NORTHWESTERN UNIVERSITY

Science in the Clinic:
HIV Research in the Era of Evidence-Based Medicine

A DISSERTATION

SUUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Sociology

by

JuLeigh Petty

Evanston, Il

December 2008

ABSTRACT

Science in the Clinic: HIV Research in the Era of Evidence-Based Medicine

JuLeigh Petty

This is a study of the conduct and consumption of statistical medical research HIV/AIDS clinics in the context of the expansion of domestic and international clinical research and evidence-based medicine. Evidence-based medicine is the most recent and most successful attempt at subjecting medical decisions to statistical measurement and control. The dissertation is divided into three substantive sections. The first examines how scientific evidence is defined and valued in formal evidence-based medicine standard of care guidelines to how evidence is produced and judged at the clinic-level. Second, I examine the organizational processes of conducting research and argue that the introduction of new jobs, hierarchies and technologies potentially influence clinical work as much as the subsequent implementation of research results. Finally, I examine how physicians and nurses actually conduct medical statistical research in local settings.

Theoretically, this dissertation expands on an emerging sociology of medical science, an area of inquiry that concerns the content of medical scientific knowledge, its underlying modes of reasoning and the processes of medical research. The mandate to produce and use numbers entails more than simply reporting the numbers. Translating material, especially human material, into quantifiable categories is a social process. This dissertation traces how staff working in dispersed locations fit the messy reality of the clinic to the uniform reporting requirements of clinical research and how, in turn, the

requirements of research alter the organization of the clinic in ways that ease the production and consumption of research results. In short, this dissertation is about making and using numbers and the rules, procedures, and infrastructure that make quantification possible. The dissertation draws on fieldwork observations and interviews in four HIV clinics in the United States of America, Uganda and South Africa.

ACKNOWLEDGEMENTS

All accomplishments are collective enterprises. This dissertation is no exception. I am lucky to have had such a talented dissertation committee including Carol Heimer, Arthur Stinchcombe and Wendy Espeland. My dissertation advisor, Carol Heimer, has been especially influential and supportive. It is from Carol that I first learned fieldwork methods, and this dissertation grew out of my work on a research project with her. I thank Arthur Stinchcombe for many interesting conversations, his kindness, and his detailed and useful comments on drafts. I am also grateful to Carol and Art for their seemingly non-academic support that is so crucial to actually getting academic work done. This support ranged from encouragement and coffees to babysitting and home-cooked meals. Wendy Espeland's teaching and scholarship have been particularly influential on my theoretical approach.

This dissertation would not have been possible without the cooperation of the physicians, nurses, staff, and patients at the HIV/AIDS clinics observed as part of my research. I am grateful for their time and insights. This dissertation was also made possible by the funding and support of the American Bar Foundation. I would also like to thank my friends at Northwestern University whose intellectual conversation and humor helped me get through graduate school: Ellen Berry, Rebecca Culyba, Alan Czaplicki, Steve Hoffman, Peter Levin, and Nate Wright.

My family has been very supportive during this long journey. My mother, Lucinda, my brother, Michael, and my uncle, Martin, are ridiculously proud of me and have helped me in countless ways. I especially want to thank them for refraining from

constantly asking that dreaded graduate school question – when are you going to be finished? I am most thankful to my husband, Chris, for supporting me and putting up with me for all these years of graduate school and dissertation work. Finally, our children, Sam and Lucy, though not particularly helpful, made the years of research and writing more pleasurable.

TABLE OF CONTENTS

LIST OF TABLES

TABLES, cont.

Chapter 1. Introduction to Statistics and Standards in Medical Work

Research staff discussing ACTG 5095 study patient:

> A study nurse and a physician researcher talk gravely about the lipodystrophy of a female study patient, agreeing that it's really "terrible" and that she is "really getting deformed." They talk about the degree of distortion – "stick-skinny legs," a huge buffalo hump, a lot of fat on her chest. They describe her as being "shaped like a cube" with her head disappearing into the hump, and say it's been getting rapidly worse over the last 4 months. This can be called a grade 3 or 4 as a symptom. They want to end her participation in ACTG 5095. According to the study protocol, a primary physician can ask to be "unblinded," and they decide that that's what they should request. If it turns out that she is taking Abacavir, they'll get her off of it. Another physician joins the conversation. They compare her symptoms to those of some other women patients and talk about remediation. She's on public aid, and they want to send her to a clinic in the hospital to do surgery to get rid of some of the fat. The physician says that he will write a note to the protocol team today reporting the severe lipodystrophy and requesting to be "unblinded." (US1 040202 JP)[1]

Results of ACTG 5095 published in the *New England Journal of Medicine*:

> *Results:* We enrolled a total of 1147 subjects with a mean baseline HIV-1 RNA level of 4.85 log10 (71,434) copies per milliliter and a mean CD4 cell count of 238 ….After a median follow-up of 32 weeks, 82 of 382 subjects in the triple-

> nucleoside group (21 percent) and 85 of 765 of those in the combined Efavirenz groups (11 percent) had virologic failure; the time to virologic failure was significantly shorter in the triple-nucleoside group (P<0.001).…Changes in the CD4 cell count and the incidence of grade 3 or grade 4 adverse events did not differ significantly between the groups. (Gulick, et al., 2004)

"ACTG 5095" was a randomized clinical trial designed and organized by the US National Institutes of Health's AIDS Clinical Trials Group (ACTG), the largest HIV clinical research organization in the world. A randomized clinical trial measures and compares the effect of a medical intervention in two or more groups. The purpose of ACTG 5095 was to determine how effective Trizivir was as a treatment for HIV compared to existing treatments. Trizivir is a "triple nucleoside," combining three different drugs – Zidovudine, Lamivudine, and Abacavir in one pill. Physicians and HIV patients had high hopes for Trizivir; if effective, it would allow for the treatment of HIV with only one pill twice a day. It was also anticipated that Trizivir would have fewer side effects than established HIV drug regimens. Subjects in the study were randomly assigned to take one of three drug regimens: Trizivir alone, Trizivir plus Efavirenz or Combivir plus Efavirenz. Efavirenz and Combivir were established components of existing standard of care HIV regimens. After the study, researchers concluded that Trizivir alone was not as effective as the existing regimens.

The patient discussed in the above fieldnotes excerpt above was one of the 1147 subjects in ACTG 5095. The research team worrying about her case was one of the dozens of research teams conducting the ACTG 5095 study in clinics across the US.

Each research team conducted the study according to the same protocol (i.e. study recipe), which is produced by a protocol team, made up of a group of clinical, research and statistical experts. It is to this group that the principal investigator plans to write. Because the study was double-blinded, the research team did not know for certain what drug regimen the study patient was taking, but they suspected Abacavir, one of the components of Trizivir. Lipodystrophy, the redistribution of body fat, is a known side effect of some HIV medications, including Abacavir.

The research staff appeared quite concerned about the patient's deformity. They talked about the severity of her side effect and discussed the possibility of treatment. Lipodystrophy is not easily reversed; they proposed surgery and discussed how the surgery would be funded. They figured out how to correspond with the protocol team. They worked to fit the protocol to the patient by asking that the protocol team "unblind" the drug regimen and get her off Abacavir. The discussion of the research team is in stark contrast to the published research results. In the language of research, the study patient's "deformity" is an "adverse event." An adverse event is defined as any medical problem that arises during drug treatment. Research staff assign grades to adverse events. Grading entails matching adverse events to a pre-defined severity scale ranging from 1 (mild) to 4 (severe). It is a technique for translating medical problems into comparable categories in order to assess treatments and produce study findings.

How did the patient with the deformity become one of the 1147 study subjects? The researchers interpreted and implemented the study protocol. They categorized, measured, assigned grades, and documented, turning "thick" narratives into quantitative accounts. This dissertation is not another study about how numbers dehumanize, though

they surely do that. Instead, it is about the work of producing and consuming statistical research. The mandate to produce and use numbers entails more than simply reporting the numbers. Translating material, especially human material, into quantifiable categories is a social process. This study traces how staff working in dispersed locations fit the messy reality of the clinic to the uniform reporting requirements of clinical research and how, in turn, the requirements of research alter the organization of the clinic in ways that ease the production and consumption of research results. In short, this dissertation is about making and using numbers and the rules, procedures, and infrastructure that make quantification possible.

Empirically, this is a study of the conduct and consumption of statistical medical research in two American and two African HIV clinics. HIV/AIDS is a fruitful case for a study of clinical research because the rise of HIV/AIDS coincided with important changes in the organization of medical research, including the expansion of domestic and international clinical research and the rise of Evidence-Based Medicine. Medical research signaled political and public commitment to a disease, and AIDS activists demanded research. At first, clinical trials were also the only way for Americans with HIV/AIDS to access antiretroviral treatment. In the past 25 years, governments and non-government organizations have built a multi-billion dollar global research program. Thousands of HIV-related clinical trials are conducted each year, in the US and around the world.

The proliferation of statistical medical research lies at the crossroads of two broad trends in medical care: scientization and standardization. Scientization refers to the expansion in the amount of scientific activity, the authority of science, and the range of

issues to which scientific evidence and expertise can be applied (Drori & Meyer, 2006). Science encourages rule-making, turning mysteries and uncertainties into risks to be managed or opportunities for efficiency (Drori & Meyer, 2006, p. 31). Medical statistical research propels and relies on standardization, which is the process of making things uniform (Timmermans & Berg, 2003). This does not imply the production of number crunching, rule-following drones. When standards finally do interact with the real world, they are enacted, modified, ignored and/or misunderstood in ways unpredicted by standard creators.

Statistics is the ultimate standardized method. Hackings likens the methods of statistical probability to the "Victorian valet, ready to be the loyal servant of the natural, biological and social science," (Hacking, 1990, p. 2). Statistical techniques imply but do not actually require good data; you can calculate the average of any list of numbers. Whether the average has anything to say about the real world is another matter. Thus, the requirement to produce meaningful statistics requires an "infrastructure of standardization" (Porter, 1995) – rules and procedures for collecting and reporting data as well as an apparatus for monitoring and enforcement to ensure that the rules and procedures are followed. My task in this dissertation is to tie together the very local standardization that occurs when standards meet the real world to the "infrastructure of standardization" (Porter, 1995). Tying together threads of inquiry from medical sociology, social studies of science and technology and organizational studies, it develops a literature on science and standards "in action." This literature attends to the negotiated processes of doing and using science and standards. Below I first describe the rise of medical statistical research and its relationship to standardization and then turn to the

literature on medical science and explain the literature on science and standards in action in more detail.

The Rise of Statistical Medical Research

HIV/AIDS arose in the era of Evidence-Based Medicine. Evidence-Based Medicine is the most recent and most successful attempt at subjecting medical decisions to statistical measurement and control. Randomized clinical trials, like ACTG 5095, are the preferred technique of Evidence-Based Medicine. The randomized clinical trial is the gold standard in medicine. In a randomized clinical trial, different medical interventions (e.g. medications, diagnostic tools) are assigned by chance to two or more groups. The effects of the medical intervention on each group are measured and compared. Medical decisions are supposed to be based on the evidence from these trials. There is still no cure for HIV/AIDS, but the disease is managed with antiretroviral drugs. Clinical trials are conducted to measure the safety and efficacy of new, experimental drugs as well as the effects of different combinations of existing drugs.

It is not immediately apparent that Evidence-Based Medicine is something new. The medical profession has long relied on science as a source of decision-making and professional authority. However, evidence-based medicine signals a real shift in the relationship between science and medicine. The 19th and early 20th century saw the rise of "scientific medicine." Based on their knowledge and use of natural sciences, physicians won authority over healthcare from competitors such as herbalists, homeopaths, midwives and chiropractors (see e.g. Bynum, 1994; Reiser, 1978; Tuchman, 1993). Evidence-based medicine differs from scientific medicine. Scientific medicine

refers to medical practice that relies on knowledge derived in the laboratory while evidence-based medicine requires that therapeutics be proven effective through statistical epidemiological methods. This entails a profound shift in the knowledge base of clinical practice from pathophysiological to epidemiological (Timmermans & Kolker, 2005). The randomized clinical trial, a method for quantitatively comparing clinical interventions, replaced the laboratory experiment, a method for uncovering disease mechanisms, as the arbiter of clinical decision-making.

This shift in the knowledge base of medicine was made possible by changes in the scale and content of medical research in the second half of the 20th century. Medical research exploded in the US after World War II. Between 1945 and 1950, US federal funding for medical research increased from 180,000 to 46 million dollars a year (Mulrow & Lohr, 2001). Total US spending on medical research reached 95 billion dollars a year in 2003 (Moses, Dorse, Matheson, & Their, 2005).

Just as more research funds were becoming available, medical research expanded from the laboratory to the clinic. Although a small group of "therapeutic reformers" urged statistical comparison of medical interventions (Marks, 1997), prior to the 1940s, nearly all medical research was basic research on cells, organs and animals. Clinical trials proliferated in the wake of pharmaceutical advancements and expanded government regulation of pharmaceuticals. The government and public had long distrusted drug company claims, and they became increasingly distrustful of the medical profession's ability to judge drug company claims, especially in the wake of Thalidomide (Porter, 1995). Physicians prescribed Thalidomide to pregnant women as a treatment for nausea; the drug caused thousands of birth defects and infant deaths. Following the Thalidomide

disaster, the U.S. Congress passed new legislation in 1962 adding the determination of efficacy to the FDA's mandate. Drugs must now be proven statistically superior to placebo in order to receive FDA marketing approval. By 2003, just over 40% of the funding for biomedical research from the National Institute of Health (NIH) and pharmaceutical companies went toward clinical trials (Moses, et al., 2005).

Clinical research is a standardizing and a standardized enterprise. Statistical clinical research propels the creation of medical standards. Indeed, what bothers critics of evidence-based medicine is the translation of evidence into rules for care. Some physicians worry about "cookbook" medicine and the loss of autonomy and discretion (Daly, 2005). Evidence-based medicine was originally imagined as an individual-level practice, where a physician encountering a case would search the medical literature, select the best and most relevant studies, judge the evidence, and based on the evidence decides how to handle the case. This model is widely recognized as untenable on a broad scale, largely because of constraints on the time and competencies of individual clinicians. Even the Evidence-Based Working Group at MacMaster University in Toronto, who originally coined the phrase, "Evidence-based medicine" recognizes that most physicians will use evidence-based tools and will not conduct their own critical analysis of individual studies (Daly, 2005). The primary tools of evidence-based medicine are secondary reviews of medical literature and clinical treatment guidelines. The Institute of Medicine defines clinical treatment guidelines as "systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances" (Field & Lohr, 1990, p. 38). In the US, there are between 1400 and 4000 existing guidelines and about 1000 new guidelines produced

each year (Timmermans & Kolker, 2005, citing Rosser, David, & Gilbart, 2001). At their core, these evidence-based tools are aimed at the reduction of human bias in judging what constitutes good care. The expectation is that medical decision-making tools will make medical practice more rational, uniform and efficient (Berg, 1997).

In addition to standards for care, the rise of clinical research has motivated standards for the conduct of research. Medical research has become increasingly standardized and rule bound in recent decades (Heimer, Petty, & Culyba, 2005). Part of the impetus toward standardization in research arises from traditional concerns with replication and reliability. Indeed, every clinical trial is governed by a specific research protocol. A typical study protocol contains the background and the hypotheses as well as the plan for the actual conduct of the clinical trial. A protocol describes who is eligible to participate in the trial, the duration of the study, the tests, procedures, and medications to be administered, and the information to be gathered (NIH, 2008).

Changes in the ethics and organization of the medical research have propelled new kinds of standards. Fox (1989, p. 493) observes that changes in "human experimentation in magnitude, complexity and potential peril" have increased concerns with the ethics and regulation of research. The proliferation of rules is linked to increasing concerns about the ethics of research, following scandals such as Nazi research and later the public revelation of unethical American research, especially the Tuskegee syphilis trial during which researchers tracked but did not treat African American men with syphilis. International standards for ethical conduct of research include the Nuremberg Code, the World Health Association's *Declaration of Helsinki* (initial version in 1964, last revision in 2004), the Council for International Organizations

of Medical Sciences' *International Ethical Guidelines for Biomedical Research involving Human Subjects* (initial version in 1982, last revision in 2002), and the pharmaceutical industry's *Good Clinical Practice* produced by the International Conference on Harmonization. In 1979 the US National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research produced the *Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research* which is codified in US 1991 "Common Rule" regulations governing medical research. Other countries have their own guidelines as well, including Uganda's 1997 *Guidelines for the Conduct of Health Research Involving Human Subjects in Uganda* (revised in 2007), and South Africa's 2005 *Guidelines on the Ethics for Medical Research*. The standards share basic requirements of informed consent and ethical review but vary more in the extent to which they address standards of care and principles of justice.

The proliferation of rules governing medical research is also linked to the centralization of medical research and the rise of the multi-center clinical trial. At one time, individual investigators initiated and designed their own research projects. Increasingly however, researchers are more likely to do "ready-made" research projects developed by teams or networks of researchers funded by both government and private sponsors. Biomedical research has shifted from small individual studies to large collaborative studies and research institutions. One indicator of this is the US NIH grant applicant success rate which is the percentage of reviewed research project grant applications that receive funding. Within the National Institute of Allergies and Infectious Diseases (NIAID), the NIH institute that funds research in HIV/AIDS and

other infectious diseases, the success rate dropped from 43.3% in 1997 to 23% in 2007 while the amount of resource funds awarded doubled, increasing from 227 million to 504 million dollars (NIH, 2007). This suggests a shift away from individual researchers toward larger projects, a trend which is also observed among European and Chinese research funders (Check & Castellani, 2004.)

AIDS came on the American scene in the early 1980s, when multicenter clinical trials were becoming *the* strategy for studying disease. A multi-center clinical trial is a prospective study of the effect of a diagnostic, therapeutic or prophylactic intervention, conducted at dispersed sites (Friedman, Furberg, & DeMets, 1998). Multicenter trials enable researchers to recruit more research subjects and to better control for geographic and demographic variations. A consequence of multicenter trials is focus on uniformity and the production of research protocols that emphasize data commensurability as much as data accuracy. So for example, research projects may require that laboratory tests be performed at a particular lab, not because that lab produces more accurate results, but so that data collected in dispersed sites will be comparable. Multicenter clinical research is more rule-bound because of the demands of reliable, uniform data.

The rise of HIV/AIDS has motivated the production of new standards and the revision of existing standards of ethical conduct and data quality. In response to early AIDS activist demands for access to experimental drugs, the US Food and Drug Association (FDA) loosened its rules about the kind of data required for drug approval and expanded pre-market access to new drugs. Later, the *Declaration of Helsinki* was revised in 1996 as a result of the controversies surrounding placebo-controlled studies of the efficacy of AZT, an antiretroviral drug, in preventing mother-to-child transmission of

HIV (Carlson, Boyd, & Webb, 2004). These placebo-controlled trials took place in developing countries and were widely criticized because AZT was already the standard of care in wealthy countries. Critics argued that researchers should not conduct studies in other countries that would be unethical to conduct in their own countries. The *Declaration of Helsinki* added language limiting the use of placebo-controlled trials, which led the FDA to change its policies so that some studies conducted outside the US are no longer required to follow the *Declaration of Helsinki* and instead must follow the *Good Clinical Practice Guidelines* produced by the drug industry and regulatory authorities in the US, European Union and Japan (Lurie & Greco, 2005).

Theoretical Thinking about Medical Research

Sociology of Medical Science

This dissertation expands on an emerging sociology of medical science. Fox (1989) proposes this term for an area of inquiry that concerns the content of medical scientific knowledge, its underlying modes of reasoning and the processes of medical research. Much of the scholarship on medical science focuses on the profession. Knowledge is the key to Parsons' (1951) distinction between the professions and other occupations. In contrast to the emphasis on profits in business, the medical profession is distinguished by established training, a scientific knowledge base, and a commitment to the collective good. Scientific knowledge is central to more recent scholarship on the medical profession, but in this work, knowledge is a site of contest and currency in the struggle for control. According to the view of professional dominance (Friedson, 1970), control over knowledge is central to professional power. Professional knowledge is

elusive and esoteric, requiring expertise and discretion. Professionals risk losing control of their work if their knowledge base is readily accessible to laypersons (Abbott, 1988). They also risk losing control if their knowledge-based interventions are perceived as regularly ineffective. Hafferty and Light (1995) observe a decline in professional dominance as other parties, "countervailing powers" such as public and private payers, drug companies, consumers, have claimed more control over medical practice and knowledge. At the same time, the medical profession has engaged in defensive rule-making to pre-empt rule-making by other parties, such as insurers and lawyers (Heimer, Petty, & Culyba, 2005). Despite the loss of autonomy for professionals that is associated with adopting standardized procedures, standards can serve as a way to defend the organization against outsiders (Porter, 1995).

From the perspective of Parsonian functionalism, the imposition of an epistemological knowledge base is a welcome development because it strengthens ties between clinical practice and the scientific knowledge base and furthers the normative goals of universality and neutrality (Timmermans & Kolker, 2005). From this perspective, rules are good for the profession. From the perspective of professional dominance, rules are a threat to discretion, which according to Friedson (1986) is the key to professional power. Rules are potentially both a source of, and a threat to, professional jurisdiction (Abbott, 1988). On the one hand, codifying professional knowledge into rules for care is a technique for claiming jurisdiction, but codified standards potentially undermine the argument that the work requires expertise (Abbott, 1988). Likewise, Hafferty and Light (1995) argue that the existence of rules for performing care potentially

undermine professional power especially when those rules are produced outside the medical profession.

The rise of evidence-based medicine has motivated speculation from some theorists about what evidence-based medicine, especially clinical treatment guidelines, means for the medical profession. Hafferty and Light (1995) frame the rise of treatment guidelines and medical effectiveness research as part of the struggle between the countervailing powers of corporate sellers of pharmaceuticals and medical devices and medical payers who want to limit expensive medical interventions. Similarly, Porter (1995) observes that the medical profession successfully prevented the intrusion of statistics on medical decision-making until the state intervened to limit drug makers' claims of efficacy. What is unclear is whether the strengthening of the medical profession's scientific footing will increase the power of the medical profession or lead to its undoing as it is "hoisted by its own petard" (Hafferty & Light, 1995, p. 143).

The rise of treatment guidelines based on medical research, as opposed to usual and customary medical traditions, threaten the autonomy of individual physicians (Hafferty & Light, 1995). Loss of autonomy may be outweighed by improvements in care and increased medical certainty. Indeed, the formal goal of medical research is to reveal and resolve uncertainties of medical practice (Fox, 1989). Nevertheless, Fox (2000) and Hafferty (2000) are skeptical of the claims that evidence-based medicine reduces medical uncertainty and propose that evidence-based medicine potentially increases uncertainty (Fox, 2000, Hafferty, 2000). Based on interviews with medical residents, Timmermans and Berg (2003) observe that evidence-based medicine is

associated with an increase in a new kind of "research-based uncertainty" regarding how to select and evaluate medical studies (see also Timmermans & Angell, 2001).

Science and Standards in Action

The scholarship on scientific knowledge and the medical profession highlights rule-following, discretion, and uncertainty. This scholarship has focused on the level of the medical profession instead of medical work and on the application of research results rather that the conduct of medical research. The first is a really important distinction because regulatory activities that increase the autonomy and authority of physicians as a group may actually reduce the autonomy and authority of individual physicians. These ideas about the consumption of medical research are usefully turned on the conduct of medical research because the process of research is so standardized and rule-bound. The conduct of clinical research is another wedge into the authority of medical professionals, and perhaps, intervenes more directly in the work of medicine than the standards or guidelines based on research results.

Though not typically framed as such, a central issue in the scholarship on the actual work of medical research is the loss of the physician researcher's discretion to provide care outside of the research protocol. Research protocols potentially conflict with standards of care. Studies of the conduct of medical research typically highlight the conflict between care and research. Medical research first received sociological attention in Fox's (1974) study of how physicians and patients on a hospital research ward resolve points of conflict between research and therapy. More recently, Mueller (1997) observed tensions around research and care in the interactions between the physicians and nurses in

an HIV/AIDS research unit. In the research unit, the physicians were especially concerned with following the research protocol while the nurses worried more about providing care.

To understand the actual processes of standardization and not just the potential effects of standardization requires that we examine medical work. Scholarship on the production of scientific facts provides a useful frame for thinking about the very local and very uniform parts of medical statistical research. Numerous laboratory studies have argued that scientific knowledge is locally situated, negotiated, and grounded in practice (see e.g. Collins & Pinch 1994; Knorr-Cetina 1983; Latour & Woolgar 1979; Latour, 1987). Most laboratory studies are micro-level studies that take the perspective of constructionism and attend to negotiation interactions. Everything is an outcome of negotiation in the laboratory – procedures, methods, outcomes, etc. In this way, laboratory studies has much in common with symbolic interactionism, which examines how meaning is negotiated and produced through interaction.

Just as social studies of science have shifted our attention from the ideal of the scientific method to "science in action" in the laboratory, a literature on "standards in action" shifts our attention to what people actually do with standards. From this scholarship arises a model of rule following that is more flexible and helps account for what Timmermans and Berg (1997) call "local universalism." The effects of a rule are underdetermined by the content of the rule (Feldmand & Pentland, 2003; March, 1988; Reynaud, 2005). In place of terms such as rule adoption or implementation, this scholarship examines "standard tinkering" (Timmermans & Berg, 2003), rule completion

(Reynaud, 2005), performance (Feldman & Pentland, 2003) and improvisation (Weick, 1998) which evoke a more active notion of rule work.

Creating and sharing scientific information is work that requires negotiation with others in organized settings (Fujimura, 1992). It is not only a local accomplishment. The production and use of research findings requires "actor networks," (Latour 1983; Latour, 1999; Latour, 1997) "social worlds," (Clarke 1990) and "infrastructural work" (Bowker, 1994). Latour (1983) writes "Scientific facts are like trains, they do not work off their rails. You can extend the rails and connect them but you cannot drive a locomotive through a field" (p. 155). Likewise, standards cannot be driven through open fields; they do not work off their rails. Producing and using new technology and knowledge requires the building of networks and infrastructure.[2] Similarly, in his work on decision-making tools in medicine, Berg (1997) observes that medical workers must "materialize" a tool's demand so that using the tool becomes less noticed and avoidable.

The ethnographies of laboratories have pointed to contingent and variable enactment of scientific techniques; the social worlds approach provides some direction for understanding how scientific work can be both local and variable and global and universal. The "social worlds" approach is the most organizational and sociological of this science studies scholarship. It arose out of the Chicago School symbolic interactionism which studied "social wholes" such as medicine, science and the arts. Social worlds are made up of groups who share commitments and resources and engage in activities in order to reach a goal (Becker, 1982; Clarke 1990; Strauss, 1978). The "social worlds of science" approach is loosely connected through an interest in standards and classification that span different social worlds of scientific activities. It focuses on

communication and coordination in the production of knowledge in multi-organizational and multi-disciplinary contexts. In this vein, Star and Griesemer (1989) discuss the creation of standardized methods and "boundary objects" to account for the coexistence of heterogeneity and cooperation. Boundary objects may be either concrete or abstract; they "have different meanings in different social worlds but their structure is common enough" to allow for communication and cooperation (Star & Griesemer, 1989, p. 393). Fujimura (1992) proposes the term "standardized packages" to refer to a more stable and less abstract set of concepts and tools.

Medical statistical research abounds with concrete and abstract boundary objects – data reporting forms, adverse events, reports, etc. that allow for the cotemporaneous conduct of research and medical treatment and the coordination between protocol teams and medical researchers. The coordination of clinic-level research differs in important ways from the biological and physical sciences. The basic sciences are collaborative, bringing together scientists from different labs and even different disciplines. They rely on standard techniques (e.g. for extracting DNA), standard reagents, standard equipment, etc., but they rarely require that every lab in a research collaborative collect the same kind of information in the same way, as collaborative (i.e. multicenter) clinical research does. Multicenter clinical research demands procedural uniformity across sites in order to standardize the behavior of scientists and non-scientists and very reactive human subjects.

Overview of the Dissertation

Physician and anthropologist, Paul Farmer, (1999) urges social scientists of medicine to "take account of basic biomedical insights" (p. 256). Likewise, science studies scholars typically immerse themselves in the content of the science. Thus, chapter two, of this dissertation begins with an overview of HIV disease and the technology for treatment and clinical research. Technical, biomedical information is not only context for understanding subsequent chapters. The characteristics and complications of HIV/AIDS disease and its technology have social consequences. HIV/AIDS research is made up of a network of natural, technical, and social entities. In addition to biomedical insights, this chapter describes the general organizational field of international HIV/AIDS research as well as the specific clinic fieldsites observed for this dissertation. The methods of fieldwork and data analysis are also discussed.

The next three chapters are empirically grounded chapters on the epistemology of statistical medical evidence and the organizational processes of standardization associated with statistical medical research. The production and consumption of scientific evidence has become an increasingly important to medical work. What counts as good evidence in medicine has changed over the past few decades, and physicians are expected to practice evidence-based medicine. However, in addition to statistical clinical research, physicians draw on the laboratory sciences, expert opinion as well as their own experience. Chapter three compares and contrasts the status of statistical clinical research in the clinic and in formalized practice guidelines. This chapter bridges medical sociologists' traditional concern with uncertainty with science studies' attention to producing scientific facts. There is more uncertainty over what counts as good evidence

at the clinic level than there is in formalized guidelines. I observe a "research-based uncertainty" (Timmermans & Angell, 2001; Timmermans & Berg, 2003) arising from ambiguity and flexibility in what counts as good statistical evidence as well as how to integrate this with clinical experience. Among the research team, uncertainty largely arises from figuring out the rules for doing research. Here there is "rule-based uncertainty" where what counts as evidence is largely a matter of determining what should get measured and documented, and how.

Chapter four looks more closely at research and rule-following. It examines the enactment of research protocols and sheds light on the general process of standardization in medicine. Medical studies are governed by research protocols. The goal of a protocol is to create comparable results from the bodies of hundreds or even thousands of different patients and from the labor of numerous researchers in dispersed clinics. Protocols specify what to do and how to do it, when to do those tasks, who should do them and whom they should do it to. Scholarship on standardization has observed that making standards fit local conditions typically involves more work than standard producers assume. Chapter four examines what researchers actually do with protocols. In order to make protocols work, researchers limit input uncertainty, tinker to fit the clinic to the protocol and the protocol to the clinic, develop protocol intelligence and negotiate with competing rule systems.

The fifth chapter turns to the organizational processes of conducting research and demonstrates that organizing the conduct of research potentially influences clinical work as much as the subsequent implementation of research results. In the course of fitting clinical research to the organization and the organization to clinical research, bureaucratic

relations are reorganized, material environments are altered, priorities are changed with increasing value placed on reliability of measurement and records, and variability is reduced as more and more aspects of clinic life are covered by standard operating procedures. An additional outcome of conducting research is institutional isomorphism; organizations become increasingly alike as they evolve to meet the requirements of a centralized set of regulators and funders. This is part of the development of an infrastructure of standardization that makes doing and consuming research possible.

Finally, the last chapter sums up the empirical chapters and draws conclusions regarding standardization, commensuration and trust. It ties together threads from the previous chapters to speculate about the implications of the rise of medical statistical evidence on medical work and the medical profession.

Chapter 2: Studying Studies

This dissertation is a study of studies. It examines the conduct, interpretation and use of medical research in HIV clinics. HIV/AIDS research is made up of a network of relations among natural, technical, and social entities. This chapter begins by describing the virus itself followed by an overview of the technologies of detection and treatment, and finally, a discussion of the organizations that house and sponsor HIV care and treatment. HIV/AIDS is a complicated disease residing in a complex organizational field. The second part of the chapter discusses how I went about studying HIV/AIDS studies. Descriptions of each of the four fieldwork sites follow, including details about where these clinics lie in the network of HIV care and research.

Network of HIV/AIDS Research

Virus and Technologies

According to estimates from the United Nations Program on HIV/AIDS (UNAIDS), in 2007 there were approximately 33.2 million people living with HIV worldwide; approximately, 22.5 million of these people were in Sub-Saharan Africa (UNAIDS, 2007). The human immunodeficiency virus (HIV) resides in bodily fluids. The main ways of contracting HIV is through unprotected vaginal or anal sex, mother-to-child transmission (which may occur in utero, during delivery or through breast milk), the use of infected blood or blood products, or the use of contaminated needles. HIV is a retrovirus which attacks CD4 T cells. CD4 T cells, also known as T-cells or T-helper cells are “white blood cells that orchestrate the immune response, signaling other cells in

the immune system to perform their special functions" (Klimas, Koneru, O'Brien, & Fletcher, 2008). A person is infected with HIV once the virus attaches to one of these cells. The virus enters the cell and replicates rapidly. In the first stage of HIV disease, sometimes called "acute HIV infection" or "primary infection" the amount of virus in the blood is extremely high and the person is very infectious. After this stage, the amount of virus declines, and there is period of "incubation" or "latency" during which the virus and the cells they attack reproduce and kill each other at a fairly equal rate (Barnett & Whiteside, 2006). During this time, people with HIV usually do not feel or look sick.

Over time, the CD4 cells lose out to the virus and the number of CD4 cells slowly declines. A healthy person has roughly 2000 CD4 cells in one milliliter (ml) of blood (Barnett & Whiteside, 2006). HIV-infected people are considered to have AIDS when their CD4 count drops below 200 per ml or when they are diagnosed with AIDS-defining illnesses. AIDS defining illnesses include opportunistic infections and rare cancers, such as Kaposi's sarcoma, which causes the purple spots which were a hallmark of AIDS in the 1980s. As CD4 cells decline, people typically experience opportunistic infections, such as pneumocystis carinii pneumonia (PCP), mycobacterium avium complex (MAC) disease, cytomegalovirus (CMV), toxoplasmosis, and candidiasis. They are called "opportunistic" infections because the organisms that cause these infections rarely affect people with healthy immune systems but they have the "opportunity" to flourish in people with weakened immune systems (Barnett & Whiteside, 2006).

The time between infection and progression to AIDS is approximately 10 years in wealthy countries. Without treatment, people live approximately 1 to 2 years from the onset of AIDS (Barnett & Whiteside, 2006). People treated with highly active

antiretroviral therapy (HAART) live much longer. When adult patients start HAART at CD4 counts less than 350, the average projected life expectancy is 24.2 years from the time of treatment initiation. (Schackman, et al., 2006). HAART, sometimes also referred to as antiretroviral therapy, combination therapy or even a drug "cocktail," entails the use of at least three antiretrovirals from at least two classes of HIV antiretroviral drugs. At the time of our fieldwork, there were four classes of antiretrovirals: nucleoside transcriptase inhibitors (NRTI), non-nucleoside transcriptase inhibitors (NNRTI), protease inhibitors (PI) and fuseon inhibitors (FI) (DHHS, 2005). By the time of this writing, two additional classes have become available: CCR5 antagonists and integrase inhibitors. Each drug class uses a different mechanism to interfere with different steps in the HIV replication process. According to the antiretroviral guidelines published by the US Public Service's Department of Health and Human Services (DHHS), there are 20 antiretroviral drugs approved by the US Food and Drug Administration for the treatment of HIV (DHHS, 2008).

HAART was a key turning point in HIV/AIDS care. Its diffusion was rapid in the U.S., with 71% of patients with HIV on HAART within two years of its availability (Cunningham, et al., 2005). As a result, AIDS related morbidity and mortality decreased dramatically (Palella, et al., 1998). The AIDS-related death rate in the US dropped 47% in 1997 alone (Henkel, 1999). As deaths and hospitalizations decreased, the locus of HIV/AIDS care shifted to outpatient clinics. By December 1996, 77.7% of HIV/AIDS patients received care in specialty HIV clinics (Wilson, et al., 2005). Although about two-thirds of the people who are infected with HIV live in sub-Saharan Africa, the diffusion of HAART has been far slower in Africa than in the US. In 2003, 2% of those

who needed HAART in Sub-Saharan Africa were receiving it; by 2006, HAART treatment reached 28% of those who needed it (WHO, UNAIDS & UNICEF, 2007). When antiretrovirals are available, triple drug therapy is the norm in both rich and poor countries. This is in contrast to malaria where people in poor countries regularly received monotherapy when combination therapy is the standard in wealthier countries (WHO, 2006). However, there is some variation in the specific types of antiretrovirals taken by people in rich and poor countries. For the most part, people in the US have access to the whole range of antiretrovirals while people in poor countries have access to the least expensive antiretrovirals. The expense of antiretrovirals is a complicated and contentious issue. Typically, poor countries are using antiretrovirals that are now off patent or they are using antiretrovirals which drug companies agree to provide at a discount. Older and cheaper drugs are not necessarily worse. However, some of these drugs are cheap because almost no one in the rich countries uses them anymore because of annoying and sometimes serious side effects.

When to start HAART can be a difficult decision for healthcare providers and policymakers. The US DHHS antiretroviral treatment guidelines recommend that persons with HIV *should* start HAART if their CD4 count drops below 200 or they have an AIDS-defining illness but *may* start HAART if their CD4 count drops below 350 (DHHS, 2008). The 2003 World Health Organization guidelines for resource poor settings recommend starting HAART when CD4 drops below 200 or whenever a patient has a “stage 4” HIV disease (WHO, 2003). The WHO recommendations contain a clinical staging scheme which ranges from one to four, with four being the most severe. The most recent revision of the WHO guidelines moves closer toward the American

DHHS guidelines (WHO, 2006). Of the four clinics discussed in this dissertation, three followed the 200 CD4 count standard; the physicians at the private American clinic had more discretion.

When effective, HAART reduces viral loads to extremely low levels, often within weeks of starting treatment. This inhibits further damage to the immune system. Over time, CD4 counts are likely to rise. However, HAART is not without drawbacks. Antiretroviral drugs have serious side effects. Common side effects include rash, diarrhea, high cholesterol, elevated levels of lactic acid, lipodystrophy (characterized by the accumulation of fatty deposits or the loss of fat on the body) and neuropathy (characterized by burning, tingling, sometimes disabling pain in the feet and legs) (DHHS, 2008). HAART has to be taken daily for life. The number of pills in a daily regimen ranges from three to more than 10 a day. Some drugs must be taken with food and other without food. Some drugs have to be taken once a day, others two or three times a day. Near perfect adherence is required for antiretrovirals to be effective. Even with near-perfect adherence, antiretrovirals lose their effectiveness over time because of drug resistance.

Drug resistance is a serious and complicated issue. Overtime, HIV mutates and becomes resistant to antiretroviral drugs that were once effective. In other words, the drugs no longer block the growth of HIV. The main reason for resistance is poor adherence. HIV treatment requires a 95% adherence rate to be effective. This level of adherence is difficult since these drugs should be taken one or more times a day for life. According to Klimas, et al. (2008, p. 528), one “missed weekend of medications can result in drug resistance.” Antiretroviral resistance is a concern for individual patients

because it limits their treatment options. Drug resistance is also a public health issue. HIV/AIDS physicians, researchers and policymakers worry about the development of widespread drug resistant HIV at the population level. An individual can spread drug resistant HIV to others; when these newly infected people seek treatment, they will have limited medication options. Widespread drug resistance is especially a public health concern for poor countries where only a few affordable drug regimens are available.

Physicians use laboratory tests to monitor the progression of HIV disease and antiretroviral side effects. The most common tests include viral load, CD4 count, complete blood count (CBC) and blood chemistry tests. CD4 and viral load tests are used as indicators of HIV disease progression as well as how well a drug or set of drugs works. Complete blood count and other blood chemistries are important for detecting side effects of HIV treatment. If a drug effectively kills HIV but also kills the person with HIV or makes them so ill that they cannot take the drug, it is not very useful. CD4, viral load, complete blood count and blood chemistries are important technologies of research. HIV clinical research is oriented toward determining which drug combinations are the most effective at killing the virus and which have fewer side effects for various categories of patients, such as treatment naïve patients, patients with particular diseases in addition to HIV/AIDS, pregnant women, those who have failed previous regimens. Laboratory tests results are quantitative indicators of efficacy and side effects. Furthermore, these tests have additional meaning for patients. They translate into labels patients pay attention to; patients want to achieve “undetectable” viral load and avoid labels like “treatment failure” and especially “AIDS.”

CD4 count refers to the absolute number of CD 4 cells. It is calculated based on the percent of CD4 cells in a blood sample. The CD4 count is an indicator of the health of the immune system; the more CD4 cells, the healthier the immune system. HIV attacks CD4 cells and uses them to produce more HIV. As mentioned above, a person with HIV is considered to have AIDS once the CD4 count drops below 200. The viral load test measures the amount of HIV in blood. The metric is copies of the virus per ml of blood. An effective HAART regimen reduces viral load dramatically, often to levels undetectable by most laboratory equipment. Depending upon the kind of test used, "undetectable" is considered less than 50 to 400 copies of HIV per ml blood. When a HAART regimen is not effective, this is referred to as "treatment failure." Treatment failure may occur soon after starting HAART or after an extended time of viral suppression. Treatment failures may be indicated by viral, immunological or clinical information. Viral failure is a lack of viral response and occurs when an antiretroviral regimen fails to reduce the viral load to undetectable or near undetectable levels. Immunological failure is a lack of immunological response and is indicated by a decrease in CD4 count. Clinical failure occurs when a person on HAART experiences an AIDS-defining illness.

To researchers and caregivers, treatment failure indicates poor adherence and/or drug resistance. Researchers and caregivers use drug resistance tests to determine the specific kind of viral mutations a person has and which kinds of drugs are likely to be effective. There are two kinds of genetic tests – genotypic and phenotypic. Genotypic tests identify specific viral mutations. Phenotypic tests determine which antiretrovirals the virus is resistant to. Another relatively complex set of tests are pharmacokinetic tests

which measure drug levels in the blood. These tests can be used as an indicator of whether a patient is adherent or, in adherent patients, an indicator of how the body is metabolizing the drug. Because individual metabolisms differ, some people may need a bigger or smaller dose to achieve viral suppression.

There is variability in the use of laboratory testing between rich and poor clinics. In the US, CD4 and viral load testing is routine. The US DHHS guidelines recommend that these tests be performed every three to six months. At the time of the fieldwork, drug resistance tests were less common, usually only occurring when a patient had failed multiple antiretroviral regimens (DHHS, 2005). These are expensive tests, costing a few hundred US dollars for the genotypic test and around 1000 US dollars for the phenotypic test in 2005. The tests were provided more often in research than in care. Genotypic and phenotypic resistance tests are almost completely unavailable outside of research in African countries. Even the far less complicated and expensive CD4 and viral load tests were more limited in poor countries. CD4 tests are a routine part of care in many HIV clinics or what the WHO refers to as "regional referral centers." Viral load testing is less routine. According to the WHO (2003) "because of the cost and technical issues associated with viral load testing, it is not currently recommended as part of the treatment guidelines" (p. 25) In poor settings, when to begin treatment and when to change treatment because of treatment failure is more likely to be based on immunological (CD4 count) or clinical indicators (observable illness, weight loss).

Organization of HIV/AIDS Clinical Research

HIV/AIDS arose as a public health issue following the enormous growth of clinical research in the US after World War II. Between 1986 and 1987, clinical trials took "center stage" in the fight against AIDS (Epstein, 1996). During those years, the NIH was setting up a nation-wide network of research centers to conduct clinical trials on AIDS drugs, and researchers announced that AZT might be effective against AIDS based on initial clinical research (Epstein, 1996). After a controversial placebo controlled trial, AZT quickly became the standard of care for AIDS with which other potential treatments must be compared. HIV/AIDS clinical research subsequently exploded, propelled by activist demands, public opinion, federal funding and the prospect of corporate profit. Between 1987 and 1999, over 68,000 volunteers participated in one of the many studies organized by the ACTG (NIAID, 1999). A recent search of the registry of publicly and privately funded clinical trials maintained by the US NIH yielded 3027 studies of HIV infections.[3] Another indicator of the scale and pace of HIV clinical research is the biannual revision of the US Department of Health and Human Services HIV/AIDS treatment guidelines; the guidelines for most other diseases are revised every few years.

HIV research is an international enterprise. Indeed, 1004 of the trials listed on the NIH's clinical trials registry are located outside the US. The US government first supported international clinical research in 1984 with a research project in Haiti and then established Project SIDA (French for "AIDS Project") in 1985 in what was then Zaire (Office of AIDS Research, 1998). In 1994, researchers published findings from a US trial (ACTG 076) that treating mother and baby with AZT reduced the rate of mother-to-child transmission by two-thirds (from 25% to 8%) (Connor, et al., 1994). AZT became

the standard of care in the US but was deemed too expensive for developing countries by the WHO. Soon after, 16 clinical trials of AZT were conducted in poor countries; the US NIH and Centers for Disease Control and Prevention (CDC) funded nine of the trials (Annas & Grodin, 1998). As mentioned in the previous chapter, this research became the subject of debate and public attention because of the use of placebos.

According to the Africa Clinical Trials Portal, an online resource for information about HIV, TB and malaria clinical trials, the "number of clinical trials being conducted in Africa is increasing dramatically" as are the number of international research collaborations (Africa Clinical Trials Portal, 2004). The Africa Clinical Trial Portal lists 101 HIV trials in Africa. The NIH registry and the Africa Clinical Trial Portal are voluntary listings. A more reliable source, for Uganda at least, is the list of trials maintained by the Ugandan National Council of Science and Technology. According to the Ugandan National Council of Science and Technology, over 120,000 Ugandans participated in 125 HIV/AIDS biomedical between 1995 and 2004.[4] Another 60,000 participated in 120 social and behavioral research projects. The Ugandan rate of participation is likely higher than the corresponding rate for other countries. The Ugandan government's openness about HIV made it especially easy for researchers to conduct research there.

Every disease field is characterized by its own alphabet soup of organizations. This is especially true for HIV/AIDS because it is a global pandemic and the target of global efforts at surveillance, prevention, treatment and research. The infrastructure of international HIV research is made up of funders, regulators and standard setters including UNAIDS, WHO, the President's Emergency Plan for AIDS Relief (PEPFAR),

US NIAID, the US Division of AIDS (DAIDS), US CDC and national health ministries. The NIH funds the largest HIV/AIDS research program in the world; it has a budget of almost three billion dollars and funds research in 90 countries (Office of AIDS Research, 2008). The NIH describes its HIV research efforts as a "complex, multi-disciplinary, multi-institute global research program" (Office of AIDS Research, 2008, p. 8). The NIH spends approximately 370 million dollars on international HIV/AIDS research annually, with most of this going to US-based researchers who collaborate with researchers in the host countries (Office of AIDS Research, 2008). In 1986, the US Congress awarded funds for the DAIDS, which is housed in the NIAID, to develop and implement the national HIV/AIDS research agenda. The NIAID now sponsors six HIV/AIDS related research networks with sites in 17 countries (DAIDS, 2007). One of these networks, ACTG, is the largest HIV clinical trial organization in the world.

In addition to the NIH, other large funders of research include the United Kingdom's Medical Research Council, which has research units in Gambia and Uganda and collaborates on research projects with other African countries including South Africa (Medical Research Council, 2008). Private funders include organizations such as the Elizabeth Glaser Pediatric AIDS Foundation. The well-known Bill and Melinda Gates Foundation is another large funder of international HIV treatment and research, spending almost two billion dollars on HIV, TB and reproductive health since its inception in 1994 (Gates Foundation, 2008). Pharmaceutical companies also fund research, but in the case of international research, they are more likely to participate by providing free antiretrovirals to study organizers. Each of these entities has its own set of standards which overlap and potentially conflict with the conduct of research.

Methods

This dissertation draws on field observations of clinical researchers "in action" and the infrastructure in which they reside. In *Science in Action*, the obligatory text on the techniques of science studies, Latour (1987) encourages researchers to enter the "black box" of scientific work and "follow the scientists." Years earlier, Merton (1952) observed that empirical study of how scientists "actually do think, feel and act" would benefit our understanding of science (quoted in Fox, 1989). In the field of research on medical science, however, Renee Fox observed a dearth of fieldwork on medical scientists as late as 1989. Exceptions include Fox's own research on medical scientists in the US, Belgium and Zaire. More recently, Mueller (1997) studied the interaction between nurses and doctors in an HIV research unit. Additional scholarship instructs us to pay attention to the organizational ties and the social worlds of scientific work (Clarke, 1990; Fujimura, 1992; Lowy, 1996).

Methods

This dissertation draws on fieldwork collected as part of Carol Heimer's "Clinic-Level Law: The 'Legalization' of Medicine in AIDS Treatment and Research."[5] The fieldwork consisted of three broad activities: interviewing staff, observing meetings, and shadowing staff (Table 1). We conducted formal and informal interviews with staff. Formal interviews were pre-arranged, recorded and transcribed. We prepared a content specific interview guide for each interview. Informal interviews were conducted on the fly as opportunities arose. While interviews are useful for understanding how people

make sense of their work, understanding work, that is what people do and how they do it, requires observation, "because most work practices are so contextualized that people cannot articulate how they do what they do, unless they are in the process of doing it" (Barley & Kunda, 2001).

Latour (1987) recommends that social studies of science "follow the scientist," and we literally followed people around – with their permission of course. We shadowed staff as they went about their work which included study visits, clinical exams, paperwork, phone calls, monitor visits, and meetings. We watched them complete paperwork and requested copies of forms and policies. Along the way, we asked clarification questions when there was time. We attended meetings of all kinds ranging from weekly research team meetings, training sessions, clinic care meetings, meetings about standard operating procedures, research grant application meetings, journal club and even prayer meetings. There was overlap in the kinds of meetings attended in each site. All of the sites held routine meetings to discuss research projects. Both of the American clinics held general weekly research team meetings in which they discussed all of their studies as a whole while the Ugandan and the South African clinics held study-specific meetings. The American private clinic was associated with an infectious disease department journal club that met weekly to discuss journal articles; none of the other clinics had such a formal journal club, though they meet more or less regularly to discuss cases and recent research results.

Observing outpatient clinical research poses some challenges because of how work is distributed. There are no centralized places where work goes on. Research work went on in exam rooms when researchers conducted study visits, in conference rooms

when research teams met to discuss on-going studies, and in individual offices and cubicles located either down the hall or down the street from the clinic. This differs from the laboratory model of research, where research takes place within a single, shared space. It also contrasts with the hospital model of units or wards anchored by nursing stations. In the outpatient clinic, there is no central location to simply hang out and watch research work. This raises logistical and substantive issues. Information about the overall organization of each site was compiled bit by bit, observing the reception area one day, the pharmacy on another, study visits on another, etc. Furthermore, negotiating access was an ongoing process as we learned about potentially interesting meetings and people to observe. With the exception of standing meetings, nearly every fieldwork encounter had to be pre-arranged. Though labor intensive, the benefit of this strategy was the creation of ties with many different staff members. Just as weak ties increase the likelihood of finding out about job (Granovetter, 1973), they also expand fieldwork opportunities. At the same time, we had one or two strong ties in each fieldsite, to the kinds of people who were engaged enough with the substance of the research projects to alert us to potentially interesting events.

Just as the medical research work is dispersed, group fieldwork entails distributed work and knowledge. For a summary of who worked where, and the kinds of fieldwork they did, see Table 1. Formal interviews were recorded and transcribed. In the case of observations and informal interviews, we used the general strategy of jotting down more or less coherent notes by hand in the field then typing up more complete notes later. Writing fieldnotes with the knowledge that they will be read by others improves their quality. We shared fieldnotes electronically and stored them on a secure central

computer server. When more than one fieldworker was present in the field, we shared notes to check that we had gotten things right. Whenever we were all in the same country, we held weekly team meetings by phone or in person. As part of my contribution to this project, I did much of the fieldwork in the American private clinic and visited the American public clinic and both of the African clinics. While I have access to the well-written fieldnotes and documents, produced and collected by the other field researchers, these visits were invaluable providing me with faces and settings to connect to the texts.

Fieldwork Analysis

My analysis attends to what people do when they conduct research and interpret research findings. I draw on data from study visits, formal and informal research staff meetings, physician journal club meetings, interviews and informal interviews with clinic and research staff. The analysis focuses on research work and attends to administrative and clinical care work only to the extent that it impinges on research or when it provides useful theoretical comparisons. This dissertation is not a comparative study of HIV research in private and public clinics or in American and African clinics. These are very different kinds of clinics and not necessarily representative of their country. Rather, it observes the same kinds of activities in different kinds of places in order to get a sense of the general activity. For example, all of the clinics worried about which patients should be in their studies and what kinds of studies they should do. Instead of broad statements, such as Ugandan medical research is like this while American research like that, this dissertation makes generalizations about the ways of doing research in resource-poor

settings or in the context of more or less strict government standards regarding treatment regimens.

Miles and Huberman (1994) observe that "coding is hard, obsessive work…[and] not nearly as much fun as getting more good stuff from the field" (p. 65). This project produced an enormous amount of fieldnotes and transcripts. Data analysis consisted of writing memos and coding, writing and then coding some more. I coded all of the fieldnotes from the American and African clinics. I used a computer program (Atlis.ti) to produce an electronic version of highlighted chunks of text, margin labels and scribbled notes. The computer program is not magic; coding remains "hard, obsessive work." The program did, however, make it easier to revise coding schemes and decreased the likelihood of lost or illegible coding notes.

My coding scheme began with a general distinction between doing research and interpreting research results. My codes focus on situations in which clinic staff engaged in a lot of worrying and figuring out such as preparing for a new study, handling adverse events, learning how to do a new procedure, judging the quality and utility of published research results, figuring out who is supposed to do some research-related task, working out how to get some piece of data required by the study protocol or how specimens should be transported to and from the lab, etc. I pay careful attention to the interactions among the researchers and between the researchers and different kinds of staff. From attention to the interactions that occur as people figure out and use new technologies, like research protocols and research results, we can see the processes through which organizational change is made.

Description of Fieldsites

This dissertation draws on fieldwork in four HIV clinics in the US, Uganda and South Africa. Uganda and South Africa were selected for their uniqueness, and the US, for its central role in developing standards of HIV treatment and research. In comparison to Uganda and South Africa, the US has a far lower rate of HIV infection, 0.6% (CDC, 2007). As discussed in chapter 1, much of HIV research either resides in or emanates from the US.

Uganda is touted as a HIV/AIDS success story. In Kampala, the major urban area of Uganda, the percent of women in prenatal clinics infected with HIV increased from 11% in 1985 to 31% in 1990 and then declined to 8.2% in 2002 (UNAIDS, 2004). The HIV prevalence rate in the whole population was 4.1% in 2003. This success is commonly attributed to the Ugandan government's early recognition of and commitment to addressing HIV/AIDS and particularly the ABC (Abstinence, Be faithful, and Condoms) approach. Although some accounts question the success of the Ugandan ABC program in reducing the rates of HIV/AIDS (Parkhurst, 2002; Wawer, Gray, Serwadda, Namukwaya, Makumbi, Li, et al., 2005), there is no doubt that Uganda has reduced it HIV/AIDS prevalence rate and has become an important center of HIV/AIDS research and treatment in Africa. Uganda was the location of early mother-to-child HIV prevention studies and the first African country to pilot an antiretroviral treatment program (Weidle, et al., 2002).

The growth of the South African HIV epidemic began later than the Uganda epidemic and has become one of the most severe HIV epidemics in the world. According to an estimate by UNAIDS, 21.5% of the population of South Africa were infected with

HIV in 2003. Much of this increase took place during the 1990s when the government was distracted by the transition from apartheid rule. In contrast to Uganda, the South African government has been reluctant to accept the explanations and treatments of HIV/AIDS produced in Western countries. AIDS denialism is not unique to South Africa. An AIDS denial movement exists in the US, and some of the sources for denialism are American, especially Peter Duesberg, a former NIH scientist who argues that AIDS is caused by non-contagious physical and chemical factors and not HIV (see e.g. Duesberg, 1995). What distinguishes South African AIDS denialism is its embrace by government leaders, who have been reluctant to fund antiretrovirals to prevent mother to child transmission and treat people infected with HIV. AIDS activists in South Africa and around the world have criticized the government, even calling for the resignation of the minister of health because of her controversial stance on nutrition as a treatment for HIV/AIDS. South Africa appears to have recently turned a corner regarding its AIDS policy; in 2008 a new health minister announced that AIDS is caused by HIV and should be treated with antiretrovirals. Despite mixed messages regarding AIDS, the South African government developed a plan to provide HIV treatment, including HAART, through government-accredited health facilities in 2003 (See South Africa Department of Health, 2003). At the time of our fieldwork, healthcare workers were still figuring out how to work with the government program.

HIV/AIDS outpatient clinics are the place to observe HIV/AIDS research and care in the US, Uganda and South Africa. This was not always the case. At one time, most HIV/AIDS care took place in hospitals, and many argued that HIV/AIDS should be handled as a primary, non-specialty disease. The rise of HIV/AIDS clinics in the US was

propelled by funding from the Robert Wood Johnson Foundation and later from the US government through the Ryan White Comprehensive AIDS Resources Emergency (CARE) Act. Furthermore, the availability of HAART shifted care from hospital inpatient units, homes, and hospices to outpatient clinics. HIV treatment, particularly HAART, also occurs in specialized HIV outpatient clinics which the WHO refers to as "regional referral centres" in African countries. The initial development of regional centers of HIV care is likely an outcome of ad hoc concentrations of expertise as well as research funding from western nations and NGOs. Regional centers are integral to the current move to "scale up" the provision of antiretroviral treatment (WHO, 2003). International funding programs distribute antiretrovirals to regional HIV centers either directly, or more often, through national health ministries.

US private clinic. The US private clinic is housed within a large, private, urban, university hospital where clinical trials and treatment go on side by side in the clinic. At the time of fieldwork, the clinic had approximately 2000 patients. The private clinic required that all patients have Medicare, Medicaid or private insurance. Patients without insurance were usually referred to a nearby public clinic. There were some exceptions for patients who had the funds to "self-pay" in lieu of insurance. Clinicians in the American private clinic had access to the full range of HIV treatments and diagnostics. The physicians were generally expected to follow one of the major sets of antiretroviral guidelines produced by the DHHS or the International AIDS Society (IAS-USA). But this was not enforced. There was no formal training in the guidelines; new physicians were told to access the guidelines on the web. According to the clinical director, "they're

experts" and "if somebody does something that's not—doesn't fit the guideline, it's inevitably for a good reason" (Interview US1 050615 JP). Unlike the three other clinics in this dissertation, the American private clinic did not distribute antiretrovirals directly to patients and was not constrained by a formulary. The exception to this is research; when research projects provided drugs, study patients picked up free antiretrovirals from the clinic.

In the American private clinic, care and research overlapped. They shared staff, equipment and exam rooms. For the most part, study visits were conducted by study nurses whose work was overseen by a physician principal investigator (PI) and supported by a study coordinator and regulatory specialist. The study nurses and physician PI did clinical work in addition to research. The clinic participated in 40 different studies and had approximately 800 study participants; almost all of the study patients were also clinic patients. Not all of the studies involved clinical interventions. Most of the research participants were enrolled in a large, observational CDC study. The clinic conducted industry-funded trials and was a member of the ACTG, a network of research sites funded by the DAIDS in the US NIH. Just over 100 patients participated in 16 ACTG studies. Approximately 50 patients were enrolled in 15 studies funded by drug companies. With a few exceptions, all of the studies were multi-site which means that the American private clinic was one of many clinics called upon to produce comparable data. In addition to the CDC, ACTG and industry studies, the clinic conducted other smaller studies initiated by individual physicians.

US public clinic. The US public clinic is a free-standing, urban, outpatient clinic affiliated with a large public hospital and a private university. The clinic received much of its funding through the Ryan White program. Through the 1990 Ryan White Comprehensive AIDS Resources Emergency (CARE) Act, the US Congress authorized federal funds for multidisciplinary HIV clinics. These multidisciplinary or "one-stop shopping" clinics, provide primary HIV treatment, dental care as well as psychological and social services, including counseling on safe sex and drug addiction. At the time of fieldwork, the clinic had a staff of almost 200 (including social services) and approximately 4000 patients. Almost half of the clinic patients were uninsured. The rest were covered by Medicare or Medicaid. The few patients that were privately insured sought care in the public clinic because their insurance did not cover the cost of antiretrovirals. Only people who lived in the surrounding counties were eligible for care at the public clinic.

The American public clinic had more constraints on antiretrovirals than did the private clinic. The public clinic established two preferred initial treatment regimens based on the US Department of Health and Human Services guidelines. According to the clinic director, they must say that they are following the DHHS guidelines on their Ryan White application. The notion that they would be "caught" is mostly theoretical since there is not much auditing about the content of regimes (US2 040526 RC). The public clinic was part of a local public health system that was experiencing severe budget shortfalls. Since antiretrovirals cost a minimum of 1200-1300 US dollars per month, controlling these costs was an early target for administrators. The clinic instituted a "no free drug" policy. Whereas in the past, patients were started on antiretrovirals without a

secure payer source, the clinic administration mandated that a payer source must be secured before any antiretrovirals were administered.

Unlike the US private clinic which used the same space for clinic and research, the public clinic separated research and care. Research took place in an attached research suite with it own offices, records storage, and examining rooms. Research visits were conducted by clinic physicians with the assistance of a study coordinator. At the time of fieldwork, the American public clinic enrolled between 500 and 600 study participants in 20 to 25 studies. Like the American private clinic, most of these patients (400) were enrolled in observational or low-intervention studies (e.g. urine studies). The US public clinic was also a member of the ACTG and enrolled 80 patients in ACTG studies.

Ugandan Clinic. The Ugandan clinic is a free-standing, urban, outpatient clinic. During an interview, the clinic administrator, knocked on the thin wall of his office and noted that the interiors of the buildings are made so that they can easily be reconfigured when the building reverts to the hospital (UG 050217 EW&CH). The walls of the interior of the clinic make a nice metaphor for the semi-permanent position of the Ugandan clinic in relation to its collaborators and funders. It is a result of overlapping collaborations between an American university, a Ugandan university, a local hospital and the Ugandan government. The organization is experiencing growing pains; it now has over 250 staff but has no core funding or infrastructure. They currently operate more as a group of over-lapping research studies and care programs each of which has its own grant. The clinic was a member of the network of research sites funded by the DAIDS in the US NIH. Because Uganda is a much smaller country than the US with far fewer hospitals

and clinics, I do not specify which network in order to help protect the confidentiality of the Ugandan site.

The clinic staff was made up of physician investigators (trained in the US or UK), medical officers (Ugandan trained medical doctors), nurses, pharmacists, psycho-social counselors, peer counselors and research staff. Half of the physician investigators were white Americans and the rest were black Ugandans. All of the medical officers were either black or Indian Ugandans. All of the nurses and support staff were black Ugandans. The clinic research subjects and patients were all black Ugandans and more or less poor. The majority of the study participants and clinic patients were women and children. Many were just getting by; others were nearly starving.

As mentioned above, Uganda has been the site of many HIV studies. The Ugandan clinic began as a research site and had only recently begun the process of expanding to provide care independent of its research programs. As a site of research, the clinic had more experience with and access to HIV treatment and monitoring technologies than would have been the case in non-research clinics in poor countries. The Ugandan clinic had access to local CD4 and viral load testing. However, they did fewer tests than would be the case in the American setting, and they sometimes worried where funds for viral load testing would come from. The clinic was short on staff and basic medical supplies.

At the time of the fieldwork, the clinic had two care programs and five on-going studies with more being planned. The clinic was in the process of expanding care programs. In comparison to the American clinics, the Ugandan clinic had fewer studies but far more patients in each study. To give a sense of the relative research work load –

the director of the Ugandan clinic described one of their CDC-funded study as a "slow" study, because they see only maybe 4 babies per day for this study (UG 050303 CH&EW). In the US clinics, this would have been considered a fast-paced study. Calculating the actual number of study patients is difficult partly because of the overlapping studies and programs but also because of the more or less formal agreement that the clinic will treat staff and staff family members. There were approximately 800 program participants and 300 research participants; these numbers grew daily as staff recruited new participants. To qualitatively describe the work load, the clinic staff was "swamped," and both study and program participants both spent long hours in the waiting room.

South African clinic. The South African clinic is a free-standing clinic with links to a religiously-affiliated hospital and the provincial flagship university. The clinic described itself as "semi-private," not a government clinic but not entirely private either. South Africa began a "roll out" of antiretroviral therapy through government clinics in 2004. This semi-private clinic was also selected to receive antiretrovirals through the government's antiretroviral program, which means that the clinic had to more or less comply with government treatment guidelines. The patients were certainly not wealthy, but most were not destitute. Patients were expected to pay 70 Rand per month to cover clinic care and antiretrovirals.[6] The patient population was mostly made up of black South Africans; there were a few patients of Indian descent and even fewer white South Africans. In contrast, the physicians were more diverse.

The South African clinic did less research than the other clinics. Also, unlike the others it was not part of a US-funded research network. It did receive funding for treatment through US programs and research funding through an American university. Most of its current research projects were conceived and designed internally and oriented toward improving care within the local clinic. At the time of the fieldwork, the primary clinical research project was a study of HIV drug resistance. The clinic was just beginning to expand its research program. As such, the South African clinic was a bit of an outlier. However, it provided the opportunity to observe organizational learning as it developed its new research program.

Conclusion

In this chapter, I have described the infrastructure of HIV/AIDS research including the virus, the technologies for treating and tracking the disease as well as the "alphabet soup" of the field. An overview of the technical aspects of the disease, including laboratory tests and pharmaceuticals, is provided because these techniques are central to the vocabulary of HIV/AIDS research. Such knowledge is necessary for understanding the very real consequences of clinical research and will make subsequent discussions of the clinical research process more comprehensible. A discussion of the HIV/AIDS research field is not only context for the reader; these organizations and standard setters matter to the HIV clinics examined in this dissertation. They are potentially sources of funding, governance and prestige to the clinics. In the following chapters, I examine how clinic staff members work both with, and around, these standards and organizational requirements.

Chapter 3. Ranking Science: The Production, Evaluation And Codification of "Good" Evidence

Evidence provides a warrant for acts and ideas (Goodman, 2003). When interpreted, it confers the rationale and authority for particular actions. Scientific evidence has become an increasingly important warrant for action. What distinguishes science from other kinds of decision-making practices is the reliance on objective evidence and the supposed exclusion of interests, opinion, religion and other sources of "human bias" (Kuhn, 1996). With the rise of evidence-based medicine, the evidentiary expectations of science are applied to medicine. Evidence-based medicine is defined as using the "best available external evidence from systematic research" along with physician experience and expertise to make medical decisions (Sackett, 1996, p. 71), where the "evidence" is primarily epidemiological. When medical treatments are evaluated through epidemiological research, which treatment is the best, and therefore what to do, is supposed to become clear based on the "objective" evidence. Indeed, evidence-based medicine proponents describe it as a means of managing uncertainty in medical care. However, reality is not so simple. While epidemiological data is formally privileged by evidence-based medicine, it competes with other kinds of information that shapes decision-making in practice including clinical experience, physician intuition, and patient preference. The requirement to use evidence-based medicine may actually increase uncertainty. The expectation that physicians apply epidemiological data to clinical practice raises interrelated epistemological and practical questions: what counts as good scientific evidence, what is its value relative to other kinds of information, and

which evidence should we implement? This chapter examines how medical professionals answer these questions by comparing and contrasting how scientific evidence is defined and valued in formal evidence-based medicine standard of care guidelines to how evidence is produced and judged at the clinic level.

Overview of the Components of Evidence-based Medicine

In common medical parlance, evidence-based medicine is nearly synonymous with clinical practice guidelines, but it also includes the production of scientific evidence, the evaluation of published research and a general orientation toward critical self-evaluation (Timmermans & Berg, 2003). This chapter examines three components of evidence-based medicine: production, evaluation and codification of "good" evidence. Scientific evidence is produced in the day to day conduct of research. This conduct is defined by research protocols which are lengthy, written instructions of how to do the study. Evaluating research occurs when people read and discuss scientific publications and presentations. This work occurs in groups, as in the case of journal clubs where physicians discuss recent articles, and individually when physicians search the medical literature to figure out how to treat a patient or a set of patients with similar symptoms. By codification, I mean the use of science to produce collective rules for the conduct of medical care, particularly written clinical practice guidelines which specify how to treat some disease. When clinical practice guidelines reach the clinic, the evidence has already been judged and translated into doable steps. In contrast to guidelines which are examples of "ready made" science (Latour, 1987) or "pre-packaged" evidence-based medicine (Timmermans & Berg, 2003), observing the day to day conduct of research and

interpretation of journal publications offers an opportunity for observing evidence-based medicine in the making.

Whether people are producing guidelines, interpreting published study results or doing research, science in its various forms is one of the kinds of evidence that medical professionals weigh when they ask, "what should I do?" Further, doing research, evaluating research and making guidelines are all occasions in which groups decide what constitutes good scientific evidence. But the participants do this in different ways and under different constraints. While the research team produces data which they hope will lead to sound, scientific results, guideline makers and journal club members judge the soundness of already produced and published (i.e. cleaned up) scientific results. Doing research, interpreting research and translating research into rules for care bring together clinical experience with scientific evidence. Guideline makers and journal club members judge which scientific evidence is true and likely to be useful in clinical setting while the research team produces scientific evidence in the clinical setting. However, unlike guideline makers, journal club members and the research unit must fit scientific evidence to particular organizational routines, local clinical experience and individual patients. The relationship between scientific evidence and clinical care is typically conceived of as a time line moving from research design and evidence collection to the presentation or publication of research results to the interpretation of presented and published evidence to the production or revision of rules for care to the implementation of evidence into practice. Instead of a timeline, this process is more properly conceived us as a cycle that begins in the clinic with data collection and ends in the clinic with implementation.[7] Thus, somewhat counter-intuitively, doing research is actually more entrenched in the

clinical setting than is interpreting scientific publications or translating science into rules for care. Moreover, as one gets closer to the messiness of the clinical setting, the status of scientific evidence becomes more uncertain.

What Counts as Evidence?

Good evidence is increasingly central to good doctoring. The issue of what counts as good evidence is of growing importance to the medical profession. Over the past 10 to 20 years, there have been numerous publications on how to judge medical research. Examples include a series of articles in the *British Journal of Medicine* as well as texts such as Greenhalgh's (2006) *How to Read a Paper: The Basics of Evidence-Based Medicine (3rd edition)* and Guyatt and Rennie's (2001) pocket-sized *User's Guide to the Medical Literature: Essentials of Evidence-Based Clinical Practice*. Judging, interpreting and producing scientific evidence have become essential skills for physicians. There is the expectation that physicians in training should ask themselves, "What's the evidence?" everyday (Eisenberg, 1999, cited in Timmermans & Berg, 2003.) Evidence-based medicine is becoming a routine part of medical education. According to a 2004-2005 American Association of Medical Colleges survey of 125 medical schools, nearly all (123) have added evidence-based medicine to their curriculum. Similarly, training in how to do research, and thus produce good evidence, is becoming more common. All 125 of the schools surveyed included biostatistics as a required part of the curriculum, and 112 included research methods (AAMC, 2005).

What counts as good evidence is a subject addressed by professional associations and regulatory bodies. Even before epidemiological research is translated into

guidelines, data quality has been regulated. Indeed, the FDA produces standards of "Good Clinical Practice" defining what constitutes good clinical research for regulatory purposes. Likewise, guideline-writing bodies produce standards for what kinds of evidence should be used in guidelines. The US Public Health Service Preventive Task Force standards for guideline development have been particularly influential (Woolf, DiGuiseppi, Atkins & Kamerow, 1996). The most recent version recommends systematic literature reviews so as to avoid selective citation by guideline writers (Harris, et al., 2001). The guideline writers must determine the bibliographic databases to be searched (e.g. *Medline* or *Cochrane Collaboration Library*), as well as the specific inclusion and exclusion criteria, such as study design, population studied, year of study, outcomes assessed, and length of follow-up. Other professional organizations and government entities have published standards for guideline development including the American Medical Association, Institute of Medicine (IOM), US Agency for Healthcare Research and Quality, and the Canadian Medical Association. Such standards are expected to change over time as measurement, equipment, and disease classification change.

The ideal "evidence" in evidence-based medicine is typically quantitative. In particular, proponents of evidence-based medicine classify information according to a hierarchy with randomized clinical trials at the top and case series followed by expert opinion at the bottom. Observational studies (e.g. cohort studies, case-control studies) fall in between (Concato, Shah & Howitz, 2000). This hierarchy is defined in published standards of evidence-based medicine and codified into clinical practice guidelines which typically contain evidence rating schemes. One of the earliest evidence rating schemes

was produced in 1986 by David Sackett, a strong proponent of evidence-based medicine. Sackett (1986; 1993) published these as "rules of evidence" borrowing the phrase from the legal standards governing the admissibility of evidence in court. Like the rating schemes for clinical practice guidelines, these legal standards differentiate between expert opinion and scientific evidence. The US Preventive Services Task Force's evidence grading scheme is particularly influential and has been widely adopted and adapted by other guideline makers. The evidence grading scheme privileges research over expert opinion and randomized clinical trials over other kinds of research designs.

Evidence-based medicine, in general, and evidence rating schemes, in particular, impose a hierarchy that potentially shifts what kind of evidence is valued in medicine. Of particular importance is statistical or epidemiological data. As stated by a group of evidence-based medicine physicians: "the best evidence means using epidemiological and biostatistical ways of thinking" (Davidoff, Haynes, Sackett & Smith, 1995, p. 1085). While the application of epidemiological data to clinical practice is heralded by evidence-based medicine proponents as a way to increase the efficiency, certainty and quality of care, it is not without its critics. Criticisms from the medical profession focus on concerns about the exclusion of clinical experience in favor of statistics and guidelines as well as questions of how to apply evidence at the clinical level (Lambert, 2006).

Furthermore, science studies scholars, feminist epistemologists and medical phenomenologists have criticized the notion of evidence on which evidence-based medicine rests: "Relying on 'the facts' or 'the evidence' to adjudicate between competing clinical practices or scientific beliefs assumes that the evaluative standards of evidence-based medicine are transparent, neutral, objective, and universal" (Goldenberg, 2006).

The idealized notion of evidence in evidence-based medicine is undermined by two central tenets of science studies. First, observation is theory-laden and second, theories are underdetermined by data. In other words, evidence, even in controlled experiments, is never purely objective; observations are shaped by the expectations of the observer, and these interpretations may even be built into the technology of observation (Kuhn, 1965; Latour & Woolgar, 1979). For example, definitions of diseases are laden with presumptions that similar syndromes have single causes and similar outcomes. Further, the interpretation of data is not obvious; evidence can be used to support a number of theories or practices (Feyerabend, 1993). In a rare documentation of the under-determination of medical practice, a 1995 issue of *The Lancet* published two different interpretations from the same research team of the same trial results for a streptokinase treatment for acute ischemic stroke (Horton, 1995 cited in Goldenberg, 2006).

While scholars of science have questioned the status of the evidence in evidence-based medicine, medical sociologists have questioned claims that evidence-based medicine reduces uncertainty. Uncertainty has been an area of inquiry in medical sociology since Fox's (1957) classic article on learning to doctor. Expanding on Fox's (1957) work on learning to manage uncertainty, Light (1979) identifies five kinds of uncertainty: uncertainty arising from superiors' expectations, uncertainty of professional knowledge, uncertainty of diagnosis, uncertainty of treatment and uncertainty of client or patient response. Strategies for controlling uncertainty include: "psyching out" superiors, mastering knowledge, acquiring clinical experience and emphasizing technique over outcome (Light, 1979). Evidence-based medicine and clinical practice guidelines in particular, are an addition to this arsenal for controlling uncertainty. Clinical practice

guidelines and review articles summarize knowledge in a particular area and identify areas where evidence is incomplete and thus may decrease, or at least identify clearly, uncertainties of professional knowledge. At the least, they increase the ability of trainees to "psych out" (Light, 1979) or convince patients and other physicians that they are competent. Clinical practice guidelines are most oriented toward controlling uncertainties of diagnosis and treatment that arise in practice. However, the extent to which evidence-based medicine reduces uncertainty is a source of debate. Recent scholarship in medical sociology counters evidence-based medicine proponents' claims that evidence-based medicine reduces uncertainty, arguing that evidence-based medicine potentially increases uncertainty (Fox, 2000; Hafferty, 2000). Based on interviews with medical residents, Timmermans and Berg (2003) observe that evidence-based medicine is associated with an increase in a new kind of "research-based uncertainty" regarding how to select and evaluate medical studies (see also Timmermans & Angell, 2001).

Scholarship on "research-based uncertainty" bridges medical sociologists' traditional concern with uncertainty with science studies' attention to evidence. This chapter further develops the connection between conceptions of uncertainty and evidence by examining the production, interpretation and codification of scientific evidence. I argue that there is variation among these components of evidence-based medicine in both the amount and the character of uncertainty. In particular, guidelines are associated with lower levels of treatment related uncertainty than is found in the production and interpretation of evidence. Furthermore, there is more uncertainty over what counts as good evidence at the clinic level than there is in formalized guidelines. In the case of interpreting evidence, I observe "research-based uncertainty" similar to what is discussed

by Timmermans and Angell (2001) and Timmermans and Berg (2003). The uncertainty arises from ambiguity and flexibility in what counts as good quality scientific evidence as well as how to integrate this with clinical experience. In contrast, among the research team, uncertainty largely arises from figuring out the rules for doing research. Here there is "rule-based uncertainty" where what counts as evidence is largely a matter of determining what should get measured and documented, and how. I further illustrate these findings below, beginning with the production of scientific evidence, which is the first component in the cycle of Evidence-Based Medicine.

Producing Evidence: Managing Rule-Based Uncertainty

Conducting clinical research is the first component of Evidence-Based Medicine. Evidence has to be produced before it can be published in medical journals or codified in guidelines. The collection of research data is governed by two general sets of standards defining what constitutes "good" science. The first, and the one that typically receives the most attention, defines "good" science in its moral sense and has to do with the protection of human research subjects. These standards are defined by U.S. federal regulations and implemented through local institutional review boards. However, it is the second set of standards to which this chapter primarily attends. The second set of standards defines "good" science in terms of the quality and trustworthiness of data. These standards stem, in part, from the FDA's regulatory responsibility to protect the general public from unsafe drugs. Data standards are also imposed by research funders. In particular, standards regarding reliability and data uniformity have proliferated with the rise of the multi-center clinical trials. Data quality standards shape the day to day

conduct of research, and it is this set of standards on which the legitimacy of clinical practice guidelines and journal articles rests.

In order to perform evidence-based medicine, a physician fits epidemiological data to individual patients. A researcher faces the opposite problem – translating individual patient information into comparable cases in order to produce epidemiological data. Data collection entails the transformation of patients into like cases to produce generally applicable generalizations about treatment. The process of data collection requires the work and coordination of groups. Indeed, "production" is a more accurate term than "collection" because scientific facts do not lie around waiting to be unambiguously discovered. Measurement is rarely as simple as laying a ruler alongside a stationary object; it is a product of social organization and discipline (Porter, 1995.)

Conducting clinical research, particularly multi-center research, is an effort at producing mechanical objectivity. Quantification is especially valued in these circumstances. For something to be quantified, each unit of a population or a measurement has to be made equal. Through a process which Berg's (1997) work terms "reshuffling spokesmanship," quantitative data is privileged over qualitative data and technology over people (see also Porter, 1995; Anspach, 1988). Research funders provide technology and training to increase the objectivity and reliability of measures. The research protocol is the central means of ensuring the collection valid and reliable data. A research protocol contains a set of rules for the actual conduct of the study, including a description of the inclusion and exclusion criteria, randomization procedures, the amount and type of drug a subject should receive, which lab tests should be performed and when, as well as what to do if a subject has a bad reaction to the drug.

These protocols specify what to do, who should do it, when to do it and how to document it. The last is particularly crucial for ensuring trustworthiness.

Filling out case report forms using information culled from medical records or "source documents" is a central way that medical information is translated into scientific evidence. Case report forms, or as the study nurses typically refer to them – "CRFs," are the primary tool for recording data in clinical research. They are essentially paper and/or computer-based questionnaires designed to collect the data as described in the study protocol. For each study, there is one set of CRFs. This same set of CRFs must be completed for each subject in each research site. By the end of a study, this amounts to hundreds of pages per study patient. I spent an afternoon observing a study nurse catching up on her CRFs:

> There is a form that asks for new medications. She said that he did start Advil but she is not going to add that. There is also a form for new signs and symptoms. Yes, he does have a new condition – sciatica. There is an appendix where she can look up the codes for medications, symptoms, diagnosis, AIDS-related clinical events, etc. She looks up lower back pain in the appendix. Pain is coded 757, numbness and tingling down leg is coded 778. The body site is the right leg. There is a code for that too. For date of onset, she writes 1-23-04. She estimates this as the date of onset given that he said that he had this pain for about 3 weeks at his 2-13-04 visit. She classifies it as a grade 2 symptom. (US1 040217 JP)

This illustrates the production of rule-based objectivity. The study nurse quickly decides that Advil, a common over-the-counter pain reliever, does not count as a new medication

for the purposes of research. She then translates the new medical condition, sciatica, into a series of numbers using written instructions supplied by the funder. In this case, translation was straightforward and simple.

However, following the protocol is not always straightforward. In fact, the study nurses spent a great deal of time during formal research meetings and informal hallway conversations discussing questions related to study patient's side effects and symptoms, such as: what grade is it (i.e. how severe is it), do we have to document it, is it related to the study drug, should we stop or change the drug regimen? Sorting out these issues was complicated for the research team (and its observers) because the data-driven questions were intermingled with care-driven issues, such as what does this mean and how should we care for the patient? This uncertainty about how to apply research projects' general instructions to individual's sets of symptoms is often forgotten at the later stages of evidence-based medicine when research results are interpreted and codified.

To elucidate the work of evidence production as well as the uncertainty associated with production, I examine how the research team measures and reports symptoms and side effects. As noted above, the data on producing evidence is drawn from observations of research team meetings, of informal conversations and of filling out paperwork. Some side effects and symptoms are defined as adverse events (AE) or serious adverse events (SAE) which require extra attention and documentation by the research team. Generally, an adverse event is any untoward medical occurrence that may present itself during treatment or administration with a pharmaceutical product, and which may or may not have a causal relationship with the treatment (Federal regulations 21 CFR Part 312). As the name suggests, an SAE is a more severe sub-category of

adverse events; SAEs are typically any new condition that results in hospitalization, disability, life-threatening illness or death. What specifically counts as an AE or SAE is defined separately for each study.

Discussing adverse events was a standard part of the research team's weekly meeting agenda. The used the same pre-printed agenda each week; discussing these topics is a requirement of their ACTG funding. The research team discussed adverse events in 57 of the 65 research meetings that my colleagues or I attended.[8] The first question posed to the group by meeting leader was typically, "any adverse events," "any toxicities" or even once "did we kill anybody this week?" (US1 040112 CH). This provided an opportunity for the study nurses to report on any medical problems experienced by their study patients. More often however, adverse events required far more than a simple report. While every research protocol addresses adverse events, ambiguity remains and it still takes some figuring out. In 35 of the 57 meetings, there was some degree of uncertainty about what to do about that weeks adverse event(s). Note that this probably underestimates the amount of figuring out that symptoms and side effects require because some of this work is accomplished during informal, hallway conversations outside of the research meeting, and while we attended the majority of research meetings, we observed a smaller proportion of these informal conversations.

"Figuring out" an adverse event involves five steps:

1. Identify side effect or symptom
2. Assess and assign grade or severity level
3. Assess and assign the relationship to the study drug (definitely, possibly or unlikely)

4. Determine if it is reportable as an adverse event or serious adverse event according to the protocol of the specific study (different studies have different criteria)
5. Determine how to respond (e.g. stop or change the study drug)

These steps are important, in part, because they are the sorts of things that study monitors pay attention to. There is potential for uncertainty at each step. In particular, this analysis pays careful attention to issues of grading and assigning relationship to the study drug.

Grading is a technique for standardizing the evaluation of symptoms or side effects that arise either as a result of a particular disease or medical treatment. The research team relied largely on the "Table for Grading the Severity of Adult and Pediatric Adverse Events" produced by the DAIDS (DAIDS, 2004).[9] There are four grades: mild (grade 1), moderate (grade 2), severe (grade 3), potentially life-threatening (grade 4). Death is its own special category. The Table is 20 pages long and is divided into clinical and laboratory adverse events. For the laboratory adverse events, grades correspond to a particular range of lab values – e.g. between 160 and 190 mg/dL is a Grade 2 LDL cholesterol. Assessing clinical adverse events is more complicated, entailing assessing the severity of visually observed events (e.g. rash, fat accumulation), the impact of symptoms on patient functioning (e.g. dressing, movement, incontinence, work, social activities), and technological measures (e.g. x-ray). Sometimes grading is somewhat circular depending less on the severity of the symptoms than on the course of treatment taken; for example, nausea is grade 3 if IV fluids are required and a new diagnosis of

diabetes is grade 3 if medication is required. Another complication is that not all possible conditions are included in the table, and particular studies might add grading scales for other conditions. Furthermore, it is not always clear who should do what in the grading process.

Grading is important because it translates individual patient symptoms and side effects into comparable categories for the purposes of assessing treatments and producing study findings. Grading is a good example of where research runs up against care. Typically, grading is not very useful for care. Furthermore, how to grade a clinical condition is not always clear to the research team:

> Study nurse K now notes that one of her patients has a grade 2 rash. She clarifies that it should be classified as grade 2 because they ordered a med for it. But they just need to treat it symptomatically and monitor it. Can they call it grade 1? No, it's grade 2, one study nurse says, and another study nurse clarifies that if you order meds for it then it has to be classified as grade 2. The rash is everywhere. Hopefully he won't swell up, his throat close, and stop breathing. They're making jokes about the grading system now. Would that be grade 4 or 5? 5 [where 5 indicates death.] (US1 031215 CH)

Grading is sometimes not taken seriously by the researchers when it seems irrelevant to patient care or to documenting what they judge is important research data. When the research team was criticized by study monitors for not grading symptoms in the CRFs, the group described their criticisms as "picky" (US1 031124 CH) and about "piddly stuff" (US1 030922 CH).

It is not that the research team does not care about the severity of symptoms or even the integrity of data. Rather, it is that they have somewhat different notions of what matters. The research team cares most about reportable adverse events, especially serious adverse events, which require extra documentation and may even be tracked by the FDA and the institutional review board. The "piddly" criticisms above were cases where the symptoms need only be reported in the CRFs, not as adverse events or serious adverse events. Furthermore, it is important to note that the research team is clearly concerned about the well-being of study patients, but some conditions, while they may be reportable on the CRFs, may not be clinically significant for the patient. An important caveat to this is the fact that the study nurses must often rely on clinic physicians to either grade symptoms or provide enough information so that the study nurses can grade the symptoms themselves. However, the study nurses cannot *require* that clinic physicians provide this information. Issues of controlling uncertainty differ for nurses and physicians, because nurses never gain the autonomy that physicians do (Light, 1979). When the doctors fail to provide enough information in the medical record, the nurses are supposed to track down the doctor to find out the information, which is time consuming.

Assigning relationship of a reported clinical change to study drug was another occasion for uncertainty. Partly, this is because the categories provided by the research forms do not necessarily match how the members of the research team think about causality. In one case, the pharmacist complained that it seems unfortunate to have to check "probably" when what you really mean is "probably not." "Probably not" is not a choice on the SAE form.[10] In another case, the group was not sure if the symptom was related to study drug and had to rely on their "gut reaction:"

> Looking at the papers, they talk about how what they have to report is the abnormal heart rhythm. The problem to put on the form is cardiac arrhythmia. If the patient hadn't been admitted to the hospital, it would have been grade 2, but because he was admitted they have to call it grade 3. They have to list all of his meds on the form …the study nurse has now written the core paragraph, but the pharmacist and the study coordinator are again considering, this time for each of the meds, how to code relationships to the symptoms. At one point, the pharmacist says "Well, M said 'not related' and she's the doctor." Oddly, this still doesn't settle the matter. They talk about how the reviewers will then say "why did you report it then?" But we have to report it if it's grade 3. This was the pharmacist talking, and now the study coordinator says no, that's not true, that technically they don't have to report it if it's not related. The study coordinator continues: "My gut reaction is that we're having this discussion because we think there's a possibility it's related. It's safer to report it." (US1 031215 CH)

A further complication in this is that assessing grade, relationship to study drug, whether to report as an AE or SAE, and what to do with the study patient are not distinct issues. They are interrelated. Here I return to the case briefly introduced in chapter one of this dissertation. The research team was concerned about a woman with lipodystrophy, a side effect of HIV/AIDS treatment in which body fat is re-distributed. They talk gravely about this, agreeing that it's really "terrible" and that the woman is "really getting deformed." They talk about the degree of distortion— "stick-skinny legs," a huge buffalo hump, a lot of fat on her chest, etc. They describe her as being "shaped

like a cube" with her head disappearing into the hump, and say it's been getting rapidly worse over the last 4 months. They talk about how this can be called grade 3 or 4 as a symptom, and they want to end her participation in the study (US1 040402 CH). Later in an informal interview, the study nurse explained that she and the PI want to change the study patient's drug regimen because of the lipodistrophy. The study nurse described it as pretty "gross" –large deposits on her neck, chest, stomach, pubic area. In the email to the protocol team, she described it as grade 3 lipodistrophy. The protocol chair responded that there was no formal scale for lipodistrophy, so what did she mean by grade 3? She responded that she and the PI had defined it as severe and decided that it corresponded to a grade 3 level.

The above example illustrates the power of numbers and how research alters what kind of information counts. The study nurse's description of the lipodystrophy as "gross" may sound disrespectful. However, in conversation with her, it became clear that she simply did not think that it was okay for this woman to live with this malformation, and she was strongly advocating for a change in drugs that might stop or reverse the lipodystrophy. The PI and study nurse code the lipodystrophy as grade 3, even though there is no formal method for or requirement to grade lipodistrophy. This was a strategy for convincing the protocol team to okay a change in the study regimen. Why is grade 3 likely to be more persuasive than "severe"? It translates descriptors – "gross," "skinny legs" and "shaped like a cube," into the language of research – "grade 3." When adjectives are transformed into numbers, it makes them appear more objective – even absent the rule-based, decision process that is supposed to underlie this transformation.

The people who develop and conduct clinical research are engaged in the production of mechanical or rule-based objectivity. One of the purposes of rule-based objectivity is to ensure the trustworthiness and certainty of statements based on these findings. However, when we look behind the data, we see that rule-based objectivity is associated with rule-based uncertainty. Clinical research is a highly rule-driven enterprise. Nevertheless, it is not always clear how to follow the rules or how to fit the rules of research with the demands of care. This uncertainty is made invisible when cleaned up, published versions of the scientific evidence produced by research units in the U.S. and around the world are codified into treatment guidelines like the ones discussed below.

Codifications of Evidence: Efforts to Produce Certainty

The dominant set of guidelines in HIV/AIDS treatment is the US Department of Health and Human Service (DHHS) Guidelines for the Use of Antiretroviral Agents in HIV-Infected Adolescents and Adults. The antiretroviral (ARV) guidelines use a rating scheme that makes it possible to track the balance between scientific evidence and expert opinion over the past decade since antiretrovirals became widely available in the US. These guidelines are especially well-suited for looking at change over time because they are updated more often than is the case for most guidelines. There have been 15 versions of the guidelines since they were first published in 1998. The ARV guidelines rate recommendations using a simplified version of the Infectious Disease Society of America and the United States Public Health Service Grading System. The “Rating Scheme for Clinical Practice” distinguishes quality of evidence and strength of recommendation:

Quality of Evidence for Recommendation:

I: At least one randomized trial with clinical endpoints,

II: Clinical trial with laboratory endpoints[11]

III: Expert Opinion.

Strengths of Recommendation:

|A: Strong, should always be offered,

B: Moderate, should usually be offered,

C: Optional,

D: Should generally not be offered, and

E: Should never be offered" (DHHS, 2008, p. 57).

"Quality of evidence" refers to what kind of evidence a recommendation is based on and the five strengths of recommendation indicate how strongly the guideline writers recommend a practice. Evidence ratings follow recommendations in the text of the guidelines. For example, the "BII" following "Asymptomatic patients with CD4+ T cell counts of 201–350 cells/mm^3 should be offered treatment. (BII)" (DHHS, 2006, p. 8) indicates that physicians should usually offer treatment based on evidence from clinical trials with laboratory endpoints.

The DHHS Guidelines evidence rating scheme provides a convenient means for distinguishing between different types of objectivity. Evidence quality I and II correspond with mechanical or rule-based objectivity and evidence quality III with disciplinary objectivity based on expert consensus. The ratings distinguish between

clinical and laboratory outcomes in clinical research. For simplicity, however, I combine quality I and II to create one category for all clinical research.[12] The DHHS Guidelines suggest a moderate increase in the role of mechanical objectivity over time. The proportion of recommendations based on scientific evidence increases from 45.5% in 1998 to a high of 68.5 % in May 1999 then falls back down then increases to 62% in 2006 (See Table 2). According to a physician in the HIV clinic, the spike in the December 1998 and May 1999 is likely a result of a publication boom in the last half of 1998 (US1 040714 JP). The publication boom was linked to the availability of new anti-HIV drugs and the release of study results timed to coincide with the World AIDS Conference in July 1998. Overall, the increase in the proportion of recommendations based on scientific evidence is partly but not entirely a matter of old recommendations, first rated as III (expert opinion), being changed to I or II following the publication of clinical trials that support the expert opinion. In 2003, the number of rated recommendations doubles. The new recommendations are even more likely to be based on scientific evidence than the old ones. Thus, this reflects a real change in the field.

The rating scheme also allows us to examine certainty as well as the relationship between certainty and kind of evidence. For my purposes, the direction of the recommendation is not important. I combined category A "should always" with E "should never" to create a single category of strong certainty. Likewise, B "should usually" and D "should generally not be" are combined to indicate a moderate level of certainty. The "optional" C rating indicates the lowest level of certainty. Evidence-based medicine has been described as a means of managing uncertainty in clinical care (Timmermans & Angell, 2001). However, the recommendations in the ARV Guidelines

are not characterized by an ever-increasing certainty about how to handle clinical situations. Between 1998 and 2006, the proportion of strong recommendations ranges from 54.5% in the first version to a low of 36.2% in May 2004 (See Table 3). This occurs even as scientific research and clinical experience increases. Technological advancement can produce new areas of uncertainty as research resolves questions in existing areas. The number of antiretrovirals has increased from 11 in 1998 to over 20 in 2006. Each new drug presents physicians with a different set of complications and side effects to manage. Furthermore, new laboratory tests (e.g. drug resistance tests and measurements of drug levels in the blood) are becoming part of the routine laboratory work of HIV care along side CD4 counts and viral loads. It is not entirely clear when these new tests should be performed and how they should be interpreted.

The relationship between kind of evidence and strength of recommendation is an indicator of the value of that evidence. If recommendations based on clinical trials are more likely than those based on expert opinion to be strongly recommended, then this is evidence that the clinical trials are more credible – at least from the point of view of the guideline producers. This is the case for most versions of the guidelines. Overall, expert opinion is underrepresented among strong recommendations and overrepresented among optional recommendations. The relationship is statistically significant for all versions of the guidelines except December 1998 (see Tables 4 through 18). Furthermore, analysis of the guideline content suggests increasing divergence in the kinds of things appropriately decided based on expert opinion versus scientific evidence over time. Expert opinion is increasingly relegated to new or fuzzy areas of HIV/AIDS treatment. In 2003, this was responding to antiretroviral treatment failure; in 2006, it was drug

resistance testing. Fuzziness is not only a product of technological advancement; it may also have social origins, such as how to increase patient compliance. One cannot measure social phenomena through clinical trials; therefore, such phenomena are likely to continue to be relegated to the fuzzy areas. Furthermore, in contrast to what early guideline makers might have assumed, things keep becoming fuzzy. Thus, the role for expert opinion is in fact unlikely to decline substantially overall.

This analysis does not tell us anything about how guideline users interpret these ratings – they may in fact, give more credence to expert opinion or ignore these ratings entirely. It does tell us that within the dominant set of HIV/AIDS treatment guidelines, disciplinary objectivity and mechanical objectivity coexist but mechanical objectivity is more legitimate, at least to the guideline producers. The evidence rating scheme is an abstraction that imposes new meaning. I, II and III are not random labels. As numbers, they impose a rank order and are "a technique for the transformation of arrangements" (Foucault, 1970, p. 146.) I and II roughly correspond to best and second best science. The randomized clinical trial is routinely described as the "gold standard" in the literature on evidence-based medicine. Clinical endpoints are preferred to laboratory endpoints. In fact, this is codified in the policy of the Federal Drug Administration for approving new drugs. What is ambiguous is whether the labels I, II and III (for trials with clinical endpoints, trials with laboratory endpoints, expert opinion, respectively) refer to best, second best and *third best* or whether they indicated best, second best and *other*. Among guideline producers, it appears that expert opinion is considered third best, something to be relied on in the absence of scientific evidence. However, as is discussed below,

practicing physicians are likely to treat expert opinion as an "other" kind of evidence that is incommensurable with and sometimes even preferable to scientific evidence.

Interpretations of Evidence: Wrestling with Research-Based Uncertainty

The data on interpreting evidence is drawn primarily from my observations of the weekly infectious disease department journal club. Journal clubs are especially interesting places for studying the interpretation of science. The journal club fits the model of how people typically think about the relationship between science and medicine; here physicians judge which science is relevant, legitimate, and true and which science should be translated into practice. In interpretive settings like a journal club, physicians act as scientific gatekeepers. Journal clubs have been a part of medical education for over a century (Linzer, 1987). The first journal clubs were organized mainly as a means of keeping up with the literature in an era when access to scientific journals was limited. Since the 1980s, the function of journal clubs has expanded from a forum for keeping up with the literature to include an emphasis on teaching critical reading skills to physicians-in-training (Linzer, 1987). There is variation among institutions and departments in the function, content and attendance of journal clubs.

In the journal club I observed, there were usually 8 to 12 participants including attending physicians, fellows, pharmacists and sometimes medical and pharmacy students. Each week, a couple of people were assigned to present an article or two of their choice; then the group discussed the article. Most of the articles were from top tier general and specialty medical journals, such as *Journal of the American Medical Association, Annals of Internal Medicine, Clinical Infectious Diseases,* and *Journal of*

AIDS. According to the physician who runs the Infectious Disease journal club, one of his goals is to get people talking, to make it interesting; he really likes it when the experienced physicians "that have been around the block many times" come (US1 050615 JP). Likewise, young physicians and pharmacists told me that they liked journal club because they got to hear what the experienced physicians have to say (US1 040324, US1 040512). Thus, learning the opinion of experts is central to the mission of a meeting that is formally about sharing the latest scientific results. Subjecting published research to "expert opinion" is the gold standard of science. Here, however, there is an important epistemological distinction to be explored between experts judging the quality of evidence and experts judging the outcomes of studies by comparison to their individual experience.

Judging the Quality of Scientific Evidence

For comparison to the ARV guidelines, I code each of the journal club articles for type of research and certainty of application (Table 19). Of the 71 publications that I observed being discussed in the course of 28 meetings, 65 were original journal articles.[13] I grouped the 65 articles into four types: epidemiological research (n=39), basic research (n=14), other research (n=6) and case reports/series (n=6). Epidemiological research includes clinical research (n=20), observational research (n=17), and meta-analysis (n=2). Clinical research refers to prospective, ideally randomized, studies where researchers measure the effect of a medical intervention – a drug in most cases. Meta-analysis is a systematic technique for combining the results of multiple studies in order to increase statistical power. Observational studies may be of prospective or retrospective design.

They may measure the effects of medical interventions or simply track the natural history of a disease; in either case, the interventions are not provided by the researchers. Basic research includes animal and laboratory research. The category of "other research" includes behavioral research and cost-benefit analysis. A case report is a description of diagnosis, treatment and outcome for a single patient whereas a case series is for multiple patients. I include case reports and series in order to examine how physicians judge different kinds of research in comparison to clinical experience and to assess what use they make of this kind of data. Case reports/series are essentially formalized reports of other physicians' clinical experience. Case reports/series are published "anecdotal evidence" which differs from scientific evidence in that the findings have not been subjected to systematic inquiry or statistical analysis; therefore, case series may be true but not necessarily generalizable.

Out of 65 journal articles, the journal club discussed the quality of 28; 7 were judged as good, 12 as okay or mixed, and 9 as weak (Table 20).[14] The journal clubs' judgment of epidemiological studies was similar to that of the evidence based rating scheme found in guidelines. In general, clinical trials were judged higher than observational research. Three out of the 20 clinical research articles were judged as good, while none of the 17 observational studies were judged good. Only one of the clinical research articles was judged as weak; five of the observational studies were judged as weak. Of the two meta-analyses, one was judged good and the other weak. However, the group exhibited a general distrust of the meta-analysis, once even jokingly referring to it as "everyone's favorite" (US1 040324 JP). Methods or data quality were

rarely discussed for the other categories of research and not at all for any of the case reports or case series.

These findings largely reflect the existing hierarchy of scientific evidence in the medical field. Observational studies are typically considered less credible than clinical trials according to proponents of evidence-based medicine (Vandenbroucke, 1998). The general worry about observational studies is that outcomes might be explained by confounding factors arising from the absence of randomization. The two generally recognized benefits of randomization are unbiased allocation of treatment to study patients and the application of statistics based on random sampling theory. Vandenbroucke (1998) has also observed that physicians trained in the critical appraisal of literature learn that randomized clinical trials are superior and often do not know how to judge observational research because there are only generalized principles and no simple, hard and fast guidelines. Similarly, the journal club participants were not critical of clinical trials as a general method, only noting when clinical trials did not meet the standards of good research.

While meta-analysis is a key component of evidence-based medicine, the journal club exhibited distrust of the method in general.[15] There is ambivalence and debate over the status of meta-analysis in the medical field. According to proponents of evidence-based medicine, meta-analysis is considered an impartial and balanced method for reviewing the medical literature systematically – as opposed to the more subjective, narrative review. Meta-analysis provides a shortcut for physicians who want to practice evidence-based medicine but lack the time to read and evaluate all of the individual studies on a given topic. On the other hand, critics worry that the selection of articles for

meta-analysis is subjective. At issue here is trust. According to one physician that I interviewed, while meta-analysis are potentially useful, especially in the case of conditions that are rare or hard to study:

> I think that it's possible to do bad meta-analyses because it's all in, you know, selection of studies that contain similar groups of patients and try to answer the questions in a similar way. And so I think that… readers are rightfully suspicious, you know, that what they're reading may or may not be true. (US1 040928 JP)

The doctor went on to describe how difficult it is to judge whether the findings of meta-analyses are trustworthy:

> You could, you know, go on what journal it's published in and who the authors are, where they come from. But to really know, you'd really have to go back and read the studies yourself. Because it… the information that is put in the article is going to be very selective and… and could bias the reader. (US1 040928 JP)

There are two different ways that research articles might be judged as weak or "untrue." They might be judged as weak according to epidemiological standards or clinical standards. This distinction has implications for the status of scientific evidence. I provide extended examples of each to illustrate this distinction. The only clinical trial judged as weak was an article on prophylaxis with fluoroquinolones, a sub-group of antibiotics, in cancer patients with neutropenia, a condition characterized by a low level of white blood cells which puts them at increased risk for infection. The authors

conclude that prophylaxis is associated with lower levels of infection-related mortality. The journal club members are critical of the findings:

> Dr. O doubts their findings. He says that all other studies show no difference in mortality outcomes with fluoroquinolone prophylaxis. Dr. G says that this looks like a "big, formal anecdote" to him. Others agree. Dr. O says that if they had had an optimal design in the first place, there wouldn't be this confusion over their results. Dr. V asks, what would be an optimal design – and pointing to the younger MDs at the other end of the table, as if to say – listen and do this research project. Dr. O said that they should have had a double blind placebo controlled trial. Then the DMB [data monitoring board] could have adjudicated between the groups. Another doctor says that they did it this way because it was easier. …. An MD asks the presenter if they had a rule in place for stopping the study. She says just the "stringent stopping rule." Dr. O says so, basically, it was "expert judgment." He says it would have been different if they had a blinded DMB. It [the difference between the groups] could have just been coincidence. (US1 050316 JP)

In another case, the group criticized a prospective, observational study of injection drug users with HIV. The study compared survival rates in injection drug users receiving treatment for HIV (referred to as Highly Active Antiretroviral Therapy or HAART) with those who were not. The authors concluded that physicians should consider starting HAART earlier than the current treatment guidelines indicate to increase the survival of HIV-positive injection drug users. In contrast to the above study,

at least a portion of the journal club supports the findings. According to one MD, this is a case of reaching the right conclusions [start HAART early] for the wrong reasons. But the methods are roundly criticized:

> Dr. G points out a selection bias in the study. These are not randomized groups. The no-HAART group is likely to be the worst drug users and the most in danger of dying. Dr. T suggested that they should have limited the group to active drug users. There is some back and forth between Dr. G [the assistant chief of infectious disease and Dr. F [a fellow training to be an infectious disease specialist] about how functioning active drug users can be – with Dr. E suggesting that sometimes they are quite functioning. Dr. G mentions that they don't see too many of the non-functioning at the clinic, but he used to see a lot at his previous institution [located in a poorer area of the city.] Dr. Q says that they also should have controlled for co-morbidities, especially Hepatitis C. … Dr. Q says that the authors reach the conclusion that you should consider early therapy, but they have p-value that is not highly statistically significant. He goes on to say that his recently published article had a highly significant p-value, noting as an aside that the p-value was probably why his own article was published in [a prestigious medical journal.] (US1 040825 JP)

Both of these articles are criticized because they do not meet the standards of epidemiological research; they are not "scientific" enough. The criticisms have to with blinding, randomization, statistics, variation in study groups, and poor outcome measures. The HAART study also shows the attention given to p-values by the group. Statistical

measures like p-values provide an easy, if imperfect, way for people who are not statistical experts to judge findings quickly. In effect, p-values provide a simple rule for judging what numbers matter. A p-value below a pre-determined value (usually .05, .01, or .001) signals that the finding is "statistically significant" or not likely to be the product of chance and thus potentially read as "real" or "true." Note however, that the reverse assumption that statistically insignificant differences are likely the product of chance, is very often wrong.

There are other issues besides scientific evidence here. The dispute over the importance of distinguishing active drug users in the HAART article raises the issue of experience. Basically, the more experienced attending physician argues with the fellow that his relatively brief experience is not generalizable.[16] Furthermore, in the case of the prophylaxis article, the study findings go against what the group already believes to be true. Prophylaxis is a source of contention between oncologists and infectious disease specialists. Oncologists worry about pleasing patients who feel unprotected when they are not on antibiotic prophylaxis; infectious disease specialists worry that prophylaxis increases the risk of hard to treat drug-resistant infections (US1 050316 JP).[17]

Evaluating study results is not just a matter of judging their scientific value. In the journal club, study findings are evaluated in the context of experience, clinical knowledge or reasoning about mechanisms. A simple example of the last is that antibiotics do not help with viral infections. Study findings were sometimes criticized for not "making sense." For example, a study whose purpose was to identify risk factors for secondary infections in patients with cancer found that the rate of secondary infection is higher in patients whose fever is caused by something other than a clinically documented

infection; the other potential causes include microbiologically documented infection or a fever of unknown origin. One of the doctors has questions about this finding and points to the table which reports that a microbiologically documented infection or a fever of unknown origin were twice as likely to lead to a secondary infection than a clinically documented infection. According to the doctor, this "doesn't make sense." He says, it might make sense statistically, but it doesn't make sense clinically; why would a clinically documented infection be different from the others? Another physician tells him that it is probably just a statistical issue. The article presenter agrees that it doesn't "make sense" (US1 050309 JP). In another case, a nutritional study concluded that taking a multivitamin containing B, C and E slowed the progression of HIV. The physician presenting the article pointed to the table which showed that the patients on multivitamins had better CD4 counts and viral loads, shrugging and saying "go figure." Another argued that the finding "doesn't make sense" (US1 040707 JP). In other words, the findings, while statistically significant, do not mesh with what is known about the mechanisms of nutrition and HIV.

In another case, journal club participants disagreed over the use of caspofungin, a new, expensive antifungal. One group of physicians, some of whom are members of the hospital formulary committee, criticized the use of caspofungin as a first line drug to treat invasive *aspergillosis* (a kind of fungal infection) in the absence of research proving it is better than existing, less expensive drugs. The drug was approved by the FDA in 2001 based on evidence of its efficacy when other standard treatments fail; it was considered a "salvage" regimen. Another physician acknowledged the absence of research support for caspofungin as a first line drug but nevertheless argued for the use of caspofungin, saying

it “makes sense” and “it works.” (US1 040901 JP) According to another physician, the cost of caspofungin therapy is ten times the cost of conventional therapy, but conventional therapy “really isn’t conventional” anymore. (US1 040901 JP) A member of the hospital formulary committee said that caspofungin has become part of the standard first line treatment as a result of anecdotal evidence and drug company marketing of caspofungin as a first line treatment for invasive *aspergillosis*. Another doctor commented that drug companies are not supposed to advertise off-market use of drugs. The formulary member smiled, raising his eyebrows and hands, and commented that once a drug gets FDA approval, it makes it easier to use if for other uses. This conversation among the journal club members represents a clash between the different kinds of information that guide medicine: statistical evidence, clinical experience (we “know” this works), pathophysiology (it “makes sense” given what we know about disease progression) and cost.

Not Judging the Quality of Scientific Evidence

In 37 of the 65 articles, the journal club did not discuss the quality of the data, which is remarkable in a setting devoted to presenting and evaluating published research. No quality discussion means different things depending on the type of article. Case reports are not science and thus one would not expect methods or data quality to be discussed. The quality of basic science articles was often not discussed. Basic science articles often report about some new kind of proposed treatment that may be in the pipeline and thus have no immediate clinical relevance; the physicians take a wait and see attitude before weighing in on such findings. More importantly, physicians are largely

not trained to judge basic science. Instead the evidence was treated as an already established fact. Practitioners must typically rely on resident experts (e.g. the Chief of the Infectious Disease department who was considered the basic science person in this setting) or external experts (e.g. journal editors, reviewers) to screen out poor quality basic science. More puzzling is the finding that quality is not discussed for roughly half of the clinical research (10 of 20) and observational research (8 of 17) articles (see Table 20). In some cases, the group simply did not spend much time discussing some articles because they did not appear to find the article relevant or interesting or simply because they ran out of time. However, in the majority of the cases of "not discussed," they treated the article more as a report of an established fact, rather than a finding to be evaluated. They jumped straight to the question of the utility the study results, skipping the step of judging the truth of the study results.

The finding that the journal club members accept epidemiological findings at face value roughly half the time suggests the existence of two competing and overlapping orientations toward epidemiological evidence. Previous scholarship on the use of evidence-based medicine by pediatric residents observed two ideal types of orientations: researcher and librarian (Timmermans & Angell, 2001; Timmermans & Berg, 2003). For the residents who used medical literature like librarians, the authority of information rested on its textual format, the status of the author and the status of the journal (Timmermans & Berg, 2003, 147). In contrast, residents with the researcher orientation critically evaluated the literature. These orientations are ideal types and not mutually exclusive categories. However, I did observe some consistency in the journal club. In general, whether or not one took on the researcher or librarian orientation appeared to be

related to expertise and career stage. Among those who took on the researcher orientation were a handful of physicians who had expertise in research methods. One physician had expertise in basic science (mentioned above), another had a master's degree in epidemiology, and others had a lot of experience conducting clinical research. The fellows, who were in training to become infectious disease specialists, typically took on a researcher approach to the presented articles. It is unclear whether this indicates a real generational shift in orientation. More likely, it is a case of physicians in training showing their superiors that they understand research design, which is an indication of the importance of research methods in medical training.

The librarian/researcher distinction in the approach to evidence-based medicine has links to science studies scholarship on scientific fact making. According to this scholarship, scientific findings are not inherently true; instead, they are made true or constructed as facts through a social process. In particular, Latour (1987) attends to the role of rhetoric in fact construction. The process of fact construction involves omitting the qualifiers – e.g. "Dr. Smith's group from University Hospital found that X causes Y," becomes simply – "X causes Y." The journal club participants made the scientific results of articles more or less fact-like when they talked about the study design, the affiliation of authors and the strengths and weaknesses of the statistics. Scientific results become most fact-like when the physicians take on the librarian orientation toward study findings. The quality of findings is not evaluated at all and the discussion skips straight to questions of implementation. Clinical guidelines are more often treated with a librarian orientation than are journal articles.

Judging the Utility of Scientific Findings

The journal club not only exhibited more uncertainty about the status of evidence than the standards of evidence-based medicine would predict, but more uncertainty about what to do with evidence. I developed a coding scheme for whether or not the group intended to follow the study findings similar to the evidence rating scheme of the ARV guidelines for certainty: should always follow (1 case), should usually follow (19 cases), should usually not follow (9 cases), should never follow (1 case), and optional follow (12 cases), not discussed (23 cases). Most of the 23 cases where the group did not discuss whether to implement findings were basic science, where results are typically not immediately applicable, and case reports, where generalizability is in question. The other cases where they did not talk about whether to implement included interventions that were outside of their province (e.g. pediatrics, ER staffing, sanitation practices on cruise ships, antimicrobials in household products) or when they ran out of time in meeting.

The difference in certainty between the antiretroviral guidelines and the journal club is stark. In any given year of the antiretroviral guidelines, somewhere between 54.5% and 36.2% of the recommendations were in the form of "should always/should never." In contrast, only two out of the 65 journal club articles were characterized by a high degree of certainty. 67.4% of the journal club articles' findings are assigned moderate certainty; 20.9% are assigned low certainty (Table 21). Note too, that the relationship between certainty and scientific evidence is not found in the journal club. In fact, the findings from case series, which are essentially formalized experience and anecdotes, are perhaps more likely than the findings from epidemiological research to be accepted by the journal club (Table 22). With case reports, the issue of translating data

into applicability differs from that of clinical trials. Data may not be as reliable but the reasoning is probably fuller – a puzzle way of thinking rather than a recipe way of thinking. However, this is based on a really small number of cases. There are only six case series and implementation is only discussed for three of these, and the fact that no evaluation was done except for positive ones is odd.

There are several reasons for this discrepancy in certainty between the guidelines and the journal articles. Guideline makers systematically select high quality, conclusive articles. In contrast, the journal club selected articles according to the individual preferences of the members, and the articles were of varying quality and relevance. However, the majority of the journal club articles were selected from top tier journals. As important, however, is the fact that the "should always/should never" language of the guidelines simply does not match the language of journal articles or how physicians interpret study findings. The findings in journal articles are rarely clearly proscriptive. Even fairly clear study findings are couched in terms like "should consider." There is usually an "except" that follows the groups' conclusions regarding whether to follow a study finding. For example, the group agreed with an article's conclusion that prophylaxis with mupirocin (an antibacterial ointment) did not prevent hospital-acquired *staphylococcus aureus* infections in non-surgical patients. However, they speculated that maybe it would be useful for patients in the intensive care unit (US1 040324 JP). In another case, the physicians generally agreed that a new TB assay would not be useful. They were, in fact, highly critical of the assay and the study. Still, they suppose that you might want to use this test in certain specific clinical situations, and while the test is not good at the population level, it might be a good marker of individual patient change (US1

040428 JP). Because of this uncertainty, it was difficult to code the journal club articles as "always" and "never". In my observations of the journal club, there is only one coded case of "should always follow" and one of "should never follow." And these cases are, in fact, more accurately described as "should almost always follow" and "should almost never follow."

The "always" article was a study of the efficacy of different HIV drug combinations. The article reported the results of ACTG 5095 (the same study discussed above in this Chapter and in Chapter 1); the journal team knew some of the data for this study was collected at the local HIV clinic. The journal team judged the article as high quality and trustworthy. The authors of the article concluded that a three NNRTI (nucleoside reverse transcriptase inhibitor) combination was not a good choice for an initial HIV treatment regimen. This was big news in the HIV field because of high hopes for Trizivir, a newly developed drug that conveniently provided three FDA approved nucleosides in one pill. The conclusion of the article was fairly prescriptive. Everyone in the journal club agreed that you should not use a triple nucleoside as an initial regimen. One doctor described this as a "conclusive study" because this was the best triple nucleoside regimen they had. The study results were already known by the group and had already changed their practice. There was uncertainty however in how to translate the finding to some individual patients. Dr. T. said that this left them in a "predicament" – what should they do if someone is doing well on Trizivir? They speculate that they could add one of two drugs (efavirenz or tenofovir) to the Trizivir regimen (US1 040505 JP).

The "never" article also raises some degree of uncertainty in the group. This was a follow-up of a clinical trial conducted in the 1930s of BCG, a tuberculosis vaccine.[18] According to the study, the vaccine had approximately 50% efficacy rate, and the effects of the vaccine did not wane over time as many had assumed. The vaccine is used in other parts of the world but not in the US. The vaccine can cause a positive TB skin test. The journal club physicians sometimes get patients from Mexico with positive TB tests who have received the BCG vaccine. They generally assume that someone who has a positive PPD (TB skin test) was really exposed to TB – not that their BCG vaccine is causing the response. In fact, as one physician notes, this is how the guidelines say to respond to positive PPDs. According to the CDC, the BCG vaccine is rarely indicated in the US, and positive PPDs in people who have received BCG should be used as evidence of a TB infection (CDC, 1996). The physicians did not plan to change their practice. However, the article did raise some uncertainty. According to one MD, this article would make her change her assumptions if not her practice. She had always assumed that BCG only provides time limited prevention (US1 040512 JP).

Even in the cases where certainty is the highest, where the group agrees that they "should nearly always" or "nearly never" follow the study findings, uncertainty still lurks. One of the responses to uncertainty about what to do is a call for better or further research. For example, in the case of the question of what to do with the HIV patients doing well on the triple nucleoside regimen, the group is waiting for more study results to find out if and how to change their regimen. In the second part of the ACTG study, they are going to compare the effects of adding Efavirenz or Tenofovir (US1 040505 JP). In the case of the TB vaccine, the group notes that further research is needed to figure out

what the baseline rate of TB infection would need to be to justify the use of this vaccine (US1 040512 JP). The call for additional research does not help with what to do about specific patients in the short run; even with unclear information, the physicians have to decide to do something.

In practice, the interpretation of scientific evidence is far more complex than the simplified categories of evidence found in the cleaned-up guideline recommendations. In particular, the journal club judged scientific results according to three different sets of standards: epidemiological (i.e. is the study design appropriate? Are there proper controls? Are the statistics good?), the established professional knowledge system (i.e. do the findings make sense given the accepted knowledge of disease mechanisms?), and physician's clinical experience. The journal club's judgment of the quality of evidence was similar to the hierarchy of scientific evidence described in the medical literature. The exception was meta-analysis. The judgments of meta-analysis were mixed and the group exhibited a general distrust of meta-analysis as a method. One of the most unexpected findings was that the journal club did not judge the quality of over half the articles that they discussed, exhibiting the "librarian" orientation observed by Timmermans and Angell (2001). As predicted by sociologists critical of evidence-based medicine (Fox, 2000; Hafferty, 2000), among care-oriented physicians, epidemiological research is not associated with greater levels of certainty of what to do. In particular, the "always" and "never" language of the ARV guidelines does not match how physicians approach research findings. There are almost always caveats and exceptions.

Conclusion

The cycle of evidence-based medicine begins and ends with the individual patient in the clinic. The evidence on which evidence-based medicine relies is culled from the bodies of individual patients, translated into data and generally applicable treatment recommendations, and then applied to a wider set of individual patients. This process entails a back and forth between the general and the particular and the neat and the messy. Abigail Zuguer, MD (2006) writes of the illusion of order and the reality of messiness: "I figured that once training was over, life would become as orderly as it was in the [medical] journals. It was a delusion born of sleeplessness: medical reality always diverges wildly from the printed record. Drugs often don't behave the way they do in studies, and patients almost never do."

Evidence-based medicine is an attempt to control the messiness and uncertainty of medicine. Clinical research is highly rule-driven work, the outcome of which is data that are subsequently translated into rules that are supposed to govern care. One of the purposes of rule-based objectivity is to ensure the trustworthiness and certainty of scientific findings. However, when we look behind the data, we see that rule-based objectivity is associated with rule-based uncertainty. It is not always clear how to follow the rules or how to fit the rules of research with the demands of care. Ironically then, evidence-based medicine creates new uncertainties. Uncertainties arise from how to make the neat categories of the study design fit the reality of the patient and the clinic. In the case of the journal club, research-based uncertainty arises from what counts as good evidence and how to apply it. Unlike the guidelines, and as predicted by sociologists critical of evidence-based medicine (Fox, 2000; Hafferty, 2000), epidemiological

research does not translate into greater levels of certainty of what to do among the journal club members. In particular, the “always” and “never” language of the ARV guidelines does not match how physicians approach research findings. There are almost always caveats and exceptions. Partly, this is a result of guideline writing being more rule-bound than interpreting science. Guideline writing does not allow for the expression of as much uncertainty. The form of the guidelines imposes a level of certainty that does not reflect clinical reality.

Chapter 4. Enacting Standards: When, If and How Research Protocols Work[19]

In *All the Kings Men*, the character of Willy Stark colorfully describes the law as "a single-bed blanket on a double bed and three folks in the bed and a cold night… There ain't ever enough blanket to cover the case, no matter how much pulling and hauling, and somebody is always going to nigh catch pneumonia" (Warren, 1946, p. 136). Codifications are nearly always inadequate for the real world. This chapter attends to the "pulling and hauling" in which researchers engage in order to fit protocols to the clinic. While some pulling and hauling is always required, some rules systems are better than others, and the amount of pulling and hauling varies depending on the adequacy of the rules.

In order to do clinical trials, researchers fit rules and procedures to individual bodies and local situations. Every clinical trial, from the smallest to the largest, is governed by a study protocol which describes how to conduct the study. As "coordinating devices," research protocols structure and sequence work (Timmermans & Berg, 2003). They specify what tasks to do, how to do them, when to do them, who should do them and to whom they should do them. The purpose for following a protocol is to create comparable results from the bodies of hundreds or even thousands of different patients and from the labor of numerous different researchers in dispersed clinics. Following the protocol is the core of research work at the clinic level, and the primary goal of clinical researchers is making the protocol work. When protocols do work, they have to be made to work. Scholarship on standardization has observed that making standards fit local conditions typically involves more work than standard producers

assume. The importation of a universal decision-making tool into a local setting requires a process of back and forth between the tool and the setting (Berg, 1997; Giddens, 1991).

Research protocols provide a good place for studying rule-following because researchers (unlike Willie Stark) are really trying to follow the rules. This commitment to rules arises in part from a general acceptance of the importance of replicable science. More important, at least to the day to day work of research, are the layers of internal and external scrutiny to which research work is subjected. This scrutiny surely shifts rule followers' priorities to those activities which are most observable and documentable, but it also means that lots of rule-following actually occurs. From the point of view of those who monitor research, work is either compliant or non-compliant, and researchers worry about getting "dinged" for non-compliance. However, rule-following is a more flexible act than the dichotomy of compliance and non-compliance suggests. This chapter examines how protocol compliance occurs, when and if it does.

Research as Rule-Work

As discussed in chapter one, medical research has become increasingly standardized and rule bound in recent decades. National and international bodies produce general sets of guidelines regarding research ethics and data quality. Private and public research sponsors produce rules for their grantees. Furthermore, each research project is governed by a study-specific protocol that details how the study should be conducted. With the expansion of multi-center research, protocols become more concerned with comparability and uniformity. As the general clinical research enterprise is progressively

more formalized and codified, the work of conducting clinical research is increasingly oriented toward rule-following.

Furthermore, the shift to large, multi-center clinical trials has implications for the content of research work. With the rise of multicenter clinical trials, researchers increasingly use research packages (protocols, SOPs, support systems) produced by distant others which may not easily translate into the work environment of the local clinic. While protocols for investigator-initiated research projects are usually written by an individual or a small group located in a single clinic, protocols for multicenter research are created through a lengthy process of committee work involving nationally and even internationally respected experts. Another effect of the rise of multicenter clinical trials is increasing concern with uniformity. Multicenter trials place a strong emphasis on standardization so that data produced in different clinics will be comparable. Thus, in addition to protocols, study sponsors provide other standardizing devices, including educational materials, computerized and paper forms, and supplies.

To a great extent, the work of conducting research is a matter of producing and following rules and procedures. Drawing on the fieldwork as well as the World Health Organization's (2005) overview of clinical research, the conduct of research at the clinic level can be described in six steps:

1. Obtain research protocol from sponsor or produce own research protocol
2. Produce locally and/or acquire from sponsor standard operating procedures (SOPs) in order to follow the research protocol (includes procedures for shipping and handling, communicating with IRB, performing lab tests and other things not specifically defined in the protocol)

3. Produce locally and/or acquire from sponsor the systems, documents, and tools needed to facilitate the conduct of the study (includes case report forms, lab equipment, computer software and hardware, fax machines needed for reporting SAEs, signature logs, administrative sheets for tracking funding, drugs and supplies, informed consent forms, adverse event reporting forms, checklists, etc.)
4. Obtain initial and periodic approval from relevant ethics committee and regulatory authorities (e.g. local IRB and FDA)
5. Conduct trial according to the protocol and SOPs
    a. Enroll subjects (includes recruiting, determining eligibility and obtaining informed consent)
    b. Administer study intervention and measuring the effect
    c. Collect and manage data
    d. Manage and report adverse events
6. Work with monitors to ensure protocol compliance

The protocol is essential to research. Every clinical trial, from the smallest to the largest, is governed by a protocol. At the clinic-level, the primary goal of clinical researchers is making the protocol work. They use strategies to encourage protocol compliance and avoid protocol "violations." To illustrate the content of protocols and the general process of clinical research, I return to the example of ACTG 5095 previously discussed in Chapter one and describe the protocol for this study in some detail. US private clinic was conducting this study at the time of fieldwork, finishing up the study just a few months before we left the field. ACTG 5095 was designed to compare the

efficacy of Trizivir to other HIV treatment regimens. Trizivir is three medications combined in one pill taken twice a day. There were high hopes for Trizivir. The low pill burden combined with few side effects was appealing to patients. Furthermore, combining three drugs in one pill means that patients pay one, rather than three, insurance co-payments. However, the results of the study were disappointing. Trizivir, alone, was less effective than the other regimens.

ACTG 5095 was a double-blinded comparison of three drug regimens: Trizivir, Trizivir plus Efavirenz and Combivir plus Efavirenz. Combivir contains two members of the NRTI class of antiretrovirals (Lamivudine and Zidovudine) in one pill. Efavirenz is a member of another class of antiretrovirals, NNRTIs. (See Chapter two for a more detailed discussion of HIV/AIDS treatment.) According to the 5095 protocol, if the patient experiences intolerable side effects or "viral failure" to their initial study regimen, they can change to a second drug regimen, described as Step 2 in the protocol. Viral failure occurs when medications do not reduce the amount of HIV in the blood to an extremely low level; in this case, the cut off level is 500 copies per ml of blood. On Step 2, the physician and the study patient select another medication regimen from one of three study arms. (See Appendix 1 for an excerpt from Step 2). Here there is much more flexibility, but also more complexity. Part of the regimen is selected by the study patients and their physician and part is defined by the protocol. In each of the three arms, study patients take a combination of two or three drugs from the NRTI class of HIV drugs as well as one or two protocol-defined drugs from another class of HIV drugs. The relatively simple list of the three arms of the study is followed by a long list of exceptions

and rules for selecting regimens. For example, patients with heart disease are not allowed to be in one of the study arms. And some drugs cannot be taken with each other.

An excerpt from the schedule of events for ACTG 5095 provides a week by week account of the work for this study and illustrates the complexity of research work (see Appendix 2). For example, at the week 12 study visit, the researcher should clinically assess signs and symptoms, document all medications, take body measurements, have the patient fill out a paper survey, and order a long list of lab tests. The "F" under week 12 indicates that the study patient should fast before the study visit. The hematology, chemistries, LFT (liver function tests), lipid panel and metabolic tests are groups of tests and amount to over 20 different lab tests. Which lab is to conduct each of these tests and how the lab specimens should be transported to the specified lab is documented in the protocol or the SOPs. All of the clinical and laboratory information is recorded in sponsor-supplied case report forms (CRF). Some clinical signs, symptoms and lab results may have to be reported to the sponsor DAIDS as "adverse events." What counts as an adverse event is defined in the protocol. Some adverse events only require documentation; in other cases, study patients may have to change drug regimens or stop the study drug regimen and become "off drug/on study." How researchers work with these often complicated protocols is discussed below.

Making Protocols Work in the Clinic

Clinical research is a technique for translating patients into numbers. Like other techniques of quantification, the randomized clinical trial can be "transported across

oceans and continents" (Porter, 1995, p. ix). Indeed, clinical research is a global enterprise. However, transporting the randomized trial around the world is an enormous amount of work. Standards, after all, finally do connect with the real world. In the course of enacting standards, researchers and patients achieve only a very "local universalism" which "emerges from localized processes of negotiations and pre-existing institutional, infrastructural, and material relations" (Timmermans & Berg, 1997, p. 275).

Quantification relies on standardization, and the research protocol is an important apparatus of standardization. The point of the protocol is to allow researchers to produce comparable results in diverse settings. The commensurability of the results depends on the extent to which the protocols are followed. As scholarship on standards in action suggests, rule-following is complicated. First of all, many rule adoptions are more symbol than substance, and some rules never intrude on the actual work of the organization (Meyer & Rowan, 1977). Even when rules are taken up by the organization, how they effect work is uncertain. The rule as written is distinct from the rule as enacted. Rules are enacted through more or less improvised performances (Weick, 1998). Rules "trigger" action but do not "over-determine" the content of action (Reynaud, 2005).

Timmermans and Berg's (1997; 2003) work on "standardization in action" provide valuable insights for examining how researchers work with protocols. Using protocols requires "tinkering," a back and forth between the work and the rule. Tinkering occurred when the researchers worked to fit real patients to protocol supplied slots. Scholarship on social studies of science, have shown that laboratory research relies on the availability of standardized materials, ranging from chemical reagents to white mice. Research on humans is different in obvious ways. Human research subjects are not

purchased through regulated supply chains. Instead, they are recruited by researchers. Through recruitment, researchers make decisions that help standardize their supply of human research subjects. In particular, researchers screen out the most unpredictable patients. After recruitment, researchers creatively tinker with the protocol and the clinic to make them fit each other. The more inadequate the rule system, the more tinkering work required.

For the most part, researchers are not mindless recipe followers but "thinking cooks." In the course of making sense of and using protocols, they potentially increase their skills in ways that make the protocol work better. In other words, both the intelligence of the researcher and of the protocol grows. Finally, while rule following may be the official responsibility of a particular individual in an organization, enacting rules almost always relies on the cooperation of other parties who may have different interests and commitments. The research staff negotiates with other staff, departments and organizations over activities ranging from the shipping of lab specimens to the content of medical care. These strategies for working with protocols are elaborated in more detail below.

Controlling the Input Stream: Finding and Making Good Study Patients

Sometimes patients simply do not comply with study regimens. They miss drug doses; they fail to accurately report missed doses; they show up late or not at all for study visits; they forget to fast. Other times, patients' bodies do not cooperate with protocol requirements. They develop new diseases or symptoms that do not fit the protocol. They have lab results that fall out of the study criteria. Through a process Pickering (1995)

calls the “mangle of practice,” objects of study exhibit agency which impacts scientific practice and outcomes. Human raw materials introduce a kind of reactivity that is different than when the raw materials are non-human (Heimer & Stevens, 1997). As Leidner (1993) observes, even in seemingly simple fast food work, it is hard to standardize work when you are working with people. One strategy is to screen and protect organizations from the social variability introduced by human raw material (Heimer & Stevens, 1997).

Recruiting research subjects is a serious matter for the research team in the American private clinic. They spend a good portion of each weekly research meeting discussing recruitment, rejoicing over successful recruits, identifying potentially “good leads” and lamenting the “good leads” that did not pan out. The research unit’s reputation and funding depends on the number of subjects they recruit. At the same time, they are penalized for recruiting patients who later drop out of the study. Thus, the quantitative pressure to recruit is mediated by qualitative judgments. The research team is under pressure to recruit not just bodies but good subjects. Researchers search for patients who fit the protocol by searching the computerized records, by alerting clinic physicians and nurses to the kinds of patients they are looking and advertising available studies to patients. In some cases, researchers shift gears and instead of searching for the right patients for the protocol, they search for the right protocol for the patient. The last scenario often occurs when patients need free or experimental treatment.

As has been observed in the sociology of medicine, patients are depersonalized and described in terms of their disease (see e.g. Anspach, 1988; Heimer & Staffen, 1998). In the case of clinical trials, researchers match disease categories and symptoms to

protocol slots. So for instance, the research team might search for "new starts" (patients who needed to start HIV treatment for the first time.) When patients were in search of a study, the research staff at the American private clinic would go around the table, each going through their inclusion checklist – Hepatitis? Neuropathy? High cholesterol? Is the patient a thin, couch potato who smokes (all indicators that they would have low bone density)? To an outsider, this conversation would seem odd with people asking almost hopefully about the existence of characteristics that would typically be viewed with dismay. Occasionally, the nurses become aware of the strangeness. For instance, upon hearing that a potential subject had neuropathy, one nurse excitedly exclaimed "good," then caught herself, clarifying that it was certainly bad for the patient but good for the study.

When researchers recruit good study subjects, they are limiting "input uncertainty." Input uncertainty refers to the unpredictable variability of patients' signs and symptoms as well as their actions (e.g. medication compliance). Healthcare organizations face "input uncertainty" in the make-up of their patient stream. High input uncertainty decreases the ability of organizations to use programmed coordinating devices, such as rules and regular meetings (Argote, 1982). One means of increasing the ease of standardization is to reduce the uncertainty of the inputs. Choosing study patients who fit certain technical and social criteria makes the protocol more workable. The more a research site is structured like the clinic imagined by the protocol writers, the easier it is to follow the research protocol. Similarly, the more actual patients are like the ideal research subject, the easier it is to follow the research protocol.

Researchers limit input uncertainty through screening out some of the patient variability. This is formally built into the research protocol through the inclusion/exclusion criteria. Common exclusion criteria in studies of HIV drugs include pregnancy, liver problems, kidney problems, lipid problems. Inclusion/exclusion criteria increase the ability of researchers to measure the effects of interventions. They also reduce the likelihood that study patients will be harmed by study interventions. They reduce the variability of patients, making it easier to follow research protocols. Unfortunately however, such criteria have the effect of reducing the applicability of studies to the wider patient population – many of whom have the conditions excluded by the study protocol.

Researchers further limit uncertainty when they select and recruit patients who are the most likely to be compliant study subjects. In order to conduct a study, good study patients must be found and trained. Good study candidates are patients who are compliant which means they take their medications and come to their clinic appointments on a regular basis. Medication compliance is determined mostly by lab tests and patient reports. Medication compliance is especially important in HIV care, where an adherence rate of 95% is required to prevent viral breakthrough and the development of drug resistance. It is especially hard to judge who is likely to be a "good" subject when recruiting "new starts" or "treatment naïve" patients; they have no record of compliance. In fact, the research staff at the American private clinic complained about how difficult "treatment naïve" studies were. In part this was because some patients were newly diagnosed and so grappling with that difficult information, but it was also because the "new starts" were still learning how to be good patients.

There are exceptions to the requirement of relative medication compliance. For example, observational studies might accept patients who do not take their medications so long as they are truthful about what medications they do or do not take. In addition, patients who have trouble keeping up with their medication might be good recruits for studies which problematize pill burden and adherence. For the research team, there is generally less penalty for medication non-compliance than appointment non-compliance. There is a lot at stake when selecting study subjects because researchers are penalized when study subjects are "lost" or withdraw. Therefore, the researchers at the American private clinic chose patients who they thought were the most likely to show up for study visits. So patients with psycho-social problems (e.g. homelessness and drug addiction) were mostly excluded because these problems reduce the ability of people to comply with a protocol. Researchers avoided patients who were too much trouble. For instance, although one patient met the inclusion/exclusion criteria, his physician did not feel the patient was a good candidate for a study, in part, because his behavior bordered on assault and he had to have a "two person watch" when he was in the clinic.

Sometimes, the research team decided that study participation would be too much for the patient to cope with. One study nurse described a "screen failure" for a study of drug regimens in people first starting HIV treatment. The potential recruit had been diagnosed with HIV only a month before and was a wreck. He cried during their meeting. His partner had attempted suicide two days after he was diagnosed. The patient was unsure if he wanted this relationship of 16 years to continue. He was depressed and taking anti-depressants. One of his anti-depressants, Welbutrin, was considered a "precautionary" but not an "exclusionary" drug in the study protocol. She said that

psychiatrically, it might have been okay for him to participate, but psychologically he was not ready to participate in a research study. The patient told her that he would go off his anti-depressants if that was the problem. She said, no you can't just stop your medication. If you are going to stop, you need to go see your doctor who can gradually take you off and get you on something else. But she told him that he didn't need the extra strain right now of a research project. She called in the social worker right away. Then she took him to meet with the pharmacist. The pharmacist agreed with her that taking Welbutrin would put him at risk. The study nurse admitted that she was relieved that this guy was a screening failure because he was really "needy" (interview with study nurse US1 031204).

Good study participants are not only found but made. After patients are enrolled in a study, researchers put a lot of work into keeping a patient on study. The researchers at the American private clinic called study patients when they missed appointments. They even called relatives when a study patient was in danger of being dropped from the study. They provided funds for transportation and parking. When some study patients moved to another city, the principal investigator even offered to pay airfare for them to travel back to the clinic. And unlike the clinic nurses, the US study nurses were relatively tolerant of tardiness for study visits, sometimes waiting an hour for a study patient to arrive. *They need the study patient.*

Keeping study patients was even more work in Uganda, where transportation and telephone access is less reliable. HIV is a stigmatized disease, and the women and their children sometimes lost the support of the child's father. HIV positive mothers often hid their HIV status from family members and neighbors which not only complicated their

lives but the conduct of HIV care and research. Clinic and research visits were time consuming, and it was hard for the women to explain their absences to family members. Likewise, they had to hide that they were taking antiretrovrials while trying to remain adherent.[20] The clinic dispatched cars and health visitors when participants missed visits, parking the car far from the home and walking so that neighbors would not see the clinic's name on the car.

Just as researchers need study patients, patients need research. Part of study patients' compliance with the protocol arises from the fact that patients need the treatment provided through studies. Nevertheless, getting study patients into the clinic was sometimes difficult. Researchers reminded study patients that complying with the protocol was in their best interest. While I was observing a study nurse, she called a study patient to schedule a study visit. The study patient's HIV medications were not working. The study nurse told me that this patient is hard to schedule, canceling and rescheduling three times for each study visit. The study nurse speculates that the patient may be more likely to come in if she knows her medications are not working. During the phone call, the study nurse tries to strike a balance between not making the study patient overly anxious about her medications not working but concerned enough to follow the protocol (which required a blood draw within 30 days of her last blood draw.) The study nurse explained that as part of the study, an expensive genotype test is conducted on all study patients whose viral load gets above a certain level. The information from this test helps doctors choose a good drug regimen. Because of its expense, regular clinic patients would normally not have access to this test until they were more ill (US1 040217 JP). In other words, she subtly reminded the patient that she needs the study – or at least this test.

In the observed clinics, there was variation in the extent to which researchers controlled the input stream. The American researchers typically had more control over their input stream than did clinicians. This was especially pronounced in the private American clinic where the only requirement for becoming a patient was having insurance. A description of the patients encountered in an afternoon observing compared to typical study patients shows this difference. I observed a physician seeing seven regular clinic patients during two hours and 45 minutes in the HIV clinic. During these few hours, I observed a tearful young man with a cocaine problem whom the physician referred to social work; a patient with cancer experiencing complications from the radiation treatment whom the physician hospitalized; a patient who had been to the ER twice with heart palpitations which they suspected might be a side-effect of his HIV medications; an extremely weak man with advanced HIV, experiencing ear pain, having some trouble with his disability benefits and unsure if he wanted to take antiretrovirals; a patient with stress, paranoia, fatigue, bladder incontinence, and mouth injury; and even a transplant patient without HIV (US1 040429 JP). In other words, there was lots of variability and very little time. The physician rushed from patient to patient and worked on chart documentation in between patients. This is in contrast to the study visits in the American private clinic where the patients were mostly stable and for the most part, not obviously ill. Furthermore, the way that the studies were assigned reduced variability. Each study nurse was assigned a few protocols at a time. And typically, each study focused a specific category of patients – treatment naïve patients, people with dementia, patients with neuropathy, patients with high blood lipids, patients with lipoatrophy, etc. Of course, unexpected illnesses and side effects did occur, but pre-categorizing study

patients reduced some of this variability. Note however, that this is substituting one sort of complexity (complex procedures for interventions, data collection, data recording, etc.) for a different kind of complexity (variability in patients).

In the American private clinic, researchers had more control over uncertainty arising from social factors than did the clinicians. Researchers could choose only the most compliant and best behaved patients for their studies. They also had more time to spend on tracking and counseling study patients. This difference in control over uncertainty was less pronounced in the Ugandan clinic and the public American clinic. Both of these clinics had general routines for training patients and increasing compliance. They set up formal arrangements for training and counseling all of the patients before they ever received HIV medications. Furthermore, the healthcare providers in the American public clinic and the Ugandan clinic were more likely to withhold medications from non-compliant patients. This was possible because medications were paid for and provided by these organizations. The organizations had more control over the provision of drugs and worried about balancing scarce resources. In contrast, the American private clinic did not provide medications directly to patients; physicians only provided prescriptions to patients. Whichever pharmacy the patient used could withhold the medication because of a lack of supply or because of the patient's ability to pay, but this was beyond the province of the physicians. Physicians in the private clinic might encourage patients to wait until they were ready to adhere to an antiretroviral regimen, but they would be unlikely to refuse a prescription to a demanding patient. However, the private clinic sometimes did withhold refill prescriptions when patients missed their clinic appointments because a patient's HIV progression should be monitored at least

every 6 month according the American standard of HIV care. Likewise, the tight control of antiretrovirals in the other clinics was also a matter of providing quality care, but it was also a means of conserving resources.

Fitting the Protocol to the Clinic and the Clinic to the Protocol

Protocols do not always accurately describe the world of the local clinic. Making a protocol work requires "standard tinkering" – a back and forth between the clinical situation and the protocol. It has long been recognized that rules may be modified in the process of implementation (Blau, 1955). Indeed the literature on medical diffusion observes that guidelines are nearly always re-invented and simplified in the course of adoption (Berwick, 2003). The scholarship on standards at work focuses attention to the space between the formal rule and the rule as enacted. Drawing on the notion of the duality of structure and agency from Giddens (1984) and Bourdieu (1977), Felder and Pentland (2003) argue that routines are made up of two parts: the abstract, general "routine in principle" and the performative, situated "routine in practice." When the principle is put into practice, it is adapted to the needs of the concrete situation. In the process of enactment, the principle is potentially modified. Likewise, Reynaud (2005), argues that rules as formal, general statements, are incomplete, only becoming complete when they are enacted as routines. Rules "trigger an action with a certain degree of predictability, but do not overdetermine it" (850). Weick's (1998) analogy of musical improvisation is useful here, where the score is akin to a rule that is enacted through performance. Timmermans and Berg (1997) observe an "extended layer of

reverberations" that are not obvious from a reading of the protocol text and only become apparent when protocols are observed in action (282).

In order to make a protocol work, the protocol and the situation are revised and re-imagined through formal and informal means. Through a process similar to what Stinchcombe (2001) terms "formality in construction," protocols are formally revised to more accurately reflect the world they govern. At the same time, protocols and situations are fitted to each other in more informal ways. In particular, researchers engage in "category work" to make clinical events and the protocol supplied ways of categorizing events overlap.

Over the course of a study, protocols writers make changes to make the protocol better fit the situation. Indeed, one nurse described keeping up with all of the updates, amendments and clarifications for all of her studies as a "fulltime job" (US1 040115 JP). She showed me a clarification memo for of a study of pregnant women on which she had just started working. After some of the participating research clinics pointed out the difficulty of getting the women to come in for a study visit towards the end of their pregnancy, the protocol team expanded the visit window from plus or minus 7 days to plus or minus 14 days for those visits after 36 weeks pregnant.

In addition to clarification memos, protocol are explained and altered through updates, amendments, and even entirely new versions. Other questions are answered through emails with the protocol team. For example, the nurse preparing for the pregnancy study looked through her files for a clarification memo about which blood tests to prioritize. In addition to the usual CD4 count, viral load and blood chemistries required by HIV care and research, pregnant women undergo standard pregnancy-related

blood tests. The study nurse suspected that there would be times when the size of the blood draw needed for care-related tests purposes combined with the amount needed for research-specific blood tests would be excessive. If this occurred, she would have to leave off some of the research lab work, and she wanted to know which she should leave off. When she did not find anything about it in the memos, she said that she would email the protocol team about it (US1 040115 JP).

Figuring out what a protocol meant required lots of clarification and discussion among the research staff and between the research staff and the protocol team. And in the course of figuring out, the meaning of the protocol is re-made. For example, one study had lots of detailed rules about eligibility. One of the exclusion criteria precluded the participation of patients taking steroids. One potential participant said he or she took a drug called DHEA (dehydroepiandrosteron) which is a metabolic steroid and probably not the kind of steroid that the protocol team was worried about excluding.[21] The study nurse read up on DHEA, consulted the clinic pharmacist and emailed the protocol team. After some back and forth with the protocol team, they decided that the use of DHEA did not exclude participation in the study (US1 040112 CH).

When researchers worked to make the protocol better fit the situation, they are engaging in a kind of informality that is actually "formality being constructed." When the abstraction system is inadequate because they lack accuracy, economy, sufficiency or scope, users can either construct a better system or patch up the system to get the right solution for the specific case at hand (Stinchcombe, 2001). Researchers engaged in both of these strategies to make protocols work.

Protocols are more flexible than they first appear. Indeed, one investigator compared a protocol to a constitution. Researchers tinker with the stream of patient events and the protocol-provided boxes to make them overlap. For example, after all the work that goes into finding and training good study patients, the research team at the American private clinic engaged in category work to keep someone "on-study." Sometimes study patients have to stop taking a study drug because of side effects or treatment failure. If there is a second line of treatment built into the protocol, the patient can change to this regimen and stay "on study" and "on drug." Otherwise, the patient can become "on study" but "off drug." The worst case is for patients to be "lost to follow up" which occurs when a study patient simply stops arriving for study visits. Researchers develop strategies to avoid the label "lost to follow up." For instance, they can mark a certain number of visits as "missed" while they track down the patient and hope that the study patient shows up again. At the American private clinic, the research staff used information from regular clinic visits to complete CRFs. For example, the American private clinic had a study patient who permanently lives in an RV, stopping by the clinic irregularly for care as he passes through the city. The research team decided to tag his file so that when he does come in they can get the extra stuff they need from him for research. The study nurse had worried they would end up having to drop him because they could not get the data, but the study coordinator told her no, that they have a six month window for doing a study visit for this project, and in any case much better late than never. The study coordinator stressed that they have a lot of flexibility about these things (US1 031027 CH).

An additional example from the American private clinic further illustrates category work. A study patient became suicidal and depressed and was subsequently admitted to a psychiatric hospital. Because the study nurses lacked privileges in the psychiatric hospital, she could not conduct a study visit. This case was further complicated because the patient told the study nurse that he wanted to quit the study because he did not want to take antiretrovirals anymore. The research staff discussed this case during a research meeting. The study coordinator and the other study nurses encourage the study nurse assigned to this study not to drop him from the study yet. They tell her to define him as having a "missed visit" and then approach him again when he is more stable. They also propose that she do a study visit by taking information from his medical charts or by talking to him over the phone – there are all kinds of ways to do a visit without actually seeing the patient. This category work was not only about avoiding penalization. There is a moral commitment to study patients and a perception that patients benefit from the extra attention provided by study nurses. How strongly this commitment is felt varies by patient and by staff member. One study nurse said, "You don't give up on a study patient until they are in the ground."[22]

The Ugandan research team engaged in category work when they discussed what should count as an "overdose" and thus a protocol departure in the study of women and infants. The mothers in a study were supposed to give their babies 0.5 ml of liquid NVP for several weeks. Sometimes, the mothers gave two doses, if the baby spit out the first dose. Someone in the meeting said, if the mother gives two doses rather than one dose, then we have to report. But as they talk more, some admit that it seems harsh that it would be counted as a protocol departure. They agree that they need to think about the

implications of 0.5 ml variation as “overdosing” and that they also need to think about the implications of anything they do for other sites involved in this research which are located in other countries. They decide that this issue has to be taken up with the principal investigator (UG 050318 CH). Whether or not this counts as an overdose has fairly large implications for clinic work. According to the protocol, researchers must report overdoses and other significant protocol breaches to the local principal investigator, the US-based principal investigator, the DAIDS medical officer, the local IRB, the US IRB and the Ugandan government IRB.

Protocols are more likely to be inadequate in the Uganda than in America. The Ugandan clinic engaged in lots of standards tinkering to make the protocol fit local conditions and to avoid protocol departures. Problems arose because complications that are uncommon in the US are routine in Uganda. More importantly, handling these complications is more difficult in Uganda than in the US. The Ugandan clinics’ big study was a mother to child transmission study comparing Nevirapine alone to Nevirapine plus an immunoprophylaxis. The research team repeatedly had to report adverse events in the newborns in the study. Newborns in Uganda sometimes have grade 1 creatinine levels (a measure of kidney function) because the babies in Uganda are often dehydrated after birth. The attitude of the protocol team, which resides in the US, is “so record it as a grade 1 and follow up, no big deal.” However, it is a big deal because it requires the research staff to do a lot of extra work documenting, tracking, and doing blood tests for something that is normal in Uganda. Furthermore, all of this is more complicated in Uganda where the faxing, photocopying, and emailing required for reporting SAEs can be more difficult (UG 050124 CH&EW).

Similarly, problems arose when the women in the Ugandan study delivered their babies at home. As the Ugandan researchers reviewed a monitoring report, they noticed that the monitor reported that they had a number of "missed doses" – where the clinic did not administer the required drug to a few study patients. The missed doses occurred because some women delivered their babies at home and then came to the clinic a week later. By this time, it was too late to administer the dose of Nevirapine, which is supposed to be dispensed to the baby at delivery. The research team was supposed to have reported missed doses of study drugs to the study monitors. But a member of the research staff complained, if we report missed doses, we will be reporting all the time. In this case, the work of keeping the protocol working was too onerous, at least according to some of the local research staff. In the Ugandan context, a woman not coming to the hospital for delivery is seen as routine – just as forgetting to fast or missing study visits in late pregnancy is understandable to the American researchers. The Ugandan research team decided to ask about how to handle these missed doses at the next meeting with the regulators (UG 050318 CH).

Fitting the protocol to the clinic and the clinic to the protocol clearly required more work in the Ugandan clinic than in the US clinics. In the Ugandan setting, the distance between the protocol and the local work conditions required something more like remodeling than "tinkering." From the point of view of protocol writers, the extra work of fitting the event to the protocol and the protocol to the event was justified for unexpected events. For the most part, the protocol users agreed with this view. However, when what is an unexpected event for the protocol writers is a routine event for the protocol users, this agreement becomes more tenuous. It was not just that the

protocol did not cover unusual events; the protocol often did not cover routine events. It did not accurately reflect the world of the Ugandan clinic.

What is considered routine or normal is complicated in international research. What is expected in one setting (e.g. electricity outages, malfunctioning fax machines and spotty email access) is unexpected in another. Consider also the case of laboratory results, seemingly the most standardized part of medical care. Research protocols define when lab results should be considered adverse events. Lab results are assigned a severity level or "graded." There are two ways of assigning grades to laboratory results. One is an absolute number; for instance, a blood cholesterol level over 300 mg/dL is considered a Grade 3 (severe) toxicity according to the 2004 DAIDS toxicity table. The second way of determining the severity of side effects is as a factor of established normal ranges. For example, the 2004 DAIDS toxicity table defines a creatinine level that falls within 1.1 and 1.3 times the UNL (upper normal limit) as Grade 1 (mild) toxicity. In the DAIDS table, what counts as the upper normal level varies for age and sex but not for location. Nevertheless, as discussed above what counts as normal does in fact vary by location.

A general point here is that the more the material environment of the local clinic differs from the universal notion of the clinic inscribed in the research protocol, the more work is required to make research protocols doable. This is mostly, but not entirely, a matter of resources. Commentators on international research (see e.g. Resnik, 1998) have observed that US standards often do not translate well to international settings. Generally, it is the case that making a tool function is more work in resource poor clinics outside of the US. For instance, a pharmacologist described the extra work required to conduct a study of Kaletra, an antiretroviral, in Uganda. Kaletra requires refrigeration,

but most Ugandans do not have a refrigerator. The Ugandan researchers developed extra guidelines to help prevent drug spoilage on hot days. They instructed study subjects to submerge the pill bottles in water, occasionally changing the water. Note though, that extra work is not just undertaken by the researchers but by the study subjects as well.

However, the amount of work required to translate research recipes into local settings is not only a matter of resources; it is about the extent to which research recipes fit existing routines. Not surprisingly, it easier to follow a rule that says "do X" if you already do X. Similarly, the biggest predictor of whether a physician follows a guideline is the extent to which the guideline resembles their pre-existing practices. Because protocol teams are dominated by Western researchers, mandated research procedures are more likely to be those of developed countries. American researchers often find that they can continue with business as usual when researchers from resource constrained countries are likely to have to learn new procedures. The case of mother to child HIV transmission provided an important exception to this observation.

While the Ugandan clinic has far fewer resources than the American private clinic, they saw many more mothers and infants with HIV. This is because of differences in organizational mission as well as country-level variations in HIV demographics; unlike Uganda, mother-to-child transmission has increasingly become a rarity in the US. The Ugandan clinic is funded primarily through grants for mother to child transmission studies. Thus, mother to child transmission studies are far more routine in the Uganda clinic than they are in US. The sheer scale of their work required that the Ugandan clinic set up routines to handle things which the American private clinic handled on an ad hoc

basis. For example, the American private clinic conducted an observational study of pregnant women that required a blood draw within four days of birth a procedure that caused logistic problems. To the study nurses, asking a woman to come to the clinic for non-therapeutic reasons within four days of delivery, perhaps with her baby in tow, seemed unreasonable. The simplest way of ensuring this blood draw would have been to have the delivering physician order one at delivery, but the American private clinic had no formal way of requiring that the hospital physicians do so. They could only request that the obstetrician linked to the clinic do the blood draw whenever she delivered a study subject's baby. Otherwise, the nurse had to figure out some way of obtaining the test herself, and in some cases, the blood draw was just skipped. In contrast, the Ugandan clinic had closer ties with the physicians in the local hospital, even monitoring the extent to which the hospital physicians followed study protocols in some cases.

Protocol Intelligence: Improving the protocol and the researcher

During the course of a study, researchers develop protocol intelligence. This is a matter of learning the content of the protocol as well as drawing on and learning extra-protocol skills and knowledge that help the protocol work better. The need for protocol intelligence is exemplified by what a study nurse told me about how she handled work when she went on vacation. She left detailed written instructions so that another study nurse could conduct a study visit for her. If following the protocol were self-evident and did not require expertise, no additional instructions would have been necessary. As it turns out, the study nurse forgot a detail that was quite important from the study patient's point of view. When the substitute study nurse injected the study drug, she injected it too

fast, and as a result, the study patient experienced pain. The study nurse assigned to the protocol had already learned about this drug; she knew that slowly injecting the drug reduced pain (US1 040106 JP). Her improvisation improved the protocol.

In the course of working with a protocol, users develop expertise. Researchers become experts with the protocol content as they repeatedly encounter new events and work to make them fit the protocol. As protocol users learn more and more about the protocol, they learn what matters most. Importantly, they learn what kinds of mistakes can be made to fit the protocol and which cannot. For example, during a research meeting at the American private clinic, a study nurse raised a question about a patient from an ACTG study who was supposed to fast, was told he needed to fast, but came in for his study visit having eaten. The study nurse reports that he does not think this is a problem; he should just write in the CRF that the patient was told to fast but forgot. According to the study nurse, since the patient is not on the drug yet, there is not actually any consequence. The study coordinator concurs with the study nurse's conclusion that it's not a problem (US1 040105 CH).

In one case, the study nurses developed embodied expertise by actually submitting their own bodies to the protocol. When a large study added chest measurements, the nurses practiced on each other. Through this process, they learned how to do the measures but also how uncomfortable and embarrassing it was to be measured. According to a study nurse,

> It takes a long time in the beginning, but once you know, it goes a little smoother….You know, you are just trying to find the landmarks, and it was just something new. It just takes a little time. Some people are heavier than others

> and you just have to try to find landmarks, as we say. You know? We've got breasts that we need to work with if it's a woman. And, then there is modesty. With women, we have to have them take their bra off and everything, so there is an issue to make sure that they are comfortable. Men are more – that's not an issue, but with women, we ask, 'Do you want a blanket, a towel?' You have to ask if they are cold, you have to make sure the door is closed. Those are considerations that you need to take. (US1 040130 CH)

By actively taking up the protocol, the nurses made the protocol work better. If the study nurses had not learned how to make the measurement process more comfortable and less embarrassing, some study patients might have dropped out.

In addition to protocol expertise, the successful conduct of a study requires extra-protocol expertise such as how to make study subjects more comfortable by injecting a drug slowly or offering a blanket. Other expertise is of a more technical nature. In the course of doing a study, the protocol users learn likely drug side effects which means they know which patient complaints to pay careful attention to. Expertise makes for intelligent rule-following. Knowing why you are following the rules increases commitment to the rules, but also means you know what parts of the rules really matter and if and how errors can be corrected. This is similar to Tannenbaum's (1994) observation of the epistemological weakness of guidelines based on statistical evidence. Without an understanding of the mechanisms of the disease and the body, physicians would not know what do to if the probabilistic guideline does not work. Similarly, in the

example discussed below, knowledge of pharmacology improves the researchers' work when the protocol fails to fit the clinic.

During a national ACTG site meeting, a laboratory manager conducted training to improve a pharmacology study. The study required calculating patients' normalized inhibitory quotient (NIQ), which is the "increase needed of a specific antiretroviral drug required to stop HIV from growing in your blood." Calculating the NIQ requires a blood test; the blood has to be drawn just before the next dose is taken, to measure the level of the drug at its lowest. In the language of pharmacology, this is the "trough" that lies in between the "peaks" of drug levels. The blood has to be collected within a relatively small window of time and they need to know exactly when the last dose was taken. It is hard to meet these standards. The approach of the trainer seemed to be that if the study nurses know what kind of data they need and how mistakes affect the data, they will do a better job. The trainer gave several scenarios of how things go wrong in ACTG study 5146. For example, a nurse may correctly tell a patient not to take their pill until they come in for the visit, but they may not tell the patient that they should adjust their schedule so that their previous pill was taken 8 to 12 hours earlier. She said that this error occurs partly because the protocol is unclear. In this case, the patient is going too long between doses. In other cases, patients are taking doses too close together. Another error is that patients take their dose before coming in rather than waiting until after their blood draw. Some of the errors actually compromise patient safety – by causing their drug concentration to drop too low; all of the errors compromise the data. She showed the various ways that these protocol deviations shift and change the shape of the drug concentration curve (the peak and trough of the level of the drug in the blood). In order

to get an accurate concentration curve, the study nurses also need to write down exactly when the blood was drawn when it actually goes into the tube – not the time that the blood is scheduled to be drawn. She said that she is surprised from the reports how often blood draws occur exactly when they are scheduled – she knows this is not actually the case. She then showed that while documenting the ideal times does not change the curve that much, it does throw off the data a bit (US1 040518 JP).

Later, I talked to the pharmacist at the American private clinic about this training. The PI at the American private clinic had initially been reluctant to do this study because the research team had never done a pharmacology study before. He and the study coordinator had trouble deciphering the protocol. They did not know what a NIQ was. However, the pharmacist at the site was eager to do research and took on the study. I explained how the presenter gave different error scenarios and showed how it would affect the curve. The pharmacist said that was a really useful way of showing the nurses why it is important to follow the protocol. She said that she thinks that she understands 5146 better because of her training in pharmacology. She knows what is important. At the same time, because of her training, she knows that some things which are errors according to the protocol really do not matter. For example, the protocol requires that the blood draw occur between 8 and 12 hours after the patient took their last dose. So if the patient took his last dose at 10 PM, he has to have a blood draw between 6 and 10 AM the next morning. The pharmacist gave the example of case where the phlebotomist draws the blood at 10:02 AM. Two minutes is not a big deal but according to the protocol, she would have to redo the blood levels (US1 040525 JP).

Negotiating Competing Systems

Negotiation is central to all kinds of work. Changing organized work "involves a lot of convincing and persuading, buying and adopting, teaching and learning" (Fujimura, 1988, p. 261). Making a standard work requires the active participation of interested parties (Timmermans & Berg, 2003). In the case of research, interested parties include study patients, the nurses and physicians who conduct study visits and collect data as well as the nurses and physicians who provide non-research treatment. Other interested parties include study sponsors (especially protocol teams), pharmacists, data managers, technical personnel (e.g. phlebotomists, laboratory and x-ray technicians), grants administrators, billing specialists, as well as other administrative, regulatory and clinical personnel. Each of these jobs and their respective departments have their own routines and rule system. While it is the research staff that is charged with implementing a protocol, they cannot do it alone. In order to follow the protocol, researchers get others to sign on to doing new tasks or existing tasks in new ways.

To transform clinical information into data, researchers enlist the help of clinicians. The study nurses in the American private had to track down additional information from physicians to complete their paperwork. In the research meetings, research nurses often talked about the difficulty of getting proper documentation from caregivers who were not part of the core research team. Grading of symptoms seemed to be a particular sticking point. The study nurses are required to track the grade or severity of symptoms – whether nausea is mild or moderate, for example – a distinction which is not important to physician caregivers.

Researchers negotiate with clinical and non-clinical routines and rule systems. Clinical rules pertain to the core part of medical work and include codified treatment guidelines, un-codified but agreed upon standards of care, as well as the ethical principles underlying the medical profession (e.g. patient-doctor confidentiality, do no harm, etc.) Healthcare organizations are also permeated by numerous sets of non-clinical routines and rules systems that govern for example, billing, the movement of patients, lab specimens and paperwork around the hospital, etc. In this section, I examine how researchers' negotiate the demands of research with non-clinical, organizational rules and routines and then turn to their potentially more serious negotiations with clinical rules and routines. Finally, I observe that the boundary between the clinical and the non-clinical, the care-related and the organizational, is not always clear.

The non-research staff typically cooperates, but sometimes they make mistakes or otherwise prevent the protocol from working. Research is not their priority and they are governed by a different rule system. For example, the research staff at the clinic had been sending specimens on ice to the hospital lab for research-related lab tests. After they didn't receive the test results, the staff found out that the hospital lab had been throwing out the specimens! According to the lab's guidelines, this sort of specimen should not be on ice. As it turned out the research staff had misinterpreted the research protocol to mean that specimens *should* be sent to the lab on ice rather than *could* be sent on ice, depending on the local lab's guidelines (US1 030922 CH).

Both of the American clinics participated in a DAIDS-funded study of when to start HAART in patients with serious opportunistic infections. Patients who signed up for the study were randomly assigned to one of two comparison groups or "arms."

Patients in the first arm started HAART within two weeks of randomization. They were treated for their opportunistic infections at the same time. In the second arm, study patients first received treatment for their opportunistic infections and later started antiretroviral treatment. The study provided no free drugs. The requirement to start HAART within two weeks of randomization posed difficulties for the American public clinic. As part of belt-tightening measures arising from a serious budget shortfall, the public clinic established a policy that no patients (with some exceptions for emergencies) would start HAART until funding for their medications had been identified. Identifying funders could be a time consuming process because many of the public clinic patients had no insurance to cover the cost the HIV medications. Pharmacists and counselors met with and collected information from patients, then sought funds from the state drug assistance program and/or pharmaceutical companies' drug assistance programs. When study patients were assigned to the first arm of the study, staff scrambled to get medication funding within the two week cut-off. This was not the case in the American private clinic where nearly all of the patients were insured.

In the American private clinic, this study met few organizational barriers but one of the non-research physicians questioned the study design and raised a potentially more serious ethical problem – conflict between the research protocol and standard of care. During a physician meeting, when one of the physician researchers described this study, another physician expressed some surprise that they would wait to start HAART. At issue here is tension between collective and individual uncertainty.[23] The first, referred to as "clinical equipoise," is the ethical basis of medical research. Clinical equipoise refers to the existence of "professional disagreement among expert clinicians about the

preferred treatment" (Freedman, 1987). At the time of this study, there was clinical equipoise; among HIV specialists, there was no consensus whether to start HAART immediately or after treating opportunistic infections. However, from the point of view of some individual physicians, waiting to start HAART was the wrong (incorrect) thing to do, and for them, enrolling a patient in this study would be the wrong (unethical) thing to do. Ethical issues are further complicated and highlighted here because patients with opportunistic infections are often very sick. In fact, one of the patients that the American private clinic enrolled in the study died before he started HAART.

Beliefs about which arm of a controlled study is "best" for patients are considered "bias" and exactly the kind of thing that research protocols seek to exclude with randomization and other procedures.[24] A proponent of clinical equipoise, Freedman (1987) observes that physicians are licensed based on the acquisition of "professionally validated knowledge, not after they reveal a superior capacity for guessing." Likewise, researcher beliefs about whether a treatment works are inadequate indicators of whether a treatment is proven effective through statistical comparison. Indeed, during our fieldwork, one study was stopped by the safety monitoring board because there was no difference between the control and treatment group. Prior to this finding, the research nurse had been really eager to get patients enrolled in this study because she thought lots of study patients were benefiting from the treatment.

On the other hand, physicians, governed by the principle of "do no harm," are understandably reluctant to implement a research-prescribed treatment that they believe, even inaccurately, is harmful to their patient. The Ugandan research team experienced some difficulty with their study after some study participants experienced serious

symptoms that, according to the research team, were almost certainly not related to the study drug. Nevertheless, some of the physicians, who were charged with actually doing the infusions, viewed these symptoms as side effects and began to distrust the drug. As a result, they suspended infusions on a few newborns. The protocol allowed them to stop an infusion in cases of liver dysfunction. But the protocol was ambiguous; it did not clearly define liver dysfunction. The protocol writers intended liver dysfunction to mean grade 3 or 4 liver function tests but the doctors were stopping when there were grade 1 or 2 liver function tests. According to the physician-investigator, the physicians were being over-cautious and some of these suspensions were inappropriate (UG 050411 JP&CH).

To be clear, this was not a case of unethical behavior by the researcher or the physicians. Medical professionals disagree; the point is that when the professionals are engaged in work governed by different rule systems, disagreements take on a different character, becoming for example "inappropriate protocol departures" in the case of research.[25] Indeed, David Sackett, *the* proponent for evidence-based medicine, admits to "cheating" once when he gave a very sick study patient the experimental treatment that he believed was best for her – in conflict with the study protocol. He later described his actions as "right in particular, wrong in general" (Enkin, 2000 citing Peto & Baigent, 1998.)

When research protocols directly conflict with clinical care, the principles of ethical research as well as specific study protocols dictate that care trump the needs of research. Indeed much of the commentary and scholarship on medical research addresses this issue. This is particularly true for HIV research in the poor settings. In practice, these ethical issues are much murkier. In the end, physicians have authority over the care

of their patients who are also research subjects, but this authority is fairly rarely invoked. Outright conflicts between quality of care and research protocols were actually fairly uncommon. And when they did occur, they were far more subtle and complicated than the typical accounts of research ethics assume.

The distinction between what impinges on the core parts of clinical work and what is imply an organizational matter is not always clear. For example, a patient's physician and a study nurse discussed how to handle a pregnant woman's blood work in the American private clinic. Both the research project and the patient's physician required the results of her viral load tests.[26] Usually, the study nurse for the research project has the blood collected for the test, sends it to a university laboratory and then shares the results with the physician(s). In this case, the physician wants the viral load test performed in the hospital lab instead because it is faster. Doing two viral load tests at once is not an option. The amount of blood needed for two viral load tests along with all the other needed blood tests cannot be safely collected from a pregnant woman at one time. The physician needs the results of a viral load test to decide whether a C-section or a vaginal birth is safest. The woman is 30 weeks pregnant; the physician fears that she could go into labor before the university lab returns the results. However, the research project requires that all the viral load tests be performed at the same university lab to ensure that their data is comparable. After some discussion, the study nurse agrees to have the test performed at the hospital. A bit later, they figure out that they have gotten it wrong; the patient is less pregnant than they thought, so it would be safe to use the slower research-approved lab. Nevertheless, the study nurse does not alter the lab order (US1 040122 JP). Apparently, good working relations are more important than missed labs.

CONCLUSION

If protocols work, they must be actively taken up by researchers and made to work. Researchers use strategies to encourage protocol compliance and avoid protocol departures. They limit input uncertainty, engage in standard tinkering, develop protocol intelligence and negotiate with competing rule systems. Researchers control the uncertainty of their input streams by selecting and training patients who most fit the study subject imagined by the protocol. In formal and informal ways, they "tinker" with the protocol and the local situation to make them fit each other. In general, the Ugandan researcher engaged in a lot more "tinkering" to make the protocol of the large American sponsored study fit the Ugandan setting. In the course of doing research, researchers learn new skills and develop protocol expertise, which make the protocol work better. Conducting research relies on the cooperation of groups of people, including nurses, physicians, phlebotomists, and laboratory technicians. As such, following a protocol entails learning and negotiating different routines and rule systems. It also depends on the extent to which the other systems actually work. When the systems are dysfunctional or nonexistent, research projects either must be scrapped or researchers must re-work or build new organizational forms. The following chapter turns to these organizational processes of research.

Chapter 5. Administering The World: How Research Shapes Clinical Work[27]

In his work on quantification, Porter (1995) observes that science is as much about "administering the world" as understanding it. This is particularly true for modern medical research which often entails multi-center clinical trials and the coordination of groups across distances. Conducting clinical research is an organizational accomplishment with organizational effects; in the course of producing research results, clinics work out new routines, acquire equipment, adopt new technologies, re-train staff and even hire new staff. This chapter examines how clinical research "administers the world" of the HIV/AIDS clinic. It demonstrates that organizing the conduct of research – particularly the introduction of new jobs, technologies and standard operating procedures – potentially influence clinical work as much as the subsequent implementation of research results. An additional outcome of conducting research is institutional isomorphism; organizations become increasingly alike as they evolve to meet the requirements of a centralized set of regulators and funders.

The tie between research and medical practice is expected to be strong. The traditional model of clinical research includes a division of labor between researchers and clinicians and a more or less linear progression from identifying research questions to conducting a clinical trial to publishing research findings to modifying medical practice to take account of those research findings. In clinical trials, research questions are expected to be clearly oriented to the medical practice. Indeed, according to the Nuremberg Code, the Belmont Report, and the Declaration of Helsinki, it is not ethically acceptable to conduct clinical trials using human subjects unless there are clearly defined

clinical objectives. Once researchers are able to offer clear answers to scientific questions, healthcare providers are expected to modify medication regimens and therapies in light of the new information.

This chapter observes that the relationship between science and clinical practice differs from the traditional model of clinical research. In particular, the effect of medical research on practice diverges from this model in magnitude, timing and locus. The magnitude of the relationship between clinical practice and research varies; clinicians do not always adapt their practice to bring it into line with the latest scientific knowledge. This point was established some time ago in studies of diffusion of innovation. In attempting to explain why some people adopted innovations earlier than others, scholars acknowledged that there is nothing automatic about making use of new scientific knowledge. Rather, these studies pointed to the fundamentally social nature of the diffusion of innovation, with opinion leaders playing an important role in encouraging others to adopt new practices.

Although diffusion studies were important in uncovering variations in how rapidly innovations were adopted, the fundamental model linking scientific research to clinical practice remained intact. Changes in practices were assumed to *follow* scientific innovation. The argument here is that the original model was wrong not just in the assuming that changes would occur everywhere at the same speed, but more fundamentally that the *causal order* was wrong and the *locus of change* was wrong. Thus, the changes in clinical practice that are associated with research occur not just because of the *results* of the research but because of the *practice* of research. These changes occur long before final research results are available. In fact, these early changes

may smooth the way for adoption of clinical research because some part of the changes that would be necessary before scientific results could be incorporated into the ongoing activities of a clinic have already been made. Research in the medical literature supports the observation that the results of clinical trials are more likely to be diffused in clinics that conduct research. According to a study of prescription rates in New York state for the two years following the announcement of the results of Pediatric AIDS Clinical Trials Group the study that showed antiretrovirals reduced mother to child transmission of HIV, women who received care in hospitals that conducted federally-funded HIV research were twice as likely to be prescribed antiretrovirals (Turner, Newschaffer, Zhang, Fanning, & Hauck, 1999). Similarly, a recent study of the treatment of patients with acute coronary symptoms observed that hospitals that enrolled patients in clinical trials were more likely to follow treatment guidelines and even had a lower mortality rate than hospitals that enrolled no patients in clinical trials (Majumdar, Roe, Peterson, Chen, Gibler, & Armstrong, 2008).

The locus of change is broader than the outcomes usually addressed by studies of the diffusion of medical innovation. Evaluations of the impact of clinical trials are typically limited to the scientific portion of the research protocol, in particular the extent to which the hypotheses, whether supported or not supported, are translated into practice. However, in addition to drug regimens, research protocols concern matters such as what equipment is used, how work is divided among professional staff, how records are kept, and what kind of information is collected from patients. During the course of a clinical study, routines are worked out, skills learned, equipment acquired, and staff hired. Thus, research results enter a world that has been re-made in ways that ease their

implementation. As Timmermans and Berg (1997) observe in their research on insurance disability claims and cancer research, medical protocols are "the means through which facts can be *produced* and, at the same time, a crucial part of the networks through which the facts can be *performed*" (p. 296, italics in original).

This chapter examines the organizational processes through which the clinic is re-shaped by the conduct of research. The four HIV clinics in this study have useful characteristics for understanding the processes through which research affects the clinic. The two US clinics as well as the Ugandan clinic conducted multi-center clinical research funded by the US NIH's DAIDS research program. DAIDS funded research plays an important role in establishing the standards of research and care in HIV/AIDS. The South African clinic did less research than the other clinics discussed here. However, they were expanding to do more research which provided the opportunity to observe organizational learning. Conversely, the Ugandan clinic was in the process of expanding from a research clinic to a clinic that provided increasingly more care independent of its research programs. Thus, the Ugandan clinic allowed for the observation of organizational "unlearning" as they figured out how to scale back and translate some of their research-related services into a sustainable treatment program. While the clinical and research programs in the US clinics were more settled, they did start new studies; preparing for new studies often required the clinic to redeploy and retrain staff, procure equipment, institute new record-keeping routines as well as make arrangements for new treatment interventions. Importantly too, the US clinics were members of DAIDS ACTG which underwent an expansive re-organization and re-competition for funds during the course of the fieldwork. Along with producing lots of anxiety among the site

members, this led to a re-iteration and re-thinking of the sites' research arrangements.

How Research Reshapes the Clinic: Four General Processes

Much of the influence of clinical trials on the practice of medicine occurs while the research is being conducted rather than after it has been completed. Although research *findings* may influence medical practice, the *practice* of research also shapes the practice of medicine. Clinical trials shape medical practice largely by altering the organizations in which both medical treatment and clinical trials take place. Some of these changes are mandated by external entities such as the sponsors of clinical trials or the scientific protocol teams that design procedures to ensure the comparability of data collected in many sites. Other changes arise as researchers, caregivers, and administrators attempt to balance doing research and providing care under local constraints.

Marc Berg's (1997) work on the rationalization of medical work offers some interesting parallels and suggests mechanisms that might account for this broad impact of clinical research on medical practice. Although decision tools are intended to be universally applicable, in fact they are typically only very locally useable, and then only with a lot of work to fit them into the setting (Berg, 1997). Protocols rely on and motivate "infrastructural work" (Bowker, 1994). "Infrastructural work" refers to creating and supporting a network of relations among social, technical, economic, and political entities. In the course of fitting clinical research to the organization and the organization to clinical research, bureaucratic relations are reorganized, material environments are altered, priorities are changed with increasing value placed on reliability of measurement

and records, and variability is reduced as more and more aspects of clinic life are covered by standard operating procedures. Below I develop and illustrate these four general processes.

Reorganizing Bureaucratic Relations

Barley (1990) observes that "Technologies are depicted as implanting or removing skills much as a surgeon would insert a pacemaker or remove a gallbladder. Rarely however, is the process of technical change so tidy" (p. 67). In order to understand how technologies alter the social order, we must observe how technologies shape not only the user's activities but their relationships with others (Barley, 1990). The technology of research, including protocols, paper forms, equipment, and techniques, potentially alter skills and tasks as well as whom one interacts with and the frequency and content of those interactions. New technologies may be used to support or challenge existing bureaucratic hierarchies. This section examines the skills and tasks as well as the relational aspects of work are reorganized through the conduct of research.

In order to conduct clinical research, clinics make changes to their workload and work practices. New staff is hired and existing staff is retrained and redeployed. For example, both US clinics employed additional nurses to assist with the extra work of research. Likewise, the Ugandan clinic employed study-specific midwives in the labor and delivery ward in the hospital. The midwives assisted in the delivery of babies, identified study patients, and worked to ensure that the research protocol was followed and the appropriate research paperwork completed correctly. Furthermore, in recent years, research has led to the creation of wholly new job categories, including study

coordinator, data monitor and regulatory personnel. There is variation among the clinics in the extent to which new staff is hired and new research-specific job categories are created. In general, the ability to add new staff and job categories is constrained by resources, so the American clinics were more likely to have all of these new personnel – and thus likely to have an easier time following the rules of research. In contrast, the Ugandan and South African clinics had to hire additional personnel or add tasks to existing personnel.

The requirements of research confer new duties on clinical staff as nurses become "study nurses" and physicians become "physician-investigators." Study protocols often specify new diagnostic tests or treatments. Arranging for and performing such tasks requires individual and organizational learning which may last beyond the life of the clinical trial. Knowing what kinds of patients need a test as well as how to obtain a test increases the odds that the test will be ordered in the future. For example, the ACTG funded a study of bone density because it is suspected that HIV impacts bone density. In order to qualify for the study, patients had to have low bone density as measured by a DEXA scan. However, the ACTG would only pay for a limited number of these expensive DEXA scans. Because the study nurse needed to be especially selective in screening patients, she learned who is likely to have low bone density – skinny, inactive smokers. Furthermore, the nurse learned whom to call and how to send patients for DEXA scans, and the technicians performing the DEXA scan learned the method prescribed in the ACTG protocol.

A key obligation of research is documentation. Research documentation relies on good clinical documentation but has extra requirements. In the US, research staff were

allotted time and space for documentation. The staff at the Ugandan clinic talked quite often about the extra work of documentation required for research. This was even a source of tension between clinical and research staff; research staff complained that they were asked to work harder than care program staff. When the Ugandan management team discussed the distinction between research and program staff, the key difference was documentation, especially filling out the serious adverse event forms required for studies (UG 050302 CH). This training in adverse event reporting is crucial for participating in research and can even impact care. In a discussion of drug-induced adverse events, unrelated to research, a South African researcher noted that you get specific training in reporting adverse events if you are involved in clinical trials. But when you don't have that training, then follow-up often doesn't occur. Most people, with the exception of pharmacists, don't know where to get the forms, where to send them, etc. (SA 060330 CH). This is about learning the general procedures for reporting adverse events as well as taking on a general belief that tracking adverse events is important.

Research offers opportunities to learn new skills and is associated with a general upskilling among staff. This effect was particularly strong in resource-poor settings. For instance, the Ugandan clinic employed counselors to assist with conducting studies. The counselors had no prior formal medical education, but received training in general health skills such as monitoring infant growth, infant feeding, interpreting lab results, in addition to research-specific skills including obtaining consent and checking research forms. An exception to upskilling occurred in Uganda when medical professionals took on the considerable administrative tasks required for research without the support staff usually employed in the US, under conditions where even making copies and sending

faxes and emails could be time consuming because of technical difficulties and power outages.

Research obligations can conflict with established clinical obligations. Indeed, some general tenets of nursing training conflict with the demands of research. For example, filling out the case report forms for data collection purposes requires unlearning the nursing charge not to double-chart. A large part of the work of research entails recording study patient information into the medical chart and then translating a subset of this information into case report forms which are submitted to the study sponsor. Furthermore, data collection requires the removal of identifying and extraneous information. Depersonalization conflicts with the ethos of nursing. As one study nurse pointed out, nurses are charged all the way through their training with the obligation to think of "Mrs. Jones in room 235" rather than depersonalizing her (US1 040112 CH).

In order to understand how technologies alter the social order, we must observe how technologies shape not only the user's activities but their relationships with others. The non-relational parts of work (skills and tasks) nearly always spill over into the relational aspects of work (Barley, 1990). When a research protocol defines what kinds of tasks can be performed by certain kinds of staff, it potentially appropriates and reconfigures existing authority relations. The introduction of new tools (e.g. research protocols) can shore up authority relations through a process which Berg (1997) calls "reinforcing bureaucratic hierarchies." For example, while study protocols may alter the kinds of drugs and tests used, they largely leave intact the established jurisdiction of certain clinical duties; only certain categories of professionals are allowed to examine patients, prescribe drugs and dispense drugs. In such cases, the introduction of research

protocols supports the established medical hierarchy.

On the other hand, conducting research potentially disrupts other established authority relations. In particular, this occurs when local researchers, following research protocols, perform work that they usually would not be authorized to do. For example, protocols require specific tests be performed at particular study visits or in particular situations. This is similar to the practice of nurses following "standing orders" which are written by physicians. A simple example is the standing order to give Tylenol if a patient's oral temperature rises above 101 degrees Fahrenheit. The distinction between standing orders and research protocols is distance. Research protocols are written by experts residing in distant locales. Like the standing order, the authority of a research protocol formally arises from experts who wrote it. Nevertheless, because of the distance between the authors and implementers of the research protocol, the *apparent* authority of local implementers increases. Thus, established notions of medical authority are called into question in the implementation of research protocols when, for instance, when nurses are asked to do work that is usually performed by physicians. Note a change in local authority in the DEXA example discussed above – it is typically physicians and not nurses who decide which patients need particular tests. Order DEXA scan if the patient is a skinny, inactive, smoker is certainly not a standing order in the clinic.

Authority relations are also challenged by the additional layers of external and internal scrutiny introduced by research. In particular, medical decisions must be documented in extra detail for additional sets of eyes. Researchers regularly submit their work to the review of external study monitors. Appointed by study sponsors, the monitor's job is to verify that the rights and well-being of human subjects are protected,

that data accurate, complete and verifiable and that the trial is in compliance with the protocol and other regulatory requirements (International Conference on Harmonization, 1997). Study monitors do not actually watch research; rather, they visit clinics periodically and for the most part, review records. Thus, researchers worry a lot about documentation. Concerned that everything be done to the rigid specifications of the protocol, researchers intervene in the work of clinical staff.

Perhaps because of previous problems with documentation, the Ugandan researchers had developed a quality assessment/quality control (QA/QC) program that took them out of the research clinic to the dispersed sites (hospital, antenatal clinics, and infant intensive care unit) where some of the research data were being generated. Staff training programs, daily reviews of records as they were being produced, follow-up on errors, retraining staff on common errors, additional checks of a random sample each month, and unannounced site visits all reinforced a developing hierarchical relation in which research requirements dominated some parts of care giving.

The new program challenged authority relations because it required QA/QC staff check the work of physicians. The person charged with creating the program said that she had to overcome resistance to point out errors to superiors. She explained that there is a big professional gap between nurses and doctors in Uganda, and:

> A nurse will never tell the doctor, 'Well, what you wrote is not correct' or 'You haven't completed this portion'. Unlike in the U.S., where if you're a QC officer, you know that it doesn't matter who you are, even if you are the nurse, you have the right to go to the doctor and tell him to clean, or, you know, correct the research record. Here, it really isn't that way, you have to empower them with

> that ability to go and ask the doctor to correct something. (UG 050308 CH)

She noticed that when the reviewers filled out the new review forms, they left the part that asked about the contact person blank. The contact person was the person who was responsible for the error, and thus, the one who needed to be contacted in order to clear it up. They left out the contact information partly because they were worried about repercussions for themselves if the contact person was a doctor or their boss. However, they also were worried about what would happen to that person and looked at it as causing a problem for someone else. She said that the first thing we had to do is to change that mentality. One change that she implemented was to use employee ID numbers on the forms instead of names, in order to make it less personal.

The intervention of research staff with care giving was largely about the production of records, but there were also consultations about the care itself, particularly when local ways of proceeding would seem inappropriate to external referees reviewing how a serious adverse event (SAE) had been managed.[28] For example, the Ugandan clinic staff discussed how to handle their concerns over the hospital staff's admission and discharge decisions. They worried that sick people were being sent home. However, they expressed concern that sending people back to the hospital from the clinic would alienate hospital staff:

> One doctor worried that the hospital staff will think we don't trust them. Another doctor says that he has used the Ministry of Health guidelines a lot to defend himself, but he's "jittery about how we send patients back to the hospital." He worries the hospital will start to think that "here comes Dr. M to tell us that this patient is not ready to go home… our relations with them will deteriorate, and the

> patients will suffer." Dr. M talks about the practicalities of getting a patient readmitted and the delicacies of negotiating this with hospital staff – we don't really want a situation in which the hospital staff has the attitude of "can you tell us what you want us to do? We think the baby's fine and you assess it differently." No one in the hospital will want to work with us again if we don't deal with this carefully. (UG 050505 CH)

The Ugandan researchers intervened more in the work of care giving than did the American researchers, and the Ugandan caregivers were generally more willing to accept the authority of researchers than would be the case with US caregivers. When it was a matter of the production of records, Ugandan caregivers mostly seemed eager to get it right. Medical decisions were more of a sticking point as indicated by the above fieldnotes excerpt. While the Ugandan research staff worried about alienating hospital staff, they were, in fact, working out a plan for intervening when they disagreed with hospitalization decisions. The American research staff was unlikely to intervene with hospital admission and discharge decisions. This is partly because the standard of care expected by the researchers is similar to that of the hospital-based physicians in the US. However, the ability of the Ugandan researchers to institute routines in the hospital is also a matter of the authority structure of research in Uganda compared to the US. In both of the US clinics, it is nurses who generally implement the study protocol. It is hard for a nurse to force a physician to perform some test. In contrast, in Uganda, physicians conduct study visits under the instruction of the researchers. The Ugandan researchers are typically physicians who have some US or British medical training, which makes

them higher status than the average Ugandan physician who has the equivalent of an American bachelor degree. While caregivers operate in a local environment, research staff in an environment that combines local and international ties, so the difference in the attractiveness of the two activities is probably greater in Uganda than in the US. Thus, one of the reasons that Ugandan caregivers were relatively willing to alter their practices was that high status physician investigators were more involved in the day-to-day conduct of research. Furthermore, research is a source of funding in a resource-poor environment like Uganda. In the US, the infusion of research resources was less important.

Altering the Material Environment of the Clinic

Research re-shapes the clinic through the introduction of new tools and technologies. Indicators of change in the material environment include new or altered medications, equipment, paper or computerized forms, or space resulting from clinical research. What is new or altered for one of the research clinics might be routine for the others because access to the technology of HIV/AIDS care varied across the clinics. The tools and technology of HIV care are discussed in detail in Chapter 2. Both of the American sites had access to the full range of technology for HIV care. However, the American public clinic was more constrained in its use of expensive new antiretrovirals, because of government funding restrictions.[29] Resources were even scarcer in the African clinics, especially the Ugandan clinic where only CD4 and liver enzyme tests and a small set of antiretrovirals were routine. Below, I discuss the material effects of research in these different environments, beginning with the provision of HIV drugs.

Research often requires tests and treatments that are not yet part of the standard of care either because they are new or unproven or because of a lack of resources in the particular clinic. The imposition of study protocol-defined medication regimens is the most obvious technical effect of HIV/AIDS research. Prescribing is a privilege held nearly exclusively by physicians. Clinical research alters how prescribing decisions are made, and there is the potential for research protocols to conflict with clinical judgment. Research funders often provide antiretrovirals for free in the US and nearly always in poor countries, and participating in research is a means for patients to get free antiretrovirals in both wealthy and poor countries. However, the impact of the research protocol on medication is greatest in settings where access to expensive HIV medications is limited and research staff and participants have no options beyond study-provided antiretrovirals. In fact, it is unclear how voluntary research participation can be in areas where antiretrovirals are largely unavailable. Furthermore, in the Ugandan clinic there was increasing concern about providing antiretrovirals beyond the study, for the rest of the research participant's life.

Research projects require the use of medical equipment ranging from basic supplies to expensive machines. The project may expect a clinic to use its own equipment or provide equipment; they may require researchers use the equipment in study-specific ways in order to standardize results. In contrast to poor countries where medical equipment may be in short supply, research-provided equipment is often a duplicate of equipment already available in US clinics. Indeed, there was a lack of basic medical equipment in the Ugandan site. For example, there was only one working blood pressure machine on the adult wing of the clinic. One of the doctors had even used superglue to

repair one of the clinic's otoscopes (a medical device for looking inside ears.) Doctors were urged to share because there was not enough equipment for each doctor to have his/her own equipment or enough to locate them in each exam room (UG 050505 CH). Research funds were used to supply blood pressure kits for the maternity ward. Previously, blood pressures simply were not measured on the ward.

In resource poor places, an organization is less likely to be able to have material flexibility, so the mandates from a project will have a bigger influence on the rest of the organization. The impact of research resources on care is likely to be stronger in Uganda than in the US with a caveat for expensive equipment and tests. The Ugandan clinic would likely retain and use any equipment purchased through a research project for subsequent treatment or research projects. However, if a machine were to stop working after the end of the study in Uganda, it would likely remain broken. Because is hard to get machines repaired, there would be a reversion to older practices. And in fact, the maternity unit's blood pressure machine (mentioned above) was already broken at the time of our fieldwork. The other exception to the effect of resources are very expensive tests, such as genotype/phenotype tests for HIV drug resistance and more recently, the extremely expensive tests to determine whether a patient has a strain of HIV with CCR5 or CXCR4 receptors.[30] The genotypic and phenotypic drug resistance tests that were once only available for research are becoming increasingly routine for patient care in the US and Europe. While clinics in poor countries might collect blood for these tests for the purposes of research, they are likely to incorporate these tests as part of their clinical routine far more slowly than clinics in wealthy countries.

In addition to tests and treatments, paperwork is part of the material environment. In fact, documentation is the core task of research. Filling out case report forms using information culled from medical records or "source documents" is a central way that medical information is translated into scientific evidence. The amount of research forms (paper and/or computer-based) is voluminous. Research sponsors provide forms for informed consent, laboratory results, adverse event reporting, psycho-social surveys, etc. For each study, there is a different set of case report forms, and this set of forms must be completed for each subject in each research site. By the end of a study, this amounts to hundreds of pages per study patient. In addition to externally mandated forms, research sites typically create additional forms (e.g. checklists) for internal use to help them meet the obligations of research.

According to Berg (1997), "materializing a tool's demands" — or in this case, materializing research's demands — requires that those advocating the use of a tool change the physical environment so that "the decision technique becomes an unavoidable (and often unnoticed) part of daily practice" (93). Altering the environment to induce rule-following may occur in more or less explicit ways. When the Ugandan clinic instituted a quality assessment and control program to reduce the number of protocol violations and documentation errors, the reviewers were instructed to pay attention not just to documentation but also to the conditions that make it possible to do documentation. The QA/QC trainer mentioned a couple of examples:

> If people don't have any place to sit while they do their work, they are less likely to fill out forms correctly or to check them. If there is not clock, they can't document the time if they don't themselves own watches. If there is not a locking

> cabinet, they can't store their work securely. You can't tell people not to leave forms in the pocket of the binder (as they have) if you don't supply them with a hole punch so that they can prepare the forms to insert in the right sections of the binder. (UG 050316 CH&EW)

The QA/QC group was in the process of providing what equipment it could. For example, one of the QA/QC reviewers had taken around boxes of supplies. The group also proposed pressuring a doctor who has promised a desk and chair but had not yet come through (UG 050316 CH&EW). Where rule infractions were related to working conditions that could not be changed, they were to document the local working conditions for external monitors.

Altering clinical materials so that they overlap with the demands of research is another way of making the demands unavoidable and unnoticed. Take for example the case of the revision of the "subjective form" in the American private clinic. The form is filled out at every patient visit (regardless of whether the patient is a research subject) and becomes part of the medical chart. It is labeled as "subjective" because it contains patient reported information that is not easily measured, e.g. fatigue and nausea. The American private clinic revised its subjective form to fit the needs of an observational study in which over 500 of the clinics 2025 patients were enrolled (US1 040218 JP). When research forms are mandated, then it gets easier to standardize and the easiest way to standardize is in favor of the most rigid requirements. It probably truly does save time, but it surely also means that data are collected that are not needed and with a specificity that is not relevant for care giving. But recreating information after the fact from

inadequate records is such a burden that everyone buys into the lesser burden or tax of routinely collecting a bit extra.

In contrast to the subjective form, lots of the work of implementing research protocols is never made unavoidable and unnoticed. In fact, it was a lot of extra work for staff. A key role of the clinic nurse manager in the American private clinic was to ensure that research did not interfere with the work of the clinic. And in the Ugandan clinic, complaints that research was more work than care was a source of contention between the research and clinical staff. But following research protocols may be a step in making the implementation of subsequent research results and treatment guidelines easier and thus less noticeable and avoidable.

What is going on here is the construction of new knowledge and new routines about medications. By conducting research, physicians and nurses learn first hand about new drug regimens; they learn what side effects to be on the look out for and how to treat them, for instance. This point is similar to the observation that the provision of drug samples is successful marketing technique for pharmaceutical companies because it increases the likelihood that physicians will learn the dosage and side effects of their drugs, and thus, increases the likelihood that they will prescribe the drug (Abramson, 2008). Research is also an opportunity to create the organizational routines needed to use a new drug. For example, in the course of conducting a study of an immune booster that is infused (injected) in newborns, the Ugandan hospital had to train staff and set up new routines including moving newborns to the special care unit for infusion and hospitalizing these newborns for longer (48 hours as opposed to 24 hours). Thus, if there are positive study results for this immune booster, the Ugandan hospital will be better positioned to

implement the technique. Likewise, through participating in research, patients learn how to take new medications.

Prioritizing the Values of Research

During the course of a medical examination, individuals are subjected to the "clinical gaze;" observation, examination and documentation establishes the individual as an analyzable object, a "case" (Foucault, 1977). Medical research goes even further. Patients become data sources, and their symptoms and side effects are quantified and documented in uniform ways. In the course of documentation, personal facts and meaning are abstracted from their local, historical context (Smith, 1990). In order to produce data, individuals are made commensurable, which makes it possible to compare and connect disparate objects. This is particularly the case for multi-center clinical trials where the protocols are as much about insuring data uniformity as data quality. What is important is not so much "truth" as compatibility between different entities (Cambrosio et al., 2006).

Research protocols re-define what counts as relevant and trustworthy information and this is translated into organizational routines and embodied expertise of the clinic. As mentioned in chapter three, research funders provide technology and training to increase the objectivity and reliability of measures. This section extends this observation by discussing clinics' organizational responses to the requirements of objectivity and commensurability. In order to conduct research, clinics set up new practices. One response to expectations of research is increasing orientation toward quantification. Through a process Berg (1997) calls "reshuffling spokesmanship," quantitative is

privileged over qualitative data. An important example of this is grading, which entails translating symptoms into a numeric point severity scale.

Another response is the privileging of machine measurement. As Anspach (1988) observes, different levels of authority are attributed to technology, professional staff, and patients; the x-ray "shows" but the patient "alleges." The measurement of fat re-distribution, a side effect of antiretrovirals, elucidates this process. Fat redistribution (known as lipodystrophy or lypoatrophy in medical terminology) can radically alter a patient's appearance, resulting in sunken cheeks, extra fat on the neck and/or stomach, the "buffalo hump," or skinny legs. Patients report that they look different, but this subjective information is insufficient for the purposes of research. Instead, data collectors are trained in anthropometrics to measure the extent of fat loss and gain. Still, this method is suspect, with some preferring the use of expensive DEXA scans to measure changes in fat distribution. Similarly, a research funder gave the Ugandan clinic a blood pressure machine to eliminate the variation associated with the manual measurement of blood pressure.

Another response to expectations of research is the adoption of new styles of record keeping. The emphasis on comparability was especially salient in the South African clinic where the expectations of research had altered how they thought about medical records. They had recently begun systematizing their medical records partly in response to research. In an unexpected twist, this was not due to a specific research protocol. Instead, it was a response to the anticipated general requirements of research. According to the head physician, in the last couple of years they started thinking more about how their clinical records, which she referred to as "the data," might be useable for

other researchers. She described their records as a data "goldmine" for researchers (SA 060214 CH.) In other words, they re-conceptualized their patients' medical records as data from a longitudinal cohort study. This impacted how they collected information and the kinds of information they collected. For example, when they revised their clinical protocols in 2003, she realized that if they stopped doing some of the things they'd been doing before, they wouldn't have the baselines for comparison. So the question this raised for her was how to balance giving care efficiently versus maintaining some continuity with what they'd be doing in the past so that they could do comparisons.

None of the other clinics were re-conceptualizing their medical records quite to this extent. However, as discussed above, the American private HIV clinic had revised its subjective form to better match the needs of a large observational study. Part of the South African's re-conception of their records had to do with the computerization of their medical records. They were quiet proud of their system, and were working to improve it. Uniform, computerized record-keeping makes it easier for investigators to do observational research comparing the outcomes of similar groups of patients who received different treatments. Indeed, one of the motives of the general movement in the US toward computerizing medical records is to increase comparability and enable research. In fact, one of the physicians at American private was disappointed that the computerized record-keeping system at the hospital did not yet have the capability of uniformly recording blood test results, something that she had hoped to study. Uniform, computerized record keeping also allows hospitals to do internal quality assessment research. An example of this would include calculating the proportion of inpatients with pneumonia who received the appropriate antibiotic treatment as defined by national

guidelines, which allows hospitals to monitor their progress and compare themselves to other hospitals.

In the other cases, practices were developed for ensuring the comparability of records in response to the requirements of specific research projects and funding agencies. For example, the Ugandan health visitor manager described how her work increased and changed in the wake of some research-mandated changes in record-keeping. Based on the expectation that the data from one of the studies conducted at the Ugandan clinic would be used as part of an US Food and Drug Administration approval process, they had someone review the way they kept their records and tell them how to do things so that they would be useable in an FDA approval process. According to the health visitor manager, "before the FDA" they used cards, but now they use forms. They had to "legalize" everything, including even the writing. If someone scribbled, they have to cross it out, correct it, date it, and sign it. She showed examples. It is extra work for her, and she often has to take work home (UG 050218 CH&EW). US research staff also complained about these nit-picky research requirements. However, in the US, similar standards of record keeping are defined and enforced by non-research entities such as the Joint Commission. While existing rules raise the potential for conflicting standards, they also increase the odds that the organization can follow additional, similar rules.

One of the effects of multi-center research is the spread of Western expectations of medical record keeping. While participating in healthcare funding programs, such as PEPFAR, also increases record keeping duties, the evaluation for treatment programs is typically less deep than it is for research. Program funders certainly expect reports, which may be quite onerous, but unlike research monitors, they rarely review individual

medical records. Research encourages a legalistic orientation to medical record keeping because of the anticipation that outsiders will make judgments about work quality based on the record. This orientation is already common in the US where medical records are increasingly treated as legal documents that demonstrate that the proper care is provided in the proper way by the proper personnel (Berg & Bowker, 1997, citing Waters & Murphy, 1979). A key part of this is an expansion in the audience for medical records from local caregivers (e.g. in the same clinic or hospital) to outsiders who use the records for non-caregiving functions, such as justification for payment or evidence of malpractice.

The practice of recording medical information for sharing information between different locations is less routine in Uganda than in the US. Historians of medicine have observed that implementing a standard medical record in the early twentieth century in American hospitals demanded new infrastructure. In particular, it required constructing a network of doctors, administrators, records, and buildings (Timmermans & Berg, 2003). Similarly, maintaining research records requires a network of personnel, objects (e.g. binders and computers) and locations (e.g. laboratories, space for completing forms and data entry).

Standardizing HIV Research and Care

The past couple of decades have seen an increase in the number of multicenter clinical trials which require a collaborative effort among dispersed sites. Multicenter trials facilitate the collection of large, representative samples of study subjects. Such trials allow the collaboration of investigators to solve important medical problems.

Multicenter trials place a strong emphasis on standardization. Differences between research sites are minimized through training, certification, as well as the testing and re-testing of equipment. Another means for reducing variability is the use of specialized centers to perform certain technical functions, such as performing laboratory tests and reading x-rays (Friedman, Furberg, & DeMets, 1998). The central means of ensuring that similar procedures are performed on every subject at each site is the study protocol. A study protocol is a recipe for the actual conduct of the study, including a description of the inclusion and exclusion criteria, randomization procedures, the schedule of the intervention(s), and what kinds of tests should be performed.

Clinical research, particularly multi-center trials promote institutional isomorphism. In their classic article, DiMaggio and Powell (1983) argue that modern organizations become increasingly like other organizations as a result of institutional isomorphism. Institutional isomorphism is distinguished from competitive isomorphism as a motivator of bureaucratization and rationalization. In competitive isomorphism, organizations operating in a capitalistic market become more similar as they strive to increase efficiency. Increasingly however, the state and the professions are the source of much of the rationalization of modern organizations. Starr (1982) observes institutional isomorphism among American hospitals. There are a number of reasons for institutional isomorphism – federal funding of healthcare, centralization, increasing scale of healthcare organizations, demise of professional dominance of physicians (Scott, Ruef, Mendel, & Caronna, 2000). Medical research is another source of institutional isomorphism among healthcare organizations. Indeed, the last half of the twentieth century saw the rise of large research institutions – typically university-affiliated

hospitals in urban centers, largely as a result of the large federal influx of federal funding for medical research after WWII. Research institutions are a subfield, and perhaps a dominating subfield, within the American healthcare system.

Institutional isomorphism is a result of coercion, mimicry and normative pressure (DiMaggio & Powell, 1983). In some cases, organizational change is response to government mandate. For example, academic research centers set up institutional review boards (IRB) to comply with U.S. federal regulations. Far more numerous are the requirements imposed on research centers in exchange for federal funding. In fact, the US government funds the largest HIV clinical trial organization in the world. The DAIDS research networks are particularly important. Lowy (1996) observes that the diffusion of cooperative clinical trials of cancer treatment contributed to the standardization of diagnostic criteria and measures of therapeutic success. Likewise, the conduct of HIV research by networks sites contributes to the standardization of HIV research and care.

Large scale research, especially federally-funded research, creates and reinforces particular sets of organizational arrangements in healthcare. For example, ACTG research units must have ties to properly certified laboratory facilities capable of performing certain kinds of HIV-specific tests, e.g. HIV viral RNA, genotypic/phenotypic.[31] Another important way the ACTG impacts research clinics is through requirements regarding staff expertise. These are an example of what Timmermans and Berg call “design standards” which specify the membership of research teams and jurisdictional areas. For example, the ACTG requires that every research site have a social worker. Each ACTG unit must have an investigator who is an expert in

virology, immunology, pathogenesis of HIV disease, complications of HIV disease and its therapies, pharmacology, women's health or outcomes research as well as local expertise in surgery, neurology, gastroenterology, endocrinology, pulmonology, obstetrics or gynecology. Thus, a primary care practice is unlikely to be selected as an ACTG clinic. These requirements reinforce the treatment of HIV as an area of specialization rather than a primary care disease. Particular studies require additional expertise. These requirements can prevent some clinics from participating in research. In fact, the American public research team ruled out doing some research projects because they did not have a hepatologist (liver specialist) on site (US2 040114 RC).

Research participation promotes the creation and adoption of standard operating procedures (SOPs). According to the clinical director of the American public clinic, everything is very standardized in ACTG studies. There are SOPs for filling out case report forms and laboratory requests, for "quality assuring" (i.e. checking) forms. They have to use black ink. All the forms have to be signed and dated. Errors must be crossed out and dated. There are also SOPs for doing specific exams. For instance, there is a training video for assessing foot vibrations (US2 040106 RC). The orientation toward creating SOPs was particularly strong in the Ugandan clinic. The clinic conducted a big standard operating procedure (SOP) writing program associated with the launching of a large study, funded by the NIH and the Ugandan Ministry of Health. Once these templates were created they could be adopted outright by other projects or at least be a starting point so that SOPs need not be created anew with each research project. The imposition of SOPs can be experienced by clinics as a hassle or welcome subsidy. Where clinics are resource poor and infrastructure is not already in place, this subsidy will be

more happily accepted and also more influential because it is creating infrastructure rather than supplanting existing infrastructure. Differences between sites on this point are about differences in resources, clearly, but also about balance between research and care giving, the balance among large and small projects, and also the lifecycle of the organization.

Observing how standards are translated beyond academic centers provides an opportunity to see what kinds of relationships are taken for granted in clinical research generally and NIH-funded HIV/AIDS research specifically. In the recent grant re-competition, American private has applied to add two clinics to its ACTG clinical trials unit. One is an American private, outpatient clinic with no hospital or university affiliations. The lack of university and hospital affiliation means that the clinic must work harder to fit the definition of an ACTG site. The clinic must make changes in order to be selected as an ACTG site. For instance, they have to link up with an IRB and make arrangements to do research-specific lab tests. The other proposed site was in Africa and had even more work to do. The American study coordinator visited the African clinic to conduct a weeklong training in how to do research. As part of this training, the African staff practiced filling out case report forms. The material was created by a study nurse. She drafted some patient scenarios based on her experience with the ACTG 5095 study and provided copies on the 5095 case report forms, so that the African staff could practice entering patient problems into case report forms.

However, the research staff suggested that the proposed African site had a good start because it had received a grant from the US Comprehensive International Program of Research on AIDS (CIPRA). CIPRA provides a good example of coercive

institutional isomorphism. The US NIH implements CIPRA to fund research planning and infrastructure building in developing countries. According the CIPRA website, the NIAID provides oversight as a condition of funding:

> providing scientific consultation; (ii) advising on good clinical and laboratory practice; (iii) assisting with preparation of regulatory and ethical clearance documents required by US Federal regulations; and (iv) providing study monitoring, including data and safety monitoring for clinical trials. [32]

CIPRA staff also helps clinics establish ties with international organizations, pharmaceutical and biotech industry organizations, and with other US government agencies (e.g., FDA, USDA, CDC) so that the clinics can get supplies and services.[33] The American private clinic had contacts with an African clinic receiving CIPRA funds. They complained that this oversight can sometimes be onerous and inappropriate. The complaints mostly had to do with the American inability to recognize existing expertise. The clinic already had high standards in some areas.

It is not only clinics, but studies that become more alike. This is partly an outcome of mimicry as new studies are modeled on existing studies through the use of protocol templates. In fact, the study coordinator at the American private clinic used the ACTG protocol template to write the protocol for an investigator-initiated study sponsored by a drug company. During a meeting to discuss the project, the drug company representative offered a template created by the drug company, but the study coordinator opted to use the ACTG template which had used before (US1 040609 JP). This is a matter of efficiency and organizational learning. When new study protocols

look like previous protocols, clinics do not have to institute an entirely new set of practices each time they take on an ACTG protocol. The nurses in American private and American public clinic described ACTG protocols as having an initial steep learning curve that flattens out; once you have done one ACTG study, it is not difficult to do another.

Conclusion

Institutions set up different organizational arrangements that influence how research and care impact each other. In the four clinics where our team has done fieldwork, we observe variation in the overlap between research and care ranging from complete overlap to almost none. In the Ugandan clinic, essentially all care was initially provided through research. Therefore, clinical needs often had to be fitted to research protocols. Now grant-funded treatment "programs" exist alongside research in the Ugandan clinic. Treatment no longer needs to fit the strictures of the research program, but the constraints imposed by treatment programs are also rather severe.

In US private clinic, there is some overlap between research and care. Both take place with the same clinic, and the research nurses and physicians do both clinical and research work. Most of the research subjects are also clinic patients. When the study subjects require HIV treatment in addition to what the study provides, they usually get that care in the clinic paid for by their own insurance. The research staff consults with subjects' HIV care provider whenever changes must be made to their medications to encourage them to make changes that comply with the research protocol if at all possible. However, medical needs come first, and the research staff would not prevent a subject

from receiving medical treatment that they really need even if it means losing a study subject. In US public clinic, research and care are kept distinct. The research unit is housed within the same building as the clinic but they operate separately. In fact, during study visits, study subjects are not to be seen by their usual (primary care) physician. According to the researchers, the purpose of this policy is to prevent research subjects from confusing research with treatment. How the South African clinic was going to resolve the overlap between research was less clear because the clinic had not yet fully developed its research program. But it was clear that the dilemmas of simultaneous, sometimes conflicting, purposes were causing great concern.

Institutions can organize their research so that it has little impact on care giving, but perhaps they should not if we want clinical research to shape practices. Conducting clinical research is not only a means for testing new treatments, or a means for poor patients to get access to drugs and procedures but also a way of increasing the likelihood that new therapies, if and when appropriated, will fit local conditions. Clinical research may be as much about organizational innovation as technical innovation (Lowy, 1996).[34] New therapies enter into a "full world" (Lowy, 1996) of constraining routines, practices and knowledge. Translating clinical research into the "full world" of medical care involves complex articulation (Epstein, 1996; Rosengarten, 2004). It is not only after but during the conduct of research that this articulation occurs. Clinical trials shape and are shaped by the healthcare organizations within which they are conducted. Even before the results of clinical studies are announced, the full world of medicine has intruded on clinical research and vice versa.

If clinical trials shape organizational routines as much medical routines, this may

help explain why scientific knowledge penetrates non-research arenas only with difficulty. If one of the main routes of influence is the indirect one through organizational practices, then less elite organizations that do not also carry out research may be less able to adopt scientific knowledge because they lack the organizational and material culture that goes along with such scientific knowledge. The clinics that segregate research, and to an even greater extent, the clinics that do no research will have much more infrastructural work to do before new research results can be implemented. In the case of the American public clinic, the director had strong ties to the research community; he also required that the clinic staff follow treatment guidelines that are updated regularly. Staff was required to discuss certain treatment decisions in a group meeting. In other words, there were mechanisms in place for distributing and enforcing new tools of care. What the public clinic has to worry about with the implementation of new research results, was getting the other parts of their network, namely state funders, to keep up. For example, with each new drug or test, reporting forms and payment mechanisms are revised.[35]

Another useful comparison in this regard is a small non-research clinic which we visited in Uganda. The clinic first began by treating TB, then expanded to AIDS-related opportunistic infections and then added a few HIV treatment regimens. They did have some access to CD4 tests through another clinic, though they had to ration these. At the time of our fieldwork, they did not use viral load tests at all. The standard model of diffusion argues for the importance of "early adopters" in motivating others to take up a new technology. Perhaps, what is more important than adoption is network building. The Ugandan research clinic expanded the HIV care infrastructure. By purchasing viral

load testing through a local lab, the clinic was building infrastructure. In fact, as discussed above, the American government has a formal program for infrastructure building in poor settings.

How, when and if technologies impact the content of work is an empirical question. In order to meet research requirements, clinics create new routines, hire new staff, train existing staff, adopt new technologies and establish new relationships with technicians and experts. The extent and permanency of these effects varies. Some research procedures are learned only temporarily. For example, there is a specialized technique for taking blood pressure that is prescribed by some ACTG protocols. The study nurse laughingly discussed the technique. The protocol requires that the study patient raise his or her arms during some part of the procedure; that the nurse do a blood pressure test on both arms, document on which arm the measure is higher and then always do that arm first during future study visits; and that the study patient be left alone in the exam room for 5 minutes in between the measures on each arm so that he or she is relaxed. The especially funny part according to the nurses was that the patients keep both feet on the floor. They joked that they felt like Catholic school nuns yelling at students to keep two feet on the floor. From study nurses' laughter about the requirements, it is apparent that this procedure does not extend beyond the study. It is also apparent that procedures are temporary when they have to be looked up in the protocols.

While the effects of some protocols are fleeting, other protocols are built into the structure of the clinic. Sewell's (1992) essay on the duality of structure as schema and resource offers an interesting frame for thinking about the permanency of research requirements. The permanency of the organizational effects of research depends upon

the extent to which the protocols are supported by clinic resources. The norms of research such as the commitment to standardization, quantification, and even the rights of individual participants, are potentially ephemeral unless they are paired with resources. Schemas only become structural when they are supported by human or non-human resources. Once codified, schemas become resources. During the course of a research project, research protocols are resources for action. The staff has no reason to look to a research protocol once the study is finished. But the protocol has lasting effects when the norms and techniques of research are embodied in the clinic's paper and computerized forms, machines, and job positions. It is actually harder than many people realize to change forms and job categories, so once they have been altered, they are unlikely to be altered back. This is particularly true because of the semi-sacred status of research protocols and research results in medicine.

Producing and using new technology and knowledge requires the building of networks and infrastructure. Like locomotives, the results of medical research "do not work off their rails" and can only be "driven through a field" with much difficulty (Latour, 1983, p. 155). Latour's (1983) example is Louis Pasteur's work with the anthrax vaccine. Pasteur "raised the world" of the anthrax vaccine through the creation of laboratories and the enforcement of laboratory standards. Recognizing that networks are essential to the use of universal protocols, Timmermans and Berg (1997) question the extent to which networks are built anew and emphasize the incorporation and transformation of existing networks. Indeed, there are few "open fields" remaining in medical care and research. Medicine, especially the field of infectious diseases, is an area more or less covered with crisscrossing tracks – perhaps more so in the US than in

Africa (at least in regard to western medicine). Conducting research is likely to have its most lasting effects when the network of ties and infrastructure built and reconfigured in the course of doing a research project are later appropriated by subsequent research projects and by care programs.

Chapter 6. Standardization, Commensuration and Trust

The image of the lone scientist is an enduring but outdated model. Even more so in clinical research than in laboratory research, research is the accomplishment of dispersed groups with ties to different organizations. HIV/AIDS clinical research takes place in far flung research sites. Much of this research is collaborative (Cohen, 2000). Data is collected in bits and pieces from multiple clinics and sometimes even on different continents. Because the conduct of clinical research is collective and distributed, it raises issues of standardization, commensuration and trust.

All the components of evidence based medicine are part of a system of standardization, which links the production of objective evidence through clinical trials with material standards (e.g. biological indicators of disease progression, drugs, equipment) and procedural standards (e.g. study protocols, clinical practice guidelines, quality control standards, etc.) Medicine is increasingly standardized but clinical research requires even more standardization than care because of the demand for uniformity across patients and even clinics, in the case of multicenter research. Standardized and standardizing devices of clinical research include study protocols, predefined category or classification systems, computer and paper reporting forms, equipment for measuring symptoms and side effects, standard operating procedures, training and certification programs as well as other staffing requirements. These devices make up what Porter (1995) calls an “infrastructure of standardization.” In addition to rules and procedures for collecting and reporting data, researchers, funders and regulators set up systems for monitoring and enforcing rule following and data quality.

The rise of evidence-based medicine signals distrust in decision making based on medical expertise. Government funders, private insurers and patient groups have grown increasingly wary of expert discretion and demand objective verification. Most commentary about the loss of discretion associated with the imposition of a mechanical or rule-based objectivity has worried about this in regard to clinical practice. However, distrust in expertise pervades the conduct of clinical research. Clinical research is governed by layers of internal and external scrutiny, which shifts researcher's priorities to those activities which are most observable and documentable. All scientific activity is characterized by a "near-obsession" with inscription (Latour, 1983). This is especially true of clinical research, where hundreds of pages of forms are compiled for each study patient. One of the broad effects of clinical research is a quantitative and even legalistic orientation toward record keeping. This documentation goes beyond the sharing and proving of data that is common to the lab notebooks used by basic scientists; it is about subordinates producing information that withstands the audits conducted by study monitors (Heimer & Espeland, 2008). They worry about being judged compliant and trustworthy. Compliance does not necessarily ensure accurate data. Likewise, rule violations do not always imply poor data. In particular, the Ugandan researchers struggled with perceived and real distrust. The Ugandan researchers worried about how their decisions would be perceived by American researchers and regulators who were accustomed to a different medical system and a different standard of care.

An important technique for communicating in the absence of trust is quantification (Porter, 1995). Evidence-based medicine is an exemplar of modern rational decision making which requires precise comparison and commensuration, the

transformation of disparate information into a common scale (Espeland, 1997; Espeland & Stevens, 1998). Statistical comparison of medical treatments requires that their effects be expressed in a common metric. Using the DAIDS toxicity tables (DAIDS, 2004), very different signs and symptoms, such as cholesterol greater than 300 mg/dl, persistent nausea resulting in minimal eating of drinking for more than 48 hours, and difficulty sleeping causing inability to perform usual social and functional activities can be transformed into grade 3 cholesterol, grade 3 nausea and grade 3 insomnia. These can be further abstracted as simply three comparable grade 3 adverse events. Using standardized reporting forms, graded adverse events are compiled and compared across study groups becoming the results that are reported in journals and which may eventually make their way into treatment guidelines.

When commensuration works, its products appear to be rational reflections of the world. However, high cholesterol, nausea and insomnia are not obviously addable entities. Indeed, patients experiencing such symptoms are unlikely to perceive that they have anything in common, short of being ill. Even this shared sense of being ill is uncertain because the person with high cholesterol may not feel ill at all and insomnia is not always perceived as a medical problem. It is only through measuring, classifying, documenting, and reporting that the process and products of research become legible. In the course of making material legible, local information is lost. In his work on state map making, Scott (1998) makes a similar point about the cadastral map, which while precise, misses important local information (e.g. soil quality) that local owners know about a piece of land. Because of these unmeasured differences, seemingly equal plots may have very different agricultural yields.

Quantification is a special kind of inscription device for producing legible scientific findings. However, like cadastral maps, numbers obscure as much as they reveal. What is obscured is local information about the objects of quantification as well as the labor of quantification. Commensuration is not easily achieved. Researchers do not simply report the data. Fitting research categories and reporting requirements to the real world entailed interpretive work and organizational changes. Researchers work to figuring out what counts as an excluding criteria, an overdose, a missed dose, an adverse event, a missed study visit, etc. For example, a study nurse, pharmacist and protocol team decided that a steroid was not *really* a steroid according to the protocol; therefore, the patient was eligible for the study. The Ugandan research team worried about whether an extra 0.5 ml of medication counted as an "overdose." Categories have consequences. Patients who have characteristics of the exclusion criteria cannot be enrolled in the study. Adverse events, overdoses and missed doses are to be reported to a long list of internal and external regulators. Some kinds of adverse events require that patients stop taking the study drug. When researchers worry about whether this is the right category, they are also worrying about whether it leads to the right outcome for the patient and for the clinic. When patients miss study visits or worse are "lost to follow-up," clinics get in trouble with their sponsor, potentially losing money and reputation.

Requiring that workers fit events to a pre-existing set of categories has more impact on the content of work than category creators typically assume. Mary Ruggie's (1992) work on diagnostic related groups elucidates the unintended consequences of categories on a grand scale. In an attempt to minimize state intervention, the US government shifted from retrospective cost reimbursement to prospective fixed prices for

defined categories of illnesses, diagnostic related groups, which paradoxically increased state control (Ruggie, 1992). In addition to increasing interpretive work, research categories beget new procedures for categorizing. Transforming patients into trustworthy and commensurate data depends on rules and standardization.

Conducting medical statistical research entails more rule-following than many people realize. The scholarship on medical discretion and autonomy in the wake of the standardization has missed this piece of the puzzle. Every clinical trial is governed by a specific research protocol. Research protocols are highly prescriptive, procedural standards. Research protocols define who is eligible to participate in the trial, which procedures and medications should be administrated, as well as when, how and by whom those procedures and medications should be administered. They also specify how to document research-related decisions and activities in standardized ways. Just as quantification is more than simply reporting numbers; implementing study protocols is more complex than simply following rules. While the goal of clinical trials is uniformity, researchers achieve a limited and local universalism.

Standards are always insufficient and inaccurate to a certain degree. Inaccuracy and insufficiency are greater when standards created in one setting are diffused to another. In a 2000 letter to world leaders defending his contrarian position on the relationship between HIV and AIDS, South African president, Thabo Mbeki, wrote "It is obvious that whatever lessons we have to and may draw from the West about the grave issue of HIV/AIDS, a simple superimposition of Western experience on African reality would be absurd and illogical" (Cohen, 2000). While the international HIV/AIDS community is critical of President Mbeki's position on the relationship between HIV and

AIDS, few would dispute his point about "simple superimposition." When research programs are diffused from one place to another, a lot of work is needed to conduct the research program because research clinics in the US and Africa reside in very different healthcare systems. What is normal and routine in one setting is abnormal and exceptional in another.

Some commentators worry that clinical trials are irrelevant to practicing physicians (Glasgow, Magid, Ritzwoller & Estabrooks, 2005) and that this is one reason for the gap between evidence-based recommendations and clinical practice identified in the Institute of Medicine's (2001) report, *Crossing the Quality Chasm.* According to other critics, "Randomized trials, especially if associated with complex and strict protocols and many exclusion criteria, often give us the ability to know much but only about a world that does not exist." (Bellomo & Bagshaw, 2006). Clinical trials emphasize efficacy at the expense of other dimensions such as patient applicability and organizational adoption (Glasgow, et al., 2005). The publicized results of clinical trials are "simplifications" (Star, 1983) which obscure patient and organizational information but the process of clinical research is very much engaged with such issues. A closer look at the day to day conduct of clinical research reveals that the outcome of clinical research is as much about organizational innovation as technical efficacy.

The world imagined by complex and strict protocols does not actually exist, but parts of this world are in fact brought into existence in the course of a clinical trial when researchers alter their world in lots of ways that makes the protocol easier to follow. Conducting clinical research entails "infrastructural work" (Bowker, 1994), the creation of a network of relations among social, technical, economic, and political entities.

Infrastructure is not built anew with each research project. Conducting medical statistical research relies on an infrastructure that is common to American academic medical centers, such as standard record keeping, medical equipment ranging from disposable syringes to stethoscopes to X-ray machines, certified laboratories, trained laboratory and technical staff, physician specialists, and a local IRB. When clinical research is conducted in settings lacking infrastructure typical of an American medical center, it is harder to fit the rules of research to the setting.

From an ethical standpoint, clinical research raises the "science care dilemma" (Fox, 1974), the concern that research protocols will conflict with medical standards of care. However, in the clinics observed in this dissertation, outright science care dilemmas were very rare. This is not to say that clinical research does not impact medical work. Rather, clinical research intervenes most with the seemingly non-medical parts of the clinic. Staff are hired and re-trained; routines are altered; new ties are established with other departments and organizations. It is through these relational and organizational processes that medical practices may finally be altered.

The Diversity of Medical Science

One of the central questions in the sociology of science concerns how science impacts other institutions. The relationship between medicine and science has long been a concern of the medical profession as well as social and historical scholarship on medicine. This dissertation problematizes the relationship between science and medicine in two ways. First, it observes that physicians draw on different kinds of scientific knowledge in different ways. In particular, it addresses the interpretation of statistical

medical research favored by Evidence-Based Medicine and pathophysiological bench research. Secondly, it distinguishes between the different components or tasks of evidence-based medicine – conducting medical statistical research, interpreting research and the guidelines that translate research into rules for care, and argues that each of these intervenes on medical work in different ways.

At issue here are broader questions about the governing of professional activities. How do professionals decide what to do? And on what grounds? Proponents of evidence-based medicine answer this question, in an apparently simply manner. Physicians should base decisions on the "best available external evidence from systematic research" using their clinical experience and expertise (Sackett, 1996, p. 71). The evidence is expected to be epidemiological. Evidence-based medicine defines a hierarchy of what is "best" which ranges from randomized clinical trials at the top to observational studies in the middle to expert opinion at the bottom. While experience and expertise are relegated to the residual – that which cannot be or has not been subjected to statistical comparison, pathophysiological bench science is not even in this hierarchy. In fact, the most lasting effect of evidence-based medicine may not be the clinical practice guidelines that everyone worries about but the shift from pathophysiological to statistical knowledge base (Timmermans & Kolker, 2005).

Pathophysiological bench research and clinical epidemiological research favored by evidence-based medicine imply very different models of medical decision making. Statistical medical research typically addresses *what* works as opposed to pathophysiological research which sheds light on *why* and *how* diseases progress and treatments work. This distinction has serious consequences. If you do not know why

something works, you are at a loss if it does not work (Tannenbaum, 1994). Statistics are after all about probability, about increasing the odds that you get it right more often than not, which means that you do occasionally get it wrong if you base decisions on statistics. This distinction between how things work and what works translates into a division of labor in some fields of applied science. For example, knowing how and why some kinds of concrete work better than others is the province of chemical engineers, while contractors simply need to know what kind of concrete to use (Stinchcombe, 2001). In contrast, most physicians are not prepared to cede control over knowledge of how and why to a medical research elite. Bench science played a larger role in the infectious disease department's journal club than many would expect. A substantial minority of the articles discussed by the infectious disease department journal club were basic science articles (14 out of 65), and the group drew on knowledge of disease mechanisms in judging whether statistical evidence made sense. The group judged medical statistical research according to the standards of research design, the established pathophysiological professional knowledge system and clinical experience.

Science is not just a product but an activity. Understanding the conduct of scientific research can help us understand scientific knowledge and vice versa (Pickering, 1992). In day-to-day medicine, conducting clinical research, interpreting published research results and translating research results into rules for care are apparently independent tasks performed by different groups. Physicians read and judge the medical literature. Research teams made up of medical and non-medical personnel work to implement study protocols. Expert panels meet to devise rules for the conduct of care. However, they are in fact interdependent processes. The importation of science-based

standards into medical care relies on huge increases in the amount of medical research, particularly the proliferation of new kinds of clinical research which are more directly applicable to clinical care. These new forms of medical research are more applicable to clinical care precisely because the studies take place in the clinic, conducted by practicing physicians on actual patients.

While inter-related, the components of the system of evidence-based medicine do not intervene equally in medical work. Scholarship on evidence-based medicine and the medical profession has attended to the potentially standardizing effects of clinical practice guidelines. In particular, commentators have worried about a loss of professional autonomy and individual discretion. It is not only at the consumption end of medical research that this occurs. In addition to published research results and clinical practice guidelines, the actual conduct of clinical research, is a third, often overlooked, route through which scientific evidence impacts medical care. In fact, the rules for conducting medical statistical research actually intervene more directly with care than the guidelines based on research results. Physicians may be guarding the front door of the clinic while medical research enters through the back door, impacting the content and organization of medical work in unexpected ways.

TABLES

Table 1: Overview of Fieldwork

	US Private Clinic (US1)	**US Public Clinic (US2)**	**Uganda clinic (UG)**	**South Africa Clinic (SA)**
Duration of fieldwork	23 months, on and off between 2003 and 2005; follow up visit in 2007	6 months in 2004; follow up visit in 2007	4 months in 2005; follow up visit in 2007	4 months in 2006; follow up visit in 2007
Fieldworkers	Carol Heimer Juleigh Petty Site visits by Lynn Gazley	Rebecca Culyba Site visits by Carol Heimer, JuLeigh Petty	Carol Heimer Enid Wamani Site visits by JuLeigh Petty, Wendy Espeland	Carol Heimer Site visits by JuLeigh Petty, Lynn Gazley, Alan Czaplicki

Guideline Tables

Table 2: Recommendations by Evidence Type[36]

	Clinical Trial		Expert Opinion		Total
Jun-98	15	46%	18	55%	33
Dec-98	35	66%	18	34%	53
May-99	37	69%	17	32%	54
Jan-00	21	45%	26	55%	47
Feb-01	23	47%	26	53%	49
Apr-01	23	47%	26	53%	49
Aug-01	23	47%	26	53%	49
Feb-02	20	44%	26	57%	46
Jul-03	60	61%	38	39%	98
Nov-03	62	60%	41	40%	103
Mar-04	64	61%	41	39%	105
Oct-04	140	63%	82	37%	222
Apr-05	147	64%	82	36%	229
Oct-05	161	65%	85	35%	246
May-06	160	62%	98	38%	258
Total	991	60%	650	40%	1641

Table 3: Recommendations by Degree of Certainty

	High		Mod		Low		Total
Jun-98	18	55%	10	30%	5	15%	33
Dec-98	29	42%	17	42%	7	16%	53
May-99	27	50%	19	35%	8	15%	54
Jan-00	17	36%	21	45%	9	19%	47
Feb-01	18	37%	22	45%	9	18%	49
Apr-01	18	37%	22	45%	9	18%	49
Aug-01	18	37%	22	45%	9	18%	49
Feb-02	16	34%	21	46%	9	20%	46
Jul-03	43	44%	28	29%	27	28%	98
Nov-03	40	39%	36	35%	27	26%	103
Mar-04	38	36%	41	39%	26	25%	105
Oct-04	86	39%	97	44%	39	18%	222
Apr-05	86	38%	104	45%	39	17%	229
Oct-05	98	40%	109	44%	39	16%	246
May-06	97	38%	118	46%	43	17%	258
Total	639	39%	689	42%	305	19%	1641

Table 4: Correlation between Type of Evidence and Certainty, June 1998

	Expert Opinion		Scientific Evidence		Total	
	N	%	N	%	N	%
High	6	33	12	80	18	55
Moderate	8	44	2	13	10	30
Low	4	22	1	7	5	15
Total	18		15		33	

CHISQ= 7.19, $p < .05$

Table 5: Correlation between Type of Evidence and Certainty, December 1998

	Expert Opinion		Scientific Evidence		Total	
	N	%	N	%	N	%
High	6	33	23	66	29	55
Moderate	8	44	9	26	17	32
Low	4	22	3	9	7	13
Total	18		35		53	

CHISQ= 5.25, $p > .05$

Table 6. Correlation between Type of Evidence and Certainty, May 1999

	Expert Opinion		Scientific Evidence		Total	
	N	%	N	%	N	%
High	4	24	23	62	27	50
Moderate	8	47	11	30	19	35
Low	5	29	3	8	8	15
Total	17		37		54	

CHISQ= 8.04, $p < .025$

Table 7. Correlation between Type of Evidence and Certainty, January 2000

	Expert Opinion		Scientific Evidence		Total	
	N	%	N	%	N	%
High	5	19	12	57	17	36
Moderate	14	54	7	33	21	45
Low	7	27	2	10	9	19
Total	26		21		47	

CHISQ= 7.55, $p < 0.025$

Table 8. Correlation between Type of Evidence and Certainty, February 2001

	Expert Opinion		Scientific Evidence		Total	
	N	%	N	%	N	%
High	5	20	13	54	18	37
Moderate	13	52	9	38	22	45
Low	7	28	2	8	9	18
Total	25		24		49	

CHISQ= 7.043, $p < 0.05$

Table 9. Correlation between Type of Evidence and Certainty, April 2001

	Expert Opinion		Scientific Evidence		Total	
	N	%	N	%	N	%
High	5	20	13	54	18	37
Moderate	13	52	9	38	22	45
Low	7	28	2	8	9	18
Total	25		24		49	

CHISQ= 7.043, $p < 0.05$

Table 10. Correlation between Type of Evidence and Certainty, August 2001

	Expert Opinion		Scientific Evidence		Total	
	N	%	N	%	N	%
High	5	20	13	54	18	37
Moderate	13	52	9	38	22	45
Low	7	28	2	8	9	18
Total	25		24		49	

CHISQ= 7.043, $p < 0.05$

Table 11. Correlation between Type of Evidence and Certainty, February 2002

	Expert Opinion		Scientific Evidence		Total	
	N	%	N	%	N	%
High	5	20	11	52	16	35
Moderate	13	52	8	38	21	46
Low	7	28	2	10	9	20
Total	25		21		46	

CHISQ= 5.915, $p > 0.05$

Table 12. Correlation between Type of Evidence and Certainty, July 2003

	Expert Opinion		Scientific Evidence		Total	
	N	%	N	%	N	%
High	10	26	33	55	43	44
Moderate	10	26	18	30	28	29
Low	18	47	9	15	27	28
Total	38		60		98	

CHISQ= 13.321, $p < 0.001$

Table 13. Correlation between Type of Evidence and Certainty, November 2003

	Expert Opinion		Scientific Evidence		Total	
	N	%	N	%	N	%
High	10	24	30	48	40	39
Moderate	13	32	23	37	36	35
Low	18	44	9	15	27	26
Total	41		62		103	

CHISQ= 11.995, $p = 0.002$

Table 14. Correlation between Type of Evidence and Certainty, March 2004

	Expert Opinion		Scientific Evidence		Total	
	N	%	N	%	N	%
High	9	22	29	45	38	36
Moderate	14	34	27	42	41	39
Low	18	44	8	13	26	25
Total	41		64		105	

CHISQ= 14.135, $p < 0.001$

Table 15. Correlation between Type of Evidence and Certainty, October 2004

	Expert Opinion		Scientific Evidence		Total	
	N	%	N	%	N	%
High	28	34	58	41	86	39
Moderate	29	35	68	49	97	44
Low	25	31	14	10	39	18
Total	82		140		222	

CHISQ= 15.13, $p < .001$

Table 16. Correlation between Type of Evidence and Certainty, April 2005

	Expert Opinion		Scientific Evidence		Total	
	N	%	N	%	N	%
High	28	34	58	40	86	38
Moderate	29	35	75	51	104	45
Low	25	30	14	10	39	17
Total	82		147		229	

CHISQ= 16.82, $p < .001$

Table 17. Correlation between Type of Evidence and Certainty, October 2005

	Expert Opinion		Scientific Evidence		Total	
	N	%	N	%	N	%
High	28	33	70	44	98	40
Moderate	32	38	77	48	109	44
Low	25	29	14	9	39	16
Total	85		161		246	

CHISQ= 17.91, $p <.001$

Table 18. Correlation between Type of Evidence and Certainty, May 2006

	Expert Opinion		Scientific Evidence		Total	
	N	%	N	%	N	%
High	27	28	70	44	97	38
Moderate	42	43	76	48	118	46
Low	29	30	14	9	43	17
Total	98		160		258	

CHISQ=20.37, $p < .001$

Journal Club Tables

Table 19. Types of Journal Club Articles

Type	N
Epidemiological:	
Clinical Research	20
Observational Research	17
Meta-analysis	2
Basic Research	14
Other Research	8
Case Report/series	6
Total	65

Table 20. Journal Club: Type of Evidence by Judged Quality

	Good	Ok/mix	Weak	ND	Total
Epidemiological:					
Clinical Research	3	6	1	10	20
Observational Research	0	4	5	8	17
Meta-analysis	1	0	1	0	2
Basic Research	2	0	2	10	14
Other Research	1	1	0	4	6
Case Report/series	0	0	0	6	6
Total	7	11	9	38	65

Table 21. Journal Club: Type of Evidence by Certainty of Implementation

	Strong (+)	Moderate (+)	Weak	Moderate (-)	Strong (-)	ND	Total
Epidemiological:							
Clinical Research	1	7	5	3	1	3	20
Observational Research	0	4	4	4	0	5	17
Meta-analysis	0	1	0	1	0	0	2
Basic Research	0	0	2	0	0	11	14
Other Research	0	5	1	2	0	1	8
Case Report/series	0	3	0	0	0	3	6
Total	1	19	12	9	1	23	65

Table 22. Journal Club: Epidemiological and Experiential Evidence by Certainty of Implementation

	Strong (+)	Moderate (+)	Weak	Moderate (-)	Strong (-)	ND	Total
Epidemiological	1 (3%)	12 (31%)	9 (23%)	8 (21%)	1 (3%)	8 (21%)	39
Case	0 (0%)	3 (50%)	0 (0%)	0 (0%)	0 (0%)	3 (50%)	6
Total	1 (2%)	15 (33%)	9 (20%)	8 (18%)	1 (2%)	11 (24%)	45

APPENDIX 1

Excerpt from ACTG 5095 Protocol
STEP 2: Open-Label Treatment

ARM D	2 or 3 NRTIs + 600 mg Efavirenz (once a day) [1]
ARM E	2 or 3 NRTIs + ATV 400 mg (once a day) [2,3] or 2 or 3 NRTIs + ATV 300 mg (once a day) + RTV 100 mg (once a day) [3,4]
ARM F	2 or 3 NRTIs + TDF 300 mg (once a day) [5]

[1]NVP may be substituted for EFV if NVP substitution occurred on Step 1 or **if** treatment-limiting intolerance to EFV **occurs** on Step 2. See the precaution in Step 1.
[2]ATV and ddI must be taken at least 1 hour apart.
[3]**TDF is not allowed to be coadministered unboosted ATV**.
[4]**RTV is not supplied by the study.**
[5]TDF is not supplied by the study for Step 2. When used with TDF, ddI should be dose-reduced to 250 mg QD. **The 3-drug combinations of ddI+3TC+TDF or ABC+3TC+TDF should not be used.**

Based on their interpretation of the genotypic resistance report, the subjects and their physicians select their regimens based on the following rules:

- Each regimen must include two or three NRTIs, including ZDV, ddI, 3TC, d4T, or ABC (all supplied by the study) individually or in fixed-dose combinations, and/or TDF (not supplied by the study) with the following exceptions:
  ZDV/d4T is not permitted for anyone

- If virus is NNRTI "sensitive," defined as the absence of resistance mutations posted on the protocol-specific web page (http://aactg.s-3.com/members/ps/5095/ps5095.htm), subjects are eligible to choose Arm D, Arm E, or Arm F (see above).

- If virus is NNRTI "resistant," defined as the presence of any of the resistance mutations posted on the protocol-specific web page (http://aactg.s-3.com/members/ps/5095/ps5095.htm), subjects are eligible for Arm E or Arm F only.

- **ATV,** EFV, and the combination of ddI/d4T are not permitted in pregnant women.

An EKG is recommended for subjects who select Arm E (ATV) if they have underlying heart disease or are receiving other drugs that may effect conduction. Those with the following conditions are ineligible for treatment on Arm E:

- symptomatic heart block
- 3rd degree heart block, even if asymptomatic
- pre-excitation syndromes
- heart rate <40 bpm
- ventricular pause length >3 sec
- QTc >500 msec
- history of syncope of undetermined origin
- cardiomyopathy

APPENDIX 2

Excerpt from ACTG 5095 Protocol Schedule of Events

Evaluations	Screen[1]	Pre-Entry[2]	Entry to Step 1[3] F	Evaluations During Study Treatment (in weeks)											
				*1	*2	4	6	8	12 F	16	20	24 F	32	40	48 F
Informed Consent	X														
Documentation of HIV-1	X														
Medical/Medication History[4]	X														
Clinical. Assessment Signs/Symptoms[5]		X	X		X	X		X	X	X	X	X	X	X	X
Prescription/Concomitant Meds[6]		X	X		X	X	X	X	X	X	X	X	X	X	X
Hematology/Chemistries/LFTs[7]		X	X		X	X	X	X	X	X	X	X	X	X	X
Lipid Panel[8]			X							X			X		X
Hepatitis Serology[21]			X[21]												
EKG															
Urinalysis w. microscopic exam			X												
Urinalysis without microscopic exam															
Pregnancy Test (Urine or Serum)[2]		X													
CD4/CD8[9]		X	X			X		X		X		X	X	X	X
HIV-1 RNA[10]	X	X	X		X	X		X	X	X	X	X	X	X	X
Genotyping[11]			X												
Plasma Storage[12]			X		X	X		X	X	X	X	X	X	X	X
PBMC Storage[12]			X									X			X
Fasting Metabolic Blood Tests[13]			X						X			X			X
Body Measurements[14]			X						X			X			X
ALLRT Self Report			X			X			X			X			X
Gyn. Status Questionnaire			X[25]							X			X		X
A5097s Neurology Substudy[15]			X	X		X			X			X			
A5166s Viral Dynamics[23]				X	X			X							

REFERENCES

Abbott, Andrew D. (1988). *The System of Professions: An Essay on the Division of Expert Labor*. Chicago, IL: University of Chicago Press.

Abramson, J. (2008). *Overdosed America: The Broken Promise of American Medicine*. New York: Harper Collins.

Africa Clinical Trials Portal. (2004). *About Africa Clinical Trials Portal.* Retrieved from, http://www.africaclinicaltrials.org/About.aspx.

American Association of Medical Colleges. (2005). *Hot Topics in Medical Education*. http://services.aamc.org/currdir/section2/04_05hottopics.pdf

Annas, G. J. & Grodin, M. A. (1998). Human rights and maternal-fetal HIV transmission prevention trials in Africa, *American Journal of Public Health*, 88, 560-563.

Anspach, R. (1988). Notes on the sociology of medical discourse: the language of case presentation. *Journal of Health and Social Behavior*, 29, 357-75.

Argote, L. (1982). Input uncertainty and organizational coordination in hospital emergency units. *Administrative Science Quarterly*, 27, 420-434.

Barley, S. (1990). The alignment of technology and structure through roles and networks, *Administrative Science Quarterly*, 35,.

Barley, S. & Kunda, G. (2001). Bringing work back in, *Organization Science*, 12(1), 76-95.

Barnett, T. & Whiteside, A. (2006). *AIDS in the 21st century: Disease and globalization*. Palgrave Macmillan: New York.

Becker, H. (1982). *Art Worlds*. Berkley: University of California Press.

Bellomo, R., & Bagshaw, S. M. (2006). Evidence-based medicine: classifying the evidence from clinical trials – the need to consider other dimensions. *Critical Care*, 10(5), 232.

Berg, M. (1997). *Rationalizing medical work: Decision-support techniques and medical practices*. Cambridge, MA: MIT Press.

Berg, M. & Bowker, G. (1997). The multiple bodies of the medical record: Toward a sociology of an artifact. *The Sociological Quarterly*, 38, 511-535.

Berwick, D. (2003). Disseminating innovations in health care, *Journal of the American Medical Association*, 289, 1969-1975.

Bijker, W. E., Hughes, T. P. & Pinch, T. J. (1990). *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. Cambridge, MA: MIT Press.

Blau, P. (1955). *The dynamics of bureaucracy*. Chicago: University of Chicago Press.

Bourdieu, P. (1977). *Outline of a theory of practice*. Cambridge, UK: Cambridge University Press.

Bowker, G. & Star, L. (1999). *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT Press.

Bowker, G. (1994). *Science on the run: Information management and industrial geophysics at Schlumberger, 1920-1940*. Cambridge, MA: MIT Press

Braithwaite, J. & Drahos, P. (2000). *Global business regulation*. Cambridge UK: Cambridge University Press.

Bynum, W. (1994). *Science and the practice of medicine in the nineteenth century*. Cambridge UK: Cambridge University Press.

Callahan, D. (2003). *What price better health: Hazards of the research imperative*. Berkley: University of California Press.

Cambrosio, A., Keating, P., Schlich, T., & Weisz, G. (2006). Regulatory objectivity and the generation and management of evidence in medicine. *Social Science and Medicine*, 63(1), 189-199.

Carlson, R. V., Boyd, K. M., & Webb, D. J. (2004). The revision of the Declaration of Helsinki: past, present and future. *British Journal of Clinical Pharmacology*, 57(6), 695-713.

Center for Disease Control and Prevention. (1996). The role of BCG vaccine in the prevention and control of tuberculosis in the United States: a joint statement by ACET and the advisory committee on immunization practices. *Morbidity and Mortality Weekly Report*, 45 (No. RR-4), 1-18.

Center for Disease Control and Prevention. (2006). *BCG Vaccine Fact Sheet*. http://www.cdc.gov/tb/pubs/tbfactsheets/BCG.htm.

Center for Disease Control and Prevention. (2007). *CDC HIV/AIDS Statistics and Surveillance*. http://www.cdc.gov/hiv/topics/surveillance/index.htm.

Chambliss, D. (1996). *Beyond caring: Hospitals, nurses and the social organization of ethics*, Chicago: University of Chicago Press.

Check, E. & Castellani, F. (2004). David versus Goliath, *Nature*, 432, 546-548

Clarke, A. (1990). A social worlds research adventure: The case of reproductive science. In S. Cozzens & T. Gieryn (Eds.), *Theories of Science in Society* (pp. 23-50). Bloomington: Indiana University Press.

Concato, J., Shah, N., & Howitz, R. I. (2000). Randomized, controlled trials, observational studies, and the hierarchy of research designs. *New England Journal of Medicine*, 342(25), 1887-1892.

Connor, E. M., Sperling, R.S., Gelber, R., Kisley, P., Scott, G., VanDyke, R., et al. (1994). Reduction of maternal-infant transmission of human immunodeficiency virus type 1 with zidovudine treatment. Pediatric AIDS Clinical Trials Group Protocol 076 Study Group. *New England Journal of Medicine*, 331(18), 1173-80.

Cunningham, W. E., Hays, R. D., Duan, Andersen, N., Nakazono, R., Bozzette, T. T., et al. (2005). The effect of socioeconomic status on the survival of people receiving care for HIV infection in the United States. *Journal of Health Care for the Poor and Underserved* 16, 655-676.

Daly, J. 2005. *Evidence-based medicine and the search for a science of clinical care*. Berkley: University of California Press.

Davidoff, B., Haynes, D., Sackett, D., & Smith, R. (1995). Evidence-based medicine, *British Medical Journal*, 310, 1085–1086.

Department of Health and Human Service Panel on Antiretroviral Guidelines for Adults and Adolescents. (2005). *Guidelines for the Use of Antiretroviral Agents among HIV-Infected Adults and Adolescents Guidelines*. http://aidsinfo.nih.gov

Department of Health and Human Service Panel on Antiretroviral Guidelines for Adults and Adolescents. (2006). *Guidelines for the Use of Antiretroviral Agents among HIV-Infected Adults and Adolescents Guidelines*. http://aidsinfo.nih.gov

Department of Health and Human Service Panel on Antiretroviral Guidelines for Adults and Adolescents. (2008). *Guidelines for the Use of Antiretroviral Agents among HIV-Infected Adults and Adolescents Guidelines*. http://aidsinfo.nih.gov

DiMaggio, P.J. & Powell, W.W. (1983). The iron cage revisited. *Administrative Science Quarterly*, 48, 147-160.

Division of AIDS. (2004). *Table for Grading the Severity of Adult and Pediatric Adverse Events*. http://www3.niaid.nih.gov/research/resources/DAIDSClinRsrch/PDF/Safety/DAIDSAEGradingTable.pdf.

Division of AIDS. (2007). *Overview of Division of AIDS*. Retrieved from http://www3.niaid.nih.gov/about/organization/daids/

Drori, G. & Meyer, J. (2006). Scientization: making a world safe for organizing. In M. Djelic (Ed.), *Transnational Governance: Institutional Dynamics of Regulation* (pp. 31-52), Cambridge UK: Cambridge University Press.

Duesberg, P. (1995). *Infectious AIDS: Have we been misled?* Berkley: North Atlantic Press.

Epstein, S. (1996). *Impure science: AIDS, activism and the politics of knowledge*. Berkley: University of California Press.

Espeland, W. N. (1998). *The struggle for water: Politics, rationality, and identity in the American southwest*. Chicago: University of Chicago Press.

Espeland, W. & Stevens, M. (1998). Commensuration as social process. *Annual Review of Sociology*, 24, 313-343.

Feldman, M. & Pentland, B. (2003). Reconceptualizing organizational change as a source of flexibility and change. *Administrative Science Quarterly*, 48(1), 94-118.

Feyerabend, P. K. (1993). *Against Method* (3rd ed.). New York: Verso.

Field, K. & Lohr, J. (1992). *Guidelines for clinical practice: From development to use*. Washington DC: National Academy Press.

Folkers, G. K. & Fauci, A. S. (2001). The AIDS research model: Implications for other infectious diseases of global health importance. *Journal of the American Medical Association*, 286(4), 458-461.

Foucault, M. (1977). *Discipline and punish: Birth of the prison*. New York: Pantheon Books.

Foucault, M. (1970). *The order of things: An archaeology of the human sciences*. New York: Pantheon Books.

Fox, R. C. (1957). Training for uncertainty. In R. K. Merton, G. G. Reader & P. Kendall (Eds.), *The Student Physician* (pp. 207-41). Cambridge: Harvard University Press.

Fox, R. C. (1974). *Experiment Perilous: Physicians and Patients Facing the Unknown*. Philadelphia: University of Pennsylvania Press.

Fox, R. C. (1979). *Essays in Medical Sociology: Journeys into the Field*. New York: Wiley.

Fox, R. C. (1989). *The Sociology of Medicine: A Participant Observer's View*. Englewood Cliffs, N.J.: Prentice Hall.

Fox, R. C. (2000). Medical uncertainty revisited. In G. Albrecht, R. Fitzpatrick, & S. Scrimshaw (Eds.), *The Handbook in Social Studies of Health and Medicine* (pp. 409-25). London: Sage Press.

Freedman, B. (1987). Equipoise and the ethics of clinical research. *New England Journal of medicine*, 317, 141-145.

Friedman, L. M., Furberg, C., & DeMets, D. 1998. *Fundamentals of clinical trials*. New York: Springer.

Freidson, E. (1986). *Professional powers: A study of the institutionalization of formal knowledge*. Chicago: University of Chicago Press.

Fujimura, Joan H. (1987). Constructing "doable" problems in cancer research: Articulating alignment. *Social Studies of Science* 17, 257-293.

Fujimura, J. (1992). Crafting science: Standardized packages, boundary objects and "translation." In A. Pickering (Ed.), *Science as Practice and Culture* (p. 168-211). Chicago: University of Chicago Press.

Gates Foundation. (2008). *Grants from Inception through 2007*. http://www.gatesfoundation.org/GlobalHealth/Pri_Diseases/HIVAIDS/Grants.

Gerbert, B., Moe, J., Saag, M., Benson, C., Jacobsen, D., Feraios, A., et al. (2001). Toward a definition of HIV expertise: A survey of experienced HIV physicians. *AIDS Patient Care STDS, 15*(6), 321-330.

Giddens, A. (1984). The constitution of society: Outline of the theory of structuration. Berkley: University of California Press.

Giddens, A. (1991). The consequences of modernity. Stanford University Press.

Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Kruger, L. (1989). *The empire of chance: How probability changed science and everyday life*. Cambridge UK: Cambridge University Press.

Glasgow, R.E., Magid, D.J., Ritzwoller, D., & Estabrooks, P.A. (2005). Practical clinical trials for translating research to practice: Design and measurement recommendations. *Medical Care*, 43(6), 551-7.

Goldenberg, M. J. (2006). On evidence and evidence-based medicine: lessons from the philosophy of science. *Social Science and Medicine*, 62(11), 2621-32.

Goodman, K. W. (2003). *Ethics and evidence-based medicine: Fallibility and responsibility in clinical science*. Cambridge UK: Cambridge University Press.

Gordon, D. (1988). Clinical science and clinical expertise: changing boundaries between art and science in medicine. In M. Lock & D. Gordon (Eds.), *Biomedicine Examined* (pp. 257–295). Norwell, MA: Kluwer Academic Publishers.

Greenhalgh, T. (2006). *How to read a paper: The basics of evidence-based medicine* (3rd ed.). Malden, MA: Blackwell Publishing.

Guyatt, G. & Rennie, D. (2001). *Pocket-sized user's guide to the medical literature: Essentials of evidence-based clinical practice*. Chicago: American Medical Association Press.

Gulick RM, Ribaudo HJ, Shikuma CM, Gulick, Ribaudo, H. J., Shikuma, C. M,, Lustgarten, M. S., Squires, K. E., Meyer, W. A., et al. (2004). Triple-nucleoside regimens versus efavirenz-containing regimens for the initial treatment of HIV-1 infection. *New England Journal of Medicine*, 350, 1850-1861

Hacking, I. (1983). *Representing and intervening*. Cambridge UK: Cambridge University Press.

Hacking, I. (1990). *The taming of chance*. Cambridge UK: Cambridge University Press.

Hafferty, F. (2000). Reconfiguring the sociology of medical education: Emerging topics and pressing issues," In C. Bird, P. Conrad & A. Fremont (Eds.). *Handbook of Medical Sociology* (pp.238-257). Upper Saddle River, NJ: Prentice Hall.

Harris, R.P, Helfand, M., Woolf, S.H., Lohr, K.N., Mulrow, C.D., Teutsch, S.M., et al. (2001). Current methods of the U.S. Preventive Services Task Force: A review of the process. *American Journal of Preventive Medicine*, 20(3S), 21-35.

Heimer, C. A, & Espeland, W. N. 2008. Side effects: Accountability and the treatment of HIV/AIDS. Presented at the annual Law and Society Annual Meeting, Montreal, CA.

Heimer, C. A. Petty, J., & Culyba, R. J. (2005). Risk and rules: The legalization of medicine. In B. Hutter and M. Powers (Eds.) *Organizational Encounters with Risk* (pp. 92-131). Cambridge UK: Cambridge University Press.

Heimer, C. A. & Staffen, L. (1998). *For the sake of the children: The social organization of responsibility in the hospital and in the home*. Chicago: University of Chicago Press.

Heimer, C. A. & Stevens, M. 1997. Caring for the Organization: Social Workers as Frontline Risk Managers in Neonatal Intensive Care Units. *Work and Occupations*, 24(2), 133-163.

Henkel, J. (1999). Attacking AIDS with a 'cocktail' therapy. *FDA Consumer*, 33(4). Retrieved from http://www.fda.gov/FDAC/499_toc.html

Horton, R. (1995). MAST-I: Agreeing to disagree. *The Lancet*, 346, 1504.

orton, R. (2003.) *Health wars: On the global frontline of modern medicine*. New York: New York Review Books.

Huque, M. F. (1988). Experiences with meta-analysis in NDA submissions. *Proceedings of the Biopharmaceutical Section of the American Statistical Association*, 2, 28-33.

International Conference on Harmonization. (1997). E6 Guidance for Industry: Good Clinical Practice. *Federal Register*, 62(90), 25691-25709E7.

Klimas, N., O'Brien, A.K., & Fletcher, M. (2008). Overview of HIV. *Psychosomatic Medicine*, 70, 523–530.

Knorr-Cetina, K. (1981). *The manufacture of knowledge: an essay on the constructivist and contextual nature of science*. Oxford: Pergamon Press.

Knorr-Cetina, K. (1983). The ethnographic study of scientific work: towards a constructivist interpretation of science studies. In K. Knorr-Cetina & M. Mulkay (Eds.), *Science observed: perspectives on the social study of science* (pp. 115-40). London: Sage Publications.

Knorr Cetina, K. (1995). Laboratory studies: The cultural approach to the study of science. In S. Jasanoff, G. E. Markle, J. C. Petersen, & T. Pinch (Eds.), *Handbook of science and technology studies* (pp. 140-66). Thousand Oaks: Sage Publications.

Knorr-Cetina, K,, & Mulkay, M. (1983). Introduction: Emerging principles of social studies of science. In K. Knorr-Cetina & M. Mulkay (Eds.), *Science observed: perspectives on the social study of science* (pp. 1-17). London: Sage Publications.

Kuhn, T. S. (1965). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Lambert, H. 2006. Accounting for EBM: notions of evidence in medicine. *Social Science and Medicine*, 62(11), 2633-2645.

Lambert, H., Gordon E. J., & Bogdan-Lovis, E. A. (2006). Introduction: gift horse or trojan horse? social science perspectives on evidence based health care. *Social Science and Medicine*, 62(11), 2613-20.

Latour, B. (1981). Is it possible to REconstruct the REsearch process? sociology of a brain peptide. In K. Knorr, R. Kohn, & R. Whitley (Eds.) *The social process of scientific investigation* (pp. 53-73). London: D. Reidel Publishing.

Latour, L. & Woolgar, S. (1979). *Laboratory life: the social construction of scientific facts*. London: Sage Press.

Latour, B. (1983). Give me a laboratory and i will raise the world. In K. Knorr-Cetina & M. Mulkay (Eds.) *Science Observed: Perspectives on the Social Study of Science* (pp. 140-170). London: Sage Publications.

Latour, B. (1987). *Science in Action*. Cambridge: Harvard University Press.

Latour, B. (1997). *On actor-network theory: a few clarifications*. Keele University, UK: Centre for Social Theory and Technology. http://amsterdam.nettime.org/Lists-Archives/nettime-l-9801/msg00019.html.

Latour, B. (1999). On recalling ANT. In J. Law & J. Hassard (Eds.), *Actor network theory and after* (pp. 15-25). Oxford: Blackwell Publishers.

Leidner, R. (1993). *Fast food, fast talk: Service work and the routinization of everyday life*. Berkley: University of California Press.

Light, D. (1979). Uncertainty and Control in Professional Training. *Journal of Health and Social Behavior* 20(4), 310-322.

Linzer, M. (1987). The journal club and medical education: over 100 years of unrecorded history. *Postgraduate Medical Journal*, 63, 475-478.

Lowy, I. (1996). *Between bench and bedside: science, healing and interleukin-2 in a cancer ward*. Cambridge: Harvard University Press.

Lurie, P. & Greco D. (2005). US exceptionalism comes to research ethics. *The Lancet*, 365 (9465), 1117-19.

Majumdar S., Roe, M., Peterson, E.D., Chen, A., Gibler, B., & Armstrong, P. (2008). Better outcomes for patients treated at hospitals that participate in clinical trials. *Archives of Internal Medicine*, 168, 657-662.

March J. (1988). *Decisions and Organizations*. Oxford: Basil Blackwell.

Marks, H. M. (1997). *The progress of experiment: Science and therapeutic reform in the Unites States, 1900-1990*. New York: Cambridge University Press.

Mathews, J. R. (1995). *Quantification and the quest for medical certainty*. Princeton, NJ: Princeton University Press.

Mathews, J. R. (1999). Practice Guidelines and Tort Reform: The Legal System Confronts the Technocratic Wish. *Journal of Health Politics, Policy and Law*, 24(2), 275-304.

Medical Research Council. (2008). *A snapshot of MRC research into HIV and AIDS*. http://www.mrc.ac.uk/OurResearch/Impact/Infections/HIVAIDS/index.htm

Millenson, M. L. (1997). *Demanding medical excellence: doctors and accountability in the information age*. Chicago: University of Chicago Press.

Moses, H., Dorsey, E. R., Matheson, D. H., & Their, S. O. (2005). Financial anatomy of biomedical research. *Journal of the American Medical Association*, 1333-1342.

Mueller, M. R. (1997). Science versus care: physicians, nurses and the dilemma of clinical research. In M. A. Elston (Ed.), *The sociology of medical science and technology* (pp. 57-78). Oxford, Blackwell Publishers.

Mulrow, C. & Lohr, K. (2001). Proof and policy from medical research evidence. *Journal of Health, Politics, Policy and Law*, 26(2), 249-266.

National Institute of Allergies and Infectious Disease. 1999. *NIAID renews funding for Adult AIDS Clinical Trial Group*. National Institutes of Health New Release. http://www3.niaid.nih.gov/news/newsreleases/1999/aactg.htm.

National Institutes of Health. (2007). *Research Project Success Rate by Institute 2007*. Retrieved November 2, 2008 from, http://grants.nih.gov/grants/award/success/Success_ByIC.cfm

National Institutes of Health. 2008. *How does clinical research work*. Retrieved May 15, 2008 from, http://clinicalresearch.nih.gov/how.html

Office of AIDS Research. (2008). FY 2008 Congressional Budget Justification. Retrieved May 15, 2008 from, http://www.oar.nih.gov/budget/fy08justification.asp

Palella, F. J. Jr, Delaney, K.M., Moorman, A.C., Loveless, M.O., Fuhrer, J, Satten, G.A., et al. (1998). Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators. *New England Journal of Medicine*, 338(13), 853-60.

Panem, S. (1988). *The AIDS bureaucracy*. Cambridge: Harvard University Press.

Parkhurst, J. (2002). Ugandan success story? Evidence and claims of HIV-1 prevention. *The Lancet*, 360 (9326), 78-80.

Parsons, T. (1951). *The social system*. Chicago: University of Chicago Press.

Perrow, C. & Guillén, M. F. (1990). *The AIDS disaster: The failure of organizations in New York and the nation*. New Haven: Yale University Press.

Pickering, A. (1984). *Constructing quarks: A sociological history of particle physics*. Chicago: University of Chicago Press.

Porter, T. (1995). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton: Princeton University Press.

Reiser, J. (1978). *Medicine and the reign of technology*. New York: Cambridge University Press.

Reynaud, B. (2005). The void at the heart of the rules: Routines in the context of rule-following. The case of the Paris metro workshop. *Industrial and Corporate Change*, 14(5), 847-871.

Resnik, D. B. (1998). The ethics of HIV research in developing nations. *Bioethics*, 286.

Rettigg, R. A. (2000). The industrialization of clinical research. *Health Affairs*, (March/April), 129-146.

Rosenberg, C. (1987). *The Care of Strangers: The Rise of America's Hospital System*. Chicago: University of Chicago Press.

Rosengarten, M. (2004). After the euphoria: HIV medical technologies from the perspective of their prescribers. *Sociology of Health and Illness*, 26(5), 575-596.

Rosser, W. W., Davis, D., & Gilbart, E. (2001). Assessing Guidelines for Use in Family Practice, *Journal of Family Practice*, 50, 974–75.

Sackett, D. L. (1986). Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest*, 89 (2 suppl.), 2S-3S.

Sackett, D. L. (1993). Rules of Evidence and Clinical Recommendations. *Canadian Journal of Cardiology*, 9(6), 487-9.

Sackett, D. L. (1996). Evidence based medicine: What it is and what it isn't. *British Journal of Medicine*, 312, 71-72.

Schackman, B. R., Gebo, K. A., Walensky, R. P, Losina, E., Muccio, T., Sax, P. E., et al. (2006). The lifetime cost of current human immunodeficiency virus care in the United States. *Medical Care*, 44, 990-7.

Scott, J. (1998). *Seeing like a state: How certain schemes to improve the human condition have failed*. New Haven, CT: Yale University Press.

Scott, R., Ruef, M., Mendel, P. & Caronna, C. (2000). *Institutional Change in Healthcare Organizations: From Professional Dominance to Managed Care*. Chicago: University of Chicago Press.

Sewell, W. H. (1992). A theory of structure: Duality, agency, and transformation. *American Journal of Sociology*, 98(1), 1-29.

South Africa Department of Health. (2003).*Operational Plan for Comprehensive HIV and AIDS Care, Management and Treatment for South Africa*. Retrieved July 1, 2008 from, http://www.info.gov.za/otherdocs/2003/aidsplan/report.pdf

Smith, D. (1990). *The conceptual practices of power: A feminist sociology of knowledge*. Toronto: University of Toronto Press.

Star, S. L. (1983). Simplification in scientific work: An example from neuroscience research. *Social Studies of Science*, 13, 205-228.

Star, S. L. & Griesemer, J. (1989). Institutional ecology, "translations" and boundary objects: Amateurs and professionals in Berkley's museum of vertebrate zoology, 1907-1939. *Social Studies of Science*, 19, 387-420.

Starr, P. (1982). *The social transformation of American Medicine*. Basic Books.

Stinchcombe, A. (2001). *When formality works: Authority and abstraction in law and organizations*. Chicago: University of Chicago Press.

Strauss, A. (1978). A Social World Perspective. In Denzin, N. (Ed.), *Studies in Symbolic Interation* (Vol. 1), (pp. 119-128). Greenwich, CT: JAI Press.

Strauss, S. E. (2005). *Evidence-based medicine: How to practice and teach EBM*. Edinburgh: Elsevier/Churchill Livingstone.

Suchman, M. (2007). Commercializing clinical trials: Risks and benefits of the CRO boom. *New England Journal of Medicine*, 357(14), 1365-1368.

Tannebaum, S. (1994). Knowing and acting in medical practice: The epistemological politics of outcomes research. *Journal of Health Politics and Law*, 19(1), 27-44.

Timmermans, S. & Berg, M. (1997). Standardization in action: Achieving local universality through medical protocols. *Social Studies of Science*, 26, 769–99.

Timmermans, S. & Kolker, E. (2005). Clinical Practice Guidelines and the Reconfiguration of Medical Knowledge, *Journal of Health and Social Behavior*, 45, (Suppl.), 177-193.

Timmermans, S. & Angell, A. (2001). Evidence-based medicine, clinical uncertainty, and learning to doctor. *Journal of Health and Social Behavior*, 42(4), 342-359.

Timmermans, S. & Berg, M. (2003). *The gold standard: The challenge of evidence based medicine and standardization in healthcare*. Philadelphia: Temple University Press.

Tuchman, A. (1993). S*cience, medicine, and the state in Germany: the case of Baden, 1815-1871*. New York: Oxford University Press.

Turner, B., Newschaffer, C. J., Zhang, D., Fanning, T., & Hauck, W. W. (1999). Translating clinical trial results into practice: The effect of an aids clinical trial on prescribed antiretroviral therapy for HIV-infected pregnant women. *Annals of Internal Medicine*, 130 (12), 979-986.

United Nations Program on HIV/AIDS. 2007. *The AIDS Epidemic Update*. Retrieved from http://data.unaids.org

Upshur, R. E. G., VanDenKerkhof, E. G., & Goel, V. (2001). Meaning and measurement: An inclusive model of evidence in health care. *Journal of Evaluation in Clinical Practice*, 7, 91–96.

Vandenbroucke, J.P. (1998). Observational research and evidence-based medicine: What should we teach young physicians? *Journal of Clinical Epidemiology*, 51(6), 467-72.

Vázquez, E. (2003). We want our Trizivir! *Positively Aware*, September/October, 27-28.

Wawer, M.J. Gray, R., Serwadda, D., Namukwaya, Z., Makumbi, F., Li, X, et al. (2005). Declines in HIV prevalence in Uganda: Not as simple as abc. Twelfth Conference on Retroviruses and Opportunistic Infections, Boston, Abstract 27LB. Retrieved from, http://www.retroconference.org/2005/cd/Abstracts/25775.htm.

Weber, M. (1978). *Economy and society: an outline of interpretive sociology*. Berkley: University of California Press.

Weick, K. (1998). Improvisation as a mindset for organizational analysis," *Organization Science*, 9(5), 543-555.

Weidle, P. J., Malambia, R., Mwebazi, C., Sozi, G., Rukundo, R., Downing, D., et al. (2002). Assessment of a pilot antiretroviral drug therapy programme in Uganda: patients' response, survival and drug resistance," *The Lancet*, 360 (9326), 34-40.

Wilson, I. B., Landon, B. E., Ding, L., Zaslavsky, A.M., Shapiro, M.F., Bozzette, S.A., et al. (2005). A national study of the relationship of care site HIV specialization to early adoption of highly active antiretroviral therapy. *Medical Care*, 43(1), 12-20.

Woolf, S.H., DiGuiseppi, C.G., Atkins, D., & Kamerow, D.B. (1996). Developing evidence-based clinical practice guidelines: lessons learned by the U.S. preventive services task force. *Annual Review of Public Health*, 17, 511-538.

Collins, H. & Pinch, T. (1994). *The golem: What everyone should know about science*. Cambridge, UK: Cambridge University Press.

World Health Organization. 2003. *Scaling up antiretroviral therapy in resource poor settings: Treatment guidelines for a public health approach*. Retrieved from http://www.who.int/hiv/pub/prev_care/en/arvrevision2003en.pdf.

World Health Organization. 2005. *Handbook for Good Clinical Practice: Guidance for Implementation*. Retrieved from, http://www.who.int/entity/medicines/areas/quality_safety/safety_efficacy/gcp1.pdf.

World Health Organization. 2006. *Antiretroviral therapy for adults and adolescents: Towards universal access. A public health approach.* Retrieved from, http://www.who.int/hiv/topics/treatment/en/index.html

World Health Organization. 2006. *News Release: WHO calls for immediate halt of provision of single-drug artemisinin malaria pills.* Retrieved November 2, 2008 from, http://www.who.int/mediacentre/news/releases/2006/pr02/en/index.html

World Health Organization, United Nations Program on HIV/AIDS & United Nations Children's Fund. (2007). *Towards universal access: scaling up priority HIV/AIDS interventions in the health sector, progress report.* http://www.who.int/hiv/mediacentre/universal_access_progress_report_en.pdf.

Zuger, A. (2006, January 3). Cleaning up the mess of medicine in the pages of posterity. *New York Times*. Retrieved from, http://www.nytimes.com

Footnotes

[1] In this dissertation, fieldnote excerpts are followed by a code identifying the clinic where the data was collected, the date it was collected and the initials of the fieldworker who collected the data. The American private clinic, American public clinic, the Ugandan clinic and the South African clinic are referred to as US1, US2, UG and SA, respectively, in the code.

[2] "Network" is a special term used by Latour (1997) and other science study scholars to evoke the "fibrous, thread-like, wiry, stringy, ropy, capillary character" of modern society. Sturdiness arises from the "netting, lacing, weaving, [and] twisting" of weak ties (Latour 1997).

[3] Retrieved June 24, 2008 from, www.Clinicaltrials.gov.

[4] These numbers are based on my notes on materials observed at the Ugandan Council on Science and Technology in Kampala, Uganda.

[5] Heimer's Clinic-Level Law project also included fieldwork in a Thai clinic, which is not included in this dissertation. The funders of this project include the American Bar Foundation, the National Science Foundation (NSF SES – 0319560) and the Russell Sage Foundation.

[6] At the start of our fieldwork in 2006, the exchange rate for Rand was around 6 Rand for 1 US dollar. Thus 70 Rand was just under 12 US dollars.

[7] While some clinical research occurs in purely research settings, much of clinical research occurs within practicing clinics. And the latter is the case in the clinic that I am writing about in this chapter.

[8] In six of the meetings, there were no adverse events to report. In another case, the meeting was a special training meeting that did not follow the agenda. In the eighth case, the fieldworker could not hear the adverse event conversation from across the room. Note that this overestimates the total number of adverse events addressed by the group because the same adverse event might be ongoing and discussed for weeks.

[9] DAIDS funded ACTG studies, which made up the bulk of research at this clinic, used this grading scheme. Industry studies used a version of the DAIDS grading table; these tables sometimes did not use numbers (grade 1, 2, 3 or 4) just severity categories (mild, moderate, severe, life threatening.)

[10] There are 3 categories to describe the relationship to study drug on the ACTG SAE form: definitely, possibly and unlikely. Note here that the pharmacist is actually confusing "probably" with "possibly," which have different meanings.

[11] Category II does include observational studies in addition to non-randomized controlled trials. In this case, clinical research simply indicates research conducted on real patients as opposed to laboratory research on animals or cells.

[12] Clinical endpoints include outcomes like heart attack and death, while laboratory endpoints refer to indicators of disease progression such as CD4 counts for AIDS or cholesterol serum levels for heart disease. Clinical endpoints are considered a better indication of the effect of an intervention because of the ambiguous relationship between laboratory tests and clinical outcomes. For

example, not everyone with high cholesterol has a heart attack; nor do all heart attack victims have high cholesterol.

[13] The other six included an FDA report, CDC guidelines, two commentary articles, a policy statement from a professional association, and an early CDC MMWR on HIV to mark the anniversary of the discovery of HIV.

[14] Articles were coded as "good" if the article was described as "good" or "strong" by one or more journal club members with no disagreement. Similarly, articles were coded as "weak" if the article was described as "weak" or "poor" by one or more journal club members with no disagreement. Articles that fell in the middle were coded as "okay." If there was disagreement about the quality of the article was coded as "mixed."

[15] There were only two meta-analysis discussed in the club, but I followed up on the distrust of meta-analysis in interviews with journal club members.

[16] Infectious disease fellows have completed medical school as well as residency training in internal medicine specialty. The purpose of a fellowship is to sub-specialize, in this case in infectious disease. During this portion of their training, fellows are supervised by attending physicians (such as the assistant chief of infectious disease) who have already completed a fellowship.

[17] Infectious disease physicians are not against all forms of antibiotic prophylaxis. For example, everyone agrees that prescribing Bactrim to prevent opportunistic infections in HIV patients with CD4 counts below 200 is appropriate.

[18] BCG, or bacille Calmette-Guérin, is a vaccine for tuberculosis disease and is used in many countries with a high prevalence of tuberculosis. BCG is not recommended for use in the United States because of the low risk of infection with tuberculosis, the variable effectiveness (CDC, 2006).

[19] In the introduction to his book on formality, Stinchcombe's (2001) observes that perhaps the book's subtitle should have been "How and Why Formality Works, If and When it Does" (1).

[20] Breastfeeding posed an especially difficult challenge. In Uganda, extended breastfeeding is expected and mothers who do not breastfeed are suspect. HIV can be transmitted to babies through breast milk, and in places with access to clean water, it is recommended that HIV positive mothers feed their infants formula. Where access to clean water is unreliable, the risk of serious illness associated formula feeding can actually be higher than the risk of HIV infection through breastfeeding. The Uganda Ministry of Health guidelines recommend exclusive breastfeeding for 6 months.

[21] According to *Medline*, DHEA is a hormone naturally produced by the human body which serves as precursor to male and female sex hormones. Some people with AIDS have low levels of this hormone. Retrieved from, http://www.nlm.nih.gov/medlineplus/druginfo/natural/patient-dhea.html

[22] Note that this partly a matter of moral performance for the fieldworker and the rest of the group. After this conversation, the study nurse turned to me and asked, are you getting this? (I had been, of course, busily scribbling notes.) I replied, which part? She helpfully repeated her statement for my benefit: "You

don't give up on a study patient until they are in the ground." I suspect that she worried that the other study nurse's frustration with a study patient looked bad.

[23] See Weijer, C. Shapiro, S.H., Cranley Glass K., and Enkin, M. 2000. "For and Against: Clinical Equipoise and Not the Uncertainty Principle is the Moral Underpinning of the Randomised Controlled Trial," BMJ, 321: 756-758.

[24] Researcher beliefs about whether a treatment works are inadequate indicators of whether a treatment actually works – at least according to statistical measures. Indeed, during our fieldwork, one study was stopped by the safety monitoring board because their was no difference between the control and treatment group. Before this, the research nurse was really eager to get patients enrolled in this study because she had seen lots of study patients benefit from the treatment.

[25] There are, however, some relevant differences in status and education between the Ugandan-trained physicians or "medical officers" and the western-trained physician-investigators; authority relationships is discussed in detail in the following chapter.

[26] A viral load test measures the amount of HIV virus in the blood and is used as an indicator of disease progression and efficacy of treatment.

[27] An early version of this chapter, titled "How Research Shapes Medical Work: Organizational Effects of Clinical Trials" was presented with Carol Heimer at the 2006 meetings of the American Sociological Association.

[28] An adverse event is any undesirable event associated with a medical product. It is defined as serious when it leads to disability, birth defect, hospitalization or death www.fda.gov/medwatch). Serious adverse events are tracked by funding agencies and IRBs.

[29] What is new and expensive changes over time. At the time of our fieldwork, Fuzeon was the new expensive drug. With CCR5 inhibitors, the new class of HIV drugs approved by the FDA in 2007, the concern is not with the cost of the drugs but with the costs of the expensive "receptor test." Receptor tests determine whether a patient with HIV has the kind of receptor that is compatible with the drug. State funded care programs next scramble to figure out how funds for the test will be distributed.

[30] As mentioned in above, the drugs only work in people with particular kinds of HIV receptors, so a blood test to determine the kind of receptors a patient is required before prescribing the drugs.

[31] "Procedures and Criteria for Selection of Individual AIDS Clinical Trials Units and External Expertise," Standard Operating Procedure, NIAID ADULT AIDS CLINICAL TRIALS GROUP, ACTG-SOP No. AACTG-122, version 2.

[32] http://www.niaid.nih.gov/daids/cipra/u01.htm, accessed November 10, 2005.

[33] COMPREHENSIVE INTERNATIONAL PROGRAM OF RESEARCH ON AIDS (CIPRA). 2001. Program Announcement, released March 15, 2001,

NOT-AI-01-018, National Institute of Allergy and Infectious Diseases(http://www.niaid.nih.gov)

[34] Technical efficiency refers to the efficiency of isolated acts (e.g. how well a drug regimen reduces viral load). Organizational efficiency refers to the ease with which a task can be performed within a complex environment (Lowy 1996, Abbott 1988).

[35] As a clinic receiving state funds, the care part of the American public clinic was highly standardized. Indeed, the reporting requirements were high – much greater than the care part of the American private clinic. The public clinic tracked patients and care so that they could give periodic accounts of their work; documentation was crucial to getting credit. In the American public clinic, getting credit for care work intruded somewhat less because the American private clinic relied on the billing unit in the attached hospital to handle claims. Importantly, there were many different insurers and not one central funder that the American clinic had to account to for care. The public clinic was so concerned with tracking patients every patient that came through their doors that the research unit even submitted a paper form for every study patient to the care part of the organization.

[36] Percents do not always add to 100 because of rounding.