Published in final edited form as:

Child Youth Serv Rev. 2010 January 1; 32(1): 103–112. doi:10.1016/j.chilgyouth.2009.07.020.

Using risk adjustment approaches in child welfare performance measurement: Applications and insights from health and mental health settings

Ramesh Raghavan*

Washington University in St. Louis, Campus Box 1196, One Brookings Dr, St. Louis, MO 63130, United States

Abstract

Federal policymaking in the last decade has dramatically expanded performance measurement within child welfare systems, and states are currently being fiscally penalized for poor performance on defined outcomes. However, in contrast to performance measurement in health settings, current policy holds child welfare systems solely responsible for meeting outcomes, largely without taking into account the effects of factors at the level of the child, and his or her social ecology, that might undermine the performance of child welfare agencies. Appropriate measurement of performance is predicated upon the ability to disentangle individual, as opposed to organizational, determinants of outcomes, which is the goal of risk adjustment methodologies. This review briefly conceptualizes and examines risk adjustment approaches in health and child welfare, suggests approaches to expanding its use to appropriately measure the performance of child welfare agencies, and highlights research gaps that diminish the appropriate use of risk adjustment approaches – and which consequently suggest the need for caution – in policymaking around performance measurement of child welfare agencies.

Keywords

Child welfare; Performance measurement; Quality improvement; Mental health; Performance assessment

1. Introduction

What factors are responsible for the variations observed between, say, hospitals when it comes to patient mortality? Are these variations solely due to the characteristics of hospitals themselves – in other words, are there “bad” hospitals and “good” hospitals, the former producing high rates of patient deaths and the latter producing low rates? Or, is part of the variation in mortality due to the characteristics of the patients treated within these hospitals? In other words, are there hospitals that treat sicker patients (such as cancer hospitals) and

those that treat relatively healthy patients (such as obstetric hospitals), and do variations in who goes into these hospitals account for the variations in mortality rates between them?

The answer to this question is of great importance to health planners and policymakers. Clearly, the goal of policy is to identify “bad” hospitals from “good” hospitals, closely monitor the former, and incentivize the latter. But the overriding goal of policy is to do this *fairly*. If variations between hospitals are solely due to the characteristics of patients found in these hospitals (cancer diagnosis, in the above example), then even the best cancer hospital will have higher mortality rates than the worst obstetric hospital. If we are to appropriately measure hospital performance, then we need to separate the relative contributions of hospitals from the contributions of the patients they see.

The need to disentangle organizational determinants of outcomes from client-level determinants of outcomes was recognized as far back as the middle of the 19th century. In 1856, with a reputation firmly established in the aftermath of the Crimean War, Florence Nightingale returned to London and began to inquire into the causes of hospital death rates. Nightingale soon realized that death rates were highest in hospitals with poor sanitation and crowded wards. But she also observed that differences in patient age, and the “state of the cases on admission” were important determinants of mortality (Nightingale & Munding, 1999). Patients with these characteristics, whom she called “special inmates,” were disproportionately represented in London hospitals with the highest death rates.

Since that time, attempts to identify the characteristics of patients that account for differences in outcomes between hospitals have led to the development of two methodologically related, but conceptually distinct, methodologies. *Risk adjustment* refers to a group of techniques used to predict outcomes (such as service use, or costs of care) based on patient or client characteristics (Blumenthal et al., 2005; Iezzoni, 2003). The level of observation is the individual patient (client), and the goal is to identify client-level characteristics that account for outcomes. *Case-mix adjustment* refers to identifying and adjusting for the differences in clinical presentation (such as diagnoses) treated by different hospitals so that hospital-level costs and outcomes are comparable (Jencks & Dobson, 1987). The level of observation is the hospital or health facility, and variables usually include not only the patient-level characteristics that are controlled for in risk adjustment approaches, but also hospital-level characteristics. The overall goal is to control for the fact that variations in patient and hospital characteristics may cause even efficient hospitals to seem high-cost compared to others. Case-mix adjustment is also used to profile providers whose performance needs to be adjusted for the mix of patients they see. Case-mix adjustment is often based on risk adjustment – a facility’s observed performance is sometimes compared to its risk adjusted expected performance (McMillen, Lee, & Jonson-Reid, 2008). The magnitude of the difference between these two is expressed in the form of a report card (performing better than expected, for example). Because both approaches share a similar methodology and differ only in explanatory variables and level of analysis, the rest of this review focuses on risk adjustment.

Risk adjustment approaches are currently used in several different ways. Researchers use risk adjustment to set insurance premium rates for health plans, which is perhaps the

dominant use of risk adjustment approaches. To do so, they use methodologies that can be implemented by the plans themselves, rely only on variables in existing administrative datasets, and take incentives and equity issues into account (Blumenthal et al., 2005; Iezzoni, 2003). An example of an equity issue is race/ethnicity – although race/ethnicity is an established predictor of service use (Harris, Gordon-Larsen, Chantala, & Udry, 2006) – risk adjustment studies do not use this variable to predict costs. To do so would be to reimburse health plans differentially depending on the racial and ethnic distribution of their enrollees, which in turn might worsen race/ethnic disparities in enrollment into health plans. Risk adjustment methodologies also do not use variables that health plans could control, such as past use of services, because plans have the potential ability to game the payment system if past use is used to set reimbursement rates.

In 1984 Medicare became the first payor to adopt a risk adjustment approach called Diagnosis Related Groups (Weissman, Wachterman, & Blumenthal, 2005), adopting a technique of using categories derived from beneficiaries' diagnostic information to predict costs of care (and, therefore, Medicare's payments to health plans and providers) (Vladeck, 1984). The past two decades have seen an explosion in the use of publicly available as well as proprietary systems that are designed to predict health care utilization, costs, and outcomes (Iezzoni, 2003).

More recently, these approaches are used in *value-based purchasing* (Deas, 2006) – which describes health care purchasers preferentially contracting with health plans that offer greater value rather than merely lower cost. Payors can compare outcomes across health plans, use risk adjustment methods to see if, after controlling for patient characteristics, some plans are 'better' (good outcomes, lower cost) than others. They can then prefer to contract with these high-value plans. Risk adjustment also informs *pay-for-performance* (Rosenthal & Dudley, 2007) – which involves tying fiscal and non-fiscal rewards and punishments to a variety of performance outcomes, such as health outcomes, patient satisfaction, scores on quality scorecards, screening rates, prescribing practices, adherence to clinical guidelines, and investments in information technology, among others. Clearly, any such performance-based contracting needs to account for differences in client characteristics; absent such adjustment, health plans have the incentive to game the system by systematically disenrolling sick patients. In theory, such gaming is prevented by using risk adjustment.

But while risk adjustment approaches enjoy widespread use as instruments to promote the performance of health systems, there is little evidence that they are being used to promote the performance of child welfare systems. In a previous publication in the *Review*, Courtney and colleagues highlighted the limitation that current performance measurement through the Child and Family Services Reviews process is not risk adjusted, and suggested the use of demographic characteristics as a first step in risk-adjusting outcomes (Courtney, Needell, & Wulczyn, 2004). Child welfare policymaking today seems to be largely uninformed by their suggestion. In an attempt to increase the appropriateness of child welfare performance measurement, this paper develops principles and applications for the use of risk adjustment techniques in child welfare systems. Following a review of the basic principles of risk adjustment, examples of their application within health and mental health are presented.

This paper ends with suggestions for translating such learning from health settings to child welfare settings, and overviews the challenges that need to be overcome in order to adapt and develop risk adjustment approaches to child welfare performance measurement.

2. Background

2.1. The conduct of risk adjustment: selected examples from health and mental health

Risk adjustment is essentially the systematic study of variation. If the factors that contribute to client-level variations in some outcome of interest can be identified, then their effects on outcomes can be understood. This understanding can lead to ways to manage these variations, either statistically or through administrative mechanisms.

Statistically, risk adjustment methodologies identify and control for client-level variables that affect some outcome of interest. In order to do this, risk adjustment approaches need to also focus on two additional developmental considerations. First, the ‘window of observation,’ or the time period over which these outcomes may occur; and, second, the purpose for which these approaches are being developed – to set capitation rates for health plans, to perform case-mix adjustment, to profile providers based on their clients’ outcomes, to set reimbursement rates for hospitals, and so on. Answers to these questions, then, are critical to the development of risk adjustment approaches (Iezzoni, 2003).

Client-level predictors of many outcomes can be classified into several domains. Sociodemographic characteristics (age, gender, race/ ethnicity) are universally used in risk adjustment methodologies. Diagnosis, as obtained from claims in the form of ICD-9 codes or recoded based on chart information forms the second domain of client-level predictors. Over the past decades, the relative lack of concordance between diagnosis and outcomes have been increasingly recognized. This happens when, for example, clients with identical diagnoses and demographic characteristics exhibit different disease progression and outcomes over time. To deal with this problem, researchers have developed actuarial combinations of demographic characteristics and diagnoses (including comorbid diagnoses) that can provide greater predictive power than the use of separate demographic and diagnostic predictors.

For example, one of the widest used methodologies is called Adjusted Clinical Groups (ACG) (Johns Hopkins University, 2007), which has been used with modest success with Medicaid beneficiaries (Adams, Bronstein, & Raskind-Hood, 2002). The core data elements of the ACG system consist of the patient’s gender, age, and ICD diagnoses assigned to them over a given year. Because many adults have more than one diagnosis, and because certain combinations of diagnoses may have worse outcomes than other combinations, these diagnoses can be aggregated into groups, and ranked according to the potential for producing adverse outcomes. To do this, the ACG system first develops 32 different diagnostic groupings called Aggregated Diagnostic Groups (ADG). Examples of these ADGs include “time-limited minor conditions,” “recurrent or persistent, unstable psychosocial conditions” and even a group of “see and reassure conditions.” These 32 ADG groups are then combined with age, gender, and other specialized predictors for newborns and pregnant women to form 93 Adjusted Clinical Groups. Examples of ACGs include “4–5

different (not elsewhere categorized) ADGs, two or more of which include at least one serious condition”; “one or more acute minor conditions only: age 2–5”; and “pregnant, not delivered, with 2–3 different ADGs, of which one or more has a serious or major condition.” A third approach within the ACG system is to aggregate all similar conditions together into Expanded Diagnosis Clusters (EDCs). An example of an EDC is “surgical,” which includes all “Dental, ENT, Eye, General Surgery, Genito-urinary, Musculoskeletal, Reconstructive” diagnoses. (Adams et al., 2002) These ACGs are combined with patient characteristics and, in some cases, from chart information to develop predictive models for patient outcomes.

Medicare’s use of the DRG, described previously, was the earliest attempt to aggregate diagnostic and client information into predictive clusters. The DRG system, much like the ACG system, classifies patients into groups such as “psychoses” (DRG 430) and “vaginal delivery” (Centers for Medicare and Medicaid Services, 2006). Today much of risk adjustment occurs using these ‘systems’ instead of simple diagnoses. In addition, there are over a dozen discrete risk adjustment methodologies, some publicly available and some proprietary, that are used by health plans and hospitals to better predict client outcomes.

Client outcomes of interest ultimately determine use of one particular risk adjustment approach over another. For example, the DRG of “psychoses” is probably a good measure of utilization, but a poor measure of cost. This is because psychoses cover a range of illnesses from affective disorders with psychotic symptoms through to schizophrenia, and require a range of management strategies from antipsychotic drugs and psychoeducation to (expensive) electroconvulsive therapy. What risk adjustment measure to use for what outcome of interest is still an empirical question, and has resulted in the specialization of risk adjustment approaches to predict one type of outcome over another. For example, an approach called the Chronic Illness and Disability Payment System (CDPS) is a methodology provided free of charge to public agencies, the purpose of which is to help Medicaid agencies to better set capitation rates by giving them more information on health expenditures (Kronick, Gilmer, Dreyfus, & Lee, 2000). The CDPS system divides health conditions into 19 diagnostic categories, and assigns cost rankings within each category. For example, the group called “psychiatric” is divided into 3 expenditure groups – high (schizophrenia), medium (bipolar disorder), and low (other depression, panic disorder).

In addition to clinical outcomes (symptom remission), risk adjustment studies have attempted to predict a variety of different types of service utilization – e.g., inpatient lengths of stay, total number of inpatient days, inpatient nursing hours, inpatient readmissions, outpatient visits for specified disorders, and outpatient visits for all disorders (Hermann, Rollins, & Chan, 2007). These techniques are also applicable toward studying a variety of costs – health care costs, costs of mental health and substance use services, costs of disabilities, and costs of inpatient use. In their review, Hermann et al. (2007) examine studies reporting on a range of other outcomes, including functioning, symptom severity, satisfaction, and quality of life. Generally, risk adjustment methodologies are not as well developed for such client-derived outcomes as they are for utilization, cost, and clinical outcomes.

The *window of observation* also determines the choice of the risk adjustment methodology. CDPS is used in predictions of Medicaid expenditures based on the past year's worth of claims. Its predictions, then, are meant to be incorporated in the next year's planning process. In contrast, hospitals often require shorter time horizons – being concerned with costs during a given length of stay that may only last a few days, for example. In such cases, another risk adjustment approach such as ACGs might be preferred.

2.2. Performance of risk adjustment approaches

Among adult samples, patient factors seem to explain greater variance in the *costs of inpatient use* and *health care costs*. Models using the Global Assessment of Functioning and clinical status measures (diagnostic and chart information) predicted 25% of the variance in inpatient costs (Leslie, Rosenheck, & White, 2000); and approaches using another risk adjustment approach called the Diagnostic Cost Group/Hierarchical Coexisting Condition model, along with clinical diagnosis, demographic information (age, gender, sex) and clinical status (data from rating scales and patient interviews) predicted over 30% of the variance in health care costs (Rosen et al., 2002). In general, the best physical health risk adjustment approaches predict less than a third of the variance in outcomes.

More analogous to the situation in child welfare is the performance of mental health risk adjustment. In contrast to the development of standardized risk adjustment approaches in physical health, there are no standard approaches to risk adjust in mental health (Hendryx, Beigel, & Doucette, 2001). In a series of studies, Ettner and colleagues documented the underperformance of many of the existing risk adjustment approaches in predicting the *costs of behavioral health services* and found that no models predicted over 10% of the variance in costs of mental health services (Ettner, Frank, Mark, & Smith, 2000; Ettner, Frank, McGuire, & Hermann, 2001; Ettner, Frank, McGuire, Newhouse, & Notman, 1998; Ettner & Notman, 1997). Simpler models containing sociodemographic characteristics and past service use outperformed a risk adjustment approach called Ambulatory Care Groups (Starfield, Weiner, Mumford, & Steinwachs, 1991) (now called Adjusted Clinical Groups) in predicting costs of utilization among adults, though not among children (Ettner & Notman, 1997). A comparative analysis of 4 risk adjustment models also revealed that none of them predicted more than 10% of the variance in mental health expenditures (Ettner et al., 1998).

Among mental health outcomes, the greatest predictive power appears to be for *functioning*, where diagnosis and clinical status accounted for nearly 40% of the variance in scores on the Medical Outcomes Study Short Form-12, or SF-12 (Ware, Kosinski, & Keller, 1996). Demographic information (age, sex, gender) and clinical status also predicted over a third of the variance in quality of life scores, and more modest predictive powers were observed when measuring outcomes of *symptom severity* and *patient satisfaction*.

Phillips et al. (2000) undertook a 2-step process to determine factors that best predicted *clinical outcomes* following receipt of mental health services among adolescents. They first convened an expert panel, which identified 11 predictors of poor outcomes.¹ They then conducted a literature review, identifying 34 studies that found factors (e.g., psychotic symptoms) that predicted worsening outcomes (e.g., recurrence of depression) among

adolescents with a given disorder (e.g., depression). They described predictive factors across 4 domains of child characteristics: (1) Clinical characteristics: Worse outcomes were seen among adolescents with diagnoses of substance use disorders, conduct disorders and psychotic symptoms; psychiatric or substance use comorbidities; and severe symptoms at baseline. (2) Clinical history: Age at onset was variably related to outcomes, though prior treatment/hospitalization was a strong predictor of worsening outcomes. (3) Adolescent characteristics: Female gender and age at intake were variably predictive of outcomes, though IQ, reading level and arithmetic scores on standardized tests predicted poor outcomes. Finally, (4) Family characteristics: Living with parents, high paternal occupational status were associated with good outcomes, while lowered family cohesion or conflict, maternal distress, paternal disengagement from the family, and high SES were associated with worse outcomes.

Because of their underperformance, few Medicaid agencies use formal risk adjustment approaches to better understand the mental health service utilization of their beneficiaries, and those that do, use them primarily to estimate costs for disabled beneficiaries. For example, the state of Washington's Medicaid agency uses the Chronic Illness and Disability Payment System (CDPS) (Kronick et al., 2000) to predict costs of their disabled beneficiaries; Colorado's Medicaid agency uses the Disability Payment System (Kronick, Dreyfus, Lee, & Zhou, 1996); Maryland's Medicaid uses the Adjusted Clinical Groups (Weiner, Starfield, Steinwachs, & Mumford, 1991) approach (Weissman et al., 2005); and Rhode Island relies on age, sex, and eligibility status. Unevaluated approaches also include Indiana's attempt to develop risk adjusted case rate models for children and adults using mental health services using a combination of diagnostic clusters, and functional levels.

2.3. Preliminary efforts to develop risk adjustment methodologies for child welfare

In the child welfare literature, there are several studies that attempt to identify associations between various predictors (at the child, family, neighborhood, systems, and social level) and outcomes (including services, and health and mental health outcomes). These studies do not, however, describe themselves as undertaking risk adjustment; these are reviewed in the next section. In this section, a much smaller set of studies that explicitly undertake risk adjustment are reviewed.

In a study using child welfare administrative data, McMillen et al. (2008) used risk adjusted spells in residential treatment of children over age 10 years to study 4 outcomes. Strong predictors (odds ratios > 2) of the 'discharge to live with a parent and stayed there' outcome included system goal of return to family, voluntary placement, prior runaway, and female gender. Predictors of the 'return to parents within 1 year of discharge' outcome included system goal of return to parents and voluntary placement. System goals were also predictive of 'discharge to less restrictive environment' outcome, and the strongest predictors of 'having a detention or runaway episode in the year post discharge' outcome were the region within which the facility was located.

¹These included comorbid mental illness and substance abuse, increased severity of behavioral problems, younger age, prior psychiatric hospitalization, (presumably) male gender, parental history of behavioral health problems, residential or school instability, "family dysfunction," extreme poverty, history of physical or sexual abuse, and exposure to violence in the home or in the community.

Wulczyn, Chen and Orlebeke (2009) have conducted case-mix studies to profile child-serving agencies on the total number of days a child spends with the care of a single human service agency (an “agency spell”). This is a procedural measure, measuring the relative efficiency of contract agencies in achieving a reunification outcome as evaluated by total lengths of stay. Child-level characteristics associated with increased rates of reunification included female gender, older age, Black or Hispanic race/ethnicity, being in a kin home prior to coming to the agency, having a change in level of care within the agency, and having family incomes above the threshold of eligibility for Title IV-E benefits. These risk ratios were used in models with an agency-level fixed effect to identify agency-level variations in performance (agencies with highest reunification rates being the “better” agencies, controlling for child and family characteristics). Longer-term outcomes (2 year reunification rates) show similar effects of child and family characteristics, with the added finding that having a sibling in care at the same time modestly reduced the rate of family reunification (Wulczyn, Orlebeke, & Melamid, 2000).

3. Applying risk adjustment approaches for child welfare performance measurement

Although associations between the conditions that lead children to come into contact with child welfare agencies are known to be associated with several adverse health and mental health outcomes in later life, relatively few studies have described themselves as using risk adjustment approaches to study child welfare outcomes. Despite this, the field offers great potential for the development and successful application of risk adjustment approaches. This is because a wealth of literature has documented the consequences of maltreatment on various poor outcomes, and has developed information on both the predictors of maltreatment and its sequelae. Both of these sets of studies can be used to identify variables, and inform methodologies, for risk adjustment.

Perhaps the best known of “associative” studies showing the consequences of maltreatment on outcomes is the Adverse Childhood Experiences study, which is a combined retrospective-prospective study of nearly 19,000 enrollees of Kaiser Permanente, a Southern California health maintenance organization (Felitti et al., 1998). The study classified adverse childhood experiences (ACE) into 3 categories of child abuse (physical, sexual, and psychological), and 4 categories of exposure to household dysfunction (substance abuse, mental illness, domestic violence toward mother, and criminal behavior occurring within the home). Findings from the ACE study are reported in over 70 publications and a for-public website (Adverse Childhood Experiences Study, 2007) and suggest that: (a) adverse childhood experiences are very common, with 64% of respondents reporting at least one ACE (Felitti et al., 1998); (b) experiencing adversity during childhood is associated in adulthood with several of the nation’s leading causes of mortality and morbidity, including risky sexual behaviors (Hillis, Anda, Felitti, & Marchbanks, 2001) and pregnancy among adolescent women (Dietz et al., 1999); sexually transmitted diseases (Hillis, Anda, Felitti, Nordenberg, & Marchbanks, 2000); attempted suicide (Dube et al., 2001); alcohol dependence (Anda et al., 2002; Dube, Anda, Felitti, Edwards, & Croft, 2002); drug dependence (Dube et al., 2003); liver disease (Dong, Dube, Felitti, Giles, & Anda, 2003);

depressive disorders (Chapman et al., 2004); obesity (Williamson, Thompson, Anda, Dietz, & Felitti, 2002); and heart disease (Dong et al., 2004) among others; and (c) childhood adversity exhibits a dosing effect, such that individuals with greater numbers of ACEs are at progressively higher risk of these conditions.

With respect to predictors, identifying particular *characteristics of children* that place them at higher risk for these conditions has proven far more difficult because of the multiplicity of variables that may mediate the relationships between child maltreatment and a particular outcome, and the shared contribution of several variables (such as genetic vulnerabilities and environmental stressors) to both child maltreatment as well as a particular health or mental health outcome. Many of these approaches (conducting cluster analysis of physically abused children in order to identify clusters associated with worse outcomes (Sabourin Ward & Haskett, 2008) for example) are probably useful in research but are impractical for large scale risk adjustment studies designed to be conducted by child welfare agencies themselves. (This is also the reason why qualitative approaches cannot be widely used in risk adjustment. Unless qualitative data are readily available in administrative data sets, it is too expensive and cumbersome to use individually-collected qualitative data in conducting risk adjustment studies on large samples.)

National origin of adopted children has been reported as a predictor of poor outcomes. A study compared 16,222 children born outside Europe who were adopted in Sweden between 1973 and 1984, with an age-adjusted comparison group of 1,026,523 children from the domestic population (Elmund, Lindblad, Vinnerljung, & Hjern, 2007). Intercounty adoptees had 5 times the odds of residential care placement and over 3 times the odds of foster care placement after the age of 10 years, when compared to domestic children. (In Sweden, residential care placement is apparently largely due to behavioral problems.) Increased child age at adoption, adoption from Latin America, a single parent adoption, and older (>35) age of the mother at birth of the child were predictors of behavioral problems necessitating entry into residential and foster care.

Because children coming into contact with the child welfare system and those in foster care are such high utilizers of psychotropic medications (McMillen et al., 2004; Raghavan et al., 2005), risk adjustment technologies developed to predict pharmacy utilization are particularly suited for applications to this population. Pharmacy based risk adjustment methodologies such as Medicaid Rx (Gilmer, Kronick, Fishman, & Ganiats, 2001) can easily be used in child welfare populations. However, in so far psychopharmacological services are under the purview of state mental health and Medicaid departments rather than child welfare agencies, its use may be inappropriate in child welfare performance measurement.

Biomarkers have assumed increasing prominence in health and mental health research. With respect to child welfare, lowered morning serum cortisol levels have been documented among children exposed to sexual abuse, physical abuse, neglect, and emotional maltreatment (Cicchetti & Rogosch, 2001a,b), although the exact profile of cortisol abnormalities remain to be determined. The majority of maltreated children in one pilot seemed to exhibit a normal cortisol pattern when followed over time (Linares et al., 2008),

which diminishes the utility of using cortisol as a biomarker for maltreatment and, possibly, in risk adjustment studies. Preliminary evidence also suggests that there may be a genetic protective effect upon anxiety disorders among adolescents exhibiting a particular allele at both an enzyme locus as well as a serotonin transporter locus; such protection may be enhanced under conditions of stressors such as sexual abuse (Olsson et al., 2007).

Characteristics of the maltreatment experience are perhaps to child welfare outcomes what diagnoses are to health and mental health outcomes. The traditional approach to examining relationships between types of maltreatment and its behavioral sequelae is to adopt a type of hierarchical classification. This contrasts with the ACE study reported above, where number of types of maltreatment are simply summed, in a cumulative approach. This cumulative approach is also referred to as a “multiple maltreatment” approach (Trickett & Trickett, 1998). In a hierarchical approach, in contrast, abuse is predicted to produce worse outcomes than neglect, sexual abuse is predicted to produce worse outcomes than physical abuse, and so on. There is some evidence for a ‘ranking’ of different types of maltreatment, with child sexual abuse producing the most emotional sequelae among a sample of youth admitted to a psychiatric facility (Boxer & Terranova, 2008).

Indeed, the phenomenology of sexual abuse may itself be a risk factor – Fergusson, Horwood, and Lynskey (1996) found that, controlling for histories of parental psychiatric morbidity, family conflict, and parental substance abuse, child sexual abuse predicted occurrence of alcohol use disorders with odds ratios that varied between 3.2 (for children experiencing contact abuse but no intercourse), and 2.7 (sexual abuse involving intercourse). Also, individuals who experience sexual abuse at the hands of an intrafamilial perpetrator, and those who experience maltreatment under the threat of force, often experience worse outcomes (Trickett et al., 1997). This suggests that perpetrator identity and the context of the maltreatment experience are important predictors to consider for risk adjustment studies.

Alternative conceptualizations of the effects of maltreatment upon outcomes rest on a multiple hazards approach. Because few children are victims of only one type of abuse, co-occurrence of maltreatment may be the norm in child welfare. This co-occurrence means that each additional type of abuse may have synergistic effects on outcomes, and may produce outcomes that are far worse than a simple count of maltreatment might suggest. Research on young adolescents, for example, suggests that multiple forms of maltreatment are only predictive of trauma symptoms when they are of a certain severity (Clemmons, Walsh, DiLillo, & Messman-Moore, 2007), suggesting a dose–response relationship predicated upon more than mere number of types of maltreatment. This conceptualization suggests that risk adjustment approaches should consider the degree of maltreatment (frequency and severity) (Higgins, 2004), rather than solely count the number of types of maltreatment.

Factors within the child’s familial and social ecology could also be used in risk adjustment studies. Much of the literature, however, focuses on factors such as the relationship between placement stability (Rubin et al., 2004; Rubin, O’Reilly, Luan, & Localio, 2007) and outcomes such as behavioral health symptoms and costs of care. It is unclear if placement changes should be considered while constructing risk adjustment approaches. First, because

placement changes can, to a degree, be controlled by a child welfare agency, they are not exogenous markers of outcomes. Second, interventions deployed by child welfare agencies, such as parent training programs, are explicitly designed to minimize the effects of unstable placements. To use a medical analogy, placement changes are not akin to diagnosis but are essentially indicators of response to treatment, the “treatment” being parent training or out-of-home placement, and similar actions taken by child welfare agencies in response to the child and his/her ecology. This is why this review does not consider strategies such as enhanced foster care (Kessler et al., 2008), Early Intervention Foster Care (treatment foster care) (Fisher, Gunnar, Chamberlain, & Reid, 2000), identifying and intervening with sibling relationships in foster care (Linares, Li, Shrout, Brody, & Pettit, 2007), or other prevention and treatment strategies as appropriate predictors for risk adjustment studies in child welfare despite their proven efficacy. Thirdly, in many cases “treatment” is either controversial or impractical. For example, children in low-income single parent families – as opposed to two-parent families – have been reported to have higher risks of being exposed to certain social and contextual factors, which in turn are associated with a variety of behavioral, welfare, and criminal justice outcomes (Fergusson, Boden, & Horwood, 2007). Attempting to “treat” such families by promoting family formation is perhaps a policy that should be approached with a great deal of caution. Impracticality is also the reason why the intergenerational transmission of psychosocial risk (abuse of the child’s birth mother when she herself was a child, for example) (Collishaw, Dunn, O’Connor, & Golding, 2007) may be a poor risk adjuster for child outcomes.

One example of a familial characteristic that can be used in risk adjustment studies is maternal substance use. Approximately 41% of birth mothers of children adopted into the United States from countries comprising the former Soviet Union (these countries were the largest donors of adoptee children to the United States for much of the 1990s) had histories of alcohol abuse. Perhaps not surprisingly, several of these adopted children had neurodevelopmental delays and other symptoms suggestive of fetal alcohol syndrome (McGuinness, McGuinness, & Dyer, 2000).

System-level factors also seem important with respect to outcomes, with these outcomes varying by region (McMillen et al., 2008). However, it remains unclear exactly what those regional differences represent, and to what extent they influence child welfare outcomes.

The bulk of this literature, focused as it is on identifying predictors of illness, is less focused on identifying predictors of wellness. To take one protective factor as an example, a study that interviewed a cohort of 188 maltreated children aged 6–12 years attending a summer camp revealed that girls who placed a high importance on their religious faith had lower internalizing symptoms compared to a comparison sample of non-maltreated girls. Also, boys who reported a greater level of attendance at religious services reported lower externalizing symptoms (Kim, 2008). It is unclear if administrative data contain information on such protective factors that could be used in risk adjustment studies.

In summary, existing literature in child welfare offers considerable information on variables and approaches that can be used to better predict, and thereby enhance, the performance of child welfare systems. This next section addresses applications of these methodologies for

child welfare performance measurement under four broad topics – what is it that we should be risk-adjusting, how should we risk-adjust these outcomes, what are some challenges in the widespread use of risk adjustment approaches, and are there any low-hanging fruit that can serve as first steps toward increasing experience with risk adjustment in child welfare settings?

4. Discussion

A comprehensive overview of the conceptualization and operationalization of performance is beyond the scope of this report. Several reviews address issues related to performance measurement in organizations (Harvard Business Review on Measuring Corporate Performance, 1998; Neely, 2007) and in health systems (Joint Commission on Accreditation of Healthcare Organizations., 2000). Government entities have released national measures for quality and accountability in behavioral health (Substance Abuse and Mental Health Services Administration, 2007). The Institute of Medicine through its *Pathways to Quality Health Care* series of books has proposed several design principles for the development of performance measures, including expanding the scope of measurement, incorporating a longer time window of measurement, assessing the entire ecology of clinical care, and emphasizing shared accountability, among others (Institute of Medicine (U.S.). Committee on Redesigning Health Insurance Performance Measures Payment and Performance Improvement Programs., 2006) It has also suggested the operationalization of performance measures to the 6 aims of quality improvement – safety, effectiveness, patient-centeredness, timeliness, efficiency, and equity (known collectively as STEEP) – as previously proposed (Institute of Medicine (U.S.). Committee on Quality of Health Care in America., 2001). The Institute has also examined fiscal approaches to quality improvement, such as pay-for-performance strategies directed toward specified areas of quality improvement, which have yielded mixed results thus far. (Institute of Medicine (U.S.). Committee on Redesigning Health Insurance Performance Measures Payment and Performance Improvement Programs., 2007).

Similar concern with quality improvement within child welfare settings have resulted in the development of performance measurement directed toward child welfare agencies. Passage of the 1997 Adoption and Safe Families Act (PL 105-89) introduced three significant changes to existing child welfare regulations (Baker & Rauber, 2001; Wulczyn, 2005). First, it directed states to develop and implement standards ensuring that children in foster care receive high-quality services to ensure their safety and health. Second, it directed the Department of Health and Human Services (HHS) to recommend to Congress a performance-based incentive funding system for federal child welfare payments; and third, (and perhaps most significantly), it mandated achievement of ‘child well-being’ in addition to prior goals of safety and permanency. In 2000, the HHS implemented Child and Family Services Reviews (CFSRs), which track states’ compliance with federal regulations on 45 outcomes distributed across goals of safety, permanency, and well-being; and carry fiscal punishment for underperforming states (United States. General Accounting Office., 2004). These outcomes are listed in Table 1. For the rest of this section, this paper focuses on achievement of the CFSR outcomes, as the most relevant metrics of performance measurement for state agencies.

4.1. Identifying appropriate outcomes for child welfare risk adjustment studies

If the CFSR outcomes are considered to be the appropriate metrics of performance assessment, then the question becomes: To which of these outcome metrics should risk adjustment methodologies be applied? In other words, what is the appropriate *dependent variable* on which risk adjustment strategies can be used to assess performance of child welfare systems?

With respect to the *safety* outcome, several studies suggest a potential role for risk adjustment techniques in predicting *recurrence of maltreatment* (Cash, 2001). (This is different from using recurrence as a criterion to develop risk assessment instruments.) Child characteristics such as younger age are known predictors of recurrence (Fluke, Yuan, & Edwards, 1999; Fryer, Fryer, & Miyoshi, 1994). Maltreatment type has been shown to display gradations in risk of recurrence (neglect, then physical abuse, then sexual abuse in order of decreasing risks of recurrence), and provision of services seems to increase risks of maltreatment recurrence (Fluke et al., 1999). Prior maltreatment also seemingly exerts a multiplicative rather than simply additive effect upon future maltreatment. Connell, Bergeron, Katz, Saunders and Tebes (2007) examined data on 22,584 children referred to child protective services in Rhode Island over a 4-year period, and found that younger children, those who were physically disabled, and whose caregivers abused substances or reported financial difficulties had greater risk of re-investigation. Children of African American and Hispanic race/ethnicity, those with history of sexual abuse, and those with past substantiated maltreatment were at lower relative risk. Clusters of risk factors have also been identified, including child vulnerabilities and exposure to the perpetrator; familial competencies such as the mother's parenting skills, and the mother's expectations of the child; and the availability and accessibility of social and agency resources (DePanfilis & Zuravin, 1999; Fuller, Wells, & Cotton, 2001; Johnson & L'Esperance, 1984). These studies suggest, then, that risk adjustment methodologies can be developed for at least two of the measures under the safety outcome (repeat maltreatment and absence of recurrence of maltreatment).

With respect to adjusting for child-level differences in measures of the *permanency* outcome, there is evidence from studies cited earlier that risk adjustment techniques can be used to compare agencies on a process measure of time (agency spells) to a particular desirable outcome (here, reunification) (Wulczyn et al., 2009, 2000). This is probably the most developed area in terms of risk adjusting child welfare outcomes. These studies focus on reunification as an ultimate outcome, and treat agency spells as a proxy for efficiencies in securing that outcome. Such use is consistent with the bulk of the literature supporting the use of risk adjustment techniques to achieve outcomes.

Risk adjustment techniques can be appropriately used to measure processes of care (such as lengths of stay), only when these processes of care are linked to an appropriate outcome (such as reunification), as in the set of studies by Wulczyn and colleagues (above). This for two reasons. First, the number of days spent within a facility is a procedural measure, not an outcome measure; applying risk adjustment methodologies to such applications makes the assumption that lower lengths of stay within an agency are *always* better. If a possibility

exists that some children may benefit from longer stays within residential care facilities (perhaps to allow time for treatments to work), then widespread use of risk adjustment may contribute to inappropriate *underutilization* of services. Under such circumstances, risk adjustment methodologies could be used to model time in remission, clinical outcomes, time to readmission, or some other measure suggestive of outcomes of care. Second, risk adjustment methodologies can be gamed when the outcome is under the control of the residential facilities themselves. Because lengths of stay within facilities are to some extent functions of the facilities themselves, facilities can artificially reduce lengths of stay (for example, by discharging and readmitting residents so as to break a single duration of stay into smaller segments). Facilities can also potentially game the system by discharging hard-to-treat clients in an attempt to maintain smaller lengths of stay, and perhaps the ultimate measure of the appropriateness of a risk adjustment method is its resistance to gaming by provider groups or facilities.

Because the costs of residential care are so high, several child welfare agencies have engaged in performance-based contracting with residential treatment providers. The state of Illinois, for example, has a project called Striving for Excellence (“Striving for Excellence: Project Overview,”) which developed an in-house risk adjustment system to develop benchmarks for two performance measures. The difference between actual performance and this risk-adjusted benchmarked performance is penalized fiscally. For example, the “Treatment Opportunity Days Rate” performance measure uses the difference between the agency’s benchmark number of days in a fiscal year that residents are “not on the run, psychiatrically hospitalized or incarcerated,” and the agency’s actual number of such days. The shortfall between actual (observed) days and ideal (benchmark) days is compensated at 75% of the per-diem rate for the bed. This type of endeavor assumes a great deal of confidence in the variables and in the approach used to construct the benchmark. If states quasi-randomly assign children to residential facilities (as in no-decline contracting, or rotating assignment of incoming children to facilities without any attempt to match children to facilities), then child-level characteristics may not greatly bias facility performance. Nevertheless, given the weak predictive power of most psychosocial variables, further study across other states may be required before statewide benchmarks can be reliably established.

Examining the extent to which children receive care consistent with one or more measures within the *child well-being outcome* is the risk adjustment application for which the greatest empirical evidence and experience exists. As described in the previous section, the risk adjustment literature has captured a variety of physical and mental health outcomes, many of which can be operationalized for use with child welfare populations using existing administrative data.

4.2. Identifying appropriate predictors for child welfare risk adjustment studies

What are the appropriate *independent variables* that could be used to predict specific outcomes of relevance to child welfare?

Risk adjustment within public settings has a major advantage that is lacking in private applications. As described in the Conceptual Overview section, race/ethnicity is not used as a predictor in risk adjustment studies even though it is an established predictor of service use

(Harris et al., 2006) because of equity issues. Such constraints may not apply to its use as a predictor in public settings, so far as risk adjustment studies are not used to set reimbursements to child-serving agencies. So long as the goal is to better understand factors that affect well-being (or some other outcome) the child's race/ethnicity may offer additional explanatory power.

Two great advantages in developing risk adjustment methodologies to child welfare are the wealth of data on predictors of child welfare outcomes (reviewed in the last section), and the wealth of experience that exists in the development of risk assessment instruments. These instruments, which are either based on expert opinion (*consensus-based*) or based on statistical models (*actuarial*), are used to predict risk of recurrence of maltreatment. A comparison of the relative performance of two consensus-based models versus one actuarial model suggested greater predictive power for the actuarial model (Baird & Wagner, 2000), and newer actuarial approaches have recently been developed using nationally representative data (Shlonsky, 2007). Actuarial risk assessment models are actually a special type of risk adjustment models, using somewhat similar methodologies, but focused on a different outcome.

Perhaps the greatest barrier to the development of risk assessment approaches – which diminishes its ability to ‘cross-walk’ to risk adjustment – is the heterogeneity of the outcome variable. Maltreatment is not a unitary construct, but is a construct whose operational definition varies across state legal systems and child welfare practices (Hutchison, 1994). This means that risk assessment methodologies are actually trying to predict a variety of different events, collectively subsumed under the term of maltreatment. (To take a medical analogy, if the definition of schizophrenia varied between Alabama and California, the ability of researchers to identify predictors of response to treatment for patients with schizophrenia would be severely compromised.) Repurposing existing risk assessment methodologies from maltreatment to outcomes that are more stable across legal and child welfare systems is a necessary first step in the building risk adjustment methodologies upon the scaffolding of risk assessment approaches.

4.3. Methodological advances to convert groups of predictors into ‘systems’

Following agreement on appropriate predictor and outcome variables, the next step toward application of risk adjustment methodologies to child welfare is specifying an appropriate model specification. Studies cited in the previous section largely tend to use individual predictors introduced into the model. However, experience from the development of risk adjustment instruments in health suggests that these individual variables do not have great predictive power because individuals with identical values on predictor variables exhibit variations in clinical outcomes and use of services. Risk adjustment approaches work best when predictors can be converted into predictive clusters, which have greater explanatory power because they better account for individual-level heterogeneity.

What these appropriate predictive clusters are is an empirical question. Current approaches in risk assessment (Shlonsky, 2007) use scaling techniques, which may be the first element of a research agenda on developing models with greater predictive power. Ultimately, it is likely that the development of risk adjustment approaches in child welfare will require

availability of pooled administrative data across several child welfare systems, establishment of collaborative relationships between agencies and researchers (or vendors), the external development of methodologies, and their subsequent incorporation into child welfare systems – in other words, the model followed in the health care field.

4.4. Stratification problems in child welfare risk adjustment

In contrast to health care settings, child welfare settings are characterized by greater levels of heterogeneity. For instance, the bulk of risk adjustment studies in health have been conducted upon largely hospital-based populations. In child welfare, however, children with high protective and health needs are found in a variety of settings – within kin families, in non-relative foster care, and in residential care, among others. This heterogeneity in service settings suggests a need for stratification when it comes to risk adjustment applications. In other words, separate risk adjustment approaches will likely need to be developed for defined service venues.

Intuitively realizing this issue, risk-adjustment researchers cited above seem to have adopted a tacit gradation of service venue. They have conducted risk adjustment studies on clients of residential treatment facilities (institutionally, perhaps akin to hospitals, to take a health analogy). A research agenda that focuses on children in foster care is perhaps the next step, to be followed by children receiving services within their own homes, whose needs are comparable to those of out-of-home children (Burns et al., 2004; Farmer et al., 2001; Kolko, Selleyo, & Brown, 1999). (Clearly, then, models in such stratifications cannot rely on placement as a proxy for unobserved effects at the level of a service venue.)

4.5. Solving the pricing problem: what are the costs of child welfare services?

The above discussion assumes that the goal of risk adjustment is to gain further explanatory power on service use. However, if a goal is to also use these approaches to implement fiscal policy (such as pay-for-performance contracting), then this will require development of a research agenda on the costs of child welfare services.

Conceptually, expenditures are the product of price and quantity ($E = P \times Q$). In health care, both the costs of a service as well as the quantity of the service are reasonably well known in most instances. For example, in most individuals, an infection caused by a bacterium can be eradicated using an antibiotic to which it is sensitive after a predetermined course of treatment. Hence, both the cost of care (the antibiotic) as well as the course of its treatment (number of doses) are known. Hence, an outcomes-based pay-for-performance model can be constructed.

In contrast, in mental health, while the cost of care may be known, the quantity of care is frequently unknown. For example, an index episode of depression among adolescents ranges between 4 and 9 months in clinical samples, and between 3 and 6.5 months in community samples (Birmaher, Arbelaez, & Brent, 2002). When exactly an adolescent's symptoms will remit is an open question. Indeed, in some cases recovery to baseline might never occur, and the disease may exhibit a heterogeneous course and outcome. This complicates rate setting because it is unclear how much this index episode of depression is going to cost a service

provider. This uncertainty then affects the payor, who cannot arrive at a reasonable mean estimate of cost around which to construct confidence intervals for a pay-for-performance system.

In child welfare today, neither is the price of service known, nor is its appropriate quantity. It is unclear what the true costs of delivering child welfare services are. Furthermore, these services are delivered by a variety of providers, and it is likely that the costs of these services are cost-shifted i.e., borne by a variety of budgets (e.g., deficiencies in Medicaid funding may be made up using child welfare dollars for children with health needs). Absent greater clarity on the costs and course of child welfare services, fiscal policy to support quality should be approached with caution.

5. Conclusion

This brief survey of applications of risk adjustment methodologies to child welfare performance assessment suggests both strengths as well as challenges. On the strengths side, there is considerable knowledge on what factors are associated with various outcomes of importance to child welfare agencies. Importantly, child welfare researchers have identified not only child-level, but also ecological predictors of outcomes. They have also begun to build methodological systems that approximate risk adjustment.

On the challenge side, much of this knowledge has not been translated into formal risk adjustment approaches that can inform policy. Given the general lack of this translation from research to policy, child welfare policies designed to punish or reward states or agencies on the basis of their performance on defined outcomes should be approached with a great deal of caution. It is unclear if these entities are solely responsible for the outcomes they manifest, and enacting regulations that treat them as if they are, may be inappropriate. Child welfare agencies that engage in performance-based contracting should also carefully consider the quality of the scientific data on which their policymaking is based.

However, there are several areas where risk adjustment approaches might conceivably be used immediately, particularly in the child well-being domain. Performance based contracting for residential care is perhaps a logical first area for the further development and introduction of high-quality risk adjustment techniques using several of the variables and approaches detailed above. Because mental health services are frequently purchased using child welfare dollars, adapting mental health risk adjustment approaches to such contracts can also be effected. Applications to the safety and permanency outcomes will require further research before widespread adoption.

As federal law and regulation moves toward increasing oversight and monitoring of state child welfare agencies, the need for appropriate risk adjustment assumes greater criticality. Given the multifactorial causes of variations in the performance of child welfare systems, only part of which may be under their control, the appropriate and careful use of risk adjustment techniques may be the only way to assure a truly high-performance child welfare system.

Acknowledgments

Support for this paper came from the National Academies, and the National Institute of Mental Health (R03 MH082117). All views expressed are those of the author, and do not necessarily reflect the views of the National Academy of Sciences, its Division of Behavioral and Social Sciences and Education, or its Board on Children Youth and Families.

References

- Adams EK, Bronstein JM, Raskind-Hood C. Adjusted clinical groups: Predictive accuracy for Medicaid enrollees in three states. *Health Care Financing Review*. 2002; 24(1):43–61. [PubMed: 12545598]
- Adverse Childhood Experiences Study. 2007. Retrieved 3/29/07, from www.acestudy.org
- Anda RF, Whitfield CL, Felitti VJ, Chapman D, Edwards VJ, Dube SR, et al. Adverse childhood experiences, alcoholic parents, and later risk of alcoholism and depression. *Psychiatric Services* (Washington, D C). 2002; 53(8):1001–1009.
- Baird C, Wagner D. The relative validity of actuarial- and consensus-based risk assessment systems. *Children and Youth Services Review*. 2000; 22(11-12):839–871. [Print Electronic; Print].
- Baker, DR.; Rauber, DB. Making sense of the ASFA regulations: A roadmap for effective implementation. Washington, D.C.: American Bar Association, National Child Welfare Resource Center on Legal and Judicial Issues, Center on Children and the Law; 2001.
- Birmaher B, Arbelaez C, Brent D. Course and outcome of child and adolescent major depressive disorder. *Child and Adolescent Psychiatric Clinics of North America*. 2002; 11(3):619–637 x. [PubMed: 12222086]
- Blumenthal D, Weissman JS, Wachterman M, Weil E, Stafford RS, Perrin JM, et al. The who, what, and why of risk adjustment: A technology on the cusp of adoption. *Journal of Health Politics, Policy and Law*. 2005; 30(3):453–473.
- Boxer P, Terranova AM. Effects of multiple maltreatment experiences among psychiatrically hospitalized youth. *Child Abuse & Neglect*. 2008; 32(6):637–647. [PubMed: 18582936]
- Burns BJ, Phillips SD, Wagner HR, Barth RP, Kolko DJ, Campbell Y, et al. Mental health need and access to mental health services by youths involved with child welfare: a national survey. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2004; 43(8):960–970. [PubMed: 15266190]
- Cash SJ. Risk assessment in child welfare: The art and science. *Children and Youth Services Review*. 2001; 23(11):811–830. [Print Electronic; Print].
- Centers for Medicare and Medicaid Services. CMS DRG Definitions Manual version 24.0. Washington, DC: 3m Health Information; 2006.
- Chapman DP, Whitfield CL, Felitti VJ, Dube SR, Edwards VJ, Anda RF. Adverse childhood experiences and the risk of depressive disorders in adulthood. *Journal of Affective Disorders*. 2004; 82(2):217–225. [PubMed: 15488250]
- Cicchetti D, Rogosch FA. Diverse patterns of neuroendocrine activity in maltreated children. *Development and Psychopathology*. 2001; 13(3):677–693. [PubMed: 11523854]
- Cicchetti D, Rogosch FA. The impact of child maltreatment and psychopathology on neuroendocrine functioning. *Development and Psychopathology*. 2001; 13(4):783–804. [PubMed: 11771908]
- Clemmons JC, Walsh K, DiLillo D, Messman-Moore TL. Unique and combined contributions of multiple child abuse types and abuse severity to adult trauma symptomatology. *Child Maltreatment*. 2007; 12(2):172–181. [PubMed: 17446570]
- Collishaw S, Dunn J, O'Connor TG, Golding J. Maternal childhood abuse and offspring adjustment over time. *Development and Psychopathology*. 2007; 19(2):367–383. [PubMed: 17459175]
- Connell CM, Bergeron N, Katz KH, Saunders L, Tebes JK. Re-referral to child protective services: The influence of child, family, and case characteristics on risk status. *Child Abuse & Neglect*. 2007; 31(5):573–588. [PubMed: 17537504]

- Courtney ME, Needell B, Wulczyn F. Unintended consequences of the push for accountability: The case of national child welfare performance standards. *Children and Youth Services Review*. 2004; 26(12):1141–1154.
- Deas TM Jr. Health care value-based purchasing. *Gastrointestinal Endoscopy Clinics of North America*. 2006; 16(4):643–656. [PubMed: 17098612]
- DePanfilis D, Zuravin SJ. Predicting child maltreatment recurrences during treatment. *Child Abuse & Neglect*. 1999; 23(8):729–743. [PubMed: 10477234]
- Dietz PM, Spitz AM, Anda RF, Williamson DF, McMahon PM, Santelli JS, et al. Unintended pregnancy among adult women exposed to abuse or household dysfunction during their childhood. *Jama*. 1999; 282(14):1359–1364. [PubMed: 10527183]
- Dong M, Dube SR, Felitti VJ, Giles WH, Anda RF. Adverse childhood experiences and self-reported liver disease: new insights into the causal pathway. *Archives of Internal Medicine*. 2003; 163(16): 1949–1956. [PubMed: 12963569]
- Dong M, Giles WH, Felitti VJ, Dube SR, Williams JE, Chapman DP, et al. Insights into causal pathways for ischemic heart disease: Adverse childhood experiences study. *Circulation*. 2004; 110(13):1761–1766. [PubMed: 15381652]
- Dube SR, Anda RF, Felitti VJ, Chapman DP, Williamson DF, Giles WH. Childhood abuse, household dysfunction, and the risk of attempted suicide throughout the life span: findings from the Adverse Childhood Experiences Study. *Jama*. 2001; 286(24):3089–3096. [PubMed: 11754674]
- Dube SR, Anda RF, Felitti VJ, Edwards VJ, Croft JB. Adverse childhood experiences and personal alcohol abuse as an adult. *Addictive Behaviors*. 2002; 27(5):713–725. [PubMed: 12201379]
- Dube SR, Felitti VJ, Dong M, Chapman DP, Giles WH, Anda RF. Childhood abuse, neglect, and household dysfunction and the risk of illicit drug use: the adverse childhood experiences study. *Pediatrics*. 2003; 111(3):564–572. [PubMed: 12612237]
- Elmund A, Lindblad F, Vinnerljung B, Hjern A. Intercountry adoptees in out-of-home care: A national cohort study. *Acta Pædiatrica*. 2007; 96(3):437–442.
- Ettner SL, Frank RG, Mark T, Smith MW. Risk adjustment of capitation payments to behavioral health care carve-outs: How well do existing methodologies account for psychiatric disability? *Health Care Management Science*. 2000; 3(2):159–169. [PubMed: 10780284]
- Ettner SL, Frank RG, McGuire TG, Hermann RC. Risk adjustment alternatives in paying for behavioral health care under Medicaid. *Health Services Research*. 2001; 36(4):793–811. [PubMed: 11508640]
- Ettner SL, Frank RG, McGuire TG, Newhouse JP, Notman EH. Risk adjustment of mental health and substance abuse payments. *Inquiry*. 1998; 35(2):223–239. [PubMed: 9719789]
- Ettner SL, Notman EH. How well do ambulatory care groups predict expenditures on mental health and substance abuse patients? *Administration and Policy in Mental Health*. 1997; 24(4):339–357. [PubMed: 9217332]
- Farmer EM, Burns BJ, Chapman MV, Phillips SD, Angold A, Costello EJ. Use of mental health services by youth in contact with social services. *Social Service Review*. 2001; 75:605–624.
- Felitti VJ, Anda RF, Nordenberg D, Williamson DF, Spitz AM, Edwards V, et al. Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The Adverse Childhood Experiences (ACE) Study. *American Journal of Preventive Medicine*. 1998; 14(4):245–258. [PubMed: 9635069]
- Fergusson DM, Boden JM, Horwood LJ. Exposure to single parenthood in childhood and later mental health, educational, economic, and criminal behavior outcomes. *Archives of General Psychiatry*. 2007; 64(9):1089–1095. [PubMed: 17768274]
- Fergusson DM, Horwood LJ, Lynskey MT. Childhood sexual abuse and psychiatric disorder in young adulthood: II. Psychiatric outcomes of childhood sexual abuse. *Journal of the American Academy of Child and Adolescent Psychiatry*. 1996; 35(10):1365–1374. [PubMed: 8885591]
- Fisher PA, Gunnar MR, Chamberlain P, Reid JB. Preventive intervention for maltreated preschool children: impact on children's behavior, neuroendocrine activity, and foster parent functioning. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2000; 39(11):1356–1364. [PubMed: 11068890]

- Fluke JD, Yuan Y-YT, Edwards M. Recurrence of maltreatment: An application of the National Child Abuse and Neglect Data System (NCANDS). *Child Abuse & Neglect*. 1999; 23(7):633–650. [PubMed: 10442829]
- Fryer GE, Fryer GE, Miyoshi TJ. A survival analysis of the revictimization of children: The case of Colorado. *Child Abuse & Neglect*. 1994; 18(12):1063–1071. [Print]. [PubMed: 7850614]
- Fuller TL, Wells SJ, Cotton EE. Predictors of maltreatment recurrence at two milestones in the life of a case. *Children and Youth Services Review*. 2001; 23(1):49–78. [Print Electronic; Print].
- Gilmer T, Kronick R, Fishman P, Ganiats TG. The Medicaid Rx model: Pharmacy-based risk adjustment for public programs. *Medical Care*. 2001; 39(11):1188–1202. [PubMed: 11606873]
- Harris KM, Gordon-Larsen P, Chantala K, Udry JR. Longitudinal trends in race/ethnic disparities in leading health indicators from adolescence to young adulthood. *Archives of Pediatrics & Adolescent Medicine*. 2006; 160(1):74–81. [PubMed: 16389215]
- Harvard Business Review on Measuring Corporate Performance. Boston, MA: Harvard Business School Press; 1998.
- Hendryx MS, Beigel A, Doucette A. Introduction: Risk-adjustment issues in mental health services. *The Journal of Behavioral Health Services & Research*. 2001; 28(3):225–234. [PubMed: 11497019]
- Hermann RC, Rollins CK, Chan JA. Risk-adjusting outcomes of mental health and substance-related care: A review of the literature. *Harvard Review of Psychiatry*. 2007; 15(2):52–69. [PubMed: 17454175]
- Higgins DJ. The importance of degree versus type of maltreatment: A cluster analysis of child abuse types. *The Journal of Psychology*. 2004; 138(4):303–324. [PubMed: 15379008]
- Hillis SD, Anda RF, Felitti VJ, Marchbanks PA. Adverse childhood experiences and sexual risk behaviors in women: A retrospective cohort study. *Family Planning Perspectives*. 2001; 33(5): 206–211. [PubMed: 11589541]
- Hillis SD, Anda RF, Felitti VJ, Nordenberg D, Marchbanks PA. Adverse childhood experiences and sexually transmitted diseases in men and women: A retrospective study. *Pediatrics*. 2000; 106(1):E11. [PubMed: 10878180]
- Hutchison, E. Child maltreatment: Can it be defined? *Child Welfare Research Review*. Barth, R.; Berrick, JD.; Gilbert, N., editors. Vol. 1. New York, NY: Columbia University Press; 1994. p. 5–27.
- Iezzoni LI. Risk adjustment for measuring health care outcomes. 2003
- Institute of Medicine (U.S.). Crossing the quality chasm: A new health system for the 21st century. Washington, D.C.: National Academy Press; 2001. Committee on Quality of Health Care in America.
- Institute of Medicine (U.S.). Performance measurement: Accelerating improvement. Washington, D.C.: National Academies Press; 2006. Committee on Redesigning Health Insurance Performance Measures Payment and Performance Improvement Programs.
- Institute of Medicine (U.S.). Rewarding provider performance: Aligning incentives in Medicare. Washington, DC: National Academies Press; 2007. Committee on Redesigning Health Insurance Performance Measures Payment and Performance Improvement Programs.
- Jencks SF, Dobson A. Refining case-mix adjustment. The research evidence. *New England Journal of Medicine*. 1987; 317(11):679–686. [PubMed: 3114635]
- Johns Hopkins University. The Johns Hopkins University ACG Case-Mix System. 2007. Retrieved 2/11/09, from <http://www.acg.jhsph.edu/index.htm>
- Johnson W, L'Esperance J. Predicting the recurrence of child abuse. *Social Work Research & Abstracts*. 1984; 20(2):21–26. [Print].
- Joint Commission on Accreditation of Healthcare Organizations. A guide to performance measurement for hospitals. Oakbrook Terrace, IL: Joint Commission on Accreditation of Healthcare Organizations; 2000.
- Kessler RC, Pecora PJ, Williams J, Hiripi E, O'Brien K, English D, et al. Effects of enhanced foster care on the long-term physical and mental health of foster care alumni. *Archives of General Psychiatry*. 2008; 65(6):625–633. [PubMed: 18519820]

- Kim J. The protective effects of religiosity on maladjustment among maltreated and nonmaltreated children. *Child Abuse & Neglect*. 2008; 32(7):711–720. [PubMed: 18617264]
- Kolko DJ, Seleyo J, Brown EJ. The treatment histories and service involvement of physically and sexually abusive families: Description, correspondence, and clinical correlates. *Child Abuse & Neglect*. 1999; 23(5):459–476. [PubMed: 10348382]
- Kronick R, Dreyfus T, Lee L, Zhou Z. Diagnostic risk adjustment for Medicaid: the disability payment system. *Health Care Financing Review*. 1996; 17(3):7–33. [PubMed: 10172665]
- Kronick R, Gilmer T, Dreyfus T, Lee L. Improving health-based payment for Medicaid beneficiaries: CDPS. *Health Care Financing Review*. 2000; 21(3):29–64. [PubMed: 11481767]
- Leslie DL, Rosenheck R, White WD. Capitated payments for mental health patients: A comparison of potential approaches in a public sector population. *The Journal of Mental Health Policy and Economics*. 2000; 3(1):35–44. [PubMed: 11967435]
- Linares LO, Li M, Shrout PE, Brody GH, Pettit GS. Placement shift, sibling relationship quality, and child outcomes in foster care: A controlled study. *Journal of Family Psychology*. 2007; 21(4):736–743. [PubMed: 18179345]
- Linares LO, Stovall-McClough KC, Li M, Morin N, Silva R, Albert A, et al. Salivary cortisol in foster children: A pilot study. *Child Abuse & Neglect*. 2008; 32(6):665–670. [PubMed: 18582935]
- McGuinness TM, McGuinness JP, Dyer JG. Risk and protective factors in children adopted from the former Soviet Union. *Journal of Pediatric Health Care*. 2000; 14(3):109–116. [PubMed: 10823969]
- McMillen JC, Lee BR, Jonson-Reid M. Outcomes for youth residential treatment programs using administrative data from the child welfare system: A risk-adjustment application. *Administration and Policy in Mental Health and Mental Health Services Research*. 2008; 35(3):189–197. [PubMed: 18176838]
- McMillen JC, Scott LD, Zima BT, Ollie MT, Munson MR, Spitznagel E. Use of mental health services among older youths in foster care. *Psychiatric Services*. 2004; 55(7):811–817. [PubMed: 15232022]
- Neely, AD. *Business Performance Measurement: Unifying Theories and Integrating Practice*. 2. New York, NY: Cambridge University Press; 2007.
- Nightingale, F.; Munding, MO. Florence Nightingale: Measuring hospital care outcomes: Excerpts from the books *Notes on matters affecting the health, efficiency, and hospital administration of the British Army founded chiefly on the experience of the late war*, and *Notes on hospitals*. Oakbrook Terrace, Ill: Joint Commission on Accreditation of Healthcare Organizations; 1999.
- Olsson CA, Byrnes GB, Anney RJ, Collins V, Hemphill SA, Williamson R, et al. COMT Val158Met and 5HTTLPR functional loci interact to predict persistence of anxiety across adolescence: Results from the Victorian Adolescent Health Cohort Study. *Genes, Brain and Behavior*. 2007; 6(7):647–652.
- Phillips SD, Hargis MB, Kramer TL, Lensing SY, Taylor JL, Burns BJ, et al. Toward a level playing field: Predictive factors for the outcomes of mental health treatment for adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2000; 39(12):1485–1495. [PubMed: 11128324]
- Raghavan R, Zima BT, Andersen RM, Leibowitz AA, Schuster MA, Landsverk J. Psychotropic medication use in a national probability sample of children in the child welfare system. *Journal of Child and Adolescent Psychopharmacology*. 2005; 15(1):97–106. [PubMed: 15741791]
- Rosen AK, Loveland SA, Anderson JJ, Hankin CS, Breckenridge JN, Berlowitz DR. Diagnostic cost groups (DCGs) and concurrent utilization among patients with substance abuse disorders. *Health Services Research*. 2002; 37(4):1079–1103. [PubMed: 12236385]
- Rosenthal MB, Dudley RA. Pay-for-performance: Will the latest payment trend improve care? *Jama*. 2007; 297(7):740–744. [PubMed: 17312294]
- Rubin DM, Alessandrini EA, Feudtner C, Mandell DS, Localio AR, Hadley T. Placement stability and mental health costs for children in foster care. *Pediatrics*. 2004; 113(5):1336–1341. [PubMed: 15121950]
- Rubin DM, O'Reilly AL, Luan X, Localio AR. The impact of placement stability on behavioral well-being for children in foster care. *Pediatrics*. 2007; 119(2):336–344. [PubMed: 17272624]

- Sabourin Ward C, Haskett ME. Exploration and validation of clusters of physically abused children. *Child Abuse & Neglect*. 2008; 32(5):577–588. [PubMed: 18511115]
- Shlonsky, A. Initial construction of an actuarial risk assessment measure using the National Survey of Child and Adolescent Well-Being. In: Haskins, R.; Wulczyn, F.; Webb, MB., editors. *Child protection: Using research to improve policy and practice*. Washington, DC: Brookings; 2007. p. 62-80.
- Starfield B, Weiner J, Mumford L, Steinwachs D. Ambulatory care groups: A categorization of diagnoses for research and management. *Health Services Research*. 1991; 26(1):53–74. [PubMed: 1901841]
- Striving for Excellence: Project Overview. Retrieved 2/26/09, from <http://www.uky.edu/SocialWork/qicpcw/documents/Illinois%20Fact%20Sheet%20Sept%202008.pdf>
- Substance Abuse and Mental Health Services Administration. National Outcome Measures. 2007. Retrieved 7/9/07, from http://nationaloutcomemeasures.samhsa.gov/outcome/index_2007.asp
- Trickett, PK.; Reiffman, A.; Horowitz, LA.; Putnam, FW.; Cicchetti, D.; Toth, SL. Developmental perspectives on trauma: Theory, research, and intervention. Rochester, NY US: University of Rochester Press; 1997. Characteristics of sexual abuse trauma and the prediction of developmental outcomes; p. 289-314.
- Trickett PK, Trickett PK. Multiple maltreatment and the development of self and emotion regulation. *Journal of Aggression, Maltreatment & Trauma*. 1998; 2(1):171–187. [Print Electronic; Print].
- United States. General Accounting Office. *Child and Family Services Reviews: States and HHS Face Challenges in Assessing and Improving State Performance*. Washington, D.C.: General Accounting Office; 2004.
- US Department of Health and Human Services Administration for Children and Families Children's Bureau. *Child and Family Services Reviews Procedures Manual*. 2004. Retrieved 1/30/09, from http://www.acf.hhs.gov/programs/cb/cwmonitoring/tools_guide/proce_manual.htm
- Vladeck BC. Medicare hospital payment by diagnosis-related groups. *Annals of Internal Medicine*. 1984; 100(4):576–591. [PubMed: 6422818]
- Ware J Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Medical Care*. 1996; 34(3):220–233. [PubMed: 8628042]
- Weiner JP, Starfield BH, Steinwachs DM, Mumford LM. Development and application of a population-oriented measure of ambulatory care case-mix. *Medical Care*. 1991; 29(5):452–472. [PubMed: 1902278]
- Weissman JS, Wachterman M, Blumenthal D. When methods meet politics: How risk adjustment became part of Medicare managed care. *Journal of Health Politics, Policy and Law*. 2005; 30(3): 475–504.
- Williamson DF, Thompson TJ, Anda RF, Dietz WH, Felitti V. Body weight and obesity in adults and self-reported abuse in childhood. *International Journal of Obesity and Related Metabolic Disorders*. 2002; 26(8):1075–1082. [PubMed: 12119573]
- Wulczyn, F. *Beyond common sense: child welfare, child well-being, and the evidence for policy reform*. New Brunswick, N.J.: AldineTransaction; 2005.
- Wulczyn F, Chen L, Orlebeke B. Evaluating contract agency performance in achieving reunification. *Children and Youth Services Review*. 2009; 31(5):506–512.
- Wulczyn F, Orlebeke B, Melamid E. Measuring contract agency performance with administrative data. *Child Welfare*. 2000; 79(5):457–474. [PubMed: 11021342]

Table 1

Child-level outcomes as measured within the Child and Family Services Reviews mechanism.

Safety	Safety outcome 1: children are, first and foremost, protected from abuse and neglect	<p>Timeliness of initiating investigations of reports of child maltreatment</p> <p>Repeat maltreatment</p> <p>Absence of recurrence of maltreatment</p> <p>Absence of maltreatment of children in foster care</p>
	Safety outcome 2: children are safely maintained in their homes whenever possible and appropriate.	<p>Services to family to protect child(ren) in home and prevent removal or re-entry into foster care</p> <p>Risk assessment and safety management</p>
Permanency	Permanency outcome 1: children have permanency and stability in their living situations.	<p>Foster care re-entries</p> <p>Stability of foster care placement</p> <p>Permanency goal for child</p> <p>Reunification, guardianship, or permanent placement with relatives</p> <p>Adoption</p> <p>Other planned permanent living arrangement</p> <p>Timeliness and permanency of reunifications</p> <p>Timeliness of adoptions</p> <p>Achieving permanency for children in foster care</p> <p>Placement stability</p>
	Permanency outcome 2: the continuity of family relationships and connections is preserved for children.	<p>Proximity of foster care placement</p> <p>Placement with siblings</p> <p>Visiting with parents and siblings in foster care</p> <p>Preserving connections</p> <p>Relative placement</p> <p>Relationship of child in care with parents</p>
Child and family well-being	Child and family well-being outcome 1: families have enhanced capacity to provide for their children's needs.	<p>Needs and services of child, parents, and foster parents</p> <p>Child and family involvement in case planning</p> <p>Caseworker visits with child</p> <p>Caseworker visits with parent(s)</p>
	<p>Child and family well-being outcome 2: children receive appropriate services to meet their educational needs.</p> <p>Child and family well-being outcome 3: children receive adequate services to meet their physical and mental health needs.</p>	<p>Educational needs of the child</p> <p>Physical health of the child</p> <p>Mental/behavioral health of the child</p>

From Children's Bureau. (US Department of Health and Human Services Administration for Children and Families Children's Bureau., 2004).