in the best-behaved weights (mean=1.00 ± 1.27) were the same

as those used in the adherence correction analysis that was employed in the first HIP

PRO study. In the second study, after correcting for non-adherence with IPCWs, we

found that constant exposure to the FallGard FG-04 hip protector resulted in a rate ratio

of 0.75 (95% C.I.: 0.17, 3.36), when comparing hip fractures in protected hips against

hip fractures in unprotected hips (Table A3.3).

Both HIP PRO studies, individually, had insufficient precision and were unable to

definitively evaluate the outcome, when comparing protected and unprotected hips. In

both studies, the study design was the same, information on the clinical endpoint was

evaluated by a hip fracture adjudication committee in a similar manner, and adherence

was assessed in three weekly unannounced visits. The only difference between the

studies were the hip protectors used, the study time periods, and different (though

overlapping) study populations and nursing homes. Given the similarities between the

studies and the few potential sources of heterogeneity, we proposed to combine the

adherence-adjusted effect estimates from the two HIP PRO studies to estimate an

overall effect. We assessed the variability between the two studies using a fixed effects

model, where we assume that the variability is exclusively due to random variation, and used inverse variance to calculate the summary estimate.

Figure A3.1 shows a forest plot of the two HIP PRO studies, as well as the combined effect estimate. There was no evidence of heterogeneity (χ^2=0.09, 1 d.f., p=0.77). The pooled data show a protective effect of hip protectors in preventing hip fractures, but the results remain statistically non-significant (RR=0.64, 95% CI: 0.23, 1.82). Even when we combine the adherence-corrected results from both HIP PRO studies, there were insufficient hip fractures to demonstrate a definitive protective effect of hip protectors. Based on previous under-powered studies of hip protectors and the results from the current analysis, we recommend that researchers use conservative estimates of hip fracture incidence when designing randomized studies of hip protectors.

Table A3.1. Selected Baseline Characteristics of Enrolled Residents (HIP PRO[1] Study 2)

Characteristic	Nursing Home Residents (N=1445)[3]
Women	1126 (77.9)
White[2]	1235 (85.5)
Age, Mean (SD), year	84.5 (7.5)
History of Osteoporosis[2]	319 (22.1)
History of Fall in Last 30 Days[2]	424 (29.3)
Marital Status[2]	
Never	228 (15.8)
Married	256 (17.7)
Widowed	875 (60.6)
Divorced/Separated	86 (5.9)
Cognitive Status[2,4]	
Normal / Mild Impairment	197 (13.6)
Moderate Impairment	207 (14.3)
Severe Impairment	1041 (72.0)

[1]Hip Impact Protection Project.

[2]Missing data (<2% of total residents) replaced with mode of distribution.

[3]Represents total N(%), unless otherwise noted

[4]Cognitive status categories based on Short Blessed Test (SBT): normal / minimal impairment (SBT = 0 – 8), moderate impairment (SBT = 9 – 19), severe impairment (SBT ≥ 20).

Table A3.2. Nursing Home Characteristics (HIP PRO[1] Study 2)

Characteristic	Nursing Home (N=51)[2]
For Profit	23 (45.1)
Total Number of Hospital Beds	
< 100 Beds	5 (9.8)
100 – 149 Beds	22 (43.1)
150 – 199 Beds	17 (33.3)
>= 200 Beds	7 (13.7)
NH Allows Hip Protectors to be Worn at	
None	11 (21.6)
Some	22 (43.4)
All	18 (35.3)

[1]Hip Impact Protection Project.

[2]Represents total N(%).

Table A3.3. Intent-to-treat (ITT) and inverse probability of censoring weighted (IPCW) analysis of 1445 men and women randomized to wear hip protectors on either the left hip or the right hip (HIP PRO[1] Study 2).

Analysis	Hip Status	No. Hip Fractures	No. Bi-weeks	Rate Ratio	95% Confidence Interval
ITT	Protected	17	38129	1.13	(0.57, 2.27)
	Unprotected	15	38129	1	-
IPCW3[2]	Protected	1.9	6415	0.75	(0.17, 3.36)
	Unprotected	15	38129	1	-

[1]Hip Impact Protection Project.

[2]Individuals censored in first bi-week in which adherence is less than 75%.

145

Figure A3.1. Forest plot of rate ratios for incidence of hip fractures in two hip protector studies conducted by the Hip Impact Protection Project (HIP PRO)
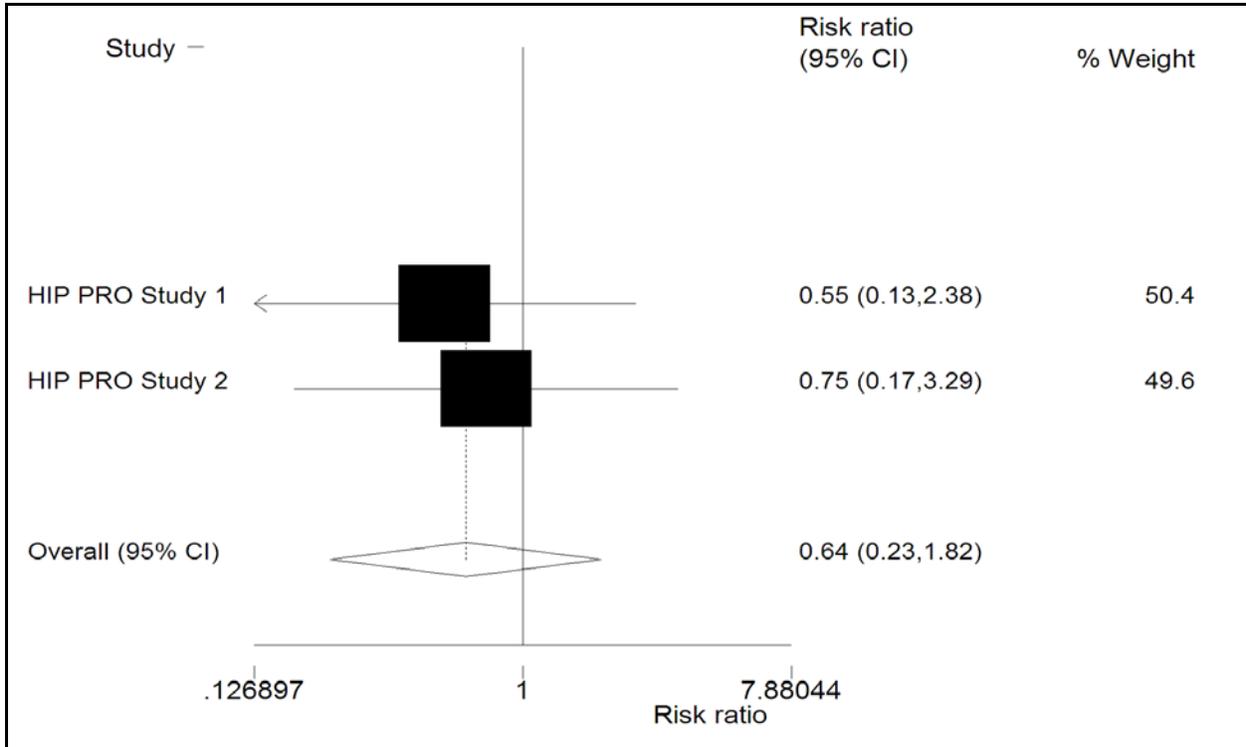
Appendix 4

A Simulation Exercise Comparing Results from an Intention-to-Treat (ITT) Analysis, an Instrumental Variable (IV) Analysis, and an Analysis Using Structural Nested Models (SNMs)

Briefly, we created J=1000 trials, each comprised of I=1000 patients, for each of K=18 simulation scenarios. The K scenarios varied the composition of the study population, which we conceptualized as being comprised of four types: 'would-be compliers,' 'always-takers,' 'never-takers,' and 'defiers.' Specifically, the first six scenarios used study populations comprised of only 'would-be compliers' and 'always-takers'; scenarios 6-12 used study populations comprised only of 'would-be compliers' and 'never-takers'; scenarios 13-18 had only 'would-be compliers' and 'defiers.' Within each trial j, a simulated data record for each individual i was comprised of three variables X, Z, and T. For context, X represented the randomization variable and was generated with a Bernoulli random variable and took values of 0 and 1, with probability 0.5, and Z represented treatment actually received and was dependent on an individual's compliance 'type' (e.g., for a 'would-be complier' treatment received, Z, is equal to treatment assigned, X, while for a defier treatment received is equal to the opposite of treatment assigned). Finally, T was a time-to-event. Note that the actual effect of treatment on T depends on treatment received, not on treatment assignment (i.e., the exclusion restriction holds). For the ITT analysis, we use a linear regression model to estimate the effect of randomization, X, on the time to an event. A linear model

is appropriate as there was no censoring. In contrast, we implement a two-stage least squares algorithm for the IV analysis by first regressing X on Z and then regressing T on the predicted value of X obtained from the first linear regression. Finally, we use g-estimation (i.e., logistic regression coupled with a grid search), as has been previously described, to estimate the factor by which assignment to treatment extends or contracts the average time to the outcome.

The results of the simulations are presented next. The simulations confirm some of the expected findings based on the theory behind the ITT, IV, and SNM approaches. First, note that in each scenario of the first set (i.e., 1-6), as the proportion of non-adherent participants increases, the ITT effect estimate is biased further toward the null (the effect estimate should be compared with the 'truth,' which is an effect on the risk difference scale of 2). A non-adherent participant can be an 'always-taker,' a 'never-taker,' or a 'defier'. Also, in the first two sets of scenarios, the adherence-correction methods only work when there is a nonzero group of adherent patients. When there are no adherent patients, the confidence intervals of the adherence corrected effect estimates become untenable. Finally, in the third set of scenarios (i.e., 13-18), the IV and SNM methods also fail to correct for non-adherence when there exist any defiers. Neither adherence-corrected method is able to recover the true risk difference, regardless of the proportion of participants that are 'defiers' (i.e., participants who always do the opposite of what they are assigned). Fortunately, individuals who perversely do the opposite of what they are assigned should be rare in real world

settings and the assumption of no defiers should readily be met. For the remainder of the discussion, we focus on the first two sets of scenarios (i.e., 1-12).

Regardless of whether a non-adherent participant is an 'always-taker' or a 'never-taker', the IV and SNM methods are always able to recover the true effect of treatment, as long as there are some 'would-be compliers' in the study population. Note, however, that the precision of the effect estimates is reduced as the proportion of non-adherent participants increase. Recall that our original purpose for undertaking the simulation exercise was to determine the target population for an SNM analysis. In this simulation, under the assumptions outlined earlier, the target population, for both the IV method and the SNM method, appears to be the same, namely the 'would-be compliers.' This similarity in the target population is only indirectly confirmed, however, based on the fact that the adherence-corrected effect estimates are similar. As noted previously, only a derivation mapping the estimand for the IV method to the estimand for the SNM method can resolve the question of whether the target population for an SNM method is the same as the IV method. While questions remain, this simulation exercise reveals that both the IV method and the SNM method appropriately correct for non-adherence. Also, on the risk difference scale, both methods produce similar effect estimates

REFERENCES

1. Mautalen CA, Vega EM, Einhorn TA. Are the etiologies of cervical and trochanteric hip fractures different? *Bone.* Mar 1996;18(3 Suppl):133S-137S.

2. Jensen J, Lundin-Olsson L, Nyberg L, Gustafson Y. Fall and injury prevention in older people living in residential care facilities - A cluster randomized trial. *Annals of Internal Medicine.* May 21 2002;136(10):733-741.

3. NCHS. Trends in Health and Aging. 2008.

4. Cummings SR, Rubin SM, Black D. The future of hip fractures in the United States. Numbers, costs, and potential effects of postmenopausal estrogen. *Clin Orthop Relat Res.* Mar 1990(252):163-166.

5. Close JC. Interdisciplinary practice in the prevention of falls--a review of working models of care. *Age Ageing.* Nov 2001;30 Suppl 4:8-12.

6. Rubenstein L. Hip protectors - A breakthrough in fracture prevention. *N Engl J Med.* Nov 23 2000;343(21):1562-1563.

7. Waldegger L, Cranney A, Man-Son-Hing M, Coyle D. Cost-effectiveness of hip protectors in institutional dwelling elderly. *Osteoporos Int.* May 2003;14(3):243-250.

8. Wolinsky FD, Fitzgerald JF, Stump TE. The effect of hip fracture on mortality, hospitalization, and functional status: a prospective study. *Am J Public Health.* Mar 1997;87(3):398-403.

9. Campbell AJ, Robertson MC, Gardner MM, Norton RN, Buchner DM. Falls prevention over 2 years: a randomized controlled trial in women 80 years and older. *Age Ageing.* Oct 1999;28(6):513-518.

10. Wallace RB, Ross JE, Huston JC, Kundel C, Woodworth G. Iowa Ficsit Trial - the Feasibility of Elderly Wearing a Hip-Joint Protective Garment to Reduce Hip-Fractures. *Journal of the American Geriatrics Society.* Mar 1993;41(3):338-340.

11. Greenspan SL, Myers ER, Kiel DP, Parker RA, Hayes WC, Resnick NM. Fall direction, bone mineral density, and function: risk factors for hip fracture in frail nursing home elderly. *Am J Med.* Jun 1998;104(6):539-545.

12. Bonjour JP, Schurch MA, Rizzoli R. Nutritional aspects of hip fractures. *Bone.* Mar 1996;18(3 Suppl):139S-144S.

13. Kiel DP, Felson DT, Anderson JJ, Wilson PWF, Moskowitz MA. Hip fracture and the use of estrogens in postmenopausal women. *N Engl J Med.* 1987;317:1169-1174.

14. Williams AR, Weiss NS, Ure CL, Ballard J, Daling JR. Effect of weight, smoking, and estrogen use on the risk of hip and forearm fractures in postmenopausal women. *Obstet Gynecol.* 1982;60:695-699.

15. Becker C, Kron M, Lindemann U, et al. Effectiveness of a multifaceted intervention on falls in nursing home residents. *J Am Geriatr Soc.* Mar 2003;51(3):306-313.

16. Kiel DP, Magaziner J, Zimmerman S, et al. Efficacy of a hip protector to prevent hip fracture in nursing home residents - The HIPPRO randomized controlled trial. *JAMA.* Jul 25 2007;298(4):413-422.

17. Melton LJ, 3rd. Epidemiology of hip fractures: implications of the exponential increase with age. *Bone.* Mar 1996;18(3 Suppl):121S-125S.

18. Gullberg B, Johnell O, Kanis JA. World-wide projections for hip fracture. *Osteoporos Int.* 1997;7(5):407-413.

19. Kannus P, Parkkari J, Sievanen H, Heinonen A, Vuori I, Jarvinen M. Epidemiology of hip fractures. *Bone.* Jan 1996;18(1 Suppl):57S-63S.

20. Rehnberg L, Nungu S, Olerud C. The incidence of femoral nect fractures in women is decreasing. *Acta Orthopaedica Scandinavica.* 1992;63((suppl 248)):92-93.

21. Keene GS, Parker MJ, Pryor GA. Mortality and morbidity after hip fractures. *Bmj.* Nov 13 1993;307(6914):1248-1250.

22. Louis M. Falls and their causes. *J Gerontol Nurs.* Mar 1983;9(3):143-149, 156.

23. Trombetti A, Herrmann F, Hoffmeyer P, Schurch MA, Bonjour JP, Rizzoli R. Survival and potential years of life lost after hip fracture in men and age-matched women. *Osteoporos Int.* Sep 2002;13(9):731-737.

24. Jacobsen SJ, Goldberg J, Miles TP, Brody JA, Stiers W, Rimm AA. Race and sex differences in mortality following fracture of the hip. *Am J Public Health.* Aug 1992;82(8):1147-1150.

25.   Cumming RG, Nevitt MC, Cummings SR. Epidemiology of hip fractures. *Epidemiol Rev.* 1997;19(2):244-257.

26.   Mossey JM, Mutran E, Knott K, Craik R. Determinants of recovery 12 months after hip fracture: the importance of psychosocial factors. *Am J Public Health.* Mar 1989;79(3):279-286.

27.   Grimley E, Prudham D, Wandless I. Prospective study of fractured proximal femur: Incidence and Outcome. *Publ Health Lond.* 1979;93:235-241.

28.   Davidson CW, Merrilees MJ, Wilkinson TJ, McKie JS, Gilchrist NL. Hip fracture mortality and morbidity--can we do better? *N Z Med J.* Jul 27 2001;114(1136):329-332.

29.   Lauritzen JB. Hip fractures: incidence, risk factors, energy absorption, and prevention. *Bone.* Jan 1996;18(1 Suppl):65S-75S.

30.   Slemenda C. Prevention of hip fractures: risk factor modification. *Am J Med.* Aug 18 1997;103(2A):65S-71S; discussion 71S-73S.

31.   Hayes WC, Myers ER, Morris JN, Gerhart TN, Yett HS, Lipsitz LA. Impact near the hip dominates fracture risk in elderly nursing home residents who fall. *Calcif Tissue Int.* Mar 1993;52(3):192-198.

32.   Cummings SR, Nevitt MC. A hypothesis: the causes of hip fractures. *J Gerontol.* Jul 1989;44(4):M107-111.

33.   Parkkari J, Kannus P, Palvanen M, et al. Majority of hip fractures occur as a result of a fall and impact on the greater trochanter of the femur: a prospective controlled hip fracture study with 206 consecutive patients. *Calcif Tissue Int.* Sep 1999;65(3):183-187.

34.   Kanis JA, McCloskey EV. Evaluation of the risk of hip fracture. *Bone.* Mar 1996;18(3 Suppl):127S-132S.

35.   Lauritzen JB, Schwarz P, Lund B, McNair P, Transbol I. Changing incidence and residual lifetime risk of common osteoporosis-related fractures. *Osteoporos Int.* May 1993;3(3):127-132.

36.   Melton LJ, 3rd. Hip fractures: a worldwide problem today and tomorrow. *Bone.* 1993;14 Suppl 1:S1-8.

37. Cummings SR, Black DM, Rubin SM. Lifetime risks of hip, Colles', or vertebral fracture and coronary heart disease among white postmenopausal women. *Arch Intern Med.* Nov 1989;149(11):2445-2448.

38. Ensrud KE, Lipschutz RC, Cauley JA, et al. Body size and hip fracture risk in older women: a prospective study. Study of Osteoporotic Fractures Research Group. *Am J Med.* Oct 1997;103(4):274-280.

39. Cummings SR, Nevitt MC, Browner WS, et al. Risk factors for hip fracture in white women. Study of Osteoporotic Fractures Research Group. *N Engl J Med.* Mar 23 1995;332(12):767-773.

40. Schindler AE, Ebert A, Friedrich E. Conversion of androstenedione to estrone by human tissue. *J Clin Endocrinol Metab.* Oct 1972;35(4):627-630.

41. Wilkinson TJ, Sainsbury R. Hip protectors. *Age Ageing.* Mar 1998;27(2):89-90.

42. Lauritzen JB, Askegaard V. Protection against hip fractures by energy absorption. *Dan Med Bull.* Feb 1992;39(1):91-93.

43. Schwartz AV, Kelsey JL, Sidney S, Grisso JA. Characteristics of falls and risk of hip fracture in elderly men. *Osteoporos Int.* 1998;8(3):240-246.

44. Cummings S, Cauley J, Palermo L, et al. Racial differences in hip axis length might explain racial differences in rates of hip fracture. *Osteoporos Int.* 1993;4:226-229.

45. Campbell AJ, Borrie MJ, Spears GF. Risk factors for falls in a community-based prospective study of people 70 years and older. *J Gerontol.* Jul 1989;44(4):M112-117.

46. Nevitt MC, Cummings SR, Kidd S, Black D. Risk factors for recurrent nonsyncopal falls. A prospective study. *Jama.* May 12 1989;261(18):2663-2668.

47. Tinetti ME, Speechley M, Ginter SF. Risk factors for falls among elderly persons living in the community. *N Engl J Med.* Dec 29 1988;319(26):1701-1707.

48. Meyer HE, Tverdal A, Falch JA. Risk factors for hip fracture in middle-aged Norwegian women and men. *Am J Epidemiol.* Jun 1 1993;137(11):1203-1211.

49. Bean N, Bennett KM, Lehmann AB. Habitus and hip fracture revisited: skeletal size, strength and cognition rather than thinness? *Age Ageing.* Nov 1995;24(6):481-484.

50. Grisso JA, Kelsey JL, Strom BL, et al. Risk factors for falls as a cause of hip fracture in women. The Northeast Hip Fracture Study Group. *N Engl J Med.* May 9 1991;324(19):1326-1331.

51. Rubenstein LZ. Falls in the elderly: a clinical approach. *West J Med.* Feb 1983;138(2):273-275.

52. Johnell O, Melton LJ, 3rd, Atkinson EJ, O'Fallon WM, Kurland LT. Fracture risk in patients with parkinsonism: a population-based study in Olmsted County, Minnesota. *Age Ageing.* Jan 1992;21(1):32-38.

53. Buchner DM, Larson EB. Falls and fractures in patients with Alzheimer-type dementia. *Jama.* Mar 20 1987;257(11):1492-1495.

54. Morris JC, Rubin EH, Morris EJ, Mandel SA. Senile dementia of the Alzheimer's type: an important risk factor for serious falls. *J Gerontol.* Jul 1987;42(4):412-417.

55. Cummings SR, Black DM, Nevitt MC, et al. Bone density at various sites for prediction of hip fractures. The Study of Osteoporotic Fractures Research Group. *Lancet.* Jan 9 1993;341(8837):72-75.

56. Riggs BL, Wahner HW, Seeman E, et al. Changes in bone mineral density of the proximal femur and spine with aging. Differences between the postmenopausal and senile osteoporosis syndromes. *J Clin Invest.* Oct 1982;70(4):716-723.

57. Hawker G. Epidemiology of arthritis and osteoporosis. In: Bradley E, Williams JI, eds. *Patterns of health care in Ontario: arthritis and related conditions.* Newy Your: Institute for Clinical Evaluative Sciences; 1998:1-10.

58. Cumming RG, Cummings SR, Nevitt MC, et al. Calcium intake and fracture risk: results from the study of osteoporotic fractures. *Am J Epidemiol.* May 15 1997;145(10):926-934.

59. Reid IR, Ames RW, Evans MC, Gamble GD, Sharpe SJ. Long-term effects of calcium supplementation on bone loss and fractures in postmenopausal women: a randomized controlled trial. *Am J Med.* Apr 1995;98(4):331-335.

60. Chapuy MC, Arlot ME, Duboeuf F, et al. Vitamin D3 and calcium to prevent hip fractures in the elderly women. *N Engl J Med.* Dec 3 1992;327(23):1637-1642.

61. Isley WL, Underwood LE, Clemmons DR. Dietary components that regulate serum somatomedin-C concentrations in humans. *J Clin Invest.* Feb 1983;71(2):175-182.

62. Johnell O, Nilsson B, Obrant K, Sernbo I. Age and sex patterns of hip fracture--changes in 30 years. *Acta Orthop Scand.* Jun 1984;55(3):290-292.

63. Lauritzen JB, Lund B. Risk of hip fracture after osteoporosis fractures. 451 women with fracture of lumbar spine, olecranon, knee or ankle. *Acta Orthop Scand.* Jun 1993;64(3):297-300.

64. Close J, Ellis M, Hooper R, Glucksman E, Jackson S, Swift C. Prevention of falls in the elderly trial (PROFET): a randomised controlled trial. *Lancet.* Jan 9 1999;353(9147):93-97.

65. Butler M, Norton R, Lee-Joe T, Cheng A, Campbell AJ. The risks of hip fracture in older people from private homes and institutions. *Age Ageing.* Sep 1996;25(5):381-385.

66. Baudoin C, Fardellone P, Bean K, Ostertag-Ezembe A, Hervy F. Clinical outcomes and mortality after hip fracture: a 2-year follow-up study. *Bone.* Mar 1996;18(3 Suppl):149S-157S.

67. Norton R, Campbell AJ, Reid IR, et al. Residential status and risk of hip fracture. *Age Ageing.* Mar 1999;28(2):135-139.

68. Ray WA, Griffin MR, Downey W. Benzodiazepines of long and short elimination half-life and the risk of hip fracture. *Jama.* Dec 15 1989;262(23):3303-3307.

69. Ray WA, Griffin MR, Malcolm E. Cyclic antidepressants and the risk of hip fracture. *Arch Intern Med.* Apr 1991;151(4):754-756.

70. Shorr RI, Griffin MR, Daugherty JR. Opiod analgesics and the risk of hip fracture in the elderly: codeine and propoxyphene. *J Gerontol.* 1992;47:M111-115.

71. Campbell AJ, Borrie MJ, Spears GF, Jackson SL, Brown JS, Fitzgerald JL. Circumstances and consequences of falls experienced by a community population 70 years and over during a prospective study. *Age Ageing.* Mar 1990;19(2):136-141.

72. Bergland A, Wyller TB. Risk factors for serious fall related injury in elderly women living at home. *Inj Prev.* Oct 2004;10(5):308-313.

73. Cryer C, Knox A, Martin D, Barlow J. Hip protector compliance among older people living in residential care homes. *Inj Prev.* Sep 2002;8(3):202-206.

74. Hayes WC, Myers ER, Robinovitch SN, Van Den Kroonenberg A, Courtney AC, McMahon TA. Etiology and prevention of age-related hip fractures. *Bone.* Jan 1996;18(1 Suppl):77S-86S.

75. Courtney AC, Wachtel EF, Myers ER, Hayes WC. Effects of loading rate on strength of the proximal femur. *Calcif Tissue Int.* Jul 1994;55(1):53-58.

76. Cameron ID. Hip protectors - Prevent fractures but adherence is a problem. *British Medical Journal.* Feb 16 2002;324(7334):375-376.

77. Forsen L, Arstad C, Sandvig S, Schuller A, Roed U, Sogaard AJ. Prevention of hip fracture by external hip protectors: an intervention in 17 nursing homes in two municipalities in Norway. *Scand J Public Health.* 2003;31(4):261-266.

78. Wiener SL, Andersson GBJ, Nyhus LM, Czech J. Force reduction by an external hip protector on the human hip after falls. *Clinical Orthopaedics and Related Research.* May 2002(398):157-168.

79. Sawka AM, Boulos P, Beattie K, et al. Do hip protectors decrease the risk of hip fracture in institutional and community-dwelling elderly? A systematic review and meta-analysis of randomized controlled trials. *Osteoporos Int.* Dec 2005;16(12):1461-1474.

80. Butler M, Norton R, Lee-Joe T, Eng M, Coggan C. Preventing falls and fall-related injuries among older people living in institutions: current practice and future opportunities. *New Zealand Medical Journal.* Sep 25 1998;111(1074):359-361.

81. Myers AH, Michelson JD, Vannatta M, Cox Q, Jinnah R. Prevention of Hip-Fractures in the Elderly - Receptivity to Protective Garments. *Archives of Gerontology and Geriatrics.* Sep-Oct 1995;21(2):179-189.

82. Hubacher M, Wettstein A. Acceptance of hip protectors for hip fracture prevention in nursing homes. *Osteoporos Int.* 2001;12(9):794-799.

83.	O'Halloran PD, Cran GW, Beringer TRO, et al. A cluster randomised controlled trial to evaluate a policy of making hip protectors available to residents of nursing homes. *Age and Ageing.* Nov 2004;33(6):582-588.

84.	van Schoor NM, Deville WL, Bouter LM, Lips P. Acceptance and compliance with external hip protectors: a systematic review of the literature. *Osteoporos Int.* Dec 2002;13(12):917-924.

85.	Becker C, Walter-Jung B, Nikolaus T. The other side of hip protectors. *Age Ageing.* Mar 2000;29(2):186.

86.	Thompson PW, Jones C. Adherence to hip protector use in elderly people requiring domicilary care is greater in fallers than non-fallers. *Age Ageing.* Sep 2000;29(5):459.

87.	Villar MTA, Hill P, Inskip H, Thompson P, Cooper C. Will elderly rest home residents wear hip protectors? *Age and Ageing.* Mar 1998;27(2):195-198.

88.	Lauritzen JB, Petersen MM, Lund B. Effect of External Hip Protectors on Hip-Fractures. *Lancet.* Jan 2 1993;341(8836):11-13.

89.	Kannus P, Parkkari J, Niemi S, et al. Prevention of hip fracture in elderly people with use of a hip protector. *N Engl J Med.* Nov 23 2000;343(21):1506-1513.

90.	Harada A, Matsui Y, Mizuno M, Tokuda H, Niino N, Ohta T. Japanese orthopedists' interests in prevention of fractures in the elderly from falls. *Osteoporosis International.* Jul 2004;15(7):560-566.

91.	Harada A, Mizuno M, Takemura M, Tokuda H, Okuizumi H, Niino N. Hip fracture prevention trial using hip protectors in Japanese nursing homes. *Osteoporos Int.* 2001;12(3):215-221.

92.	Cameron ID, Kurrle SE, Cumming RG, Quine S. Proximal femoral fracture while wearing correctly applied hip protectors. *Age and Ageing.* Jan 2000;29(1):85-86.

93.	Cameron ID, Venman J, Kurrle SE, et al. Hip protectors in aged-care facilities: a randomized trial of use by individual higher-risk residents. *Age Ageing.* Nov 2001;30(6):477-481.

94.	van Schoor NM, Smit JH, Twisk JWR, Bouter LM, Lips P. Prevention of hip fractures by external hip protectors - A randomized controlled trial. *JAMA.* Apr 16 2003;289(15):1957-1962.

95. Birks YF, Hildreth R, Campbell P, Sharpe C, Torgerson DJ, Watt I. Randomised controlled trial of hip protectors for the prevention of second hip fractures. *Age Ageing.* Jul 2003;32(4):442-444.

96. Cameron ID, Cumming RG, Kurrle SE, et al. A randomised trial of hip protector use by frail older women living in their own homes. *Inj Prev.* Jun 2003;9(2):138-141.

97. van Schoor NM, Smit JH, Twisk JWR, Bouter LM, Lips P. External hip protectors and risk of hip fracture - Reply. *Jama-Journal of the American Medical Association.* Aug 20 2003;290(7):884-885.

98. Kurrle SE, Cameron ID, Quine S, Cumming RG. Adherence with hip protectors: a proposal for standardised definitions. *Osteoporosis International.* Jan 2004;15(1):1-4.

99. O'Halloran PD. Effectiveness of hip protectors - It's more the effectiveness of compliance. *British Medical Journal.* Mar 25 2006;332(7543):729-729.

100. van Schoor NM, Smit JH, Asma G, Bouter LM, Lips P. Effectiveness of external hip protectors in preventing hip fractures: A randomized clinical trial. *Journal of Bone and Mineral Research.* Sep 2002;17:S207-S207.

101. Parkkari J, Heikkila J, Kannus P. Acceptability and compliance with wearing energy-shunting hip protectors: a 6-month prospective follow-up in a Finnish nursing home. *Age and Ageing.* Mar 1998;27(2):225-229.

102. Forsen L, Sandvig S, Schuller A, Sogaard AJ. Compliance with external hip protectors in nursing homes in Norway. *Inj Prev.* Dec 2004;10(6):344-349.

103. Cameron ID, Cumming RG, Kurrle SE. Prevention of hip fracture with use of a hip protector. *N Engl J Med.* Mar 15 2001;344(11):855-856; author reply 856-857.

104. O'Halloran PD, Murray LJ, Cran GW, Dunlop L, Kernohan G, Beringer TRO. The effect of type of hip protector and resident characteristics on adherence to use of hip protectors in nursing and residential homes - an exploratory study. *International Journal of Nursing Studies.* May 2005;42(4):387-397.

105. Butler M, Coggan C, Norton R. A qualitative investigation into the receptivity to hip protective underwear among staff and residents of residential institutions. *N Z Med J.* Oct 9 1998;111(1075):383-385.

106. Hildreth R. Review: hip protectors reduce hip fractures after falls in elderly people living in institutions or supported home environments. *Evid Based Nurs.* Jan 2002;5(1):23.

107. van Schoor NM, Asma G, Smit JH, Bouter LM, Lips P. The Amsterdam Hip Protector Study: compliance and determinants of compliance. *Osteoporos Int.* Jun 2003;14(4):353-359.

108. Kurrle SE, Cameron ID, Quine S. Predictors of adherence with the recommended use of hip protectors. *J Gerontol A Biol Sci Med Sci.* Sep 2004;59(9):M958-961.

109. Parker MJ, Gillespie WJ, Gillespie LD. Effectiveness of hip protectors for preventing hip fractures in elderly people: systematic review. *BMJ.* Mar 11 2006;332(7541):571-573.

110. Ekman A, Mallmin H, Michaelsson N, Ljunghall S. External hip protectors to prevent osteoporotic hip fractures. *Lancet.* Aug 23 1997;350(9077):563-564.

111. Koike T, Orito Y, Tada M, et al. A cluster randomized controlled trial to evaluate the effects of hip protectors against risk factors for hip fracture in nursing homes. *Journal of Bone and Mineral Research.* Sep 2006;21:S293-S293.

112. Birks YF, Porthouse J, Addie C, et al. Randomized controlled trial of hip protectors among women living in the community. *Osteoporos Int.* Sep 2004;15(9):701-706.

113. Birge SJ, Nyhus LM, Wiener SL. Prevention of nursing home (NH) hip fractures with external hip protectors (EHP). *Gerontologist.* Oct 2001;41:298-298.

114. Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Stat Med.* May 15 2004;23(9):1351-1375.

115. Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics.* Jun 2000;56(2):455-463.

116. Jantti PO, Aho HJ, Maki-Jokela PL, Heikinheimo RJ. Hip protectors and hip fractures. *Age Ageing.* Nov 1998;27(6):758-759.

117.	Campbell MK, Grimshaw JM. Cluster randomised trials: time for improvement. The implications of adopting a cluster design are still largely being ignored. *Bmj.* Oct 31 1998;317(7167):1171-1172.

118.	Meyer G, Warnke A, Bender R, Muhlhauser I. Effect on hip fractures of increased use of hip protectors in nursing homes: cluster randomised controlled trial. *BMJ.* Jan 11 2003;326(7380):76.

119.	Chan DK, Hillier G, Coore M, et al. Effectiveness and acceptability of a newly designed hip protector: a pilot study. *Archives of Gerontology and Geriatrics.* Jan-Feb 2000;30(1):25-34.

120.	Hernan MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology.* Jul 2006;17(4):360-372.

121.	Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol.* Aug 2000;29(4):722-729.

122.	Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology.* Sep 2000;11(5):550-560.

123.	Robins JM, Blevins D, Ritter G, Wulfsohn M. G-estimation of the effect of prophylaxis therapy for Pneumocystis carinii pneumonia on the survival of AIDS patients. *Epidemiology.* Jul 1992;3(4):319-336.

124.	Cole SR, Chu H. Effect of acyclovir on herpetic ocular recurrence using a structural nested model. *Contemp Clin Trials.* Jun 2005;26(3):300-310.

125.	Zimmerman S, Magaziner J, Birge SJ, Barton BA, Kronsberg SS, Kiel DP. Adherence to hip protectors and implications for U.S. long-term care settings. *J Am Med Dir Assoc.* Feb 2010;11(2):106-115.

126.	Blessed G, Tomlinson BE, Roth M. The association between quantitative measures of dementia and of senile change in the cerebral grey matter of elderly subjects. *Br J Psychiatry.* 1968;114(512):797-811.

127.	Hernan MA. A definition of causal effect for epidemiological research. *J Epidemiol Community Health.* Apr 2004;58(4):265-271.

128.	Dunn G, Maracy M, Tomenson B. Estimating treatment effects from randomized clinical trials with noncompliance and loss to follow-up: the role of instrumental variable methods. *Stat Methods Med Res.* Aug 2005;14(4):369-395.

129. Korhonen PA, Laird NM, Palmgren J. Correcting for non-compliance in randomized trials: an application to the ATBC Study. *Stat Med.* Nov 15 1999;18(21):2879-2897.

130. Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Instrumental variables: application and limitations. *Epidemiology.* May 2006;17(3):260-267.

131. Cole SR, Hernan MA, Margolick JB, Cohen MH, Robins JM. Marginal structural models for estimating the effect of highly active antiretroviral therapy initiation on CD4 cell count. *Am J Epidemiol.* Sep 1 2005;162(5):471-478.

132. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics.* Sep 2000;56(3):779-788.

133. Hernan MA, Cole SR, Margolick JB, Cohen M, Robins JM. Structural accelerated failure time models for survival analysis in studies with time-varying treatments. *Pharmacoepidemiology and Drug Safety.* 2005;14:477-491.

134. Sommer A, Zeger SL. On Estimating Efficacy from Clinical Trials. *Stat Med.* 1991 1991;10:45-52.

135. Scheiner LB, Rubin DB. Intention-to-treat analysis and the goals of clinical trials. *Clinical Pharmacology & Therapeutics.* 1995;57(1):6-15.

136. Greenland S, Lanes S, Jara M. Estimating effects from randomized trials with discontinuations: the need for intent-to-treat design and G-estimation. *Clin Trials.* 2008;5(1):5-13.

137. Newhouse JP, McClellan M. Econometrics in outcomes research: the use of instrumental variables. *Annu Rev Public Health.* 1998;19:17-34.

138. Joffe MM. Administrative and artificial censoring in censored regression models. *Stat Med.* Aug 15 2001;20(15):2287-2304.

139. Cuzick J, Edwards R, Segnan N. Adjusting for non-compliance and contamination in randomized clinical trials. *Stat Med.* May 15 1997;16(9):1017-1029.

140. Angrist JD, Imbens GW, Rubin DB. Identification of Causal Effects using Instrumental Variables. *Journal of the American Statistical Association.* 1996;91(434):444-455.

141. Greenland S, Brumback B. An overview of relations among causal modelling methods. *Int J Epidemiol.* Oct 2002;31(5):1030-1037.

142. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology.* Jan 1999;10(1):37-48.

143. Marcus SM, Gibbons RD. Estimating the Efficacy of Receiving Treatment in Randomized Clinical Trials with Noncompliance. *Health Services & Outcomes Research Methodology.* 2001 2001;2:247-258.

144. Nagelkerke N, Fidler V, Bernsen R, Borgdorff M. Estimating treatment effects in randomized clinical trials in the presence of non-compliance. *Stat Med.* Jul 30 2000;19(14):1849-1864.

145. Cain LE, Cole SR, Greenland S, et al. Effect of highly active antiretroviral therapy on incident AIDS using calendar period as an instrumental variable. *Am J Epidemiol.* May 1 2009;169(9):1124-1132.

146. Siris ES, Selby PL, Saag KG, Borgstrom F, Herings RM, Silverman SL. Impact of osteoporosis treatment adherence on fracture rates in North America and Europe. *Am J Med.* Feb 2009;122(2 Suppl):S3-13.

147. Cain LE, Cole SR. Inverse probability-of-censoring weights for the correction of time-varying noncompliance in the effect of randomized highly active antiretroviral therapy on incident AIDS or death. *Stat Med.* May 30 2009;28(12):1725-1738.

148. Toh S, Hernan MA. Causal inference from longitudinal studis with baseline randomization. *Int J Biostat.* 2008;4(1):Article 22.

149. Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology.* Sep 2000;11(5):561-570.

150. Cook NR, Cole SR, Hennekens CH. Use of a marginal structural model to determine the effect of aspirin on cardiovascular mortality in the Physicians' Health Study. *Am J Epidemiol.* Jun 1 2002;155(11):1045-1053.

151. D'Agostino RB, Jr., Lee M, Belanger AJ. Relation of Pooled Logistic Regression to Time Dependent Cox Regression Analysis: The Framingham Heart Study. *Stat Med.* 1990;9:1501-1515.

152. Harrell FE, Jr. *Regression modeling strategies.* New York: Springer; 2001.

153. Cole SR, Hernan MA, Anastos K, al. e. Determining the effect of highly active antiretroviral therapy on changes in human immunodeficiency virus type 1 RNA viral load using marginal structural left-censored mean model. *Am J Epidemiol.* 2007;166:219-227.

154. Kiel DP, Birge SJ, Magaziner J, Zimmerman SI, Brown KM, Barton BA. A multicenter, randomized, controlled, trial of the efficacy of an energy absorbing hip protector in US nursing home (NH) residents for the prevention of hip fracture: HIPPRO. *J Am Geriatr Soc.* Apr 2006;54(4):S15-S16.

155. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Sturmer T. Variable selection for propensity score models. *Am J Epidemiol.* 2006;163(12):1149-1156.

156. Cole SR, Frangakis CE. The consistency statement in causal inference: a definition or an assumption? *Epidemiology.* Jan 2009;20(1):3-5.

157. Hernan MA, Taubman SL. Does obesity shorten life? The importance of well-defined interventions to causal questions. *Int J Obes.* 2008;32:S8-S14.

158. Hernan MA. Invited commentary: Hypothetical interventions to define causal effects - afterthought or prerequisite? *Am J Epidemiol.* 2005;162(7):618-620.

159. Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *J Am Stat Assoc.* 1995;90(429):106-121.

160. Matsuyama Y. A comparison of the results of intent-to-treat, per-protocol, and g-estimation in the presence of non-random treatment changes in a time-to-event non-inferiority trial. *Stat Med.* Jun 15 2010.

161. Cole SR, Hernan MA, Robins JM, et al. Effect of highly active antiretroviral therapy on time to acquired immunodeficiency syndrome or death using marginal structural models. *Am J Epidemiol.* Oct 1 2003;158(7):687-694.

162. Mark SD, Robins JM. A method for the analysis of randomized trials with compliance information: an application to the Multiple Risk Factor Intervention Trial. *Control Clin Trials.* Apr 1993;14(2):79-97.

163. Cox DR, Oakes D. *Analysis of Survival Data.* London: Chapman & Hall; 1984.

164.	Kalbfleishch JD, Prentice RL. *The Statistical Analysis of Failure Time Data.* New Jersey: John Wiley & Sons, Inc.; 2002.

165.	Wei LJ. The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Stat Med.* Oct-Nov 1992;11(14-15):1871-1879.

166.	Cole SR, Haitao C, Lei N. Nonparametric estimator of relative time with application to the Acyclovir Prevention Trial. *Clin Trials.* Aug 2009;6(4):320-328.

167.	Robins J, Tsiatis A. Correcting for non-compliance in randomized trials using rank perserving structural failure time models. *Commun. Statist Theory Methods.* 1991;20(8):23.

168.	Witteman JC, D'Agostino RB, Stijnen T, et al. G-estimation of causal effects: isolated systolic hypertension and cardiovascular death in the Framingham Heart Study. *Am J Epidemiol.* Aug 15 1998;148(4):390-401.

169.	Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* 1952;47:23.

170.	Hernan MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health.* 2006;60:578-586.

171.	Greenland S, Robins JM. Conceptual problems in the definition and interpretation of attributable fractions. *Am J Epidemiol.* Dec 1988;128(6):1185-1197.

172.	Hernan MA. The hazards of hazard ratios. *Epidemiology.* Jan 2010;21(1):13-15.

173.	Bandura A. *Social Learning Theory.* Englewood Cliffs, NJ: Prentice Hall; 1977.

174.	Rosenstock IM, Strecher VJ, Becker MH. Social learning theory and the Health Belief Model. *Health Educ Q.* Summer 1988;15(2):175-183.

175.	Janz NK, Becker MH, Hartman PE. Contingency contracting to enhance patient compliance: a review. *Patient Educ Couns.* 1984;5(4):165-178.

176.	Leventhal H, Cameron L. Behavioral theories and the problem of compliance. *Patient Educ Couns.* 1987;10:117-138.

177. Green CA. What can patient health education coordinators learn from ten years of compliance research? *Patient Educ Couns.* Oct 1987;10(2):167-174.

178. Koike T, Orito Y, Toyoda H, et al. Prevention of hip fracture in aged-care facilities with increased use of hip protectors in Japan. *Journal of Bone and Mineral Research.* Sep 2005;20(9):S291-S292.