

Challenges and Opportunities in Nurse Credentialing Research Design

Matthew D. McHugh, Richard E. Hawkins, Paul E. Mazmanian, Patrick S. Romano,
Herbert L. Smith, and Joanne Spetz

August 19, 2014

The views expressed in this discussion paper are those of the authors and not necessarily of the authors' organizations or of the Institute of Medicine. The paper is intended to help inform and stimulate discussion. It has not been subjected to the review procedures of the Institute of Medicine and is not a report of the Institute of Medicine or of the National Research Council.

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES

Advising the nation • Improving health

Copyright 2014 by the National Academy of Sciences. All rights reserved.

Challenges and Opportunities in Nurse Credentialing Research Design

Matthew D. McHugh, University of Pennsylvania; Richard E. Hawkins, American Medical Association; Paul E. Mazmanian, Virginia Commonwealth University; Patrick S. Romano, University of California, Davis; Herbert L. Smith, University of Pennsylvania; and Joanne Spetz, University of California, San Francisco¹

Many of the approximately 3 million registered nurses (RNs) in the U.S (IOM, 2011; HRSA, 2013) pursue additional credentials beyond their basic nursing education. Health care organizations, such as hospitals or nursing homes, also pursue credentials. The intent of credentialing is to demonstrate that a nurse or institution has met established standards that can be uniformly tested and validated. For example, individual nurses often pursue certification in a particular area of nursing, such as critical care or oncology, to establish their competence. Organizations may pursue an accreditation that distinguishes them as having specific characteristics related to quality, services provided, and/or their mission. Health care institutions and organizations, for example, can pursue accreditation through the Joint Commission, which confirms that the institution has met safety and quality standards. The Magnet Recognition Program of the American Nurses Credentialing Center (ANCC) recognizes organizations that demonstrate professional nurse work environments associated with high-quality nursing and good patient outcomes.

Credentialing may improve quality, or it may signal differences in quality, to consumers and employers. Credentialing research can focus on the process of credentialing or the value of the credential as perceived or experienced by stakeholders, including individual nurses, employers, patients, or the general public. A notable aspect of credentialing in health care is that there are so many stakeholders. The productive value that can arise from the pursuit of credentials is thus indexed not only by something negotiated between nurses and employers, such as a wage, but also in the various measures of outcomes for patients. Whether credentialing within nursing actually improves or signals better quality depends a great deal on researchers' ability to produce convincing evidence that there is an effect on health care outcomes. To design studies that would provide such evidence, we must consider an underlying set of challenges that might narrowly be classified as "methods" issues, but are in fact much broader. We present several questions meant to guide and challenge researchers conducting credentialing research and highlight some of the issues that must be considered when designing credentialing studies specific to nursing research.

WHAT DOES A CREDENTIAL REPRESENT?

A credential is an indication that an individual, group, or organization has been evaluated by a qualified and objective third-party credentialing body and was determined to have met standards that are defined, published, psychometrically sound, and legally defensible (Hickey et

¹ Suggested citation: McHugh, M. D., R. E. Hawkins, P. E. Mazmanian, P. S. Romano, H. L. Smith, and J. Spetz. 2014. *Challenges and opportunities in nurse credentialing research design*. Discussion Paper, Institute of Medicine, Washington, DC. <http://nam.edu/wp-content/uploads/2015/06/credentialingresearchdesign>.

al., 2014; NOCA, 2005; U.S. Department of Labor, 2014). There are different types of credentials, however, and the differences between them are important to consider in relation to what might be expected in terms of outcomes attributable to each. The National Organization for Competency Assurance (NOCA)² defines four major types of credentials: licensure, professional certification, registration, and accreditation (NOCA, 2005). The definitions are shown in Box 1.

Box 1 Definitions of Types of Credentials	
Licensure	The mandatory process by which a governmental agency grants time-limited permission to an individual to engage in a given occupation after verifying that he/she has met predetermined and standardized criteria and offers title protection for those who meet the criteria
Certification	The voluntary process by which a non-governmental entity grants a time-limited recognition and use of a credential to an individual after verifying that he or she has met predetermined and standardized criteria. It is the vehicle that a profession or occupation uses to differentiate among its members, using standards, sometimes developed through a consensus-driven process, based on existing legal and psychometric requirements.
Registration	[First, registration is defined as] the process by which a governmental agency grants a time-limited status on a registry, determined by specified knowledge-based requirements (e.g., experience, education, examinations), thereby authorizing those individuals to practice, similar to licensure. Its purpose is to maintain a continuous record of past and current occupational status of that individual and to provide title protection. A second meaning of registration is a listing of practitioners maintained by a governmental entity, without educational, experiential, or competency-based requirements; for example, maintaining a list of practitioners on a state registry. A third use of the term registration is a professional designation defined by a governmental entity in professional regulations or rules. However, the governmental regulatory body does not itself maintain a listing or registry of those who purport to meet registration requirements. Verification and authentication of such individuals are left to the employer of the individual claiming to be registered.
Accreditation	A voluntary process by which a nongovernmental agency grants a time-limited recognition to an institution, organization, business, or other entity after verifying that it has met predetermined and standardized criteria.
NOTE: Box 1 was created based on definitions of credential types, which originally appear in a 2005 report by the National Organization for Competency Assurance.	
SOURCE: NOCA, 2005.	

² Since 2009, NOCA has been known as the Institute for Credentialing Excellence (ICE).

At the individual provider level, licensure, certification, and registration are intended to recognize levels of achievement and competence. Generally speaking, licensure indicates attainment of a minimal threshold of competence necessary to provide care in an unsupervised context. Professional certification, on the other hand, generally includes requirements for advanced training and additional assessment(s). There are at least three meanings for registration, as shown in Box 1. Educational programs or health care institutions seek accreditation as a marker signaling to the public and various stakeholders that the institution has met agreed-on standards. The different types of credentialing are often linked. For example, in medicine, licensure and American Board of Medical Specialties certification standards include requirements for completion of an accredited training program (Sharp et al., 2002; USMLE, 2013). There can also be linkage across the individual and organization levels, whereby the organizational credential may depend, in part, on the credentialing status of individuals within that organization. Those who conduct credentialing research and those who use science in their program and policy decisions must understand the underlying differences in the types of credentials, as well as the psychometric properties of various assessment approaches for each.

In addition to considering the different definitions of the various credentials, investigators who conduct credentialing research must ask the following: *What does the focus of my research—the credential—represent conceptually?* Human capital theory suggests that the process of attaining education is evidence of acquired skill (Becker, 1962; Schultz, 1961; Sweetland, 1996; Weisbrod, 1962). Thus, the credentialing process should facilitate skill acquisition through additional education, training, and assessment. The credential signals that this process has occurred and that skills have been attained. At the organizational level, the credentialing process should lead to organizational change, and the credential indicates attainment of specified organizational and/or cultural benchmarks. Viewed from a human capital perspective, the person or organization is fundamentally changed for having gone through the credentialing process and is expected to perform better because of the acquired skill or knowledge.

There is an alternate perspective—signals of educational and skill attainment (e.g., credentials) might be a marker of intrinsic traits that those who pursue the credential have, rather than specific skills attained (Arrow, 1973; Spence, 1973; Stiglitz, 1975; Weiss, 1995). For the individual, these traits might be intelligence, motivation, or wealth, and for the organization, these traits might be baseline resources, the market environment, or mix of patients. From this perspective, the credential is similar to a seal of approval from the Good Housekeeping Institute or the Underwriters Laboratories, which indicates the inherent safety and quality of products.

From the human capital perspective, credentialing is attractive because it can encourage improvement in the person's or organization's performance. We see support for this perspective in education generally where evidence suggests, for example, that the greatest benefits of college education are not seen by those who were better off to begin with, but rather by those least likely or able to pursue a college education (Hout, 2012). Research questions that arise from this perspective include the following: *What are the results of investing in human capital, and to whom do they accrue? What are the implications for how credentialing should be organized and financed?*

If, however, the credential is a signal of preexisting traits, then encouraging the pursuit of credentialing by those who are not credentialed and likely would not have pursued credentialing may not lead to improved performance. This is because these persons or organizations do not possess the innate traits that deliver the difference in performance seen between those with and

those without the credential. In this case, the credential would be valuable as a signal because the credential differentiates the preexisting traits associated with quality for consumers, colleagues, competitors, and employers. If we assume that the credential accurately reflects quality, this information can improve efficiency and overall quality in the health care market.

Let us extend this discussion of what credentialing represents to an organizational credentialing example: Magnet hospitals. Studies have shown that, with few exceptions (e.g., Trinkoff et al., 2010), care environments in hospitals recognized by the ANCC's Magnet Recognition Program are better than those in non-Magnet hospitals (Kelly et al., 2011; Lacey et al., 2007; Lake and Friese, 2006). Furthermore, those differences are associated with better outcomes for nurses (Aiken et al., 2008; Upenieks, 2003) and patients (Dunton et al., 2007; Lake et al., 2010; McHugh et al., 2013). The program is respected and well known; Magnet recognition has been included as a criterion for national hospital ranking and quality benchmarking programs such as *U.S. News and World Report's* Best Hospitals rankings and the Leapfrog Group hospital ratings (Drenkard, 2010; Leapfrog Group, 2011; McClure and Hinshaw, 2002). For hospitals that wish to transform their work environment, the Magnet recognition process may provide a replicable blueprint. Choosing to undergo the process required to be recognized as a Magnet hospital, however, is not random, that is, it appears that some hospitals are more likely than others to pursue and ultimately attain Magnet recognition. The program is voluntary, and there are costs associated with engaging in the process. Evidence suggests that Magnet hospitals are more likely to be large, academic teaching hospitals located in places where there is a higher concentration of other Magnet hospitals (Abraham et al., 2011; Powers and Sanders, 2013).

This situation presents difficulties not only for researchers who are trying to understand the relationship between Magnet status and outcomes but also for administrators trying to determine whether Magnet recognition would be a worthwhile pursuit for their institution. Magnet recognition could be an indicator of preexisting quality. This idea would align with the signaling theory perspective discussed earlier. If this is the case, then the hospitals that could most benefit from the improvements attributed to Magnet recognition are exactly those hospitals least likely to participate. Another possibility is that the benefits seen in Magnet hospitals are an outcome of going through the Magnet recognition process, that is, going through the "Magnet journey" is transformational. It is likely that this is not an either/or situation; the credential probably signals a baseline level of performance and presence of requisite traits, and the process of attaining the credential also fundamentally changes the organization through acquisition of knowledge and skills. The balance between these two factors likely varies across institutions and credentialing programs. Both are likely relevant to acquiring a credential and, to different degrees, account for any observed outcomes differences.

Researchers should consider these phenomena, which introduce analytic challenges because of the alternative possibilities of what the credential represents and the often voluntary self-selection process (Berk, 1983; Heckman, 1979). To isolate the effect of the accreditation process itself, organizations would have to be randomly assigned to either engage in the Magnet journey or not. Conducting such a randomized experiment would be extremely challenging and is likely infeasible; thus, we must rely on observational study approaches. Although the randomized experiment has long been viewed as the gold standard of research design, it is not without limitations (Black, 1996). Randomized trials often have limited generalizability, and the results may not be sustained when applied in real-world settings (Dreyer et al., 2010). Most policy and practice decisions do not (and could not) rely on having perfect information drawn

from randomized experiments. The inability to conduct such an experiment should not discourage our pursuit of credentialing research.

Other research approaches, however, can be implemented to improve researchers' ability to assess the likely causality of the relationships measured in observational research. These approaches can help to identify and isolate the confounding effects of any preexisting advantage that individuals or organizations that pursue credentialing have from the credential itself. Typical approaches for addressing causation include conditioning on variables (matching, stratification, weighting, regression, or some combination); instrumental variables; and repeated data (interrupted time series, regression discontinuity, panel data) (Shadish et al., 2002; Winship and Morgan, 2007).

We see some examples of these analytic approaches in nurse credentialing research. Conditioning on variables in a regression framework to adjust for confounding is one of the more common analytic approaches for estimating causal effects in the social sciences (Winship and Morgan, 2007). The approach involves regressing the outcome of interest on the independent variable with the addition of control variables. Lake and colleagues (2010), for example, evaluated the relationship between hospital Magnet status and patient falls, including teaching intensity, hospital size, and ownership as control variables. Results indicate that even when controlling for hospital characteristics that differ between Magnet and non-Magnet hospitals and could account for outcomes differences, the fall rate was 5 percent lower in Magnet compared to non-Magnet hospitals. This approach, however, depends on the assumption (among others) that the model is correctly specified (i.e., it includes all relevant variables and takes the appropriate functional form) in order to obtain unbiased estimates of any causal effect. This may or may not be the case; claims of causality based on regression approaches that adjust for confounding alone are some of the most limited, even under the best circumstances (Freedman, 2010).

Matching is an intuitively appealing approach because the primary goal is to obtain treatment and control groups (e.g., credentialed vs. noncredentialed) that are as similar as possible in regard to the distribution of covariates (Rubin, 2006; Stuart, 2010). In an experimental framework, random assignment to the treatment or control group guarantees that the distribution of covariates is balanced between groups. Matching (and similar approaches involving weighting or subclassification) allows the researcher to create comparable treatment and control samples based on similarity of covariate distribution. Aiken and colleagues (1994) used a matching approach to examine the effects of Magnet status on inpatient mortality among Medicare beneficiaries. They matched Magnet and non-Magnet hospitals using a propensity score approach that balanced characteristics of the hospitals, including number of beds, average daily census, advanced technology service availability, area population size, and percentage of board-certified physicians, among other characteristics. They found significantly lower Medicare mortality rates for the Magnet hospitals compared to their otherwise similar non-Magnet counterparts.

The matching approach creates an apples-to-apples comparison so that, at least relative to observed variables, the treatment and control groups differ on average only in terms of their treatment status. In that sense, it is similar to an experiment because it focuses our attention on the effect of a treatment and moves us away from the effects of other factors on the outcome of interest. Matching reduces bias that is associated with imbalance in the covariate distributions between treated and control groups and helps to identify the population where the treatment effect would realistically be expected (Rubin and Thomas, 1996; Smith, 1997). Restricting observations within matching to treated cases and, at best, to a handful of close controls may

mean that the sample size of a matching analysis will only be a fraction of that relative to a multivariate regression. However, the truncation in the range of controlled covariates means that statistical efficiency is not compromised and may even be improved (Stuart, 2010). Matching can also be combined with regression (doubly robust methods) to provide additional protection from misspecified models (Funk et al., 2011). One approach to matching involves identifying a control group that “just barely” failed to meet standards required for attaining a credential and comparing it to a group that “just barely” attained that credential. This general approach, called regression discontinuity design (Imbens and Lemieux, 2008), has been used to study the impact of unionization on outcomes (DiNardo and Lee, 2004), which is another organizational characteristic that is not randomly selected.

The matching framework shares an important limitation with regression analysis and other methods for adjusting observational data: the potential presence of unmeasured factors. It is not uncommon to see skepticism regarding the effects of credentialing, especially when it is reasonable to assume that there are unmeasured aspects of nurses (or hospitals) that factor into both those who seek credentials and those who are productive. A good practice for evaluating such omitted-variable bias is to conduct sensitivity analyses to quantify how large the influence of an unmeasured factor would need to be in order to call into question findings that suggest a treatment effect (Rosenbaum, 2002). This practice helps quantify how sensitive (or resistant) the finding of treatment effect is to an unmeasured covariate. Although not often done in general, at least one study in nurse credentialing carried out an explicit sensitivity analysis (Witkoski Stimpfel et al., 2014). This practice should be more regularly carried out.

Another analytic approach is to use longitudinal data in repeated measures analysis; as noted earlier, examples include interrupted time series, regression discontinuity, and panel data methods (Shadish et al., 2002). Studies with longitudinal data allow the researcher to evaluate temporality (i.e., whether cause precedes the effect)—an essential criterion for establishing a causal relationship (Hill, 1965). A number of studies have taken advantage of panel data to examine the effect of credentialing at the organizational level (Boyle et al., 2012; Everhart et al., 2014; Jayawardhana et al., 2014), but fewer studies have used this approach to study individual-level credentialing. Using longitudinal data, one can include an individual-level or organization-level “fixed effect,” which is a dummy variable that controls for unknown time-invariant characteristics of the individual or organization.

One approach to addressing the nonrandom selection of individuals or organizations into credentialing is to calculate selectivity variables from regressions of the probability of credentialing. The most common implementation of this method, typically referred to as a Heckman sample selection model, involves estimating a probit regression in the first stage, with the dependent variable measuring the presence of a credential (Heckman, 1979). The inverse Mills ratio and related variables from this first regression can be added to the outcomes regressions to correct for selectivity (Lee, 1982). The success of this approach depends, at least in part, on the inclusion of explanatory variables in the first-stage probit regression, which can be excluded from the second-stage regression. Although exclusion restrictions are not required because the first-stage selection model is formally identified by its nonlinear functional form, convincing exclusion restrictions (i.e., variables that affect credentialing but are not directly associated with the outcome) lend more credence to the selection approach.

A second method for addressing nonrandom selection of credentialing involves using simultaneous equations to estimate the likelihood of credentialing and the effect of credentialing on outcomes; this approach is usually called instrumental variables (IV) regression or two-stage

least squares (2SLS) regression (Greene, 2011). This approach is similar to that of a Heckman sample selection model, except that the Heckman model incorporates selectivity variables in the second-stage equation, whereas standard instrumental variables techniques do not. A related approach is called two-stage residual inclusion, which is used when the outcome of interest is not a continuous variable (i.e., when estimating logistic or probit regressions) (Terza, 2008). The effectiveness of this approach depends on identifying appropriate IVs. An IV should be associated with treatment, should not affect the outcome except by treatment, and should not share any causes with the outcome (Angrist et al., 1996; Newhouse and McClellan, 1998). Jayawardhana and colleagues (2014) used an IV approach to show that, although becoming a Magnet-recognized hospital was associated with increased costs, these costs were likely offset by increased revenues.

The identification of appropriate IVs can be challenging. The success of instrumental variables regression is highly dependent on the strength of instruments applied, because the use of weak instruments can bias results (Bound et al., 1995). There is a lively debate in the literature regarding how one can determine if instruments are “strong enough,” and many tests are available to test the strength of proposed IVs (e.g., Andrews and Stock, 2005; Baum et al., 2003; Chernozhukov and Hansen, 2008; Dufour, 2003; Shea, 1997; Staiger and Stock, 1997; Stock et al., 2002). Jayawardhana and colleagues (2014) used the proportion of Magnet hospitals within each hospital market from the prior year as an instrument for a hospital becoming a Magnet, with the expectation that a hospital’s pursuit of a credential might be affected by its competitors’ credentials, but a hospital’s outcomes will not be affected directly by its competitors’ credentials. Other possibilities for instruments include using time or geography; if, for example, a credentialing program was rolled out earlier in some areas than in others for an exogenous reason, this would function as an IV. Despite the increase in the adoption of Magnet recognition as an organizational improvement strategy, little is known beyond cross-sectional studies (Abraham et al., 2011; Powers and Sanders, 2013) about factors that influence or impede the likelihood that a hospital will go through the Magnet recognition process. A similar lack of knowledge exists about the motivations for pursuing individual credentialing. More research is needed to answer the question: *Who pursues, or does not pursue, credentialing and why?* This information would help us understand whether there is an advantage to encouraging and supporting the credentialing of individuals or organizations that would otherwise be less likely to engage in the process on their own. This research would also help to identify the specific selection factors that should be controlled in analysis of outcomes (including through their use as IVs).

Another strategy that can be used to assess whether nonrandom selection is an issue leverages longitudinal data by testing the counterfactual directly, that is, discovering whether particular quality outcomes are present before becoming credentialed to learn whether the outcomes are the result of the credentialing process (Chamberlain, 1982). For example, one study (McHugh et al., 2013) determined that outcomes did not differ significantly between hospitals that eventually gained Magnet status and non-Magnet hospitals; in other words, the Magnet outcomes advantage did not appear to predate Magnet recognition. This kind of investigation could be expanded.

ARE ALL CREDENTIALS EQUAL?

A principal concern in conducting research, particularly for researchers trying to make causal inferences, is clearly defining the condition that is being studied—in our case, the credential. Regardless of the underlying rigor of the study, confidence in evidence is limited when credentials represent groupings of different items or characteristics. For individual nurse credentialing, the problem is compounded by many credentialing bodies and innumerable credentials (Cary, 2001). The 2008 National Sample Survey of Registered Nurses, for example, captured hundreds of possible credentials, ranging from functional credentials such as Basic Life Support, which most nurses have, to very specialized credentials such as clinical genetic nursing certification (HRSA, 2010). This is not to say that this diversity is bad for practice and quality, but it increases the difficulty for researchers to clearly identify and attribute a credentialing effect in the face of so much heterogeneity. A further complication is that some credentialing processes do not adhere to the objective standards required of the types of credentialing programs we have addressed herein. For example, some programs require only attendance, completion, or a passing grade on a knowledge-based assessment for awarding a certificate. Such certificates are ordinarily not seen as credentials because recipients are not required to demonstrate competence according to professional or trade standards (NOCA, 2005). In attempting to answer the research question, *Does credentialing have an effect on outcomes?*, it may be that certain credentials matter (or matter more) and that some do not.

In addition, there are significant areas of overlap and duplication. In some specialties, a nurse may be certified by multiple organizations. Thus, it may be difficult to discern which credential matters. This ambiguity presents an opportunity for credentialing organizations and credentialing oversight bodies (e.g., the National Commission for Certifying Agencies) to coordinate their efforts toward standardization and consistent definitions that are clear to researchers, data producers, and the health care community.

Organizational credentialing faces a somewhat different, but equally difficult, issue. The organizational credential is broad and encompasses multiple components. The basis for Magnet recognition, for example, includes five key domains: transformational leadership; structural empowerment; exemplary professional practice; new knowledge, innovations, and improvements; and empirical outcomes. Natural questions that arise include the following: *Are all components equally important for outcomes? Is there synergy in the collective set? What and how should credentialing standards change as evidence evolves?* To answer such questions, we must have available good data and valid, reliable measures that can identify and quantify the variation in different domains across institutions.

HOW DOES CREDENTIALING WORK TO PRODUCE BETTER OUTCOMES?

Improving our understanding of the process through which credentialing has its effect on outcomes would improve our assessment of the causal effects of credentialing (Hedström and Ylikoski, 2010; Imai et al., 2011). Such research would focus on the following question: *If credentialing leads to better outcomes, why?* If a plausible mechanism for why and how credentialing leads to better outcomes cannot be identified, evidence supporting a causal relationship is less compelling. An understanding of mechanisms also helps to understand and clarify theories of causal relationships, linking how we think the world works with evidence about what is actually happening.

Credentialing can influence the quality of care in several ways. In the case of individual nurse credentialing, it is possible that quality of care may differ between credentialed nurses and otherwise similar nurses without the credential because the credentialed nurses provide care differently. To get to this question of whether there are differences in process, we need to establish evidence that: (1) credentialed nurses work differently, and (2) there is a difference in outcomes that could plausibly be attributed to how the nurses work differently. Once an outcomes difference is established, then evaluations of differences in what credentialed and noncredentialed nurses do, and how they do it, may shed light on why patients of credentialed nurses have better outcomes. Qualitative or mixed-methods studies that involve in-depth interviews and direct observation of how credentialed and noncredentialed nurses practice differently may be useful.

Research has been conducted regarding the pathways through which organizational credentialing has its effect. In one study of Magnet hospitals (McHugh et al., 2013), investigators evaluated whether differences in the nurse work environment, a defining characteristic of Magnet hospitals associated with better outcomes (Aiken et al., 2011; Aiken et al., 2008; IOM, 2003; Kutney-Lee et al., 2009), are what account for the mortality advantage in Magnet hospitals. This turned out to be the case. Using linked data from nurses, patients, and hospitals, the investigators found significantly lower odds of death following surgery for patients cared for in Magnet compared to non-Magnet hospitals, after accounting for patient and hospital characteristics. Furthermore, this mortality advantage was partly explained by differences in the work environment as reported by nurses. Another study showed that differences in the quality of the nurse work environment explain differences in nurse-reported care quality seen between Magnet and non-Magnet hospitals (Witkoski Stimpfel et al., 2014). Longitudinal studies are needed to evaluate these processes. For instance, a longitudinal study could examine how changes in the nurse work environment (suggested as a mechanism explaining Magnet differences) vary as a hospital achieves Magnet status and how those changes relate to changes in outcomes.

A variant of the question about the mechanisms through which credentials affect change asks: *At what level do the mechanisms operate?* For example, it may be that, although the credentialing is at the individual level, the causal relationship and the target for intervention are at the aggregate level. The consideration for outcomes may thus be less about whether an individual nurse is credentialed, and more about whether the organization has many credentialed nurses. In other words, it may take a critical mass of credentialed nurses to influence outcomes, regardless of any individual nurse's credentialing status. If that is the case, then differences in the concentration of credentialed nurses between hospitals matter more for outcomes than differences between credentialed and noncredentialed nurses within hospitals. Another consideration is how the credentialing composition of nurses aligns with that of other health care providers. There may be synergy as more providers across disciplines become credentialed.

The research evaluating relationships between individual credentialing and outcomes has largely been carried out using individual-level data, aggregated to the unit or hospital level. This approach focuses on the proportion of nurses with a credential as an attribute of the clinical environment (e.g., the hospital or unit) (Kendall-Gallagher et al., 2011; Kendall-Gallagher and Blegen, 2009; Newhouse et al., 2005). The question of whether the key mechanisms operate at the aggregate versus individual level affects the recommendations suggested by research findings. For example, if having a high proportion of credentialed nurses is important, then it would be essential to understand the factors that make a workplace attractive to credentialed

nurses, as well as what differentiates hospitals and other employers that place a value on credentialing.

Another key issue is how credentialing works in different contexts. For individual nurses, the benefits of their credentialing may depend on the type, quality, and possibly the credential of their workplace, among other factors. Organizational credentialing is not limited to hospitals; the importance of credentialing may vary for different settings (e.g., hospital, nursing home, or ambulatory care). Joint Commission accreditation and ANCC's Magnet recognition program are examples of organizational credentialing that also occur outside the United States. Our understanding of how these credentialing initiatives work will be advanced with study of whether, and how, they translate to better outcomes in other countries.

WILL THERE BE DATA AND FUNDING?

In addition to an optimal design and advanced methods, we need two other key ingredients to develop a program of research: data and funding. A lack of data often dictates the gap between what we want to do as researchers and what is possible. For research on the impact of individual credentials, the gap is quite large. There are no data collection systems for identifying and tracking individual nurse credentialing. The National Sample Survey of Registered Nurses provided a general overview of the credentialing status of the nursing workforce in the United States. These data were useful for policy planning and labor research. This survey was discontinued as of 2008, however, despite it being a vital source of nursing workforce information and a potential resource for shedding light on additional questions with regard to the relationship between individual credentialing and employment setting, or nurses' income (Spetz, 2013). Survey research by individual investigators has also been utilized to investigate questions using information about individual nurse credentials (Kendall-Gallagher et al., 2011; Miller and Boyle, 2008). Such surveys, however, require substantial funding and are often not primarily focused on credentialing.

There are efforts to create a minimum data set on the nurse workforce that would depend on states collaborating in data collection efforts as part of the license renewal process (Spetz, 2013). A significant ramp-up in participation, funding, and political will are needed to make this minimum data set a viable option. Although questions about registered nurse credentialing are not included in the recommendations of the Forum of State Nursing Workforce Centers (Spetz, 2013), this information is an important component to include if and when the minimum data set is developed. Another opportunity is through the National Database of Nursing Quality Indicators (NDNQI),³ which collects hospital and unit-level data on nurses, nurse work environment characteristics, and quality outcomes. Specialty certification information is available at both the individual (Boyle et al., 2012; Miller and Boyle, 2008) and organizational level (Lake et al., 2010), which has been used for research. The NDNQI data, however, represent a select group of hospitals that elect to (and have the resources to) participate in the benchmarking enterprise. In addition, confidentiality restrictions currently limit investigators' ability to link the NDNQI data with external outcomes data and data that could be used to adjust for confounding. Examples do exist of successful research on the relationship between an individual health care provider's credentialing and patient outcomes outside of nursing, which demonstrate the utility of linked data (Chen et al., 2006; Curtis et al., 2009; Silber et al., 2002). Using the more widely available data on physician credentials, for example, Silber and

³ NDNQI was formerly owned by the American Nurses Association. It is now owned by Press-Ganey.

colleagues found provider-level differences in outcomes for patients cared for by board-certified anesthesiologists compared with those who were not board certified (Silber et al., 2002). Currently, research that involves linking individual nurses with their patients across a large number of hospitals is virtually impossible to execute. Excessive data restrictions increase the cost and difficulty of conducting essential research on the relationship between credentialing and patient outcomes and dissuade researchers from pursuing credentialing research questions.

As hospitals and other health care delivery systems begin to pursue big data and analytics solutions, along with implementation of electronic medical records and documentation systems, there are significant opportunities for advancing credentialing data. This effort requires the thoughtful integration of standardized human resources information with the collection and maintenance of credentialing information. Data collection and data systems should be developed with the preceding discussion in mind, and data systems and elements should be designed to capture information that would allow researchers to carry out studies that eliminate, or at least shed light on, some of the current constraints around causal inference. For instance, much of the research on Magnet hospital status is carried out with sparse information from the ANCC that, at most, identifies when a hospital achieved Magnet status, when and how many times a hospital was redesignated (hospitals must undergo periodic reviews to maintain the recognition), and whether and when a hospital lost Magnet status. There is no information pertaining to why a hospital did not maintain its Magnet status, nor is there information on the hospitals that pursued Magnet but did not achieve it. These pieces of information would facilitate evaluations that clarify issues of self-selection, as well as causal mechanisms and relationships with outcomes.

Funding for nurse credentialing research is even more limited than the available data. Of course, funding is more likely to be available if there is an evidence base suggesting that the topic is likely to have a meaningful impact. The increasing emphasis on patient safety and quality of care suggests that there is potential to attract funding for research that explores the role of credentialing individuals or organizations in improving safety and other domains of quality. Nevertheless, given the foundational research needed to gain the attention of federal funders such as the National Institutes of Health, other funding strategies are necessary to carry out the kind of research we have discussed. One possibility is that credentialing organizations allocate a proportion of their revenues to research by independent investigators.

CONCLUSION

Greater attention is being paid to the health workforce as a principle mechanism to achieve organizational change and improve health care quality. Much of the capacity of the workforce to contribute to high-quality and efficient care is determined by education, licensing, and certification and whether and how organizations value these factors. Our current understanding based on nurse credentialing research hints at the potential for important gains in quality, safety, and cost-effectiveness. This research, however, is relatively nascent and has yet to take full advantage of all available analytic approaches and conceptual considerations. Carefully designed and reasonably funded research using good quality data could provide improved evidence that would increase legitimacy and guide improvements in a credentialing process. In addition to the areas of inquiry posed throughout this discussion, priority questions include the following:

- How should credentialing programs be designed and implemented to improve patient outcomes? What are the critical components of these programs?

- What motivates individuals and organizations to pursue credentialing?
- Should credentialing be viewed as a journey or as an achievement? In other words, should all eligible individuals and organizations be encouraged to take the journey, or should only the best be recognized for their achievement?
- Given limited resources, as well as the monetary and time commitments involved, is credentialing cost-effective from different stakeholders' perspectives?
- Who should pay for credentialing? Should it be viewed as a private good with its own market-based rewards? Or should it be viewed as a public good that deserves public investment?

Answers to these and other questions based on findings from rigorous studies can be used to shape credentialing programs and increase the value of investment in credentialing by individuals and organizations.

REFERENCES

- Abraham, J., B. Jerome-D'Emilia, and J. W. Begun. 2011. The diffusion of Magnet hospital recognition. *Health Care Management Review* 36(94):306–314.
- Aiken, L. H., J. P. Cimiotti, D. M. Sloane, H. L. Smith, L. Flynn, and D. F. Neff. 2011. Effects of nurse staffing and nurse education on patient deaths in hospitals with different nurse work environments. *Medical Care* 49(12):1047–1053.
- Aiken L. H., S. P. Clarke, D. M. Sloane, E. T. Lake, and T. Cheney. 2008. Effects of hospital care environment on patient mortality and nurse outcomes. *Journal of Nursing Administration* 38(5):223–229.
- Aiken, L. H., H. L. Smith, and E. T. Lake. 1994. Lower Medicare mortality among a set of hospitals known for good nursing care. *Medical Care* 32(8):771–787.
- Andrews, D. W. K., and J. H. Stock. 2005. Inference with weak instruments. NBER working paper no. t0313, <http://ssrn.com/abstract=795264> (accessed June 23, 2014).
- Angrist, J. D., G. W. Imbens, and D. B. Rubin. 1996. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91(434):444–455.
- Arrow, K. J. 1973. Higher education as a filter. *Journal of Public Economics* 2(3):193–216.
- Baum, C. F., M. E. Schaffer, and S. Stillman. 2003. Instrumental variables and GMM: Estimation and testing. Working paper no. 545, <http://www.stata.com/meeting/2nasug/wp545.pdf> (accessed June 23, 2014).
- Becker, G. S. 1962. Investment in human capital: A theoretical analysis. *Journal of Political Economy* 70(5):9–49.
- Berk, R. A. 1983. An introduction to sample selection bias in sociological data. *American Sociological Review* 48(3):386–398.
- Black, N. 1996. Why we need observational studies to evaluate the effectiveness of health care. *British Medical Journal* 312(7040):1215–1218.
- Bound, J., D. A. Jaeger, and R. M. Baker. 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association* 90(430):443–450.

- Boyle, D. K., B. J. Gajewski and P. A. Miller. 2012. A longitudinal analysis of nursing specialty certification by Magnet® status and patient unit type. *Journal of Nursing Administration* 42(12):567–573.
- Cary, A. H. 2001. Certified registered nurses: Results of the study of the certified workforce. *American Journal of Nursing* 101(1):44–52.
- Chamberlain, G. 1982. Multivariate regression models for panel data. *Journal of Econometrics* 18(1):5–46.
- Chen, J., S. S. Rathore, Y. Wang, M. J. Radford, and H. M. Krumholz. 2006. Physician board certification and the care and outcomes of elderly patients with acute myocardial infarction. *Journal of General Internal Medicine* 21(3):238–244.
- Chernozhukov, V., and C. Hansen. 2008. The reduced form: A simple approach to inference with weak instruments. *Economics Letters* 100(1):68–71.
- Curtis, J. P., J. J. Luebbert, Y. Wang, S. S. Rathore, J. Chen, P. A. Heidenreich, S. C. Hammill, R. I. Lampert, H. M. Krumholz. 2009. Association of physician certification and outcomes among patients receiving an implantable cardioverter-defibrillator. *Journal of the American Medical Association* 301(16):1661–1670.
- DiNardo, J., and D. S. Lee. 2004. Economic impacts of new unionization on private sector employers: 1984–2001. *Quarterly Journal of Economics* 119(4):1383–1441.
- Drenkard, K. 2010. The business case for Magnet. *Journal of Nursing Administration* 40(6):263–271.
- Dreyer, N. A., S. R. Tunis, M. Berger, D. Ollendorf, P. Mattox, and R. Gliklich. 2010. Why observational studies should be among the tools used in comparative effectiveness research. *Health Affairs* 29(10):1818–1825.
- Dufour, J. M. 2003. Identification, weak instruments, and statistical inference in econometrics. *Canadian Journal of Economics/Revue canadienne d'économie* 36(4):767–808.
- Dunton, N., B. Gajewski, S. Klaus, and B. Pierson. 2007. The relationship of nursing workforce characteristics to patient outcomes. *Online Journal of Issues in Nursing* 12(3):1–15.
- Everhart, D., J. R. Schumacher, R. P. Duncan, A. G. Hall, D. F. Neff, and R. I. Shorr. 2014. Determinants of hospital fall rate trajectory groups: A longitudinal assessment of nurse staffing and organizational characteristics. *Health Care Management Review*, doi: 10.1097/HMR.000000000000013.
- Freedman, D. A. 2010. *Statistical models and causal inference: A dialogue with the social sciences*. New York: Cambridge University Press.
- Funk, M. J., D. Westreich, C. Wiesen, T. Stürmer, M. A. Brookhart, and M. Davidian. 2011. Doubly robust estimation of causal effects. *American Journal of Epidemiology* 173(7):761–767.
- Greene, W. H. 2011. *Econometric analysis*, 7th ed. Upper Saddle River, NJ: Prentice Hall.
- Heckman, J. J. 1979. Sample selection bias as a specification error. *Econometrica* 47(1):153–161.
- Hedström, P., and P. Ylikoski. 2010. Causal mechanisms in the social sciences. *Annual Review of Sociology* 36:49–67.
- Hickey, J. V., L. R. Unruh, R. P. Newhouse, M. Koithan, M. E. Johantgen, R. G. Hughes, K. B. Haller, V. A. Lundmark. 2014. Credentialing: The need for a national research agenda. *Nursing Outlook* 62(2):119–127.
- Hill, A. B. 1965. The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine* 58(5):295–300.

- Hout, M. 2012. Social and economic returns to college education in the United States. *Annual Review of Sociology* 38:379–400.
- HRSA (Health Resources and Services Administration). 2010. *The registered nurse population: Initial findings from the 2008 National Sample Survey of Registered Nurses*. Washington, DC: U.S. Department of Health and Human Services, Health Resources and Services Administration.
- HRSA. 2013. *The U.S. nursing workforce: Trends in supply and education*, <http://bhpr.hrsa.gov/healthworkforce/reports/nursingworkforce/nursingworkforcefullreport.pdf> (accessed August 4, 2014).
- Imai, K., L. Keele, D. Tingley, and T. Yamamoto. 2011. Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review* 105(4):765–789.
- Imbens, G. W., and T. Lemieux. 2008. Regression discontinuity designs: A guide to practice. *Journal of Econometrics* 142(2):615–635.
- IOM (Institute of Medicine). 2003. *Keeping patients safe: Transforming the work environment of nurses*. Washington, DC: The National Academies Press.
- IOM. 2011. *The future of nursing: Leading change, advancing health*. Washington, DC: The National Academies Press.
- Jayawardhana, J., J. M. Welton, and R. C. Lindrooth. 2014. Is there a business case for Magnet hospitals? Estimates of the cost and revenue implications of becoming a Magnet. *Medical Care* 52(5):400–406.
- Kelly, L. A., M. D. McHugh, and L. H. Aiken. 2011. Nurse outcomes in Magnet® and non-Magnet® hospitals. *Journal of Nursing Administration* 41(10):428–433.
- Kendall-Gallagher, D., L. H. Aiken, D. M. Sloane, and J. P. Cimiotti. 2011. Nurse specialty certification, inpatient mortality, and failure to rescue. *Journal of Nursing Scholarship* 43(2):188–194.
- Kendall-Gallagher, D., and M. A. Blegen. 2009. Competence and certification of registered nurses and safety of patients in intensive care units. *American Journal of Critical Care* 18(2):106–113.
- Kutney-Lee, A., M. D. McHugh, D. M. Sloane, J. P. Cimiotti, L. Flynn, D. F. Neff, and L. H. Aiken. 2009. Nursing: A key to patient satisfaction. *Health Affairs (Millwood)* 28(4):w669–w677.
- Lacey, S. R., K. S. Cox, K. C. Lorfing, S. L. Teasley, C. A. Carroll, and K. Sexton. 2007. Nursing support, workload, and intent to stay in Magnet, Magnet-aspiring, and non-Magnet hospitals. *Journal of Nursing Administration* 37(4):199–205.
- Lake, E. T., and C. R. Friese. 2006. Variations in nursing practice environments: Relation to staffing and hospital characteristics. *Nursing Research* 55(1):1–9.
- Lake, E. T., J. Shang, S. Klaus, and N. E. Dunton. 2010. Patient falls: Association with hospital Magnet status and nursing unit staffing. *Research in Nursing and Health* 33(5):413–425.
- Leapfrog Group. 2011. *Hospitals reporting Magnet status to the 2011 Leapfrog Hospital Survey*, <http://www.leapfroggroup.org/MagnetRecognition> (accessed April 25, 2014).
- Lee, L-F. 1982. Some approaches to the correction of selectivity bias. *Review of Economic Studies* 49(3):355–372.
- McClure, M. L., and A. S. Hinshaw. 2002. *Magnet hospitals revisited: Attraction and retention of professional nurses*. Washington, DC: American Nurses Association.

- McHugh, M. D., L. A. Kelly, H. L. Smith, E. S. Wu, J. Vanak, and L. H. Aiken. 2013. Lower mortality in Magnet hospitals. *Medical Care* 51(5):382–388.
- Miller, P. A., and D. K. Boyle. 2008. Nursing specialty certification: A measure of expertise. *Nursing Management* 39(10):10–16.
- Newhouse, J. P., and M. McClellan. 1998. Econometrics in outcomes research: The use of instrumental variables. *Annual Review of Public Health* 19:17–34.
- Newhouse, R. P., M. Johantgen, P. J. Pronovost, and E. Johnson. 2005. Perioperative nurses and patient outcomes: Mortality, complications, and length of stay. *AORN Journal* 81(3):508–509.
- NOCA (National Organization for Competency Assurance). 2005. *The NOCA guide to understanding credentialing concepts*. Washington, DC: National Organization for Competency Assurance.
- Powers, T. L., and T. J. Sanders. 2013. Environmental and organizational influences on Magnet hospital recognition. *Journal of Healthcare Management* 58(5):353–366.
- Rosenbaum, P. R. 2002. *Observational studies*. New York: Springer.
- Rubin, D. B. 2006. *Matched sampling for causal effects*. New York: Cambridge University Press.
- Rubin, D. B., and N. Thomas. 1996. Matching using estimated propensity scores: Relating theory to practice. *Biometrics* 52(1):249–264.
- Schultz, T. W. 1961. Investment in human capital. *American Economic Review* 51(1):1–17.
- Shadish, W. R., T. D. Cook, and D. T. Campbell. 2002. *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Sharp, L. K., P. G. Bashook, M. S. Lipsky, S. D. Horowitz, and S. H. Miller. 2002. Specialty board certification and clinical outcomes: The missing link. *Academic Medicine* 77(6):534–542.
- Shea, J. 1997. Instrument relevance in multivariate linear models: A simple measure. *Review of Economics and Statistics* 79(2):348–352.
- Silber, J. H., S. K. Kennedy, O. Even-Shoshan, W. Chen, R. E. Mosher, A. M. Showman, and D. E. Longnecker. 2002. Anesthesiologist board certification and patient outcomes. *Anesthesiology* 96(5):1044–1052.
- Smith, H. L. 1997. Matching with multiple controls to estimate treatment effects in observational studies. *Sociological Methodology* 27:325–353.
- Spence, M. 1973. Job market signaling. *Quarterly Journal of Economics* 87(3):355–374.
- Spetz, J. 2013. The research and policy importance of nursing sample surveys and minimum data sets. *Policy, Politics, and Nursing Practice* 14(1):33–40.
- Staiger, D., and J. H. Stock. 1997. Instrumental variables with weak instruments. *Econometrica* 65(3):557–586.
- Stiglitz, J. E. 1975. The theory of “screening,” education, and the distribution of income. *American Economic Review* 65(3):283–300.
- Stock, J. H., J. H. Wright, and M. Yogo. 2002. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics* 20(4):518–529.
- Stuart, E. A. 2010. Matching methods for causal inference: A review and a look forward. *Statistical Science* 25(1):1–21.
- Sweetland, S. R. 1996. Human capital theory: Foundations of a field of inquiry. *Review of Educational Research* 66(3):341–359.

- Terza, J. V., A. Basu, and P. J. Rathouz. 2008. Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. *Journal of Health Economics* 27(3):531–543.
- Trinkoff, A. M., M. Johantgen, C. L. Storr, K. Han, Y. Liang, A. P. Gurses, and S. Hopkinson. 2010. A comparison of working conditions among nurses in Magnet and non-Magnet hospitals. *Journal of Nursing Administration* 40(7/8):309–315.
- Upenieks, V. V. 2003. The interrelationship of organizational characteristics of Magnet hospitals, nursing leadership, and nursing job satisfaction. *Health Care Manager* 22(2):83–98.
- U.S. Department of Labor, Employment and Training Administration. 2014. *Credential resource guide*, <http://wdr.doleta.gov/directives/attach/TEGL15-10a2.pdf> (accessed May 15, 2014).
- USMLE (United States Medical Licensing Examination). 2013. *United States Medical Licensing Examination (USMLE) 2013 bulletin of information*, <http://www.usmle.org/pdfs/bulletin/2013bulletin.pdf> (accessed August 4, 2014).
- Weisbrod, B. A. 1962. Education and investment in human capital. *Journal of Political Economy* 70(5):106–123.
- Weiss, A. 1995. Human capital vs. signalling explanations of wages. *Journal of Economic Perspectives* 9(4):133–154.
- Winship, C., and S. L. Morgan. 2007. *Counterfactuals and causal inference: Methods and principles for social research*. New York: Cambridge University Press.
- Witkoski Stimpfel, A., J. Rosen, and M. D. McHugh. 2014. Understanding the role of the professional practice environment on quality of care in Magnet and Non-Magnet hospitals. *Journal of Nursing Administration* 44(1):10–16.