

Running head: DIAGNOSTIC ERRORS AND ISABEL

Diagnostic Errors and ISABEL: A Web-Based Diagnostic Reminder System

Janice Vaughn MSN, RN, CNL

University of New Hampshire

Diagnostic Errors and ISABEL: A Web-Based Diagnostic Reminder System

Background

Diagnostic error rates have not significantly changed in the last forty years (Kirch & Schafii, 1996; Shojania, Burton, McDonald, & Goldman, 2003). In modern medicine, there exists a gap between the theoretical promise of appropriate or correct diagnosis via high-tech advances and the delivery of such in real life everyday practice (Lundberg, 1998). The best method of identifying diagnostic discrepancies is the autopsy, long recognized as the clinical evaluation tool allowing for comparison of pre and postmortem findings (Graber, 2005; Lundberg; Nemetz et al., 1989; Shojania et al.). Unfortunately, autopsy rates have fallen steadily since 1965. Rates averaged from 30-40% in the 1960s and by 1994 had decreased to less than 6%, while nursing home rates may be as low as 1 per 1000 (Lundberg; Shojania et al.). Autopsy rates have declined for several reasons: medical, legal, social, and economic. Specifically age at death, physical location of death, cost of autopsy, the influence of advancing mean age at death, and the increasing number of deaths occurring within the nursing home setting are factors influencing decline in autopsy rate (Nemetz et al.). Unfortunately, lacking autopsy, diagnostic errors many times remain hidden not only to families but also more importantly to practitioners who would benefit from lessons learned.

Shojania et al. (2003) conducted a systematic review of the literature targeting 53 studies of consecutive autopsy series or random samples from such series from 1966 to 2002. For purposes of the review a major error was defined as a clinically missed diagnosis involving a principal underlying disease or primary cause of death and class I errors were defined as major errors that, had they been detected in time, would have effected patient outcome, minimally resulting in the patient being discharged from the hospital alive. Synthesizing data, they found

the median error rates for major errors equal to 23.5% and class I errors equal to 9.0%. They noted a small but insignificant decrease in diagnostic errors over time. Authors concluded by estimating the current rate of major errors occurring at institutions within the U.S. falling between 8.4% to 24.4% and class one errors between 4.1% to 6.7%. Alarming, these rates indicate that one in twenty patients may experience a fatal diagnostic error.

Are Diagnostic Errors Addressed?

Clinical diagnosis is primarily a cognitive process, making it difficult to analyze via such methods as root cause analysis. Generating hypotheses, followed by the weighing and prioritizing of them are the two main steps identified in diagnostic assessment (Schiff et al, 2005). Also, diagnosis is subject to influence by the affective state of the clinician and even the decision maker may not be completely aware of how his or her diagnostic decision was reached. Metacognition, heuristics, and debiasing all contribute to the process of clinical decision-making (Graber, 2005; Graber, Franklin, & Gordon, 2005).

Schiff et al. (2005) notes that diagnostic errors are double to quadruple the number of medication errors as demonstrated by studies of malpractice claims, yet they represent an underemphasized and understudied area of patient-safety. According to Zhan and Miller (2003) only one of 93 safety projects funded by the Agency for Healthcare Research and Quality (AHRQ) is focused on diagnostic error and no AHRQ Patient Safety Indicators directly measure failure to diagnose. Graber (2005) identifies three factors in explaining why diagnostic errors go unnoticed. Diagnostic errors are difficult to identify, insurers and quality organizations have not focused on diagnostic error, and finally providers do not perceive their own diagnostic capabilities as subject to error. When faced with the statistics, practitioners typically do not recognize their own potential culpability, frequently believing the error rate to be reflective of

careless work by “other” providers and their own diagnoses to be correct. Why do practitioners not focus more attention on their own diagnostic capacities? Graber attributes this to two factors: inadequate calibration and overconfidence. Inadequate calibration refers to lack of follow-up information necessary to assess accuracy, including the unavailability of definitive diagnostic information, breakdown in the communication loop, and the previously discussed decreased autopsy rate, depriving clinicians of an important quality tool. Overconfidence is exemplified by the overestimation of one’s ability to function flawlessly. Studies in overconfidence by Kruger and Dunning (as cited by Graber) disconcertingly demonstrate that the least skilled are in fact the most overconfident. In a study by Friedman et al. (2005), experienced residents and internists reported a high degree of confidence 25% of the time, when, in actuality, they had reached an incorrect diagnosis.

Graber (2005) further postulates that physicians assign “working” diagnoses because they are uncomfortable with uncertainty and also because they experience pressure from patients. Dr. Bergsagel, a senior oncologist, (as quoted by Leonhardt, 2006) notes that once a clinical diagnostic path is taken, it is difficult to back off of. Graber and Mathews (2007) refer to this as ‘premature closure’, where further possibilities are not considered. This phenomenon is also referred to as anchoring, i.e., the tendency of decision makers to remain loyal to their current idea and resistant to change even when confronted by compelling external evidence (Friedman et al., 2005). While diagnostic errors are currently not addressed in a systematic manner, it is critical to take active steps not only to identify diagnostic errors, perhaps by reversing the trend in decreasing autopsy rates, but also to proactively take steps congruent with the prevention of diagnostic errors. This paper will examine the efficacy of ISABEL, a web-based diagnostic reminder tool, in providing diagnostic support.

The Case of Isabel Maude

In 1999, then three-year-old Isabel Maude was diagnosed with chicken pox. A few days after her initial diagnosis, Isabel developed additional symptoms including a high fever, vomiting, diarrhea, severe pain, and discoloration of her existing rash. Visits to both the family pediatrician and the local emergency department failed to diagnosis Necrotising Fasciitis, a life threatening bacterial infection, which, although occurring rarely, is a well-defined complication of chicken pox. When Isabel's condition deteriorated, her parents made a return visit to the ED. Minutes after their arrival, Isabel suffered cardiac arrest. After being stabilized, Isabel was transferred to a London hospital where she spent the next eight weeks, four of them on life support in the Pediatric Intensive Care Unit. Following Isabel's recovery, her father, Jason Maude, teamed with Dr. Joseph Britto, the attending physician who had cared for Isabel in the PICU to create a web-based clinical diagnosis decision support system and founded Isabel Healthcare (Isabel Healthcare, 2007c).

Design

ISABEL offers two main functions: the Isabel Diagnosis Reminder System (IDRS), designed to provide a list of possible diagnoses, causative drugs, and related diagnoses in response to a patient's significant clinical features; second is the Isabel Knowledge Mobilizing System (IKMS), allowing the immediate retrieval of information from textbooks, journals, and even supporting images (Isabel Healthcare, 2007b). This paper focuses on the diagnostic support component, IDRS.

ISABEL currently provides support in pediatrics, neonatology, internal medicine, surgery, obstetrics, gynecology, toxicology, and most recently, bioterrorism. Drop down menus prompt the user to quickly select an age group, gender, and geographic region. The user can

choose to search diagnoses by clinical specialty or by allowing a general search option. Developers recommend a simple list entry of pertinent clinical findings, however case history text may be pasted in or transferred via interface with an electronic medical record system (ISABEL Healthcare, 2007a,b).

After entering necessary data, a list of possible diagnoses, selected from a data-base of over 11,000 items, is presented sorted by body system. A maximum of thirty possible diagnoses may be presented, ten per page. Diagnoses are not ranked by order of importance or probability since developers believe that there will always be patient information, which has not been entered into the system, therefore clinicians retain the ultimate responsibility of clinical judgment and decision making. It is important to note that Isabel is designed to be a reminder system only, improving the process by providing possible diagnoses that may not have been considered. The clinician still needs to utilize his or her clinical skill and judgment to determine the actual diagnosis and best plan of care, including additional tests and or physical examination (Isabel Healthcare, 2007b).

Cost

Cost information is not available at Isabel Healthcare's website, however the CEO of the company, Dr. Joseph Britto, readily supplied me with the information via e-mail. Cost to the individual subscriber is \$2.00 per day. Hospitals pay 51 cents a day per hospital bed or \$3.50 per admission (J. Britto, personal communication, August 28, 2007).

Efficacy

A study of earlier computer based diagnostic systems, e.g., Dxplain, Iliad, Meditel, and QMR, by Berner et al. (1994) demonstrated limitations in sensitivity and specificity and also inefficiency, incurring too much time to input necessary data. ISABEL developers claim to have

overcome these inadequacies. ISABEL utilizes statistical natural language processing which enables software to recognize unstructured data, making the necessary data entry fast by alleviating the need for controlled vocabulary. The practitioner is only required to input a short list of pertinent clinical findings, which in turn generates a list of possible diagnoses (Isabel Healthcare, 2007b).

Ramnarayan, Tomlinson, Rao, Coren, Winrow, and Britto (2003) conducted an initial performance evaluation aimed at demonstrating the sensitivity of the tool. At the time of the study, late 2000, ISABEL was available free of charge on the internet. The Knowledge Mobilizing System component was not yet in existence and the system was only available for pediatrics. For purposes of the study, ISABEL was accurate if the expected or final diagnosis was included in the output list of possible diagnoses. In stage I of the study, ninety-nine hypothetical cases were used and in stage II, one hundred actual cases of children presenting to acute pediatric units were used. ISABEL provided the correct diagnosis in 91% of the hypothetical cases and 95% of the actual cases.

An Indian study entered the key clinical and laboratory findings from the records of 200 patients admitted to the pediatric intensive care unit into the ISABEL diagnostic tool. Researchers sought to determine the sensitivity of the tool as established by the presence of the final diagnosis in the result list generated by ISABEL. Data established a sensitivity of 80.5%, the most common diagnoses missed were septicemia, tuberculosis, tetanus, renal failure, and anemia. Researchers noted that the sensitivity demonstrated was lower than sensitivity findings from a previous study conducted in the UK. Authors suggest that sensitivity may suffer due to wrong words or misspellings entered by non-English speaking clinicians. Authors advice enhancing the program to reject wrongly spelled words or jargon that will not be recognized

(Bavdekar & Pawar, 2005). I would also postulate that the Indian study cohort contained a higher prevalence of the conditions that researchers indicated were likely to be missed by ISABEL, e.g., tuberculosis and tetanus, as compared to the UK cohort. Authors do note that during the period of their study, developers were in the process of implementing a change to ISABEL to allow entering the geographical location of the patient, which may result in improved sensitivity of the tool (Bavdekar & Pawar).

Briggs and Fitch (2005), in a study funded by Britain's Department of Health (DoH), surveyed registered users of ISABEL, via an e-mailed questionnaire, to determine the existing use of ISABEL and identify impediments to its wider use in the National Health Service (NHS). At the time of the survey, ISABEL was still free of charge and limited to pediatrics. Data gathered included type of clinician, computer equipment and internet access available at point of service, frequency of ISABEL consultation, and opinion of information provided by ISABEL. Responses showed that the vast majority of pediatricians, 72%, used ISABEL only occasionally, less than once per week; and only 5% used ISABEL on a daily basis. Data supported the conclusion that ISABEL has a small number of users who have a high opinion of the tool, however, even this group typically utilizes it less than once per week. Reasons cited for non-use included: time constraints, lack of access to a computer, and most interesting, clinicians' belief that they personally do not need diagnostic assistance, confirming Graber's (2005) observations of clinicians' attitudes concerning their own diagnostic ability and performance.

Ramnarayan et al. (2006) used simulated pediatric cases of various complexities to compare diagnoses made independently against those arrived at with the use of ISABEL. Fifty-two British clinicians with different levels of experience recorded their clinical decisions for twelve separate cases before and again after system consultation. Researchers used the change in

number of diagnostic errors of omission (DEO) as the primary indicator for measuring ISABEL's effectiveness. For purposes of the study, a DEO was defined as failure to include all clinically important diagnoses for the case, as previously determined by a panel of two expert pediatric consultants. Researchers also calculated additional time required for utilization of ISABEL. Median extra time required for the use of ISABEL was one minute, consistent with findings by Graber and Mathews (2007).

Numbers of DEOs are presented in a table, the mean DEO pre-ISABEL equal to 5.50 (per twelve cases) and post-ISABEL 4.98. Authors note a reduction of .52 DEOs per 12 cases. Results were statistically significant, $p < .001$ (Ramnarayan et al., 2006). I calculated the absolute rate reduction to be 4.3% and number needed to treat at 23.25. A study by Gilchrist et al. (2005) found that clinicians see an average of 20.1 patients per day. Integrating this figure with numbers generated from the study, ISABEL has the potential to prevent one DEO per day for a practicing pediatrician, if ISABEL is utilized for every patient.

I do have reservations concerning potential bias in this study, as well as the initial performance evaluation conducted in 2000. One of the researchers, Joseph F. Britto is CEO and co-founder of Isabel Healthcare Inc. Additionally, Amanda Tomlinson, is employed by and Dr. Ramnarayan consults part-time to Isabel Healthcare. All three hold stock options (Ramnarayan et al., 2006). While the primary objective of this study was to measure the change in number of DEOs following intervention with ISABEL in the diagnostic process, researchers also sought to obtain a secondary objective, a subtle measure of the change in quality of clinical decisions. For this they utilized a measurement tool developed in an earlier study by many of the same researchers, including Britto, Tomlinson, and Ramnarayan. From this earlier study, researchers

concluded that the tool they developed is both reliable and valid for assessing the value of the ISABEL system (Ramnarayan & Kapoor et al., 2003).

Adult disease conditions were added to the ISABEL database in January 2005. Subsequently an initial performance evaluation was undertaken for the use of ISABEL in adult emergency medicine. Researchers utilized data from 464 patients, mean age 49.4 years, presenting to emergency departments of three British teaching hospitals, to conduct this observational study. A research assistant entered patient's age and gender, from dropdown menus. Clinical information entered included positive clinical findings and test results, symptoms, and salient past illnesses. Two outcome measures were predetermined: the number of cases where the final discharge diagnosis for admitted patients was included in the list of possibilities generated by ISABEL. The second outcome measure was the number of cases in which ISABEL provided all clinically important diagnoses as established by an expert panel for purposes of the study (Ramnarayan et al, 2007).

Researchers concluded that all clinically important diagnoses were displayed in 92% of cases and the discharge diagnosis for those patients admitted was displayed 95% of the time (Ramnarayan et al., 2007). Again ISABEL demonstrated a high degree of sensitivity, consistent with findings in earlier studies. While this is a fairly large randomized study, again, I believe it is worth noting that three members of the research team, Ramnarayan, Cronje, and Britto, are employees of ISABEL Healthcare. (Ramnarayan et al., 2007).

Graber and Mathews (2007) conducted a simple study to evaluate the speed and accuracy of ISABEL. Like the initial evaluation performed by Ramnarayan and Tomlinson et al. (2003), the tool was determined to be accurate if the correct diagnosis was included in the generated list of possible diagnoses. Their methodology utilized fifty sequential case records from

Massachusetts General Hospital for patients over the age of ten. Data was entered into the ISABEL tool by two different methods: manually entering key clinical findings, and pasting in the entire case history. While both methods proved to be fast, typically requiring less than one minute per case, the recommended method of entering key findings yielded the best results generating correct diagnosis in 48 of 50 cases, yielding a sensitivity of 96%. The two diagnoses missed were not part of ISABEL's database at the time of the study. This is similar to earlier findings by Ramnarayan and Tomlinson et al. of 91% and 95% sensitivity for hypothetical and actual pediatric cases during the initial evaluation of ISABEL.

Available research proved to be limited. I did find two other sources of information. One was an article derived from the work of a Master's dissertation in the UK. The article proved to be a very abbreviated version, and from the information available the work appeared to have several serious flaws, as judged by this author. Another source found was information from a poster presentation of the impact of a web-based diagnostic reminder system, specifically ISABEL, presented at the American Medical Informatics Association 2006 annual conference. Again information concerning the methodology of deriving their conclusions was extremely limited making critical evaluation of evidence by this author impossible.

Conclusion

From the data available, it is evident that ISABEL demonstrates a high degree of sensitivity. Although research is limited at this time, three of four studies, including initial evaluations in both pediatrics and emergency medicine, as well as a general study in adult medicine, indicate that the ISABEL diagnostic reminder system offers the correct diagnosis between 91% and 96% of the time. While, the fourth study, conducted in India, found the sensitivity of the tool to be only 80.5%, I do believe there are several possible reasons for this

discrepancy, as discussed earlier, which may have been overcome by modifications and enhancements to the tool. Clearly, additional large RCTs need to be conducted with patient populations in various geographical locations. Distinct advantages of ISABEL over its predecessors are improved sensitivity, efficiency, typically requiring approximately one minute to enter key findings, and ease of use, i.e., no training is required.

Diagnostic error remains a significant threat to patient safety. Ten to thirty percent of medical errors are diagnostic in nature (Schiff et al., 2005), yet many go unnoticed due to lack of follow-up information necessary to assess accuracy, including declining autopsy rates (Lundberg, 1998; Shojania et al., 2003). Additionally, insurers and quality organizations have yet to focus their efforts on reducing diagnostic error (Graber, 2005; Schiff et al., 2005; Zhan & Miller, 2003). In my opinion, ISABEL is fast, accurate, and has the potential to reduce diagnostic error. While ISABEL does not provide the diagnostic answer, I believe being presented with a list of possible diagnoses for consideration is a valuable asset for opening the mind of busy clinicians to alternative and perhaps obscure possibilities. However, any tool is only effective if it is utilized. The main impediment to its use may be clinician's sense of infallibility in their own personal diagnostic abilities and their resistance to consider alternate possibilities once they have arrived at a conclusion, i.e. premature closure or anchoring (Briggs & Fitch, 2005; Friedman et al., 2005; Graber; Leonhardt, 2006). All clinicians need to own the problem of diagnostic error and proactively take steps to affect a reduction in the rate thereof.

From a Personal Standpoint

At age eight, during the spring of second grade, my daughter began to complain of stomach pain. In the beginning it was not severe, she was able to continue with the normal activities of an average eight-year-old, i.e., school time and backyard play. I reported her distress

to our pediatrician, an individual, highly respected by both the community and myself. Over time there were several visits and phone consultations, always with the same complaint, stomach pain. After ascertaining that there were no other significant symptoms, e.g., diarrhea, constipation, or vomiting, her pediatrician felt there was no cause for concern. Life went on and the frequency of her stomach complaints seemed to increase. There were several more visits and phone calls and emotional causation seemed to emerge as a working diagnosis, i.e. her stomach pains were caused by emotional distress, very common in young children. Since she had a particularly grouchy second grade teacher, the diagnosis seemed plausible.

Second grade ended and the summer began, life should have improved; yet it did not. Additional visits brought no answers. When third grade began, her health status declined rapidly in terms of severity and frequency of her symptoms; I requested a referral to Boston Children's Hospital. While granting me a referral, the pediatrician reassured me that she had no physical problems. After eight weeks at a world-class institution, with invasive testing, no causative factor could be established. A gastroenterologist pronounced nothing was wrong and she would need to "learn to live with her pain", perhaps by attending a pain clinic. That very same day, a simple low cost, non-invasive ultrasound revealed an ovarian tumor. Surgery the next day removed a germ cell tumor, specifically dysgerminoma.

Would a clinician have any reason to suspect a germ cell tumor in an 8 year-old girl as a likely diagnosis? Absolutely not, remarkably the incidence of malignant germ cell tumors in a female during their first decade of life is only about 1 in 1,000,000 per year (Smith et al., 2006). There was no fault with the clinicians involved, it was simply too rare of an occurrence to enter into the cognitive process as a possibility.

I signed up for a trial membership, with no training it took me less than thirty seconds to enter the age, gender, geographical location, and pertinent clinical findings, in this case, “stomach pain” was the only item entered. In the first test, I used “general” as the specialty and the possible diagnoses did not include the correct diagnosis, i.e., germ cell tumor. Even when the heading of “neoplastic diseases” was expanded, the possible diagnoses given were, leukemia, liver neoplasms, and familial polyposis syndrome. However, in a second test, I repeated the age, gender, geographical location, the same clinical finding, but entered “oncology” as the specialty. Germ cell tumor, the correct diagnosis, was one of six possible diagnoses presented.

References

- Bavdekar, S. B., & Pawar, M. (2005). Evaluation of an internet-delivered pediatric diagnosis support system (ISABEL) in a tertiary care center in India. *Indian Pediatrics, 42*(11), 1086-1091.
- Berner, E. S., Webster, G. D., Shugerman, A. A., Jackson, J. R., Algina, J., Baker, A. L. et al., (1994). Performance of four computer-based diagnostic systems. *New England Journal of Medicine, 330*(25), 1792-1806.
- Briggs, J. S., & Fitch, C. J. (2005). The ISABEL user survey. *Medical Informatics and the Internet in Medicine, 30*(2), 167-172.
- Friedman, C. P., Guido, G., Gatti, M. S., Franz, T. M., Murphy, G. C., Wolf, F. M., et al. (2005). Do physicians know when their diagnoses are correct? *Journal of General Internal Medicine, 20*(4), 334-339.
- Gilchrist, V., McCord, G., Schrop, S. L., King, B. D., McCormick, K. F., Oprandi, A. M., et al. (2005). Physician activities during time out of the examination room. *Annals of Family Medicine, 3*, 494-499.
- Graber, M. (2005). Diagnostic errors in medicine: A case of neglect. *Journal on Quality and Patient Safety, 31*(2), 106- 113.
- Graber, M. L., Granklin, N., & Gordon, R. (2005). Diagnostic error in internal medicine. *Archives of Internal Medicine, 165*(13), 1493-1499.
- Graber, M. L., & Mathews, A. (2007). *More on googling for a diagnosis: Performance of a web-based clinical diagnosis support system for internists*. Manuscript submitted for publication.
- Isabel Healthcare. (2007a). Demo. Retrieved from www.isablehealthcare.com/

- Isabel Healthcare. (2007b). FAQs. Retrieved from www.isabelhealthcare.com/info/faqs.jsp
- Isabel Healthcare. (2007c). Isabel's story. Retrieved from www.isabelhealthcare.com/info/Isabel_Story.doc
- Kirch, W., & Schafii, C. (1996). Misdiagnosis at a university hospital in 4 medical eras: Report on 400 cases. *Medicine*, 75(1), 29-40.
- Leonhardt, D. (2006, February 22). Why doctors so often get it wrong. *The New York Times*.
- Lundberg, G. (1998). Low-tech autopsies in the era of high-tech medicine. *Journal of the American Medical Association*, 280(14), 1273-1274.
- Nemetz, P. N., Ballard, D. J., Beard, C. M., Ludwig, J., Tangalos, E. G., Kokmen, E., et al. (1989). An anatomy of the autopsy, Olmsted County, 1935 through 1985. *Mayo Clinic Proceedings*, 64(9), 1055-1064.
- Nichols, L., Aronica, P., & Babe, C. (1998). Are autopsies obsolete? *American Journal of Clinical Pathology*, 110(2), 210-218.
- Ramnarayan, P., Cronje, N., Brown, R., Negus, R., Coode, B., Moss, P., et al. (2007). *Validation of a diagnostic reminder system in emergency medicine: A multi-centre study*. Manuscript submitted for publication.
- Ramnarayan, P., Kapoor, R. R., Coren, M., Nanduri, V., Tomlinson, A. L., Taylor, P. M., et al. (2003). Measuring the impact of diagnostic decision support on the quality of clinical decision making: Development of a reliable and valid composite score. *Journal of the American Medical Informatics Association*, 10(6), 563-572.
- Ramnarayan, P., Roberts, G. C., Coren, M., Nanduri, V., Tomlinson, A. L., Taylor, P. M.,

- et al. (2006). Assessment of the potential impact of a reminder system on the reduction of diagnostic errors: A quasi-experimental study. *BMC Medical Informatics and Decision Making*, 6, 22-27.
- Ramnarayan, P., Tomlinson, A., Rao, A., Coren, M., Winrow, A., & Britto, J. (2003). ISABEL: A web-based differential diagnostic aid for paediatrics: Results from an initial performance evaluation. *Archives of Disease in Childhood*, 88(5), 408-413.
- Schiff, G. D., Seijeoung, K., Abrams, R., Cosby, K., Lambert, B., Elstein, A. S., et al. (2005). Diagnosing diagnosis errors: Lessons from a multi-institutional collaborative project. Agency for Healthcare Research and Quality. Retrieved from <http://www.ahrq.gov/downloads/pub/advances/vol2/Schiff.pdf>
- Shojania, K. G., Burton, E. C., McDonald, K. M., & Goldman, L. (2003). Changes in rates of autopsy-detected diagnostic errors over time: A systematic review. *Journal of the American Medical Association*, 289(21), 2849-2856.
- Smith, H. O., Berwick, M., Verschraegen, C. F., Wiggins, C., Lansing, L., Muller, C. Y. et al. (2006). Incidence and survival rates for female malignant germ cell tumors. *Obstetrics and Gynecology*, 107(5), 1075-1085.
- Zhan, C., & Miller, M. R. (2003). Administrative data based patient safety research: A critical review. *Quality and Safety in Health Care*, 12(Suppl 2), ii58-ii63.